Lecture Notes on Linear Algebra

A K Lal S Pati

February 10, 2015

Contents

Chapter 1

Introduction to Matrices

1.1 Definition of a Matrix

Definition 1.1.1 (Matrix)**.** *A rectangular array of numbers is called a matrix.*

The horizontal arrays of a matrix are called its ROWS and the vertical arrays are called its COLUMNS. A matrix is said to have the ORDER $m \times n$ if it has m rows and n columns. An $m \times n$ matrix A can be represented in either of the following forms:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{or} \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

where a_{ij} is the entry at the intersection of the i^{th} row and j^{th} column. In a more concise manner, we also write $A_{m\times n} = [a_{ij}]$ or $A = [a_{ij}]_{m\times n}$ or $A = [a_{ij}]$. We shall mostly be concerned with matrices having real numbers, denoted $\mathbb{R}$, as entries. For example, if $A = \begin{bmatrix} 1 & 3 & 7 \\ 4 & 5 & 6 \end{bmatrix}$ then $a_{11} = 1,\ a_{12} = 3,\ a_{13} = 7,\ a_{21} = 4,\ a_{22} = 5,$ and $a_{23} = 6.$

A matrix having only one column is called a COLUMN VECTOR; and a matrix with only one row is called a ROW VECTOR. WHENEVER A VECTOR IS USED, IT SHOULD BE UNDERSTOOD FROM THE CONTEXT WHETHER IT IS A ROW VECTOR OR A COLUMN VECTOR. ALSO, ALL THE VECTORS WILL BE REPRESENTED BY BOLD LETTERS.

Definition 1.1.2 (Equality of two Matrices)**.** *Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ having the same order $m \times n$ are equal if $a_{ij} = b_{ij}$ for each $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.*

In other words, two matrices are said to be equal if they have the same order and their corresponding entries are equal.

Example 1.1.3. *The linear system of equations $2x + 3y = 5$ and $3x + 2y = 5$ can be identified with the matrix $\begin{bmatrix} 2 & 3 & : & 5 \\ 3 & 2 & : & 5 \end{bmatrix}$. Note that x and y are indeterminate and we can think of x being associated with the first column and y being associated with the second column.*

1.1.1 Special Matrices

Definition 1.1.4. 1. *A matrix in which each entry is zero is called a zero-matrix, denoted by* $\mathbf{0}$. *For example,*

$$\mathbf{0}_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \mathbf{0}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

2. *A matrix that has the same number of rows as the number of columns, is called a square matrix. A square matrix is said to have order* n *if it is an* $n \times n$ *matrix.*

3. *The entries* $a_{11}, a_{22}, \ldots, a_{nn}$ *of an* $n \times n$ *square matrix* $A = [a_{ij}]$ *are called the diagonal entries (the principal diagonal) of* A.

4. *A square matrix* $A = [a_{ij}]$ *is said to be a diagonal matrix if* $a_{ij} = 0$ *for* $i \neq j$. *In other words, the non-zero entries appear only on the principal diagonal. For example, the zero matrix* $\mathbf{0}_n$ *and* $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ *are a few diagonal matrices.*

   A diagonal matrix D *of order* n *with the diagonal entries* $d_1, d_2, \ldots, d_n$ *is denoted by* $D = diag(d_1, \ldots, d_n)$. *If* $d_i = d$ *for all* $i = 1, 2, \ldots, n$ *then the diagonal matrix* D *is called a* **scalar matrix**.

5. *A scalar matrix* A *of order* n *is called an* IDENTITY MATRIX *if* $d = 1$. *This matrix is denoted by* I_n.

   For example, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *and* $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. *The subscript* n *is suppressed in case the order is clear from the context or if no confusion arises.*

6. *A square matrix* $A = [a_{ij}]$ *is said to be an* UPPER TRIANGULAR *matrix if* $a_{ij} = 0$ *for* $i > j$.

   A square matrix $A = [a_{ij}]$ *is said to be a* LOWER TRIANGULAR *matrix if* $a_{ij} = 0$ *for* $i < j$.

   A square matrix A *is said to be* TRIANGULAR *if it is an upper or a lower triangular matrix.*

   For example, $\begin{bmatrix} 0 & 1 & 4 \\ 0 & 3 & -1 \\ 0 & 0 & -2 \end{bmatrix}$ is upper triangular, $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ is lower triangular.

Exercise 1.1.5. *Are the following matrices upper triangular, lower triangular or both?*

1. $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$

2. *The square matrices* $\mathbf{0}$ *and* I *or order* n.

3. *The matrix* $diag(1, -1, 0, 1)$.

1.2 Operations on Matrices

Definition 1.2.1 (Transpose of a Matrix). *The transpose of an* $m \times n$ *matrix* $A = [a_{ij}]$ *is defined as the* $n \times m$ *matrix* $B = [b_{ij}]$, *with* $b_{ij} = a_{ji}$ *for* $1 \le i \le m$ *and* $1 \le j \le n$. *The transpose of* A *is denoted by* A^t.

That is, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ then $A^t = \begin{bmatrix} 1 & 0 \\ 4 & 1 \\ 5 & 2 \end{bmatrix}$. Thus, the transpose of a row vector is a column vector and vice-versa.

Theorem 1.2.2. *For any matrix* A, $\quad (A^t)^t = A$.

Proof. Let $A = [a_{ij}]$, $A^t = [b_{ij}]$ and $(A^t)^t = [c_{ij}]$. Then, the definition of transpose gives

$$c_{ij} = b_{ji} = a_{ij} \quad \text{for all} \quad i, j$$

and the result follows. □

Definition 1.2.3 (Addition of Matrices). *let* $A = [a_{ij}]$ *and* $B = [b_{ij}]$ *be two* $m \times n$ *matrices. Then the sum* $A + B$ *is defined to be the matrix* $C = [c_{ij}]$ *with* $c_{ij} = a_{ij} + b_{ij}$.

Note that, we define the sum of two matrices only when the order of the two matrices are same.

Definition 1.2.4 (Multiplying a Scalar to a Matrix). *Let* $A = [a_{ij}]$ *be an* $m \times n$ *matrix. Then for any element* $k \in \mathbb{R}$, *we define* $kA = [ka_{ij}]$.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ and $k = 5$, then $5A = \begin{bmatrix} 5 & 20 & 25 \\ 0 & 5 & 10 \end{bmatrix}$.

Theorem 1.2.5. *Let* A, B *and* C *be matrices of order* $m \times n$, *and let* $k, \ell \in \mathbb{R}$. *Then*

1. $A + B = B + A$ *(commutativity).*

2. $(A + B) + C = A + (B + C)$ *(associativity).*

3. $k(\ell A) = (k\ell)A$.

4. $(k + \ell)A = kA + \ell A$.

Proof. Part 1.
Let $A = [a_{ij}]$ and $B = [b_{ij}]$. Then

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = [b_{ij}] + [a_{ij}] = B + A$$

as real numbers commute.

The reader is required to prove the other parts as all the results follow from the properties of real numbers. □

Definition 1.2.6 (Additive Inverse)**.** *Let A be an $m \times n$ matrix.*

1. *Then there exists a matrix B with $A + B = \mathbf{0}$. This matrix B is called the additive inverse of A, and is denoted by $-A = (-1)A$.*

2. *Also, for the matrix $\mathbf{0}_{m\times n}$, $A + \mathbf{0} = \mathbf{0} + A = A$. Hence, the matrix $\mathbf{0}_{m\times n}$ is called the additive identity.*

Exercise 1.2.7.

1. *Find a 3×3 non-zero matrix A satisfying $A = A^t$.*

2. *Find a 3×3 non-zero matrix A such that $A^t = -A$.*

3. *Find the 3×3 matrix $A = [a_{ij}]$ satisfying $a_{ij} = 1$ if $i \neq j$ and 2 otherwise.*

4. *Find the 3×3 matrix $A = [a_{ij}]$ satisfying $a_{ij} = 1$ if $|i - j| \leq 1$ and 0 otherwise.*

5. *Find the 4×4 matrix $A = [a_{ij}]$ satisfying $a_{ij} = i + j$.*

6. *Find the 4×4 matrix $A = [a_{ij}]$ satisfying $a_{ij} = 2^{i+j}$.*

7. *Suppose $A + B = A$. Then show that $B = \mathbf{0}$.*

8. *Suppose $A + B = \mathbf{0}$. Then show that $B = (-1)A = [-a_{ij}]$.*

9. *Let $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 1 & 2 \end{bmatrix}$. Compute $A + B^t$ and $B + A^t$.*

1.2.1 Multiplication of Matrices

Definition 1.2.8 (Matrix Multiplication / Product)**.** *Let $A = [a_{ij}]$ be an $m \times n$ matrix and $B = [b_{ij}]$ be an $n \times r$ matrix. The product AB is a matrix $C = [c_{ij}]$ of order $m \times r$, with*

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

That is, if $A_{m\times n} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$ and $B_{n\times r} = \begin{bmatrix} \vdots & b_{1j} & \vdots \\ \vdots & b_{2j} & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & b_{mj} & \vdots \end{bmatrix}$ then

$$AB = [(AB)_{ij}]_{m\times r} \text{ and } (AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Observe that the product AB is defined if and only if
THE NUMBER OF COLUMNS OF $A =$ THE NUMBER OF ROWS OF B.

For example, if $A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$ and $B = \begin{bmatrix} \alpha & \beta & \gamma & \delta \\ x & y & z & t \\ u & v & w & s \end{bmatrix}$ then

$$AB = \begin{bmatrix} a\alpha + bx + cu & a\beta + by + cv & a\gamma + bz + cw & a\delta + bt + cs \\ d\alpha + ex + fu & d\beta + ey + fv & d\gamma + ez + fw & d\delta + et + fs \end{bmatrix}. \tag{1.2.1}$$

Observe that in Equation (1.2.1), the first row of AB can be re-written as

$$a \cdot \begin{bmatrix} \alpha & \beta & \gamma & \delta \end{bmatrix} + b \cdot \begin{bmatrix} x & y & z & t \end{bmatrix} + c \cdot \begin{bmatrix} u & v & w & s \end{bmatrix}.$$

That is, if $\mathrm{Row}_i(B)$ denotes the i-th row of B for $1 \le i \le 3$, then the matrix product AB can be re-written as

$$AB = \begin{bmatrix} a \cdot \ \mathrm{Row}_1(B) + b \cdot \ \mathrm{Row}_2(B) + c \cdot \ \mathrm{Row}_3(B) \\ d \cdot \ \mathrm{Row}_1(B) + e \cdot \ \mathrm{Row}_2(B) + f \cdot \ \mathrm{Row}_3(B) \end{bmatrix}. \tag{1.2.2}$$

Similarly, observe that if $\mathrm{Col}_j(A)$ denotes the j-th column of A for $1 \le j \le 3$, then the matrix product AB can be re-written as

$$\begin{aligned} AB = \Big[\ & \mathrm{Col}_1(A) \cdot \alpha + \mathrm{Col}_2(A) \cdot x + \mathrm{Col}_3(A) \cdot u, \\ & \mathrm{Col}_1(A) \cdot \beta + \mathrm{Col}_2(A) \cdot y + \mathrm{Col}_3(A) \cdot v, \\ & \mathrm{Col}_1(A) \cdot \gamma + \mathrm{Col}_2(A) \cdot z + \mathrm{Col}_3(A) \cdot w \\ & \mathrm{Col}_1(A) \cdot \delta + \mathrm{Col}_2(A) \cdot t + \mathrm{Col}_3(A) \cdot s \big]. \end{aligned} \tag{1.2.3}$$

Remark 1.2.9. *Observe the following:*

1. *In this example, while AB is defined, the product BA is not defined.*

   However, for square matrices A and B of the same order, both the product AB and BA are defined.

2. *The product AB corresponds to operating on the* rows of the matrix B *(see Equation (1.2.2)). This is* ROW METHOD *for calculating the matrix product.*

3. *The product AB also corresponds to operating on the* columns of the matrix A *(see Equation (1.2.3)). This is* COLUMN METHOD *for calculating the matrix product.*

4. *Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices. Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the rows of A and $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ are the columns of B. If the product AB is defined, then check that*

$$AB = [A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p] = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_n B \end{bmatrix}.$$

Example 1.2.10. *Let* $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. *Use the row/column method of matrix multiplication to*

1. *find the second row of the matrix* AB.
   Solution: *Observe that the second row of* AB *is obtained by multiplying the second row of* A *with* B. *Hence, the second row of* AB *is*

$$1 \cdot [1, 0, -1] + 0 \cdot [0, 0, 1] + 1 \cdot [0, -1, 1] = [1, -1, 0].$$

2. *find the third column of the matrix* AB.
   Solution: *Observe that the third column of* AB *is obtained by multiplying* A *with the third column of* B. *Hence, the third column of* AB *is*

$$-1 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Definition 1.2.11 (Commutativity of Matrix Product)**.** *Two square matrices* A *and* B *are said to commute if* $AB = BA$.

Remark 1.2.12. *Note that if* A *is a square matrix of order* n *and if* B *is a scalar matrix of order* n *then* $AB = BA$. *In general, the matrix product is not commutative. For example, consider* $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. *Then check that the matrix product*

$$AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = BA.$$

Theorem 1.2.13. *Suppose that the matrices* A, B *and* C *are so chosen that the matrix multiplications are defined.*

1. *Then* $(AB)C = A(BC)$. *That is, the matrix multiplication is associative.*

2. *For any* $k \in \mathbb{R}$, $(kA)B = k(AB) = A(kB)$.

3. *Then* $A(B + C) = AB + AC$. *That is, multiplication distributes over addition.*

4. *If* A *is an* $n \times n$ *matrix then* $AI_n = I_nA = A$.

5. *For any square matrix* A *of order* n *and* $D = diag(d_1, d_2, \ldots, d_n)$, *we have*

   - *the first row of* DA *is* d_1 *times the first row of* A;
   - *for* $1 \le i \le n$, *the* i^{th} *row of* DA *is* d_i *times the* i^{th} *row of* A.

   A similar statement holds for the columns of A *when* A *is multiplied on the right by* D.

Proof. Part 1. Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{n\times p}$ and $C = [c_{ij}]_{p\times q}$. Then

$$(BC)_{kj} = \sum_{\ell=1}^{p} b_{k\ell}c_{\ell j} \quad \text{and} \quad (AB)_{i\ell} = \sum_{k=1}^{n} a_{ik}b_{k\ell}.$$

Therefore,

$$\begin{aligned}
\big(A(BC)\big)_{ij} &= \sum_{k=1}^{n} a_{ik}\big(BC\big)_{kj} = \sum_{k=1}^{n} a_{ik}\big(\sum_{\ell=1}^{p} b_{k\ell}c_{\ell j}\big) \\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{p} a_{ik}\big(b_{k\ell}c_{\ell j}\big) = \sum_{k=1}^{n}\sum_{\ell=1}^{p}\big(a_{ik}b_{k\ell}\big)c_{\ell j} \\
&= \sum_{\ell=1}^{p}\big(\sum_{k=1}^{n} a_{ik}b_{k\ell}\big)c_{\ell j} = \sum_{\ell=1}^{t}\big(AB\big)_{i\ell}c_{\ell j} \\
&= \big((AB)C\big)_{ij}.
\end{aligned}$$

Part 5. For all $j = 1, 2, \ldots, n$, we have

$$(DA)_{ij} = \sum_{k=1}^{n} d_{ik}a_{kj} = d_i a_{ij}$$

as $d_{ik} = 0$ whenever $i \neq k$. Hence, the required result follows.

The reader is required to prove the other parts. □

Exercise 1.2.14. *1. Find a 2×2 non-zero matrix A satisfying $A^2 = \mathbf{0}$.*

2. *Find a 2×2 non-zero matrix A satisfying $A^2 = A$ and $A \neq I_2$.*

3. *Find 2×2 non-zero matrices A, B and C satisfying $AB = AC$ but $B \neq C$. That is, the cancelation law doesn't hold.*

4. *Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Compute $A + 3A^2 - A^3$ and $aA^3 + bA + cA^2$.*

5. *Let A and B be two matrices. If the matrix addition $A + B$ is defined, then prove that $(A+B)^t = A^t + B^t$. Also, if the matrix product AB is defined then prove that $(AB)^t = B^tA^t$.*

6. *Let $A = [a_1, a_2, \ldots, a_n]$ and $B^t = [b_1, b_2, \ldots, b_n]$. Then check that order of AB is 1×1, whereas BA has order $n\times n$. Determine the matrix products AB and BA.*

7. *Let A and B be two matrices such that the matrix product AB is defined.*

   (a) *If the first row of A consists entirely of zeros, prove that the first row of AB also consists entirely of zeros.*

   (b) *If the first column of B consists entirely of zeros, prove that the first column of AB also consists entirely of zeros.*

(c) *If A has two identical rows then the corresponding rows of AB are also identical.*

(d) *If B has two identical columns then the corresponding columns of AB are also identical.*

8. *Let* $A = \begin{bmatrix} 1 & 1 & -2 \\ 1 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$. *Use the row/column method of matrix multiplication to compute the*

   (a) *first row of the matrix AB.*

   (b) *third row of the matrix AB.*

   (c) *first column of the matrix AB.*

   (d) *second column of the matrix AB.*

   (e) *first column of B^tA^t.*

   (f) *third column of B^tA^t.*

   (g) *first row of B^tA^t.*

   (h) *second row of B^tA^t.*

9. *Let A and B be the matrices given in Exercise 1.2.14.8. Compute $A - A^t$, $(3AB)^t - 4B^tA$ and $3A - 2A^t$.*

10. *Let n be a positive integer. Compute A^n for the following matrices:*

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

    Can you guess a formula for A^n and prove it by induction?

11. *Construct the matrices A and B satisfying the following statements.*

    (a) *The matrix product AB is defined but BA is not defined.*

    (b) *The matrix products AB and BA are defined but they have different orders.*

    (c) *The matrix products AB and BA are defined and they have the same order but $AB \neq BA$.*

12. *Let A be a 3×3 matrix satisfying* $A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a+b \\ b-c \\ 0 \end{bmatrix}$. *Determine the matrix A.*

13. *Let A be a 2×2 matrix satisfying* $A \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \cdot b \\ a \end{bmatrix}$. *Can you construct the matrix A satisfying the above? Why!*

1.2.2 Inverse of a Matrix

Definition 1.2.15 (Inverse of a Matrix)**.** *Let A be a square matrix of order n.*

1. *A square matrix B is said to be a* LEFT INVERSE *of A if $BA = I_n$.*
2. *A square matrix C is called a* RIGHT INVERSE *of A, if $AC = I_n$.*
3. *A matrix A is said to be* INVERTIBLE *(or is said to have an* INVERSE*) if there exists a matrix B such that $AB = BA = I_n$.*

Lemma 1.2.16. *Let A be an $n \times n$ matrix. Suppose that there exist $n \times n$ matrices B and C such that $AB = I_n$ and $CA = I_n$, then $B = C$.*

Proof. Note that
$$C = CI_n = C(AB) = (CA)B = I_nB = B.$$
$\square$

Remark 1.2.17. 1. *From the above lemma, we observe that if a matrix A is invertible, then the inverse is unique.*

2. *As the inverse of a matrix A is unique, we denote it by A^{-1}. That is, $AA^{-1} = A^{-1}A = I$.*

Example 1.2.18. 1. *Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.*

   (a) *If $ad - bc \neq 0$. Then verify that $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.*

   (b) *If $ad-bc = 0$ then prove that either $[a\ b] = \alpha[c\ d]$ for some $\alpha \in \mathbb{R}$ or $[a\ c] = \beta[b\ d]$ for some $\beta \in \mathbb{R}$. Hence, prove that A is not invertible.*

   (c) *In particular, the inverse of $\begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix}$ equals $\frac{1}{2}\begin{bmatrix} 7 & -3 \\ -4 & 2 \end{bmatrix}$. Also, the matrices $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 4 & 0 \end{bmatrix}$ and $\begin{bmatrix} 4 & 2 \\ 6 & 3 \end{bmatrix}$ do not have inverses.*

2. *Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix}$. Then $A^{-1} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$.*

Theorem 1.2.19. *Let A and B be two matrices with inverses A^{-1} and B^{-1}, respectively. Then*

1. *$(A^{-1})^{-1} = A$.*
2. *$(AB)^{-1} = B^{-1}A^{-1}$.*
3. *$(A^t)^{-1} = (A^{-1})^t$.*

Proof. Proof of Part 1.
By definition $AA^{-1} = A^{-1}A = I$. Hence, if we denote A^{-1} by B, then we get $AB = BA = I$. Thus, the definition, implies $B^{-1} = A$, or equivalently $(A^{-1})^{-1} = A$.

Proof of Part 2.
Verify that $(AB)(B^{-1}A^{-1}) = I = (B^{-1}A^{-1})(AB)$.

Proof of Part 3.
We know $AA^{-1} = A^{-1}A = I$. Taking transpose, we get

$$(AA^{-1})^t = (A^{-1}A)^t = I^t \iff (A^{-1})^t A^t = A^t (A^{-1})^t = I.$$

Hence, by definition $(A^t)^{-1} = (A^{-1})^t$. □

We will again come back to the study of invertible matrices in Sections 2.2 and 2.5.

Exercise 1.2.20. *1. Let A be an invertible matrix and let r be a positive integer. Prove that $(A^{-1})^r = A^{-r}$.*

2. *Find the inverse of* $\begin{bmatrix} -\cos(\theta) & \sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ *and* $\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$.

3. *Let $A_1, A_2, \ldots, A_r$ be invertible matrices. Prove that the product $A_1 A_2 \cdots A_r$ is also an invertible matrix.*

4. *Let $\mathbf{x}^t = [1, 2, 3]$ and $\mathbf{y}^t = [2, -1, 4]$. Prove that $\mathbf{x}\mathbf{y}^t$ is not invertible even though $\mathbf{x}^t\mathbf{y}$ is invertible.*

5. *Let A be an $n \times n$ invertible matrix. Then prove that*

   (a) *A cannot have a row or column consisting entirely of zeros.*

   (b) *any two rows of A cannot be equal.*

   (c) *any two columns of A cannot be equal.*

   (d) *the third row of A cannot be equal to the sum of the first two rows, whenever $n \geq 3$.*

   (e) *the third column of A cannot be equal to the first column minus the second column, whenever $n \geq 3$.*

6. *Suppose A is a 2×2 matrix satisfying $(I + 3A)^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Determine the matrix A.*

7. *Let A be a 3×3 matrix such that $(I - A)^{-1} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$. Determine the matrix A [Hint: See Example 1.2.18.2 and Theorem 1.2.19.1].*

8. *Let A be a square matrix satisfying $A^3 + A - 2I = \mathbf{0}$. Prove that $A^{-1} = \frac{1}{2}\left(A^2 + I\right)$.*

9. *Let $A = [a_{ij}]$ be an invertible matrix and let p be a nonzero real number. Then determine the inverse of the matrix $B = [p^{i-j} a_{ij}]$.*

1.3 Some More Special Matrices

Definition 1.3.1. *1. A matrix A over $\mathbb{R}$ is called symmetric if $A^t = A$ and skew-symmetric if $A^t = -A$.*

2. A matrix A is said to be orthogonal if $AA^t = A^tA = I$.

Example 1.3.2. *1. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -1 \\ 3 & -1 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ -2 & 3 & 0 \end{bmatrix}$. Then A is a symmetric matrix and B is a skew-symmetric matrix.*

2. Let $A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}$. Then A is an orthogonal matrix.

3. Let $A = [a_{ij}]$ be an $n \times n$ matrix with a_{ij} equal to 1 if $i - j = 1$ and 0, otherwise. Then $A^n = \mathbf{0}$ and $A^\ell \neq \mathbf{0}$ for $1 \leq \ell \leq n-1$. The matrices A for which a positive integer k exists such that $A^k = \mathbf{0}$ are called NILPOTENT *matrices. The least positive integer k for which $A^k = \mathbf{0}$ is called the* ORDER OF NILPOTENCY.

4. Let $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. Then $A^2 = A$. The matrices that satisfy the condition that $A^2 = A$ are called IDEMPOTENT *matrices.*

Exercise 1.3.3. *1. Let A be a real square matrix. Then $S = \frac{1}{2}(A + A^t)$ is symmetric, $T = \frac{1}{2}(A - A^t)$ is skew-symmetric, and $A = S + T$.*

2. Show that the product of two lower triangular matrices is a lower triangular matrix. A similar statement holds for upper triangular matrices.

3. Let A and B be symmetric matrices. Show that AB is symmetric if and only if $AB = BA$.

4. Show that the diagonal entries of a skew-symmetric matrix are zero.

5. Let A, B be skew-symmetric matrices with $AB = BA$. Is the matrix AB symmetric or skew-symmetric?

6. Let A be a symmetric matrix of order n with $A^2 = \mathbf{0}$. Is it necessarily true that $A = \mathbf{0}$?

7. Let A be a nilpotent matrix. Prove that there exists a matrix B such that $B(I + A) = I = (I + A)B$ [Hint: If $A^k = \mathbf{0}$ then look at $I - A + A^2 - \cdots + (-1)^{k-1}A^{k-1}$].

1.3.1 Submatrix of a Matrix

Definition 1.3.4. *A matrix obtained by deleting some of the rows and/or columns of a matrix is said to be a submatrix of the given matrix.*

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$, a few submatrices of A are

$$[1], [2], \begin{bmatrix} 1 \\ 0 \end{bmatrix}, [1 \ 5], \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix}, \ A.$$

But the matrices $\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 4 \\ 0 & 2 \end{bmatrix}$ are not submatrices of A. (The reader is advised to give reasons.)

Let A be an $n \times m$ matrix and B be an $m \times p$ matrix. Suppose $r < m$. Then, we can decompose the matrices A and B as $A = [P \ Q]$ and $B = \begin{bmatrix} H \\ K \end{bmatrix}$; where P has order $n \times r$ and H has order $r \times p$. That is, the matrices P and Q are submatrices of A and P consists of the first r columns of A and Q consists of the last $m - r$ columns of A. Similarly, H and K are submatrices of B and H consists of the first r rows of B and K consists of the last $m - r$ rows of B. We now prove the following important theorem.

Theorem 1.3.5. *Let $A = [a_{ij}] = [P \ Q]$ and $B = [b_{ij}] = \begin{bmatrix} H \\ K \end{bmatrix}$ be defined as above. Then*

$$AB = PH + QK.$$

Proof. First note that the matrices PH and QK are each of order $n \times p$. The matrix products PH and QK are valid as the order of the matrices P, H, Q and K are respectively, $n \times r$, $r \times p$, $n \times (m - r)$ and $(m - r) \times p$. Let $P = [P_{ij}]$, $Q = [Q_{ij}]$, $H = [H_{ij}]$, and $K = [k_{ij}]$. Then, for $1 \leq i \leq n$ and $1 \leq j \leq p$, we have

$$\begin{aligned} (AB)_{ij} &= \sum_{k=1}^{m} a_{ik} b_{kj} = \sum_{k=1}^{r} a_{ik} b_{kj} + \sum_{k=r+1}^{m} a_{ik} b_{kj} \\ &= \sum_{k=1}^{r} P_{ik} H_{kj} + \sum_{k=r+1}^{m} Q_{ik} K_{kj} \\ &= (PH)_{ij} + (QK)_{ij} = (PH + QK)_{ij}. \end{aligned}$$

□

Remark 1.3.6. *Theorem 1.3.5 is very useful due to the following reasons:*

1. *The order of the matrices P, Q, H and K are smaller than that of A or B.*
2. *It may be possible to block the matrix in such a way that a few blocks are either identity matrices or zero matrices. In this case, it may be easy to handle the matrix product using the block form.*

3. *Or when we want to prove results using induction, then we may assume the result for $r \times r$ submatrices and then look for $(r+1) \times (r+1)$ submatrices, etc.*

For example, if $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$, Then

$$AB = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} [e\ f] = \begin{bmatrix} a+2c & b+2d \\ 2a+5c & 2b+5d \end{bmatrix}.$$

If $A = \begin{bmatrix} 0 & -1 & 2 \\ 3 & 1 & 4 \\ -2 & 5 & -3 \end{bmatrix}$, then A can be decomposed as follows:

$$A = \left[\begin{array}{c|cc} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \text{ or } \quad A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right], \text{ or}$$

$$A = \left[\begin{array}{cc|c} 0 & -1 & 2 \\ \hline 3 & 1 & 4 \\ -2 & 5 & -3 \end{array}\right] \text{ and so on.}$$

Suppose $A = \begin{array}{c} \\ n_1 \\ n_2 \end{array} \begin{array}{c} \begin{array}{cc} m_1 & m_2 \end{array} \\ \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \end{array}$ and $B = \begin{array}{c} \\ r_1 \\ r_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{bmatrix} E & F \\ G & H \end{bmatrix} \end{array}$. Then the matrices $P,\ Q,\ R,\ S$ and $E,\ F,\ G,\ H$, are called the blocks of the matrices A and B, respectively.

Even if $A+B$ is defined, the orders of P and E **may not be same** and hence, we may not be able to add A and B in the block form. But, if $A+B$ and $P+E$ is defined then $A+B = \begin{bmatrix} P+E & Q+F \\ R+G & S+H \end{bmatrix}$.

Similarly, if the product AB is defined, the product PE need not be defined. Therefore, we can talk of matrix product AB as block product of matrices, if both the products AB and PE are defined. And in this case, we have $AB = \begin{bmatrix} PE+QG & PF+QH \\ RE+SG & RF+SH \end{bmatrix}$.

That is, ONCE A PARTITION OF A IS FIXED, THE PARTITION OF B HAS TO BE PROPERLY CHOSEN FOR PURPOSES OF BLOCK ADDITION OR MULTIPLICATION.

Exercise 1.3.7. 1. *Complete the proofs of Theorems 1.2.5 and 1.2.13.*

2. *Let* $A = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$ *and* $C = \begin{bmatrix} 2 & 2 & 2 & 6 \\ 2 & 1 & 2 & 5 \\ 3 & 3 & 4 & 10 \end{bmatrix}$. *Compute*

   (a) *the first row of* AC,

   (b) *the first row of* $B(AC)$,

(c) *the second row of* $B(AC)$, *and*

(d) *the third row of* $B(AC)$.

(e) *Let* $\mathbf{x}^t = [1, 1, 1, -1]$. *Compute the matrix product* $C\mathbf{x}$.

3. *Let* $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ *and* $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. *Determine the* 2×2 *matrix*

   (a) A *such that the* $\mathbf{y} = A\mathbf{x}$ *gives rise to counter-clockwise rotation through an angle* α.

   (b) B *such that* $\mathbf{y} = B\mathbf{x}$ *gives rise to the reflection along the line* $y = (\tan\gamma)\mathbf{x}$.

   Now, let C *and* D *be two* 2×2 *matrices such that* $\mathbf{y} = C\mathbf{x}$ *gives rise to counter-clockwise rotation through an angle* β *and* $\mathbf{y} = D\mathbf{x}$ *gives rise to the reflection along the line* $y = (\tan\delta)\,\mathbf{x}$, *respectively. Then prove that*

   (c) $\mathbf{y} = (AC)\mathbf{x}$ *or* $\mathbf{y} = (CA)\mathbf{x}$ *give rise to counter-clockwise rotation through an angle* $\alpha + \beta$.

   (d) $\mathbf{y} = (BD)\mathbf{x}$ *or* $\mathbf{y} = (DB)\mathbf{x}$ *give rise to rotations. Which angles do they represent?*

   (e) *What can you say about* $\mathbf{y} = (AB)\mathbf{x}$ *or* $\mathbf{y} = (BA)\mathbf{x}$ *?*

4. *Let* $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $B = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ *and* $C = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. *If* $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ *and* $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ *then geometrically interpret the following:*

   (a) $\mathbf{y} = A\mathbf{x}$, $\mathbf{y} = B\mathbf{x}$ *and* $\mathbf{y} = C\mathbf{x}$.

   (b) $\mathbf{y} = (BC)\mathbf{x}$, $\mathbf{y} = (CB)\mathbf{x}$, $\mathbf{y} = (BA)\mathbf{x}$ *and* $\mathbf{y} = (AB)\mathbf{x}$.

5. *Consider the two coordinate transformations*

   $$\begin{array}{ll} x_1 & = a_{11}y_1 + a_{12}y_2 \\ x_2 & = a_{21}y_1 + a_{22}y_2 \end{array} \quad \text{and} \quad \begin{array}{ll} y_1 & = b_{11}z_1 + b_{12}z_2 \\ y_2 & = b_{21}z_1 + b_{22}z_2 \end{array}.$$

   (a) *Compose the two transformations to express* x_1, x_2 *in terms of* z_1, z_2.

   (b) *If* $\mathbf{x}^t = [x_1,\ x_2]$, $\mathbf{y}^t = [y_1,\ y_2]$ *and* $\mathbf{z}^t = [z_1,\ z_2]$ *then find matrices* A, B *and* C *such that* $\mathbf{x} = A\mathbf{y}$, $\mathbf{y} = B\mathbf{z}$ *and* $\mathbf{x} = C\mathbf{z}$.

   (c) *Is* $C = AB$?

6. *Let* A *be an* $n \times n$ *matrix. Then* **trace** *of* A, *denoted* $tr(A)$, *is defined as*

   $$tr(A) = a_{11} + a_{22} + \cdots a_{nn}.$$

   (a) *Let* $A = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$ *and* $B = \begin{bmatrix} 4 & -3 \\ -5 & 1 \end{bmatrix}$. *Compute* $tr(A)$ *and* $tr(B)$.

(b) *Then for two square matrices, A and B of the same order, prove that*

i. $tr\ (A+B) = tr\ (A) + tr\ (B)$.

ii. $tr\ (AB) = tr\ (BA)$.

(c) *Prove that there do not exist matrices A and B such that $AB - BA = cI_n$ for any $c \neq 0$.*

7. *Let A and B be two $m \times n$ matrices with real entries. Then prove that*

(a) $A\mathbf{x} = \mathbf{0}$ *for all $n \times 1$ vector $\mathbf{x}$ with real entries implies $A = \mathbf{0}$, the zero matrix.*

(b) $A\mathbf{x} = B\mathbf{x}$ *for all $n \times 1$ vector $\mathbf{x}$ with real entries implies $A = B$.*

8. *Let A be an $n \times n$ matrix such that $AB = BA$ for all $n \times n$ matrices B. Show that $A = \alpha I$ for some $\alpha \in \mathbb{R}$.*

9. *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \end{bmatrix}$.

(a) *Find a matrix B such that $AB = I_2$.*

(b) *What can you say about the number of such matrices? Give reasons for your answer.*

(c) *Does there exist a matrix C such that $CA = I_3$? Give reasons for your answer.*

10. *Let* $A = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right]$ *and* $B = \left[\begin{array}{cc|cc} 1 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 \\ \hline 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{array}\right]$. *Compute the matrix product AB using the block matrix multiplication.*

11. *Let* $A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$. *If P, Q, R and S are symmetric, is the matrix A symmetric? If A is symmetric, is it necessary that the matrices P, Q, R and S are symmetric?*

12. *Let A be an $(n+1) \times (n+1)$ matrix and let* $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & c \end{bmatrix}$, *where A_{11} is an $n \times n$ invertible matrix and c is a real number.*

(a) *If $p = c - A_{21}A_{11}^{-1}A_{12}$ is non-zero, prove that*

$$B = \begin{bmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{p} \begin{bmatrix} A_{11}^{-1}A_{12} \\ -1 \end{bmatrix} \begin{bmatrix} A_{21}A_{11}^{-1} & -1 \end{bmatrix}$$

is the inverse of A.

(b) *Find the inverse of the matrices* $\left[\begin{array}{cc|c} 0 & -1 & 2 \\ 1 & 1 & 4 \\ \hline -2 & 1 & 1 \end{array}\right]$ *and* $\left[\begin{array}{cc|c} 0 & -1 & 2 \\ 3 & 1 & 4 \\ \hline -2 & 5 & -3 \end{array}\right]$.

13. *Let* $\mathbf{x}$ *be an* $n \times 1$ *matrix satisfying* $\mathbf{x}^t\mathbf{x} = 1$.

    (a) *Define* $A = I_n - 2\mathbf{x}\mathbf{x}^t$. *Prove that* A *is symmetric and* $A^2 = I$. *The matrix* A *is commonly known as the* HOUSEHOLDER MATRIX.

    (b) *Let* $\alpha \neq 1$ *be a real number and define* $A = I_n - \alpha\mathbf{x}\mathbf{x}^t$. *Prove that* A *is symmetric and invertible [Hint: the inverse is also of the form* $I_n + \beta\mathbf{x}\mathbf{x}^t$ *for some value of* β*].*

14. *Let* A *be an* $n \times n$ *invertible matrix and let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two* $n \times 1$ *matrices. Also, let* β *be a real number such that* $\alpha = 1 + \beta\mathbf{y}^t A^{-1}\mathbf{x} \neq 0$. *Then prove the famous Shermon-Morrison formula*

$$(A + \beta\mathbf{x}\mathbf{y}^t)^{-1} = A^{-1} - \frac{\beta}{\alpha} A^{-1}\mathbf{x}\mathbf{y}^t A^{-1}.$$

    This formula gives the information about the inverse when an invertible matrix is modified by a rank one matrix.

15. *Let* J *be an* $n \times n$ *matrix having each entry* 1.

    (a) *Prove that* $J^2 = nJ$.

    (b) *Let* $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$. *Prove that there exist* $\alpha_3, \beta_3 \in \mathbb{R}$ *such that*

$$(\alpha_1 I_n + \beta_1 J) \cdot (\alpha_2 I_n + \beta_2 J) = \alpha_3 I_n + \beta_3 J.$$

    (c) *Let* $\alpha, \beta \in \mathcal{R}$ *with* $\alpha \neq 0$ *and* $\alpha + n\beta \neq 0$ *and define* $A = \alpha I_n + \beta J$. *Prove that* A *is invertible.*

16. *Let* A *be an upper triangular matrix. If* $A^*A = AA^*$ *then prove that* A *is a diagonal matrix. The same holds for lower triangular matrix.*

1.4 Summary

In this chapter, we started with the definition of a matrix and came across lots of examples. In particular, the following examples were important:

1. The zero matrix of size $m \times n$, denoted $\mathbf{0}_{m\times n}$ or $\mathbf{0}$.
2. The identity matrix of size $n \times n$, denoted I_n or I.
3. Triangular matrices
4. Hermitian/Symmetric matrices
5. Skew-Hermitian/skew-symmetric matrices
6. Unitary/Orthogonal matrices

We also learnt product of two matrices. Even though it seemed complicated, it basically tells the following:

1. Multiplying by a matrix on the left to a matrix A is same as row operations.
2. Multiplying by a matrix on the right to a matrix A is same as column operations.

Chapter 2

System of Linear Equations

2.1 Introduction

Let us look at some examples of linear systems.

1. Suppose $a, b \in \mathbb{R}$. Consider the system $ax = b$.
   (a) If $a \neq 0$ then the system has a UNIQUE SOLUTION $x = \frac{b}{a}$.
   (b) If $a = 0$ and
      i. $b \neq 0$ then the system has NO SOLUTION.
      ii. $b = 0$ then the system has INFINITE NUMBER OF SOLUTIONS, namely all $x \in \mathbb{R}$.

2. Consider a system with 2 equations in 2 unknowns. The equation $ax + by = c$ represents a line in $\mathbb{R}^2$ if either $a \neq 0$ or $b \neq 0$. Thus the solution set of the system

   $$a_1 x + b_1 y = c_1, \; a_2 x + b_2 y = c_2$$

   is given by the points of intersection of the two lines. The different cases are illustrated by examples (see Figure 1).

   (a) UNIQUE SOLUTION
   $x + 2y = 1$ and $x + 3y = 1$. The unique solution is $(x, y)^t = (1, 0)^t$.
   Observe that in this case, $a_1 b_2 - a_2 b_1 \neq 0$.

   (b) INFINITE NUMBER OF SOLUTIONS
   $x + 2y = 1$ and $2x + 4y = 2$. The solution set is $(x, y)^t = (1 - 2y, y)^t = (1, 0)^t + y(-2, 1)^t$ with y arbitrary as both the equations represent the same line. Observe the following:
      i. Here, $a_1 b_2 - a_2 b_1 = 0, \; a_1 c_2 - a_2 c_1 = 0$ and $b_1 c_2 - b_2 c_1 = 0$.
      ii. The vector $(1, 0)^t$ corresponds to the solution $x = 1, y = 0$ of the given system whereas the vector $(-2, 1)^t$ corresponds to the solution $x = -2, y = 1$ of the system $x + 2y = 0, 2x + 4y = 0$.

(c) NO SOLUTION
$x+2y=1$ and $2x+4y=3$. The equations represent a pair of parallel lines and hence there is no point of intersection. Observe that in this case, $a_1b_2-a_2b_1=0$ but $a_1c_2-a_2c_1 \neq 0$.

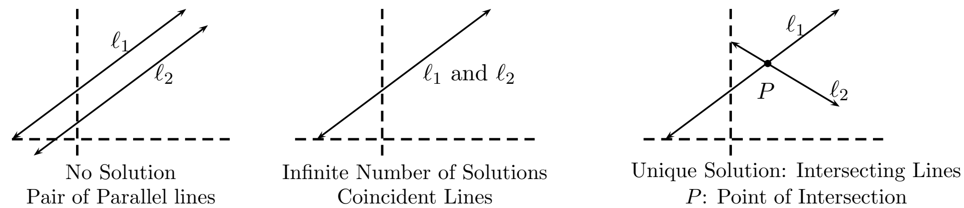

Figure 1 : Examples in 2 dimension.

3. As a last example, consider 3 equations in 3 unknowns.
A linear equation $ax+by+cz=d$ represent a plane in $\mathbb{R}^3$ provided $(a,b,c) \neq (0,0,0)$. Here, we have to look at the points of intersection of the three given planes.

(a) UNIQUE SOLUTION
Consider the system $x+y+z=3, \;\; x+4y+2z=7$ and $4x+10y-z=13$. The unique solution to this system is $(x,y,z)^t=(1,1,1)^t$; *i.e.* THE THREE PLANES INTERSECT AT A POINT.

(b) INFINITE NUMBER OF SOLUTIONS
Consider the system $x+y+z=3, \;\; x+2y+2z=5$ and $3x+4y+4z=11$. The solution set is $(x,y,z)^t=(1,2-z,z)^t=(1,2,0)^t+z(0,-1,1)^t$, with z arbitrary. Observe the following:

i. Here, the three planes intersect in a line.

ii. The vector $(1,2,0)^t$ corresponds to the solution $x=1, y=2$ and $z=0$ of the linear system $x+y+z=3, \;\; x+2y+2z=5$ and $3x+4y+4z=11$. Also, the vector $(0,-1,1)^t$ corresponds to the solution $x=0, y=-1$ and $z=1$ of the linear system $x+y+z=0, \;\; x+2y+2z=0$ and $3x+4y+4z=0$.

(c) NO SOLUTION
The system $x+y+z=3, \;\; x+2y+2z=5$ and $3x+4y+4z=13$ has no solution. In this case, we get three parallel lines as intersections of the above planes, namely

i. a line passing through $(1,2,0)$ with direction ratios $(0,-1,1)$,

ii. a line passing through $(3,1,0)$ with direction ratios $(0,-1,1)$, and

iii. a line passing through $(-1,4,0)$ with direction ratios $(0,-1,1)$.

The readers are advised to supply the proof.

Definition 2.1.1 (Linear System)**.** *A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \quad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{2.1.1}$$

where for $1 \leq i \leq n$, and $1 \leq j \leq m$; $a_{ij}, b_i \in \mathbb{R}$. Linear System (2.1.1) is called HOMOGENEOUS *if $b_1 = 0 = b_2 = \cdots = b_m$ and* NON-HOMOGENEOUS *otherwise.*

We rewrite the above equations in the form $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

The matrix A is called the COEFFICIENT matrix and the block matrix $[A \;\; \mathbf{b}]$, is called the AUGMENTED matrix of the linear system (2.1.1).

Remark 2.1.2.

1. *The first column of the augmented matrix corresponds to the coefficients of the variable x_1.*
2. *In general, the j^{th} column of the augmented matrix corresponds to the coefficients of the variable x_j, for $j = 1, 2, \ldots, n$.*
3. *The $(n+1)^{th}$ column of the augmented matrix consists of the vector $\mathbf{b}$.*
4. *The i^{th} row of the augmented matrix represents the i^{th} equation for $i = 1, 2, \ldots, m$.*

   That is, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, the entry a_{ij} of the coefficient matrix A corresponds to the i^{th} linear equation and the j^{th} variable x_j.

Definition 2.1.3. *For a system of linear equations $A\mathbf{x} = \mathbf{b}$, the system $A\mathbf{x} = \mathbf{0}$ is called the* ASSOCIATED HOMOGENEOUS SYSTEM.

Definition 2.1.4 (Solution of a Linear System)**.** *A solution of $A\mathbf{x} = \mathbf{b}$ is a column vector $\mathbf{y}$ with entries $y_1, y_2, \ldots, y_n$ such that the linear system (2.1.1) is satisfied by substituting y_i in place of x_i. The collection of all solutions is called the* SOLUTION SET *of the system.*

That is, if $\mathbf{y}^t = [y_1, y_2, \ldots, y_n]$ is a solution of the linear system $A\mathbf{x} = \mathbf{b}$ then $A\mathbf{y} = \mathbf{b}$ holds. For example, from Example 3.3a, we see that the vector $\mathbf{y}^t = [1, 1, 1]$ is a solution of the system $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 2 \\ 4 & 10 & -1 \end{bmatrix}$, $\mathbf{x}^t = [x, y, z]$ and $\mathbf{b}^t = [3, 7, 13]$.

We now state a theorem about the solution set of a homogeneous system. The readers are advised to supply the proof.

Theorem 2.1.5. *Consider the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. Then*

1. *The zero vector,* $\mathbf{0} = (0, \ldots, 0)^t$, *is always a solution, called the* TRIVIAL *solution.*

2. *Suppose* $\mathbf{x}_1, \mathbf{x}_2$ *are two solutions of* $A\mathbf{x} = \mathbf{0}$. *Then* $k_1\mathbf{x}_1 + k_2\mathbf{x}_2$ *is also a solution of* $A\mathbf{x} = \mathbf{0}$ *for any* $k_1, k_2 \in \mathbb{R}$.

Remark 2.1.6. 1. *A non-zero solution of* $A\mathbf{x} = \mathbf{0}$ *is called a* NON-TRIVIAL *solution.*

2. *If* $A\mathbf{x} = \mathbf{0}$ *has a non-trivial solution, say* $\mathbf{y} \neq \mathbf{0}$ *then* $\mathbf{z} = c\mathbf{y}$ *for every* $c \in \mathbb{R}$ *is also a solution. Thus, the existence of a non-trivial solution of* $A\mathbf{x} = \mathbf{0}$ *is equivalent to having an infinite number of solutions for the system* $A\mathbf{x} = \mathbf{0}$.

3. *If* $\mathbf{u}, \mathbf{v}$ *are two distinct solutions of* $A\mathbf{x} = \mathbf{b}$ *then one has the following:*

   (a) $\mathbf{u} - \mathbf{v}$ *is a solution of the system* $A\mathbf{x} = \mathbf{0}$.

   (b) *Define* $\mathbf{x}_h = \mathbf{u} - \mathbf{v}$. *Then* $\mathbf{x}_h$ *is a solution of the homogeneous system* $A\mathbf{x} = \mathbf{0}$.

   (c) *That is, any two solutions of* $A\mathbf{x} = \mathbf{b}$ *differ by a solution of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$.

   (d) *Or equivalently, the set of solutions of the system* $A\mathbf{x} = \mathbf{b}$ *is of the form,* $\{\mathbf{x}_0 + \mathbf{x}_h\}$; *where* $\mathbf{x}_0$ *is a particular solution of* $A\mathbf{x} = \mathbf{b}$ *and* $\mathbf{x}_h$ *is a solution of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$.

2.1.1 A Solution Method

Example 2.1.7. *Solve the linear system* $y + z = 2,\ 2x + 3z = 5,\ x + y + z = 3$.

Solution: In this case, the augmented matrix is $\begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$ and the solution method proceeds along the following steps.

1. Interchange 1^{st} and 2^{nd} equation.

$$\begin{aligned} 2x + 3z &= 5 \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad \begin{bmatrix} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

2. Replace 1^{st} equation by 1^{st} equation times $\frac{1}{2}$.

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

3. Replace 3^{rd} equation by 3^{rd} equation minus the 1^{st} equation.

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ y - \tfrac{1}{2}z &= \tfrac{1}{2} \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

4. Replace 3^{rd} equation by 3^{rd} equation minus the 2^{nd} equation.

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ -\tfrac{3}{2}z &= -\tfrac{3}{2} \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

5. Replace 3^{rd} equation by 3^{rd} equation times $\frac{-2}{3}$.

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ z &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

The last equation gives $z = 1$. Using this, the second equation gives $y = 1$. Finally, the first equation gives $x = 1$. Hence the solution set is $\{(x, y, z)^t : (x, y, z) = (1, 1, 1)\}$, A UNIQUE SOLUTION.

In Example 2.1.7, observe that certain operations on equations (rows of the augmented matrix) helped us in getting a system in Item 5, which was easily solvable. We use this idea to define elementary row operations and equivalence of two linear systems.

Definition 2.1.8 (Elementary Row Operations). *Let A be an $m \times n$ matrix. Then the elementary row operations are defined as follows:*

1. *R_{ij}: Interchange of the i^{th} and the j^{th} row of A.*
2. *For $c \neq 0$, $R_k(c)$: Multiply the k^{th} row of A by c.*
3. *For $c \neq 0$, $R_{ij}(c)$: Replace the j^{th} row of A by the j^{th} row of A plus c times the i^{th} row of A.*

Definition 2.1.9 (Equivalent Linear Systems). *Let $[A \;\; \mathbf{b}]$ and $[C \;\; \mathbf{d}]$ be augmented matrices of two linear systems. Then the two linear systems are said to be equivalent if $[C \;\; \mathbf{d}]$ can be obtained from $[A \;\; \mathbf{b}]$ by application of a finite number of elementary row operations.*

Definition 2.1.10 (Row Equivalent Matrices). *Two matrices are said to be row-equivalent if one can be obtained from the other by a finite number of elementary row operations.*

Thus, note that linear systems at each step in Example 2.1.7 are equivalent to each other. We also prove the following result that relates elementary row operations with the solution set of a linear system.

Lemma 2.1.11. *Let $C\mathbf{x} = \mathbf{d}$ be the linear system obtained from $A\mathbf{x} = \mathbf{b}$ by application of a single elementary row operation. Then $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ have the same solution set.*

Proof. We prove the result for the elementary row operation $R_{jk}(c)$ with $c \neq 0$. The reader is advised to prove the result for other elementary operations.

In this case, the systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ vary only in the k^{th} equation. Let $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ be a solution of the linear system $A\mathbf{x} = \mathbf{b}$. Then substituting for α_i's in place of x_i's in the k^{th} and j^{th} equations, we get

$$a_{k1}\alpha_1 + a_{k2}\alpha_2 + \cdots a_{kn}\alpha_n = b_k, \text{ and } a_{j1}\alpha_1 + a_{j2}\alpha_2 + \cdots a_{jn}\alpha_n = b_j.$$

Therefore,

$$(a_{k1} + ca_{j1})\alpha_1 + (a_{k2} + ca_{j2})\alpha_2 + \cdots + (a_{kn} + ca_{jn})\alpha_n = b_k + cb_j. \tag{2.1.2}$$

But then the k^{th} equation of the linear system $C\mathbf{x} = \mathbf{d}$ is

$$(a_{k1} + ca_{j1})x_1 + (a_{k2} + ca_{j2})x_2 + \cdots + (a_{kn} + ca_{jn})x_n = b_k + cb_j. \tag{2.1.3}$$

Therefore, using Equation (2.1.2), $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is also a solution for k^{th} Equation (2.1.3).

Use a similar argument to show that if $(\beta_1, \beta_2, \ldots, \beta_n)$ is a solution of the linear system $C\mathbf{x} = \mathbf{d}$ then it is also a solution of the linear system $A\mathbf{x} = \mathbf{b}$. Hence, the required result follows. $\square$

The readers are advised to use Lemma 2.1.11 as an induction step to prove the main result of this subsection which is stated next.

Theorem 2.1.12. *Two equivalent linear systems have the same solution set.*

2.1.2 Gauss Elimination Method

We first define the Gauss elimination method and give a few examples to understand the method.

Definition 2.1.13 (Forward/Gauss Elimination Method)**.** *The Gaussian elimination method is a procedure for solving a linear system $A\mathbf{x} = \mathbf{b}$ (consisting of m equations in n unknowns) by bringing the augmented matrix*

$$[A \;\; \mathbf{b}] = \left[\begin{array}{cccccc|c} a_{11} & a_{12} & \cdots & a_{1m} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} & \cdots & a_{mn} & b_m \end{array}\right]$$

to an upper triangular form

$$\left[\begin{array}{cccccc|c} c_{11} & c_{12} & \cdots & c_{1m} & \cdots & c_{1n} & d_1 \\ 0 & c_{22} & \cdots & c_{2m} & \cdots & c_{2n} & d_2 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & c_{mm} & \cdots & c_{mn} & d_m \end{array}\right]$$

by application of elementary row operations. This elimination process is also called the forward elimination method.

We have already seen an example before defining the notion of row equivalence. We give two more examples to illustrate the Gauss elimination method.

Example 2.1.14. *Solve the following linear system by Gauss elimination method.*

$$x+y+z=3,\ x+2y+2z=5,\ 3x+4y+4z=11$$

Solution: Let $A=\begin{bmatrix}1&1&1\\1&2&2\\3&4&4\end{bmatrix}$ and $\mathbf{b}=\begin{bmatrix}3\\5\\11\end{bmatrix}$. The Gauss Elimination method starts with the augmented matrix $[A\ \ \mathbf{b}]$ and proceeds as follows:

1. Replace 2^{nd} equation by 2^{nd} equation minus the 1^{st} equation.

$$\begin{array}{rl} x+y+z & =3\\ y+z & =2\\ 3x+4y+4z & =11\end{array} \qquad \begin{bmatrix}1&1&1&3\\0&1&1&2\\3&4&4&11\end{bmatrix}.$$

2. Replace 3^{rd} equation by 3^{rd} equation minus 3 times 1^{st} equation.

$$\begin{array}{rl} x+y+z & =3\\ y+z & =2\\ y+z & =2\end{array} \qquad \begin{bmatrix}1&1&1&3\\0&1&1&2\\0&1&1&2\end{bmatrix}.$$

3. Replace 3^{rd} equation by 3^{rd} equation minus the 2^{nd} equation.

$$\begin{array}{rl} x+y+z & =3\\ y+z & =2\end{array} \qquad \begin{bmatrix}1&1&1&3\\0&1&1&2\\0&0&0&0\end{bmatrix}.$$

Thus, the solution set is $\{(x,y,z)^t : (x,y,z)=(1,2-z,z)\}$ or equivalently $\{(x,y,z)^t : (x,y,z)=(1,2,0)+z(0,-1,1)\}$, with z arbitrary. In other words, the system has INFINITE NUMBER OF SOLUTIONS. Observe that the vector $\mathbf{y}^t=(1,2,0)$ satisfies $A\mathbf{y}=\mathbf{b}$ and the vector $\mathbf{z}^t=(0,-1,1)$ is a solution of the homogeneous system $A\mathbf{x}=\mathbf{0}$.

Example 2.1.15. *Solve the following linear system by Gauss elimination method.*

$$x+y+z=3,\ x+2y+2z=5,\ 3x+4y+4z=12$$

Solution: Let $A=\begin{bmatrix}1&1&1\\1&2&2\\3&4&4\end{bmatrix}$ and $\mathbf{b}=\begin{bmatrix}3\\5\\12\end{bmatrix}$. The Gauss Elimination method starts with the augmented matrix $[A\ \ \mathbf{b}]$ and proceeds as follows:

1. Replace 2^{nd} equation by 2^{nd} equation minus the 1^{st} equation.

$$\begin{array}{rl} x+y+z & =3\\ y+z & =2\\ 3x+4y+4z & =12\end{array} \qquad \begin{bmatrix}1&1&1&3\\0&1&1&2\\3&4&4&12\end{bmatrix}.$$

2. Replace 3^{rd} equation by 3^{rd} equation minus 3 times 1^{st} equation.

$$\begin{array}{ll} x+y+z & =3 \\ y+z & =2 \\ y+z & =3 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 \end{bmatrix}.$$

3. Replace 3^{rd} equation by 3^{rd} equation minus the 2^{nd} equation.

$$\begin{array}{ll} x+y+z & =3 \\ y+z & =2 \\ 0 & =1 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The third equation in the last step is

$$0x + 0y + 0z = 1.$$

This can never hold for any value of x, y, z. Hence, the system has NO SOLUTION.

Remark 2.1.16. *Note that to solve a linear system $A\mathbf{x} = \mathbf{b}$, one needs to apply only the row operations to the augmented matrix $[A \;\; \mathbf{b}]$.*

Definition 2.1.17 (Row Echelon Form of a Matrix)**.** *A matrix C is said to be in the row echelon form if*

1. *the rows consisting entirely of zeros appears after the non-zero rows,*
2. *the first non-zero entry in a non-zero row is 1. This term is called the* LEADING TERM *or a* LEADING 1. *The column containing this term is called the* LEADING COLUMN.
3. *In any two successive non-zero rows, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.*

Example 2.1.18. *The matrices* $\begin{bmatrix} 0 & 1 & 4 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *and* $\begin{bmatrix} 1 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ *are in row-echelon form. Whereas, the matrices*

$$\begin{bmatrix} 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

are not in row-echelon form.

Definition 2.1.19 (Basic, Free Variables)**.** *Let $A\mathbf{x} = \mathbf{b}$ be a linear system consisting of m equations in n unknowns. Suppose the application of Gauss elimination method to the augmented matrix $[A \;\; \mathbf{b}]$ yields the matrix $[C \;\; \mathbf{d}]$.*

1. *Then the variables corresponding to the leading columns* (in the first n columns of $[C \;\; \mathbf{d}]$) *are called the* BASIC *variables.*

2. *The variables which are not basic are called* FREE *variables.*

The free variables are called so as they can be assigned arbitrary values. Also, the basic variables can be written in terms of the free variables and hence the value of basic variables in the solution set depend on the values of the free variables.

Remark 2.1.20. *Observe the following:*

1. *In Example 2.1.14, the solution set was given by*

$$(x, y, z) = (1, 2 - z, z) = (1, 2, 0) + z(0, -1, 1), \quad \textit{with } z \ \textit{ arbitrary.}$$

   That is, we had x, y as two basic variables and z as a free variable.

2. *Example 2.1.15 didn't have any solution because the row-echelon form of the augmented matrix had a row of the form $[0, 0, 0, 1]$.*

3. *Suppose the application of row operations to $[A \ \ \mathbf{b}]$ yields the matrix $[C \ \ \mathbf{d}]$ which is in row echelon form. If $[C \ \ \mathbf{d}]$ has r non-zero rows then $[C \ \ \mathbf{d}]$ will consist of r leading terms or r leading columns. Therefore,* THE LINEAR SYSTEM $A\mathbf{x} = \mathbf{b}$ WILL HAVE r BASIC VARIABLES AND $n - r$ FREE VARIABLES.

Before proceeding further, we have the following definition.

Definition 2.1.21 (Consistent, Inconsistent)**.** *A linear system is called* CONSISTENT *if it admits a solution and is called* INCONSISTENT *if it admits no solution.*

We are now ready to prove conditions under which the linear system $A\mathbf{x} = \mathbf{b}$ is consistent or inconsistent.

Theorem 2.1.22. *Consider the linear system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix and $\mathbf{x}^t = (x_1, x_2, \ldots, x_n)$. If one obtains $[C \ \ \mathbf{d}]$ as the row-echelon form of $[A \ \ \mathbf{b}]$ with $\mathbf{d}^t = (d_1, d_2, \ldots, d_m)$ then*

1. *$A\mathbf{x} = \mathbf{b}$ is inconsistent (has no solution) if $[C \ \ \mathbf{d}]$ has a row of the form $[\mathbf{0}^t \ 1]$, where $\mathbf{0}^t = (0, \ldots, 0)$.*

2. *$A\mathbf{x} = \mathbf{b}$ is consistent (has a solution) if $[C \ \ \mathbf{d}]$ has* NO ROW OF THE FORM $[\mathbf{0}^t \ 1]$. *Furthermore,*

   (a) *if the number of variables equals the number of leading terms then $A\mathbf{x} = \mathbf{b}$ has* A UNIQUE SOLUTION.

   (b) *if the number of variables is strictly greater than the number of leading terms then $A\mathbf{x} = \mathbf{b}$ has* INFINITE NUMBER OF SOLUTIONS.

Proof. PART 1: The linear equation corresponding to the row $[\mathbf{0}^t \ 1]$ equals

$$0x_1 + 0x_2 + \cdots + 0x_n = 1.$$

Obviously, this equation has no solution and hence the system $C\mathbf{x} = \mathbf{d}$ has no solution. Thus, by Theorem 2.1.12, $A\mathbf{x} = \mathbf{b}$ has no solution. That is, $A\mathbf{x} = \mathbf{b}$ is inconsistent.

PART 2: Suppose $[C \;\; \mathbf{d}]$ has r non-zero rows. As $[C \;\; \mathbf{d}]$ is in row echelon form there exist positive integers $1 \leq i_1 < i_2 < \ldots < i_r \leq n$ such that entries $c_{\ell i_\ell}$ for $1 \leq \ell \leq r$ are leading terms. This in turn implies that the variables x_{i_j}, for $1 \leq j \leq r$ are the basic variables and the remaining $n-r$ variables, say $x_{t_1}, x_{t_2}, \ldots, x_{t_{n-r}}$, are free variables. So for each ℓ, $1 \leq \ell \leq r$, one obtains $x_{i_\ell} + \sum\limits_{k > i_\ell} c_{\ell k} x_k = d_\ell$ ($k > i_\ell$ in the summation as $[C \;\; \mathbf{d}]$ is a matrix in the row reduced echelon form). Or equivalently,

$$x_{i_\ell} = d_\ell - \sum_{j=\ell+1}^{r} c_{\ell i_j} x_{i_j} - \sum_{s=1}^{n-r} c_{\ell t_s} x_{t_s} \quad \text{for} \quad 1 \leq l \leq r.$$

Hence, a solution of the system $C\mathbf{x} = \mathbf{d}$ is given by

$$x_{t_s} = 0 \text{ for } \; s = 1, \ldots, n-r \text{ and } \; x_{i_r} = d_r, x_{i_{r-1}} = d_{r-1} - d_r, \ldots, x_{i_1} = d_1 - \sum_{j=2}^{r} c_{\ell i_j} d_j.$$

Thus, by Theorem 2.1.12 the system $A\mathbf{x} = \mathbf{b}$ is consistent. In case of Part 2a, there are no free variables and hence the unique solution is given by

$$x_n = d_n, x_{n-1} = d_{n-1} - d_n, \ldots, x_1 = d_1 - \sum_{j=2}^{n} c_{\ell i_j} d_j.$$

In case of Part 2b, there is at least one free variable and hence $A\mathbf{x} = \mathbf{b}$ has infinite number of solutions. Thus, the proof of the theorem is complete. □

We omit the proof of the next result as it directly follows from Theorem 2.1.22.

Corollary 2.1.23. *Consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then*

1. *$A\mathbf{x} = \mathbf{0}$ is always consistent as $\mathbf{0}$ is a solution.*

2. *If $m < n$ then $n - m > 0$ and there will be at least $n - m$ free variables. Thus $A\mathbf{x} = \mathbf{0}$ has infinite number of solutions. Or equivalently, $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.*

We end this subsection with some applications related to geometry.

Example 2.1.24. 1. *Determine the equation of the line/circle that passes through the points $(-1, 4), (0, 1)$ and $(1, 4)$.*

Solution: *The general equation of a line/circle in 2-dimensional plane is given by $a(x^2 + y^2) + bx + cy + d = 0$, where a, b, c and d are the unknowns. Since this curve passes through the given points, we have*

$$\begin{aligned} a((-1)^2 + 4^2) + (-1)b + 4c + d &= \;= 0 \\ a((0)^2 + 1^2) + (0)b + 1c + d &= \;= 0 \\ a((1)^2 + 4^2) + (1)b + 4c + d &= \;= 0. \end{aligned}$$

Solving this system, we get $(a,b,c,d) = (\frac{3}{13}d, 0, -\frac{16}{13}d, d)$. *Hence, taking* $d = 13$, *the equation of the required circle is*

$$3(x^2+y^2) - 16y + 13 = 0.$$

2. *Determine the equation of the plane that contains the points* $(1,1,1),(1,3,2)$ *and* $(2,-1,2)$.
Solution: *The general equation of a plane in 3-dimensional space is given by* $ax + by + cz + d = 0$, *where* a, b, c *and* d *are the unknowns. Since this plane passes through the given points, we have*

$$\begin{aligned} a+b+c+d &= \; = 0 \\ a+3b+2c+d &= \; = 0 \\ 2a-b+2c+d &= \; = 0. \end{aligned}$$

Solving this system, we get $(a,b,c,d) = (-\frac{4}{3}d, -\frac{d}{3}, -\frac{2}{3}d, d)$. *Hence, taking* $d = 3$, *the equation of the required plane is* $-4x - y + 2z + 3 = 0$.

3. *Let* $A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & -1 & 0 \\ 0 & -3 & 4 \end{bmatrix}$.

   (a) *Find a non-zero* $\mathbf{x}^t \in \mathbb{R}^3$ *such that* $A\mathbf{x} = 2\mathbf{x}$.

   (b) *Does there exist a non-zero vector* $\mathbf{y}^t \in \mathbb{R}^3$ *such that* $A\mathbf{y} = 4\mathbf{y}$?

Solution of Part 3a: *Solving for* $A\mathbf{x} = 2\mathbf{x}$ *is same as solving for* $(A - 2I)\mathbf{x} = \mathbf{0}$. *This leads to the augmented matrix* $\begin{bmatrix} 0 & 3 & 4 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{bmatrix}$. *Check that a non-zero solution is given by* $\mathbf{x}^t = (1,0,0)$.

Solution of Part 3b: *Solving for* $A\mathbf{y} = 4\mathbf{y}$ *is same as solving for* $(A - 4I)\mathbf{y} = \mathbf{0}$. *This leads to the augmented matrix* $\begin{bmatrix} -2 & 3 & 4 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & -3 & 0 & 0 \end{bmatrix}$. *Check that a non-zero solution is given by* $\mathbf{y}^t = (2,0,1)$.

Exercise 2.1.25. 1. *Determine the equation of the curve* $y = ax^2 + bx + c$ *that passes through the points* $(-1,4),(0,1)$ *and* $(1,4)$.

2. *Solve the following linear system.*

   (a) $x+y+z+w = 0$, $x-y+z+w = 0$ *and* $-x+y+3z+3w = 0$.

   (b) $x+2y = 1$, $x+y+z = 4$ *and* $3y+2z = 1$.

   (c) $x+y+z = 3$, $x+y-z = 1$ *and* $x+y+7z = 6$.

   (d) $x+y+z = 3$, $x+y-z = 1$ *and* $x+y+4z = 6$.

(e) $x+y+z=3,\ x+y-z=1,\ x+y+4z=6$ *and* $x+y-4z=-1$.

3. *For what values of c and k, the following systems have i) no solution, ii) a unique solution and iii) infinite number of solutions.*

   (a) $x+y+z=3,\ \ x+2y+cz=4,\ \ 2x+3y+2cz=k.$

   (b) $x+y+z=3,\ x+y+2cz=7,\ x+2y+3cz=k.$

   (c) $x+y+2z=3,\ x+2y+cz=5,\ x+2y+4z=k.$

   (d) $kx+y+z=1,\ x+ky+z=1,\ x+y+kz=1.$

   (e) $x+2y-z=1,\ 2x+3y+kz=3,\ x+ky+3z=2.$

   (f) $x-2y=1,\ x-y+kz=1,\ ky+4z=6.$

4. *For what values of a, does the following systems have i) no solution, ii) a unique solution and iii) infinite number of solutions.*

   (a) $x+2y+3z=4,\ \ 2x+5y+5z=6,\ \ 2x+(a^2-6)z=a+20.$

   (b) $x+y+z=3,\ 2x+5y+4z=a,\ 3x+(a^2-8)z=12.$

5. *Find the condition(s) on x,y,z so that the system of linear equations given below (in the unknowns a,b and c) is consistent?*

   (a) $a+2b-3c=x,\ 2a+6b-11c=y,\ a-2b+7c=z$

   (b) $a+b+5c=x,\ a+3c=y,\ 2a-b+4c=z$

   (c) $a+2b+3c=x,\ 2a+4b+6c=y,\ 3a+6b+9c=z$

6. *Let A be an $n\times n$ matrix. If the system $A^2\mathbf{x}=\mathbf{0}$ has a non trivial solution then show that $A\mathbf{x}=\mathbf{0}$ also has a non trivial solution.*

7. *Prove that we need to have 5 set of distinct points to specify a general conic in 2-dimensional plane.*

8. *Let $\mathbf{u}^t=(1,1,-2)$ and $\mathbf{v}^t=(-1,2,3)$. Find condition on x,y and z such that the system $c\mathbf{u}^t+d\mathbf{v}^t=(x,y,z)$ in the unknowns c and d is consistent.*

2.1.3 Gauss-Jordan Elimination

The Gauss-Jordan method consists of first applying the Gauss Elimination method to get the row-echelon form of the matrix $[A\ \ \mathbf{b}]$ and then further applying the row operations as follows. For example, consider Example 2.1.7. We start with Step 5 and apply row operations once again. But this time, we start with the 3^{rd} row.

I. Replace 2^{nd} equation by 2^{nd} equation minus the 3^{rd} equation.

$$\begin{aligned} x+\tfrac{3}{2}z &= \tfrac{5}{2} \\ y &= 2 \\ z &= 1 \end{aligned} \qquad\qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

II. Replace 1^{st} equation by 1^{st} equation minus $\frac{3}{2}$ times 3^{rd} equation.

$$\begin{aligned} x &= 1 \\ y &= 1 \\ z &= 1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

III. Thus, the solution set equals $\{(x, y, z)^t : (x, y, z) = (1, 1, 1)\}$.

Definition 2.1.26 (Row-Reduced Echelon Form)**.** *A matrix C is said to be in the row-reduced echelon form or reduced row echelon form if*

1. *C is already in the row echelon form;*

2. *the leading column containing the leading 1 has every other entry zero.*

A matrix which is in the row-reduced echelon form is also called a row-reduced echelon matrix.

Example 2.1.27. *Let* $A = \begin{bmatrix} 0 & 1 & 4 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. *Then A and B are in row echelon form. If C and D are the row-reduced echelon forms of A and B, respectively then* $C = \begin{bmatrix} 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *and* $D = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

Definition 2.1.28 (Back Substitution/Gauss-Jordan Method)**.** *The procedure to get The row-reduced echelon matrix from the row-echelon matrix is called the* BACK SUBSTITUTION*. The elimination process applied to obtain the row-reduced echelon form of the augmented matrix is called the* GAUSS-JORDAN ELIMINATION METHOD*.*

That is, the Gauss-Jordan elimination method consists of both the forward elimination and the backward substitution.

Remark 2.1.29. *Note that the row reduction involves only row operations and proceeds from* LEFT TO RIGHT*. Hence, if A is a matrix consisting of first s columns of a matrix C, then the row-reduced form of A will consist of the first s columns of the row-reduced form of C.*

The proof of the following theorem is beyond the scope of this book and is omitted.

Theorem 2.1.30. *The row-reduced echelon form of a matrix is unique.*

Remark 2.1.31. *Consider the linear system $A\mathbf{x} = \mathbf{b}$. Then Theorem 2.1.30 implies the following:*

1. *The application of the Gauss Elimination method to the augmented matrix may yield different matrices even though it leads to the same solution set.*

2. *The application of the Gauss-Jordan method to the augmented matrix yields the same matrix and also the same solution set even though we may have used different sequence of row operations.*

Example 2.1.32. *Consider $A\mathbf{x} = \mathbf{b}$, where A is a 3×3 matrix. Let $[C \;\; \mathbf{d}]$ be the row-reduced echelon form of $[A \;\; \mathbf{b}]$. Also, assume that the first column of A has a non-zero entry. Then the possible choices for the matrix $[C \;\; \mathbf{d}]$ with respective solution sets are given below:*

1. $\begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \end{bmatrix}$. *$A\mathbf{x} = \mathbf{b}$ has a* UNIQUE SOLUTION, *$(x, y, z) = (d_1, d_2, d_3)$.*

2. $\begin{bmatrix} 1 & 0 & \alpha & d_1 \\ 0 & 1 & \beta & d_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & \alpha & 0 & d_1 \\ 0 & 0 & 1 & d_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ *or* $\begin{bmatrix} 1 & \alpha & \beta & d_1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. *$A\mathbf{x} = \mathbf{b}$ has* NO SOLUTION *for any choice of α, β.*

3. $\begin{bmatrix} 1 & 0 & \alpha & d_1 \\ 0 & 1 & \beta & d_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \alpha & 0 & d_1 \\ 0 & 0 & 1 & d_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \alpha & \beta & d_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. *$A\mathbf{x} = \mathbf{b}$ has* INFINITE NUMBER OF SOLUTIONS *for every choice of α, β.*

Exercise 2.1.33. 1. *Let $A\mathbf{x} = \mathbf{b}$ be a linear system in 2 unknowns. What are the possible choices for the row-reduced echelon form of the augmented matrix $[A \;\; \mathbf{b}]$?*

2. *Find the row-reduced echelon form of the following matrices:*

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 3 \\ 3 & 0 & 7 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 1 \\ -2 & 0 & 3 \\ -5 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -1 & -2 & 3 \\ 3 & 3 & -3 & -3 \\ 1 & 1 & 2 & 2 \\ -1 & -1 & 2 & -2 \end{bmatrix}.$$

3. *Find all the solutions of the following system of equations using Gauss-Jordan method. No other method will be accepted.*

$$\begin{array}{rrrrrrl} x & +\,y & & -\,2\,u & +\,v & & = 2 \\ & & z & +\,u & +\,2\,v & & = 3 \\ & & & & v & +\,w & = 3 \\ & & & & v & +\,2\,w & = 5 \end{array}$$

2.2 Elementary Matrices

In the previous section, we solved a system of linear equations with the help of either the Gauss Elimination method or the Gauss-Jordan method. These methods required us to make row operations on the augmented matrix. Also, we know that (see Section 1.2.1) the row-operations correspond to multiplying a matrix on the left. So, in this section, we try to understand the matrices which helped us in performing the row-operations and also use this understanding to get some important results in the theory of square matrices.

Definition 2.2.1. *A square matrix E of order n is called an* ELEMENTARY MATRIX *if it is obtained by applying exactly one row operation to the identity matrix, I_n.*

Remark 2.2.2. *Fix a positive integer n. Then the elementary matrices of order n are of three types and are as follows:*

1. *E_{ij} corresponds to the interchange of the i^{th} and the j^{th} row of I_n.*
2. *For $c \neq 0$, $E_k(c)$ is obtained by multiplying the k^{th} row of I_n by c.*
3. *For $c \neq 0$, $E_{ij}(c)$ is obtained by replacing the j^{th} row of I_n by the j^{th} row of I_n plus c times the i^{th} row of I_n.*

Example 2.2.3. 1. *In particular, for $n = 3$ and a real number $c \neq 0$, one has*

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \; E_1(c) = \begin{bmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; \text{and } E_{32}(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}.$$

2. *Let $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 4 & 5 & 6 \\ 2 & 0 & 3 & 4 \end{bmatrix}$. Then B is obtained from A by the interchange of 2^{nd} and 3^{rd} row. Verify that*

$$E_{23}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 4 & 5 & 6 \\ 2 & 0 & 3 & 4 \end{bmatrix} = B.$$

3. *Let $A = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$. Then $B = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ is the row-reduced echelon form of A. The readers are advised to verify that*

$$B = E_{32}(-1) \cdot E_{21}(-1) \cdot E_3(1/3) \cdot E_{23}(2) \cdot E_{23} \cdot E_{12}(-2) \cdot E_{13} \cdot A.$$

Or equivalently, check that

$$\begin{aligned}
E_{13}A &= A_1 = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \end{bmatrix}, \; E_{12}(-2)A_1 = A_2 = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & -2 & 1 & -1 \\ 0 & 1 & 1 & 2 \end{bmatrix}, \\
E_{23}A_2 &= A_3 = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & -2 & 1 & -1 \end{bmatrix}, \; E_{23}(2)A_3 = A_4 = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 3 & 3 \end{bmatrix}, \\
E_3(1/3)A_4 &= A_5 = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \; E_{21}(-1)A_5 = A_6 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \\
E_{32}(-1)A_6 &= B = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.
\end{aligned}$$

Remark 2.2.4. *Observe the following:*

1. *The inverse of the elementary matrix E_{ij} is the matrix E_{ij} itself. That is, $E_{ij}E_{ij} = I = E_{ij}E_{ij}$.*
2. *Let $c \neq 0$. Then the inverse of the elementary matrix $E_k(c)$ is the matrix $E_k(1/c)$. That is, $E_k(c)E_k(1/c) = I = E_k(1/c)E_k(c)$.*
3. *Let $c \neq 0$. Then the inverse of the elementary matrix $E_{ij}(c)$ is the matrix $E_{ij}(-c)$. That is, $E_{ij}(c)E_{ij}(-c) = I = E_{ij}(-c)E_{ij}(c)$.*

   That is, all the elementary matrices are invertible and the inverses are also elementary matrices.
4. *Suppose the row-reduced echelon form of the augmented matrix $[A \;\; \mathbf{b}]$ is the matrix $[C \;\; \mathbf{d}]$. As row operations correspond to multiplying on the left with elementary matrices, we can find elementary matrices, say $E_1, E_2, \ldots, E_k$, such that*

   $$E_k \cdot E_{k-1} \cdots E_2 \cdot E_1 \cdot [A \;\; \mathbf{b}] = [C \;\; \mathbf{d}].$$

   That is, the Gauss-Jordan method (or Gauss Elimination method) is equivalent to multiplying by a finite number of elementary matrices on the left to $[A \;\; \mathbf{b}]$.

We are now ready to prove a equivalent statements in the study of invertible matrices.

Theorem 2.2.5. *Let A be a square matrix of order n. Then the following statements are equivalent.*

1. *A is invertible.*
2. *The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
3. *The row-reduced echelon form of A is I_n.*

4. *A is a product of elementary matrices.*

Proof. $1 \Longrightarrow 2$

As A is invertible, we have $A^{-1}A = I_n = AA^{-1}$. Let $\mathbf{x}_0$ be a solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then, $A\mathbf{x}_0 = \mathbf{0}$ and Thus, we see that $\mathbf{0}$ is the only solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

$2 \Longrightarrow 3$

Let $\mathbf{x}^t = [x_1, x_2, \ldots, x_n]$. As $\mathbf{0}$ is the only solution of the linear system $A\mathbf{x} = \mathbf{0}$, the final equations are $x_1 = 0, x_2 = 0, \ldots, x_n = 0$. These equations can be rewritten as

$$\begin{array}{rcl}
1 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \cdots + 0 \cdot x_n & = & 0 \\
0 \cdot x_1 + 1 \cdot x_2 + 0 \cdot x_3 + \cdots + 0 \cdot x_n & = & 0 \\
0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 + \cdots + 0 \cdot x_n & = & 0 \\
\vdots & = & \vdots \\
0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \cdots + 1 \cdot x_n & = & 0.
\end{array}$$

That is, the final system of homogeneous system is given by $I_n \cdot \mathbf{x} = \mathbf{0}$. Or equivalently, the row-reduced echelon form of the augmented matrix $[A \quad \mathbf{0}]$ is $[I_n \quad \mathbf{0}]$. That is, the row-reduced echelon form of A is I_n.

$3 \Longrightarrow 4$

Suppose that the row-reduced echelon form of A is I_n. Then using Remark 2.2.4.4, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_1 E_2 \cdots E_k A = I_n. \tag{2.2.4}$$

Now, using Remark 2.2.4, the matrix E_j^{-1} is an elementary matrix and is the inverse of E_j for $1 \leq j \leq k$. Therefore, successively multiplying Equation (2.2.4) on the left by $E_1^{-1}, E_2^{-1}, \ldots, E_k^{-1}$, we get

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

and thus A is a product of elementary matrices.

$4 \Longrightarrow 1$

Suppose $A = E_1 E_2 \cdots E_k$; where the E_i's are elementary matrices. As the elementary matrices are invertible (see Remark 2.2.4) and the product of invertible matrices is also invertible, we get the required result. $\square$

As an immediate consequence of Theorem 2.2.5, we have the following important result.

Theorem 2.2.6. *Let A be a square matrix of order n.*

1. *Suppose there exists a matrix C such that $CA = I_n$. Then A^{-1} exists.*

2. *Suppose there exists a matrix B such that $AB = I_n$. Then A^{-1} exists.*

Proof. Suppose there exists a matrix C such that $CA = I_n$. Let $\mathbf{x}_0$ be a solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then $A\mathbf{x}_0 = \mathbf{0}$ and

$$\mathbf{x}_0 = I_n \cdot \mathbf{x}_0 = (CA)\mathbf{x}_0 = C(A\mathbf{x}_0) = C\mathbf{0} = \mathbf{0}.$$

That is, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Hence, using Theorem 2.2.5, the matrix A is invertible.

Using the first part, it is clear that the matrix B in the second part, is invertible. Hence

$$AB = I_n = BA.$$

Thus, A is invertible as well. □

Remark 2.2.7. *Theorem 2.2.6 implies the following:*

1. *"if we want to show that a square matrix A of order n is invertible, it is enough to show the existence of*
   (a) *either a matrix B such that $AB = I_n$*
   (b) *or a matrix C such that $CA = I_n$.*
2. *Let A be an invertible matrix of order n. Suppose there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $E_1 E_2 \cdots E_k A = I_n$. Then $A^{-1} = E_1 E_2 \cdots E_k$.*

Remark 2.2.7 gives the following method of computing the inverse of a matrix.

Summary: Let A be an $n \times n$ matrix. Apply the Gauss-Jordan method to the matrix $[A \;\; I_n]$. Suppose the row-reduced echelon form of the matrix $[A \;\; I_n]$ is $[B \;\; C]$. If $B = I_n$, then $A^{-1} = C$ or else A is not invertible.

Example 2.2.8. *Find the inverse of the matrix* $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ *using the Gauss-Jordan method.*

Solution: *let us apply the Gauss-Jordan method to the matrix* $\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$.

1. $\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_{13}} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$
2. $\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right] \xrightarrow{R_{31}(-1)} \xrightarrow{R_{32}(-1)} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$
3. $\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right] \xrightarrow{R_{21}(-1)} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$.

Thus, the inverse of the given matrix is $\begin{bmatrix} 0 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Exercise 2.2.9. 1. *Find the inverse of the following matrices using the Gauss-Jordan method.*

(i) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 7 \end{bmatrix}$, (ii) $\begin{bmatrix} 1 & 3 & 3 \\ 2 & 3 & 2 \\ 2 & 4 & 7 \end{bmatrix}$, (iii) $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, (iv) $\begin{bmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$.

2. *Which of the following matrices are elementary?*

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

3. *Let* $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. *Find the elementary matrices* E_1, E_2, E_3 *and* E_4 *such that* $E_4 \cdot E_3 \cdot E_2 \cdot E_1 \cdot A = I_2$.

4. *Let* $B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix}$. *Determine elementary matrices* E_1, E_2 *and* E_3 *such that* $E_3 \cdot E_2 \cdot E_1 \cdot B = I_3$.

5. *In Exercise 2.2.9.3, let* $C = E_4 \cdot E_3 \cdot E_2 \cdot E_1$. *Then check that* $AC = I_2$.

6. *In Exercise 2.2.9.4, let* $C = E_3 \cdot E_2 \cdot E_1$. *Then check that* $BC = I_3$.

7. *Find the inverse of the three matrices given in Example 2.2.3.3.*

8. *Let* A *be a* 1×2 *matrix and* B *be a* 2×1 *matrix having positive entries. Which of* BA *or* AB *is invertible? Give reasons.*

9. *Let* A *be an* $n \times m$ *matrix and* B *be an* $m \times n$ *matrix. Prove that*

   (a) *the matrix* $I - BA$ *is invertible if and only if the matrix* $I - AB$ *is invertible [Hint: Use Theorem 2.2.5.2].*

   (b) $(I - BA)^{-1} = I + B(I - AB)^{-1}A$ *whenever* $I - AB$ *is invertible.*

   (c) $(I - BA)^{-1}B = B(I - AB)^{-1}$ *whenever* $I - AB$ *is invertible.*

   (d) $(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B$ *whenever* A, B *and* $A + B$ *are all invertible.*

We end this section by giving two more equivalent conditions for a matrix to be invertible.

Theorem 2.2.10. *The following statements are equivalent for an* $n \times n$ *matrix* A.

1. *A is invertible.*

2. *The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.*

3. *The system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$.*

Proof. $1 \Longrightarrow 2$

Observe that $\mathbf{x}_0 = A^{-1}\mathbf{b}$ is the unique solution of the system $A\mathbf{x} = \mathbf{b}$.

$2 \Longrightarrow 3$

The system $A\mathbf{x} = \mathbf{b}$ has a solution and hence by definition, the system is consistent.

$3 \Longrightarrow 1$

For $1 \le i \le n$, define $\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i\text{th position}}, 0, \ldots, 0)^t$, and consider the linear system $A\mathbf{x} = \mathbf{e}_i$. By assumption, this system has a solution, say $\mathbf{x}_i$, for each i, $1 \le i \le n$. Define a matrix $B = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$. That is, the i^{th} column of B is the solution of the system $A\mathbf{x} = \mathbf{e}_i$. Then

$$AB = A[\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n] = [A\mathbf{x}_1, A\mathbf{x}_2 \ldots, A\mathbf{x}_n] = [\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_n] = I_n.$$

Therefore, by Theorem 2.2.6, the matrix A is invertible. □

We now state another important result whose proof is immediate from Theorem 2.2.10 and Theorem 2.2.5 and hence the proof is omitted.

Theorem 2.2.11. *Let A be an $n \times n$ matrix. Then the two statements given below cannot hold together.*

1. *The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.*

2. *The system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.*

Exercise 2.2.12. 1. *Let A and B be two square matrices of the same order such that $B = PA$ for some invertible matrix P. Then, prove that A is invertible if and only if B is invertible.*

2. *Let A and B be two $m \times n$ matrices. Then prove that the two matrices A, B are row-equivalent if and only if $B = PA$, where P is product of elementary matrices. When is this P unique?*

3. *Let $\mathbf{b}^t = [1, 2, -1, -2]$. Suppose A is a 4×4 matrix such that the linear system $A\mathbf{x} = \mathbf{b}$ has no solution. Mark each of the statements given below as* TRUE *or* FALSE*?*

   (a) *The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

   (b) *The matrix A is invertible.*

   (c) *Let $\mathbf{c}^t = [-1, -2, 1, 2]$. Then the system $A\mathbf{x} = \mathbf{c}$ has no solution.*

   (d) *Let B be the row-reduced echelon form of A. Then*

      i. *the fourth row of B is $[0, 0, 0, 0]$.*

      ii. *the fourth row of B is $[0, 0, 0, 1]$.*

iii. *the third row of B is necessarily of the form* $[0,0,0,0]$.

iv. *the third row of B is necessarily of the form* $[0,0,0,1]$.

v. *the third row of B is necessarily of the form* $[0,0,1,\alpha]$, *where α is any real number.*

2.3 Rank of a Matrix

In the previous section, we gave a few equivalent conditions for a square matrix to be invertible. We also used the Gauss-Jordan method and the elementary matrices to compute the inverse of a square matrix A. In this section and the subsequent sections, we will mostly be concerned with $m \times n$ matrices.

Let A by an $m \times n$ matrix. Suppose that C is the row-reduced echelon form of A. Then the matrix C is unique (see Theorem 2.1.30). Hence, we use the matrix C to define the rank of the matrix A.

Definition 2.3.1 (Row Rank of a Matrix)**.** *Let C be the row-reduced echelon form of a matrix A. The number of non-zero rows in C is called the row-rank of A.*

For a matrix A, we write 'row-rank (A)' to denote the row-rank of A. By the very definition, it is clear that row-equivalent matrices have the same row-rank. Thus, the number of non-zero rows in either the row echelon form or the row-reduced echelon form of a matrix are equal. Therefore, we just need to get the row echelon form of the matrix to know its rank.

Example 2.3.2. 1. *Determine the row-rank of* $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix}$.

Solution: *The row-reduced echelon form of A is obtained as follows:*

$$\begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

The final matrix has 3 non-zero rows. Thus row-rank$(A) = 3$. This also follows from the third matrix.

2. *Determine the row-rank of* $A = \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix}$.

Solution: *row-rank$(A) = 2$ as one has the following:*

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The following remark related to the augmented matrix is immediate as computing the rank only involves the row operations (also see Remark 2.1.29).

Remark 2.3.3. *Let $A\mathbf{x} = \mathbf{b}$ be a linear system with m equations in n unknowns. Then the row-reduced echelon form of A agrees with the first n columns of $[A\ \ \mathbf{b}]$, and hence*

$$\textit{row-rank}(A) \leq \textit{row-rank}([A\ \ \mathbf{b}]).$$

Now, consider an $m \times n$ matrix A and an elementary matrix E of order n. Then the product AE corresponds to applying column transformation on the matrix A. Therefore, for each elementary matrix, there is a corresponding column transformation as well. We summarize these ideas as follows.

Definition 2.3.4. *The column transformations obtained by right multiplication of elementary matrices are called* COLUMN OPERATIONS.

Example 2.3.5. *Let* $A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 0 & 3 & 2 \\ 3 & 4 & 5 & 3 \end{bmatrix}$. *Then*

$$A \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 3 & 0 & 2 \\ 3 & 5 & 4 & 3 \end{bmatrix} \quad \textit{and} \quad A \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 0 \\ 3 & 4 & 5 & 0 \end{bmatrix}.$$

Remark 2.3.6. *After application of a finite number of elementary column operations (see Definition 2.3.4) to a matrix A, we can obtain a matrix B having the following properties:*

1. *The first nonzero entry in each column is 1, called the leading term.*
2. *Column(s) containing only 0's comes after all columns with at least one non-zero entry.*
3. *The first non-zero entry (the leading term) in each non-zero column moves down in successive columns.*

We define COLUMN-RANK *of A as the number of non-zero columns in B.*

It will be proved later that row-rank(A) = column-rank(A). Thus we are led to the following definition.

Definition 2.3.7. *The number of non-zero rows in the row-reduced echelon form of a matrix A is called the* RANK *of A, denoted rank(A).*

we are now ready to prove a few results associated with the rank of a matrix.

Theorem 2.3.8. *Let A be a matrix of rank r. Then there exist a finite number of elementary matrices $E_1, E_2, \ldots, E_s$ and $F_1, F_2, \ldots, F_\ell$ such that*

$$E_1 E_2 \ldots E_s\ A\ F_1 F_2 \ldots F_\ell = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Proof. Let C be the row-reduced echelon matrix of A. As $\text{rank}(A) = r$, the first r rows of C are non-zero rows. So by Theorem 2.1.22, C will have r leading columns, say $i_1, i_2, \ldots, i_r$. Note that, for $1 \le s \le r$, the i_s^{th} column will have 1 in the s^{th} row and zero, elsewhere.

We now apply column operations to the matrix C. Let D be the matrix obtained from C by successively interchanging the s^{th} and i_s^{th} column of C for $1 \le s \le r$. Then D has the form $\begin{bmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where B is a matrix of an appropriate size. As the $(1,1)$ block of D is an identity matrix, the block $(1,2)$ can be made the zero matrix by application of column operations to D. This gives the required result. $\square$

The next result is a corollary of Theorem 2.3.8. It gives the solution set of a homogeneous system $A\mathbf{x} = \mathbf{0}$. One can also obtain this result as a particular case of Corollary 2.1.23.2 as by definition $\text{rank}(A) \le m$, the number of rows of A.

Corollary 2.3.9. *Let A be an $m \times n$ matrix. Suppose rank$(A) = r < n$. Then $A\mathbf{x} = \mathbf{0}$ has infinite number of solutions. In particular, $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.*

Proof. By Theorem 2.3.8, there exist elementary matrices $E_1, \ldots, E_s$ and $F_1, \ldots, F_\ell$ such that $E_1 E_2 \cdots E_s \; A \; F_1 F_2 \cdots F_\ell = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. Define $P = E_1 E_2 \cdots E_s$ and $Q = F_1 F_2 \cdots F_\ell$. Then the matrix $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. As E_i's for $1 \le i \le s$ correspond only to row operations, we get $AQ = \left[\begin{array}{c|c} C & \mathbf{0} \end{array}\right]$, where C is a matrix of size $m \times r$. Let $Q_1, Q_2, \ldots, Q_n$ be the columns of the matrix Q. Then check that $AQ_i = \mathbf{0}$ for $i = r+1, \ldots, n$. Hence, the required results follows (use Theorem 2.1.5). $\square$

Exercise 2.3.10.
1. *Determine ranks of the coefficient and the augmented matrices that appear in Exercise 2.1.25.2.*
2. *Let P and Q be invertible matrices such that the matrix product PAQ is defined. Prove that rank$(PAQ) =$ rank(A).*
3. *Let $A = \begin{bmatrix} 2 & 4 & 8 \\ 1 & 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Find P and Q such that $B = PAQ$.*
4. *Let A and B be two matrices. Prove that*
   - *(a) if $A + B$ is defined, then rank$(A+B) \le$ rank$(A) +$ rank(B),*
   - *(b) if AB is defined, then rank$(AB) \le$ rank(A) and rank$(AB) \le$ rank(B).*
5. *Let A be a matrix of rank r. Then prove that there exists invertible matrices B_i, C_i such that*
   $B_1 A = \begin{bmatrix} R_1 & R_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $AC_1 = \begin{bmatrix} S_1 & \mathbf{0} \\ S_3 & \mathbf{0} \end{bmatrix}$, $B_2 A C_2 = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ *and* $B_3 A C_3 = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$,
   where the $(1,1)$ block of each matrix is of size $r \times r$. Also, prove that A_1 is an invertible matrix.

6. *Let A be an $m \times n$ matrix of rank r. Then prove that A can be written as $A = BC$, where both B and C have rank r and B is of size $m \times r$ and C is of size $r \times n$.*

7. *Let A and B be two matrices such that AB is defined and rank(A) = rank(AB). Then prove that $A = ABX$ for some matrix X. Similarly, if BA is defined and rank (A) = rank (BA), then $A = YBA$ for some matrix Y. [Hint: Choose invertible matrices P, Q satisfying $PAQ = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $P(AB) = (PAQ)(Q^{-1}B) = \begin{bmatrix} A_2 & A_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Now find R an invertible matrix with $P(AB)R = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Define $X = R \begin{bmatrix} C^{-1}A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1}$.]*

8. *Suppose the matrices B and C are invertible and the involved partitioned products are defined, then prove that*
$$\begin{bmatrix} A & B \\ C & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & C^{-1} \\ B^{-1} & -B^{-1}AC^{-1} \end{bmatrix}.$$

9. *Suppose $A^{-1} = B$ with $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$. Also, assume that A_{11} is invertible and define $P = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Then prove that*

   (a) A is row-equivalent to the matrix $\begin{bmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$,

   (b) P is invertible and $B = \begin{bmatrix} A_{11}^{-1} + (A_{11}^{-1}A_{12})P^{-1}(A_{21}A_{11}^{-1}) & -(A_{11}^{-1}A_{12})P^{-1} \\ -P^{-1}(A_{21}A_{11}^{-1}) & P^{-1} \end{bmatrix}$.

We end this section by giving another equivalent condition for a square matrix to be invertible. To do so, we need the following definition.

Definition 2.3.11. *A $n \times n$ matrix A is said to be of* FULL RANK *if rank$(A) = n$.*

Theorem 2.3.12. *Let A be a square matrix of order n. Then the following statements are equivalent.*

1. *A is invertible.*
2. *A has full rank.*
3. *The row-reduced form of A is I_n.*

Proof. $1 \Longrightarrow 2$

Let if possible rank$(A) = r < n$. Then there exists an invertible matrix P (a product of elementary matrices) such that $PA = \begin{bmatrix} B_1 & B_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where B_1 is an $r \times r$ matrix. Since A is invertible, let $A^{-1} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$, where C_1 is an $r \times n$ matrix. Then
$$P = PI_n = P(AA^{-1}) = (PA)A^{-1} = \begin{bmatrix} B_1 & B_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} B_1C_1 + B_2C_2 \\ \mathbf{0} \end{bmatrix}.$$

Thus, P has $n - r$ rows consisting of only zeros. Hence, P cannot be invertible. A contradiction. Thus, A is of full rank.

$2 \Longrightarrow 3$

Suppose A is of full rank. This implies, the row-reduced echelon form of A has all non-zero rows. But A has as many columns as rows and therefore, the last row of the row-reduced echelon form of A is $[0, 0, \ldots, 0, 1]$. Hence, the row-reduced echelon form of A is I_n.

$3 \Longrightarrow 1$

Using Theorem 2.2.5.3, the required result follows. □

2.4 Existence of Solution of $A\mathbf{x} = \mathbf{b}$

In Section 2.2, we studied the system of linear equations in which the matrix A was a square matrix. We will now use the rank of a matrix to study the system of linear equations even when A is not a square matrix. Before proceeding with our main result, we give an example for motivation and observations. Based on these observations, we will arrive at a better understanding, related to the existence and uniqueness results for the linear system $A\mathbf{x} = \mathbf{b}$.

Consider a linear system $A\mathbf{x} = \mathbf{b}$. Suppose the application of the Gauss-Jordan method has reduced the augmented matrix $[A \;\; \mathbf{b}]$ to

$$[C \;\; \mathbf{d}] = \begin{bmatrix} 1 & 0 & 2 & -1 & 0 & 0 & 2 & 8 \\ 0 & 1 & 1 & 3 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then to get the solution set, we observe the following.

Observations:

1. The number of non-zero rows in C is 4. This number is also equal to the number of non-zero rows in $[C \;\; \mathbf{d}]$. So, there are 4 leading columns/basic variables.

2. The leading terms appear in columns $1, 2, 5$ and 6. Thus, the respective variables x_1, x_2, x_5 and x_6 are the basic variables.

3. The remaining variables, x_3, x_4 and x_7 are free variables.

Hence, the solution set is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 8 - 2x_3 + x_4 - 2x_7 \\ 1 - x_3 - 3x_4 - 5x_7 \\ x_3 \\ x_4 \\ 2 + x_7 \\ 4 - x_7 \\ x_7 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \\ 0 \\ 0 \\ 2 \\ 4 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix},$$

where x_3, x_4 and x_7 are arbitrary.

Let $\mathbf{x}_0 = \begin{bmatrix} 8 \\ 1 \\ 0 \\ 0 \\ 2 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{u}_3 = \begin{bmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$.

Then it can easily be verified that $C\mathbf{x}_0 = \mathbf{d}$, and for $1 \leq i \leq 3$, $C\mathbf{u}_i = \mathbf{0}$. Hence, it follows that $A\mathbf{x}_0 = \mathbf{d}$, and for $1 \leq i \leq 3$, $A\mathbf{u}_i = \mathbf{0}$.

A similar idea is used in the proof of the next theorem and is omitted. The proof appears on page 87 as Theorem 3.3.26.

Theorem 2.4.1 (Existence/Non-Existence Result)**.** *Consider a linear system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix, and $\mathbf{x}$, $\mathbf{b}$ are vectors of orders $n \times 1$, and $m \times 1$, respectively. Suppose rank $(A) = r$ and rank$([A \;\; \mathbf{b}]) = r_a$. Then exactly one of the following statement holds:*

1. *If $r < r_a$, the linear system has no solution.*

2. *if $r_a = r$, then the linear system is consistent. Furthermore,*

   (a) *if $r = n$ then the solution set contains a unique vector $\mathbf{x}_0$ satisfying $A\mathbf{x}_0 = \mathbf{b}$.*

   (b) *if $r < n$ then the solution set has the form*

   $$\{\mathbf{x}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r} \;:\; k_i \in \mathbb{R},\ 1 \leq i \leq n-r\},$$

   where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$ for $1 \leq i \leq n-r$.

Remark 2.4.2. *Let A be an $m \times n$ matrix. Then Theorem 2.4.1 implies that*

1. *the linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if rank$(A) =$ rank$([A \;\; \mathbf{b}])$.*

2. *the vectors $\mathbf{u}_i$, for $1 \leq i \leq n-r$, correspond to each of the free variables.*

Exercise 2.4.3. *In the introduction, we gave 3 figures (see Figure 2) to show the cases that arise in the Euclidean plane (2 equations in 2 unknowns). It is well known that in the case of Euclidean space (3 equations in 3 unknowns), there*

1. *is a figure to indicate the system has a unique solution.*

2. *are 4 distinct figures to indicate the system has no solution.*

3. *are 3 distinct figures to indicate the system has infinite number of solutions.*

Determine all the figures.

2.5 Determinant

In this section, we associate a number with each square matrix. To do so, we start with the following notation. Let A be an $n \times n$ matrix. Then for each positive integers α_i's $1 \le i \le k$ and β_j's for $1 \le j \le \ell$, we write $A(\alpha_1, \ldots, \alpha_k \big| \beta_1, \ldots, \beta_\ell)$ to mean that submatrix of A, that is obtained by deleting the rows corresponding to α_i's and the columns corresponding to β_j's of A.

Example 2.5.1. *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 7 \end{bmatrix}$. *Then* $A(1|2) = \begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix}$, $A(1|3) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ *and* $A(1,2|1,3) = [4]$.

With the notations as above, we have the following inductive definition of determinant of a matrix. This definition is commonly known as the expansion of the determinant along the first row. The students with a knowledge of symmetric groups/permutations can find the definition of the determinant in Appendix 7.1.15. It is also proved in Appendix that the definition given below does correspond to the expansion of determinant along the first row.

Definition 2.5.2 (Determinant of a Square Matrix)**.** *Let A be a square matrix of order n. The determinant of A, denoted* $\det(A)$ *(or $|A|$) is defined by*

$$\det(A) = \begin{cases} a, & \text{if } A = [a] \ (n = 1), \\ \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1|j)\big), & \text{otherwise.} \end{cases}$$

Example 2.5.3. 1. *Let* $A = [-2]$. *Then* $\det(A) = |A| = -2$.

2. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *Then,* $\det(A) = |A| = a \ \big|A(1|1)\big| - b \ \big|A(1|2)\big| = ad - bc$. *For example, if* $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ *then* $\det(A) = \begin{vmatrix} 1 & 2 \\ 3 & 5 \end{vmatrix} = 1 \cdot 5 - 2 \cdot 3 = -1$.

3. *Let* $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. *Then,*

$$\begin{aligned} \det(A) = |A| &= a_{11} \det(A(1|1)) - a_{12} \det(A(1|2)) + a_{13} \det(A(1|3)) \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) \\ &\qquad + a_{13}(a_{21}a_{32} - a_{31}a_{22}) \end{aligned} \tag{2.5.1}$$

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$. *Then* $|A| = 1 \cdot \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} + 3 \cdot \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 4 - 2(3) + 3(1) = 1$.

Exercise 2.5.4. *Find the determinant of the following matrices.*

$$i) \begin{bmatrix} 1 & 2 & 7 & 8 \\ 0 & 4 & 3 & 2 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}, \quad ii) \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 2 & 0 & 5 \\ 6 & -7 & 1 & 0 \\ 3 & 2 & 0 & 6 \end{bmatrix}, \quad iii) \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}.$$

Definition 2.5.5 (Singular, Non-Singular)**.** *A matrix A is said to be a* SINGULAR *if* $\det(A) = 0$. *It is called* NON-SINGULAR *if* $\det(A) \neq 0$.

We omit the proof of the next theorem that relates the determinant of a square matrix with row operations. The interested reader is advised to go through Appendix 7.2.

Theorem 2.5.6. *Let A be an $n \times n$ matrix. If*

1. *B is obtained from A by interchanging two rows then* $\det(B) = -\det(A)$,

2. *B is obtained from A by multiplying a row by c then* $\det(B) = c\det(A)$,

3. *B is obtained from A by replacing the jth row by jth row plus c times the ith row, where $i \neq j$ then* $\det(B) = \det(A)$,

4. *all the elements of one row or column of A are* 0 *then* $\det(A) = 0$,

5. *two rows of A are equal then* $\det(A) = 0$.

6. *A is a triangular matrix then* $\det(A)$ *is product of diagonal entries.*

Since $\det(I_n) = 1$, where I_n is the $n \times n$ identity matrix, the following remark gives the determinant of the elementary matrices. The proof is omitted as it is a direct application of Theorem 2.5.6.

Remark 2.5.7. *Fix a positive integer n. Then*

1. $\det(E_{ij}) = -1$, *where E_{ij} corresponds to the interchange of the i^{th} and the j^{th} row of I_n.*

2. *For $c \neq 0$,* $\det(E_k(c)) = c$, *where $E_k(c)$ is obtained by multiplying the k^{th} row of I_n by c.*

3. *For $c \neq 0$,* $\det(E_{ij}(c)) = 1$, *where $E_{ij}(c)$ is obtained by replacing the j^{th} row of I_n by the j^{th} row of I_n plus c times the i^{th} row of I_n.*

Remark 2.5.8. *Theorem 2.5.6.1 implies that "one can also calculate the determinant by expanding along any row." Hence, the computation of determinant using the k-th row for $1 \leq k \leq n$ is given by*

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det\big(A(k|j)\big).$$

Example 2.5.9. 1. *Let* $A = \begin{bmatrix} 2 & 2 & 6 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{bmatrix}$. *Determine* $\det(A)$.

Solution: *Check that* $\begin{vmatrix} 2 & 2 & 6 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{vmatrix} \xrightarrow{R_1(2)} \begin{vmatrix} 1 & 1 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{vmatrix} \xrightarrow[R_{31}(-1)]{R_{21}(-1)} \begin{vmatrix} 1 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{vmatrix}$. *Thus, using Theorem 2.5.6,* $\det(A) = 2 \cdot 1 \cdot 2 \cdot (-1) = -4$.

2. *Let* $A = \begin{bmatrix} 2 & 2 & 6 & 8 \\ 1 & 1 & 2 & 4 \\ 1 & 3 & 2 & 6 \\ 3 & 3 & 5 & 8 \end{bmatrix}$. *Determine* $\det(A)$.

Solution: *The successive application of row operations* $R_1(2), R_{21}(-1), R_{31}(-1), R_{41}(-3), R_{23}$ *and* $R_{34}(-4)$ *and the application of Theorem 2.5.6 implies*

$$\det(A) = 2 \cdot (-1) \cdot \begin{vmatrix} 1 & 1 & 3 & 4 \\ 0 & 2 & -1 & 2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -4 \end{vmatrix} = -16.$$

Observe that the row operation $R_1(2)$ *gives* 2 *as the first product and the row operation* R_{23} *gives* -1 *as the second product.*

Remark 2.5.10. 1. *Let* $\mathbf{u}^t = (u_1, u_2)$ *and* $\mathbf{v}^t = (v_1, v_2)$ *be two vectors in* $\mathbb{R}^2$. *Consider the parallelogram on vertices* $P = (0,0)^t, Q = \mathbf{u}, R = \mathbf{u}+\mathbf{v}$ *and* $S = \mathbf{v}$ *(see Figure 3). Then Area* $(PQRS) = |u_1v_2 - u_2v_1|$, *the absolute value of* $\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$.

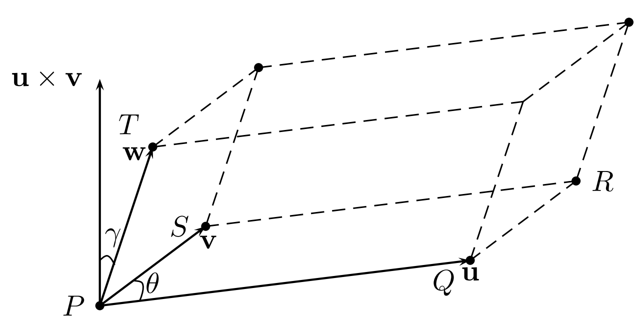

Figure 3: Parallelepiped with vertices P, Q, R *and* S *as base*

Recall the following: The dot product of $\mathbf{u}^t = (u_1, u_2)$ *and* $\mathbf{v}^t = (v_1, v_2)$, *denoted* $\mathbf{u} \bullet \mathbf{v}$, *equals* $\mathbf{u} \bullet \mathbf{v} = u_1v_1 + u_2v_2$, *and the length of a vector* $\mathbf{u}$, *denoted* $\ell(\mathbf{u})$ *equals* $\ell(\mathbf{u}) = \sqrt{u_1^2 + u_2^2}$. *Also, if* θ *is the angle between* $\mathbf{u}$ *and* $\mathbf{v}$ *then we know that* $\cos(\theta) =$

$\frac{\mathbf{u}\bullet\mathbf{v}}{\ell(\mathbf{u})\ell(\mathbf{v})}$. *Therefore*

$$\begin{aligned} Area(PQRS) &= \ell(\mathbf{u})\ell(\mathbf{v})\sin(\theta) = \ell(\mathbf{u})\ell(\mathbf{v})\sqrt{1-\left(\frac{\mathbf{u}\bullet\mathbf{v}}{\ell(\mathbf{u})\ell(\mathbf{v})}\right)^2} \\ &= \sqrt{\ell(\mathbf{u})^2+\ell(v)^2-(\mathbf{u}\bullet\mathbf{v})^2} = \sqrt{(u_1v_2-u_2v_1)^2} \\ &= |u_1v_2-u_2v_1|. \end{aligned}$$

That is, in $\mathbb{R}^2$, the determinant is $\pm$ times the area of the parallelogram.

2. *Consider Figure 3 again. Let $\mathbf{u}^t=(u_1,u_2,u_3), \mathbf{v}^t=(v_1,v_2,v_3)$ and $\mathbf{w}^t=(w_1,w_2,w_3)$ be three vectors in $\mathbb{R}^3$. Then $\mathbf{u}\bullet\mathbf{v}=u_1v_1+u_2v_2+u_3v_3$ and the cross product of $\mathbf{u}$ and $\mathbf{v}$, denoted $\mathbf{u}\times\mathbf{v}$, equals*

$$\mathbf{u}\times\mathbf{v} = (u_2v_3-u_3v_2, u_3v_1-u_1v_3, u_1v_2-u_2v_1).$$

The vector $\mathbf{u}\times\mathbf{v}$ is perpendicular to the plane containing both $\mathbf{u}$ and $\mathbf{v}$. Note that if $u_3=v_3=0$, then we can think of $\mathbf{u}$ and $\mathbf{v}$ as vectors in the XY-plane and in this case $\ell(\mathbf{u}\times\mathbf{v})=|u_1v_2-u_2v_1|=Area(PQRS)$. Hence, if γ is the angle between the vector $\mathbf{w}$ and the vector $\mathbf{u}\times\mathbf{v}$, then

$$volume\ (P) = Area(PQRS)\cdot height = |\mathbf{w}\bullet(\mathbf{u}\times\mathbf{v})| = \pm\begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

In general, for any $n\times n$ matrix A, it can be proved that $|\det(A)|$ is indeed equal to the volume of the n-dimensional parallelepiped. The actual proof is beyond the scope of this book.

Exercise 2.5.11. *In each of the questions given below, use Theorem 2.5.6 to arrive at your answer.*

1. *Let $A=\begin{bmatrix} a & b & c \\ e & f & g \\ h & j & \ell \end{bmatrix}, B=\begin{bmatrix} a & b & \alpha c \\ e & f & \alpha g \\ h & j & \alpha\ell \end{bmatrix}$ and $C=\begin{bmatrix} a & b & \alpha a+\beta b+c \\ e & f & \alpha e+\beta f+g \\ h & j & \alpha h+\beta j+\ell \end{bmatrix}$ for some complex numbers α and β. Prove that $\det(B)=\alpha\det(A)$ and $\det(C)=\det(A)$.*

2. *Let $A=\begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 1 \\ 1 & 5 & 3 \end{bmatrix}$ and $B=\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix}$. Prove that 3 divides $\det(A)$ and* $\det(B)=0$.

2.5.1 Adjoint of a Matrix

Definition 2.5.12 (Minor, Cofactor of a Matrix)**.** *The number $\det(A(i|j))$ is called the $(i,j)^{th}$ minor of A. We write $A_{ij}=\det(A(i|j))$. The $(i,j)^{th}$ cofactor of A, denoted C_{ij}, is the number $(-1)^{i+j}A_{ij}$.*

Definition 2.5.13 (Adjoint of a Matrix)**.** *Let A be an $n \times n$ matrix. The matrix $B = [b_{ij}]$ with $b_{ij} = C_{ji}$, for $1 \leq i, j \leq n$ is called the Adjoint of A, denoted $Adj(A)$.*

Example 2.5.14. *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$. *Then* $Adj(A) = \begin{bmatrix} 4 & 2 & -7 \\ -3 & -1 & 5 \\ 1 & 0 & -1 \end{bmatrix}$ *as*

$$C_{11} = (-1)^{1+1} A_{11} = 4, C_{21} = (-1)^{2+1} A_{21} = 2, \ldots, C_{33} = (-1)^{3+3} A_{33} = -1.$$

Theorem 2.5.15. *Let A be an $n \times n$ matrix. Then*

1. *for $1 \leq i \leq n$, $\sum\limits_{j=1}^{n} a_{ij}\ C_{ij} = \sum\limits_{j=1}^{n} a_{ij}(-1)^{i+j}\ A_{ij} = \det(A)$,*

2. *for $i \neq \ell$, $\sum\limits_{j=1}^{n} a_{ij}\ C_{\ell j} = \sum\limits_{j=1}^{n} a_{ij}(-1)^{\ell+j}\ A_{\ell j} = 0$, and*

3. *$A(Adj(A)) = \det(A) I_n$. Thus,*

$$\textit{whenever } \det(A) \neq 0 \ \textit{ one has } \ A^{-1} = \frac{1}{\det(A)} Adj(A). \tag{2.5.2}$$

Proof. PART 1: It directly follows from Remark 2.5.8 and the definition of the cofactor.

PART 2: Fix positive integers i, ℓ with $1 \leq i \neq \ell \leq n$. And let $B = [b_{ij}]$ be a square matrix whose ℓ^{th} row equals the i^{th} row of A and the remaining rows of B are the same as that of A.

Then by construction, the i^{th} and ℓ^{th} rows of B are equal. Thus, by Theorem 2.5.6.5, $\det(B) = 0$. As $A(\ell|j) = B(\ell|j)$ for $1 \leq j \leq n$, using Remark 2.5.8, we have

$$\begin{aligned} 0 = \det(B) &= \sum_{j=1}^{n} (-1)^{\ell+j} b_{\ell j} \det\big(B(\ell|j)\big) = \sum_{j=1}^{n} (-1)^{\ell+j} a_{ij} \det\big(B(\ell|j)\big) \\ &= \sum_{j=1}^{n} (-1)^{\ell+j} a_{ij} \det\big(A(\ell|j)\big) = \sum_{j=1}^{n} a_{ij} C_{\ell j}. \end{aligned} \tag{2.5.3}$$

This completes the proof of Part 2.

PART 3:, Using Equation (2.5.3) and Remark 2.5.8, observe that

$$\Big[A\big(\text{Adj}(A)\big)\Big]_{ij} = \sum_{k=1}^{n} a_{ik} \big(\text{Adj}(A)\big)_{kj} = \sum_{k=1}^{n} a_{ik} C_{jk} = \begin{cases} 0, & \text{if } i \neq j, \\ \det(A), & \text{if } i = j. \end{cases}$$

Thus, $A(Adj(A)) = \det(A) I_n$. Therefore, if $\det(A) \neq 0$ then $A\left(\frac{1}{\det(A)} Adj(A)\right) = I_n$. Hence, by Theorem 2.2.6,

$$A^{-1} = \frac{1}{\det(A)} Adj(A).$$

$\square$

Example 2.5.16. *For* $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$, $Adj(A) = \begin{bmatrix} -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -3 & 1 \end{bmatrix}$ *and* $\det(A) = -2$. *Thus, by Theorem 2.5.15.3,* $A^{-1} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 \\ 1/2 & 3/2 & -1/2 \end{bmatrix}$.

The next corollary is a direct consequence of Theorem 2.5.15.3 and hence the proof is omitted.

Corollary 2.5.17. *Let A be a non-singular matrix. Then*

$$\big(Adj(A)\big)A = \det(A)\ I_n \quad \textit{and} \quad \sum_{i=1}^{n} a_{ij}\ C_{ik} = \begin{cases} \det(A), & \textit{if } j = k, \\ 0, & \textit{if } j \neq k. \end{cases}$$

The next result gives another equivalent condition for a square matrix to be invertible.

Theorem 2.5.18. *A square matrix A is non-singular if and only if A is invertible.*

Proof. Let A be non-singular. Then $\det(A) \neq 0$ and hence $A^{-1} = \frac{1}{\det(A)} Adj(A)$ as .

Now, let us assume that A is invertible. Then, using Theorem 2.2.5, $A = E_1 E_2 \cdots E_k$, a product of elementary matrices. Also, by Remark 2.5.7, $\det(E_i) \neq 0$ for each i, $1 \leq i \leq k$. Thus, by a repeated application of the first three parts of Theorem 2.5.6 gives $\det(A) \neq 0$. Hence, the required result follows. $\square$

We are now ready to prove a very important result that related the determinant of product of two matrices with their determinants.

Theorem 2.5.19. *Let A and B be square matrices of order n. Then*

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

Proof. **Step 1.** Let A be non-singular. Then by Theorem 2.5.15.3, A is invertible. Hence, using Theorem 2.2.5, $A = E_1 E_2 \cdots E_k$, a product of elementary matrices. Then a repeated application of the first three parts of Theorem 2.5.6 gives

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k B) = \det(E_1)\det(E_2 \cdots E_k B) \\ &= \det(E_1)\det(E_2)\det(E_3 \cdots E_k B) \\ &= \det(E_1 E_2)\det(E_3)\det(E_4 \cdots E_k B) \\ &= \vdots \\ &= \det(E_1 E_2 \cdots E_k)\det(B) = \det(A)\det(B). \end{aligned}$$

Thus, if A is non-singular then $\det(AB) = \det(A)\det(B)$. This will be used in the second step.

Step 2. Let A be singular. Then using Theorem 2.5.18 A is not invertible. Hence, there exists an invertible matrix P such that $PA = C$, where $C = \begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}$. So, $A = P^{-1}C$ and therefore

$$\begin{aligned} \det(AB) &= \det((P^{-1}C)B) = \det(P^{-1}(CB)) = \det\left(P^{-1}\begin{bmatrix} C_1B \\ \mathbf{0} \end{bmatrix}\right) \\ &= \det(P^{-1}) \cdot \det\left(\begin{bmatrix} C_1B \\ \mathbf{0} \end{bmatrix}\right) \qquad \text{as } P^{-1} \text{ is non-singular} \\ &= \det(P) \cdot 0 = 0 = 0 \cdot \det(B) = \det(A)\det(B). \end{aligned}$$

Thus, the proof of the theorem is complete. □

The next result relates the determinant of a matrix with the determinant of its transpose. As an application of this result, determinant can be computed by expanding along any column as well.

Theorem 2.5.20. *Let A be a square matrix. Then* $\det(A) = \det(A^t)$.

Proof. If A is a non-singular, Corollary 2.5.17 gives $\det(A) = \det(A^t)$.

If A is singular, then by Theorem 2.5.18, A is not invertible. Therefore, A^t is also not invertible (as A^t is invertible implies $A^{-1} = \left((A^t)^{-1}\right)^t)$). Thus, using Theorem 2.5.18 again, $\det(A^t) = 0 = \det(A)$. Hence the required result follows. □

2.5.2 Cramer's Rule

Let A be a square matrix. Then using Theorem 2.2.10 and Theorem 2.5.18, one has the following result.

Theorem 2.5.21. *Let A be a square matrix. Then the following statements are equivalent:*

1. *A is invertible.*
2. *The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.*
3. $\det(A) \neq 0$.

Thus, $A\mathbf{x} = \mathbf{b}$ has a unique solution FOR EVERY $\mathbf{b}$ if and only if $\det(A) \neq 0$. The next theorem gives a direct method of finding the solution of the linear system $A\mathbf{x} = \mathbf{b}$ when $\det(A) \neq 0$.

Theorem 2.5.22 (Cramer's Rule)**.** *Let A be an $n \times n$ matrix. If* $\det(A) \neq 0$ *then the unique solution of the linear system $A\mathbf{x} = \mathbf{b}$ is*

$$x_j = \frac{\det(A_j)}{\det(A)}, \quad \textit{for } j = 1, 2, \ldots, n,$$

where A_j is the matrix obtained from A by replacing the jth column of A by the column vector $\mathbf{b}$.

Proof. Since $\det(A) \neq 0$, A is invertible and hence the row-reduced echelon form of A is I. Thus, for some invertible matrix P,

$$\text{RREF}[A|\mathbf{b}] = P[A|\mathbf{b}] = [PA|P\mathbf{b}] = [I|\mathbf{d}],$$

where $\mathbf{d} = A\mathbf{b}$. Hence, the system $A\mathbf{x} = \mathbf{b}$ has the unique solution $x_j = \mathbf{d}_j$, for $1 \leq j \leq n$. Also,

$$[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = I = PA = [PA[:,1], PA[:,2], \ldots, PA[:,n]].$$

Thus,

$$\begin{aligned} PA_j &= P[A[:,1], \ldots, A[:,j-1], \mathbf{b}, A[:,j+1], \ldots, A[:,n]] \\ &= [PA[:,1], \ldots, PA[:,j-1], P\mathbf{b}, PA[:,j+1], \ldots, PA[:,n]] \\ &= [\mathbf{e}_1, \ldots, \mathbf{e}_{j-1}, \mathbf{d}, \mathbf{e}_{j+1}, \ldots, \mathbf{e}_n] \end{aligned}$$

and hence $\det(PA_j) = \mathbf{d}_j$, for $1 \leq j \leq n$. Therefore,

$$\frac{\det(A_j)}{\det(A)} = \frac{\det(P)\det(A_j)}{\det(P)\det(A)} = \frac{\det(PA_j)}{\det(PA)} = \frac{\mathbf{d}_j}{1} = \mathbf{d}_j.$$

Hence, $x_j = \frac{\det(A_j)}{\det(A)}$ and the required result follows. □

In Theorem 2.5.22 $A_1 = \begin{bmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{bmatrix}$, $A_2 = \begin{bmatrix} a_{11} & b_1 & a_{13} & \cdots & a_{1n} \\ a_{21} & b_2 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & b_n & a_{n3} & \cdots & a_{nn} \end{bmatrix}$ and so on till $A_n = \begin{bmatrix} a_{11} & \cdots & a_{1n-1} & b_1 \\ a_{12} & \cdots & a_{2n-1} & b_2 \\ \vdots & \ddots & \vdots & \vdots \\ a_{1n} & \cdots & a_{nn-1} & b_n \end{bmatrix}$.

Example 2.5.23. *Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule, where* $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$ *and* $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Solution: *Check that* $\det(A) = 1$ *and* $\mathbf{x}^t = (-1, 1, 0)$ *as*

$$x_1 = \begin{vmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{vmatrix} = -1, \; x_2 = \begin{vmatrix} 1 & 1 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 1, \quad \textit{and} \;\; x_3 = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \end{vmatrix} = 0.$$

2.6 Miscellaneous Exercises

Exercise 2.6.1.

1. *Show that a triangular matrix A is invertible if and only if each diagonal entry of A is non-zero.*

2. *Let A be an orthogonal matrix. Prove that* $\det A = \pm 1$.

3. *Prove that every* 2×2 *matrix* A *satisfying* $tr(A) = 0$ *and* $\det(A) = 0$ *is a nilpotent matrix.*

4. *Let* A *and* B *be two non-singular matrices. Are the matrices* $A + B$ *and* $A - B$ *non-singular? Justify your answer.*

5. *Let* A *be an* $n \times n$ *matrix. Prove that the following statements are equivalent:*

   (a) A *is not invertible.*

   (b) $rank(A) \neq n$.

   (c) $\det(A) = 0$.

   (d) A *is not row-equivalent to* I_n.

   (e) *The homogeneous system* $A\mathbf{x} = \mathbf{0}$ *has a non-trivial solution.*

   (f) *The system* $A\mathbf{x} = \mathbf{b}$ *is either inconsistent or it is consistent and in this case it has an infinite number of solutions.*

   (g) A *is not a product of elementary matrices.*

6. *For what value(s) of* λ *does the following systems have non-trivial solutions? Also, for each value of* λ*, determine a non-trivial solution.*

   (a) $(\lambda - 2)x + y = 0, \; x + (\lambda + 2)y = 0.$

   (b) $\lambda x + 3y = 0, \; (\lambda + 6)y = 0.$

7. *Let* $x_1, x_2, \ldots, x_n$ *be fixed reals numbers and define* $A = [a_{ij}]_{n \times n}$ *with* $a_{ij} = x_i^{j-1}$. *Prove that* $\det(A) = \prod\limits_{1 \leq i < j \leq n} (x_j - x_i)$. *This matrix is usually called the Van-der monde matrix.*

8. *Let* $A = [a_{ij}]_{n \times n}$ *with* $a_{ij} = \max\{i, j\}$. *Prove that* $\det A = (-1)^{n-1} n$.

9. *Let* $A = [a_{ij}]_{n \times n}$ *with* $a_{ij} = \frac{1}{i+j-1}$. *Using induction, prove that* A *is invertible. This matrix is commonly known as the Hilbert matrix.*

10. *Solve the following system of equations by Cramer's rule.*
    i) $x + y + z - w = 1, \; x + y - z + w = 2, \; 2x + y + z - w = 7, \; x + y + z + w = 3.$
    ii) $x - y + z - w = 1, \; x + y - z + w = 2, \; 2x + y - z - w = 7, \; x - y - z + w = 3.$

11. *Suppose* $A = [a_{ij}]$ *and* $B = [b_{ij}]$ *are two* $n \times n$ *matrices with* $b_{ij} = p^{i-j} a_{ij}$ *for* $1 \leq i, j \leq n$ *for some non-zero* $p \in \mathbb{R}$. *Then compute* $\det(B)$ *in terms of* $\det(A)$.

12. *The position of an element* a_{ij} *of a determinant is called even or odd according as* $i + j$ *is even or odd. Show that*

    (a) *If all the entries in odd positions are multiplied with* -1 *then the value of the determinant doesn't change.*

    (b) *If all entries in even positions are multiplied with* -1 *then the determinant* *i. does not change if the matrix is of even order.*

ii. is multiplied by -1 if the matrix is of odd order.

13. *Let A be a Hermitian ($A^* = \overline{A^t}$) matrix. Prove that $\det A$ is a real number.*

14. *Let A be an $n \times n$ matrix. Then A is invertible if and only if $Adj(A)$ is invertible.*

15. *Let A and B be invertible matrices. Prove that $Adj(AB) = Adj(B)Adj(A)$.*

16. *Let $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be a rectangular matrix with A a square matrix of order n and $|A| \neq 0$. Then show that rank $(P) = n$ if and only if $D = CA^{-1}B$.*

2.7 Summary

In this chapter, we started with a system of linear equations $A\mathbf{x} = \mathbf{b}$ and related it to the augmented matrix $[A\,|\mathbf{b}]$. We applied row operations to $[A\,|\mathbf{b}]$ to get its row echelon form and the row-reduced echelon forms. Depending on the row echelon matrix, say $[C\,|\mathbf{d}]$, thus obtained, we had the following result:

1. If $[C\,|\mathbf{d}]$ has a row of the form $[\mathbf{0}\,|1]$ then the linear system $A\mathbf{x} = \mathbf{b}$ has not solution.

2. Suppose $[C\,|\mathbf{d}]$ does not have any row of the form $[\mathbf{0}\,|1]$ then the linear system $A\mathbf{x} = \mathbf{b}$ has at least one solution.

   (a) If the number of leading terms equals the number of unknowns then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution.

   (b) If the number of leading terms is less than the number of unknowns then the system $A\mathbf{x} = \mathbf{b}$ has an infinite number of solutions.

The following conditions are equivalent for an $n \times n$ matrix A.

1. A is invertible.

2. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

3. The row reduced echelon form of A is I.

4. A is a product of elementary matrices.

5. The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.

6. The system $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.

7. $\text{rank}(A) = n$.

8. $\det(A) \neq 0$.

Suppose the matrix A in the linear system $A\mathbf{x} = \mathbf{b}$ is of size $m \times n$. Then exactly one of the following statement holds:

1. if rank$(A) <$ rank$([A\,|\mathbf{b}])$, then the system $A\mathbf{x} = \mathbf{b}$ has no solution.
2. if rank$(A) =$ rank$([A\,|\mathbf{b}])$, then the system $A\mathbf{x} = \mathbf{b}$ is consistent. Furthermore,
   (a) if rank$(A) = n$ then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution.
   (b) if rank$(A) < n$ then the system $A\mathbf{x} = \mathbf{b}$ has an infinite number of solutions.

We also dealt with the following type of problems:

1. Solving the linear system $A\mathbf{x} = \mathbf{b}$. In the next chapter, we will see that this leads us to the question "is the vector $\mathbf{b}$ a linear combination of the columns of A"?
2. Solving the linear system $A\mathbf{x} = \mathbf{0}$. In the next chapter, we will see that this leads us to the question "are the columns of A linearly independent/dependent"?
   (a) If $A\mathbf{x} = \mathbf{0}$ has a unique solution, the trivial solution, then the columns of A are linear independent.
   (b) If $A\mathbf{x} = \mathbf{0}$ has an infinite number of solutions then the columns of A are linearly dependent.
3. Let $\mathbf{b}^t = [b_1, b_2, \ldots, b_m]$. Find conditions of the b_i's such that the linear system $A\mathbf{x} = \mathbf{b}$ always has a solution. Observe that for different choices of $\mathbf{x}$ the vector $A\mathbf{x}$ gives rise to vectors that are linear combination of the columns of A. This idea will be used in the next chapter, to get the geometrical representation of the linear span of the columns of A.

Chapter 3

Finite Dimensional Vector Spaces

3.1 Finite Dimensional Vector Spaces

Recall that the set of real numbers were denoted by $\mathbb{R}$ and the set of complex numbers were denoted by $\mathbb{C}$. Also, we wrote $\mathbb{F}$ to denote either the set $\mathbb{R}$ or the set $\mathbb{C}$.

Let A be an $m \times n$ complex matrix. Then using Theorem 2.1.5, we see that the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$, denoted V, satisfies the following properties:

1. The vector $\mathbf{0} \in V$ as $A\mathbf{0} = \mathbf{0}$.

2. If $\mathbf{x} \in V$ then $A(\alpha\mathbf{x}) = \alpha(A\mathbf{x}) = \mathbf{0}$ for all $\alpha \in \mathbb{C}$. Hence, $\alpha\mathbf{x} \in V$ for any complex number α. In particular, $-\mathbf{x} \in V$ whenever $\mathbf{x} \in V$.

3. Let $\mathbf{x}, \mathbf{y} \in V$. Then for any $\alpha, \beta \in \mathbb{C}$, $\alpha\mathbf{x}, \beta\mathbf{y} \in V$ and $A(\alpha\mathbf{x} + \beta\mathbf{y}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$. In particular, $\mathbf{x} + \mathbf{y} \in V$ and $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$. Also, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.

That is, the solution set of a homogeneous linear system satisfies some nice properties. We use these properties to define a set and devote this chapter to the study of the structure of such sets. We will also see that the set of real numbers, $\mathbb{R}$, the Euclidean plane, $\mathbb{R}^2$ and the Euclidean space, $\mathbb{R}^3$, are examples of this set. We start with the following definition.

Definition 3.1.1 (Vector Space)**.** *A vector space over* $\mathbb{F}$, *denoted* $V(\mathbb{F})$ *or in short* V *(if the field* $\mathbb{F}$ *is clear from the context), is a non-empty set, satisfying the following axioms:*

1. VECTOR ADDITION: *To every pair* $\mathbf{u}, \mathbf{v} \in V$ *there corresponds a unique element* $\mathbf{u} \oplus \mathbf{v}$ *in* V *(called the addition of vectors) such that*

   (a) $\mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u}$ *(Commutative law).*

   (b) $(\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} = \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w})$ *(Associative law).*

   (c) There is a unique element $\mathbf{0}$ *in* V *(the zero vector) such that* $\mathbf{u} \oplus \mathbf{0} = \mathbf{u}$, *for every* $\mathbf{u} \in V$ *(called* **the additive identity***).*

   (d) For every $\mathbf{u} \in V$ *there is a unique element* $-\mathbf{u} \in V$ *such that* $\mathbf{u} \oplus (-\mathbf{u}) = \mathbf{0}$ *(called* **the additive inverse***).*

2. Scalar Multiplication: *For each* $\mathbf{u} \in V$ *and* $\alpha \in \mathbb{F}$, *there corresponds a unique element* $\alpha \odot \mathbf{u}$ *in* V *(called the scalar multiplication) such that*

   (a) $\alpha \cdot (\beta \odot \mathbf{u}) = (\alpha\beta) \odot \mathbf{u}$ *for every* $\alpha, \beta \in \mathbb{F}$ *and* $\mathbf{u} \in V$.

   (b) $1 \odot \mathbf{u} = \mathbf{u}$ *for every* $\mathbf{u} \in V$, *where* $1 \in \mathbb{R}$.

3. Distributive Laws: relating vector addition with scalar multiplication *For any* $\alpha, \beta \in \mathbb{F}$ *and* $\mathbf{u}, \mathbf{v} \in V$, *the following distributive laws hold:*

   (a) $\alpha \odot (\mathbf{u} \oplus \mathbf{v}) = (\alpha \odot \mathbf{u}) \ \oplus \ (\alpha \odot \mathbf{v})$.

   (b) $(\alpha + \beta) \odot \mathbf{u} = (\alpha \odot \mathbf{u}) \ \oplus \ (\beta \odot \mathbf{u})$.

Note: the number 0 is the element of $\mathbb{F}$ whereas $\mathbf{0}$ is the zero vector.

Remark 3.1.2. *The elements of* $\mathbb{F}$ *are called* SCALARS, *and that of* V *are called* VECTORS. *If* $\mathbb{F} = \mathbb{R}$, *the vector space is called a* REAL VECTOR SPACE. *If* $\mathbb{F} = \mathbb{C}$, *the vector space is called a* COMPLEX VECTOR SPACE.

Some interesting consequences of Definition 3.1.1 is the following useful result. Intuitively, these results seem to be obvious but for better understanding of the axioms it is desirable to go through the proof.

Theorem 3.1.3. *Let* V *be a vector space over* $\mathbb{F}$. *Then*

1. $\mathbf{u} \oplus \mathbf{v} = \mathbf{u}$ *implies* $\mathbf{v} = \mathbf{0}$.

2. $\alpha \odot \mathbf{u} = \mathbf{0}$ *if and only if either* $\mathbf{u}$ *is the zero vector or* $\alpha = 0$.

3. $(-1) \odot \mathbf{u} = -\mathbf{u}$ *for every* $\mathbf{u} \in V$.

Proof. Part 1: For each $\mathbf{u} \in V$, by Axiom 3.1.1.1d there exists $-\mathbf{u} \in V$ such that $-\mathbf{u} \oplus \mathbf{u} = \mathbf{0}$. Hence, $\mathbf{u} \oplus \mathbf{v} = \mathbf{u}$ is equivalent to

$$-\mathbf{u} \oplus (\mathbf{u} \oplus \mathbf{v}) = -\mathbf{u} \oplus \mathbf{u} \Longleftrightarrow (-\mathbf{u} \oplus \mathbf{u}) \oplus \mathbf{v} = \mathbf{0} \Longleftrightarrow \mathbf{0} \oplus \mathbf{v} = \mathbf{0} \Longleftrightarrow \mathbf{v} = \mathbf{0}.$$

Part 2: As $\mathbf{0} = \mathbf{0} \oplus \mathbf{0}$, using Axiom 3.1.1.3, we have

$$\alpha \odot \mathbf{0} = \alpha \odot (\mathbf{0} \oplus \mathbf{0}) = (\alpha \odot \mathbf{0}) \ \oplus \ (\alpha \odot \mathbf{0}).$$

Thus, for any $\alpha \in \mathbb{F}$, Axiom 3.1.1.3a gives $\alpha \odot \mathbf{0} = \mathbf{0}$. In the same way,

$$0 \odot \mathbf{u} = (0 + 0) \odot \mathbf{u} = (0 \odot \mathbf{u}) \ \oplus (0 \odot \mathbf{u}).$$

Hence, using Axiom 3.1.1.3a, one has $0 \odot \mathbf{u} = \mathbf{0}$ for any $\mathbf{u} \in V$.

Now suppose $\alpha \odot \mathbf{u} = \mathbf{0}$. If $\alpha = 0$ then the proof is over. Therefore, let us assume $\alpha \neq 0$ (note that α is a real or complex number, hence $\frac{1}{\alpha}$ exists and

$$\mathbf{0} = \frac{1}{\alpha} \odot \mathbf{0} = \frac{1}{\alpha} \odot (\alpha \odot \mathbf{u}) = (\frac{1}{\alpha}\ \alpha) \odot \mathbf{u} = 1 \odot \mathbf{u} = \mathbf{u}$$

as $1 \odot \mathbf{u} = \mathbf{u}$ for every vector $\mathbf{u} \in V$. Thus, if $\alpha \neq 0$ and $\alpha \odot \mathbf{u} = \mathbf{0}$ then $\mathbf{u} = \mathbf{0}$.

Part 3: As $\mathbf{0} = 0\mathbf{u} = (1 + (-1))\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}$, one has $(-1)\mathbf{u} = -\mathbf{u}$. □

Example 3.1.4. *The readers are advised to justify the statements made in the examples given below.*

1. *Let A be an $m \times n$ matrix with complex entries and suppose rank$(A) = r \le n$. Let V denote the solution set of $A\mathbf{x} = \mathbf{0}$. Then using Theorem 2.4.1, we know that V contains at least the trivial solution, the $\mathbf{0}$ vector. Thus, check that the set V satisfies all the axioms stated in Definition 3.1.1 (some of them were proved to motivate this chapter).*

2. *The set $\mathbb{R}$ of real numbers, with the usual addition and multiplication of real numbers (i.e., $\oplus \equiv +$ and $\odot \equiv \cdot$) forms a vector space over $\mathbb{R}$.*

3. *Let $\mathbb{R}^2 = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}$. Then for $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, define*
$$(x_1, x_2) \oplus (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \quad \text{and} \quad \alpha \odot (x_1, x_2) = (\alpha x_1, \alpha x_2).$$
Then $\mathbb{R}^2$ is a real vector space.

4. *Let $\mathbb{R}^n = \{(a_1, a_2, \ldots, a_n) : a_i \in \mathbb{R}, 1 \le i \le n\}$ be the set of n-tuples of real numbers. For $\mathbf{u} = (a_1, \ldots, a_n)$, $\mathbf{v} = (b_1, \ldots, b_n)$ in V and $\alpha \in \mathbb{R}$, we define*
$$\mathbf{u} \oplus \mathbf{v} = (a_1 + b_1, \ldots, a_n + b_n) \quad \text{and} \quad \alpha \odot \mathbf{u} = (\alpha a_1, \ldots, \alpha a_n)$$
(called component wise operations). Then V is a real vector space. This vector space $\mathbb{R}^n$ is called **the real vector space of n-tuples**.

   Recall that the symbol i represents the complex number $\sqrt{-1}$.

5. *Consider the set $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ of complex numbers and let $\mathbf{z}_1 = x_1 + iy_1$ and $\mathbf{z}_2 = x_2 + iy_2$. Define*
$$\mathbf{z}_1 \oplus \mathbf{z}_2 = (x_1 + x_2) + i(y_1 + y_2), \quad \text{and}$$
   (a) *for any $\alpha \in \mathbb{R}$, define $\alpha \odot \mathbf{z}_1 = (\alpha x_1) + i(\alpha y_1)$. Then $\mathbb{C}$ is a real vector space as the scalars are the real numbers.*
   (b) *$(\alpha + i\beta) \odot (x_1 + iy_1) = (\alpha x_1 - \beta y_1) + i(\alpha y_1 + \beta x_1)$ for any $\alpha + i\beta \in \mathbb{C}$. Here, the scalars are complex numbers and hence $\mathbb{C}$ forms a complex vector space.*

6. *Let $\mathbb{C}^n = \{(z_1, z_2, \ldots, z_n) : z_i \in \mathbb{C}, 1 \le i \le n\}$. For $(z_1, \ldots, z_n), (w_1, \ldots, w_n) \in \mathbb{C}^n$ and $\alpha \in \mathbb{F}$, define*
$$\begin{aligned} (z_1, \ldots, z_n) \oplus (w_1, \ldots, w_n) &= (z_1 + w_1, \ldots, z_n + w_n), \quad \text{and} \\ \alpha \odot (z_1, \ldots, z_n) &= (\alpha z_1, \ldots, \alpha z_n). \end{aligned}$$
Then it can be verified that $\mathbb{C}^n$ forms a vector space over $\mathbb{C}$ (called complex vector space) as well as over $\mathbb{R}$ (called real vector space). Whenever there is no mention of scalars, it will always be assumed to be $\mathbb{C}$, the complex numbers.

Remark 3.1.5. *If the scalars are* $\mathbb{C}$ *then* $i(1,0) = (i,0)$ *is allowed. Whereas, if the scalars are* $\mathbb{R}$ *then* $i(1,0) \neq (i,0)$.

7. *Fix a positive integer* n *and let* $\mathcal{P}_n(\mathbb{R})$ *denote the set of all polynomials in* x *of degree* $\leq n$ *with coefficients from* $\mathbb{R}$. *Algebraically,*

$$\mathcal{P}_n(\mathbb{R}) = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : a_i \in \mathbb{R}, 0 \leq i \leq n\}.$$

Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n \in \mathcal{P}_n(\mathbb{R})$ *for some* $a_i, b_i \in \mathbb{R}$, $0 \leq i \leq n$. *It can be verified that* $\mathcal{P}_n(\mathbb{R})$ *is a real vector space with the addition and scalar multiplication defined by*

$$\begin{aligned} f(x) \oplus g(x) &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n, \quad \textit{and} \\ \alpha \odot f(x) &= \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_n x^n \quad \textit{for } \alpha \in \mathbb{R}. \end{aligned}$$

8. *Let* $\mathcal{P}(\mathbb{R})$ *be the set of all polynomials with real coefficients. As any polynomial* $a_0 + a_1x + \cdots + a_mx^m$ *also equals* $a_0 + a_1x + \cdots + a_mx^m + 0 \cdot x^{m+1} + \cdots + 0 \cdot x^p$, *whenever* $p > m$, *let* $f(x) = a_0 + a_1x + \cdots + a_px^p, g(x) = b_0 + b_1x + \cdots + b_px^p \in \mathcal{P}(\mathbb{R})$ *for some* $a_i, b_i \in \mathbb{R}$, $0 \leq i \leq p$. *So, with vector addition and scalar multiplication is defined below (called coefficient-wise),* $\mathcal{P}(\mathbb{R})$ *forms a real vector space.*

$$\begin{aligned} f(x) \oplus g(x) &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_p + b_p)x^p \quad \textit{and} \\ \alpha \odot f(x) &= \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_p x^p \quad \textit{for } \alpha \in \mathbb{R}. \end{aligned}$$

9. *Let* $\mathcal{P}(\mathbb{C})$ *be the set of all polynomials with complex coefficients. Then with respect to vector addition and scalar multiplication defined coefficient-wise, the set* $\mathcal{P}(\mathbb{C})$ *forms a vector space.*

10. *Let* $V = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$. *This is* NOT *a vector space under usual operations of addition and scalar multiplication (why?). But* $\mathbb{R}^+$ *is a real vector space with* 1 *as the additive identity if we define vector addition and scalar multiplication by*

$$\mathbf{u} \oplus \mathbf{v} = \mathbf{u} \cdot \mathbf{v} \quad \textit{and} \quad \alpha \odot \mathbf{u} = \mathbf{u}^\alpha \quad \textit{for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^+ \textit{ and } \alpha \in \mathbb{R}.$$

11. *Let* $V = \{(x,y) : x, y \in \mathbb{R}\}$. *For any* $\alpha \in \mathbb{R}$ *and* $\mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2) \in V$, *let*

$$\mathbf{x} \oplus \mathbf{y} = (x_1 + y_1 + 1, x_2 + y_2 - 3) \quad \textit{and} \quad \alpha \odot \mathbf{x} = (\alpha x_1 + \alpha - 1, \alpha x_2 - 3\alpha + 3).$$

Then V *is a real vector space with* $(-1, 3)$ *as the additive identity.*

12. *Let* $M_2(\mathbb{C})$ *denote the set of all* 2×2 *matrices with complex entries. Then* $M_2(\mathbb{C})$ *forms a vector space with vector addition and scalar multiplication defined by*

$$A \oplus B = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \oplus \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 & a_2 + b_2 \\ a_3 + b_3 & a_4 + b_4 \end{bmatrix}, \quad \alpha \odot A = \begin{bmatrix} \alpha a_1 & \alpha a_2 \\ \alpha a_3 & \alpha a_4 \end{bmatrix}.$$

13. *Fix positive integers m and n and let $M_{m\times n}(\mathbb{C})$ denote the set of all $m \times n$ matrices with complex entries. Then $M_{m\times n}(\mathbb{C})$ is a vector space with vector addition and scalar multiplication defined by*

$$A \oplus B = [a_{ij}] \oplus [b_{ij}] = [a_{ij} + b_{ij}], \qquad \alpha \odot A = \alpha \odot [a_{ij}] = [\alpha a_{ij}].$$

In case $m = n$, the vector space $M_{m\times n}(\mathbb{C})$ will be denoted by $M_n(\mathbb{C})$.

14. *Let $C([-1,1])$ be the set of all real valued continuous functions on the interval $[-1,1]$. Then $C([-1,1])$ forms a real vector space if for all $x \in [-1,1]$, we define*

$$\begin{aligned} (f \oplus g)(x) &= f(x) + g(x) \text{ for all } f, g \in C([-1,1]) \text{ and} \\ (\alpha \odot f)(x) &= \alpha f(x) \text{ for all } \alpha \in \mathbb{R} \text{ and } f \in C([-1,1]). \end{aligned}$$

15. *Let V and W be vector spaces over $\mathbb{F}$, with operations $(+, \bullet)$ and $(\oplus, \odot)$, respectively. Let $V \times W = \{(\mathbf{v}, \mathbf{w}) : \mathbf{v} \in V, \mathbf{w} \in W\}$. Then $V \times W$ forms a vector space over $\mathbb{F}$, if for every $(\mathbf{v}_1, \mathbf{w}_1), (\mathbf{v}_2, \mathbf{w}_2) \in V \times W$ and $\alpha \in \mathbb{R}$, we define*

$$\begin{aligned} (\mathbf{v}_1, \mathbf{w}_1) \oplus' (\mathbf{v}_2, \mathbf{w}_2) &= (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 \oplus \mathbf{w}_2), \text{ and} \\ \alpha \circ (\mathbf{v}_1, \mathbf{w}_1) &= (\alpha \bullet \mathbf{v}_1, \alpha \odot \mathbf{w}_1). \end{aligned}$$

$\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{w}_1 \oplus \mathbf{w}_2$ on the right hand side mean vector addition in V and W, respectively. Similarly, $\alpha \bullet \mathbf{v}_1$ and $\alpha \odot \mathbf{w}_1$ correspond to scalar multiplication in V and W, respectively.

From now on, we will use '$\mathbf{u} + \mathbf{v}$' for '$\mathbf{u} \oplus \mathbf{v}$' and '$\alpha \cdot \mathbf{u}$ or $\alpha\mathbf{u}$' for '$\alpha \odot \mathbf{u}$'.

Exercise 3.1.6. 1. *Verify all the axioms are satisfied in all the examples of vector spaces considered in Example 3.1.4.*

2. *Prove that the set $M_{m\times n}(\mathbb{R})$ for fixed positive integers m and n forms a real vector space with usual operations of matrix addition and scalar multiplication.*

3. *Let $V = \{(x, y) : x, y \in \mathbb{R}^2\}$. For $\mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2) \in V$, define*

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2) \text{ and } \alpha\mathbf{x} = (\alpha x_1, 0)$$

for all $\alpha \in \mathbb{R}$. Is V a vector space? Give reasons for your answer.

4. *Let $a, b \in \mathbb{R}$ with $a < b$. Then prove that $C([a,\ b])$, the set of all complex valued continuous functions on $[a,\ b]$ forms a vector space if for all $x \in [a,\ b]$, we define*

$$\begin{aligned} (f \oplus g)(x) &= f(x) + g(x) \text{ for all } f, g \in C([a,\ b]) \text{ and} \\ (\alpha \odot f)(x) &= \alpha f(x) \text{ for all } \alpha \in \mathbb{R} \text{ and } f \in C([a,\ b]). \end{aligned}$$

5. *Prove that $C(\mathbb{R})$, the set of all real valued continuous functions on $\mathbb{R}$ forms a vector space if for all $x \in \mathbb{R}$, we define*

$$\begin{aligned} (f \oplus g)(x) &= f(x) + g(x) \text{ for all } f, g \in C(\mathbb{R}) \text{ and} \\ (\alpha \odot f)(x) &= \alpha f(x) \text{ for all } \alpha \in \mathbb{R} \text{ and } f \in C(\mathbb{R}). \end{aligned}$$

3.1.1 Subspaces

Definition 3.1.7 (Vector Subspace)**.** *Let S be a* NON-EMPTY SUBSET *of V. The set S over $\mathbb{F}$ is said to be a subspace of $V(\mathbb{F})$ if S in itself is a vector space, where the vector addition and scalar multiplication are the same as that of $V(\mathbb{F})$.*

Example 3.1.8. *1. Let $V(\mathbb{F})$ be a vector space. Then the sets given below are subspaces of V. They are called* TRIVIAL SUBSPACES.

   (a) $S = \{\mathbf{0}\}$, consisting only of the zero vector $\mathbf{0}$ and

   (b) $S = V$, the whole space.

2. *Let $S = \{(x, y, z) \in \mathbb{R}^3 : x + 2y - z = 0\}$. Then S is a subspace of $\mathbb{R}^3$ (S is a plane in $\mathbb{R}^3$ passing through the origin).*

3. *Let $S = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0, x - y - z = 0\}$. Then S is a subspace of $\mathbb{R}^3$ (S is a line in $\mathbb{R}^3$ passing through the origin).*

4. *Let $S = \{(x, y, z) \in \mathbb{R}^3 : z - 3x = 0\}$. Then S is a subspace of $\mathbb{R}^3$.*

5. *The vector space $\mathcal{P}_n(\mathbb{R})$ is a subspace of the vector space $\mathcal{P}(\mathbb{R})$.*

6. *Prove that $S = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 3\}$ is not a subspace of $\mathbb{R}^3$ (S is still a plane in $\mathbb{R}^3$ but it does not pass through the origin).*

7. *Prove that $W = \{(x, 0) \in \mathbb{R}^2 \ : x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.*

8. *Let $W = \{(x, 0) \in V \ : x \in \mathbb{R}\}$, where V is the vector space of Example 3.1.4.11. Then $(x, 0) \oplus (y, 0) = (x + y + 1, -3) \notin W$. Hence W is not a subspace of V but $S = \{(x, 3) \ : x \in \mathbb{R}\}$ is a subspace of V. Note that the zero vector $(-1, 3) \in V$.*

9. *Let $W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{C}) \ : a = \overline{d} \right\}$. Then the condition $a = \overline{d}$ forces us to have $\alpha = \overline{\alpha}$ for any scalar $\alpha \in \mathbb{C}$. Hence,*

   (a) W is not a vector subspace of the complex vector space $M_2(\mathbb{C})$, but

   (b) W is a vector subspace of the real vector space $M_2(\mathbb{C})$.

We are now ready to prove a very important result in the study of vector subspaces. This result basically tells us that if we want to prove that a non-empty set W is a subspace of a vector space $V(\mathbb{F})$ then we just need to verify only one condition. That is, we don't have to prove all the axioms stated in Definition 3.1.1.

Theorem 3.1.9. *Let $V(\mathbb{F})$ be a vector space and let W be a non-empty subset of V. Then W is a subspace of V if and only if $\alpha\mathbf{u} + \beta\mathbf{v} \in W$ whenever $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in W$.*

Proof. Let W be a subspace of V and let $\mathbf{u}, \mathbf{v} \in W$. Then for every $\alpha, \beta \in \mathbb{F}$, $\alpha\mathbf{u}, \beta\mathbf{v} \in W$ and hence $\alpha\mathbf{u} + \beta\mathbf{v} \in W$.

Now, let us assume that $\alpha\mathbf{u} + \beta\mathbf{v} \in W$ whenever $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in W$. Need to show, W is a subspace of V. To do so, observe the following:

1. Taking $\alpha = 1$ and $\beta = 1$, we see that $\mathbf{u} + \mathbf{v} \in W$ for every $\mathbf{u}, \mathbf{v} \in W$.
2. Taking $\alpha = 0$ and $\beta = 0$, we see that $\mathbf{0} \in W$.
3. Taking $\beta = 0$, we see that $\alpha\mathbf{u} \in W$ for every $\alpha \in \mathbb{F}$ and $\mathbf{u} \in W$ and hence using Theorem 3.1.3.3, $-\mathbf{u} = (-1)\mathbf{u} \in W$ as well.
4. The commutative and associative laws of vector addition hold as they hold in V.
5. The axioms related with scalar multiplication and the distributive laws also hold as they hold in V.

Thus, we have the required result. □

Exercise 3.1.10. *1. Determine all the subspaces of $\mathbb{R}, \mathbb{R}^2$ and $\mathbb{R}^3$.*

2. Prove that a line in $\mathbb{R}^2$ is a subspace if and only if it passes through $(0,0) \in \mathbb{R}^2$.

3. Let $V = \{(a,b) : a, b \in \mathbb{R}\}$. Is V a vector space over $\mathbb{R}$ if $(a,b) \oplus (c,d) = (a+c, 0)$ and $\alpha \odot (a,b) = (\alpha a, 0)$? Give reasons for your answer.

4. Let $V = \mathbb{R}$. Define $x \oplus y = x - y$ and $\alpha \odot x = -\alpha x$. Which vector space axioms are not satisfied here?

5. Which of the following are correct statements (why!)?

(a) $S = \{(x,y,z) \in \mathbb{R}^3 : z = x^2\}$ is a subspace of $\mathbb{R}^3$.

(b) $S = \{\alpha\mathbf{x} : \alpha \in \mathbb{F}\}$ forms a vector subspace of $V(\mathbb{F})$ for each fixed $\mathbf{x} \in V$.

(c) $S = \{\alpha(1,1,1) + \beta(1,-1,0) : \alpha, \beta \in \mathbb{R}\}$ is a vector subspace of $\mathbb{R}^3$.

(d) All the sets given below are subspaces of $C([-1,1])$ (see Example 3.1.4.14).

i. $W = \{f \in C([-1,1]) : f(1/2) = 0\}$.

ii. $W = \{f \in C([-1,1]) : f(0) = 0, f(1/2) = 0\}$.

iii. $W = \{f \in C([-1,1]) : f(-1/2) = 0, f(1/2) = 0\}$.

iv. $W = \{f \in C([-1,1]) : f'(\frac{1}{4})$ exists $\}$.

(e) All the sets given below are subspaces of $\mathcal{P}(\mathbb{R})$?

i. $W = \{f(x) \in \mathcal{P}(\mathbb{R}) : \deg(f(x)) = 3\}$.

ii. $W = \{f(x) \in \mathcal{P}(\mathbb{R}) : \deg(f(x)) = 0\}$.

iii. $W = \{f(x) \in \mathcal{P}(\mathbb{R}) : f(1) = 0\}$.

iv. $W = \{f(x) \in \mathcal{P}(\mathbb{R}) : f(0) = 0, f(1) = 0\}$.

(f) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ is a subspace of $\mathbb{R}^3$.

(g) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$. Then $\{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$ is a subspace of $\mathbb{R}^3$.

6. Which of the following are subspaces of $\mathbb{R}^n(\mathbb{R})$?

(a) $\{(x_1, x_2, \ldots, x_n) : x_1 \geq 0\}$.
(b) $\{(x_1, x_2, \ldots, x_n) : x_1 + 2x_2 = 4x_3\}$.
(c) $\{(x_1, x_2, \ldots, x_n) : x_1$ *is rational* $\}$.
(d) $\{(x_1, x_2, \ldots, x_n) : x_1 = x_3^2\}$.
(e) $\{(x_1, x_2, \ldots, x_n) :$ *either* x_1 *or* x_2 *or both are* $0\}$.
(f) $\{(x_1, x_2, \ldots, x_n) : |x_1| \leq 1\}$.

7. *Which of the following are subspaces of* $i)\mathbb{C}^n(\mathbb{R})$ $ii)\mathbb{C}^n(\mathbb{C})$*?*

(a) $\{(z_1, z_2, \ldots, z_n) : z_1$ *is real* $\}$.
(b) $\{(z_1, z_2, \ldots, z_n) : z_1 + z_2 = \overline{z_3}\}$.
(c) $\{(z_1, z_2, \ldots, z_n) : \mid z_1 \mid = \mid z_2 \mid\}$.

8. *Let* $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$. *Are the sets given below subspaces of* $\mathbb{R}^3$*?*

(a) $W = \{\mathbf{x}^t \in \mathbb{R}^3 : A\mathbf{x} = \mathbf{0}\}$.
(b) $W = \{\mathbf{b}^t \in \mathbb{R}^3 :$ *there exists* $\mathbf{x}^t \in \mathbb{R}^3$ *with* $A\mathbf{x} = \mathbf{b}\}$.
(c) $W = \{\mathbf{x}^t \in \mathbb{R}^3 : \mathbf{x}^t A = \mathbf{0}\}$.
(d) $W = \{\mathbf{b}^t \in \mathbb{R}^3 :$ *there exists* $\mathbf{x}^t \in \mathbb{R}^3$ *with* $\mathbf{x}^t A = \mathbf{b}^t\}$.

9. *Fix a positive integer* n. *Then* $M_n(\mathbb{R})$ *is a real vector space with usual operations of matrix addition and scalar multiplication. Prove that the sets* $W \subset M_n(\mathbb{R})$, *given below, are subspaces of* $M_n(\mathbb{R})$.

(a) $W = \{A : A^t = A\}$, *the set of symmetric matrices.*
(b) $W = \{A : A^t = -A\}$, *the set of skew-symmetric matrices.*
(c) $W = \{A : A$ *is an upper triangular matrix*$\}$.
(d) $W = \{A : A$ *is a lower triangular matrix*$\}$.
(e) $W = \{A : A$ *is a diagonal matrix*$\}$.
(f) $W = \{A : trace(A) = 0\}$.
(g) $W = \{A = (a_{ij}) : a_{11} + a_{22} = 0\}$.
(h) $W = \{A = (a_{ij}) : a_{21} + a_{22} + \cdots + a_{2n} = 0\}$.

10. *Fix a positive integer* n. *Then* $M_n(\mathbb{C})$ *is a complex vector space with usual operations of matrix addition and scalar multiplication. Are the sets* $W \subset M_n(\mathbb{C})$, *given below, subspaces of* $M_n(\mathbb{C})$*? Give reasons.*

(a) $W = \{A : A^* = A\}$, *the set of Hermitian matrices.*
(b) $W = \{A : A^* = -A\}$, *the set of skew-Hermitian matrices.*

(c) $W = \{A : A$ *is an upper triangular matrix*$\}$.

(d) $W = \{A : A$ *is a lower triangular matrix*$\}$.

(e) $W = \{A : A$ *is a diagonal matrix*$\}$.

(f) $W = \{A : trace(A) = 0\}$.

(g) $W = \{A = (a_{ij}) : a_{11} + \overline{a_{22}} = 0\}$.

(h) $W = \{A = (a_{ij}) : a_{21} + a_{22} + \cdots + a_{2n} = 0\}$.

What happens if $M_n(\mathbb{C})$ *is a real vector space?*

11. *Prove that the following sets are not subspaces of* $M_n(\mathbb{R})$.

(a) $G = \{A \in M_n(\mathbb{R}) : \det(A) = 0\}$.

(b) $G = \{A \in M_n(\mathbb{R}) : \det(A) \neq 0\}$.

(c) $G = \{A \in M_n(\mathbb{R}) : \det(A) = 1\}$.

3.1.2 Linear Span

Definition 3.1.11 (Linear Combination)**.** *Let* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ *be a collection of vectors from a vector space* $V(\mathbb{F})$. *A vector* $\mathbf{u} \in V$ *is said to be a linear combination of the vectors* $\mathbf{u}_1, \ldots, \mathbf{u}_n$ *if we can find scalars* $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ *such that* $\mathbf{u} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_n\mathbf{u}_n$.

Example 3.1.12. 1. *Is* $(4,5,5)$ *a linear combination of* $(1,0,0)$, $(2,1,0)$, *and* $(3,3,1)$*?*
Solution: *The vector* $(4,5,5)$ *is a linear combination if the linear system*

$$a(1,0,0) + b(2,1,0) + c(3,3,1) = (4,5,5) \tag{3.1.1}$$

in the unknowns $a, b, c \in \mathbb{R}$ *has a solution. The augmented matrix of Equation (3.1.1) equals* $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 1 & 5 \end{bmatrix}$ *and it has the solution* $\alpha_1 = 4, \alpha_2 = -10$ *and* $\alpha_3 = 5$.

2. *Is* $(4,5,5)$ *a linear combination of the vectors* $(1,2,3), (-1,1,4)$ *and* $(3,3,2)$?
Solution: *The vector* $(4,5,5)$ *is a linear combination if the linear system*

$$a(1,2,3) + b(-1,1,4) + c(3,3,2) = (4,5,5) \tag{3.1.2}$$

in the unknowns $a, b, c \in \mathbb{R}$ *has a solution. The row reduced echelon form of the augmented matrix of Equation (3.1.2) equals* $\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. *Thus, one has an infinite number of solutions. For example,* $(4,5,5) = 3(1,2,3) - (-1,1,4)$.

3. *Is* $(4,5,5)$ *a linear combination of the vectors* $(1,2,1), (1,0,-1)$ *and* $(1,1,0)$.
Solution: *The vector* $(4,5,5)$ *is a linear combination if the linear system*

$$a(1,2,1) + b(1,0,-1) + c(1,1,0) = (4,5,5) \tag{3.1.3}$$

in the unknowns $a, b, c \in \mathbb{R}$ *has a solution. An application of Gauss elimination method to Equation (3.1.3) gives* $\begin{bmatrix} 1 & 1 & 1 & 4 \\ 0 & 1 & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$. *Thus, Equation (3.1.3) has no solution and hence* $(4, 5, 5)$ *is not a linear combination of the given vectors.*

Exercise 3.1.13. *1. Prove that every* $\mathbf{x} \in \mathbb{R}^3$ *is a unique linear combination of the vectors* $(1,0,0)$, $(2,1,0)$, *and* $(3,3,1)$.

2. *Find condition(s) on* x, y *and* z *such that* (x, y, z) *is a linear combination of* $(1,2,3), (-1,1,4)$ *and* $(3,3,2)$?

3. *Find condition(s) on* x, y *and* z *such that* (x, y, z) *is a linear combination of the vectors* $(1,2,1), (1,0,-1)$ *and* $(1,1,0)$.

Definition 3.1.14 (Linear Span)**.** *Let* $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *be a non-empty subset of a vector space* $V(\mathbb{F})$. *The linear span of* S *is the set defined by*

$$L(S) \quad = \quad \{\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_n \mathbf{u}_n : \alpha_i \in \mathbb{F}, 1 \le i \le n\}$$

If S *is an empty set we define* $L(S) = \{\mathbf{0}\}$.

Example 3.1.15. *1. Let* $S = \{(1,0),(0,1)\} \subset \mathbb{R}^2$. *Determine* $L(S)$.
Solution: *By definition, the required linear span is*

$$L(S) = \{a(1,0) + b(0,1) : a, b \in \mathbb{R}\} = \{(a,b) : a, b \in \mathbb{R}\} = \mathbb{R}^2. \tag{3.1.4}$$

2. *For each* $S \subset \mathbb{R}^3$, *determine the geometrical representation of* $L(S)$.

   (a) $S = \{(1,1,1),(2,1,3)\}$.
   Solution: *By definition, the required linear span is*

$$L(S) = \{a(1,1,1) + b(2,1,3) : a, b \in \mathbb{R}\} = \{(a+2b, a+b, a+3b) : a, b \in \mathbb{R}\}. \tag{3.1.5}$$

   Note that finding all vectors of the form $(a+2b, a+b, a+3b)$ *is equivalent to finding conditions on* x, y *and* z *such that* $(a+2b, a+b, a+3b) = (x, y, z)$, *or equivalently, the system*

$$a + 2b = x, a + b = y, a + 3b = z$$

   always has a solution. Check that the row reduced form of the augmented matrix equals $\begin{bmatrix} 1 & 0 & 2y - x \\ 0 & 1 & x - y \\ 0 & 0 & z + y - 2x \end{bmatrix}$. *Thus, we need* $2x - y - z = 0$ *and hence*

$$L(S) = \{a(1,1,1) + b(2,1,3) : a, b \in \mathbb{R}\} = \{(x,y,z) \in \mathbb{R}^3 : 2x - y - z = 0\}. \tag{3.1.6}$$

   Equation (3.1.5) is called an algebraic representation of $L(S)$ *whereas Equation (3.1.6) gives its geometrical representation as a subspace of* $\mathbb{R}^3$.

(b) $S = \{(1,2,1),(1,0,-1),(1,1,0)\}$.
Solution: *As in Example 3.1.15.2, we need to find condition(s) on x, y, z such that the linear system*

$$a(1,2,1) + b(1,0,-1) + c(1,1,0) = (x,y,z) \tag{3.1.7}$$

in the unknowns a, b, c is always consistent. An application of Gauss elimination method to Equation (3.1.7) gives $\begin{bmatrix} 1 & 1 & 1 & x \\ 0 & 1 & \frac{1}{2} & \frac{2x-y}{3} \\ 0 & 0 & 0 & x-y+z \end{bmatrix}$. *Thus,*

$$L(S) = \{(x,y,z) : \ x - y + z = 0\}.$$

(c) $S = \{(1,2,3),(-1,1,4),(3,3,2)\}$.
Solution: *We need to find condition(s) on x, y, z such that the linear system*

$$a(1,2,3) + b(-1,1,4) + c(3,3,2) = (x,y,z)$$

in the unknowns a, b, c is always consistent. An application of Gauss elimination method gives $5x - 7y + 3z = 0$ as the required condition. Thus,

$$L(S) = \{(x,y,z) : \ 5x - 7y + 3z = 0\}.$$

3. $S = \{(1,2,3,4),(-1,1,4,5),(3,3,2,3)\} \subset \mathbb{R}^4$. *Determine $L(S)$.*
Solution: *The readers are advised to show that*

$$L(S) = \{(x,y,z,w) : 2x - 3y + w = 0, \ 5x - 7y + 3z = 0\}.$$

Exercise 3.1.16. *For each of the sets S, determine the geometric representation of $L(S)$.*

1. $S = \{-1\} \subset \mathbb{R}$.
2. $S = \{\frac{1}{10^4}\} \subset \mathbb{R}$.
3. $S = \{\sqrt{15}\} \subset \mathbb{R}$.
4. $S = \{(1,0,0),(0,1,0),(0,0,1)\} \subset \mathbb{R}^3$.
5. $S = \{(1,0,1),(0,1,0),(3,0,3)\} \subset \mathbb{R}^3$.
6. $S = \{(1,0,1),(1,1,0),(3,-4,3)\} \subset \mathbb{R}^3$.
7. $S = \{(1,2,1),(2,0,1),(1,1,1)\} \subset \mathbb{R}^3$.
8. $S = \{(1,0,1,1),(0,1,0,1),(3,0,3,1)\} \subset \mathbb{R}^4$.

Definition 3.1.17 (Finite Dimensional Vector Space)**.** *A vector space $V(\mathbb{F})$ is said to be finite dimensional if we can find a subset S of V, having finite number of elements, such that $V = L(S)$. If such a subset does not exist then V is called an infinite dimensional vector space.*

Example 3.1.18.

1. *The set* $\{(1,2),(2,1)\}$ *spans* $\mathbb{R}^2$ *and hence* $\mathbb{R}^2$ *is a finite dimensional vector space.*

2. *The set* $\{1, 1+x, 1-x+x^2, x^3, x^4, x^5\}$ *spans* $\mathcal{P}_5(\mathbb{C})$ *and hence* $\mathcal{P}_5(\mathbb{C})$ *is a finite dimensional vector space.*

3. *Fix a positive integer* n *and consider the vector space* $\mathcal{P}_n(\mathbb{R})$. *Then* $\mathcal{P}_n(\mathbb{C})$ *is a finite dimensional vector space as* $\mathcal{P}_n(\mathbb{C}) = L(\{1, x, x^2, \ldots, x^n\})$.

4. *Recall* $\mathcal{P}(\mathbb{C})$, *the vector space of all polynomials with complex coefficients. Since degree of a polynomial can be any large positive integer,* $\mathcal{P}(\mathbb{C})$ *cannot be a finite dimensional vector space. Indeed, checked that* $\mathcal{P}(\mathbb{C}) = L(\{1, x, x^2, \ldots, x^n, \ldots\})$.

Lemma 3.1.19 (Linear Span is a Subspace)**.** *Let* S *be a non-empty subset of a vector space* $V(\mathbb{F})$. *Then* $L(S)$ *is a subspace of* $V(\mathbb{F})$.

Proof. By definition, $S \subset L(S)$ and hence $L(S)$ is non-empty subset of V. Let $\mathbf{u}, \mathbf{v} \in L(S)$. Then, there exist a positive integer n, vectors $\mathbf{w}_i \in S$ and scalars $\alpha_i, \beta_i \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 + \cdots + \alpha_n\mathbf{w}_n$ and $\mathbf{v} = \beta_1\mathbf{w}_1 + \beta_2\mathbf{w}_2 + \cdots + \beta_n\mathbf{w}_n$. Hence,

$$a\mathbf{u} + b\mathbf{v} = (a\alpha_1 + b\beta_1)\mathbf{w}_1 + \cdots + (a\alpha_n + b\beta_n)\mathbf{w}_n \in L(S)$$

for every $a, b \in \mathbb{F}$ as $a\alpha_i + b\beta_i \in \mathbb{F}$ for $i = 1, \ldots, n$. Thus using Theorem 3.1.9, $L(S)$ is a vector subspace of $V(\mathbb{F})$. □

Remark 3.1.20. *Let* W *be a subspace of a vector space* $V(\mathbb{F})$. *If* $S \subset W$ *then* $L(S)$ *is a subspace of* W *as* W *is a vector space in its own right.*

Theorem 3.1.21. *Let* S *be a non-empty subset of a vector space* V. *Then* $L(S)$ *is the smallest subspace of* V *containing* S.

Proof. For every $\mathbf{u} \in S$, $\mathbf{u} = 1.\mathbf{u} \in L(S)$ and hence $S \subseteq L(S)$. To show $L(S)$ is the smallest subspace of V containing S, consider any subspace W of V containing S. Then by Remark 3.1.20, $L(S) \subseteq W$ and hence the result follows. □

Exercise 3.1.22.

1. *Find all the vector subspaces of* $\mathbb{R}^2$ *and* $\mathbb{R}^3$.

2. *Prove that* $\{(x,y,z) \in \mathbb{R}^3 : ax + by + cz = d\}$ *is a subspace of* $\mathbb{R}^3$ *if and only if* $d = 0$.

3. *Let* W *be a set that consists of all polynomials of degree* 5. *Prove that* W *is not a subspace* $\mathcal{P}(\mathbb{R})$.

4. *Determine all vector subspaces of* V, *the vector space in Example 3.1.4.11.*

5. *Let* P *and* Q *be two subspaces of a vector space* V.

   (a) *Prove that* $P \cap Q$ *is a subspace of* V.

   (b) *Give examples of* P *and* Q *such that* $P \cup Q$ *is not a subspace of* V.

   (c) *Determine conditions on* P *and* Q *such that* $P \cup Q$ *a subspace of* V?

(d) *Define* $P+Q=\{\mathbf{u}+\mathbf{v}:\mathbf{u}\in P,\mathbf{v}\in Q\}$. *Prove that* $P+Q$ *is a subspace of* V.

(e) *Prove that* $L(P\cup Q)=P+Q$.

6. *Let* $x_1=(1,0,0),\ x_2=(1,1,0),\ x_3=(1,2,0),\ x_4=(1,1,1)$ *and let* $S=\{x_1,x_2,x_3,x_4\}$. *Determine all* x_i *such that* $L(S)=L(S\setminus\{x_i\})$.

7. *Let* $P=L\{(1,0,0),(1,1,0)\}$ *and* $Q=L\{(1,1,1)\}$ *be subspaces of* $\mathbb{R}^3$. *Show that* $P+Q=\mathbb{R}^3$ *and* $P\cap Q=\{\mathbf{0}\}$. *If* $\mathbf{u}\in\mathbb{R}^3$, *determine* $\mathbf{u}_P,\mathbf{u}_Q$ *such that* $\mathbf{u}=\mathbf{u}_P+\mathbf{u}_Q$ *where* $\mathbf{u}_P\in P$ *and* $\mathbf{u}_Q\in Q$. *Is it necessary that* $\mathbf{u}_P$ *and* $\mathbf{u}_Q$ *are unique?*

8. *Let* $P=L\{(1,-1,0),(1,1,0)\}$ *and* $Q=L\{(1,1,1),(1,2,1)\}$ *be subspaces of* $\mathbb{R}^3$. *Show that* $P+Q=\mathbb{R}^3$ *and* $P\cap Q\neq\{\mathbf{0}\}$. *Also, find a vector* $\mathbf{u}\in\mathbb{R}^3$ *such that* $\mathbf{u}$ *cannot be written uniquely in the form* $\mathbf{u}=\mathbf{u}_P+\mathbf{u}_Q$ *where* $\mathbf{u}_P\in P$ *and* $\mathbf{u}_Q\in Q$.

In this section, we saw that a vector space has infinite number of vectors. Hence, one can start with any finite collection of vectors and obtain their span. It means that any vector space contains infinite number of other vector subspaces. Therefore, the following questions arise:

1. What are the conditions under which, the linear span of two distinct sets are the same?

2. Is it possible to find/choose vectors so that the linear span of the chosen vectors is the whole vector space itself?

3. Suppose we are able to choose certain vectors whose linear span is the whole space. Can we find the minimum number of such vectors?

We try to answer these questions in the subsequent sections.

3.2 Linear Independence

Definition 3.2.1 (Linear Independence and Dependence)**.** *Let* $S=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m\}$ *be a non-empty subset of a vector space* $V(\mathbb{F})$. *The set* S *is said to be linearly independent if the system of equations*

$$\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\cdots+\alpha_m\mathbf{u}_m=\mathbf{0}, \tag{3.2.1}$$

in the unknowns α_i*'s* $1\le i\le m$, *has only the trivial solution. If the system (3.2.1) has a non-trivial solution then the set* S *is said to be linearly dependent.*

Example 3.2.2. *Is the set* S *a linear independent set? Give reasons.*

1. *Let* $S=\{(1,2,1),(2,1,4),(3,3,5)\}$.
   Solution: *Consider the linear system* $a(1,2,1)+b(2,1,4)+c(3,3,5)=(0,0,0)$ *in the unknowns* a,b *and* c. *It can be checked that this system has infinite number of solutions. Hence* S *is a linearly dependent subset of* $\mathbb{R}^3$.

2. *Let* $S = \{(1,1,1),(1,1,0),(1,0,1)\}$.
   Solution: *Consider the system* $a(1,1,1) + b(1,1,0) + c(1,0,1) = (0,0,0)$ *in the unknowns* a, b *and* c. *Check that this system has only the trivial solution. Hence* S *is a linearly independent subset of* $\mathbb{R}^3$.

In other words, if $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a non-empty subset of a vector space V, then one needs to solve the linear system of equations

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0} \tag{3.2.2}$$

in the unknowns $\alpha_1, \ldots, \alpha_n$. If $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ is THE ONLY SOLUTION of (3.2.2), then S is a linearly independent subset of V. Otherwise, the set S is a linearly dependent subset of V. We are now ready to state the following important results. The proof of only the first part is given. The reader is required to supply the proof of other parts.

Proposition 3.2.3. *Let* $V(\mathbb{F})$ *be a vector space.*

1. *Then the zero-vector cannot belong to a linearly independent set.*

2. *A non-empty subset of a linearly independent set of* V *is also linearly independent.*

3. *Every set containing a linearly dependent set of* V *is also linearly dependent.*

Proof. Let $S = \{\mathbf{0} = \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a set consisting of the zero vector. Then for any $\gamma \neq o$, $\gamma \mathbf{u}_1 + o \mathbf{u}_2 + \cdots + 0 \mathbf{u}_n = \mathbf{0}$. Hence, the system $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}$, has a non-trivial solution $(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\gamma, 0 \ldots, 0)$. Thus, the set S is linearly dependent. $\square$

Theorem 3.2.4. *Let* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ *be a linearly independent subset of a vector space* $V(\mathbb{F})$. *If for some* $\mathbf{v} \in V$, *the set* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{v}\}$ *is a linearly dependent, then* $\mathbf{v}$ *is a linear combination of* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$.

Proof. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}\}$ is linearly dependent, there exist scalars $c_1, \ldots, c_{p+1}$, NOT ALL ZERO, such that

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p + c_{p+1} \mathbf{v} = \mathbf{0}. \tag{3.2.3}$$

CLAIM: $c_{p+1} \neq 0$.
Let if possible $c_{p+1} = 0$. As the scalars in Equation (3.2.3) are not all zero, the linear system $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0}$ in the unknowns $\alpha_1, \ldots, \alpha_p$ has a non-trivial solution $(c_1, \ldots, c_p)$. This by definition of linear independence implies that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly dependent, a contradiction to our hypothesis. Thus, $c_{p+1} \neq 0$ and we get

$$\mathbf{v} = -\frac{1}{c_{p+1}}(c_1 \mathbf{v}_1 + \cdots + c_p \mathbf{v}_p) \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$$

as $-\frac{c_i}{c_{p+1}} \in \mathbb{F}$ for $1 \leq i \leq p$. Thus, the result follows. $\square$

We now state a very important corollary of Theorem 3.2.4 without proof. The readers are advised to supply the proof for themselves.

Corollary 3.2.5. *Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a subset of a vector space $V(\mathbb{F})$. If S is linearly*

1. *dependent then there exists a k, $2 \le k \le n$ with $L(\mathbf{u}_1, \ldots, \mathbf{u}_k) = L(\mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$.*
2. *independent and there is a vector $\mathbf{v} \in V$ with $\mathbf{v} \notin L(S)$ then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}\}$ is also a linearly independent subset of V.*

Exercise 3.2.6.

1. *Consider the vector space $\mathbb{R}^2$. Let $\mathbf{u}_1 = (1, 0)$. Find all choices for the vector $\mathbf{u}_2$ such that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent subset of $\mathbb{R}^2$. Does there exist vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ such that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent subset of $\mathbb{R}^2$?*
2. *Let $S = \{(1,1,1,1), (1,-1,1,2), (1,1,-1,1)\} \subset \mathbb{R}^4$. Does $(1,1,2,1) \in L(S)$? Furthermore, determine conditions on x, y, z and u such that $(x, y, z, u) \in L(S)$.*
3. *Show that $S = \{(1,2,3), (-2,1,1), (8,6,10)\} \subset \mathbb{R}^3$ is linearly dependent.*
4. *Show that $S = \{(1,0,0), (1,1,0), (1,1,1)\} \subset \mathbb{R}^3$ is linearly independent.*
5. *Prove that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a linearly independent subset of $V(\mathbb{F})$ if and only if $\{\mathbf{u}_1, \mathbf{u}_1 + \mathbf{u}_2, \ldots, \mathbf{u}_1 + \cdots + \mathbf{u}_n\}$ is linearly independent subset of $V(\mathbb{F})$.*
6. *Find 3 vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^4$ such that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent whereas $\{\mathbf{u}, \mathbf{v}\}, \{\mathbf{u}, \mathbf{w}\}$ and $\{\mathbf{v}, \mathbf{w}\}$ are linearly independent.*
7. *What is the maximum number of linearly independent vectors in $\mathbb{R}^3$?*
8. *Show that any set of k vectors in $\mathbb{R}^3$ is linearly dependent if $k \ge 4$.*
9. *Is $\{(1,0), (i,0)\}$ a linearly independent subset of $\mathbb{C}^2$ $(\mathbb{R})$?*
10. *Suppose V is a collection of vectors such that $V(\mathbb{C})$ as well as $V(\mathbb{R})$ are vector spaces. Prove that the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, i\mathbf{u}_1, \ldots, i\mathbf{u}_k\}$ is a linearly independent subset of $V(\mathbb{R})$ if and only if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linear independent subset of $V(\mathbb{C})$.*
11. *Let M be a subspace of V and let $\mathbf{u}, \mathbf{v} \in V$. Define $K = L(M, \mathbf{u})$ and $H = L(M, \mathbf{v})$. If $\mathbf{v} \in K$ and $\mathbf{v} \notin M$ prove that $\mathbf{u} \in H$.*
12. *Let $A \in M_n(\mathbb{R})$ and let $\mathbf{x}$ and $\mathbf{y}$ be two non-zero vectors such that $A\mathbf{x} = 3\mathbf{x}$ and $A\mathbf{y} = 2\mathbf{y}$. Prove that $\mathbf{x}$ and $\mathbf{y}$ are linearly independent.*
13. *Let $A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & -1 & 3 \\ 3 & -2 & 5 \end{bmatrix}$. Determine non-zero vectors $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ such that $A\mathbf{x} = 6\mathbf{x}$, $A\mathbf{y} = 2\mathbf{y}$ and $A\mathbf{z} = -2\mathbf{z}$. Use the vectors $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ obtained here to prove the following.*
    - (a) *$A^2\mathbf{v} = 4\mathbf{v}$, where $\mathbf{v} = c\mathbf{y} + d\mathbf{z}$ for any real numbers c and d.*
    - (b) *The set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly independent.*
    - (c) *Let $P = [\mathbf{x},\ \mathbf{y},\ \mathbf{z}]$ be a 3×3 matrix. Then P is invertible.*

(d) *Let* $D = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$. *Then* $AP = PD$.

14. *Let* P *and* Q *be subspaces of* $\mathbb{R}^n$ *such that* $P + Q = \mathbb{R}^n$ *and* $P \cap Q = \{\mathbf{0}\}$. *Prove that each* $\mathbf{u} \in \mathbb{R}^n$ *is uniquely expressible as* $\mathbf{u} = \mathbf{u}_P + \mathbf{u}_Q$, *where* $\mathbf{u}_P \in P$ *and* $\mathbf{u}_Q \in Q$.

3.3 Bases

Definition 3.3.1 (Basis of a Vector Space). *A basis of a vector space* V *is a subset* $\mathcal{B}$ *of* V *such that* $\mathcal{B}$ *is a linearly independent set in* V *and the linear span of* $\mathcal{B}$ *is* V. *Also, any element of* $\mathcal{B}$ *is called a basis vector.*

Remark 3.3.2. *Let* $\mathcal{B}$ *be a basis of a vector space* $V(\mathbb{F})$. *Then for each* $\mathbf{v} \in V$, *there exist vectors* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathcal{B}$ *such that* $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i$, *where* $\alpha_i \in \mathbb{F}$, *for* $1 \le i \le n$. *By convention, the linear span of an empty set is* $\{\mathbf{0}\}$. *Hence, the empty set is a basis of the vector space* $\{\mathbf{0}\}$.

Lemma 3.3.3. *Let* $\mathcal{B}$ *be a basis of a vector space* $V(\mathbb{F})$. *Then each* $\mathbf{v} \in V$ *is a unique linear combination of the basis vectors.*

Proof. On the contrary, assume that there exists $\mathbf{v} \in V$ that is can be expressed in at least two ways as linear combination of basis vectors. That means, there exists a positive integer p, scalars $\alpha_i, \beta_i \in \mathbb{F}$ and $\mathbf{v}_i \in \mathcal{B}$ such that

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_p\mathbf{v}_p \text{ and } \mathbf{v} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_p\mathbf{v}_p.$$

Equating the two expressions of $\mathbf{v}$ leads to the expression

$$(\alpha_1 - \beta_1)\mathbf{v}_1 + (\alpha_2 - \beta_2)\mathbf{v}_2 + \cdots + (\alpha_p - \beta_p)\mathbf{v}_p = \mathbf{0}. \tag{3.3.1}$$

Since the vectors are from $\mathcal{B}$, by definition (see Definition 3.3.1) the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a linearly independent subset of V. This implies that the linear system $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ in the unknowns $c_1, c_2, \ldots, c_p$ has only the trivial solution. Thus, each of the scalars $\alpha_i - \beta_i$ appearing in Equation (3.3.1) must be equal to 0. That is, $\alpha_i - \beta_i = 0$ for $1 \le i \le p$. Thus, for $1 \le i \le p$, $\alpha_i = \beta_i$ and the result follows. $\square$

Example 3.3.4. 1. *The set* $\{1\}$ *is a basis of the vector space* $\mathbb{R}(\mathbb{R})$.

2. *The set* $\{(1,1), (1,-1)\}$ *is a basis of the vector space* $\mathbb{R}^2(\mathbb{R})$.

3. *Fix a positive integer* n *and let* $\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{ith\ place}, 0, \ldots, 0) \in \mathbb{R}^n$ *for* $1 \le i \le n$. *Then* $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *is called the* **standard basis** *of* $\mathbb{R}^n$.

   (a) $\mathcal{B} = \{e_1\} = \{1\}$ *is a standard basis of* $\mathbb{R}(\mathbb{R})$.

   (b) $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ *with* $\mathbf{e}_1 = (1,0)$ *and* $\mathbf{e}_2 = (0,1)$ *is the standard basis of* $\mathbb{R}^2$.

(c) $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ *with* $\mathbf{e}_1 = (1,0,0), \mathbf{e}_2 = (0,1,0)$ *and* $\mathbf{e}_3 = (0,0,1)$ *is the standard basis of* $\mathbb{R}^3$.

(d) $\mathcal{B} = \{(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)\}$ *is the standard basis of* $\mathbb{R}^4$.

4. *Let* $V = \{(x,y,0) : x,y \in \mathbb{R}\} \subset \mathbb{R}^3$. *Then* $\mathcal{B} = \{(2,0,0), (1,3,0)\}$ *is a basis of* V.

5. *Let* $V = \{(x,y,z) \in \mathbb{R}^3 : x+y-z=0\}$ *be a vector subspace of* $\mathbb{R}^3$. *As each element* $(x,y,z) \in V$ *satisfies* $x+y-z=0$. *Or equivalently* $z = x+y$ *and hence*

$$(x,y,z) = (x,y,x+y) = (x,0,x) + (0,y,y) = x(1,0,1) + y(0,1,1).$$

Hence $\{(1,0,1), (0,1,1)\}$ *forms a basis of* V.

6. *Let* $V = \{a+ib : a,b \in \mathbb{R}\}$ *be a complex vector space. Then any element* $a+ib \in V$ *equals* $a+ib = (a+ib)\cdot 1$. *Hence a basis of* V *is* $\{1\}$.

7. *Let* $V = \{a+ib : a,b \in \mathbb{R}\}$ *be a real vector space. Then* $\{1, i\}$ *is a basis of* $V(\mathbb{R})$ *as* $a+ib = a\cdot 1 + b\cdot i$ *for* $a,b \in \mathbb{R}$ *and* $\{1,i\}$ *is a linearly independent subset of* $V(\mathbb{R})$.

8. *In* $\mathbb{C}^2$, $(a+ib, c+id) = (a+ib)(1,0) + (c+id)(0,1)$. *So,* $\{(1,0),(0,1)\}$ *is a basis of the complex vector space* $\mathbb{C}^2$.

9. *In case of the real vector space* $\mathbb{C}^2$, $(a+ib, c+id) = a(1,0) + b(i,0) + c(0,1) + d(0,i)$. *Hence* $\{(1,0),(i,0),(0,1),(0,i)\}$ *is a basis.*

10. $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *is the standard basis of* $\mathbb{C}^n$. *But* $\mathcal{B}$ *is not a basis of the real vector space* $\mathbb{C}^n$.

Before coming to the end of this section, we give an algorithm to obtain a basis of any finite dimensional vector space V. This will be done by a repeated application of Corollary 3.2.5. The algorithm proceeds as follows:

Step 1: Let $\mathbf{v}_1 \in V$ with $\mathbf{v}_1 \neq \mathbf{0}$. Then $\{\mathbf{v}_1\}$ is linearly independent.

Step 2: If $V = L(\mathbf{v}_1)$, we have got a basis of V. Else, pick $\mathbf{v}_2 \in V$ such that $\mathbf{v}_2 \notin L(\mathbf{v}_1)$. Then by Corollary 3.2.5.2, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

Step i: Either $V = L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i)$ or $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) \neq V$.
In the first case, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$ is a basis of V. In the second case, pick $\mathbf{v}_{i+1} \in V$ with $\mathbf{v}_{i+1} \notin L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i)$. Then, by Corollary 3.2.5.2, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i+1}\}$ is linearly independent.

This process will finally end as V is a finite dimensional vector space.

Exercise 3.3.5. 1. *Let* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ *be basis vectors of a vector space* V. *Then prove that whenever* $\sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i = \mathbf{0}$, *we must have* $\alpha_i = 0$ *for each* $i = 1, \ldots, n$.

2. *Find a basis of* $\mathbb{R}^3$ *containing the vector* $(1,1,-2)$.

3. *Find a basis of* $\mathbb{R}^3$ *containing the vector* $(1,1,-2)$ *and* $(1,2,-1)$.

4. *Is it possible to find a basis of* $\mathbb{R}^4$ *containing the vectors* $(1,1,1,-2)$, $(1,2,-1,1)$ *and* $(1,-2,7,-11)$*?*

5. *Let* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ *be a subset of a vector space* $V(\mathbb{F})$*. Suppose* $L(S) = V$ *but* S *is not a linearly independent set. Then prove that each vector in* V *can be expressed in more than one way as a linear combination of vectors from* S.

6. *Show that the set* $\{(1,0,1),(1,i,0),(1,1,1-i)\}$ *is a basis of* $\mathbb{C}^3$.

7. *Find a basis of the real vector space* $\mathbb{C}^n$ *containing the basis* $\mathcal{B}$ *given in Example 10.*

8. *Find a basis of* $V = \{(x,y,z,u) \in \mathbb{R}^4 : x-y-z=0, x+z-u=0\}$.

9. *Let* $A = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$*. Find a basis of* $V = \{\mathbf{x}^t \in \mathbb{R}^5 : A\mathbf{x} = \mathbf{0}\}$.

10. *Prove that* $\{1, x, x^2, \ldots, x^n, \ldots\}$ *is a basis of the vector space* $\mathcal{P}(\mathbb{R})$*. This basis has an infinite number of vectors. This is also called the standard basis of* $\mathcal{P}(\mathbb{R})$.

11. *Let* $\mathbf{u}^t = (1,1,-2), \mathbf{v}^t = (-1,2,3)$ *and* $\mathbf{w}^t = (1,10,1)$*. Find a basis of* $L(\mathbf{u},\mathbf{v},\mathbf{w})$*. Determine a geometrical representation of* $L(\mathbf{u},\mathbf{v},\mathbf{w})$*?*

12. *Prove that* $\{(1,0,0),(1,1,0),(1,1,1)\}$ *is a basis of* $\mathbb{C}^3$*. Is it a basis of* $\mathbb{C}^3(\mathbb{R})$*?*

3.3.1 Dimension of a Finite Dimensional Vector Space

We first prove a result which helps us in associating a non-negative integer to every finite dimensional vector space.

Theorem 3.3.6. *Let* V *be a vector space with basis* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$*. Let* m *be a positive integer with* $m > n$*. Then the set* $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\} \subset V$ *is linearly dependent.*

Proof. We need to show that the linear system

$$\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2 + \cdots + \alpha_m\mathbf{w}_m = \mathbf{0} \tag{3.3.2}$$

in the unknowns $\alpha_1, \alpha_2, \ldots, \alpha_m$ has a non-trivial solution. We start by expressing the vectors $\mathbf{w}_i$ in terms of the basis vectors $\mathbf{v}_j$'s.

As $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of V, for each $\mathbf{w}_i \in V$, $1 \le i \le m$, there exist unique scalars a_{ij}, $1 \le i \le n$, $1 \le j \le m$, such that

$$\begin{aligned}
\mathbf{w}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \cdots + a_{n1}\mathbf{v}_n, \\
\mathbf{w}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \cdots + a_{n2}\mathbf{v}_n, \\
\vdots &= \vdots \\
\mathbf{w}_m &= a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + \cdots + a_{nm}\mathbf{v}_n.
\end{aligned}$$

Hence, Equation (3.3.2) can be rewritten as

$$\alpha_1 \left(\sum_{j=1}^{n} a_{j1}\mathbf{v}_j\right) + \alpha_2 \left(\sum_{j=1}^{n} a_{j2}\mathbf{v}_j\right) + \cdots + \alpha_m \left(\sum_{j=1}^{n} a_{jm}\mathbf{v}_j\right) = \mathbf{0}.$$

Or equivalently,

$$\left(\sum_{i=1}^{m} \alpha_i a_{1i}\right)\mathbf{v}_1 + \left(\sum_{i=1}^{m} \alpha_i a_{2i}\right)\mathbf{v}_2 + \cdots + \left(\sum_{i=1}^{m} \alpha_i a_{ni}\right)\mathbf{v}_n = \mathbf{0}. \tag{3.3.3}$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis, using Exercise 3.3.5.1, we get

$$\sum_{i=1}^{m} \alpha_i a_{1i} = \sum_{i=1}^{m} \alpha_i a_{2i} = \cdots = \sum_{i=1}^{m} \alpha_i a_{ni} = 0.$$

Therefore, finding α_i's satisfying Equation (3.3.2) reduces to solving the homogeneous system $A\alpha = \mathbf{0}$ where $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}$ and $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$.

Since $n < m$, Corollary 2.1.23.2 (here the matrix A is $m \times n$) implies that $A\alpha = \mathbf{0}$ has a non-trivial solution. hence Equation (3.3.2) has a non-trivial solution and thus $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a linearly dependent set. □

Corollary 3.3.7. *Let $\mathcal{B}_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\mathcal{B}_2 = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be two bases of a finite dimensional vector space V. Then $m = n$.*

Proof. Let if possible, $m > n$. Then by Theorem 3.3.6, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly dependent subset of V, contradicting the assumption that $\mathcal{B}_2$ is a basis of V. Hence we must have $m \le n$. A similar argument implies $n \le m$ and hence $m = n$. □

Let V be a finite dimensional vector space. Then Corollary 3.3.7 implies that the number of elements in any basis of V is the same. This number is used to define the dimension of any finite dimensional vector space.

Definition 3.3.8 (Dimension of a Finite Dimensional Vector Space)**.** *Let V be a finite dimensional vector space. Then the dimension of V, denoted* $\dim(V)$*, is the number of elements in a basis of V.*

Note that Corollary 3.2.5.2 can be used to generate a basis of any non-trivial finite dimensional vector space.

Example 3.3.9. *The dimension of vector spaces in Example 3.3.4 are as follows:*

1. $\dim(\mathbb{R}) = 1$ *in Example 3.3.4.1.*

2. $\dim(\mathbb{R}^2) = 2$ *in Example 3.3.4.2.*

3. $\dim(V) = 2$ *in Example 3.3.4.4.*

4. $\dim(V) = 2$ *in Example 3.3.4.5.*

5. $\dim(V) = 1$ *in Example 3.3.4.6.*

6. $\dim(V) = 2$ *in Example 3.3.4.7.*

7. $\dim(\mathbb{C}^2) = 2$ *in Example 3.3.4.8.*

8. $\dim(\mathbb{C}^2(\mathbb{R})) = 4$ *in Example 3.3.4.9.*

9. *For fixed positive integer* n, $\dim(\mathbb{R}^n) = n$ *in Example 3.3.4.3 and in Example 3.3.4.10, one has* $\dim(\mathbb{C}^n) = n$ *and* $\dim(\mathbb{C}^n(\mathbb{R})) = 2n$.

Thus, we see that the dimension of a vector space dependents on the set of scalars.

Example 3.3.10. *Let* V *be the set of all functions* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *with the property that* $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ *and* $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *and* $\alpha \in \mathbb{R}$. *For any* $f, g \in V$, *and* $t \in \mathbb{R}$, *define*

$$(f \oplus g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) \quad \textit{and} \quad (t \odot f)(\mathbf{x}) = f(t\mathbf{x}).$$

Then it can be easily verified that V *is a real vector space. Also, for* $1 \leq i \leq n$, *define the functions* $\mathbf{e}_i(\mathbf{x}) = \mathbf{e}_i\big((x_1, x_2, \ldots, x_n)\big) = x_i$. *Then it can be easily verified that the set* $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *is a basis of* V *and hence* $\dim(V) = n$.

The next theorem follows directly from Corollary 3.2.5.2 and Theorem 3.3.6. Hence, the proof is omitted.

Theorem 3.3.11. *Let* S *be a linearly independent subset of a finite dimensional vector space* V. *Then the set* S *can be extended to form a basis of* V.

Theorem 3.3.11 is equivalent to the following statement:
Let V be a vector space of dimension n. Suppose, we have found a linearly independent subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ of V with $r < n$. Then it is possible to find vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ in V such that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of V. Thus, one has the following important corollary.

Corollary 3.3.12. *Let* V *be a vector space of dimension* n. *Then*

1. *any set consisting of* n *linearly independent vectors forms a basis of* V.

2. *any subset* S *of* V *having* n *vectors with* $L(S) = V$ *forms a basis of* V.

Exercise 3.3.13. 1. *Determine* $\dim(\mathcal{P}_n(\mathbb{R}))$. *Is* $\dim(\mathcal{P}(\mathbb{R}))$ *finite?*

2. *Let* W_1 *and* W_2 *be two subspaces of a vector space* V *such that* $W_1 \subset W_2$. *Show that* $W_1 = W_2$ *if and only if* $\dim(W_1) = \dim(W_2)$.

3. *Consider the vector space $C([-\pi,\ \pi])$. For each integer n, define $\mathbf{e}_n(x) = \sin(nx)$. Prove that $\{\mathbf{e}_n : n = 1, 2, \ldots\}$ is a linearly independent set.*
   [Hint: For any positive integer ℓ, consider the set $\{\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_\ell}\}$ and the linear system
   $$\alpha_1 \sin(k_1 x) + \alpha_2 \sin(k_2 x) + \cdots + \alpha_\ell \sin(k_\ell x) = \mathbf{0} \quad \text{for all} \quad x \in [-\pi,\ \pi]$$
   in the unknowns $\alpha_1, \ldots, \alpha_n$. Now for suitable values of m, consider the integral
   $$\int_{-\pi}^{\pi} \sin(mx)\left(\alpha_1 \sin(k_1 x) + \alpha_2 \sin(k_2 x) + \cdots + \alpha_\ell \sin(k_\ell x)\right)\ dx$$
   to get the required result.]

4. *Determine a basis and dimension of $W = \{(x, y, z, w) \in \mathbb{R}^4 : x + y - z + w = 0\}$.*

5. *Let W_1 be a subspace of a vector space V. If $\dim(V) = n$ and $\dim(W_1) = k$ with $k \geq 1$ then prove that there exists a subspace W_2 of V such that $W_1 \cap W_2 = \{\mathbf{0}\}$, $W_1 + W_2 = V$ and $\dim(W_2) = n - k$. Also, prove that for each $\mathbf{v} \in V$ there exist unique vectors $\mathbf{w}_1 \in W_1$ and $\mathbf{w}_2 \in W_2$ such that $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$. The subspace W_2 is called the complementary subspace of W_1 in V.*

6. *Is the set, $W = \{p(x) \in \mathcal{P}_4(\mathbb{R})\ :\ p(-1) = p(1) = 0\}$ a subspace of $\mathcal{P}_4(\mathbb{R})$? If yes, find its dimension.*

3.3.2 Application to the study of $\mathbb{C}^n$

In this subsection, we will study results that are intrinsic to the understanding of linear algebra, especially results associated with matrices. We start with a few exercises that should have appeared in previous sections of this chapter.

Exercise 3.3.14. 1. *Let $V = \{A \in M_2(\mathbb{C}) :\ tr(A) = 0\}$, where $tr(A)$ stands for the trace of the matrix A. Show that V is a real vector space and find its basis. Is $W = \left\{\begin{bmatrix} a & b \\ c & -a \end{bmatrix} : c = \overline{-b}\right\}$ a subspace of V?*

2. *In each of the questions given below, determine whether the given set is a vector space or not? If it is a vector space, find the dimension and a basis.*
   (a) $sl_n(\mathbb{R}) = \{A \in M_n(\mathbb{R})\ :\ tr(A) = 0\}$.
   (b) $S_n(\mathbb{R}) = \{A \in M_n(\mathbb{R})\ :\ A = A^t\}$.
   (c) $A_n(\mathbb{R}) = \{A \in M_n(\mathbb{R})\ :\ A + A^t = \mathbf{0}\}$.
   (d) $sl_n(\mathbb{C}) = \{A \in M_n(\mathbb{C})\ :\ tr(A) = 0\}$.
   (e) $S_n(\mathbb{C}) = \{A \in M_n(\mathbb{C})\ :\ A = A^*\}$.
   (f) $A_n(\mathbb{C}) = \{A \in M_n(\mathbb{C})\ :\ A + A^* = \mathbf{0}\}$.

3. *Does there exist an $A \in M_2(\mathbb{C})$ satisfying $A^2 \neq \mathbf{0}$ but $A^3 = \mathbf{0}$.*

4. *Prove that there does not exist an $A \in M_n(\mathbb{C})$ satisfying $A^n \neq \mathbf{0}$ but $A^{n+1} = \mathbf{0}$. That is, if A is an $n \times n$ nilpotent matrix then the order of nilpotency $\leq n$.*

5. *Let $A \in M_n(\mathbb{C})$ be a triangular matrix. Then the rows/columns of A are linearly independent subset of $\mathbb{C}^n$ if and only if $a_{ii} \neq 0$ for $1 \leq i \leq n$.*

6. *Prove that the rows/columns of $A \in M_n(\mathbb{C})$ are linearly independent if and only if* $\det(A) \neq 0$.

7. *Prove that the rows/columns of $A \in M_n(\mathbb{C})$ span $\mathbb{C}^n$ if and only if A is an invertible matrix.*

8. *Let A be a skew-symmetric matrix of odd order. Prove that the rows/columns of A are linearly dependent. Hint: What is* $\det(A)$?

We now define subspaces that are associated with matrices.

Definition 3.3.15. *Let $A \in M_{m\times n}(\mathbb{C})$ and let $R_1, R_2, \ldots, R_m \in \mathbb{C}^n$ be the rows of A and $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{C}^m$ be its columns. We define*

1. **Column Space**(A), *denoted Col(A), as Col$(A) = L(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{C}^n\} \subset \mathbb{C}^m$,*

2. **Column Space**(A^*), *as Col$(A^*) = \{A^*\mathbf{x} : \mathbf{x} \in \mathbb{C}^m\} \subset \mathbb{C}^n$,*

3. **Null Space**(A), *denoted $\mathcal{N}(A)$, as $\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \mathbf{0}\}$.*

4. **Range**(A), *denoted $\mathcal{R}(A)$, as* $\mathbf{Im}(A) = \mathcal{R}(A) = \{\mathbf{y} : A\mathbf{x} = \mathbf{y}$ *for some* $\mathbf{x} \in \mathbb{C}^n\}$.

Note that the "column space" of A consists of all $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution. Hence, $\text{Col}(A) = \text{Im}(A)$. We illustrate the above definitions with the help of an example and then ask the readers to solve the exercises that appear after the example.

Example 3.3.16. *Compute the above mentioned subspaces for* $A = \begin{bmatrix} 1 & 1 & 1 & -2 \\ 1 & 2 & -1 & 1 \\ 1 & -2 & 7 & -11 \end{bmatrix}$.

Solution: *Verify the following*

1. $\mathcal{R}(A) = L(R_1, R_2, R_3) = \{(x, y, z, u) \in \mathbb{C}^4 : 3x - 2y = z, 5x - 3y + u = 0\} = \mathcal{C}(A^*)$

2. $\mathcal{C}(A) = L(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \{(x, y, z) \in \mathbb{C}^3 : \ 4x - 3y - z = 0\} = \mathcal{R}(A^*)$

3. $\mathcal{N}(A) = \{(x, y, z, u) \in \mathbb{C}^4 : x + 3z - 5u = 0, y - 2z + 3u = 0\}$.

4. $\mathcal{N}(A^*) = \{(x, y, z) \in \mathbb{C}^3 : x + 4z = 0, y - 3z = 0\}$.

Exercise 3.3.17. 1. *Let $A \in M_{m\times n}(\mathbb{C})$. Then prove that*

   (a) *$\mathcal{R}(A)$ is a subspace of $\mathbb{C}^n$,*
   (b) *$\mathcal{C}(A)$ is a subspace of $\mathbb{C}^m$,*
   (c) *$\mathcal{N}(A)$ is a subspace of $\mathbb{C}^n$,*
   (d) *$\mathcal{N}(A^t)$ is a subspace of $\mathbb{C}^m$,*

(e) $\mathcal{R}(A) = \mathcal{C}(A^t)$ *and* $\mathcal{C}(A) = \mathcal{R}(A^t)$.

2. *Let* $A = \begin{bmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 2 & 2 & 2 & 4 \\ 2 & -2 & 4 & 0 & 8 \\ 4 & 2 & 5 & 6 & 10 \end{bmatrix}$ *and* $B = \begin{bmatrix} 2 & 4 & 0 & 6 \\ -1 & 0 & -2 & 5 \\ -3 & -5 & 1 & -4 \\ -1 & -1 & 1 & 2 \end{bmatrix}$.

   (a) *Find the row-reduced echelon forms of* A *and* B.

   (b) *Find* P_1 *and* P_2 *such that* $P_1 A$ *and* $P_2 B$ *are in row-reduced echelon form.*

   (c) *Find a basis each for the row spaces of* A *and* B.

   (d) *Find a basis each for the range spaces of* A *and* B.

   (e) *Find bases of the null spaces of* A *and* B.

   (f) *Find the dimensions of all the vector subspaces so obtained.*

Lemma 3.3.18. *Let* $A \in M_{m\times n}(\mathbb{C})$ *and let* $B = EA$ *for some elementary matrix* E. *Then* $\mathcal{R}(A) = \mathcal{R}(B)$ *and* $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(B))$.

Proof. We prove the result for the elementary matrix $E_{ij}(c)$, where $c \neq 0$ and $1 \le i < j \le m$. The readers are advised to prove the results for other elementary matrices. Let $R_1, R_2, \ldots, R_m$ be the rows of A. Then $B = E_{ij}(c)A$ implies

$$\begin{aligned}
\mathcal{R}(B) &= L(R_1, \ldots, R_{i-1}, R_i + cR_j, R_{i+1}, \ldots, R_m) \\
&= \{\alpha_1 R_1 + \cdots + \alpha_{i-1}R_{i-1} + \alpha_i(R_i + cR_j) + \cdots \\
&\qquad\qquad +\alpha_m R_m : \ \alpha_\ell \in \mathbb{R}, 1 \le \ell \le m\} \\
&= \left\{ \sum_{\ell=1}^{m} \alpha_\ell R_\ell + \alpha_i(cR_j) : \alpha_\ell \in \mathbb{R}, 1 \le \ell \le m \right\} \\
&= \left\{ \sum_{\ell=1}^{m} \beta_\ell R_\ell : \beta_\ell \in \mathbb{R}, 1 \le \ell \le m \right\} = L(R_1, \ldots, R_m) = \mathcal{R}(A)
\end{aligned}$$

Hence, the proof of the lemma is complete. $\square$

We omit the proof of the next result as the proof is similar to the proof of Lemma 3.3.18.

Lemma 3.3.19. *Let* $A \in M_{m\times n}(\mathbb{C})$ *and let* $C = AE$ *for some elementary matrix* E. *Then* $\mathcal{C}(A) = \mathcal{C}(C)$ *and* $\dim(\mathcal{C}(A)) = \dim(\mathcal{C}(C))$.

The first and second part of the next result are a repeated application of Lemma 3.3.18 and Lemma 3.3.19, respectively. Hence the proof is omitted. This result is also helpful in finding a basis of a subspace of $\mathbb{C}^n$.

Corollary 3.3.20. *Let* $A \in M_{m\times n}(\mathbb{C})$. *If*

1. B *is in row-reduced echelon form of* A *then* $\mathcal{R}(A) = \mathcal{R}(B)$. *In particular, the non-zero rows of* B *form a basis of* $\mathcal{R}(A)$ *and* $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(B)) =$ *Row rank*(A).

2. *the application of column operations gives a matrix C that has the form given in Remark 2.3.6, then* $\dim(\mathcal{C}(A)) = \dim(\mathcal{C}(C)) =$ *Column rank*(A) *and the non-zero columns of C form a basis of $\mathcal{C}(A)$.*

Before proceeding with applications of Corollary 3.3.20, we first prove that for any $A \in M_{m\times n}(\mathbb{C})$, Row rank$(A) =$ Column rank(A).

Theorem 3.3.21. *Let $A \in M_{m\times n}(\mathbb{C})$. Then Row rank*$(A) =$ *Column rank*(A).

Proof. Let $R_1, R_2, \ldots, R_m$ be the rows of A and $C_1, C_2, \ldots, C_n$ be the columns of A. Let Row rank$(A) = r$. Then by Corollary 3.3.20.1, $\dim\big(L(R_1, R_2, \ldots, R_m)\big) = r$. Hence, there exists vectors

$$\mathbf{u}_1^t = (u_{11}, \ldots, u_{1n}), \mathbf{u}_2^t = (u_{21}, \ldots, u_{2n}), \ldots, \mathbf{u}_r^t = (u_{r1}, \ldots, u_{rn}) \in \mathbb{R}^n$$

with

$$R_i \in L(\mathbf{u}_1^t, \mathbf{u}_2^t, \ldots, \mathbf{u}_r^t) \in \mathbb{R}^n, \quad \text{for all } \ i, 1 \le i \le m.$$

Therefore, there exist real numbers α_{ij}, $1 \le i \le m$, $1 \le j \le r$ such that

$$R_1 = \alpha_{11}\mathbf{u}_1^t + \cdots + \alpha_{1r}\mathbf{u}_r^t = \left(\sum_{i=1}^r \alpha_{1i}u_{i1}, \sum_{i=1}^r \alpha_{1i}u_{i2}, \ldots, \sum_{i=1}^r \alpha_{1i}u_{in}\right),$$

$$R_2 = \alpha_{21}\mathbf{u}_1^t + \cdots + \alpha_{2r}\mathbf{u}_r^t = \left(\sum_{i=1}^r \alpha_{2i}u_{i1}, \sum_{i=1}^r \alpha_{2i}u_{i2}, \ldots, \sum_{i=1}^r \alpha_{2i}u_{in}\right),$$

and so on, till

$$R_m = \alpha_{m1}\mathbf{u}_1^t + \cdots + \alpha_{mr}\mathbf{u}_r^t = \left(\sum_{i=1}^r \alpha_{mi}u_{i1}, \sum_{i=1}^r \alpha_{mi}u_{i2}, \ldots, \sum_{i=1}^r \alpha_{mi}u_{in}\right).$$

So,

$$C_1 = \begin{bmatrix} \sum\limits_{i=1}^r \alpha_{1i}u_{i1} \\ \vdots \\ \sum\limits_{i=1}^r \alpha_{mi}u_{i1} \end{bmatrix} = u_{11}\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix} + u_{21}\begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix} + \cdots + u_{r1}\begin{bmatrix} \alpha_{1r} \\ \alpha_{2r} \\ \vdots \\ \alpha_{mr} \end{bmatrix}.$$

In general, for $1 \le j \le n$, we have

$$C_j = \begin{bmatrix} \sum\limits_{i=1}^r \alpha_{1i}u_{ij} \\ \vdots \\ \sum\limits_{i=1}^r \alpha_{mi}u_{ij} \end{bmatrix} = u_{1j}\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{bmatrix} + u_{2j}\begin{bmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{m2} \end{bmatrix} + \cdots + u_{rj}\begin{bmatrix} \alpha_{1r} \\ \alpha_{2r} \\ \vdots \\ \alpha_{mr} \end{bmatrix}.$$

Therefore, $C_1, C_2, \ldots, C_n$ are linear combination of the r vectors

$$(\alpha_{11}, \alpha_{21}, \ldots, \alpha_{m1})^t, (\alpha_{12}, \alpha_{22}, \ldots, \alpha_{m2})^t, \ldots, (\alpha_{1r}, \alpha_{2r}, \ldots, \alpha_{mr})^t.$$

Thus, by Corollary 3.3.20.2, Column rank$(A) = \dim\big(\mathcal{C}(A)\big) \le r =$ Row rank(A). A similar argument gives Row rank$(A) \le$ Column rank(A). Hence, we have the required result. $\square$

Let M and N be two subspaces a vector space $V(\mathbb{F})$. Then recall that (see Exercise 3.1.22.5d) $M+N = \{\mathbf{u}+\mathbf{v} : \mathbf{u} \in M,\ \mathbf{v} \in N\}$ is the smallest subspace of V containing both M and N. We now state a very important result that relates the dimensions of the three subspaces M, N and $M+N$ (for a proof, see Appendix 7.3.1).

Theorem 3.3.22. *Let M and N be two subspaces of a finite dimensional vector space $V(\mathbb{F})$. Then*

$$\dim(M) + \dim(N) = \dim(M+N) + \dim(M \cap N). \tag{3.3.4}$$

Let S be a subset of $\mathbb{R}^n$ and let $V = L(S)$. Then Theorem 3.3.6 and Corollary 3.3.20.1 to obtain a basis of V. The algorithm proceeds as follows:

1. Construct a matrix A whose rows are the vectors in S.

2. Apply row operations on A to get B, a matrix in row echelon form.

3. Let $\mathcal{B}$ be the set of non-zero rows of B. Then $\mathcal{B}$ is a basis of $L(S) = V$.

Example 3.3.23. *1. Let $S = \{(1,1,1,1), (1,1,-1,1), (1,1,0,1), (1,-1,1,1)\} \subset \mathbb{R}^4$. Find a basis of $L(S)$.*

Solution: *Here* $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}$. *Then* $B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ *is the row echelon form of A and hence $\mathcal{B} = \{(1,1,1,1), (0,1,0,0), (0,0,1,0)\}$ is a basis of $L(S)$. Observe that the non-zero rows of B can be obtained, using the first, second and fourth or the first, third and fourth rows of A. Hence the subsets $\{(1,1,1,1),\ (1,1,0,1),\ (1,-1,1,1)\}$ and $\{(1,1,1,1),\ (1,1,-1,1),\ (1,-1,1,1)\}$ of S are also bases of $L(S)$.*

2. Let $V = \{(v,w,x,y,z) \in \mathbb{R}^5 : v+x+z = 3y\}$ and $W = \{(v,w,x,y,z) \in \mathbb{R}^5 : w-x = z, v = y\}$ be two subspaces of $\mathbb{R}^5$. Find bases of V and W containing a basis of $V \cap W$.

Solution: *Let us find a basis of $V \cap W$. The solution set of the linear equations*

$$v + x - 3y + z = 0, \quad w - x - z = 0 \quad \text{and} \quad v = y$$

is

$$(v,w,x,y,z)^t = (y, 2y, x, y, 2y - x)^t = y(1,2,0,1,2)^t + x(0,0,1,0,-1)^t.$$

Thus, a basis of $V \cap W$ is $\mathcal{B} = \{(1,2,0,1,2), (0,0,1,0,-1)\}$. Similarly, a basis of V is $\mathcal{B}_1 = \{(-1,0,1,0,0), (0,1,0,0,0), (3,0,0,1,0), (-1,0,0,0,1)\}$ and that of W is $\mathcal{B}_2 = \{(1,0,0,1,0), (0,1,1,0,0), (0,1,0,0,1)\}$. To find a basis of V containing a basis of $V \cap W$, form a matrix whose rows are the vectors in $\mathcal{B}$ and $\mathcal{B}_1$ (see the first matrix in Equation(3.3.5)) and apply row operations without disturbing the first two rows *that have come from $\mathcal{B}$. Then after a few row operations, we get*

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.3.5}$$

Thus, a required basis of V is $\{(1,2,0,1,2),(0,0,1,0,-1),(0,1,0,0,0),(0,0,0,1,3)\}$. *Similarly, a required basis of W is* $\{(1,2,0,1,2),(0,0,1,0,-1),(0,0,-1,0,1)\}$.

Exercise 3.3.24. *1. If M and N are 4-dimensional subspaces of a vector space V of dimension 7 then show that M and N have at least one vector in common other than the zero vector.*

2. *Let $V = \{(x,y,z,w) \in \mathbb{R}^4 : x+y-z+w=0, x+y+z+w=0, x+2y=0\}$ and $W = \{(x,y,z,w) \in \mathbb{R}^4 : x-y-z+w=0, x+2y-w=0\}$ be two subspaces of $\mathbb{R}^4$. Find bases and dimensions of V, W, $V \cap W$ and $V+W$.*

3. *Let W_1 and W_2 be two subspaces of a vector space V. If $\dim(W_1)+\dim(W_2) > \dim(V)$, then prove that $W_1 \cap W_2$ contains a non-zero vector.*

4. *Give examples to show that the Column Space of two row-equivalent matrices need not be same.*

5. *Let $A \in M_{m\times n}(\mathbb{C})$ with $m < n$. Prove that the columns of A are linearly dependent.*

6. *Suppose a sequence of matrices $A = B_0 \longrightarrow B_1 \longrightarrow \cdots \longrightarrow B_{k-1} \longrightarrow B_k = B$ satisfies $\mathcal{R}(B_l) \subset \mathcal{R}(B_{l-1})$ for $1 \le l \le k$. Then prove that $\mathcal{R}(B) \subset \mathcal{R}(A)$.*

Before going to the next section, we prove the rank-nullity theorem and the main theorem of system of linear equations (see Theorem 2.4.1).

Theorem 3.3.25 (Rank-Nullity Theorem)**.** *For any matrix $A \in M_{m\times n}(\mathbb{C})$,*

$$\dim(\mathcal{C}(A)) + \dim(\mathcal{N}(A)) = n.$$

Proof. Let $\dim(\mathcal{N}(A)) = r < n$ and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ be a basis of $\mathcal{N}(A)$. Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a linearly independent subset in $\mathbb{R}^n$, there exist vectors $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ (see Corollary 3.2.5.2) such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$. Then by definition,

$$\begin{aligned} \mathcal{C}(A) &= L(A\mathbf{u}_1, A\mathbf{u}_2, \ldots, A\mathbf{u}_n) \\ &= L(\mathbf{0}, \ldots, \mathbf{0}, A\mathbf{u}_{r+1}, A\mathbf{u}_{r+2}, \ldots, A\mathbf{u}_n) = L(A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n). \end{aligned}$$

We need to prove that $\{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ is a linearly independent set. Consider the linear system

$$\alpha_1 A\mathbf{u}_{r+1} + \alpha_2 A\mathbf{u}_{r+2} + \cdots + \alpha_{n-r} A\mathbf{u}_n = \mathbf{0}. \tag{3.3.6}$$

in the unknowns $\alpha_1, \ldots, \alpha_{n-r}$. This linear system is equivalent to

$$A(\alpha_1 \mathbf{u}_{r+1} + \alpha_2 \mathbf{u}_{r+2} + \cdots + \alpha_{n-r} \mathbf{u}_n) = \mathbf{0}.$$

Hence, by definition of $\mathcal{N}(A)$, $\alpha_1 \mathbf{u}_{r+1} + \cdots + \alpha_{n-r} \mathbf{u}_n \in \mathcal{N}(A) = L(\mathbf{u}_1, \ldots, \mathbf{u}_r)$. Therefore, there exists scalars $\beta_i,\ 1 \le i \le r$ such that

$$\alpha_1 \mathbf{u}_{r+1} + \alpha_2 \mathbf{u}_{r+2} + \cdots + \alpha_{n-r} \mathbf{u}_n = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \cdots + \beta_r \mathbf{u}_r.$$

Or equivalently,

$$\beta_1 \mathbf{u}_1 + \cdots + \beta_r \mathbf{u}_r - \alpha_1 \mathbf{u}_{r+1} - \cdots - \alpha_{n-r} \mathbf{u}_n = \mathbf{0}. \tag{3.3.7}$$

As $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a linearly independent set, the only solution of Equation (3.3.7) is

$$\alpha_i = 0 \text{ for } \ 1 \le i \le n-r \ \text{ and } \ \beta_j = 0 \text{ for } \ 1 \le j \le r.$$

In other words, we have shown that the only solution of Equation (3.3.6) is the trivial solution ($\alpha_i = 0$ for all $i,\ 1 \le i \le n-r$). Hence, the set $\{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ is a linearly independent and is a basis of $\mathcal{C}(A)$. Thus

$$\dim(\mathcal{C}(A)) + \dim(\mathcal{N}(A)) = (n-r) + r = n$$

and the proof of the theorem is complete. □

Theorem 3.3.25 is part of what is known as the fundamental theorem of linear algebra (see Theorem 5.2.15). As the final result in this direction, We now prove the main theorem on linear systems stated on page 48 (see Theorem 2.4.1) whose proof was omitted.

Theorem 3.3.26. *Consider a linear system* $A\mathbf{x} = \mathbf{b}$, *where* A *is an* $m \times n$ *matrix, and* $\mathbf{x},\ \mathbf{b}$ *are vectors of orders* $n \times 1$, *and* $m \times 1$, *respectively. Suppose rank* $(A) = r$ *and rank*$([A \ \ \mathbf{b}]) = r_a$. *Then exactly one of the following statement holds:*

1. *If* $r < r_a$, *the linear system has no solution.*
2. *if* $r_a = r$, *then the linear system is consistent. Furthermore,*
   (a) *if* $r = n$, *then the solution set of the linear system has a unique* $n \times 1$ *vector* $\mathbf{x}_0$ *satisfying* $A\mathbf{x}_0 = \mathbf{b}$.
   (b) *if* $r < n$, *then the set of solutions of the linear system is an infinite set and has the form*

   $$\{\mathbf{x}_0 + k_1 \mathbf{u}_1 + k_2 \mathbf{u}_2 + \cdots + k_{n-r} \mathbf{u}_{n-r} \ : \ k_i \in \mathbb{R},\ 1 \le i \le n-r\},$$

   where $\mathbf{x}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r}$ *are* $n \times 1$ *vectors satisfying* $A\mathbf{x}_0 = \mathbf{b}$ *and* $A\mathbf{u}_i = \mathbf{0}$ *for* $1 \le i \le n-r$.

Proof. Proof of Part 1. As $r < r_a$, the $(r+1)$-th row of the row-reduced echelon form of $[A \;\; \mathbf{b}]$ has the form $[\mathbf{0}, \;\; 1]$. Thus, by Theorem 1, the system $A\mathbf{x} = \mathbf{b}$ is inconsistent.

Proof of Part 2a and Part 2b. As $r = r_a$, using Corollary 3.3.20, $\mathcal{C}(A) = \mathcal{C}([A, \;\; \mathbf{b}])$. Hence, the vector $\mathbf{b} \in \mathcal{C}(A)$ and therefore there exist scalars $c_1, c_2, \ldots, c_n$ such that $\mathbf{b} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots c_n\mathbf{a}_n$, where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the columns of A. That is, we have a vector $\mathbf{x}_0^t = [c_1, c_2, \ldots, c_n]$ that satisfies $A\mathbf{x} = \mathbf{b}$.

If in addition $r = n$, then the system $A\mathbf{x} = \mathbf{b}$ has no free variables in its solution set and thus we have a unique solution (see Theorem 2.1.22.2a).

Whereas the condition $r < n$ implies that the system $A\mathbf{x} = \mathbf{b}$ has $n - r$ free variables in its solution set and thus we have an infinite number of solutions (see Theorem 2.1.22.2b). To complete the proof of the theorem, we just need to show that the solution set in this case has the form $\{\mathbf{x}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r} \; : \; k_i \in \mathbb{R}, \; 1 \le i \le n-r\}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$ for $1 \le i \le n - r$.

To get this, note that using the rank-nullity theorem (see Theorem 3.3.25) $\text{rank}(A) = r$ implies that $\dim(\mathcal{N}(A)) = n - r$. Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-r}\}$ be a basis of $\mathcal{N}(A)$. Then by definition $A\mathbf{u}_i = \mathbf{0}$ for $1 \le i \le n - r$ and hence

$$A(\mathbf{x}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r}) = A\mathbf{x}_0 + k_1\mathbf{0} + \cdots + k_{n-r}\mathbf{0} = \mathbf{b}.$$

Thus, the required result follows. □

Example 3.3.27. *Let* $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 & 3 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$ *and* $V = \{\mathbf{x}^t \in \mathbb{R}^7 : A\mathbf{x} = \mathbf{0}\}$. *Find a basis and dimension of* V.

Solution: *Observe that* x_1, x_3 *and* x_6 *are the basic variables and the rest are the free variables. Writing the basic variables in terms of free variables, we get*

$$x_1 = x_7 - x_2 - x_4 - x_5, \; x_3 = 2x_7 - 2x_4 - 3x_5 \;\; \text{and} \;\; x_6 = -x_7.$$

Hence,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} x_7 - x_2 - x_4 - x_5 \\ x_2 \\ 2x_7 - 2x_4 - 3x_5 \\ x_4 \\ x_5 \\ -x_7 \\ x_7 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}. \tag{3.3.8}$$

Therefore, if we let $\mathbf{u}_1^t = \Big[-1, 1, 0, 0, 0, 0, 0\Big]$, $\mathbf{u}_2^t = \Big[-1, 0, -2, 1, 0, 0, 0\Big]$, $\mathbf{u}_3^t = \Big[-1, 0, -3, 0, 1, 0, 0\Big]$ *and* $\mathbf{u}_4^t = \Big[1, 0, 2, 0, 0, -1, 1\Big]$ *then* $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ *is the basis of* V. *The reasons are as follows:*

1. *For Linear independence, we consider the homogeneous system*

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 = \mathbf{0} \tag{3.3.9}$$

in the unknowns c_1, c_2, c_3 and c_4. Then relating the unknowns with the free variables x_2, x_4, x_5 and x_7 and then comparing Equations (3.3.8) and (3.3.9), we get

   (a) $c_1 = 0$ *as the* 2*-nd coordinate consists only of* c_1.

   (b) $c_2 = 0$ *as the* 4*-th coordinate consists only of* c_2.

   (c) $c_3 = 0$ *as the* 5*-th coordinate consists only of* c_3.

   (d) $c_4 = 0$ *as the* 7*-th coordinate consists only of* c_4.

   Hence, the set S is linearly independent.

2. $L(S) = V$ *is obvious as any vector of V has the form mentioned as the first equality in Equation (3.3.8).*

The understanding built in Example 3.3.27 gives us the following remark.

Remark 3.3.28. *The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-r}$ in Theorem 3.3.26.2b correspond to expressing the solution set with the help of the free variables. This is done by writing the basic variables in terms of the free variables and then writing the solution set in such a way that each $\mathbf{u}_i$ corresponds to a specific free variable.*

The following are some of the consequences of the rank-nullity theorem. The proof is left as an exercise for the reader.

Exercise 3.3.29. 1. *Let A be an $m \times n$ real matrix. Then*

   (a) if $n > m$, then the system $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions,

   (b) if $n < m$, then there exists a non-zero vector $\mathbf{b} = (b_1, b_2, \ldots, b_m)^t$ such that the system $A\mathbf{x} = \mathbf{b}$ does not have any solution.

2. *The following statements are equivalent for an $m \times n$ matrix A.*

   (a) Rank $(A) = k$.

   (b) There exist a set of k rows of A that are linearly independent.

   (c) There exist a set of k columns of A that are linearly independent.

   (d) $\dim(\mathcal{C}(A)) = k$.

   (e) There exists a $k \times k$ submatrix B of A with $\det(B) \neq 0$ and determinant of every $(k+1) \times (k+1)$ submatrix of A is zero.

   (f) There exists a linearly independent subset $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$ of $\mathbb{R}^m$ such that the system $A\mathbf{x} = \mathbf{b}_i$ for $1 \leq i \leq k$ is consistent.

   (g) $\dim(\mathcal{N}(A)) = n - k$.

3.4 Ordered Bases

Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis of a vector space V. As $\mathcal{B}$ is a set, there is no ordering of its elements. In this section, we want to associate an order among the vectors in any basis of V as this helps in getting a better understanding about finite dimensional vector spaces and its relationship with matrices.

Definition 3.4.1 (Ordered Basis)**.** *Let V be a vector space of dimension n. Then an ordered basis for V is a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ together with a one-to-one correspondence between the sets $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\{1, 2, 3, \ldots, n\}$.*

If the ordered basis has $\mathbf{u}_1$ as the first vector, $\mathbf{u}_2$ as the second vector and so on, then we denote this by writing the ordered basis as $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$.

Example 3.4.2.

1. *Consider the vector space $\mathcal{P}_2(\mathbb{R})$ with basis $\{1-x, 1+x, x^2\}$. Then one can take either $\mathcal{B}_1 = \left(1-x, 1+x, x^2\right)$ or $\mathcal{B}_2 = \left(1+x, 1-x, x^2\right)$ as ordered bases. Also for any element $a_0 + a_1x + a_2x^2 \in \mathcal{P}_2(\mathbb{R})$, one has*

$$a_0 + a_1x + a_2x^2 = \frac{a_0 - a_1}{2}(1-x) + \frac{a_0 + a_1}{2}(1+x) + a_2x^2.$$

   Thus, $a_0 + a_1x + a_2x^2$ in the ordered basis

   (a) *$\mathcal{B}_1$, has $\frac{a_0-a_1}{2}$ as the coefficient of the first element, $\frac{a_0+a_1}{2}$ as the coefficient of the second element and a_2 as the coefficient the third element of $\mathcal{B}_1$.*

   (b) *$\mathcal{B}_2$, has $\frac{a_0+a_1}{2}$ as the coefficient of the first element, $\frac{a_0-a_1}{2}$ as the coefficient of the second element and a_2 as the coefficient the third element of $\mathcal{B}_2$.*

2. *Let $V = \{(x, y, z) : x + y = z\}$ and let $\mathcal{B} = \{(-1, 1, 0), (1, 0, 1)\}$ be a basis of V. Then check that $(3, 4, 7) = 4(-1, 1, 0) + 7(1, 0, 1) \in V$.*

That is, as ordered bases $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$, $(\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n, \mathbf{u}_1)$ and $(\mathbf{u}_n, \mathbf{u}_{n-1}, \ldots, \mathbf{u}_2, \mathbf{u}_1)$ are different even though they have the same set of vectors as elements. To proceed further, we now define the notion of coordinates of a vector depending on the chosen ordered basis.

Definition 3.4.3 (Coordinates of a Vector)**.** *Let $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis of a vector space V and let $\mathbf{v} \in V$. Suppose*

$$\mathbf{v} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_n\mathbf{v}_n \text{ for some scalars } \beta_1, \beta_2, \ldots, \beta_n.$$

Then the tuple $(\beta_1, \beta_2, \ldots, \beta_n)^t$ is called the coordinate of the vector $\mathbf{v}$ with respect to the ordered basis $\mathcal{B}$ and is denoted by $[\mathbf{v}]_{\mathcal{B}} = (\beta_1, \ldots, \beta_n)^t$, A COLUMN VECTOR.

Example 3.4.4.

1. *In Example 3.4.2.1, let $p(x) = a_0 + a_1x + a_2x^2$. Then*

$$[p(x)]_{\mathcal{B}_1} = \begin{bmatrix} \frac{a_0-a_1}{2} \\ \frac{a_0+a_1}{2} \\ a_2 \end{bmatrix}, [p(x)]_{\mathcal{B}_2} = \begin{bmatrix} \frac{a_0+a_1}{2} \\ \frac{a_0-a_1}{2} \\ a_2 \end{bmatrix} \text{ and } [p(x)]_{\mathcal{B}_3} = \begin{bmatrix} a_2 \\ \frac{a_0-a_1}{2} \\ \frac{a_0+a_1}{2} \end{bmatrix}.$$

2. *In Example 3.4.2.2,* $\big[(3,4,7)\big]_{\mathcal{B}} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$ *and* $\big[(x,y,z)\big]_{\mathcal{B}} = \big[(z-y,y,z)\big]_{\mathcal{B}} = \begin{bmatrix} y \\ z \end{bmatrix}$.

3. *Let the ordered bases of* $\mathbb{R}^3$ *be* $\mathcal{B}_1 = \big((1,0,0),(0,1,0),(0,0,1)\big)$, $\mathcal{B}_2 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ *and* $\mathcal{B}_3 = \big((1,1,1),(1,1,0),(1,0,0)\big)$. *Then*

$$\begin{aligned} (1,-1,1) &= 1\cdot(1,0,0) + (-1)\cdot(0,1,0) + 1\cdot(0,0,1). \\ &= 2\cdot(1,0,0) + (-2)\cdot(1,1,0) + 1\cdot(1,1,1). \\ &= 1\cdot(1,1,1) + (-2)\cdot(1,1,0) + 2\cdot(1,0,0). \end{aligned}$$

Therefore, if we write $\mathbf{u} = (1,-1,1)$, *then*

$$[\mathbf{u}]_{\mathcal{B}_1} = (1,-1,1)^t, \ [\mathbf{u}]_{\mathcal{B}_2} = (2,-2,1)^t, \ [\mathbf{u}]_{\mathcal{B}_3} = (1,-2,2)^t.$$

In general, let V be an n-dimensional vector space with $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Since $\mathcal{B}_1$ is a basis of V, there exist unique scalars a_{ij}, $1 \le i,j \le n$, such that

$$\mathbf{v}_i = \sum_{l=1}^{n} a_{li}\mathbf{u}_l, \quad \text{or equivalently,} \quad [\mathbf{v}_i]_{\mathcal{B}_1} = (a_{1i}, a_{2i}, \ldots, a_{ni})^t \text{ for } \ 1 \le i \le n.$$

Suppose $\mathbf{v} \in V$ with $[\mathbf{v}]_{\mathcal{B}_2} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$. Then

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} a_{ji}\mathbf{u}_j \right) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ji}\alpha_i \right) \mathbf{u}_j.$$

Since $\mathcal{B}_1$ is a basis this representation of $\mathbf{v}$ in terms of $\mathbf{u}_i$'s is unique. So,

$$[\mathbf{v}]_{\mathcal{B}_1} = \left(\sum_{i=1}^{n} a_{1i}\alpha_i, \sum_{i=1}^{n} a_{2i}\alpha_i, \ldots, \sum_{i=1}^{n} a_{ni}\alpha_i \right)^t = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = A[\mathbf{v}]_{\mathcal{B}_2},$$

where $A = \Big[[\mathbf{v}_1]_{\mathcal{B}_1}, [\mathbf{v}_2]_{\mathcal{B}_1}, \ldots, [\mathbf{v}_n]_{\mathcal{B}_1} \Big]$. Hence, we have proved the following theorem.

Theorem 3.4.5. *Let* V *be an* n*-dimensional vector space with bases* $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ *and* $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. *Define an* $n \times n$ *matrix* A *by* $A = \Big[[\mathbf{v}_1]_{\mathcal{B}_1}, [\mathbf{v}_2]_{\mathcal{B}_1}, \ldots, [\mathbf{v}_n]_{\mathcal{B}_1} \Big]$. *Then,* A *is an invertible matrix (see Exercise 3.3.14.7) and*

$$[\mathbf{v}]_{\mathcal{B}_1} = A[\mathbf{v}]_{\mathcal{B}_2} \ \textit{ for all } \ \mathbf{v} \in V.$$

Theorem 3.4.5 states that the coordinates of a vector with respect to different bases are related via an invertible matrix A.

Example 3.4.6. *Let* $\mathcal{B}_1 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ *and* $\mathcal{B}_2 = \big((1,1,1),(1,-1,1),(1,1,0)\big)$ *be two bases of* $\mathbb{R}^3$.

1. *Then* $[(x,y,z)]_{\mathcal{B}_1} = (x-y, y-z, z)^t$ *and* $[(x,y,z)]_{\mathcal{B}_2} = (\frac{y-x}{2}+z, \frac{x-y}{2}, x-z)^t$.

2. *Check that* $A = \Big[[(1,1,1)]_{\mathcal{B}_1}, [(1,-1,1)]_{\mathcal{B}_1}, [(1,1,0)]_{\mathcal{B}_1}\Big] = \begin{bmatrix} 0 & 2 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ *as*

$$\begin{aligned} [(1,1,1)]_{\mathcal{B}_1} &= 0\cdot(1,0,0) + 0\cdot(1,1,0) + 1\cdot(1,1,1) = (0,0,1)^t, \\ [(1,-1,1)]_{\mathcal{B}_1} &= 2\cdot(1,0,0) + (-2)\cdot(1,1,0) + 1\cdot(1,1,1) = (2,-2,1)^t \text{ and} \\ [(1,1,0)]_{\mathcal{B}_1} &= 0\cdot(1,0,0) + 1\cdot(1,1,0) + 0\cdot(1,1,1) = (0,1,0)^t. \end{aligned}$$

3. *Thus, for any* $(x,y,z) \in \mathbb{R}^3$,

$$[(x,y,z)]_{\mathcal{B}_1} = \begin{bmatrix} x-y \\ y-z \\ z \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{y-x}{2}+z \\ \frac{x-y}{2} \\ x-z \end{bmatrix} = A\ \ [(x,y,z)]_{\mathcal{B}_2}.$$

4. *Observe that the matrix* A *is invertible and hence* $[(x,y,z)]_{\mathcal{B}_2} = A^{-1}\ \ [(x,y,z)]_{\mathcal{B}_1}$.

In the next chapter, we try to understand Theorem 3.4.5 again using the ideas of 'linear transformations/functions'.

Exercise 3.4.7. 1. *Consider the vector space* $\mathcal{P}_3(\mathbb{R})$.

(a) *Prove that* $\mathcal{B}_1 = (1-x, 1+x^2, 1-x^3, 3+x^2-x^3)$ *and* $\mathcal{B}_2 = (1, 1-x, 1+x^2, 1-x^3)$ *are bases of* $\mathcal{P}_3(\mathbb{R})$.

(b) *Find the coordinates of* $\mathbf{u} = 1 + x + x^2 + x^3$ *with respect to* $\mathcal{B}_1$ *and* $\mathcal{B}_2$.

(c) *Find the matrix* A *such that* $[\mathbf{u}]_{\mathcal{B}_2} = A[\mathbf{u}]_{\mathcal{B}_1}$.

(d) *Let* $\mathbf{v} = a_0 + a_1x + a_2x^2 + a_3x^3$. *Then verify that*

$$[\mathbf{v}]_{\mathcal{B}_1} = \begin{bmatrix} -a_1 \\ -a_0 - a_1 + 2a_2 - a_3 \\ -a_0 - a_1 + a_2 - 2a_3 \\ a_0 + a_1 - a_2 + a_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 + a_1 - a_2 + a_3 \\ -a_1 \\ a_2 \\ -a_3 \end{bmatrix} = [\mathbf{v}]_{\mathcal{B}_2}.$$

2. *Let* $\mathcal{B} = \big((2,1,0), (2,1,1), (2,2,1)\big)$ *be an ordered basis of* $\mathbb{R}^3$. *Determine the coordinates of* $(1,2,1)$ *and* $(4,-2,2)$ *with respect* $\mathcal{B}$.

3.5 Summary

In this chapter, we started with the definition of vector spaces over $\mathbb{F}$, the set of scalars. The set $\mathbb{F}$ was either $\mathbb{R}$, the set of real numbers or $\mathbb{C}$, the set of complex numbers.

It was important to note that given a non-empty set V of vectors with a set $\mathbb{F}$ of scalars, we need to do the following:

1. first define vector addition and scalar multiplication and

2. then verify the axioms in Definition 3.1.1.

If all the axioms are satisfied then V is a vector space over $\mathbb{F}$. To check whether a non-empty subset W of a vector space V over $\mathbb{F}$ is a subspace of V, we only need to check whether $\mathbf{u}+\mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$ and $\alpha\mathbf{u} \in W$ for all $\alpha \in \mathbb{F}$ and $\mathbf{u} \in W$.

We then came across the definition of linear combination of vectors and the linear span of vectors. It was also shown that the linear span of a subset S of a vector space V is the smallest subspace of V containing S. Also, to check whether a given vector $\mathbf{v}$ is a linear combination of the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, we need to solve the linear system

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{v}$$

in the unknowns $c_1, \ldots, c_n$. This corresponds to solving the linear system $A\mathbf{x} = \mathbf{b}$. It was also shown that the geometrical representation of the linear span of $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is equivalent to finding conditions on the coordinates of the vector $\mathbf{b}$ such that the linear system $A\mathbf{x} = \mathbf{b}$ is consistent, where the matrix A is formed with the coordinates of the vector $\mathbf{u}_i$ as the i-th column of the matrix A.

By definition, $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linearly independent subset in $V(\mathbb{F})$ if the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution in $\mathbb{F}$, else S is linearly dependent, where the matrix A is formed with the coordinates of the vector $\mathbf{u}_i$ as the i-th column of the matrix A.

We then had the notion of the basis of a finite dimensional vector space V and the following results were proved.

1. A linearly independent set can be extended to form a basis of V.

2. Any two bases of V have the same number of elements.

This number was defined as the dimension of V and we denoted it by $\dim(V)$.

The following conditions are equivalent for an $n \times n$ matrix A.

1. A is invertible.

2. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

3. The row reduced echelon form of A is I.

4. A is a product of elementary matrices.

5. The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.

6. The system $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.

7. $\text{rank}(A) = n$.

8. $\det(A) \neq 0$.

9. The row space of A is $\mathbb{R}^n$.

10. The column space of A is $\mathbb{R}^n$.

11. The rows of A form a basis of $\mathbb{R}^n$.

12. The columns of A form a basis of $\mathbb{R}^n$.

13. The null space of A is $\{\mathbf{0}\}$.

Let A be an $m \times n$ matrix. Then we proved the rank-nullity theorem which states that $\text{rank}(A) + \text{nullity}(A) = n$, the number of columns. This implied that if $\text{rank}(A) = r$ then the solution set of the linear system $A\mathbf{x} = \mathbf{b}$ is of the form $\mathbf{x}_0 + c_1\mathbf{u}_1 + \cdots + c_{n-r}\mathbf{u}_{n-r}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$ for $1 \leq i \leq n-r$. Also, the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-r}$ are linearly independent.

Let V be a vector space of $\mathbb{R}^n$ for some positive integer n with $\dim(V) = k$. Then V may not have a standard basis. Even if V may have a basis that looks like an standard basis, our problem may force us to look for some other basis. In such a case, it is always helpful to fix an ordered basis $\mathcal{B}$ and then express each vector in V as a linear combination of elements from $\mathcal{B}$. This idea helps us in writing each element of V as a column vector of size k. We will also see its use in the study of linear transformations and the study of eigenvalues and eigenvectors.

Chapter 4

Linear Transformations

4.1 Definitions and Basic Properties

In this chapter, it will be shown that if V is a real vector space with $\dim(V) = n$ then V looks like $\mathbb{R}^n$. On similar lines a complex vector space of dimension n has all the properties that are satisfied by $\mathbb{C}^n$. To do so, we start with the definition of functions over vector spaces that commute with the operations of vector addition and scalar multiplication.

Definition 4.1.1 (Linear Transformation, Linear Operator). *Let V and W be vector spaces over the same scalar set $\mathbb{F}$. A function (map) $T : V \longrightarrow W$ is called a linear transformation if for all $\alpha \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$ the function T satisfies*

$$T(\alpha \cdot \mathbf{u}) = \alpha \odot T(\mathbf{u}) \text{ and } T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) \oplus T(\mathbf{v}),$$

where $+, \cdot$ are binary operations in V and $\oplus, \odot$ are the binary operations in W. In particular, if $W = V$ then the linear transformation T is called a linear operator.

We now give a few examples of linear transformations.

Example 4.1.2.

1. *Define $T : \mathbb{R} \longrightarrow \mathbb{R}^2$ by $T(x) = (x, 3x)$ for all $x \in \mathbb{R}$. Then T is a linear transformation as*

$$T(\alpha x) = (\alpha x, 3\alpha x) = \alpha(x, 3x) = \alpha T(x) \text{ and}$$
$$T(x + y) = (x + y, 3(x + y) = (x, 3x) + (y, 3y) = T(x) + T(y).$$

2. *Let V, W and Z be vector spaces over $\mathbb{F}$. Also, let $T : V \longrightarrow W$ and $S : W \longrightarrow Z$ be linear transformations. Then, for each $\mathbf{v} \in V$, the composition of T and S is defined by $S \circ T(\mathbf{v}) = S\big(T(\mathbf{v})\big)$. It is easy to verify that $S \circ T$ is a linear transformation. In particular, if $V = W$, one writes T^2 in place of $T \circ T$.*

3. *Let $\mathbf{x}^t = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Then for a fixed vector $\mathbf{a}^t = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$, define $T : \mathbb{R}^n \longrightarrow \mathbb{R}$ by $T(\mathbf{x}^t) = \sum\limits_{i=1}^{n} a_i x_i$ for all $\mathbf{x}^t \in \mathbb{R}^n$. Then T is a linear transformation. In particular,*

(a) $T(\mathbf{x}^t) = \sum\limits_{i=1}^{n} x_i$ *for all* $\mathbf{x}^t \in \mathbb{R}^n$ *if* $a_i = 1$, *for* $1 \le i \le n$.

(b) *if* $\mathbf{a} = \mathbf{e}_i$ *for a fixed* i, $1 \le i \le n$, *one can define* $T_i(\mathbf{x}^t) = x_i$ *for all* $\mathbf{x}^t \in \mathbb{R}^n$.

4. *Define* $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ *by* $T(x, y) = (x + y, 2x - y, x + 3y)$. *Then* T *is a linear transformation with* $T(1, 0) = (1, 2, 1)$ *and* $T(0, 1) = (1, -1, 3)$.

5. *Let* $A \in M_{m\times n}(\mathbb{C})$. *Define a map* $T_A : \mathbb{C}^n \longrightarrow \mathbb{C}^m$ *by* $T_A(\mathbf{x}^t) = A\mathbf{x}$ *for every* $\mathbf{x}^t = (x_1, x_2, \ldots, x_n) \in \mathbb{C}^n$. *Then* T_A *is a linear transformation. That is, every* $m \times n$ *complex matrix defines a linear transformation from* $\mathbb{C}^n$ *to* $\mathbb{C}^m$.

6. *Define* $T : \mathbb{R}^{n+1} \longrightarrow \mathcal{P}_n(\mathbb{R})$ *by* $T(a_1, a_2, \ldots, a_{n+1}) = a_1 + a_2 x + \cdots + a_{n+1} x^n$ *for* $(a_1, a_2, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. *Then* T *is a linear transformation.*

7. *Fix* $A \in M_n(\mathbb{C})$. *Then* $T_A : M_n(\mathbb{C}) \longrightarrow M_n(\mathbb{C})$ *and* $S_A : M_n(\mathbb{C}) \longrightarrow \mathbb{C}$ *are both linear transformations, where*

$$T_A(B) = BA^* \;\; \text{and} \;\; S_A(B) = tr(BA^*) \quad \text{for every} \quad B \in M_n(\mathbb{C}).$$

Before proceeding further with some more definitions and results associated with linear transformations, we prove that any linear transformation sends the zero vector to a zero vector.

Proposition 4.1.3. *Let* $T : V \longrightarrow W$ *be a linear transformation. Suppose that* $\mathbf{0}_V$ *is the zero vector in* V *and* $\mathbf{0}_W$ *is the zero vector of* W. *Then* $T(\mathbf{0}_V) = \mathbf{0}_W$.

Proof. Since $\mathbf{0}_V = \mathbf{0}_V + \mathbf{0}_V$, we have

$$T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = T(\mathbf{0}_V) + T(\mathbf{0}_V).$$

So $T(\mathbf{0}_V) = \mathbf{0}_W$ as $T(\mathbf{0}_V) \in W$. □

From now on, we write $\mathbf{0}$ for both the zero vector of the domain and codomain.

Definition 4.1.4 (Zero Transformation)**.** *Let* V *and* W *be two vector spaces over* $\mathbb{F}$ *and define* $T : V \longrightarrow W$ *by* $T(\mathbf{v}) = \mathbf{0}$ *for every* $\mathbf{v} \in V$. *Then* T *is a linear transformation and is usually called the zero transformation, denoted* $\mathbf{0}$.

Definition 4.1.5 (Identity Operator)**.** *Let* V *be a vector space over* $\mathbb{F}$ *and define* $T : V \longrightarrow V$ *by* $T(\mathbf{v}) = \mathbf{v}$ *for every* $\mathbf{v} \in V$. *Then* T *is a linear transformation and is usually called the Identity transformation, denoted* I.

Definition 4.1.6 (Equality of two Linear Operators)**.** *Let* V *be a vector space and let* $T, S : V \longrightarrow V$ *be a linear operators. The operators* T *and* S *are said to be equal if* $T(\mathbf{x}) = S(\mathbf{x})$ *for all* $\mathbf{x} \in V$.

We now prove a result that relates a linear transformation T with its value on a basis of the domain space.

Theorem 4.1.7. *Let V and W be two vector spaces over $\mathbb{F}$ and let $T : V \longrightarrow W$ be a linear transformation. If $\mathcal{B} = \big(\mathbf{u}_1, \ldots, \mathbf{u}_n\big)$ is an ordered basis of V then for each $\mathbf{v} \in V$, the vector $T(\mathbf{v})$ is a linear combination of $T(\mathbf{u}_1), \ldots, T(\mathbf{u}_n) \in W$. That is, we have full information of T if we know $T(\mathbf{u}_1), \ldots, T(\mathbf{u}_n) \in W$, the image of basis vectors in W.*

Proof. As $\mathcal{B}$ is a basis of V, for every $\mathbf{v} \in V$, we can find $c_1, \ldots, c_n \in \mathbb{F}$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$, or equivalently $[\mathbf{v}]_{\mathcal{B}} = (\alpha_1, \ldots, \alpha_n)^t$. Hence, by definition

$$T(\mathbf{v}) = T(c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n) = c_1T(\mathbf{u}_1) + \cdots + c_nT(\mathbf{u}_n).$$

That is, we just need to know the vectors $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$ in W to get $T(\mathbf{v})$ as $[\mathbf{v}]_{\mathcal{B}} = (\alpha_1, \ldots, \alpha_n)^t$ is known in V. Hence, the required result follows. $\square$

Exercise 4.1.8.

1. *Are the maps $T : V \longrightarrow W$ given below, linear transformations?*
   (a) *Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^3$ with $T(x, y) = (x + y + 1, 2x - y, x + 3y)$.*
   (b) *Let $V = W = \mathbb{R}^2$ with $T(x, y) = (x - y, x^2 - y^2)$.*
   (c) *Let $V = W = \mathbb{R}^2$ with $T(x, y) = (x - y, |x|)$.*
   (d) *Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^4$ with $T(x, y) = (x + y, x - y, 2x + y, 3x - 4y)$.*
   (e) *Let $V = W = \mathbb{R}^4$ with $T(x, y, z, w) = (z, x, w, y)$.*
2. *Which of the following maps $T : M_2(\mathbb{R}) \longrightarrow M_2(\mathbb{R})$ are linear operators?*
   (a) $T(A) = A^t$ (b) $T(A) = I + A$ (c) $T(A) = A^2$
   (d) $T(A) = BAB^{-1}$, *where B is a fixed 2×2 matrix.*
3. *Prove that a map $T : \mathbb{R} \longrightarrow \mathbb{R}$ is a linear transformation if and only if there exists a unique $c \in \mathbb{R}$ such that $T(\mathbf{x}) = c\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}$.*
4. *Let $A \in M_n(\mathbb{C})$ and define $T_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ by $T_A(\mathbf{x}^t) = A\mathbf{x}$ for every $\mathbf{x}^t \in \mathbb{C}^n$. Prove that for any positive integer k, $T_A^k(\mathbf{x}^t) = A^k\mathbf{x}$.*
5. *Use matrices to give examples of linear operators $T, S : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ that satisfy:*
   (a) $T \neq \mathbf{0}$, $T^2 \neq \mathbf{0}$, $T^3 = \mathbf{0}$.
   (b) $T \neq \mathbf{0}$, $S \neq \mathbf{0}$, $S \circ T \neq \mathbf{0}$, $T \circ S = \mathbf{0}$.
   (c) $S^2 = T^2$, $S \neq T$.
   (d) $T^2 = I$, $T \neq I$.
6. *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear operator with $T \neq \mathbf{0}$ and $T^2 = \mathbf{0}$. Prove that there exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that the set $\{\mathbf{x}, T(\mathbf{x})\}$ is linearly independent.*
7. *Fix a positive integer p and let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a linear operator with $T^k \neq \mathbf{0}$ for $1 \leq k \leq p$ and $T^{p+1} = \mathbf{0}$. Then prove that there exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that the set $\{\mathbf{x}, T(\mathbf{x}), \ldots, T^p(\mathbf{x})\}$ is linearly independent.*

8. *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation with $T(\mathbf{x}_0) = \mathbf{y}_0$ for some $\mathbf{x}_0 \in \mathbb{R}^n$ and $\mathbf{y}_0 \in \mathbb{R}^m$. Define $T^{-1}(\mathbf{y}_0) = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{y}_0\}$. Then prove that for every $\mathbf{x} \in T^{-1}(\mathbf{y}_0)$ there exists $\mathbf{z} \in T^{-1}(\mathbf{0})$ such that $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$. Also, prove that $T^{-1}(\mathbf{y}_0)$ is a subspace of $\mathbb{R}^n$ if and only if $\mathbf{0} \in T^{-1}(\mathbf{y}_0)$.*

9. *Define a map $T : \mathbb{C} \longrightarrow \mathbb{C}$ by $T(z) = \overline{z}$, the complex conjugate of z. Is T a linear transformation over $\mathbb{C}(\mathbb{R})$?*

10. *Prove that there exists infinitely many linear transformations $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ such that $T(1,-1,1) = (1,2)$ and $T(-1,1,2) = (1,0)$?*

11. *Does there exist a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ such that $T(1,0,1) = (1,2)$, $T(0,1,1) = (1,0)$ and $T(1,1,1) = (2,3)$?*

12. *Does there exist a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ such that $T(1,0,1) = (1,2)$, $T(0,1,1) = (1,0)$ and $T(1,1,2) = (2,3)$?*

13. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be defined by $T(x,y,z) = (2x+3y+4z, x+y+z, x+y+3z)$. Find the value of k for which there exists a vector $\mathbf{x}^t \in \mathbb{R}^3$ such that $T(\mathbf{x}^t) = (9,3,k)$.*

14. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be defined by $T(x,y,z) = (2x-2y+2z, -2x+5y+2z, 8x+y+4z)$. Find a vector $\mathbf{x}^t \in \mathbb{R}^3$ such that $T(\mathbf{x}^t) = (1,1,-1)$.*

15. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be defined by $T(x,y,z) = (2x+y+3z, 4x-y+3z, 3x-2y+5z)$. Determine non-zero vectors $\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t \in \mathbb{R}^3$ such that $T(\mathbf{x}^t) = 6\mathbf{x}$, $T(\mathbf{y}^t) = 2\mathbf{y}$ and $T(\mathbf{z}^t) = -2\mathbf{z}$. Is the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ linearly independent?*

16. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be defined by $T(x,y,z) = (2x+3y+4z, -y, -3y+4z)$. Determine non-zero vectors $\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t \in \mathbb{R}^3$ such that $T(\mathbf{x}^t) = 2\mathbf{x}$, $T(\mathbf{y}^t) = 4\mathbf{y}$ and $T(\mathbf{z}^t) = -\mathbf{z}$. Is the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ linearly independent?*

17. *Let n be any positive integer. Prove that there does not exist a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ such that $T(1,1,-2) = \mathbf{x}^t$, $T(-1,2,3) = \mathbf{y}^t$ and $T(1,10,1) = \mathbf{z}^t$ where $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Does there exist real numbers c, d such that $\mathbf{z} = c\mathbf{x} + d\mathbf{y}$ and T is indeed a linear transformation?*

18. *Find all functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ that fixes the line $y = x$ and sends (x_1, y_1) for $x_1 \neq y_1$ to its mirror image along the line $y = x$. Or equivalently, f satisfies*

    (a) *$f(x,x) = (x,x)$ and*

    (b) *$f(x,y) = (y,x)$ for all $(x,y) \in \mathbb{R}^2$.*

19. *Consider the complex vector space $\mathbb{C}^3$ and let $f : \mathbb{C}^3 \longrightarrow \mathbb{C}^3$ be a linear transformation. Suppose there exist non-zero vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^3$ such that $f(\mathbf{x}) = \mathbf{x}$, $f(\mathbf{y}) = (1+i)\mathbf{y}$ and $f(\mathbf{z}) = (2+3i)\mathbf{z}$. Then prove that*

    (a) *the vectors $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ are linearly independent subset of $\mathbb{C}^3$.*

    (b) *the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ form a basis of $\mathbb{C}^3$.*

4.2 Matrix of a linear transformation

In the previous section, we learnt the definition of a linear transformation. We also saw in Example 4.1.2.5 that for each $A \in M_{m\times n}(\mathbb{C})$, there exists a linear transformation $T_A : \mathbb{C}^n \longrightarrow \mathbb{C}^m$ given by $T_A(\mathbf{x}^t) = A\mathbf{x}$ for each $\mathbf{x}^t \in \mathbb{C}^n$. In this section, we prove that every linear transformation over finite dimensional vector spaces corresponds to a matrix. Before proceeding further, we advise the reader to recall the results on ordered basis, studied in Section 3.4.

Let V and W be finite dimensional vector spaces over $\mathbb{F}$ with dimensions n and m, respectively. Also, let $\mathcal{B}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B}_2 = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be ordered bases of V and W, respectively. If $T : V \longrightarrow W$ is a linear transformation then Theorem 4.1.7 implies that $T(\mathbf{v}) \in W$ is a linear combination of the vectors $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$. So, let us find the coordinate vectors $[T(\mathbf{v}_j)]_{\mathcal{B}_2}$ for each $j = 1, 2, \ldots, n$. Let us assume that

$$[T(\mathbf{v}_1)]_{\mathcal{B}_2} = (a_{11}, \ldots, a_{m1})^t, [T(\mathbf{v}_2)]_{\mathcal{B}_2} = (a_{12}, \ldots, a_{m2})^t, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}_2} = (a_{1n}, \ldots, a_{mn})^t.$$

Or equivalently,

$$T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + a_{2j}\mathbf{w}_2 + \cdots + a_{mj}\mathbf{w}_m = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i \text{ for } j = 1, 2, \ldots, n. \tag{4.2.1}$$

Therefore, for a fixed $\mathbf{x} \in V$, if $[\mathbf{x}]_{\mathcal{B}_1} = (x_1, x_2, \ldots, x_n)^t$ then

$$T(\mathbf{x}) = T\left(\sum_{j=1}^{n} x_j\mathbf{v}_j\right) = \sum_{j=1}^{n} x_j T(\mathbf{v}_j) = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij}\mathbf{w}_i\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}x_j\right)\mathbf{w}_i. \tag{4.2.2}$$

Hence, using Equation (4.2.2), the coordinates of $T(\mathbf{x})$ with respect to the basis $\mathcal{B}_2$ equals

$$[T(\mathbf{x})]_{\mathcal{B}_2} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{mj}x_j \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\,[\mathbf{x}]_{\mathcal{B}_1},$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \left[[T(\mathbf{v}_1)]_{\mathcal{B}_2}, [T(\mathbf{v}_2)]_{\mathcal{B}_2}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}_2}\right]. \tag{4.2.3}$$

The above observations lead to the following theorem and the subsequent definition.

Theorem 4.2.1. *Let V and W be finite dimensional vector spaces over $\mathbb{F}$ with dimensions n and m, respectively. Let $T : V \longrightarrow W$ be a linear transformation. Also, let $\mathcal{B}_1$ and $\mathcal{B}_2$ be ordered bases of V and W, respectively. Then there exists a matrix $A \in M_{m\times n}(\mathbb{F})$, denoted $A = T[\mathcal{B}_1, \mathcal{B}_2]$, with $A = \left[[T(\mathbf{v}_1)]_{\mathcal{B}_2}, [T(\mathbf{v}_2)]_{\mathcal{B}_2}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}_2}\right]$ such that*

$$[T(\mathbf{x})]_{\mathcal{B}_2} = A\ \ [x]_{\mathcal{B}_1}.$$

Definition 4.2.2 (Matrix of a Linear Transformation). *Let V and W be finite dimensional vector spaces over $\mathbb{F}$ with dimensions n and m, respectively. Let $T : V \longrightarrow W$ be a linear transformation. Then the matrix $T[\mathcal{B}_1, \mathcal{B}_2]$ is called the matrix of the linear transformation with respect to the ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$.*

Remark 4.2.3. *Let $\mathcal{B}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B}_2 = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be ordered bases of V and W, respectively. Also, let $T : V \longrightarrow W$ be a linear transformation. Then writing $T[\mathcal{B}_1, \mathcal{B}_2]$ in place of the matrix A, Equation (4.2.1) can be rewritten as*

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} T[\mathcal{B}_1, \mathcal{B}_2]_{ij} \mathbf{w}_i, \quad \text{for } 1 \le j \le n. \tag{4.2.4}$$

We now give a few examples to understand the above discussion and Theorem 4.2.1.

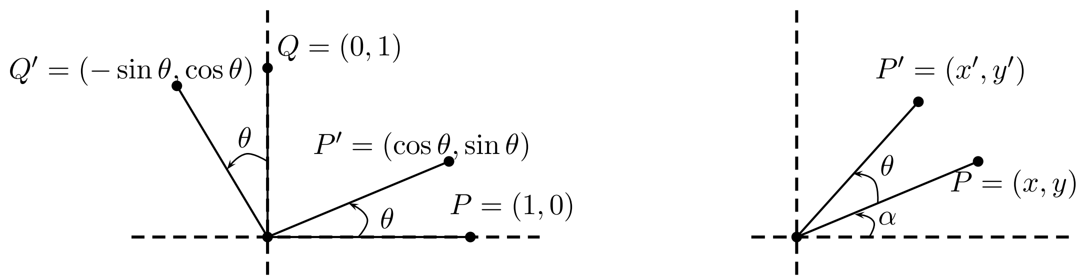

Figure 4.1: Counter-clockwise Rotation by an angle θ

Example 4.2.4. 1. *Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a function that counterclockwise rotates every point in $\mathbb{R}^2$ by an angle $\theta, 0 \le \theta < 2\pi$. Then using Figure 4.1 it can be checked that $x' = OP' \cos(\alpha + \theta) = OP\big(\cos\alpha\cos\theta - \sin\alpha\sin\theta\big) = x\cos\theta - y\sin\theta$ and similarly $y' = x\sin\theta + y\cos\theta$. Or equivalently, if $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2)$ is the standard ordered basis of $\mathbb{R}^2$, then using $T(1,0) = (\cos\theta, \sin\theta)$ and $T(0,1) = (-\sin\theta, \cos\theta)$, we get*

$$T[\mathcal{B}, \mathcal{B}] = \Big[[T(1,0)]_{\mathcal{B}}, [T(0,1)]_{\mathcal{B}}\Big] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{4.2.5}$$

2. *Let $\mathcal{B}_1 = \big((1,0),(0,1)\big)$ and $\mathcal{B}_2 = \big((1,1),(1,-1)\big)$ be two ordered bases of $\mathbb{R}^2$. Then Compute $T[\mathcal{B}_1, \mathcal{B}_1]$ and $T[\mathcal{B}_2, \mathcal{B}_2]$ for the linear transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by $T(x,y) = (x+y, x-2y)$.*

   Solution: *Observe that for $(x,y) \in \mathbb{R}^2$, $[(x,y)]_{\mathcal{B}_1} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $[(x,y)]_{\mathcal{B}_2} = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}$. Also, $T(1,0) = (1,1)$, $T(0,1) = (1,-2)$, $T(1,1) = (2,-1)$ and $T(1,-1) = (0,3)$. Thus, we have*

$$T[\mathcal{B}_1, \mathcal{B}_1] = \big[[T\big(1,0\big)]_{\mathcal{B}_1}, [T\big(0,1\big)]_{\mathcal{B}_1}\big] = \big[[(1,1)]_{\mathcal{B}_1}, [(1,-2)]_{\mathcal{B}_1}\big] = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \quad \textit{and}$$

$$T[\mathcal{B}_2, \mathcal{B}_2] = \big[[T\big(1,1\big)]_{\mathcal{B}_2}, [T\big(1,-1\big)]_{\mathcal{B}_2}\big] = \big[[(2,-1)]_{\mathcal{B}_2}, [(0,3)]_{\mathcal{B}_2}\big] = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

Hence, we see that

$$[T(x,y)]_{\mathcal{B}_1} = \big[(x+y, x-2y)\big]_{\mathcal{B}_1} = \begin{bmatrix} x+y \\ x-2y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \textit{and}$$

$$[T(x,y)]_{\mathcal{B}_2} = \big[(x+y, x-2y)\big]_{\mathcal{B}_2} = \begin{bmatrix} \frac{2x-y}{2} \\ \frac{3y}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}$$

3. *Let* $\mathcal{B}_1 = \big((1,0,0),(0,1,0),(0,0,1)\big)$ *and* $\mathcal{B}_2 = \big((1,0),(0,1)\big)$ *be ordered bases of* $\mathbb{R}^3$ *and* $\mathbb{R}^2$*, respectively. Define* $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ *by* $T(x,y,z) = (x+y-z, x+z)$*. Then*

$$T[\mathcal{B}_1, \mathcal{B}_2] = \Big[[(1,0,0)]_{\mathcal{B}_2}, [(0,1,0)]_{\mathcal{B}_2}, [(0,0,1)]_{\mathcal{B}_2}\Big] = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Check that $[T(x,y,z)]_{\mathcal{B}_2} = (x+y-z, x+z)^t = T[\mathcal{B}_1, \mathcal{B}_2]\ [(x,y,z)]_{\mathcal{B}_1}$.

4. *Let* $\mathcal{B}_1 = \big((1,0,0),(0,1,0),(0,0,1)\big)$, $\mathcal{B}_2 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ *be two ordered bases of* $\mathbb{R}^3$*. Define* $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ *by* $T(\mathbf{x}^t) = \mathbf{x}$ *for all* $\mathbf{x}^t \in \mathbb{R}^3$*. Then*

$$\begin{aligned} [T(1,0,0)]_{\mathcal{B}_2} &= 1\cdot(1,0,0) + 0\cdot(1,1,0) + 0\cdot(1,1,1) = (1,0,0)^t, \\ [T(0,1,0)]_{\mathcal{B}_2} &= -1\cdot(1,0,0) + 1\cdot(1,1,0) + 0\cdot(1,1,1) = (-1,1,0)^t, \ \textit{and} \\ [T(0,0,1)]_{\mathcal{B}_2} &= 0\cdot(1,0,0) + (-1)\cdot(1,1,0) + 1\cdot(1,1,1) = (0,-1,1)^t. \end{aligned}$$

Thus, check that

$$\begin{aligned} T[\mathcal{B}_1, \mathcal{B}_2] &= [[T(1,0,0)]_{\mathcal{B}_2},\ [T(0,1,0)]_{\mathcal{B}_2},\ [T(0,0,1)]_{\mathcal{B}_2}] \\ &= [(1,0,0)^t,\ (-1,1,0)^t,\ (0,-1,1)^t] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \\ T[\mathcal{B}_2, \mathcal{B}_1] &= [[T(1,0,0)]_{\mathcal{B}_1},\ [T(1,1,0)]_{\mathcal{B}_1},\ [T(1,1,1)]_{\mathcal{B}_1}] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned}$$

$T[\mathcal{B}_1, \mathcal{B}_1] = I_3 = T[\mathcal{B}_2, \mathcal{B}_2]$ *and* $T[\mathcal{B}_2, \mathcal{B}_1]^{-1} = T[\mathcal{B}_1, \mathcal{B}_2]$.

Remark 4.2.5. 1. *Let* V *and* W *be finite dimensional vector spaces over* $\mathbb{F}$ *with order bases* $\mathcal{B}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *and* $\mathcal{B}_2$ *of* V *and* W*, respectively. If* $T : V \longrightarrow W$ *is a linear transformation then*

(a) $T[\mathcal{B}_1, \mathcal{B}_2] = \big[[T(\mathbf{v}_1)]_{\mathcal{B}_2}, [T(\mathbf{v}_2)]_{\mathcal{B}_2}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}_2}\big]$.

(b) $[T(\mathbf{x})]_{\mathcal{B}_2} = T[\mathcal{B}_1, \mathcal{B}_2]\ [\mathbf{x}]_{\mathcal{B}_1}$ *for all* $\mathbf{x} \in V$*. That is, the coordinate vector of* $T(\mathbf{x}) \in W$ *is obtained by multiplying the matrix of the linear transformation with the coordinate vector of* $\mathbf{x} \in V$.

2. *Let* $A \in M_{m\times n}(\mathbb{R})$*. Then* A *induces a linear transformation* $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *defined by* $T_A(\mathbf{x}^t) = A\mathbf{x}$ *for all* $\mathbf{x}^t \in \mathbb{R}^n$*. Let* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *be the standard ordered bases of* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*, respectively. Then it can be easily verified that* $T_A[\mathcal{B}_1, \mathcal{B}_2] = A$.

Exercise 4.2.6. 1. *Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a linear transformation that reflects every point in $\mathbb{R}^2$ about the line $y = mx$. Find its matrix with respect to the standard ordered basis of $\mathbb{R}^2$.*

2. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be a linear transformation that reflects every point in $\mathbb{R}^3$ about the X-axis. Find its matrix with respect to the standard ordered basis of $\mathbb{R}^3$.*

3. *Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be a linear transformation that counterclockwise rotates every point in $\mathbb{R}^3$ around the positive Z-axis by an angle $\theta, 0 \le \theta < 2\pi$. Prove that T is a linear operator and find its matrix with respect to the standard ordered basis of $\mathbb{R}^3$.[Hint: Is $\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$ the required matrix?]*

4. *Define a function $D : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathcal{P}_n(\mathbb{R})$ by*

$$D(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}.$$

Prove that D is a linear operator and find the matrix of D with respect to the standard ordered basis of $\mathcal{P}_n(\mathbb{R})$. Observe that the image of D is contained in $\mathcal{P}_{n-1}(\mathbb{R})$.

5. *Let T be a linear operator in $\mathbb{R}^2$ satisfying $T(3,4) = (0,1)$ and $T(-1,1) = (2,3)$. Let $\mathcal{B} = \big((1,0),(1,1)\big)$ be an ordered basis of $\mathbb{R}^2$. Compute $T[\mathcal{B},\mathcal{B}]$.*

6. *For each linear transformation given in Example 4.1.2, find its matrix of the linear transform with respect to standard ordered bases.*

4.3 Rank-Nullity Theorem

We are now ready to related the rank-nullity theorem (see Theorem 3.3.25 on 86) with the rank-nullity theorem for linear transformation. To do so, we first define the range space and the null space of any linear transformation.

Definition 4.3.1 (Range Space and Null Space)**.** *Let V be finite dimensional vector space over $\mathbb{F}$ and let W be any vector space over $\mathbb{F}$. Then for a linear transformation $T : V \longrightarrow W$, we define*

1. *$\mathcal{C}(T) = \{T(\mathbf{x}) : \mathbf{x} \in V\}$ as the range space of T and*

2. *$\mathcal{N}(T) = \{\mathbf{x} \in V : T(\mathbf{x}) = \mathbf{0}\}$ as the null space of T.*

We now prove some results associated with the above definitions.

Proposition 4.3.2. *Let V be a vector space over $\mathbb{F}$ with basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Also, let W be a vector spaces over $\mathbb{F}$. Then for any linear transformation $T : V \longrightarrow W$,*

1. *$\mathcal{C}(T) = L(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n))$ is a subspace of W and $\dim(\mathcal{C}(T) \le \dim(W)$.*

2. *$\mathcal{N}(T)$ is a subspace of V and $\dim(\mathcal{N}(T) \le \dim(V)$.*

3. *The following statements are equivalent.*

   (a) *T is one-one.*

   (b) *$\mathcal{N}(T) = \{\mathbf{0}\}$.*

   (c) *$\{T(u_i) : 1 \le i \le n\}$ is a basis of $\mathcal{C}(T)$.*

4. $\dim(\mathcal{C}(T) = \dim(V)$ *if and only if* $\mathcal{N}(T) = \{\mathbf{0}\}$.

Proof. Parts 1 and 2 The results about $\mathcal{C}(T)$ and $\mathcal{N}(T)$ can be easily proved. We thus leave the proof for the readers.
We now assume that T is one-one. We need to show that $\mathcal{N}(T) = \{\mathbf{0}\}$.
Let $\mathbf{u} \in \mathcal{N}(T)$. Then by definition, $T(\mathbf{u}) = \mathbf{0}$. Also for any linear transformation (see Proposition 4.1.3), $T(\mathbf{0}) = \mathbf{0}$. Thus $T(\mathbf{u}) = T(\mathbf{0})$. So, T is one-one implies $\mathbf{u} = \mathbf{0}$. That is, $\mathcal{N}(T) = \{\mathbf{0}\}$.

Let $\mathcal{N}(T) = \{\mathbf{0}\}$. We need to show that T is one-one. So, let us assume that for some $\mathbf{u}, \mathbf{v} \in V$, $T(\mathbf{u}) = T(\mathbf{v})$. Then, by linearity of T, $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$. This implies, $\mathbf{u} - \mathbf{v} \in \mathcal{N}(T) = \{\mathbf{0}\}$. This in turn implies $\mathbf{u} = \mathbf{v}$. Hence, T is one-one.

The other parts can be similarly proved. $\square$

Remark 4.3.3. *1. $\mathcal{C}(T)$ is called the* RANGE SPACE *and $\mathcal{N}(T)$ the* NULL SPACE *of T.*

2. $\dim(\mathcal{C}(T)$ *is denoted by $\rho(T)$ and is called the rank of T.*

3. $\dim(\mathcal{N}(T)$ *is denoted by $\nu(T)$ and is called the nullity of T.*

Example 4.3.4. *Determine the range and null space of the linear transformation*

$$T : \mathbb{R}^3 \longrightarrow \mathbb{R}^4 \quad \text{with} \quad T(x, y, z) = (x - y + z, y - z, x, 2x - 5y + 5z).$$

Solution: *By Definition*

$$\begin{aligned}
\mathcal{R}(T) &= L\big((1,0,1,2), (-1,1,0,-5), (1,-1,0,5)\big) \\
&= L\big((1,0,1,2), (1,-1,0,5)\big) \\
&= \{\alpha(1,0,1,2) + \beta(1,-1,0,5) \ : \alpha, \beta \in \mathbb{R}\} \\
&= \{(\alpha + \beta, -\beta, \alpha, 2\alpha + 5\beta) \ : \alpha, \beta \in \mathbb{R}\} \\
&= \{(x, y, z, w) \in \mathbb{R}^4 \ : x + y - z = 0, 5y - 2z + w = 0\}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{N}(T) &= \{(x, y, z) \in \mathbb{R}^3 \ : T(x, y, z) = \mathbf{0}\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \ : (x - y + z, y - z, x, 2x - 5y + 5z) = \mathbf{0}\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \ : x - y + z = 0, y - z = 0, \\
&\qquad\qquad\qquad x = 0, 2x - 5y + 5z = 0\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \ : y - z = 0, x = 0\} \\
&= \{(0, y, y) \in \mathbb{R}^3 \ : y \in \mathbb{R}\} = L((0, 1, 1))
\end{aligned}$$

Exercise 4.3.5. *1. Define a linear operator $D : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathcal{P}_n(\mathbb{R})$ by*

$$D(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

Describe $\mathcal{N}(D)$ and $\mathcal{C}(D)$. Note that $\mathcal{C}(D) \subset \mathcal{P}_{n-1}(\mathbb{R})$.

2. *Let $T : V \longrightarrow W$ be a linear transformation. If $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ is linearly independent subset in $\mathcal{C}(T)$ then prove that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$ is linearly independent.*

3. *Define a linear operator $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $T(1,0,0) = (0,0,1)$, $T(1,1,0) = (1,1,1)$ and $T(1,1,1) = (1,1,0)$. Then*
   (a) *determine $T(x,y,z)$ for $x,y,z \in \mathbb{R}$.*
   (b) *determine $\mathcal{C}(T)$ and $\mathcal{N}(T)$. Also calculate $\rho(T)$ and $\nu(T)$.*
   (c) *prove that $T^3 = T$ and find the matrix of T with respect to the standard basis.*

4. *Find a linear operator $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ for which $\mathcal{C}(T) = L\big((1,2,0),(0,1,1),(1,3,1)\big)$?*

5. *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of a vector space $V(\mathbb{F})$. If $W(\mathbb{F})$ is a vector space and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n \in W$ then prove that there exists a unique linear transformation $T : V \longrightarrow W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for all $i = 1, 2, \ldots, n$.*

We now state the rank-nullity theorem for linear transformation. The proof of this result is similar to the proof of Theorem 3.3.25 and it also follows from Proposition 4.3.2. Hence, we omit the proof.

Theorem 4.3.6 (Rank Nullity Theorem)**.** *Let V be a finite dimensional vector space and let $T : V \longrightarrow W$ be a linear transformation. Then $\rho(T) + \nu(T) = \dim(V)$. That is,*

$$\dim(\mathcal{R}(T)) + \dim(\mathcal{N}(T)) = \dim(V).$$

Theorem 4.3.7. *Let V and W be finite dimensional vector spaces over $\mathbb{F}$ and let $T : V \longrightarrow W$ be a linear transformation. Also assume that T is one-one and onto. Then*

1. *for each $\mathbf{w} \in W$, the set $T^{-1}(\mathbf{w})$ is a set consisting of a single element.*

2. *the map $T^{-1} : W \longrightarrow V$ defined by $T^{-1}(\mathbf{w}) = \mathbf{v}$ whenever $T(\mathbf{v}) = \mathbf{w}$ is a linear transformation.*

Proof. Since T is onto, for each $\mathbf{w} \in W$ there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. So, the set $T^{-1}(\mathbf{w})$ is non-empty.

Suppose there exist vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. Then the assumption, T is one-one implies $\mathbf{v}_1 = \mathbf{v}_2$. This completes the proof of Part 1.

We are now ready to prove that T^{-1}, as defined in Part 2, is a linear transformation. Let $\mathbf{w}_1, \mathbf{w}_2 \in W$. Then by Part 1, there exist unique vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$ and $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$. Or equivalently, $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. So, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, $T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$. Hence, by definition, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, $T^{-1}(\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \alpha_1T^{-1}(\mathbf{w}_1) + \alpha_2T^{-1}(\mathbf{w}_2)$. Thus the proof of Part 2 is over. $\square$

Definition 4.3.8 (Inverse Linear Transformation)**.** *Let V and W be finite dimensional vector spaces over $\mathbb{F}$ and let $T : V \longrightarrow W$ be a linear transformation. If the map T is one-one and onto, then the map $T^{-1} : W \longrightarrow V$ defined by*

$$T^{-1}(\mathbf{w}) = \mathbf{v} \quad \textit{whenever } T(\mathbf{v}) = \mathbf{w}$$

is called the inverse of the linear transformation T.

Example 4.3.9. 1. *Let $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be defined by $T(x, y) = (x + y, x - y)$. Then $T^{-1} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is defined by $T^{-1}(x, y) = (\frac{x+y}{2}, \frac{x-y}{2})$. One can see that*

$$\begin{aligned} T \circ T^{-1}(x, y) &= T(T^{-1}(x, y)) = T(\frac{x+y}{2}, \frac{x-y}{2}) \\ &= (\frac{x+y}{2} + \frac{x-y}{2}, \frac{x+y}{2} - \frac{x-y}{2}) = (x, y) = I(x, y), \end{aligned}$$

where I is the identity operator. Hence, $T \circ T^{-1} = I$. Verify that $T^{-1} \circ T = I$. Thus, the map T^{-1} is indeed the inverse of T.

2. *For $(a_1, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$, define the linear transformation $T : \mathbb{R}^{n+1} \longrightarrow \mathcal{P}_n(\mathbb{R})$ by*

$$T(a_1, a_2, \ldots, a_{n+1}) = a_1 + a_2 x + \cdots + a_{n+1} x^n.$$

Then it can be checked that $T^{-1} : \mathcal{P}_n(\mathbb{R}) \longrightarrow \mathbb{R}^{n+1}$ is defined by $T^{-1}(a_1 + a_2 x + \cdots + a_{n+1} x^n) = (a_1, a_2, \ldots, a_{n+1})$ for all $a_1 + a_2 x + \cdots + a_{n+1} x^n \in \mathcal{P}_n(\mathbb{R})$.

Using the Rank-nullity theorem, we give a short proof of the following result.

Corollary 4.3.10. *Let V be a finite dimensional vector space and let $T : V \longrightarrow V$ be a linear operator. Then the following statements are equivalent.*

1. *T is one-one.*
2. *T is onto.*
3. *T is invertible.*

Proof. By Proposition 4.3.2, T is one-one if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$. By Theorem 4.3.6 $\mathcal{N}(T) = \{\mathbf{0}\}$ implies $\dim(\mathcal{C}(T)) = \dim(V)$. Or equivalently, T is onto.

Now, we know that T is invertible if T is one-one and onto. But we have just shown that T is one-one if and only if T is onto. Thus, we have the required result. □

Remark 4.3.11. *Let V be a finite dimensional vector space and let $T : V \longrightarrow V$ be a linear operator. If either T is one-one or T is onto then T is invertible.*

Exercise 4.3.12. 1. *Let V be a finite dimensional vector space and let $T : V \longrightarrow W$ be a linear transformation. Then prove that*

   (a) *$\mathcal{N}(T)$ and $\mathcal{C}(T)$ are also finite dimensional.*

   (b) i. *if $\dim(V) < \dim(W)$ then T cannot be onto.*

ii. if $\dim(V) > \dim(W)$ *then* T *cannot be one-one.*

2. *Let* V *be a vector space of dimension* n *and let* $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *be an ordered basis of* V. *For* $i = 1, \ldots, n$, *let* $\mathbf{w}_i \in V$ *with* $[\mathbf{w}_i]_{\mathcal{B}} = [a_{1i},\ a_{2i}, \ldots, a_{ni}\ ]^t$. *Also, let* $A = [a_{ij}]$. *Then prove that* $\mathbf{w}_1, \ldots, \mathbf{w}_n$ *is a basis of* V *if and only if* A *is invertible.*

3. *Let* $T, S : V \longrightarrow V$ *be linear transformations with* $\dim(V) = n$.

   (a) *Show that* $\mathcal{C}(T+S) \subset \mathcal{C}(T) + \mathcal{C}(S)$. *Deduce that* $\rho(T+S) \leq \rho(T) + \rho(S)$.

   (b) *Now, use Theorem 4.3.6 to prove* $\nu(T+S) \geq \nu(T) + \nu(S) - n$.

4. *Let* $z_1, z_2, \ldots, z_k$ *be* k *distinct complex numbers and define a linear transformation* $T : \mathcal{P}_n(\mathbb{C}) \longrightarrow \mathbb{C}^k$ *by* $T\big(P(z)\big) = \big(P(z_1), P(z_2), \ldots, P(z_k)\big)$. *For each* $k \geq 1$, *determine* $\dim(\mathcal{C}(T))$.

5. *Fix* $A \in M_n(\mathbb{R})$ *satisfying* $A^2 = A$ *and define* $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *by* $T_A(\mathbf{v}^t) = A\mathbf{v}$, *for all* $\mathbf{v}^t \in \mathbb{R}^n$. *Then prove that*

   (a) $T_A \circ T_A = T_A$. *Equivalently,* $T_A \circ (I - T_A) = \mathbf{0}$, *where* $I : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *is the identity map and* $\mathbf{0} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *is the zero map.*

   (b) $\mathcal{N}(T_A) \cap \mathcal{C}(T_A) = \{\mathbf{0}\}$.

   (c) $\mathbb{R}^n = \mathcal{C}(T_A) + \mathcal{N}(T_A)$. *[Hint:* $\mathbf{x} = T_A(\mathbf{x}) + (I - T_A)(\mathbf{x})$*]*

4.4 Similarity of Matrices

Let V be a finite dimensional vector space with ordered basis $\mathcal{B}$. Then we saw that any linear operator $T : V \longrightarrow V$ corresponds to a square matrix of order $\dim(V)$ and this matrix was denoted by $T[\mathcal{B}, \mathcal{B}]$. In this section, we will try to understand the relationship between $T[\mathcal{B}_1, \mathcal{B}_1]$ and $T[\mathcal{B}_2, \mathcal{B}_2]$, where $\mathcal{B}_1$ and $\mathcal{B}_2$ are distinct ordered bases of V. This will enable us to understand the reason for defining THE MATRIX PRODUCT SOMEWHAT DIFFERENTLY.

Theorem 4.4.1 (Composition of Linear Transformations)**.** *Let* V, W *and* Z *be finite dimensional vector spaces with ordered bases* $\mathcal{B}_1, \mathcal{B}_2$ *and* $\mathcal{B}_3$, *respectively. Also, let* $T : V \longrightarrow W$ *and* $S : W \longrightarrow Z$ *be linear transformations. Then the composition map* $S \circ T : V \longrightarrow Z$ *(see Figure 4.2) is a linear transformation and*

$$\begin{tikzcd}(V, \mathcal{B}_1, n) \arrow[r, "{T[\mathcal{B}_1, \mathcal{B}_2]_{m\times n}}"] \arrow[rr, bend right, "{(S\circ T)[\mathcal{B}_1, \mathcal{B}_3]_{p\times n} = S[\mathcal{B}_2, \mathcal{B}_3]\cdot T[\mathcal{B}_1, \mathcal{B}_2]}"'] & (W, \mathcal{B}_2, m) \arrow[r, "{S[\mathcal{B}_2, \mathcal{B}_3]_{p\times m}}"] & (Z, \mathcal{B}_3, p)\end{tikzcd}$$

Figure 4.2: Composition of Linear Transformations

$$(S \circ T)\ [\mathcal{B}_1, \mathcal{B}_3] = S[\mathcal{B}_2, \mathcal{B}_3]\ \cdot\ T[\mathcal{B}_1, \mathcal{B}_2].$$

Proof. Let $\mathcal{B}_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, $\mathcal{B}_2 = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ and $\mathcal{B}_3 = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$ be ordered bases of V, W and Z, respectively. Then using Equation (4.2.4), we have

$$\begin{aligned}(S \circ T)\ (\mathbf{u}_t) &= S(T(\mathbf{u}_t)) = S\left(\sum_{j=1}^{m}(T[\mathcal{B}_1, \mathcal{B}_2])_{jt}\mathbf{v}_j\right) = \sum_{j=1}^{m}(T[\mathcal{B}_1, \mathcal{B}_2])_{jt}S(\mathbf{v}_j) \\ &= \sum_{j=1}^{m}(T[\mathcal{B}_1, \mathcal{B}_2])_{jt}\sum_{k=1}^{p}(S[\mathcal{B}_2, \mathcal{B}_3])_{kj}\mathbf{w}_k = \sum_{k=1}^{p}(\sum_{j=1}^{m}(S[\mathcal{B}_2, \mathcal{B}_3])_{kj}(T[\mathcal{B}_1, \mathcal{B}_2])_{jt})\mathbf{w}_k \\ &= \sum_{k=1}^{p}(S[\mathcal{B}_2, \mathcal{B}_3]\ \ T[\mathcal{B}_1, \mathcal{B}_2])_{kt}\mathbf{w}_k.\end{aligned}$$

Thus, using matrix multiplication, the t-th column of $(S \circ T)\ [\mathcal{B}_1, \mathcal{B}_3]$ is given by

$$[(S \circ T)\ (\mathbf{u}_t)]_{\mathcal{B}_3} = \begin{bmatrix} \left(S[\mathcal{B}_2, \mathcal{B}_3]\ T[\mathcal{B}_1, \mathcal{B}_2]\right)_{1t} \\ \left(S[\mathcal{B}_2, \mathcal{B}_3]\ T[\mathcal{B}_1, \mathcal{B}_2]\right)_{2t} \\ \vdots \\ \left(S[\mathcal{B}_2, \mathcal{B}_3]\ T[\mathcal{B}_1, \mathcal{B}_2]\right)_{pt} \end{bmatrix} = S[\mathcal{B}_2, \mathcal{B}_3] \begin{bmatrix} T[\mathcal{B}_1, \mathcal{B}_2]_{1t} \\ T[\mathcal{B}_1, \mathcal{B}_2]_{2t} \\ \vdots \\ T[\mathcal{B}_1, \mathcal{B}_2]_{pt} \end{bmatrix}.$$

Hence, $(S \circ T)[\mathcal{B}_1, \mathcal{B}_3] = \left[[(S \circ T)(u_1)]_{\mathcal{B}_3}, \ldots, [(S \circ T)(u_n)]_{\mathcal{B}_3}\right] = S[\mathcal{B}_2, \mathcal{B}_3] \cdot T[\mathcal{B}_1, \mathcal{B}_2]$ and the proof of the theorem is over. ☐

Proposition 4.4.2. *Let V be a finite dimensional vector space and let $T, S : V \longrightarrow V$ be two linear operators. Then $\nu(T) + \nu(S) \geq \nu(T \circ S) \geq \max\{\nu(T), \nu(S)\}$.*

Proof. We first prove the second inequality.
Suppose $\mathbf{v} \in \mathcal{N}(S)$. Then $(T \circ S)(\mathbf{v}) = T(S(\mathbf{v}) = T(\mathbf{0}) = \mathbf{0}$ gives $\mathcal{N}(S) \subset \mathcal{N}(T \circ S)$. Therefore, $\nu(S) \leq \nu(T \circ S)$.

We now use Theorem 4.3.6 to see that the inequality $\nu(T) \leq \nu(T \circ S)$ is equivalent to showing $\mathcal{C}(T \circ S) \subset \mathcal{C}(T)$. But this holds true as $\mathcal{C}(S) \subset V$ and hence $T(\mathcal{C}(S)) \subset T(V)$. Thus, the proof of the second inequality is over.

For the proof of the first inequality, assume that $k = \nu(S)$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis of $\mathcal{N}(S)$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathcal{N}(T \circ S)$ as $T(\mathbf{0}) = \mathbf{0}$. So, let us extend it to get a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ of $\mathcal{N}(T \circ S)$.

Claim: $\{S(\mathbf{u}_1), S(\mathbf{u}_2), \ldots, S(\mathbf{u}_\ell)\}$ is a linearly independent subset of $\mathcal{N}(T)$.

It is easily seen that $\{S(\mathbf{u}_1), \ldots, S(\mathbf{u}_\ell)\}$ is a subset of $\mathcal{N}(T)$. So, let us solve the linear system $c_1 S(\mathbf{u}_1) + c_2 S(\mathbf{u}_2) + \cdots + c_\ell S(\mathbf{u}_\ell) = \mathbf{0}$ in the unknowns $c_1, c_2, \ldots, c_\ell$. This system is equivalent to $S(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_\ell\mathbf{u}_\ell) = \mathbf{0}$. That is, $\sum\limits_{i=1}^{\ell} c_i\mathbf{u}_i \in \mathcal{N}(S)$. Hence, $\sum\limits_{i=1}^{\ell} c_i\mathbf{u}_i$ is a unique linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Thus,

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_\ell\mathbf{u}_\ell = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k \qquad (4.4.1)$$

for some scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$. But by assumption, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is a basis of $\mathcal{N}(T \circ S)$ and hence linearly independent. Therefore, the only solution of Equation (4.4.1) is given by $c_i = 0$ for $1 \leq i \leq \ell$ and $\alpha_j = 0$ for $1 \leq j \leq k$.

Thus, $\{S(\mathbf{u}_1), S(\mathbf{u}_2), \ldots, S(\mathbf{u}_\ell)\}$ is a linearly independent subset of $\mathcal{N}(T)$ and so $\nu(T) \geq \ell$. Hence, $\nu(T \circ S) = k + \ell \leq \nu(S) + \nu(T)$. ☐

Remark 4.4.3. *Using Theorem 4.3.6 and Proposition 4.4.2, we see that if A and B are two $n \times n$ matrices then*

$$\min\{\rho(A), \rho(B)\} \geq \rho(AB) \geq n - \rho(A) - \rho(B).$$

Let V be a finite dimensional vector space and let $T : V \longrightarrow V$ be an invertible linear operator. Then using Theorem 4.3.7, the map $T^{-1} : V \longrightarrow V$ is a linear operator defined by $T^{-1}(\mathbf{u}) = \mathbf{v}$ whenever $T(\mathbf{v}) = \mathbf{u}$. The next result relates the matrix of T and T^{-1}. The reader is required to supply the proof (use Theorem 4.4.1).

Theorem 4.4.4 (Inverse of a Linear Transformation)**.** *Let V be a finite dimensional vector space with ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$. Also let $T : V \longrightarrow V$ be an invertible linear operator. Then the matrix of T and T^{-1} are related by $T[\mathcal{B}_1, \mathcal{B}_2]^{-1} = T^{-1}[\mathcal{B}_2, \mathcal{B}_1]$.*

Exercise 4.4.5. *Find the matrix of the linear transformations given below.*

1. *Define $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $T(1, 1, 1) = (1, -1, 1)$, $T(1, -1, 1) = (1, 1, -1)$ and $T(1, 1, -1) = (1, 1, 1)$. Find $T[\mathcal{B}, \mathcal{B}]$, where $\mathcal{B} = \big((1, 1, 1), (1, -1, 1), (1, 1, -1)\big)$. Is T an invertible linear operator?*

2. *Let $\mathcal{B} = \big(1, x, x^2, x^3\big)$ be an ordered basis of $\mathcal{P}_3(\mathbb{R})$. Define $T : \mathcal{P}_3(\mathbb{R}) \longrightarrow \mathcal{P}_3(\mathbb{R})$ by*

$$T(1) = 1, T(x) = 1 + x, T(x^2) = (1 + x)^2 \text{ and } T(x^3) = (1 + x)^3.$$

Prove that T is an invertible linear operator. Also, find $T[\mathcal{B}, \mathcal{B}]$ and $T^{-1}[\mathcal{B}, \mathcal{B}]$.

We end this section with definition, results and examples related with the notion of isomorphism. The result states that for each fixed positive integer n, every real vector space of dimension n is isomorphic to $\mathbb{R}^n$ and every complex vector space of dimension n is isomorphic to $\mathbb{C}^n$.

Definition 4.4.6 (Isomorphism)**.** *Let V and W be two vector spaces over $\mathbb{F}$. Then V is said to be isomorphic to W if there exists a linear transformation $T : V \longrightarrow W$ that is one-one, onto and invertible. We also denote it by $V \cong W$.*

Theorem 4.4.7. *Let V be a vector space over $\mathbb{R}$. If $\dim(V) = n$ then $V \cong \mathbb{R}^n$.*

Proof. Let $\mathcal{B}$ be the standard ordered basis of $\mathbb{R}^n$ and let $\mathcal{B}_1 = \big(\mathbf{v}_1, \ldots, \mathbf{v}_n\big)$ be an ordered basis of V. Define a map $T : V \longrightarrow \mathbb{R}^n$ by $T(\mathbf{v}_i) = \mathbf{e}_i$ for $1 \leq i \leq n$. Then it can be easily verified that T is a linear transformation that is one-one, onto and invertible (the image of a basis vector is a basis vector). Hence, the result follows. □

A similar idea leads to the following result and hence we omit the proof.

Theorem 4.4.8. *Let V be a vector space over $\mathbb{C}$. If $\dim(V) = n$ then $V \cong \mathbb{C}^n$.*

Example 4.4.9.

1. *The standard ordered basis of $\mathcal{P}_n(\mathbb{C})$ is given by $\big(1, x, x^2, \ldots, x^n\big)$. Hence, define $T : \mathcal{P}_n(\mathbb{C}) \longrightarrow \mathbb{C}^{n+1}$ by $T(x^i) = \mathbf{e}_{i+1}$ for $0 \leq i \leq n$. In general, verify that $T(a_0 + a_1 x + \cdots + a_n x^n) = (a_0, a_1, \ldots, a_n)$ and T is linear transformation which is one-one, onto and invertible. Thus, the vector space $\mathcal{P}_n(\mathbb{C}) \cong \mathbb{C}^{n+1}$.*

2. *Let $V = \{(x,y,z,w) \in \mathbb{R}^4 : x - y + z - w = 0\}$. Suppose that $\mathcal{B}$ is the standard ordered basis of $\mathbb{R}^3$ and $\mathcal{B}_1 = \big((1,1,0,0), (-1,0,1,0), (1,0,0,1)\big)$ is the ordered basis of V. Then $T : V \longrightarrow \mathbb{R}^3$ defined by $T(\mathbf{v}) = T(y-z+w, y, z, w) = (y, z, w)$ is a linear transformation and $T[\mathcal{B}_1, \mathcal{B}] = I_3$. Thus, T is one-one, onto and invertible.*

4.5 Change of Basis

Let V be a vector space with ordered bases $\mathcal{B}_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Also, recall that the identity linear operator $I : V \longrightarrow V$ is defined by $I(\mathbf{x}) = \mathbf{x}$ for every $\mathbf{x} \in V$. If

$$I[\mathcal{B}_2, \mathcal{B}_1] = \big[[I(\mathbf{v}_1)]_{\mathcal{B}_1}, [I(\mathbf{v}_2)]_{\mathcal{B}_1}, \ldots, [I(\mathbf{v}_n)]_{\mathcal{B}_1}\big] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

then by definition of $I[\mathcal{B}_2, \mathcal{B}_1]$, we see that $\mathbf{v}_i = I(\mathbf{v}_i) = \sum\limits_{j=1}^{n} a_{ji}\mathbf{u}_j$ for all $i, 1 \le i \le n$. Thus, we have proved the following result which also appeared in another form in Theorem 3.4.5.

Theorem 4.5.1 (Change of Basis Theorem)**.** *Let V be a finite dimensional vector space with ordered bases $\mathcal{B}_1 = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\mathcal{B}_2 = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Suppose $\mathbf{x} \in V$ with $[\mathbf{x}]_{\mathcal{B}_1} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$ and $[\mathbf{x}]_{\mathcal{B}_2} = (\beta_1, \beta_2, \ldots, \beta_n)^t$. Then $[\mathbf{x}]_{\mathcal{B}_1} = I[\mathcal{B}_2, \mathcal{B}_1] \ [\mathbf{x}]_{\mathcal{B}_2}$. Or equivalently,*

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}.$$

Remark 4.5.2. *Observe that the identity linear operator $I : V \longrightarrow V$ is invertible and hence by Theorem 4.4.4 $I[\mathcal{B}_2, \mathcal{B}_1]^{-1} = I^{-1}[\mathcal{B}_1, \mathcal{B}_2] = I[\mathcal{B}_1, \mathcal{B}_2]$. Therefore, we also have $[\mathbf{x}]_{\mathcal{B}_2} = I[\mathcal{B}_1, \mathcal{B}_2] \ [\mathbf{x}]_{\mathcal{B}_1}$.*

Let V be a finite dimensional vector space with ordered bases $\mathcal{B}_1$ and $\mathcal{B}_2$. Then for any linear operator $T : V \longrightarrow V$ the next result relates $T[\mathcal{B}_1, \mathcal{B}_1]$ and $T[\mathcal{B}_2, \mathcal{B}_2]$.

Theorem 4.5.3. *Let $\mathcal{B}_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and $\mathcal{B}_2 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ be two ordered bases of a vector space V. Also, let $A = [a_{ij}] = I[\mathcal{B}_2, \mathcal{B}_1]$ be the matrix of the identity linear operator. Then for any linear operator $T : V \longrightarrow V$*

$$T[\mathcal{B}_2, \mathcal{B}_2] = A^{-1} \cdot T[\mathcal{B}_1, \mathcal{B}_1] \cdot A = I[\mathcal{B}_1, \mathcal{B}_2] \cdot T[\mathcal{B}_1, \mathcal{B}_1] \cdot I[\mathcal{B}_2, \mathcal{B}_1]. \qquad (4.5.2)$$

Proof. The proof uses Theorem 4.4.1 by representing $T[\mathcal{B}_1, \mathcal{B}_2]$ as $(I \circ T)[\mathcal{B}_1, \mathcal{B}_2]$ and $(T \circ I)[\mathcal{B}_1, \mathcal{B}_2]$, where I is the identity operator on V (see Figure 4.3). By Theorem 4.4.1, we have

$$\begin{aligned} T[\mathcal{B}_1, \mathcal{B}_2] &= (I \circ T)[\mathcal{B}_1, \mathcal{B}_2] = I[\mathcal{B}_1, \mathcal{B}_2] \cdot T[\mathcal{B}_1, \mathcal{B}_1] \\ &= (T \circ I)[\mathcal{B}_1, \mathcal{B}_2] = T[\mathcal{B}_2, \mathcal{B}_2] \cdot I[\mathcal{B}_1, \mathcal{B}_2]. \end{aligned}$$

$$
\begin{tikzcd}
(V,\mathcal{B}_1) \arrow[r, "{T[\mathcal{B}_1,\mathcal{B}_1]}"] \arrow[d, "{I[\mathcal{B}_1,\mathcal{B}_2]}"'] \arrow[dr, bend left, "I\circ T"] \arrow[dr, bend right, "T\circ I"'] & (V,\mathcal{B}_1) \arrow[d, "{I[\mathcal{B}_1,\mathcal{B}_2]}"] \\
(V,\mathcal{B}_2) \arrow[r, "{T[\mathcal{B}_2,\mathcal{B}_2]}"'] & (V,\mathcal{B}_2)
\end{tikzcd}
$$

Figure 4.3: Commutative Diagram for Similarity of Matrices

Thus, using $I[\mathcal{B}_2,\mathcal{B}_1] = I[\mathcal{B}_1,\mathcal{B}_2]^{-1}$, we get $I[\mathcal{B}_1,\mathcal{B}_2]\ T[\mathcal{B}_1,\mathcal{B}_1]\ I[\mathcal{B}_2,\mathcal{B}_1] = T[\mathcal{B}_2,\mathcal{B}_2]$ and the result follows. $\square$

Let $T : V \longrightarrow V$ be a linear operator on V. If $\dim(V) = n$ then each ordered basis $\mathcal{B}$ of V gives rise to an $n \times n$ matrix $T[\mathcal{B},\mathcal{B}]$. Also, we know that for any vector space we have infinite number of choices for an ordered basis. So, as we change an ordered basis, the matrix of the linear transformation changes. Theorem 4.5.3 tells us that all these matrices are related by an invertible matrix (see Remark 4.5.2). Thus we are led to the following remark and the definition.

Remark 4.5.4. *The Equation (4.5.2) shows that* $T[\mathcal{B}_2,\mathcal{B}_2] = I[\mathcal{B}_1,\mathcal{B}_2]\cdot T[\mathcal{B}_1,\mathcal{B}_1]\cdot I[\mathcal{B}_2,\mathcal{B}_1]$. *Hence, the matrix* $I[\mathcal{B}_1,\mathcal{B}_2]$ *is called the* $\mathcal{B}_1 : \mathcal{B}_2$ *change of basis matrix.*

Definition 4.5.5 (Similar Matrices)**.** *Two square matrices B and C of the same order are said to be similar if there exists a non-singular matrix P such that $P^{-1}BP = C$ or equivalently $BP = PC$.*

Example 4.5.6.

1. *Let* $\mathcal{B}_1 = \big(1+x, 1+2x+x^2, 2+x\big)$ *and* $\mathcal{B}_2 = \big(1, 1+x, 1+x+x^2\big)$ *be ordered bases of* $\mathcal{P}_2(\mathbb{R})$. *Then* $I(a+bx+cx^2) = a+bx+cx^2$. *Thus,*

$$
\begin{aligned}
I[\mathcal{B}_2,\mathcal{B}_1] &= [[1]_{\mathcal{B}_1}, [1+x]_{\mathcal{B}_1}, [1+x+x^2]_{\mathcal{B}_1}] = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and} \\
I[\mathcal{B}_1,\mathcal{B}_2] &= [[1+x]_{\mathcal{B}_2}, [1+2x+x^2]_{\mathcal{B}_2}, [2+x]_{\mathcal{B}_2}] = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.
\end{aligned}
$$

   Also, verify that $I[\mathcal{B}_1,\mathcal{B}_2]^{-1} = I[\mathcal{B}_2,\mathcal{B}_1]$.

2. *Let* $\mathcal{B}_1 = \big((1,0,0),(1,1,0),(1,1,1)\big)$ *and* $\mathcal{B}_2 = \big(1,1,-1),(1,2,1),(2,1,1)\big)$ *be two ordered bases of* $\mathbb{R}^3$. *Define* $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ *by* $T(x,y,z) = (x+y, x+y+2z, y-z)$. *Then* $T[\mathcal{B}_1,\mathcal{B}_1] = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 1 & 4 \\ 0 & 1 & 0 \end{bmatrix}$ *and* $T[\mathcal{B}_2,\mathcal{B}_2] = \begin{bmatrix} -4/5 & 1 & 8/5 \\ -2/5 & 2 & 9/5 \\ 8/5 & 0 & -1/5 \end{bmatrix}$. *Also, check that*

$$I[\mathcal{B}_2, \mathcal{B}_1] = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \text{ and}$$

$$T[\mathcal{B}_1, \mathcal{B}_1]\; I[\mathcal{B}_2, \mathcal{B}_1] = I[\mathcal{B}_2, \mathcal{B}_1]\; T[\mathcal{B}_2, \mathcal{B}_2] = \begin{bmatrix} 2 & -2 & -2 \\ -2 & 4 & 5 \\ 2 & 1 & 0 \end{bmatrix}.$$

Exercise 4.5.7. *1. Let V be an n-dimensional vector space and let $T : V \longrightarrow V$ be a linear operator. Suppose T has the property that $T^{n-1} \neq \mathbf{0}$ but $T^n = \mathbf{0}$.*

(a) Prove that there exists $\mathbf{u} \in V$ with $\{\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\}$, a basis of V.

(b) For $\mathcal{B} = \big(\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\big)$ prove that $T[\mathcal{B}, \mathcal{B}] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$

(c) Let A be an $n \times n$ matrix satisfying $A^{n-1} \neq \mathbf{0}$ but $A^n = \mathbf{0}$. Then prove that A is similar to the matrix given in Part 1b.

2. Define $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $T(x, y, z) = (x + y + 2z, x - y - 3z, 2x + 3y + z)$. Let $\mathcal{B}$ be the standard basis and $\mathcal{B}_1 = \big((1, 1, 1), (1, -1, 1), (1, 1, 2)\big)$ be another ordered basis of $\mathbb{R}^3$. Then find the

(a) matrices $T[\mathcal{B}, \mathcal{B}]$ and $T[\mathcal{B}_1, \mathcal{B}_1]$.

(b) matrix P such that $P^{-1}T[\mathcal{B}, \mathcal{B}]\, P = T[\mathcal{B}_1, \mathcal{B}_1]$.

3. Define $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ by $T(x, y, z) = (x, x + y, x + y + z)$. Let $\mathcal{B}$ be the standard basis and $\mathcal{B}_1 = \big((1, 0, 0), (1, 1, 0), (1, 1, 1)\big)$ be another ordered basis of $\mathbb{R}^3$. Then find the

(a) matrices $T[\mathcal{B}, \mathcal{B}]$ and $T[\mathcal{B}_1, \mathcal{B}_1]$.

(b) matrix P such that $P^{-1}T[\mathcal{B}, \mathcal{B}]\, P = T[\mathcal{B}_1, \mathcal{B}_1]$.

4. Let $\mathcal{B}_1 = \big((1, 2, 0), (1, 3, 2), (0, 1, 3)\big)$ and $\mathcal{B}_2 = \big((1, 2, 1), (0, 1, 2), (1, 4, 6)\big)$ be two ordered bases of $\mathbb{R}^3$. Find the change of basis matrix

(a) P from $\mathcal{B}_1$ to $\mathcal{B}_2$.

(b) Q from $\mathcal{B}_2$ to $\mathcal{B}_1$.

(c) from the standard basis of $\mathbb{R}^3$ to $\mathcal{B}_1$. What do you notice?

Is it true that $PQ = I = QP$? Give reasons for your answer.

4.6 Summary

Chapter 5

Inner Product Spaces

5.1 Introduction

In the previous chapters, we learnt about vector spaces and linear transformations that are maps (functions) between vector spaces. In this chapter, we will start with the definition of inner product that helps us to view vector spaces geometrically.

5.2 Definition and Basic Properties

In $\mathbb{R}^2$ and $\mathbb{R}^3$, we had a notion of dot product between two vectors. In particular, if $\mathbf{x}^t = (x_1, x_2, x_3)$, $\mathbf{y}^t = (y_1, y_2, y_3)$ are two vectors in $\mathbb{R}^3$ then their dot product was defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

Note that for any $\mathbf{x}^t, \mathbf{y}^t, \mathbf{z}^t \in \mathbb{R}^3$ and $\alpha \in \mathbb{R}$, the dot product satisfied the following conditions:

$$\mathbf{x} \cdot (\mathbf{y} + \alpha \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \alpha \mathbf{x} \cdot \mathbf{z}, \ \mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}, \ \text{ and } \ \mathbf{x} \cdot \mathbf{x} \geq 0.$$

Also, $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. So, in this chapter, we generalize the idea of dot product for arbitrary vector spaces. This generalization is commonly known as inner product which is our starting point for this chapter.

Definition 5.2.1 (Inner Product)**.** *Let V be a vector space over $\mathbb{F}$. An inner product over V, denoted by $\langle \, , \, \rangle$, is a map from $V \times V$ to $\mathbb{F}$ satisfying*

1. $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$, *for all* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ *and* $a, b \in \mathbb{F}$,
2. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, *the complex conjugate of* $\langle \mathbf{u}, \mathbf{v} \rangle$, *for all* $\mathbf{u}, \mathbf{v} \in V$ *and*
3. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ *for all* $\mathbf{u} \in V$ *and equality holds if and only if* $\mathbf{u} = \mathbf{0}$.

Definition 5.2.2 (Inner Product Space)**.** *Let V be a vector space with an inner product $\langle \, , \, \rangle$. Then $(V, \langle \, , \, \rangle)$ is called an inner product space (in short,* IPS*).*

Example 5.2.3. *The first two examples given below are called the* STANDARD INNER PRODUCT *or the* DOT PRODUCT *on* $\mathbb{R}^n$ *and* $\mathbb{C}^n$*, respectively. From now on, whenever an inner product is not mentioned, it will be assumed to be the standard inner product.*

1. *Let* $V = \mathbb{R}^n$. *Define* $\langle u, v\rangle = u_1v_1 + \cdots + u_nv_n = \mathbf{u}^t\mathbf{v}$ *for all* $\mathbf{u}^t = (u_1, \ldots, u_n), \mathbf{v}^t = (v_1, \ldots, v_n) \in V$. *Then it can be easily verified that* $\langle\ ,\ \rangle$ *satisfies all the three conditions of Definition 5.2.1. Hence,* $\big(\mathbb{R}^n, \langle\ ,\ \rangle\big)$ *is an inner product space.*

2. *Let* $\mathbf{u}^t = (u_1, \ldots, u_n), \mathbf{v}^t = (v_1, \ldots, v_n)$ *be two vectors in* $\mathbb{C}^n(\mathbb{C})$. *Define* $\langle u, v\rangle = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n} = \mathbf{u}^*\mathbf{v}$. *Then it can be easily verified that* $\big(\mathbb{C}^n, \langle\ ,\ \rangle\big)$ *is an inner product space.*

3. *Let* $V = \mathbb{R}^2$ *and let* $A = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$. *Define* $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^t A\mathbf{y}$ *for* $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^2$. *Then prove that* $\langle\ ,\ \rangle$ *is an inner product. Hint:* $\langle \mathbf{x}, \mathbf{y}\rangle = 4x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$ *and* $\langle \mathbf{x}, \mathbf{x}\rangle = (x_1 - x_2)^2 + 3x_1^2 + x_2^2$.

4. *Prove that* $\langle \mathbf{x}, \mathbf{y}\rangle = 10x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2 + x_2y_3 + x_3y_2 + x_3y_3$ *defines an inner product in* $\mathbb{R}^3$, *where* $\mathbf{x}^t = (x_1, x_2, x_3)$ *and* $\mathbf{y}^t = (y_1, y_2, y_3) \in \mathbb{R}^3$.

5. *For* $\mathbf{x}^t = (x_1, x_2), \mathbf{y}^t = (y_1, y_2) \in \mathbb{R}^2$, *we define three maps that satisfy at least one condition out of the three conditions for an inner product. Determine the condition which is not satisfied. Give reasons for your answer.*

   (a) $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1$.
   (b) $\langle \mathbf{x}, \mathbf{y}\rangle = x_1^2 + y_1^2 + x_2^2 + y_2^2$.
   (c) $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1^3 + x_2y_2^3$.

6. *For* $A, B \in M_n(\mathbb{R})$, *define* $\langle A, B\rangle = tr(AB^t)$. *Then*
$$\langle A + B, C\rangle = tr\big((A + B)C^t\big) = tr(AC^t) + tr(BC^t) = \langle A, C\rangle + \langle B, C\rangle.$$
$$\langle A, B\rangle = tr(AB^t) = tr(\ (AB^t)^t\ ) = tr(BA^t) = \langle B, A\rangle.$$
If $A = (a_{ij})$, *then* $\langle A, A\rangle = tr(AA^t) = \sum\limits_{i=1}^{n}(AA^t)_{ii} = \sum\limits_{i,j=1}^{n} a_{ij}a_{ij} = \sum\limits_{i,j=1}^{n} a_{ij}^2$ *and therefore,* $\langle A, A\rangle > 0$ *for all non-zero matrix* A.

Exercise 5.2.4. 1. *Verify that inner products defined in Examples 3 and 4, are indeed inner products.*

2. *Let* $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ *for every vector* $\mathbf{y}$ *of an inner product space* V. *prove that* $\mathbf{x} = \mathbf{0}$.

Definition 5.2.5 (Length/Norm of a Vector)**.** *Let* V *be a vector space. Then for any vector* $\mathbf{u} \in V$, *we define the length (norm) of* $\mathbf{u}$, *denoted* $\|\mathbf{u}\|$, *by* $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$, *the positive square root. A vector of norm 1 is called a unit vector.*

Example 5.2.6. 1. *Let* V *be an inner product space and* $\mathbf{u} \in V$. *Then for any scalar* α, *it is easy to verify that* $\|\alpha\mathbf{u}\| = \big|\alpha\big| \cdot \|\mathbf{u}\|$.

2. *Let* $\mathbf{u}^t = (1,-1,2,-3) \in \mathbb{R}^4$. *Then* $\|\mathbf{u}\| = \sqrt{1+1+4+9} = \sqrt{15}$. *Thus,* $\frac{1}{\sqrt{15}}\mathbf{u}$ *and* $-\frac{1}{\sqrt{15}}\mathbf{u}$ *are vectors of norm 1 in the vector subspace* $L(\mathbf{u})$ *of* $\mathbb{R}^4$. *Or equivalently,* $\frac{1}{\sqrt{15}}\mathbf{u}$ *is a unit vector in the direction of* $\mathbf{u}$.

Exercise 5.2.7. 1. *Let* $\mathbf{u}^t = (-1,1,2,3,7) \in \mathbb{R}^5$. *Find all* $\alpha \in \mathbb{R}$ *such that* $\|\alpha\mathbf{u}\| = 1$.

2. *Let* $\mathbf{u}^t = (-1,1,2,3,7) \in \mathbb{C}^5$. *Find all* $\alpha \in \mathbb{C}$ *such that* $\|\alpha\mathbf{u}\| = 1$.

3. *Prove that* $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\big(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\big)$, *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. *This equality is commonly known as the* PARALLELOGRAM LAW *as in a parallelogram the sum of square of the lengths of the diagonals equals twice the sum of squares of the lengths of the sides.*

4. *Prove that for any two continuous functions* $f(x), g(x) \in C([-1,\ 1])$, *the map* $\langle f(x), g(x)\rangle = \int_{-1}^{1} f(x)\cdot g(x)dx$ *defines an inner product in* $C([-1,\ 1])$.

5. *Fix an ordered basis* $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ *of a complex vector space* V. *Prove that* $\langle\ ,\ \rangle$ *defined by* $\langle \mathbf{u}, \mathbf{v}\rangle = \sum\limits_{i=1}^{n} a_i\overline{b_i}$, *whenever* $[\mathbf{u}]_{\mathcal{B}} = (a_1, \ldots, a_n)^t$ *and* $[\mathbf{v}]_{\mathcal{B}} = (b_1, \ldots, b_n)^t$ *is indeed an inner product in* V.

A very useful and a fundamental inequality concerning the inner product is due to Cauchy and Schwarz. The next theorem gives the statement and a proof of this inequality.

Theorem 5.2.8 (Cauchy-Bunyakovskii-Schwartz inequality)**.** *Let* $V(\mathbb{F})$ *be an inner product space. Then for any* $\mathbf{u}, \mathbf{v} \in V$

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\ \|\mathbf{v}\|. \tag{5.2.1}$$

Equality holds in Equation (5.2.1) if and only if the vectors $\mathbf{u}$ *and* $\mathbf{v}$ *are linearly dependent. Furthermore, if* $\mathbf{u} \neq \mathbf{0}$, *then in this case* $\mathbf{v} = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$.

Proof. If $\mathbf{u} = \mathbf{0}$, then the inequality (5.2.1) holds trivially. Hence, let $\mathbf{u} \neq \mathbf{0}$. Also, by the third property of inner product, $\langle \lambda\mathbf{u}+\mathbf{v}, \lambda\mathbf{u}+\mathbf{v}\rangle \geq 0$ for all $\lambda \in \mathbb{F}$. In particular, for $\lambda = -\dfrac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}$,

$$\begin{aligned} 0 &\leq \langle \lambda\mathbf{u}+\mathbf{v}, \lambda\mathbf{u}+\mathbf{v}\rangle = \lambda\overline{\lambda}\|\mathbf{u}\|^2 + \lambda\langle \mathbf{u}, \mathbf{v}\rangle + \overline{\lambda}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \\ &= \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\|\mathbf{u}\|^2 - \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}\langle \mathbf{u}, \mathbf{v}\rangle - \frac{\overline{\langle \mathbf{v}, \mathbf{u}\rangle}}{\|\mathbf{u}\|^2}\langle \mathbf{v}, \mathbf{u}\rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{v}\|^2 - \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|^2}{\|\mathbf{u}\|^2}. \end{aligned}$$

Or, in other words $|\langle \mathbf{v}, \mathbf{u}\rangle|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2$ and the proof of the inequality is over.

If $\mathbf{u} \neq \mathbf{0}$ then $\langle \lambda\mathbf{u}+\mathbf{v}, \lambda\mathbf{u}+\mathbf{v}\rangle = 0$ if and only of $\lambda\mathbf{u}+\mathbf{v} = \mathbf{0}$. Hence, equality holds in (5.2.1) if and only if $\lambda = -\dfrac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|^2}$. That is, $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent and in this case $\mathbf{v} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$. $\square$

Let V be a real vector space. Then for every $\mathbf{u}, \mathbf{v} \in V$, the Cauchy-Schwarz inequality (see (5.2.1)) implies that $-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|} \leq 1$. Also, we know that $\cos : [0, \pi] \longrightarrow [-1, \ 1]$ is an one-one and onto function. We use this idea, to relate inner product with the angle between two vectors in an inner product space V.

Definition 5.2.9 (Angle between two vectors). *Let V be a real vector space and let $\mathbf{u}, \mathbf{v} \in V$. Suppose θ is the angle between $\mathbf{u}, \mathbf{v}$. We define*

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}.$$

1. *The real number θ, with $0 \leq \theta \leq \pi$, and satisfying $\cos\theta = \dfrac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$ is called the angle between the two vectors $\mathbf{u}$ and $\mathbf{v}$ in V.*

Definition 5.2.10 (Orthogonal Vectors). *Let V be a vector space.*

1. *The vectors $\mathbf{u}$ and $\mathbf{v}$ in V are said to be orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Orthogonality corresponds to perpendicularity.*
2. *A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ in V is called mutually orthogonal if $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for all $1 \leq i \neq j \leq n$.*

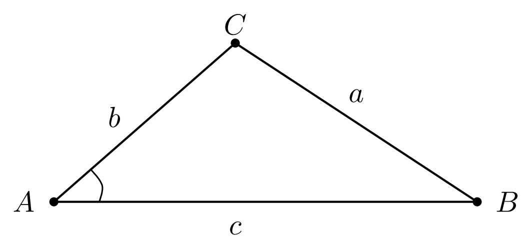

Figure 2: Triangle with vertices A, B and C

Before proceeding further with one more definition, recall that if ABC are vertices of a triangle (see Figure 5.2) then $\cos(A) = \frac{b^2+c^2-a^2}{2bc}$. We prove this as our next result.

Lemma 5.2.11. *Let A, B and C be the sides of a triangle in a real inner product space V then*

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}.$$

Proof. Let the coordinates of the vertices A, B and C be $\mathbf{0}$, $\mathbf{u}$ and $\mathbf{v}$, respectively. Then $\vec{AB} = \mathbf{u}, \vec{AC} = \mathbf{v}$ and $\vec{BC} = \mathbf{v} - \mathbf{u}$. Thus, we need to prove that

$$\cos(A) = \frac{\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2}{2\|\mathbf{v}\|\|\mathbf{u}\|}.$$

Now, using the properties of an inner product and Definition 5.2.9, it follows that

$$\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2 = 2 \, \langle \mathbf{u}, \mathbf{v} \rangle = 2 \, \|\mathbf{v}\|\|\mathbf{u}\| \, \cos(A).$$

Thus, the required result follows. ◻

Definition 5.2.12 (Orthogonal Complement)**.** *Let W be a subset of a vector space V with inner product $\langle\ ,\ \rangle$. Then the orthogonal complement of W in V, denoted $W^{\perp}$, is defined by*

$$W^{\perp} = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{w} \rangle = 0, \quad \textit{for all} \quad \mathbf{w} \in W\}.$$

Exercise 5.2.13. *Let W be a subset of a vector space V. Then prove that $W^{\perp}$ is a subspace of V.*

Example 5.2.14.

1. *Let $\mathbb{R}^4$ be endowed with the standard inner product. Fix two vectors $\mathbf{u}^t = (1,1,1,1), \mathbf{v}^t = (1,1,-1,0) \in \mathbb{R}^4$. Determine two vectors $\mathbf{z}$ and $\mathbf{w}$ such that $\mathbf{u} = \mathbf{z} + \mathbf{w}$, $\mathbf{z}$ is parallel to $\mathbf{v}$ and $\mathbf{w}$ is orthogonal to $\mathbf{v}$.*
   Solution: *Let $\mathbf{z}^t = k\mathbf{v}^t = (k,k,-k,0)$, for some $k \in \mathbb{R}$ and let $\mathbf{w}^t = (a,b,c,d)$. As $\mathbf{w}$ is orthogonal to $\mathbf{v}$, $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ and hence $a+b-c=0$. Thus, $c = a+b$ and*

   $$(1,1,1,1) = \mathbf{u}^t = \mathbf{z}^t + \mathbf{w}^t = (k,k,-k,0) + (a,b,a+b,d).$$

   Comparing the corresponding coordinates, we get

   $$d = 1, \ a+k = 1, b+k = 1 \ \textit{ and } \ a+b-k = 1.$$

   Solving for a, b and k gives $a = b = \frac{2}{3}$ and $k = \frac{1}{3}$. Thus, $\mathbf{z}^t = \frac{1}{3}(1,1,-1,0)$ and $\mathbf{w}^t = \frac{1}{3}(2,2,4,3)$.

2. *Let $\mathbb{R}^3$ be endowed with the standard inner product and let $P = (1,1,1)$, $Q = (2,1,3)$ and $R = (-1,1,2)$ be three vertices of a triangle in $\mathbb{R}^3$. Compute the angle between the sides PQ and PR.*
   Solution: Method 1: *The sides are represented by the vectors*

   $$\vec{PQ} = (2,1,3) - (1,1,1) = (1,0,2), \vec{PR} = (-2,0,1) \ \textit{and} \ \vec{RQ} = (-3,0,-1).$$

   As $\langle \vec{PQ}, \vec{PR} \rangle = 0$, the angle between the sides PQ and PR is $\dfrac{\pi}{2}$.

   Method 2: $\|PQ\| = \sqrt{5}, \|PR\| = \sqrt{5}$ *and* $\|QR\| = \sqrt{10}$. *As*

   $$\|QR\|^2 = \|PQ\|^2 + \|PR\|^2,$$

   by Pythagoras theorem, the angle between the sides PQ and PR is $\dfrac{\pi}{2}$.

We end this section by stating and proving the fundamental theorem of linear algebra. To do this, recall that for a matrix $A \in M_n(\mathbb{C})$, A^* denotes the conjugate transpose of A, $\mathcal{N}(A) = \{\mathbf{v} \in \mathbb{C}^n : A\mathbf{v} = \mathbf{0}\}$ denotes the null space of A and $\mathcal{R}(A) = \{A\mathbf{v} : \mathbf{v} \in \mathbb{C}^n\}$ denotes the range space of A. The readers are also advised to go through Theorem 3.3.25 (the rank-nullity theorem for matrices) before proceeding further as the first part is stated and proved there.

Theorem 5.2.15 (Fundamental Theorem of Linear Algebra)**.** *Let A be an $n \times n$ matrix with complex entries and let $\mathcal{N}(A)$ and $\mathcal{R}(A)$ be defined as above. Then*

1. $\dim(\mathcal{N}(A)) + \dim(\mathcal{R}(A)) = n$.
2. $\mathcal{N}(A) = \big(\mathcal{R}(A^*)\big)^{\perp}$ *and* $\mathcal{N}(A^*) = \big(\mathcal{R}(A)\big)^{\perp}$.
3. $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^*))$.

Proof. PART 1: Proved in Theorem 3.3.25.

PART 2: We first prove that $\mathcal{N}(A) \subset \mathcal{R}(A^*)^{\perp}$. Let $\mathbf{x} \in \mathcal{N}(A)$. Then $A\mathbf{x} = \mathbf{0}$ and

$$0 = \langle A\mathbf{x}, \mathbf{u}\rangle = \mathbf{u}^* A\mathbf{x} = (A^*\mathbf{u})^*\mathbf{x} = \langle \mathbf{x}, A^*\mathbf{u}\rangle$$

for all $\mathbf{u} \in \mathbb{C}^n$. Thus, $\mathbf{x} \in \mathcal{R}(A^*)^{\perp}$ and hence $\mathcal{N}(A) \subset \mathcal{R}(A^*)^{\perp}$.

We now prove that $\mathcal{R}(A^*)^{\perp} \subset \mathcal{N}(A)$. Let $\mathbf{x} \in \mathcal{R}(A^*)^{\perp}$. Then for every $\mathbf{y} \in \mathbb{C}^n$,

$$0 = \langle \mathbf{x}, A^*\mathbf{y}\rangle = (A^*\mathbf{y})^*\mathbf{x} = \mathbf{y}^*(A^*)^*\mathbf{x} = \mathbf{y}^* A\mathbf{x} = \langle A\mathbf{x}, \mathbf{y}\rangle.$$

In particular, for $\mathbf{y} = A\mathbf{x}$, we get $\|A\mathbf{x}\|^2 = 0$ and hence $A\mathbf{x} = \mathbf{0}$. That is, $\mathbf{x} \in \mathcal{N}(A)$. Thus, the proof of the first equality in Part 2 is over. We omit the second equality as it proceeds on the same lines as above.

PART 3: Use the first two parts to get the result.

Hence the proof of the fundamental theorem is complete. ☐

For more information related with the fundamental theorem of linear algebra the interested readers are advised to see the article "The Fundamental Theorem of Linear Algebra, Gilbert Strang, *The American Mathematical Monthly*, Vol. 100, No. 9, Nov., 1993, pp. 848 - 855."

Exercise 5.2.16. *1. Answer the following questions when $\mathbb{R}^3$ is endowed with the standard inner product.*

(a) Let $\mathbf{u}^t = (1, 1, 1)$. Find vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ that are orthogonal to $\mathbf{u}$ and to each other.

(b) Find the equation of the line that passes through the point $(1, 1, -1)$ and is parallel to the vector $(a, b, c) \neq (0, 0, 0)$.

(c) Find the equation of the plane that contains the point $(1, 1-1)$ and the vector $(a, b, c) \neq (0, 0, 0)$ is a normal vector to the plane.

(d) Find area of the parallelogram with vertices $(0, 0, 0)$, $(1, 2, -2)$, $(2, 3, 0)$ and $(3, 5, -2)$.

(e) Find the equation of the plane that contains the point $(2, -2, 1)$ and is perpendicular to the line with parametric equations $x = t - 1, y = 3t + 2, z = t + 1$.

(f) Let $P = (3, 0, 2), Q = (1, 2, -1)$ and $R = (2, -1, 1)$ be three points in $\mathbb{R}^3$.

i. Find the area of the triangle with vertices P, Q and R.

ii. Find the area of the parallelogram built on vectors $\vec{PQ}$ and $\vec{QR}$.

iii. Find a nonzero vector orthogonal to the triangle with vertices P, Q and R.

iv. Find all vectors $\mathbf{x}$ orthogonal to $\vec{PQ}$ and $\vec{QR}$ with $\|\mathbf{x}\| = \sqrt{2}$.

v. *Choose one of the vectors* $\mathbf{x}$ *found in part 1(f)iv. Find the volume of the parallelepiped built on vectors* $\vec{PQ}$ *and* $\vec{QR}$ *and* $\mathbf{x}$. *Do you think the volume would be different if you choose the other vector* $\mathbf{x}$*?*

(g) *Find the equation of the plane that contains the lines* $(x, y, z) = (1, 2, -2) + t(1, 1, 0)$ *and* $(x, y, z) = (1, 2, -2) + t(0, 1, 2)$.

(h) *Let* $\mathbf{u}^t = (1, -1, 1)$ *and* $\mathbf{v}^t = (1, k, 1)$. *Find* k *such that the angle between* $\mathbf{u}$ *and* $\mathbf{v}$ *is* $\pi/3$.

(i) *Let* p_1 *be a plane that passes through the point* $A = (1, 2, 3)$ *and has* $\check{\mathbf{n}} = (2, -1, 1)$ *as its normal vector. Then*

i. *find the equation of the plane* p_2 *which is parallel to* p_1 *and passes through the point* $(-1, 2, -3)$.

ii. *calculate the distance between the planes* p_1 *and* p_2.

(j) *In the parallelogram* $ABCD$, $AB \| DC$ *and* $AD \| BC$ *and* $A = (-2, 1, 3), B = (-1, 2, 2), C = (-3, 1, 5)$. *Find*

i. *the coordinates of the point* D,

ii. *the cosine of the angle* BCD.

iii. *the area of the triangle* ABC

iv. *the volume of the parallelepiped determined by the vectors* AB, AD *and the vector* $(0, 0, -7)$.

(k) *Find the equation of a plane that contains the point* $(1, 1, 2)$ *and is orthogonal to the line with parametric equation* $x = 2 + t, y = 3$ *and* $z = 1 - t$.

(l) *Find a parametric equation of a line that passes through the point* $(1, -2, 1)$ *and is orthogonal to the plane* $x + 3y + 2z = 1$.

2. *Let* $\{\mathbf{e}_1^t, \mathbf{e}_2^t, \ldots, \mathbf{e}_n^t\}$ *be the standard basis of* $\mathbb{R}^n$. *Then prove that with respect to the standard inner product on* $\mathbb{R}^n$, *the vectors* $\mathbf{e}_i$ *satisfy the following:*

(a) $\|\mathbf{e}_i\| = 1$ *for* $1 \le i \le n$.

(b) $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ *for* $1 \le i \ne j \le n$.

3. *Let* $\mathbf{x}^t = (x_1, x_2)$, $\mathbf{y}^t = (y_1, y_2) \in \mathbb{R}^2$. *Then* $\langle \mathbf{x}, \mathbf{y} \rangle = 4x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$ *defines an inner product. Use this inner product to find*

(a) *the angle between* $e_1^t = (1, 0)$ *and* $e_2^t = (0, 1)$.

(b) $\mathbf{v} \in \mathbb{R}^2$ *such that* $\langle \mathbf{v}, (1, 0)^t \rangle = 0$.

(c) *vectors* $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^2$ *such that* $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ *and* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

4. *Does there exist an inner product in* $\mathbb{R}^2$ *such that*

$$\|(1, 2)\| = \|(2, -1)\| = 1 \quad \textit{and} \quad \langle (1, 2),\ (2, -1) \rangle = 0?$$

[Hint: Consider a symmetric matrix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. *Define* $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^t A \mathbf{x}$. *Use the given conditions to get a linear system of* 3 *equations in the unknowns* a, b, c. *Solve this system.]*

5. *Let* $W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$. *Find a basis of* $W^\perp$.

6. *Let* W *be a subspace of a finite dimensional inner product space* V. *Prove that* $(W^\perp)^\perp = W$.

7. *Let* $\mathbf{x}^t = (x_1, x_2, x_3),\ \mathbf{y}^t = (y_1, y_2, y_3) \in \mathbb{R}^3$. *Show that*

$$\langle \mathbf{x}, \mathbf{y} \rangle = 10x_1y_1 + 3x_1y_2 + 3x_2y_1 + 4x_2y_2 + x_2y_3 + x_3y_2 + 3x_3y_3$$

is an inner product in $\mathbb{R}^3(\mathbb{R})$. *With respect to this inner product, find the angle between the vectors* $(1, 1, 1)$ *and* $(2, -5, 2)$.

8. *Recall the inner product space* $M_{n\times n}(\mathbb{R})$ *(see Example 5.2.3.6). Determine* $W^\perp$ *for the subspace* $W = \{A \in M_{n\times n}(\mathbb{R}) :\ A^t = A\}$.

9. *Prove that* $\langle f(x), g(x) \rangle = \int\limits_{-\pi}^{\pi} f(x) \cdot g(x)dx$ *defines an inner product in* $C[-\pi,\ \pi]$. *Define* $\mathbf{1}(x) = 1$ *for all* $x \in [-\pi,\ \pi]$. *Prove that*

$$S = \{\mathbf{1}\} \cup \{\cos(mx) : m \geq 1\} \cup \{\sin(nx) : n \geq 1\}$$

is a linearly independent subset of $C[-\pi,\ \pi]$.

10. *Let* V *be an inner product space. Prove the* TRIANGLE INEQUALITY

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \quad \text{for every} \quad \mathbf{u}, \mathbf{v} \in V.$$

11. *Let* $z_1, z_2, \ldots, z_n \in \mathbb{C}$. *Use the Cauchy-Schwarz inequality to prove that*

$$|z_1 + z_2 + \cdots + z_n| \leq \sqrt{n(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2)}.$$

When does the equality hold?

12. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. *Prove the following:*

    (a) $\langle \mathbf{x}, \mathbf{y} \rangle = 0 \Longleftrightarrow \|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ *(*PYTHAGORAS THEOREM*).*

    (b) $\|\mathbf{x}\| = \|\mathbf{y}\| \Longleftrightarrow \langle \mathbf{x} + \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = 0$ *($\mathbf{x}$ and $\mathbf{y}$ form adjacent sides of a rhombus as the diagonals $\mathbf{x} + \mathbf{y}$ and $\mathbf{x} - \mathbf{y}$ are orthogonal).*

    (c) $4\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2$ *(*POLARIZATION IDENTITY*).*
    Are the above results true if $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n(\mathbb{C})$?

13. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n(\mathbb{C})$. *Prove that*

    (a) $4\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 + i\|\mathbf{x} + i\mathbf{y}\|^2 - i\|\mathbf{x} - i\mathbf{y}\|^2$.

(b) *If* $\mathbf{x} \neq \mathbf{0}$ *then* $\|\mathbf{x} + i\mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{x}\|^2$, *even though* $\langle \mathbf{x}, i\mathbf{x} \rangle \neq 0$.

(c) $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ *whenever* $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ *and* $\|\mathbf{x} + i\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{y}\|^2$.

14. *Let* $\langle \ , \ \rangle$ *denote the standard inner product on* $\mathbb{C}^n(\mathbb{C})$ *and let* $A \in M_n(\mathbb{C})$. *That is,* $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y}$ *for all* $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{C}^n$. *Prove that* $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^*\mathbf{y} \rangle$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

15. *Let* $(V, \langle \ , \ \rangle)$ *be an* n*-dimensional inner product space and let* $\mathbf{u} \in V$ *be a fixed vector with* $\|\mathbf{u}\| = 1$. *Then give reasons for the following statements.*

    (a) *Let* $S^\perp = \{\mathbf{v} \in V \ : \ \langle \mathbf{v}, \mathbf{u} \rangle = 0\}$. *Then* $\dim(S^\perp) = n - 1$.

    (b) *Let* $0 \neq \alpha \in \mathbb{F}$. *Then* $S = \{\mathbf{v} \in V \ : \ \langle \mathbf{v}, \mathbf{u} \rangle = \alpha\}$ *is not a subspace of* V.

    (c) *Let* $\mathbf{v} \in V$. *Then* $\mathbf{v} = \mathbf{v}_0 + \alpha\mathbf{u}$ *for a vector* $\mathbf{v}_0 \in S^\perp$ *and a scalar* α. *That is,* $V = L(\mathbf{u}, S^\perp)$.

5.2.1 Basic Results on Orthogonal Vectors

We start this subsection with the definition of an orthonormal set. Then a theorem is proved that implies that the coordinates of a vector with respect to an orthonormal basis are just the inner products with the basis vectors.

Definition 5.2.17 (Orthonormal Set)**.** *Let* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *be a set of non-zero, mutually orthogonal vectors in an inner product space* V. *Then* S *is called an orthonormal set if* $\|\mathbf{v}_i\| = 1$ *for* $1 \leq i \leq n$. *If* S *is also a basis of* V *then* S *is called an orthonormal basis of* V.

Example 5.2.18. 1. *Consider* $\mathbb{R}^2$ *with the standard inner product. Then a few orthonormal sets in* $\mathbb{R}^2$ *are* $\{(1,0),(0,1)\}$, $\left\{\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\right\}$ *and* $\left\{\frac{1}{\sqrt{5}}(2,1), \frac{1}{\sqrt{5}}(1,-2)\right\}$.

2. *Let* $\mathbb{R}^n$ *be endowed with the standard inner product. Then by Exercise 5.2.16.2, the standard ordered basis* $(\mathbf{e}_1^t, \mathbf{e}_2^t, \ldots, \mathbf{e}_n^t)$ *is an orthonormal set.*

Theorem 5.2.19. *Let* V *be an inner product space and let* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *be a set of non-zero, mutually orthogonal vectors of* V.

1. *Then the set* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *is linearly independent.*

2. *Let* $\mathbf{v} = \sum\limits_{i=1}^n \alpha_i \mathbf{u}_i \in V$. *Then* $\|\mathbf{v}\|^2 = \|\sum\limits_{i=1}^n \alpha_i \mathbf{u}_i\|^2 = \sum\limits_{i=1}^n |\alpha_i|^2 \|\mathbf{u}_i\|^2$;

3. *Let* $\mathbf{v} = \sum\limits_{i=1}^n \alpha_i \mathbf{u}_i$. *If* $\|\mathbf{u}_i\| = 1$ *for* $1 \leq i \leq n$ *then* $\alpha_i = \langle \mathbf{v}, \mathbf{u}_i \rangle$ *for* $1 \leq i \leq n$. *That is,*

$$\mathbf{v} = \sum_{i=1}^n \langle \mathbf{v}, \mathbf{u}_i \rangle \mathbf{u}_i \quad \textit{and} \quad \|\mathbf{v}\|^2 = \sum_{i=1}^n |\langle \mathbf{v}, \mathbf{u}_i \rangle|^2.$$

4. *Let* $\dim(V) = n$. *Then* $\langle \mathbf{v}, \mathbf{u}_i \rangle = 0$ *for all* $i = 1, 2, \ldots, n$ *if and only if* $\mathbf{v} = \mathbf{0}$.

Proof. Consider the linear system

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{0} \tag{5.2.2}$$

in the unknowns $c_1, c_2, \ldots, c_n$. As $\langle \mathbf{0}, \mathbf{u}\rangle = 0$ for each $\mathbf{u} \in V$ and $\langle \mathbf{u}_j, \mathbf{u}_i\rangle = 0$ for all $j \neq i$, we have

$$0 = \langle \mathbf{0}, \mathbf{u}_i\rangle = \langle c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_i\rangle = \sum_{j=1}^{n} c_j\langle \mathbf{u}_j, \mathbf{u}_i\rangle = c_i\langle \mathbf{u}_i, \mathbf{u}_i\rangle.$$

As $\mathbf{u}_i \neq \mathbf{0}$, $\langle \mathbf{u}_i, \mathbf{u}_i\rangle \neq 0$ and therefore $c_i = 0$ for $1 \leq i \leq n$. Thus, the linear system (5.2.2) has only the trivial solution. Hence, the proof of Part 1 is complete.

For Part 2, we use a similar argument to get

$$\begin{aligned} \|\sum_{i=1}^{n} \alpha_i\mathbf{u}_i\|^2 &= \left\langle \sum_{i=1}^{n} \alpha_i\mathbf{u}_i, \sum_{i=1}^{n} \alpha_i\mathbf{u}_i \right\rangle = \sum_{i=1}^{n} \alpha_i \left\langle \mathbf{u}_i, \sum_{j=1}^{n} \alpha_j\mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \overline{\alpha_j}\,\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \sum_{i=1}^{n} \alpha_i\overline{\alpha_i}\,\langle \mathbf{u}_i, \mathbf{u}_i\rangle = \sum_{i=1}^{n} |\alpha_i|^2\|\mathbf{u}_i\|^2. \end{aligned}$$

Note that $\langle \mathbf{v}, \mathbf{u}_i\rangle = \left\langle \sum_{j=1}^{n} \alpha_j\mathbf{u}_j, \mathbf{u}_i \right\rangle = \sum_{j=1}^{n} \alpha_j\langle \mathbf{u}_j, \mathbf{u}_i\rangle = \alpha_j$. Thus, the proof of Part 3 is complete.

Part 4 directly follows using Part 3 as the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis of V. Therefore, we have obtained the required result. ☐

In view of Theorem 5.2.19, we inquire into the question of extracting an orthonormal basis from a given basis. In the next section, we describe a process (called the Gram-Schmidt Orthogonalization process) that generates an orthonormal set from a given set containing finitely many vectors.

Remark 5.2.20. *The last two parts of Theorem 5.2.19 can be rephrased as follows: Let $\mathcal{B} = \big(\mathbf{v}_1, \ldots, \mathbf{v}_n\big)$ be an ordered orthonormal basis of an inner product space V and let $\mathbf{u} \in V$. Then*

$$[\mathbf{u}]_{\mathcal{B}} = \left(\langle \mathbf{u}, \mathbf{v}_1\rangle, \langle \mathbf{u}, \mathbf{v}_2\rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n\rangle\right)^t.$$

Exercise 5.2.21.

1. *Let $\mathcal{B} = \left(\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\right)$ be an ordered basis of $\mathbb{R}^2$. Determine $[(2,3)]_{\mathcal{B}}$. Also, compute $[(x,y)]_{\mathcal{B}}$.*

2. *Let $\mathcal{B} = \left(\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2),\right)$ be an ordered basis of $\mathbb{R}^3$. Determine $[(2,3,1)]_{\mathcal{B}}$. Also, compute $[(x,y,z)]_{\mathcal{B}}$.*

3. *Let $\mathbf{u}^t = (u_1,u_2,u_3)$, $\mathbf{v}^t = (v_1,v_2,v_3)$ be two vectors in $\mathbb{R}^3$. Then recall that their cross product, denoted $\mathbf{u} \times \mathbf{v}$, equals*

$$\mathbf{u}^t \times \mathbf{v}^t = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1).$$

Use this to find an orthonormal basis of $\mathbb{R}^3$ containing the vector $\frac{1}{\sqrt{6}}(1,2,1)$.

4. *Let* $\mathbf{u}^t = (1, -1, -2)$. *Find vectors* $\mathbf{v}^t, \mathbf{w}^t \in \mathbb{R}^3$ *such that* $\mathbf{v}$ *and* $\mathbf{w}$ *are orthogonal to* $\mathbf{u}$ *and to each other as well.*

5. *Let* A *be an* $n \times n$ *orthogonal matrix. Prove that the rows/columns of* A *form an orthonormal basis of* $\mathbb{R}^n$.

6. *Let* A *be an* $n \times n$ *unitary matrix. Prove that the rows/columns of* A *form an orthonormal basis of* $\mathbb{C}^n$.

7. *Let* $\{\mathbf{u}_1^t, \mathbf{u}_2^t, \ldots, \mathbf{u}_n^t\}$ *be an orthonormal basis of* $\mathbb{R}^n$. *Prove that the* $n \times n$ *matrix* $A = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ *is an orthogonal matrix.*

5.3 Gram-Schmidt Orthogonalization Process

Suppose we are given two non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ in a plane. Then in many instances, we need to decompose the vector $\mathbf{v}$ into two components, say $\mathbf{y}$ and $\mathbf{z}$, such that $\mathbf{y}$ is a vector parallel to $\mathbf{u}$ and $\mathbf{z}$ is a vector perpendicular (orthogonal) to $\mathbf{u}$. We do this as follows (see Figure 5.3):
Let $\hat{\mathbf{u}} = \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$. Then $\hat{\mathbf{u}}$ is a unit vector in the direction of $\mathbf{u}$. Also, using trigonometry, we know that $\cos(\theta) = \frac{\|\vec{OQ}\|}{\|\vec{OP}\|}$ and hence $\|\vec{OQ}\| = \|\vec{OP}\| \, \cos(\theta)$. Or using Definition 5.2.9,

$$\|\vec{OQ}\| = \|\mathbf{v}\| \, \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{v}\| \, \|\mathbf{u}\|} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|},$$

where we need to take the absolute value of the right hand side expression as the length of a vector is always a positive quantity. Thus, we get

$$\vec{OQ} = \|\vec{OQ}\| \, \hat{\mathbf{u}} = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle \, \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Thus, we see that $\mathbf{y} = \vec{OQ} = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle \, \frac{\mathbf{u}}{\|\mathbf{u}\|}$ and $\mathbf{z} = \mathbf{v} - \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle \, \frac{\mathbf{u}}{\|\mathbf{u}\|}$. It is easy to verify that $\mathbf{v} = \mathbf{y} + \mathbf{z}$, $\mathbf{y}$ is parallel to $\mathbf{u}$ and $\mathbf{z}$ is orthogonal to $\mathbf{u}$. In literature, the vector $\mathbf{y} = \vec{OQ}$ is often called the orthogonal projection of the vector $\mathbf{v}$ on $\mathbf{u}$ and is denoted by $\text{Proj}_{\mathbf{u}}(\mathbf{v})$. Thus,

$$\text{Proj}_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle \, \frac{\mathbf{u}}{\|\mathbf{u}\|} \quad \text{and} \quad \|\text{Proj}_{\mathbf{u}}(\mathbf{v})\| = \|\vec{OQ}\| = \left| \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|} \right|. \tag{5.3.1}$$

Moreover, the distance of the vector $\mathbf{u}$ from the point P equals $\|\vec{OR}\| = \|\vec{PQ}\| = \|\mathbf{v} - \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \rangle \, \frac{\mathbf{u}}{\|\mathbf{u}\|}\|$.

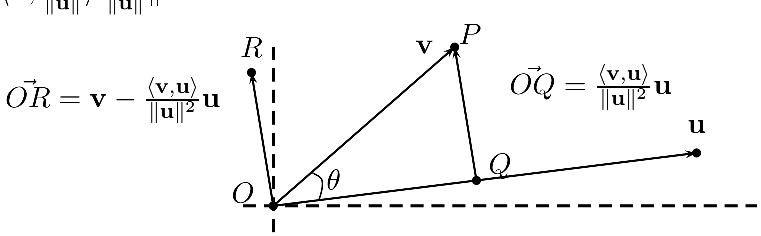

Figure 3: Decomposition of vector $\mathbf{v}$

Also, note that $\hat{\mathbf{u}}$ is a unit vector in the direction of $\mathbf{u}$ and $\hat{\mathbf{z}} = \frac{\mathbf{z}}{\|\mathbf{z}\|}$ is a unit vector orthogonal to $\hat{\mathbf{u}}$. This idea is generalized to study the Gram-Schmidt Orthogonalization process which is given as the next result. Before stating this result, we look at the following example to understand the process.

Example 5.3.1. *1. In Example 5.2.14.1, we note that* $Proj_{\mathbf{v}}(\mathbf{u}) = (\mathbf{u}\cdot\mathbf{v})\dfrac{\mathbf{v}}{\|\mathbf{v}\|^2}$ *is parallel to* $\mathbf{v}$ *and* $\mathbf{u} - Proj_{\mathbf{v}}(\mathbf{u})$ *is orthogonal to* $\mathbf{v}$*. Thus,*

$$\vec{\mathbf{z}} = Proj_{\mathbf{v}}(\mathbf{u}) = \frac{1}{3}(1,1,-1,0)^t \quad \textit{and} \quad \vec{\mathbf{w}} = (1,1,1,1)^t - \vec{\mathbf{z}} = \frac{1}{3}(2,2,4,3)^t.$$

2. Let $\mathbf{u}^t = (1,1,1,1), \mathbf{v}^t = (1,1,-1,0)$ *and* $\mathbf{w}^t = (1,1,0,-1)$ *be three vectors in* $\mathbb{R}^4$*. Write* $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ *where* $\mathbf{v}_1$ *is parallel to* $\mathbf{u}$ *and* $\mathbf{v}_2$ *is orthogonal to* $\mathbf{u}$*. Also, write* $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3$ *such that* $\mathbf{w}_1$ *is parallel to* $\mathbf{u}$, $\mathbf{w}_2$ *is parallel to* $\mathbf{v}_2$ *and* $\mathbf{w}_3$ *is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}_2$*.*

Solution : *Note that*

(a) $\mathbf{v}_1 = Proj_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^t$ *is parallel to* $\mathbf{u}$ *and*

(b) $\mathbf{v}_2 = \mathbf{v} - \frac{1}{4}\mathbf{u} = \frac{1}{4}(3,3,-5,-1)^t$ *is orthogonal to* $\mathbf{u}$*.*

Note that $Proj_{\mathbf{u}}(\mathbf{w})$ *is parallel to* $\mathbf{u}$ *and* $Proj_{\mathbf{v}_2}(\mathbf{w})$ *is parallel to* $\mathbf{v}_2$*. Hence, we have*

(a) $\mathbf{w}_1 = Proj_{\mathbf{u}}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^t$ *is parallel to* $\mathbf{u}$*,*

(b) $\mathbf{w}_2 = Proj_{\mathbf{v}_2}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{v}_2\rangle \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|^2} = \frac{7}{44}(3,3,-5,-1)^t$ *is parallel to* $\mathbf{v}_2$ *and*

(c) $\mathbf{w}_3 = \mathbf{w} - \mathbf{w}_1 - \mathbf{w}_2 = \frac{3}{11}(1,1,2,-4)^t$ *is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}_2$*.*

That is, from the given vector subtract all the orthogonal components that are obtained as orthogonal projections. If this new vector is non-zero then this vector is orthogonal to the previous ones.

Theorem 5.3.2 (Gram-Schmidt Orthogonalization Process)**.** *Let V be an inner product space. Suppose* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *is a set of linearly independent vectors in V. Then there exists a set* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *of vectors in V satisfying the following:*

1. $\|\mathbf{v}_i\| = 1$ *for* $1 \le i \le n$*,*
2. $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ *for* $1 \le i \ne j \le n$ *and*
3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ *for* $1 \le i \le n$*.*

Proof. We successively define the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ as follows.

Step 1: $\mathbf{v}_1 = \dfrac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$.

Step 2: Calculate $\mathbf{w}_2 = \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{v}_1\rangle \mathbf{v}_1$, and let $\mathbf{v}_2 = \dfrac{\mathbf{w}_2}{\|\mathbf{w}_2\|}$.

Step 3: Obtain $\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1\rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2\rangle \mathbf{v}_2$, and let $\mathbf{v}_3 = \dfrac{\mathbf{w}_3}{\|\mathbf{w}_3\|}$.

Step i: In general, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}$ are already obtained, we compute

$$\mathbf{w}_i = \mathbf{u}_i - \langle \mathbf{u}_i, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_i, \mathbf{v}_2 \rangle \mathbf{v}_2 - \cdots - \langle \mathbf{u}_i, \mathbf{v}_{i-1} \rangle \mathbf{v}_{i-1}. \tag{5.3.2}$$

As the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linearly independent, it can be verified that $\|\mathbf{w}_i\| \neq 0$ and hence we define $\mathbf{v}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$.

We prove this by induction on n, the number of linearly independent vectors. For $n = 1$, $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$. As $\mathbf{u}$ is an element of a linearly independent set, $\mathbf{u}_1 \neq \mathbf{0}$ and thus $\mathbf{v}_1 \neq \mathbf{0}$ and

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \langle \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \rangle = \frac{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} = 1.$$

Hence, the result holds for $n = 1$.

Let the result hold for all $k \leq n-1$. That is, suppose we are given any set of $k,\ 1 \leq k \leq n-1$ linearly independent vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of V. Then by the inductive assumption, there exists a set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of vectors satisfying the following:

1. $\|\mathbf{v}_i\| = 1$ for $1 \leq i \leq k$,
2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i \neq j \leq k$, and
3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ for $1 \leq i \leq k$.

Now, let us assume that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a linearly independent subset of V. Then by the inductive assumption, we already have vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ satisfying

1. $\|\mathbf{v}_i\| = 1$ for $1 \leq i \leq n-1$,
2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $1 \leq i \neq j \leq n-1$, and
3. $L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i)$ for $1 \leq i \leq n-1$.

Using (5.3.2), we define

$$\mathbf{w}_n = \mathbf{u}_n - \langle \mathbf{u}_n, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_n, \mathbf{v}_2 \rangle \mathbf{v}_2 - \cdots - \langle \mathbf{u}_n, \mathbf{v}_{n-1} \rangle \mathbf{v}_{n-1}. \tag{5.3.3}$$

We first show that $\mathbf{w}_n \notin L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. This will imply that $\mathbf{w}_n \neq \mathbf{0}$ and hence $\mathbf{v}_n = \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}$ is well defined. Also, $\|\mathbf{v}_n\| = 1$.

On the contrary, assume that $\mathbf{w}_n \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. Then, by definition, there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$, not all zero, such that

$$\mathbf{w}_n = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_{n-1} \mathbf{v}_{n-1}.$$

So, substituting $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_{n-1} \mathbf{v}_{n-1}$ for $\mathbf{w}_n$ in (5.3.3), we get

$$\mathbf{u}_n = \big(\alpha_1 + \langle \mathbf{u}_n, \mathbf{v}_1 \rangle\big) \mathbf{v}_1 + \big(\alpha_2 + \langle \mathbf{u}_n, \mathbf{v}_2 \rangle\big) \mathbf{v}_2 + \cdots + (\big(\alpha_{n-1} + \langle \mathbf{u}_n, \mathbf{v}_{n-1} \rangle\big) \mathbf{v}_{n-1}.$$

That is, $\mathbf{u}_n \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. But $L(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) = L(\mathbf{u}_1, \ldots, \mathbf{u}_{n-1})$ using the third induction assumption. Hence $\mathbf{u}_n \in L(\mathbf{u}_1, \ldots, \mathbf{u}_{n-1})$. A contradiction to the given assumption that the set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is linearly independent.

Also, it can be easily verified that $\langle \mathbf{v}_n, \mathbf{v}_i \rangle = 0$ for $1 \leq i \leq n-1$. Hence, by the principle of mathematical induction, the proof of the theorem is complete. $\square$

We illustrate the Gram-Schmidt process by the following example.

Example 5.3.3. 1. *Let* $\{(1,-1,1,1),(1,0,1,0),(0,1,0,1)\} \subset \mathbb{R}^4$. *Find a set* $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ *that is orthonormal and* $L(\,(1,-1,1,1),(1,0,1,0),(0,1,0,1)\,) = L(\mathbf{v}_1^t,\mathbf{v}_2^t,\mathbf{v}_3^t)$.

Solution: *Let* $\mathbf{u}_1^t = (1,0,1,0), \mathbf{u}_2^t = (0,1,0,1)$ *and* $\mathbf{u}_3^t = (1,-1,1,1)$. *Then* $\mathbf{v}_1^t = \frac{1}{\sqrt{2}}(1,0,1,0)$. *Also,* $\langle \mathbf{u}_2, \mathbf{v}_1\rangle = 0$ *and hence* $\mathbf{w}_2 = \mathbf{u}_2$. *Thus,* $\mathbf{v}_2^t = \frac{1}{\sqrt{2}}(0,1,0,1)$ *and*

$$\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3,\mathbf{v}_1\rangle \mathbf{v}_1 - \langle \mathbf{u}_3,\mathbf{v}_2\rangle \mathbf{v}_2 = (0,-1,0,1)^t.$$

Therefore, $\mathbf{v}_3^t = \frac{1}{\sqrt{2}}(0,-1,0,1)$.

2. *Find an orthonormal set in* $\mathbb{R}^3$ *containing* $(1,2,1)$.

Solution: *Let* $(x,y,z) \in \mathbb{R}^3$ *with* $\big\langle (1,2,1),(x,y,z)\big\rangle = 0$. *Then* $x+2y+z=0$ *or equivalently,* $x = -2y-z$. *Thus,*

$$(x,y,z) = (-2y-z,y,z) = y(-2,1,0) + z(-1,0,1).$$

Observe that the vectors $(-2,1,0)$ *and* $(-1,0,1)$ *are both orthogonal to* $(1,2,1)$ *but are not orthogonal to each other.*

Method 1: *Consider* $\{\frac{1}{\sqrt{6}}(1,2,1),(-2,1,0),(-1,0,1)\} \subset \mathbb{R}^3$ *and apply the Gram-Schmidt process to get the result.*

Method 2: *This method can be used only if the vectors are from* $\mathbb{R}^3$. *Recall that in* $\mathbb{R}^3$, *the cross product of two vectors* $\mathbf{u}$ *and* $\mathbf{v}$, *denoted* $\mathbf{u}\times\mathbf{v}$, *is a vector that is orthogonal to both the vectors* $\mathbf{u}$ *and* $\mathbf{v}$. *Hence, the vector*

$$(1,2,1)\times(-2,1,0) = (0-1,-2-0,1+4) = (-1,-2,5)$$

is orthogonal to the vectors $(1,2,1)$ *and* $(-2,1,0)$ *and hence the required orthonormal set is* $\{\frac{1}{\sqrt{6}}(1,2,1), \frac{-1}{\sqrt{5}}(2,-1,0), \frac{-1}{\sqrt{30}}(1,2,-5)\}$.

Remark 5.3.4. 1. *Let* V *be a vector space. Then the following holds.*

(a) *Let* $\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\}$ *be a linearly independent subset of* V. *Then Gram-Schmidt orthogonalization process gives an orthonormal set* $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ *of* V *with*

$$L(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_i) = L(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_i) \text{ for } 1\le i\le k.$$

(b) *Let* W *be a subspace of* V *with a basis* $\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\}$. *Then* $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ *is also a basis of* W.

(c) *Suppose* $\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$ *is a linearly dependent subset of* V. *Then there exists a smallest* k, $2\le k\le n$ *such that* $\mathbf{w}_k = \mathbf{0}$.

Idea of the proof: *Linear dependence (see Corollary 3.2.5) implies that there exists a smallest* k, $2\le k\le n$ *such that*

$$L(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k) = L(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{k-1}).$$

Also, by Gram-Schmidt orthogonalization process

$$L(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}) = L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}).$$

Thus, $\mathbf{u}_k \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1})$ *and hence by Remark 5.2.20*

$$\mathbf{u}_k = \langle \mathbf{u}_k, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}_k, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}_k, \mathbf{v}_{k-1} \rangle \mathbf{v}_{k-1}.$$

So, by definition $\mathbf{w}_k = \mathbf{0}$.

2. *Let* S *be a countably infinite set of linearly independent vectors. Then one can apply the Gram-Schmidt process to get a countably infinite orthonormal set.*

3. *Consider* $\mathbb{R}^n$ *with the standard inner product and let* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *be an orthonormal set. Then, we see that*

   (a) $\|\mathbf{v}_i\| = 1$ *is equivalent to* $\mathbf{v}_i^t \mathbf{v}_i = 1$, *for* $1 \le i \le n$,

   (b) $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ *is equivalent to* $\mathbf{v}_i^t \mathbf{v}_j = 0$, *for* $1 \le i \ne j \le n$.

   Hence, we see that

$$A^t A = \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_n^t \end{bmatrix} [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = \begin{bmatrix} \mathbf{v}_1^t \mathbf{v}_1 & \mathbf{v}_1^t \mathbf{v}_2 & \cdots & \mathbf{v}_1^t \mathbf{v}_n \\ \mathbf{v}_2^t \mathbf{v}_1 & \mathbf{v}_2^t \mathbf{v}_2 & \cdots & \mathbf{v}_2^t \mathbf{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_n^t \mathbf{v}_1 & \mathbf{v}_n^t \mathbf{v}_2 \rangle & \cdots & \mathbf{v}_n^t \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n.$$

   Since $A^t A = I_n$, *it follows that* $AA^t = I_n$. *But,*

$$AA^t = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_n^t \end{bmatrix} = \mathbf{v}_1 \mathbf{v}_1^t + \mathbf{v}_2 \mathbf{v}_2^t + \cdots + \mathbf{v}_n \mathbf{v}_n^t.$$

   Now, for each $i, 1 \le i \le n$, *the matrix* $\mathbf{v}_i \mathbf{v}_i^t$ *has the following properties:*

   (a) *it is symmetric;*

   (b) *it is idempotent; and*

   (c) *it has rank one.*

4. *The first two properties imply that the matrix* $\mathbf{v}_i \mathbf{v}_i^t$, *for each* $i, 1 \le i \le n$ *is a projection operator. That is, the identity matrix is the sum of projection operators, each of rank* 1.

5. *Now define a linear transformation* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *by* $T(\mathbf{x}) = (\mathbf{v}_i \mathbf{v}_i^t)\mathbf{x} = (\mathbf{v}_i^t \mathbf{x})\mathbf{v}_i$ *is a projection operator on the subspace* $L(\mathbf{v}_i)$.

6. *Now, let us fix* $k, 1 \leq k \leq n$. *Then, it can be observed that the linear transformation* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *defined by* $T(\mathbf{x}) = (\sum_{i=1}^{k} \mathbf{v}_i \mathbf{v}_i^t)\mathbf{x} = \sum_{i=1}^{k} (\mathbf{v}_i^t \mathbf{x})\mathbf{v}_i$ *is a projection operator on the subspace* $L(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. *We will use this idea in Subsection 5.4.1 .*

Definition 5.2.12 started with a subspace of an inner product space V and looked at its complement. We now look at the orthogonal complement of a subset of an inner product space V and the results associated with it.

Definition 5.3.5 (Orthogonal Subspace of a Set)**.** *Let* V *be an inner product space. Let* S *be a non-empty subset of* V. *We define*

$$S^{\perp} = \{\mathbf{v} \in V \ : \langle \mathbf{v}, \mathbf{s} \rangle = 0 \text{ for all } \mathbf{s} \in S\}.$$

Example 5.3.6. *Let* $V = \mathbb{R}$.

1. $S = \{0\}$. *Then* $S^{\perp} = \mathbb{R}$.

2. $S = \mathbb{R}$, *Then* $S^{\perp} = \{0\}$.

3. *Let* S *be any subset of* $\mathbb{R}$ *containing a non-zero real number. Then* $S^{\perp} = \{0\}$.

4. *Let* $S = \{(1, 2, 1)\} \subset \mathbb{R}^3$. *Then using Example 5.3.3.2,* $S^{\perp} = L(\{(-2, 1, 0), (-1, 0, 1)\})$.

We now state the result which gives the existence of an orthogonal subspace of a finite dimensional inner product space.

Theorem 5.3.7. *Let* S *be a subset of a finite dimensional inner product space* V, *with inner product* $\langle \ , \ \rangle$. *Then*

1. $S^{\perp}$ *is a subspace of* V.

2. *Let* $W = L(S)$. *Then the subspaces* W *and* $S^{\perp} = W^{\perp}$ *are complementary. That is,* $V = W + S^{\perp} = W + W^{\perp}$.

3. *Moreover,* $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ *for all* $\mathbf{w} \in W$ *and* $\mathbf{u} \in S^{\perp}$.

Proof. We leave the prove of the first part to the reader. The prove of the second part is as follows:
Let $\dim(V) = n$ and $\dim(W) = k$. Let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be a basis of W. By Gram-Schmidt orthogonalization process, we get an orthonormal basis, say $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of W. Then, for any $\mathbf{v} \in V$,

$$\mathbf{v} - \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{v}_i \rangle \mathbf{v}_i \in S^{\perp}.$$

So, $V \subset W + S^{\perp}$. Hence, $V = W + S^{\perp}$. We now need to show that $W \cap S^{\perp} = \{\mathbf{0}\}$.

To do this, let $\mathbf{v} \in W \cap S^{\perp}$. Then $\mathbf{v} \in W$ and $\mathbf{v} \in S^{\perp}$. Hence, be definition, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. That is, $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = 0$ implying $\mathbf{v} = \mathbf{0}$ and hence $W \cap S^{\perp} = \{\mathbf{0}\}$.

The third part is a direct consequence of the definition of $S^{\perp}$. $\square$

Exercise 5.3.8. *1. Let A be an $n \times n$ orthogonal matrix. Then prove that*

   (a) the rows of A form an orthonormal basis of $\mathbb{R}^n$.

   (b) the columns of A form an orthonormal basis of $\mathbb{R}^n$.

   (c) for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$, $\langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.

   (d) for any vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$.

2. *Let A be an $n \times n$ unitary matrix. Then prove that*

   (a) the rows/columns of A form an orthonormal basis of the complex vector space $\mathbb{C}^n$.

   (b) for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{n \times 1}$, $\langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.

   (c) for any vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$.

3. *Let A and B be two $n \times n$ orthogonal matrices. Then prove that AB and BA are both orthogonal matrices. Prove a similar result for unitary matrices.*

4. *Prove the statements made in Remark 5.3.4.3 about orthogonal matrices. State and prove a similar result for unitary matrices.*

5. *Let A be an $n \times n$ upper triangular matrix. If A is also an orthogonal matrix then A is a diagonal matrix with diagonal entries ± 1.*

6. *Determine an orthonormal basis of $\mathbb{R}^4$ containing the vectors $(1, -2, 1, 3)$ and $(2, 1, -3, 1)$.*

7. *Consider the real inner product space $C[-1, 1]$ with $\langle f, g \rangle = \int\limits_{-1}^{1} f(t)g(t)dt$. Prove that the polynomials $1, x, \frac{3}{2}x^2 - \frac{1}{2}, \frac{5}{2}x^3 - \frac{3}{2}x$ form an orthogonal set in $C[-1, 1]$. Find the corresponding functions $f(x)$ with $\|f(x)\| = 1$.*

8. *Consider the real inner product space $C[-\pi, \pi]$ with $\langle f, g \rangle = \int\limits_{-\pi}^{\pi} f(t)g(t)dt$. Find an orthonormal basis for $L(x, \sin x, \sin(x + 1))$.*

9. *Let M be a subspace of $\mathbb{R}^n$ and $\dim M = m$. A vector $x \in \mathbb{R}^n$ is said to be orthogonal to M if $\langle x, y \rangle = 0$ for every $y \in M$.*

   (a) How many linearly independent vectors can be orthogonal to M?

   (b) If $M = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$, determine a maximal set of linearly independent vectors orthogonal to M in $\mathbb{R}^3$.

10. *Determine an orthogonal basis of $L(\{(1, 1, 0, 1), (-1, 1, 1, -1), (0, 2, 1, 0), (1, 0, 0, 0)\})$ in $\mathbb{R}^4$.*

11. *Let $\mathbb{R}^n$ be endowed with the standard inner product. Suppose we have a vector $\mathbf{x}^t = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ with $\|\mathbf{x}\| = 1$.*

(a) *Then prove that the set* $\{\mathbf{x}\}$ *can always be extended to form an orthonormal basis of* $\mathbb{R}^n$.

(b) *Let this basis be* $\{\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. *Suppose* $\mathcal{B} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ *is the standard basis of* $\mathbb{R}^n$ *and let* $A = \begin{bmatrix} [\mathbf{x}]_{\mathcal{B}}, & [\mathbf{x}_2]_{\mathcal{B}}, & \ldots, & [\mathbf{x}_n]_{\mathcal{B}} \end{bmatrix}$. *Then prove that* A *is an orthogonal matrix.*

12. *Let* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n, n \geq 1$ *with* $\|\mathbf{u}\| = \|\mathbf{w}\| = 1$. *Prove that there exists an orthogonal matrix* A *such that* $A\mathbf{v} = \mathbf{w}$. *Prove also that* A *can be chosen such that* $\det(A) = 1$.

5.4 Orthogonal Projections and Applications

Recall that given a k-dimensional vector subspace of a vector space V of dimension n, one can always find an $(n-k)$-dimensional vector subspace W_0 of V (see Exercise 3.3.13.5) satisfying

$$W + W_0 = V \quad \text{and} \quad W \cap W_0 = \{\mathbf{0}\}.$$

The subspace W_0 is called the complementary subspace of W in V. We first use Theorem 5.3.7 to get the complementary subspace in such a way that the vectors in different subspaces are orthogonal. That is, $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ for all $\mathbf{w} \in W$ and $\mathbf{v} \in W_0$. We then use this to define an important class of linear transformations on an inner product space, called orthogonal projections.

Definition 5.4.1 (Orthogonal Complement and Orthogonal Projection)**.** *Let* W *be a subspace of a finite dimensional inner product space* V.

1. *Then* $W^{\perp}$ *is called the orthogonal complement of* W *in* V. *We represent it by writing* $V = W \oplus W^{\perp}$ *in place of* $V = W + W^{\perp}$.

2. *Also, for each* $\mathbf{v} \in V$ *there exist unique vectors* $\mathbf{w} \in W$ *and* $\mathbf{u} \in W^{\perp}$ *such that* $\mathbf{v} = \mathbf{w} + \mathbf{u}$. *We use this to define*

$$P_W : V \longrightarrow V \ \text{ by } \ P_W(\mathbf{v}) = \mathbf{w}.$$

Then P_W *is called the orthogonal projection of* V *onto* W.

Exercise 5.4.2. *Let* W *be a subspace of a finite dimensional inner product space* V. *Use* $V = W \oplus W^{\perp}$ *to define the orthogonal projection operator* $P_{W^{\perp}}$ *of* V *onto* $W^{\perp}$. *Prove that the maps* P_W *and* $P_{W^{\perp}}$ *are indeed linear transformations. What can you say about* $P_W + P_{W^{\perp}}$*?*

Example 5.4.3. 1. *Let* $V = \mathbb{R}^3$ *and* $W = \{(x, y, z) \in \mathbb{R}^3 : x + y - z = 0\}$. *Then it can be easily verified that* $\{(1, 1, -1)\}$ *is a basis of* $W^{\perp}$ *as for each* $(x, y, z) \in W$, *we have* $x + y - z = 0$ *and hence*

$$\langle (x, y, z), (1, 1, -1) \rangle = x + y - z = 0 \ \text{ for each } \ (x, y, z) \in W.$$

Also, using Equation (5.3.1), for every $\mathbf{x}^t = (x, y, z) \in \mathbb{R}^3$*, we have* $\mathbf{u} = \frac{x+y-z}{3}(1, 1, -1)$, $\mathbf{w} = (\frac{2x-y+z}{3}, \frac{-x+2y+z}{3}, \frac{x+y+2z}{3})$ *and* $\mathbf{x} = \mathbf{w} + \mathbf{u}$*. Let*

$$A = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad and \quad B = \frac{1}{3}\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}.$$

Then by definition, $P_W(\mathbf{x}) = \mathbf{w} = A\mathbf{x}$ *and* $P_{W^\perp}(\mathbf{x}) = \mathbf{u} = B\mathbf{x}$*. Observe that* $A^2 = A, B^2 = B$, $A^t = A$, $B^t = B$, $A \cdot B = \mathbf{0}_3$, $B \cdot A = \mathbf{0}_3$ *and* $A + B = I_3$*, where* $\mathbf{0}_3$ *is the zero matrix of size* 3×3 *and* I_3 *is the identity matrix of size* 3*. Also, verify that rank*$(A) = 2$ *and rank*$(B) = 1$.

2. *Let* $W = L(\ (1, 2, 1)\ ) \subset \mathbb{R}^3$*. Then using Example 5.3.3.2, and Equation (5.3.1), we get*

$$W^\perp = L(\{(-2, 1, 0), (-1, 0, 1)\}) = L(\{(-2, 1, 0), (1, 2, -5)\}),$$

$\mathbf{u} = (\frac{5x-2y-z}{6}, \frac{-2x+2y-2z}{6}, \frac{-x-2y+5z}{6})$ *and* $\mathbf{w} = \frac{x+2y+z}{6}(1, 2, 1)$ *with* $(x, y, z) = \mathbf{w} + \mathbf{u}$. *Hence, for*

$$A = \frac{1}{6}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad and \quad B = \frac{1}{6}\begin{bmatrix} 5 & -2 & -1 \\ -2 & 2 & -2 \\ -1 & -2 & 5 \end{bmatrix},$$

we have $P_W(\mathbf{x}) = \mathbf{w} = A\mathbf{x}$ *and* $P_{W^\perp}(\mathbf{x}) = \mathbf{u} = B\mathbf{x}$*. Observe that* $A^2 = A, B^2 = B$, $A^t = A$ *and* $B^t = B$, $A \cdot B = \mathbf{0}_3$, $B \cdot A = \mathbf{0}_3$ *and* $A + B = I_3$*, where* $\mathbf{0}_3$ *is the zero matrix of size* 3×3 *and* I_3 *is the identity matrix of size* 3*. Also, verify that rank*$(A) = 1$ *and rank*$(B) = 2$.

We now prove some basic properties related to orthogonal projections. We also need the following definition.

Definition 5.4.4 (Self-Adjoint Transformation/Operator)**.** *Let* V *be an inner product space with inner product* $\langle\ ,\ \rangle$*. A linear transformation* $T : V \longrightarrow V$ *is called a self-adjoint operator if* $\langle T(\mathbf{v}), \mathbf{u}\rangle = \langle \mathbf{v}, T(\mathbf{u})\rangle$ *for every* $\mathbf{u}, \mathbf{v} \in V$.

The example below gives an indication that the self-adjoint operators and Hermitian matrices are related. It also shows that the vector spaces $\mathbb{C}^n$ and $\mathbb{R}^n$ can be decomposed in terms of the null space and range space of Hermitian matrices. These examples also follow directly from the fundamental theorem of linear algebra.

Example 5.4.5. 1. *Let* A *be an* $n \times n$ *real symmetric matrix and define* $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *by* $T_A(\mathbf{x}) = A\mathbf{x}$ *for every* $\mathbf{x}^t \in \mathbb{R}^n$.

(a) *T_A is a self adjoint operator.*
As $A = A^t$*, for every* $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^n$,

$$\langle T_A(\mathbf{x}), \mathbf{y}\rangle = (\mathbf{y}^t)A\mathbf{x} = (\mathbf{y}^t)A^t\mathbf{x} = (A\mathbf{y})^t\mathbf{x} = \langle \mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, T_A(\mathbf{y})\rangle.$$

(b) $\mathcal{N}(T_A) = \mathcal{R}(T_A)^{\perp}$ *follows from Theorem 5.2.15 as* $A = A^t$. *But we do give a proof for completeness.*
Let $\mathbf{x} \in \mathcal{N}(T_A)$. *Then* $T_A(\mathbf{x}) = \mathbf{0}$ *and* $\langle \mathbf{x}, T_A(\mathbf{u})\rangle = \langle T_A(\mathbf{x}), \mathbf{u}\rangle = 0$. *Thus,* $\mathbf{x} \in \mathcal{R}(T_A)^{\perp}$ *and hence* $\mathcal{N}(T_A) \subset \mathcal{R}(T_A)^{\perp}$.
Let $\mathbf{x} \in \mathcal{R}(T_A)^{\perp}$. *Then* $0 = \langle \mathbf{x}, T_A(\mathbf{y})\rangle = \langle T_A(\mathbf{x}), \mathbf{y}\rangle$ *for every* $\mathbf{y} \in \mathbb{R}^n$. *Hence, by Exercise 2* $T_A(\mathbf{x}) = \mathbf{0}$. *That is,* $\mathbf{x} \in \mathcal{N}(A)$ *and hence* $\mathcal{R}(T_A)^{\perp} \subset \mathcal{N}(T_A)$.

(c) $\mathbb{R}^n = \mathcal{N}(T_A) \oplus \mathcal{R}(T_A)$ *as* $\mathcal{N}(T_A) = \mathcal{R}(T_A)^{\perp}$.

(d) *Thus* $\mathcal{N}(A) = Im(A)^{\perp}$, *or equivalently,* $\mathbb{R}^n = \mathcal{N}(A) \oplus Im(A)$.

2. *Let A be an $n \times n$ Hermitian matrix. Define* $T_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ *defined by* $T_A(\mathbf{z}) = A\mathbf{z}$ *for all* $\mathbf{z}^t \in \mathbb{C}^n$. *Then using arguments similar to the arguments in Example 5.4.5.1, prove the following:*

   (a) T_A *is a self-adjoint operator.*

   (b) $\mathcal{N}(T_A) = \mathcal{R}(T_A)^{\perp}$ *and* $\mathbb{C}^n = \mathcal{N}(T_A) \oplus \mathcal{R}(T_A)$.

   (c) $\mathcal{N}(A) = Im(A)^{\perp}$ *and* $\mathbb{C}^n = \mathcal{N}(A) \oplus Im(A)$.

We now state and prove the main result related with orthogonal projection operators.

Theorem 5.4.6. *Let W be a vector subspace of a finite dimensional inner product space V and let* $P_W : V \longrightarrow V$ *be the orthogonal projection operator of V onto W.*

1. *Then* $\mathcal{N}(P_W) = \{\mathbf{v} \in V : P_W(\mathbf{v}) = \mathbf{0}\} = W^{\perp} = \mathcal{R}(P_{W^{\perp}})$.

2. *Then* $\mathcal{R}(P_W) = \{P_W(\mathbf{v}) : \mathbf{v} \in V\} = W = \mathcal{N}(P_{W^{\perp}})$.

3. *Then* $P_W \circ P_W = P_W$, $P_{W^{\perp}} \circ P_{W^{\perp}} = P_{W^{\perp}}$.

4. *Let* $\mathbf{0}_V$ *denote the zero operator on V defined by* $\mathbf{0}_V(\mathbf{v}) = \mathbf{0}$ *for all* $\mathbf{v} \in V$. *Then* $P_{W^{\perp}} \circ P_W = \mathbf{0}_V$ *and* $P_W \circ P_{W^{\perp}} = \mathbf{0}_V$.

5. *Let* I_V *denote the identity operator on V defined by* $I_V(\mathbf{v}) = \mathbf{v}$ *for all* $\mathbf{v} \in V$. *Then* $I_V = P_W \oplus P_{W^{\perp}}$, *where we have written $\oplus$ instead of $+$ to indicate the relationship* $P_{W^{\perp}} \circ P_W = \mathbf{0}_V$ *and* $P_W \circ P_{W^{\perp}} = \mathbf{0}_V$.

6. *The operators* P_W *and* $P_{W^{\perp}}$ *are self-adjoint.*

Proof. PART 1: Let $\mathbf{u} \in W^{\perp}$. As $V = W \oplus W^{\perp}$, we have $\mathbf{u} = \mathbf{0} + \mathbf{u}$ for $\mathbf{0} \in W$ and $\mathbf{u} \in W^{\perp}$. Hence by definition, $P_W(\mathbf{u}) = \mathbf{0}$ and $P_{W^{\perp}}(\mathbf{u}) = \mathbf{u}$. Thus, $W^{\perp} \subset \mathcal{N}(P_W)$ and $W^{\perp} \subset \mathcal{R}(P_{W^{\perp}})$.

Also, suppose that $\mathbf{v} \in \mathcal{N}(P_W)$ for some $\mathbf{v} \in V$. As $\mathbf{v}$ has a unique expression as $\mathbf{v} = \mathbf{w} + \mathbf{u}$ for some $\mathbf{w} \in W$ and some $\mathbf{u} \in W^{\perp}$, by definition of P_W, we have $P_W(\mathbf{v}) = \mathbf{w}$. As $\mathbf{v} \in \mathcal{N}(P_W)$, by definition, $P_W(\mathbf{v}) = \mathbf{0}$ and hence $\mathbf{w} = \mathbf{0}$. That is, $\mathbf{v} = \mathbf{u} \in W^{\perp}$. Thus, $\mathcal{N}(P_W) \subset W^{\perp}$.

One can similarly show that $\mathcal{R}(P_{W^{\perp}}) \subset W^{\perp}$. Thus, the proof of the first part is complete.

PART 2: Similar argument as in the proof of Part 1.

PART 3, PART 4 AND PART 5: Let $\mathbf{v} \in V$ and let $\mathbf{v} = \mathbf{w} + \mathbf{u}$ for some $\mathbf{w} \in W$ and $\mathbf{u} \in W^{\perp}$. Then by definition,

$$\begin{aligned}
(P_W \circ P_W)(\mathbf{v}) &= P_W\big(P_W(\mathbf{v})\big) = P_W(\mathbf{w}) = \mathbf{w} \;\&\; P_W(\mathbf{v}) = \mathbf{w}, && (5.4.1)\\
(P_{W^\perp} \circ P_W)(\mathbf{v}) &= P_{W^\perp}\big(P_W(\mathbf{v})\big) = P_{W^\perp}(\mathbf{w}) = \mathbf{0} \text{ and} && (5.4.2)\\
(P_W \oplus P_{W^\perp})(\mathbf{v}) &= P_W(\mathbf{v}) + P_{W^\perp}(\mathbf{v}) = \mathbf{w} + \mathbf{u} = \mathbf{v} = I_V(\mathbf{v}). && (5.4.3)
\end{aligned}$$

Hence, applying Exercise 2 to Equations (5.4.1), (5.4.2) and (5.4.3), respectively, we get $P_W \circ P_W = P_W$, $P_{W^\perp} \circ P_W = \mathbf{0}_V$ and $I_V = P_W \oplus P_{W^\perp}$.

PART 6: Let $\mathbf{u} = \mathbf{w}_1 + \mathbf{x}_1$ and $\mathbf{v} = \mathbf{w}_2 + \mathbf{x}_2$, where $\mathbf{w}_1, \mathbf{w}_2 \in W$ and $\mathbf{x}_1, \mathbf{x}_2 \in W^{\perp}$. Then, by definition $\langle \mathbf{w}_i, \mathbf{x}_j \rangle = 0$ for $1 \le i, j \le 2$. Thus,

$$\langle P_W(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{w}_1, \mathbf{v} \rangle = \langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \langle \mathbf{u}, \mathbf{w}_2 \rangle = \langle \mathbf{u}, P_W(\mathbf{v}) \rangle$$

and the proof of the theorem is complete. $\square$

The next theorem is a generalization of Theorem 5.4.6 when a finite dimensional inner product space V can be written as $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, where W_i's are vector subspaces of V. That is, for each $\mathbf{v} \in V$ there exist unique vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ such that

1. $\mathbf{v}_i \in W_i$ for $1 \le i \le k$,
2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for each $\mathbf{v}_i \in W_i, \mathbf{v}_j \in W_j, 1 \le i \ne j \le k$ and
3. $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_k$.

We omit the proof as it basically uses arguments that are similar to the arguments used in the proof of Theorem 5.4.6.

Theorem 5.4.7. *Let V be a finite dimensional inner product space and let $W_1, W_2, \ldots, W_k$ be vector subspaces of V such that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Then for each i, j, $1 \le i \ne j \le k$, there exist orthogonal projection operators $P_{W_i} : V \longrightarrow V$ of V onto W_i satisfying the following:*

1. $\mathcal{N}(P_{W_i}) = W_i^{\perp} = W_1 \oplus W_2 \oplus \cdots \oplus W_{i-1} \oplus W_{i+1} \oplus \cdots \oplus W_k$.
2. $\mathcal{R}(P_{W_i}) = W_i$.
3. $P_{W_i} \circ P_{W_i} = P_{W_i}$.
4. $P_{W_i} \circ P_{W_j} = \mathbf{0}_V$.
5. P_{W_i} *is a self-adjoint operator, and*
6. $I_V = P_{W_1} \oplus P_{W_2} \oplus \cdots \oplus P_{W_k}$.

Remark 5.4.8. *1. By Exercise 5.4.2, P_W is a linear transformation.*

2. By Theorem 5.4.6, we observe the following:

(a) *The orthogonal projection operators P_W and $P_{W^\perp}$ are idempotent operators.*

(b) *The orthogonal projection operators P_W and $P_{W^\perp}$ are also self-adjoint operators.*

(c) *Let $\mathbf{v} \in V$. Then $\mathbf{v} - P_W(\mathbf{v}) = (I_V - P_W)(\mathbf{v}) = P_{W^\perp}(\mathbf{v}) \in W^\perp$. Thus,*

$$\langle \mathbf{v} - P_W(\mathbf{v}), \mathbf{w} \rangle = 0 \ \textit{ for every } \ \mathbf{v} \in V \ \textit{ and } \mathbf{w} \in W.$$

(d) *Using Remark 5.4.8.2c, $P_W(\mathbf{v}) - \mathbf{w} \in W$ for each $\mathbf{v} \in V$ and $\mathbf{w} \in W$. Thus,*

$$\begin{aligned} \|\mathbf{v} - \mathbf{w}\|^2 &= \|\mathbf{v} - P_W(\mathbf{v}) + P_W(\mathbf{v}) - \mathbf{w}\|^2 \\ &= \|\mathbf{v} - P_W(\mathbf{v})\|^2 + \|P_W(\mathbf{v}) - \mathbf{w}\|^2 \\ &\qquad\qquad + 2\langle \mathbf{v} - P_W(\mathbf{v}), P_W(\mathbf{v}) - \mathbf{w} \rangle \\ &= \|\mathbf{v} - P_W(\mathbf{v})\|^2 + \|P_W(\mathbf{v}) - \mathbf{w}\|^2. \end{aligned}$$

Therefore,

$$\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - P_W(\mathbf{v})\|$$

and equality holds if and only if $\mathbf{w} = P_W(\mathbf{v})$. Since $P_W(\mathbf{v}) \in W$, we see that

$$d(\mathbf{v}, W) = \inf \{\|\mathbf{v} - \mathbf{w}\| \ : \mathbf{w} \in W\} = \|\mathbf{v} - P_W(\mathbf{v})\|.$$

That is, $P_W(\mathbf{v})$ is the vector nearest to $\mathbf{v} \in W$. This can also be stated as: the vector $P_W(\mathbf{v})$ solves the following minimization problem:

$$\inf_{\mathbf{w} \in W} \|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - P_W(\mathbf{v})\|.$$

Exercise 5.4.9. 1. *Let $A \in M_n(\mathbb{R})$ be an idempotent matrix and define $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by $T_A(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v}^t \in \mathbb{R}^n$. Recall the following results from Exercise 4.3.12.5.*

(a) $T_A \circ T_A = T_A$

(b) $\mathcal{N}(T_A) \cap \mathcal{R}(T_A) = \{\mathbf{0}\}$.

(c) $\mathbb{R}^n = \mathcal{R}(T_A) + \mathcal{N}(T_A)$.

The linear map T_A need not be an orthogonal projection operator as $\mathcal{R}(T_A)^\perp$ need not be equal to $\mathcal{N}(T_A)$. Here T_A is called a projection operator of $\mathbb{R}^n$ onto $\mathcal{R}(T_A)$ along $\mathcal{N}(T_A)$.

(d) *If A is also symmetric then prove that T_A is an orthogonal projection operator.*

(e) *Which of the above results can be generalized to an $n \times n$ complex idempotent matrix A? Give reasons for your answer.*

2. *Find all 2×2 real matrices A such that $A^2 = A$. Hence or otherwise, determine all projection operators of $\mathbb{R}^2$.*

5.4.1 Matrix of the Orthogonal Projection

The minimization problem stated above arises in lot of applications. So, it is very helpful if the matrix of the orthogonal projection can be obtained under a given basis.

To this end, let W be a k-dimensional subspace of $\mathbb{R}^n$ with $W^\perp$ as its orthogonal complement. Let $P_W : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be the orthogonal projection of $\mathbb{R}^n$ onto W. Then Remark 5.3.4.6 implies that we just need to know an orthonormal basis of W. So, let $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ be an orthonormal basis of W. Thus, the matrix of P_W equals $\sum_{i=1}^k \mathbf{v}_i\mathbf{v}_i^t$. Hence, we have proved the following theorem.

Theorem 5.4.10. *Let W be a k-dimensional subspace of $\mathbb{R}^n$ and let P_W be the corresponding orthogonal projection of $\mathbb{R}^n$ onto W. Also assume that $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an orthonormal ordered basis of W. Define $A = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$, an $n \times k$ matrix. Then the matrix of P_W in the standard ordered basis of $\mathbb{R}^n$ is $AA^t = \sum_{i=1}^k \mathbf{v}_i\mathbf{v}_i^t$ (a symmetric matrix).*

We illustrate the above theorem with the help of an example. One can also see Example 5.4.3.

Example 5.4.11. *Let $W = \{(x,y,z,w) \in \mathbb{R}^4 : x = y, z = w\}$ be a subspace of W. Then an orthonormal ordered basis of W and $W^\perp$ is $\left(\frac{1}{\sqrt{2}}(1,1,0,0), \frac{1}{\sqrt{2}}(0,0,1,1)\right)$ and $\left(\frac{1}{\sqrt{2}}(1,-1,0,0), \frac{1}{\sqrt{2}}(0,0,1,-1)\right)$, respectively. Let $P_W : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be an orthogonal projection of $\mathbb{R}^4$ onto W. Then*

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } P_W[\mathcal{B},\mathcal{B}] = AA^t = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

where $\mathcal{B} = \left(\frac{1}{\sqrt{2}}(1,1,0,0), \frac{1}{\sqrt{2}}(0,0,1,1), \frac{1}{\sqrt{2}}(1,-1,0,0), \frac{1}{\sqrt{2}}(0,0,1,-1)\right)$. Verify that

1. *$P_W[\mathcal{B},\mathcal{B}]$ is symmetric,*
2. *$(P_W[\mathcal{B},\mathcal{B}])^2 = P_W[\mathcal{B},\mathcal{B}]$ and*
3. *$\big(I_4 - P_W[\mathcal{B},\mathcal{B}]\big)P_W[\mathcal{B},\mathcal{B}] = \mathbf{0} = P_W[\mathcal{B},\mathcal{B}]\big(I_4 - P_W[\mathcal{B},\mathcal{B}]\big)$.*

Also, $[(x,y,z,w)]_{\mathcal{B}} = \left(\frac{x+y}{\sqrt{2}}, \frac{z+w}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}, \frac{z-w}{\sqrt{2}}\right)^t$ and hence

$$P_W\big((x,y,z,w)\big) = \frac{x+y}{2}(1,1,0,0) + \frac{z+w}{2}(0,0,1,1)$$

is the closest vector to the subspace W for any vector $(x,y,z,w) \in \mathbb{R}^4$.

Exercise 5.4.12.

1. *Show that for any non-zero vector $\mathbf{v}^t \in \mathbb{R}^n$, rank$(\mathbf{v}\mathbf{v}^t) = 1$.*
2. *Let W be a subspace of an inner product space V and let $P : V \longrightarrow V$ be the orthogonal projection of V onto W. Let $\mathcal{B}$ be an orthonormal ordered basis of V. Then prove that $(P[\mathcal{B},\mathcal{B}])^t = P[\mathcal{B},\mathcal{B}]$.*

3. *Let $W_1 = \{(x,0) : x \in \mathbb{R}\}$ and $W_2 = \{(x,x) : x \in \mathbb{R}\}$ be two subspaces of $\mathbb{R}^2$. Let P_{W_1} and P_{W_2} be the corresponding orthogonal projection operators of $\mathbb{R}^2$ onto W_1 and W_2, respectively. Compute $P_{W_1} \circ P_{W_2}$ and conclude that the composition of two orthogonal projections need not be an orthogonal projection?*

4. *Let W be an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. Suppose $\mathcal{B}$ is an orthogonal ordered basis of $\mathbb{R}^n$ obtained by extending an orthogonal ordered basis of W. Define*

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n \quad \text{by } T(\mathbf{v}) = \mathbf{w}_0 - \mathbf{w}$$

whenever $\mathbf{v} = \mathbf{w} + \mathbf{w}_0$ for some $\mathbf{w} \in W$ and $\mathbf{w}_0 \in W^{\perp}$. Then

   (a) *prove that T is a linear transformation,*

   (b) *find $T[\mathcal{B},\mathcal{B}]$ and*

   (c) *prove that $T[\mathcal{B},\mathcal{B}]$ is an orthogonal matrix.*

T is called the reflection operator along $W^{\perp}$.

5.5 QR Decomposition*

The next result gives the proof of the QR decomposition for real matrices. A similar result holds for matrices with complex entries. The readers are advised to prove that for themselves. This decomposition and its generalizations are helpful in the numerical calculations related with eigenvalue problems (see Chapter 6).

Theorem 5.5.1 (QR Decomposition). *Let A be a square matrix of order n with real entries. Then there exist matrices Q and R such that Q is orthogonal and R is upper triangular with $A = QR$.*

In case, A is non-singular, the diagonal entries of R can be chosen to be positive. Also, in this case, the decomposition is unique.

Proof. We prove the theorem when A is non-singular. The proof for the singular case is left as an exercise.

Let the columns of A be $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis of $\mathbb{R}^n$ and hence the Gram-Schmidt orthogonalization process gives an ordered basis (see Remark 5.3.4), say $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ of $\mathbb{R}^n$ satisfying

$$\left.\begin{array}{c} L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i), \\ \|\mathbf{v}_i\| = 1, \ \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0, \end{array}\right\} \text{ for } 1 \leq i \neq j \leq n. \tag{5.5.4}$$

As $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{x}_i \in L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i)$, we can find $\alpha_{ji}, 1 \leq j \leq i$ such that

$$\mathbf{x}_i = \alpha_{1i}\mathbf{v}_1 + \alpha_{2i}\mathbf{v}_2 + \cdots + \alpha_{ii}\mathbf{v}_i = \left[(\alpha_{1i}, \ldots, \alpha_{ii}, 0 \ldots, 0)^t\right]_{\mathcal{B}}. \tag{5.5.5}$$

Now define $Q = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$ and $R = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_{nn} \end{bmatrix}$. Then by Exercise 5.3.8.4, Q is an orthogonal matrix and using (5.5.5), we get

$$\begin{aligned} QR &= [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_{nn} \end{bmatrix} \\ &= \left[\alpha_{11}\mathbf{v}_1,\ \alpha_{12}\mathbf{v}_1 + \alpha_{22}\mathbf{v}_2, \ldots, \sum_{i=1}^{n} \alpha_{in}\mathbf{v}_i\right] \\ &= [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] = A. \end{aligned}$$

Thus, we see that $A = QR$, where Q is an orthogonal matrix (see Remark 5.3.4.1) and R is an upper triangular matrix.

The proof doesn't guarantee that for $1 \leq i \leq n$, α_{ii} is positive. But this can be achieved by replacing the vector $\mathbf{v}_i$ by $-\mathbf{v}_i$ whenever α_{ii} is negative.

Uniqueness: suppose $Q_1R_1 = Q_2R_2$ then $Q_2^{-1}Q_1 = R_2R_1^{-1}$. Observe the following properties of upper triangular matrices.

1. The inverse of an upper triangular matrix is also an upper triangular matrix, and
2. product of upper triangular matrices is also upper triangular.

Thus the matrix $R_2R_1^{-1}$ is an upper triangular matrix. Also, by Exercise 5.3.8.3, the matrix $Q_2^{-1}Q_1$ is an orthogonal matrix. Hence, by Exercise 5.3.8.5, $R_2R_1^{-1} = I_n$. So, $R_2 = R_1$ and therefore $Q_2 = Q_1$. $\square$

Let $A = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$ be an $n \times k$ matrix with rank $(A) = r$. Then by Remark 5.3.4.1c , the Gram-Schmidt orthogonalization process applied to $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ yields a set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ of orthonormal vectors of $\mathbb{R}^n$ and for each i, $1 \leq i \leq r$, we have

$$L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j), \quad \text{for some } \ j, \ \ i \leq j \leq k.$$

Hence, proceeding on the lines of the above theorem, we have the following result.

Theorem 5.5.2 (Generalized QR Decomposition). *Let A be an $n \times k$ matrix of rank r. Then $A = QR$, where*

1. *$Q = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r]$ is an $n \times r$ matrix with $Q^tQ = I_r$,*
2. *$L(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r) = L(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$, and*
3. *R is an $r \times k$ matrix with rank $(R) = r$.*

Example 5.5.3. 1. *Let* $A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. *Find an orthogonal matrix* Q *and an upper triangular matrix* R *such that* $A = QR$.

Solution: *From Example 5.3.3, we know that*

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,0,1,0), \ \mathbf{v}_2 = \frac{1}{\sqrt{2}}(0,1,0,1) \text{ and } \mathbf{v}_3 = \frac{1}{\sqrt{2}}(0,-1,0,1). \tag{5.5.6}$$

We now compute $\mathbf{w}_4$. *If we denote* $\mathbf{u}_4 = (2,1,1,1)^t$ *then*

$$\mathbf{w}_4 = \mathbf{u}_4 - \langle \mathbf{u}_4, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_4, \mathbf{v}_2 \rangle \mathbf{v}_2 - \langle \mathbf{u}_4, \mathbf{v}_3 \rangle \mathbf{v}_3 = \frac{1}{2}(1,0,-1,0)^t. \tag{5.5.7}$$

Thus, using Equations (5.5.6), (5.5.7) and $Q = \left[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\right]$, *we get*

$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$ *and* $R = \begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & \frac{3}{\sqrt{2}} \\ 0 & \sqrt{2} & 0 & \sqrt{2} \\ 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & \frac{-1}{\sqrt{2}} \end{bmatrix}$. *The readers are advised to check that* $A = QR$ *is indeed correct.*

2. *Let* $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & -2 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 1 \end{bmatrix}$. *Find a* 4×3 *matrix* Q *satisfying* $Q^t Q = I_3$ *and an upper triangular matrix* R *such that* $A = QR$.

Solution: *Let us apply the Gram Schmidt orthogonalization process to the columns of* A. *That is, apply the process to the subset* $\{(1,-1,1,1), (1,0,1,0), (1,-2,1,2), (0,1,0,1)\}$ *of* $\mathbb{R}^4$.

Let $\mathbf{u}_1 = (1,-1,1,1)$. *Define* $\mathbf{v}_1 = \frac{1}{2}\mathbf{u}_1$. *Let* $\mathbf{u}_2 = (1,0,1,0)$. *Then*

$$\mathbf{w}_2 = (1,0,1,0) - \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1 = (1,0,1,0) - \mathbf{v}_1 = \frac{1}{2}(1,1,1,-1).$$

Hence, $\mathbf{v}_2 = \frac{1}{2}(1,1,1,-1)$. *Let* $\mathbf{u}_3 = (1,-2,1,2)$. *Then*

$$\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2 = \mathbf{u}_3 - 3\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}.$$

So, we again take $\mathbf{u}_3 = (0,1,0,1)$. *Then*

$$\mathbf{w}_3 = \mathbf{u}_3 - \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 - \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2 = \mathbf{u}_3 - 0\mathbf{v}_1 - 0\mathbf{v}_2 = \mathbf{u}_3.$$

So, $\mathbf{v}_3 = \frac{1}{\sqrt{2}}(0,1,0,1)$. *Hence,*

$$Q = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{-1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{-1}{2} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \text{and} \quad R = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & \sqrt{2} \end{bmatrix}.$$

The readers are advised to check the following:

(a) *rank* $(A) = 3$,

(b) $A = QR$ *with* $Q^tQ = I_3$, *and*

(c) R *a* 3×4 *upper triangular matrix with rank* $(R) = 3$.

5.6 Summary

Chapter 6

Eigenvalues, Eigenvectors and Diagonalization

6.1 Introduction and Definitions

In this chapter, the linear transformations are from the complex vector space $\mathbb{C}^n$ to itself. Observe that in this case, the matrix of the linear transformation is an $n \times n$ matrix. So, in this chapter, all the matrices are square matrices and *a vector* $\mathbf{x}$ means $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ for some positive integer n.

Example 6.1.1. *Let A be a real symmetric matrix. Consider the following problem:*

$$\text{Maximize (Minimize) } \mathbf{x}^t A\mathbf{x} \text{ such that } \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{x}^t\mathbf{x} = 1.$$

To solve this, consider the Lagrangian

$$L(\mathbf{x}, \lambda) = \mathbf{x}^t A\mathbf{x} - \lambda(\mathbf{x}^t\mathbf{x} - 1) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j - \lambda(\sum_{i=1}^{n} x_i^2 \quad - 1).$$

Partially differentiating $L(\mathbf{x}, \lambda)$ with respect to x_i for $1 \le i \le n$, we get

$$\frac{\partial L}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + \cdots + 2a_{1n}x_n - 2\lambda x_1,$$

$$\frac{\partial L}{\partial x_2} = 2a_{21}x_1 + 2a_{22}x_2 + \cdots + 2a_{2n}x_n - 2\lambda x_2,$$

and so on, till

$$\frac{\partial L}{\partial x_n} = 2a_{n1}x_1 + 2a_{n2}x_2 + \cdots + 2a_{nn}x_n - 2\lambda x_n.$$

Therefore, to get the points of extremum, we solve for

$$(0, 0, \ldots, 0)^t = (\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2}, \ldots, \frac{\partial L}{\partial x_n})^t = \frac{\partial L}{\partial \mathbf{x}} = 2(A\mathbf{x} - \lambda\mathbf{x}).$$

We therefore need to find a $\lambda \in \mathbb{R}$ and $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for the extremal problem.

Let A be a matrix of order n. In general, we ask the question:
For what values of $\lambda \in \mathbb{F}$, there exist a non-zero vector $\mathbf{x} \in \mathbb{F}^n$ such that

$$A\mathbf{x} = \lambda \mathbf{x}? \tag{6.1.1}$$

Here, $\mathbb{F}^n$ stands for either the vector space $\mathbb{R}^n$ over $\mathbb{R}$ or $\mathbb{C}^n$ over $\mathbb{C}$. Equation (6.1.1) is equivalent to the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}.$$

By Theorem 2.4.1, this system of linear equations has a non-zero solution, if

$$\text{rank } (A - \lambda I) < n, \quad \text{or equivalently} \quad \det(A - \lambda I) = 0.$$

So, to solve (6.1.1), we are forced to choose those values of $\lambda \in \mathbb{F}$ for which $\det(A - \lambda I) = 0$. Observe that $\det(A - \lambda I)$ is a polynomial in λ of degree n. We are therefore lead to the following definition.

Definition 6.1.2 (Characteristic Polynomial, Characteristic Equation)**.** *Let A be a square matrix of order n. The polynomial* $\det(A - \lambda I)$ *is called the characteristic polynomial of A and is denoted by $p_A(\lambda)$ (in short, $p(\lambda)$, if the matrix A is clear from the context). The equation $p(\lambda) = 0$ is called the characteristic equation of A. If $\lambda \in \mathbb{F}$ is a solution of the characteristic equation $p(\lambda) = 0$, then λ is called a characteristic value of A.*

Some books use the term EIGENVALUE *in place of characteristic value.*

Theorem 6.1.3. *Let $A \in \mathbb{M}_n(\mathbb{F})$. Suppose $\lambda = \lambda_0 \in \mathbb{F}$ is a root of the characteristic equation. Then there exists a non-zero $\mathbf{v} \in \mathbb{F}^n$ such that $A\mathbf{v} = \lambda_0 \mathbf{v}$.*

Proof. Since λ_0 is a root of the characteristic equation, $\det(A - \lambda_0 I) = 0$. This shows that the matrix $A - \lambda_0 I$ is singular and therefore by Theorem 2.4.1 the linear system

$$(A - \lambda_0 I_n)\mathbf{x} = \mathbf{0}$$

has a non-zero solution. □

Remark 6.1.4. *Observe that the linear system $A\mathbf{x} = \lambda \mathbf{x}$ has a solution $\mathbf{x} = \mathbf{0}$ for every $\lambda \in \mathbb{F}$. So, we consider only those $\mathbf{x} \in \mathbb{F}^n$ that are non-zero and are also solutions of the linear system $A\mathbf{x} = \lambda \mathbf{x}$.*

Definition 6.1.5 (Eigenvalue and Eigenvector)**.** *Let $A \in \mathbb{M}_n(\mathbb{F})$ and let the linear system $A\mathbf{x} = \lambda \mathbf{x}$ has a non-zero solution $\mathbf{x} \in \mathbb{F}^n$ for some $\lambda \in \mathbb{F}$. Then*

1. *$\lambda \in \mathbb{F}$ is called an eigenvalue of A,*
2. *$\mathbf{x} \in \mathbb{F}^n$ is called an eigenvector corresponding to the eigenvalue λ of A, and*
3. *the tuple $(\lambda, \mathbf{x})$ is called an eigen-pair.*

Remark 6.1.6. *To understand the difference between a characteristic value and an eigenvalue, we give the following example.*

Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. *Then* $p_A(\lambda) = \lambda^2 + 1$. *Also, define the linear operator* $T_A : \mathbb{F}^2 \longrightarrow \mathbb{F}^2$ *by* $T_A(\mathbf{x}) = A\mathbf{x}$ *for every* $\mathbf{x} \in \mathbb{F}^2$.

1. *Suppose* $\mathbb{F} = \mathbb{C}$, *i.e.,* $A \in \mathbb{M}_2(\mathbb{C})$. *Then the roots of* $p(\lambda) = 0$ *in* $\mathbb{C}$ *are* $\pm i$. *So,* A *has* $(i, (1, i)^t)$ *and* $(-i, (i, 1)^t)$ *as eigen-pairs.*

2. *If* $A \in \mathbb{M}_2(\mathbb{R})$, *then* $p(\lambda) = 0$ *has no solution in* $\mathbb{R}$. *Therefore, if* $\mathbb{F} = \mathbb{R}$, *then* A *has no eigenvalue but it has* $\pm i$ *as characteristic values.*

Remark 6.1.7.

1. *Let* $A \in \mathbb{M}_n(\mathbb{F})$. *Suppose* $(\lambda, \mathbf{x})$ *is an eigen-pair of* A. *Then for any* $c \in \mathbb{F}$, $c \neq 0$, $(\lambda, c\mathbf{x})$ *is also an eigen-pair for* A. *Similarly, if* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ *are linearly independent eigenvectors of* A *corresponding to the eigenvalue* λ, *then* $\sum\limits_{i=1}^{r} c_i \mathbf{x}_i$ *is also an eigenvector of* A *corresponding to* λ *if at least one* $c_i \neq 0$. *Hence, if* S *is a collection of eigenvectors, it is implicitly understood that the set* S *is* LINEARLY INDEPENDENT.

2. *Suppose* $p_A(\lambda_0) = 0$ *for some* $\lambda_0 \in \mathbb{F}$. *Then* $A - \lambda_0 I$ *is singular. If rank* $(A - \lambda_0 I) = r$ *then* $r < n$. *Hence, by Theorem 2.4.1 on page 48, the system* $(A - \lambda_0 I)\mathbf{x} = \mathbf{0}$ *has* $n - r$ *linearly independent solutions. That is,* A *has* $n - r$ *linearly independent eigenvectors corresponding to* λ_0 *whenever rank* $(A - \lambda_0 I) = r$.

Example 6.1.8.

1. *Let* $A = diag(d_1, d_2, \ldots, d_n)$ *with* $d_i \in \mathbb{R}$ *for* $1 \leq i \leq n$. *Then* $p(\lambda) = \prod\limits_{i=1}^{n} (\lambda - d_i)$ *and the eigen-pairs are* $(d_1, \mathbf{e}_1), (d_2, \mathbf{e}_2), \ldots, (d_n, \mathbf{e}_n)$.

2. *Let* $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. *Then* $p(\lambda) = (1 - \lambda)^2$. *Hence, the characteristic equation has roots* $1, 1$. *That is,* 1 *is a repeated eigenvalue. But the system* $(A - I_2)\mathbf{x} = \mathbf{0}$ *for* $\mathbf{x} = (x_1, x_2)^t$ *implies that* $x_2 = 0$. *Thus,* $\mathbf{x} = (x_1, 0)^t$ *is a solution of* $(A - I_2)\mathbf{x} = \mathbf{0}$. *Hence using Remark 6.1.7.1,* $(1, 0)^t$ *is an eigenvector. Therefore, note that* 1 IS A REPEATED EIGENVALUE WHEREAS THERE IS ONLY ONE EIGENVECTOR.

3. *Let* $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. *Then* $p(\lambda) = (1 - \lambda)^2$. *Again,* 1 *is a repeated root of* $p(\lambda) = 0$. *But in this case, the system* $(A - I_2)\mathbf{x} = \mathbf{0}$ *has a solution for every* $\mathbf{x}^t \in \mathbb{R}^2$. *Hence, we can* CHOOSE ANY TWO LINEARLY INDEPENDENT VECTORS $\mathbf{x}^t, \mathbf{y}^t$ *from* $\mathbb{R}^2$ *to get* $(1, \mathbf{x})$ *and* $(1, \mathbf{y})$ *as the two eigen-pairs. In general, if* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^n$ *are linearly independent vectors then* $(1, \mathbf{x}_1)$, $(1, \mathbf{x}_2)$, $\ldots, (1, \mathbf{x}_n)$ *are eigen-pairs of the identity matrix,* I_n.

4. *Let* $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. *Then* $p(\lambda) = (\lambda - 3)(\lambda + 1)$ *and its roots are* $3, -1$. *Verify that the eigen-pairs are* $(3, (1, 1)^t)$ *and* $(-1, (1, -1)^t)$. *The readers are advised to prove the linear independence of the two eigenvectors.*

5. *Let* $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. *Then* $p(\lambda) = \lambda^2 - 2\lambda + 2$ *and its roots are* $1+i, 1-i$. *Hence, over* $\mathbb{R}$, *the matrix* A *has no eigenvalue. Over* $\mathbb{C}$, *the reader is required to show that the eigen-pairs are* $(1+i, (i,1)^t)$ *and* $(1-i, (1,i)^t)$.

Exercise 6.1.9. 1. *Find the eigenvalues of a triangular matrix.*

2. *Find eigen-pairs over* $\mathbb{C}$, *for each of the following matrices:* $\begin{bmatrix} 1 & 1+i \\ 1-i & 1 \end{bmatrix}$, $\begin{bmatrix} i & 1+i \\ -1+i & i \end{bmatrix}$, $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ *and* $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$.

3. *Let* A *and* B *be similar matrices.*

   (a) *Then prove that* A *and* B *have the same set of eigenvalues.*

   (b) *If* $B = PAP^{-1}$ *for some invertible matrix* P *then prove that* $P\mathbf{x}$ *is an eigenvector of* B *if and only if* $\mathbf{x}$ *is an eigenvector of* A.

4. *Let* $A = (a_{ij})$ *be an* $n \times n$ *matrix. Suppose that for all* i, $1 \le i \le n$, $\sum\limits_{j=1}^{n} a_{ij} = a$. *Then prove that* a *is an eigenvalue of* A. *What is the corresponding eigenvector?*

5. *Prove that the matrices* A *and* A^t *have the same set of eigenvalues. Construct a* 2×2 *matrix* A *such that the eigenvectors of* A *and* A^t *are different.*

6. *Let* A *be a matrix such that* $A^2 = A$ *(*A *is called an idempotent matrix). Then prove that its eigenvalues are either* 0 *or* 1 *or both.*

7. *Let* A *be a matrix such that* $A^k = \mathbf{0}$ *(*A *is called a nilpotent matrix) for some positive integer* $k \ge 1$. *Then prove that its eigenvalues are all* 0.

8. *Compute the eigen-pairs of the matrices* $\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$ *and* $\begin{bmatrix} 2 & i \\ i & 0 \end{bmatrix}$.

Theorem 6.1.10. *Let* $A = [a_{ij}]$ *be an* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$, *not necessarily distinct. Then* $\det(A) = \prod\limits_{i=1}^{n} \lambda_i$ *and* $tr(A) = \sum\limits_{i=1}^{n} a_{ii} = \sum\limits_{i=1}^{n} \lambda_i$.

Proof. Since $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the n eigenvalues of A, by definition,

$$\det(A - \lambda I_n) = p(\lambda) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n). \tag{6.1.2}$$

(6.1.2) is an identity in λ as polynomials. Therefore, by substituting $\lambda = 0$ in (6.1.2), we get

$$\det(A) = (-1)^n (-1)^n \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i.$$

Also,

$$\det(A-\lambda I_n) = \begin{bmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{bmatrix} \tag{6.1.3}$$

$$= a_0 - \lambda a_1 + \lambda^2 a_2 + \cdots + (-1)^{n-1}\lambda^{n-1} a_{n-1} + (-1)^n \lambda^n \tag{6.1.4}$$

for some $a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}$. Note that a_{n-1}, the coefficient of $(-1)^{n-1}\lambda^{n-1}$, comes from the product

$$(a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda).$$

So, $a_{n-1} = \sum\limits_{i=1}^{n} a_{ii} = \mathrm{tr}(A)$ by definition of trace.

But , from (6.1.2) and (6.1.4), we get

$$a_0 - \lambda a_1 + \lambda^2 a_2 + \cdots + (-1)^{n-1}\lambda^{n-1} a_{n-1} + (-1)^n\lambda^n = (-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n). \tag{6.1.5}$$

Therefore, comparing the coefficient of $(-1)^{n-1}\lambda^{n-1}$, we have

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

Hence, we get the required result. $\square$

Exercise 6.1.11. 1. *Let A be a skew symmetric matrix of order $2n+1$. Then prove that 0 is an eigenvalue of A.*

2. *Let A be a 3×3 orthogonal matrix ($AA^t = I$). If $\det(A) = 1$, then prove that there exists a non-zero vector $\mathbf{v} \in \mathbb{R}^3$ such that $A\mathbf{v} = \mathbf{v}$.*

Let A be an $n \times n$ matrix. Then in the proof of the above theorem, we observed that the characteristic equation $\det(A-\lambda I) = 0$ is a polynomial equation of degree n in λ. Also, for some numbers $a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}$, it has the form

$$\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^2 + \cdots a_1\lambda + a_0 = 0.$$

Note that, in the expression $\det(A-\lambda I) = 0$, λ is an element of $\mathbb{F}$. Thus, we can only substitute λ by elements of $\mathbb{F}$.

It turns out that the expression

$$A^n + a_{n-1}A^{n-1} + a_{n-2}A^2 + \cdots a_1 A + a_0 I = \mathbf{0}$$

holds true as a matrix identity. This is a celebrated theorem called the Cayley Hamilton Theorem. We state this theorem without proof and give some implications.

Theorem 6.1.12 (Cayley Hamilton Theorem)**.** *Let A be a square matrix of order n. Then A satisfies its characteristic equation. That is,*

$$A^n + a_{n-1}A^{n-1} + a_{n-2}A^2 + \cdots a_1 A + a_0 I = \mathbf{0}$$

holds true as a matrix identity.

Some of the implications of Cayley Hamilton Theorem are as follows.

Remark 6.1.13. 1. *Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then its characteristic polynomial is $p(\lambda) = \lambda^2$. Also, for the function, $f(x) = x$, $f(0) = 0$, and $f(A) = A \neq \mathbf{0}$. This shows that the condition $f(\lambda) = 0$ for each eigenvalue λ of A does not imply that $f(A) = \mathbf{0}$.*

2. *let A be a square matrix of order n with characteristic polynomial $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^2 + \cdots a_1\lambda + a_0$.*

   (a) *Then for any positive integer ℓ, we can use the division algorithm to find numbers $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ and a polynomial $f(\lambda)$ such that*

   $$\begin{aligned} \lambda^\ell &= f(\lambda)\big(\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^2 + \cdots a_1\lambda + a_0\big) \\ &\quad + \alpha_0 + \lambda\alpha_1 + \cdots + \lambda^{n-1}\alpha_{n-1}. \end{aligned}$$

   Hence, by the Cayley Hamilton Theorem,

   $$A^\ell = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1}.$$

   That is, we just need to compute the powers of A till $n-1$.
   In the language of graph theory, it says the following:
   "Let G be a graph on n vertices. Suppose there is no path of length $n-1$ or less from a vertex v to a vertex u of G. Then there is no path from v to u of any length. That is, the graph G is disconnected and v and u are in different components."

   (b) *If A is non-singular then $a_n = \det(A) \neq 0$ and hence*

   $$A^{-1} = \frac{-1}{a_n}[A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1 I].$$

   This matrix identity can be used to calculate the inverse.
   Note that the vector A^{-1} (as an element of the vector space of all $n \times n$ matrices) is a linear combination of the vectors $I, A, \ldots, A^{n-1}$.

Exercise 6.1.14. *Find inverse of the following matrices by using the Cayley Hamilton Theorem*

$$i) \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 1 & 1 & 2 \end{bmatrix} \quad ii) \begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad iii) \begin{bmatrix} 1 & -2 & -1 \\ -2 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Theorem 6.1.15. *If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix A with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, then the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is linearly independent.*

Proof. The proof is by induction on the number m of eigenvalues. The result is obviously true if $m = 1$ as the corresponding eigenvector is non-zero and we know that any set containing exactly one non-zero vector is linearly independent.

Let the result be true for m, $1 \leq m < k$. We prove the result for $m+1$. We consider the equation

$$c_1 x_1 + c_2 x_2 + \cdots + c_{m+1} x_{m+1} = \mathbf{0} \tag{6.1.6}$$

for the unknowns $c_1, c_2, \ldots, c_{m+1}$. We have

$$\begin{aligned} \mathbf{0} = A\mathbf{0} &= A(c_1 x_1 + c_2 x_2 + \cdots + c_{m+1} x_{m+1}) \\ &= c_1 A x_1 + c_2 A x_2 + \cdots + c_{m+1} A x_{m+1} \\ &= c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 + \cdots + c_{m+1} \lambda_{m+1} x_{m+1}. \end{aligned} \tag{6.1.7}$$

From Equations (6.1.6) and (6.1.7), we get

$$c_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + c_3(\lambda_3 - \lambda_1)\mathbf{x}_3 + \cdots + c_{m+1}(\lambda_{m+1} - \lambda_1)\mathbf{x}_{m+1} = \mathbf{0}.$$

This is an equation in m eigenvectors. So, by the induction hypothesis, we have

$$c_i(\lambda_i - \lambda_1) = 0 \quad \text{for} \quad 2 \leq i \leq m+1.$$

But the eigenvalues are distinct implies $\lambda_i - \lambda_1 \neq 0$ for $2 \leq i \leq m+1$. We therefore get $c_i = 0$ for $2 \leq i \leq m+1$. Also, $\mathbf{x}_1 \neq \mathbf{0}$ and therefore (6.1.6) gives $c_1 = 0$.

Thus, we have the required result. □

We are thus lead to the following important corollary.

Corollary 6.1.16. *The eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Exercise 6.1.17. 1. *Let $A, B \in \mathbb{M}_n(\mathbb{R})$. Prove that*

(a) *if λ is an eigenvalue of A then λ^k is an eigenvalue of A^k for all $k \in \mathbb{Z}^+$.*

(b) *if A is invertible and λ is an eigenvalue of A then $\dfrac{1}{\lambda}$ is an eigenvalue of A^{-1}.*

(c) *if A is nonsingular then BA^{-1} and $A^{-1}B$ have the same set of eigenvalues.*

(d) *AB and BA have the same non-zero eigenvalues.*

In each case, what can you say about the eigenvectors?

2. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be an invertible matrix and let $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y}^t A^{-1}\mathbf{x} \neq 0$. Define $B = \mathbf{x}\mathbf{y}^t A^{-1}$. Then prove that*

(a) *$\lambda_0 = \mathbf{y}^t A^{-1}\mathbf{x}$ is an eigenvalue of B of multiplicity 1.*

(b) *0 is an eigenvalue of B of multiplicity $n-1$ [Hint: Use Exercise 6.1.17.1d].*

(c) *$1 + \alpha\lambda_0$ is an eigenvalue of $I + \alpha B$ of multiplicity 1, for any $\alpha \in \mathbb{R}$, $\alpha \neq 0$.*

(d) *1 is an eigenvalue of $I + \alpha B$ of multiplicity $n-1$, for any $\alpha \in \mathbb{R}$.*

(e) $\det(A + \alpha \mathbf{x}\mathbf{y}^t)$ *equals* $(1 + \alpha\lambda_0)\det(A)$ *for any* $\alpha \in \mathbb{R}$. *This result is known as the Shermon-Morrison formula for determinant.*

3. *Let* $A, B \in \mathbb{M}_2(\mathbb{R})$ *such that* $\det(A) = \det(B)$ *and* $tr(A) = tr(B)$.

   (a) *Do A and B have the same set of eigenvalues?*

   (b) *Give examples to show that the matrices A and B need not be similar.*

4. *Let* $A, B \in \mathbb{M}_n(\mathbb{R})$. *Also, let* $(\lambda_1, \mathbf{u})$ *be an eigen-pair for A and* $(\lambda_2, \mathbf{v})$ *be an eigen-pair for B.*

   (a) *If* $\mathbf{u} = \alpha\mathbf{v}$ *for some* $\alpha \in \mathbb{R}$ *then* $(\lambda_1 + \lambda_2, \mathbf{u})$ *is an eigen-pair for* $A + B$.

   (b) *Give an example to show that if* $\mathbf{u}$ *and* $\mathbf{v}$ *are linearly independent then* $\lambda_1 + \lambda_2$ *need not be an eigenvalue of* $A + B$.

5. *Let* $A \in \mathbb{M}_n(\mathbb{R})$ *be an invertible matrix with eigen-pairs* $(\lambda_1, \mathbf{u}_1), (\lambda_2, \mathbf{u}_2), \ldots, (\lambda_n, \mathbf{u}_n)$. *Then prove that* $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *forms a basis of* $\mathbb{R}^n(\mathbb{R})$. *If* $[\mathbf{b}]_{\mathcal{B}} = (c_1, c_2, \ldots, c_n)^t$ *then the system* $A\mathbf{x} = \mathbf{b}$ *has the unique solution*

$$\mathbf{x} = \frac{c_1}{\lambda_1}\mathbf{u}_1 + \frac{c_2}{\lambda_2}\mathbf{u}_2 + \cdots + \frac{c_n}{\lambda_n}\mathbf{u}_n.$$

6.2 Diagonalization

Let $A \in \mathbb{M}_n(\mathbb{F})$ and let $T_A : \mathbb{F}^n \longrightarrow \mathbb{F}^n$ be the corresponding linear operator. In this section, we ask the question "does there exist a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that $T_A[\mathcal{B}, \mathcal{B}]$, the matrix of the linear operator T_A with respect to the ordered basis $\mathcal{B}$, is a diagonal matrix." it will be shown that for a certain class of matrices, the answer to the above question is in affirmative. To start with, we have the following definition.

Definition 6.2.1 (Matrix Digitalization)**.** *A matrix A is said to be diagonalizable if there exists a non-singular matrix P such that* $P^{-1}AP$ *is a diagonal matrix.*

Remark 6.2.2. *Let* $A \in \mathbb{M}_n(\mathbb{F})$ *be a diagonalizable matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *By definition, A is similar to a diagonal matrix* $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *as similar matrices have the same set of eigenvalues and the eigenvalues of a diagonal matrix are its diagonal entries.*

Example 6.2.3. *Let* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. *Then we have the following:*

1. *Let* $V = \mathbb{R}^2$. *Then A has no real eigenvalue (see Example 6.1.7 and hence A doesn't have eigenvectors that are vectors in* $\mathbb{R}^2$. *Hence, there does not exist any non-singular* 2×2 *real matrix P such that* $P^{-1}AP$ *is a diagonal matrix.*

2. *In case, $V = \mathbb{C}^2(\mathbb{C})$, the two complex eigenvalues of A are $-i, i$ and the corresponding eigenvectors are $(i,1)^t$ and $(-i,1)^t$, respectively. Also, $(i,1)^t$ and $(-i,1)^t$ can be taken as a basis of $\mathbb{C}^2(\mathbb{C})$. Define $U = \frac{1}{\sqrt{2}}\begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$. Then*

$$U^*AU = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}.$$

Theorem 6.2.4. *Let $A \in \mathbb{M}_n(\mathbb{R})$. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.*

Proof. Let A be diagonalizable. Then there exist matrices P and D such that

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Or equivalently, $AP = PD$. Let $P = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Then $AP = PD$ implies that

$$A\mathbf{u}_i = d_i\mathbf{u}_i \quad \text{for} \quad 1 \le i \le n.$$

Since $\mathbf{u}_i$'s are the columns of a non-singular matrix P, using Corollary 4.3.10, they form a linearly independent set. Thus, we have shown that if A is diagonalizable then A has n linearly independent eigenvectors.

Conversely, suppose A has n linearly independent eigenvectors $\mathbf{u}_i$, $1 \le i \le n$ with eigenvalues λ_i. Then $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$. Let $P = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$. Since $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly independent, by Corollary 4.3.10, P is non-singular. Also,

$$\begin{aligned} AP &= [A\mathbf{u}_1, A\mathbf{u}_2, \ldots, A\mathbf{u}_n] = [\lambda_1\mathbf{u}_1, \lambda_2\mathbf{u}_2, \ldots, \lambda_n\mathbf{u}_n] \\ &= [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \lambda_n \end{bmatrix} = PD. \end{aligned}$$

Therefore, the matrix A is diagonalizable. ☐

Corollary 6.2.5. *If the eigenvalues of a $A \in \mathbb{M}_n(\mathbb{R})$ are distinct then A is diagonalizable.*

Proof. As $A \in \mathbb{M}_n(\mathbb{R})$, it has n eigenvalues. Since all the eigenvalues of A are distinct, by Corollary 6.1.16, the n eigenvectors are linearly independent. Hence, by Theorem 6.2.4, A is diagonalizable. ☐

Corollary 6.2.6. *Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of $A \in \mathbb{M}_n(\mathbb{R})$ and let $p(\lambda)$ be its characteristic polynomial. Suppose that for each i, $1 \le i \le k$, $(x-\lambda_i)^{m_i}$ divides $p(\lambda)$ but $(x-\lambda_i)^{m_i+1}$ does not divides $p(\lambda)$ for some positive integers m_i. Then prove that A is diagonalizable if and only if $\dim\big(\ker(A-\lambda_i I)\big) = m_i$ for each $i, 1 \le i \le k$. Or equivalently A is diagonalizable if and only if rank$(A-\lambda_i I) = n - m_i$ for each $i, 1 \le i \le k$.*

Proof. As A is diagonalizable, by Theorem 6.2.4, A has n linearly independent eigenvalues. Also, by assumption, $\sum\limits_{i=1}^{k} m_i = n$ as $\deg(p(\lambda)) = n$. Hence, for each eigenvalue $\lambda_i,\ 1 \le i \le k$, A has exactly m_i linearly independent eigenvectors. Thus, for each $i,\ 1 \le i \le k$, the homogeneous linear system $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$ has exactly m_i linearly independent vectors in its solution set. Therefore, $\dim\big(\ker(A - \lambda_i I)\big) \ge m_i$. Indeed $\dim\big(\ker(A - \lambda_i I)\big) = m_i$ for $1 \le i \le k$ follows from a simple counting argument.

Now suppose that for each $i,\ 1 \le i \le k,\ \dim\big(\ker(A - \lambda_i I)\big) = m_i$. Then for each $i,\ 1 \le i \le k$, we can choose m_i linearly independent eigenvectors. Also by Corollary 6.1.16, the eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence A has $n = \sum\limits_{i=1}^{k} m_i$ linearly independent eigenvectors. Hence by Theorem 6.2.4, A is diagonalizable. ☐

Example 6.2.7. 1. *Let* $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. *Then* $p_A(\lambda) = (2-\lambda)^2(1-\lambda)$. *Hence, the eigenvalues of* A *are* $1, 2, 2$. *Verify that* $\big(1, (1, 0, -1)^t\big)$ *and* $\big((2, (1, 1, -1)^t\big)$ *are the only eigen-pairs. That is, the matrix* A *has exactly one eigenvector corresponding to the repeated eigenvalue* 2. *Hence, by Theorem 6.2.4,* A *is not diagonalizable.*

2. *Let* $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. *Then* $p_A(\lambda) = (4-\lambda)(1-\lambda)^2$. *Hence,* A *has eigenvalues* $1, 1, 4$. *Verify that* $\mathbf{u}_1 = (1, -1, 0)^t$ *and* $\mathbf{u}_2 = (1, 0, -1)^t$ *are eigenvectors corresponding to* 1 *and* $\mathbf{u}_3 = (1, 1, 1)^t$ *is an eigenvector corresponding to the eigenvalue* 4. *As* $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ *are linearly independent, by Theorem 6.2.4,* A *is diagonalizable.*

   Note that the vectors $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *(corresponding to the eigenvalue* 1*) are not orthogonal. So, in place of* $\mathbf{u}_1, \mathbf{u}_2$*, we will take the orthogonal vectors* $\mathbf{u}_2$ *and* $\mathbf{w} = 2\mathbf{u}_1 - \mathbf{u}_2$ *as eigenvectors. Now define* $U = [\frac{1}{\sqrt{3}}\mathbf{u}_3, \frac{1}{\sqrt{2}}\mathbf{u}_2, \frac{1}{\sqrt{6}}\mathbf{w}] = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$. *Then* U *is an orthogonal matrix and* $U^*AU = \text{diag}(4, 1, 1)$.

   Observe that A *is a symmetric matrix. In this case, we chose our eigenvectors to be mutually orthogonal. This result is true for any real symmetric matrix* A. *This result will be proved later.*

Exercise 6.2.8. 1. *Are the matrices* $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ *and* $B = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ *for some* $\theta, 0 \le \theta \le 2\pi$*, diagonalizable?*

2. *Find the eigen-pairs of* $A = [a_{ij}]_{n\times n}$*, where* $a_{ij} = a$ *if* $i = j$ *and* b*, otherwise.*

3. *Let* $A \in \mathbb{M}_n(\mathbb{R})$ *and* $B \in \mathbb{M}_m(\mathbb{R})$. *Suppose* $C = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix}$. *Then prove that* C *is diagonalizable if and only if both* A *and* B *are diagonalizable.*

4. *Let* $T : \mathbb{R}^5 \longrightarrow \mathbb{R}^5$ *be a linear operator with rank* $(T - I) = 3$ *and*
$$\mathcal{N}(T) = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 \mid x_1 + x_4 + x_5 = 0,\ x_2 + x_3 = 0\}.$$
   (a) *Determine the eigenvalues of* T*?*
   (b) *Find the number of linearly independent eigenvectors corresponding to each eigenvalue?*
   (c) *Is* T *diagonalizable? Justify your answer.*

5. *Let* A *be a non-zero square matrix such that* $A^2 = \mathbf{0}$*. Prove that* A *cannot be diagonalized. [Hint: Use Remark 6.2.2.]*

6. *Are the following matrices diagonalizable?*
$$i)\ \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 2 & 3 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix},\quad ii)\ \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix},\quad iii)\ \begin{bmatrix} 1 & -3 & 3 \\ 0 & -5 & 6 \\ 0 & -3 & 4 \end{bmatrix} \text{ and } iv)\ \begin{bmatrix} 2 & i \\ i & 0 \end{bmatrix}.$$

6.3 Diagonalizable Matrices

In this section, we will look at some special classes of square matrices that are diagonalizable. Recall that for a matrix $A = [a_{ij}]$, $A^* = [\overline{a_{ji}}] = \overline{A^t} = \overline{A}^t$, is called the conjugate transpose of A. We also recall the following definitions.

Definition 6.3.1 (Special Matrices)**.** 1. *A matrix* $A \in \mathbb{M}_n(\mathbb{C})$ *is called*

   (a) *a Hermitian matrix if* $A^* = A$*.*
   (b) *a unitary matrix if* $A\,A^* = A^*A = I_n$*.*
   (c) *a skew-Hermitian matrix if* $A^* = -A$*.*
   (d) *a normal matrix if* $A^*A = AA^*$*.*

2. *A matrix* $A \in \mathbb{M}_n(\mathbb{R})$ *is called*

   (a) *a symmetric matrix if* $A^t = A$*.*
   (b) *an orthogonal matrix if* $A\,A^t = A^tA = I_n$*.*
   (c) *a skew-symmetric matrix if* $A^t = -A$*.*

Note that a symmetric matrix is always Hermitian, a skew-symmetric matrix is always skew-Hermitian and an orthogonal matrix is always unitary. Each of these matrices are normal. If A is a unitary matrix then $A^* = A^{-1}$.

Example 6.3.2. 1. *Let* $B = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix}$*. Then* B *is skew-Hermitian.*

2. *Let* $A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$*. Then* A *is a unitary matrix and* B *is a normal matrix. Note that* $\sqrt{2}A$ *is also a normal matrix.*

Definition 6.3.3 (Unitary Equivalence). *Let $A, B \in \mathbb{M}_n(\mathbb{C})$. They are called unitarily equivalent if there exists a unitary matrix U such that $A = U^*BU$. As U is a unitary matrix, $U^* = U^{-1}$. Hence, A is also unitarily similar to B.*

Exercise 6.3.4.

1. *Let A be a square matrix such that UAU^* is a diagonal matrix for some unitary matrix U. Prove that A is a normal matrix.*

2. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then $A = \frac{1}{2}(A+A^*)+\frac{1}{2}(A-A^*)$, where $\frac{1}{2}(A+A^*)$ is the Hermitian part of A and $\frac{1}{2}(A-A^*)$ is the skew-Hermitian part of A. Recall that a similar result was given in Exercise 1.3.3.1.*

3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Prove that $A - A^*$ is always skew-Hermitian.*

4. *Every square matrix can be uniquely expressed as $A = S + iT$, where both S and T are Hermitian matrices.*

5. *Does there exist a unitary matrix U such that $U^{-1}AU = B$ where*
$$A = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}.$$

Theorem 6.3.5. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then*

1. *the eigenvalues, $\lambda_i, 1 \le i \le n$, of A are real.*

2. *A is unitarily diagonalizable. That is, there exists a unitary matrix U such that $U^*AU = D$; where $D = diag(\lambda_1, \ldots, \lambda_n)$. In other words, the eigenvectors of A form an orthonormal basis of $\mathbb{C}^n$.*

Proof. For the proof of Part 1, let $(\lambda, \mathbf{x})$ be an eigen-pair. Then $A\mathbf{x} = \lambda\mathbf{x}$ and $A^* = A$ implies that $\mathbf{x}^*A = \mathbf{x}^*A^* = (A\mathbf{x})^* = (\lambda\mathbf{x})^* = \overline{\lambda}\mathbf{x}^*$. Hence,

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

As $\mathbf{x}$ is an eigenvector, $\mathbf{x} \neq \mathbf{0}$ and therefore $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x} \neq 0$. Thus $\lambda = \overline{\lambda}$. That is, λ is a real number.

For the proof of Part 2, we use induction on n, the size of the matrix. The result is clearly true for $n = 1$. Let the result be true for $n = k - 1$. we need to prove the result for $n = k$.

Let $(\lambda_1, \mathbf{x})$ be an eigen-pair of a $k \times k$ matrix A with $\|\mathbf{x}\| = 1$. Then by Part 1, $\lambda_1 \in \mathbb{R}$. As $\{\mathbf{x}\}$ is a linearly independent set, by Theorem 3.3.11 and the Gram-Schmidt Orthogonalization process, we get an orthonormal basis $\{\mathbf{x}, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of $\mathbb{C}^k$. Let $U_1 = [\mathbf{x}, \mathbf{u}_2, \ldots, \mathbf{u}_k]$ (the vectors $\mathbf{x}, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are columns of the matrix U_1). Then U_1 is a unitary matrix. In particular, $\mathbf{u}_i^*\mathbf{x} = 0$, for $2 \le i \le k$. Therefore, for $2 \le i \le k$,

$$\mathbf{x}^*(A\mathbf{u}_i) = (A\mathbf{u}_i)^*\mathbf{x} = (\mathbf{u}_i^*A^*)\mathbf{x} = \mathbf{u}_i^*(A^*\mathbf{x}) = \mathbf{u}_i^*(A\mathbf{x}) = \mathbf{u}_i^*(\lambda_1\mathbf{x}) = \lambda_1(\mathbf{u}_i^*\mathbf{x}) = 0 \text{ and}$$

$$
\begin{aligned}
U_1^* A U_1 &= U_1^*[A\mathbf{x},\ A\mathbf{u}_2,\ \cdots, A\mathbf{u}_k] = \begin{bmatrix} \mathbf{x}^* \\ \mathbf{u}_2^* \\ \vdots \\ \mathbf{u}_k^* \end{bmatrix} [\lambda_1 \mathbf{x},\ A\mathbf{u}_2,\ \cdots, A\mathbf{u}_k] \\
&= \begin{bmatrix} \lambda_1 \mathbf{x}^*\mathbf{x} & \cdots & \mathbf{x}^* A \mathbf{u}_k \\ \mathbf{u}_2^*(\lambda_1 \mathbf{x}) & \cdots & \mathbf{u}_2^*(A\mathbf{u}_k) \\ \vdots & \ddots & \vdots \\ \mathbf{u}_k^*(\lambda_1 \mathbf{x}) & \cdots & \mathbf{u}_k^*(A\mathbf{u}_k) \end{bmatrix} = \left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & \\ \vdots & B \\ \mathbf{0} & \end{array}\right],
\end{aligned}
$$

where B is a $(k-1) \times (k-1)$ matrix. As $(U_1^* A U_1)^* = U_1^* A U_1$ and $\lambda_1 \in \mathbb{R}$, the matrix B is also Hermitian. Therefore, by induction hypothesis there exists a $(k-1) \times (k-1)$ unitary matrix U_2 such that $U_2^* B U_2 = D_2 = \mathrm{diag}(\lambda_2, \ldots, \lambda_k)$, where $\lambda_i \in \mathbb{R}$, for $2 \le i \le k$ are the eigenvalues of B. Define $U = U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix}$. Then U is a unitary matrix and

$$
\begin{aligned}
U^* A U &= \left(U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right)^* A \left(U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right) \\
&= \left(\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^* \end{bmatrix} U_1^* \right) A \left(U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \right) = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^* \end{bmatrix} (U_1^* A U_1) \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2^* \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & B \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & U_2^* B U_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D_2 \end{bmatrix}.
\end{aligned}
$$

Observe that $\lambda_2, \ldots, \lambda_n$ are also the eigenvalues of A. Thus, $U^* A U$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_k$, the eigenvalues of A. Hence, the result follows. □

Corollary 6.3.6. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be a symmetric matrix. Then*

1. *the eigenvalues of A are all real,*

2. *the eigenvectors can be chosen to have real entries and*

3. *the eigenvectors also form an orthonormal basis of $\mathbb{R}^n$.*

Proof. As A is symmetric, A is also a Hermitian matrix. Hence, by Theorem 6.3.5, the eigenvalues of A are all real. Let $(\lambda,\ \mathbf{x})$ be an eigen-pair of A. Suppose $\mathbf{x}^t \in \mathbb{C}^n$. Then there exist $\mathbf{y}^t, \mathbf{z}^t \in \mathbb{R}^n$ such that $\mathbf{x} = \mathbf{y} + i\mathbf{z}$. So,

$$A\mathbf{x} = \lambda \mathbf{x} \Longrightarrow A(\mathbf{y} + i\mathbf{z}) = \lambda(\mathbf{y} + i\mathbf{z}).$$

Comparing the real and imaginary parts, we get $A\mathbf{y} = \lambda \mathbf{y}$ and $A\mathbf{z} = \lambda \mathbf{z}$. Thus, we can choose the eigenvectors to have real entries.

The readers are advised to prove the orthonormality of the eigenvectors (see the proof of Theorem 6.3.5). □

Exercise 6.3.7. 1. *Let A be a skew-Hermitian matrix. Then the eigenvalues of A are either zero or purely imaginary. Also, the eigenvectors corresponding to distinct eigenvalues are mutually orthogonal. [Hint: Carefully see the proof of Theorem 6.3.5.]*

2. *Let A be a normal matrix with $(\lambda, \mathbf{x})$ as an eigen-pair. Then*

   (a) *$(A^*)^k\mathbf{x}$ for $k \in \mathbb{Z}^+$ is also an eigenvector corresponding to λ.*

   (b) *$(\overline{\lambda}, \mathbf{x})$ is an eigen-pair for A^*. [Hint: Verify $\|A^*\mathbf{x} - \overline{\lambda}\mathbf{x}\|^2 = \|A\mathbf{x} - \lambda\mathbf{x}\|^2$.]*

3. *Let A be an $n \times n$ unitary matrix. Then*

   (a) *the rows of A form an orthonormal basis of $\mathbb{C}^n$.*

   (b) *the columns of A form an orthonormal basis of $\mathbb{C}^n$.*

   (c) *for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{n\times 1}$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.*

   (d) *for any vector $\mathbf{x} \in \mathbb{C}^{n\times 1}$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$.*

   (e) *$|\lambda| = 1$ for any eigenvalue λ of A.*

   (f) *the eigenvectors $\mathbf{x}, \mathbf{y}$ corresponding to distinct eigenvalues λ and μ satisfy $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. That is, if $(\lambda, \mathbf{x})$ and $(\mu, \mathbf{y})$ are eigen-pairs with $\lambda \neq \mu$, then $\mathbf{x}$ and $\mathbf{y}$ are mutually orthogonal.*

4. *Show that the matrices $A = \begin{bmatrix} 4 & 4 \\ 0 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 10 & 9 \\ -4 & -2 \end{bmatrix}$ are similar. Is it possible to find a unitary matrix U such that $A = U^*BU$?*

5. *Let A be a 2×2 orthogonal matrix. Then prove the following:*

   (a) *if $\det(A) = 1$, then $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for some θ, $0 \le \theta < 2\pi$. That is, A counterclockwise rotates every point in $\mathbb{R}^2$ by an angle θ.*

   (b) *if $\det A = -1$, then $A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ for some θ, $0 \le \theta < 2\pi$. That is, A reflects every point in $\mathbb{R}^2$ about a line passing through origin. Determine this line. Or equivalently, there exists a non-singular matrix P such that $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.*

6. *Let A be a 3×3 orthogonal matrix. Then prove the following:*

   (a) *if $\det(A) = 1$, then A is a rotation about a fixed axis, in the sense that A has an eigen-pair $(1, \mathbf{x})$ such that the restriction of A to the plane $\mathbf{x}^\perp$ is a two dimensional rotation in $\mathbf{x}^\perp$.*

   (b) *if $\det A = -1$, then A corresponds to a reflection through a plane P, followed by a rotation about the line through origin that is orthogonal to P.*

7. *Let* $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. *Find a non-singular matrix* P *such that* $P^{-1}AP = diag\ (4, 1, 1)$. *Use this to compute* A^{301}.

8. *Let* A *be a Hermitian matrix. Then prove that rank*(A) *equals the number of non-zero eigenvalues of* A.

Remark 6.3.8. *Let* A *and* B *be the* 2×2 *matrices in Exercise 6.3.7.4. Then* A *and* B *were similar matrices but they were not unitarily equivalent. In numerical calculations, unitary transformations are preferred as compared to similarity transformations due to the following main reasons:*

1. *Exercise 6.3.7.3d implies that an orthonormal change of basis does not alter the sum of squares of the absolute values of the entries. This need not be true under a non-singularity change of basis.*

2. *For a unitary matrix* U, $U^{-1} = U^*$ *and hence unitary equivalence is computationally simpler.*

3. *Also there is no round-off error in the operation of "conjugate transpose".*

We next prove the Schur's Lemma and use it to show that normal matrices are unitarily diagonalizable. The proof is similar to the proof of Theorem 6.3.5. We give it again so that the readers have a better understanding of unitary transformations.

Lemma 6.3.9. *(Schur's Lemma) Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Then* A *is unitarily similar to an upper triangular matrix.*

Proof. We will prove the result by induction on n. The result is clearly true for $n = 1$. Let the result be true for $n = k - 1$. we need to prove the result for $n = k$.

Let $(\lambda_1, \mathbf{x})$ be an eigen-pair of a $k \times k$ matrix A with $\|\mathbf{x}\| = 1$. Let us extend the set $\{\mathbf{x}\}$, a linearly independent set, to form an orthonormal basis $\{\mathbf{x}, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k\}$ (using *Gram-Schmidt Orthogonalization*) of $\mathbb{C}^k$. Then $U_1 = [\mathbf{x}\ \mathbf{u}_2\ \cdots \mathbf{u}_k]$ is a unitary matrix and

$$U_1^* A U_1 = U_1^* [A\mathbf{x}\ A\mathbf{u}_2\ \cdots A\mathbf{u}_k] = \begin{bmatrix} \mathbf{x}^* \\ \mathbf{u}_2^* \\ \vdots \\ \mathbf{u}_k^* \end{bmatrix} [\lambda_1 \mathbf{x}\ A\mathbf{u}_2\ \cdots A\mathbf{u}_k] = \left[\begin{array}{c|c} \lambda_1 & * \\ \hline \mathbf{0} & \\ \vdots & B \\ \mathbf{0} & \end{array}\right],$$

where B is a $(k-1) \times (k-1)$ matrix. By induction hypothesis there exists a $(k-1) \times (k-1)$ unitary matrix U_2 such that $U_2^* B U_2$ is an upper triangular matrix with diagonal entries $\lambda_2, \ldots, \lambda_k$, the eigenvalues of B. Define $U = U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix}$. Then check that U is a unitary matrix and $U^* A U$ is an upper triangular matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_k$, the eigenvalues of the matrix A. Hence, the result follows. ☐

In Lemma 6.3.9, it can be observed that whenever A is a normal matrix then the matrix B is also a normal matrix. It is also known that if T is an upper triangular matrix that satisfies $TT^* = T^*T$ then T is a diagonal matrix (see Exercise 16). Thus, it follows that normal matrices are diagonalizable. We state it as a remark.

Remark 6.3.10 (The Spectral Theorem for Normal Matrices). *Let A be an $n \times n$ normal matrix. Then there exists an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ of $\mathbb{C}^n(\mathbb{C})$ such that $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $1 \le i \le n$. In particular, if $U - [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ then U^*AU is a diagonal matrix.*

Exercise 6.3.11.

1. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be an invertible matrix. Prove that $AA^t = PDP^t$, where P is an orthogonal and D is a diagonal matrix with positive diagonal entries.*

2. *Let* $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 2 & -1 & \sqrt{2} \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ *and* $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$. *Prove that A and B are unitarily equivalent via the unitary matrix U. Hence, conclude that the upper triangular matrix obtained in the "Schur's Lemma" need not be unique.*

3. *Prove Remark 6.3.10.*

4. *Let A be a normal matrix. If all the eigenvalues of A are 0 then prove that $A = \mathbf{0}$. What happens if all the eigenvalues of A are 1?*

5. *Let A be an $n \times n$ matrix. Prove that if A is*

   (a) Hermitian and $\mathbf{x}A\mathbf{x}^* = 0$ for all $\mathbf{x} \in \mathbb{C}^n$ then $A = \mathbf{0}$.

   (b) a real, symmetric matrix and $\mathbf{x}A\mathbf{x}^t = 0$ for all $\mathbf{x} \in \mathbb{R}^n$ then $A = \mathbf{0}$.

   Do these results hold for arbitrary matrices?

We end this chapter with an application of the theory of diagonalization to the study of conic sections in analytic geometry and the study of maxima and minima in analysis.

6.4 Sylvester's Law of Inertia and Applications

Definition 6.4.1 (Bilinear Form). *Let A be an $n \times n$ real symmetric matrix. A bilinear form in $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t$ is an expression of the type*

$$Q(\mathbf{x}, \mathbf{y}) = \mathbf{y}^t A \mathbf{x} = \sum_{i,j=1}^{n} a_{ij} x_i y_j.$$

Definition 6.4.2 (Sesquilinear Form). *Let A be an $n \times n$ Hermitian matrix. A sesquilinear form in $\mathbf{x} = (x_1, x_2, \ldots, x_n)^*$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^*$ is given by*

$$H(\mathbf{x}, \mathbf{y}) = \mathbf{y}^* A \mathbf{x} = \sum_{i,j=1}^{n} a_{ij} x_i \overline{y_j}.$$

Observe that if $A = I_n$ then the bilinear (sesquilinear) form reduces to the standard real (complex) inner product. Also, it can be easily seen that $H(\mathbf{x}, \mathbf{y})$ is 'linear' in $\mathbf{x}$, the first component and 'conjugate linear' in $\mathbf{y}$, the second component. The expression $Q(\mathbf{x}, \mathbf{x})$ is called the quadratic form and $H(\mathbf{x}, \mathbf{x})$ the Hermitian form. We generally write $Q(\mathbf{x})$ and $H(\mathbf{x})$ in place of $Q(\mathbf{x}, \mathbf{x})$ and $H(\mathbf{x}, \mathbf{x})$, respectively. It can be easily shown that for any choice of $\mathbf{x}$, the Hermitian form $H(\mathbf{x})$ is a real number. Hence, for any real number α, the equation $H(\mathbf{x}) = \alpha$, represents a conic in $\mathbb{C}^n$.

Example 6.4.3. *Let* $A = \begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix}$. *Then* $A^* = A$ *and for* $\mathbf{x} = (x_1, x_2)^*$,

$$\begin{aligned} H(\mathbf{x}) &= \mathbf{x}^* A \mathbf{x} = (\overline{x}_1, \overline{x}_2) \begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \overline{x}_1 x_1 + 2\overline{x}_2 x_2 + (2-i)\overline{x}_1 x_2 + (2+i)\overline{x}_2 x_1 \\ &= |x_1|^2 + 2|x_2|^2 + 2Re[(2-i)\overline{x}_1 x_2] \end{aligned}$$

where 'Re' denotes the real part of a complex number. This shows that for every choice of $\mathbf{x}$ *the Hermitian form is always real. Why?*

The main idea of this section is to express $H(\mathbf{x})$ as sum of squares and hence determine the possible values that it can take. Note that if we replace $\mathbf{x}$ by $c\mathbf{x}$, where c is any complex number, then $H(\mathbf{x})$ simply gets multiplied by $|c|^2$ and hence one needs to study only those $\mathbf{x}$ for which $\|\mathbf{x}\| = 1$, i.e., $\mathbf{x}$ is a normalized vector.

Let $A^* = A \in \mathbb{M}_n(\mathbb{C})$. Then by Theorem 6.3.5, the eigenvalues $\lambda_i, 1 \le i \le n$, of A are real and there exists a unitary matrix U such that $U^*AU = D \equiv diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Now define, $\mathbf{z} = (z_1, z_2, \ldots, z_n)^* = U^*\mathbf{x}$. Then $\|\mathbf{z}\| = 1$, $\mathbf{x} = U\mathbf{z}$ and

$$H(\mathbf{x}) = \mathbf{z}^* U^* A U \mathbf{z} = \mathbf{z}^* D \mathbf{z} = \sum_{i=1}^{n} \lambda_i |z_i|^2 = \sum_{i=1}^{p} \left| \sqrt{|\lambda_i|}\, z_i \right|^2 - \sum_{i=p+1}^{r} \left| \sqrt{|\lambda_i|}\, z_i \right|^2. \tag{6.4.1}$$

Thus, the possible values of $H(\mathbf{x})$ depend only on the eigenvalues of A. Since U is an invertible matrix, the components z_i's of $\mathbf{z} = U^*\mathbf{x}$ are commonly known as linearly independent linear forms. Also, note that in Equation (6.4.1), the number p (respectively $r - p$) seems to be related to the number of eigenvalues of A that are positive (respectively negative). This is indeed true. That is, in any expression of $H(\mathbf{x})$ as a sum of n absolute squares of linearly independent linear forms, the number p (respectively $r - p$) gives the number of positive (respectively negative) eigenvalues of A. This is stated as the next lemma and it popularly known as the 'Sylvester's law of inertia'.

Lemma 6.4.4. *Let* $A \in \mathbb{M}_n(\mathbb{C})$ *be a Hermitian matrix and let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)^*$. *Then every Hermitian form* $H(\mathbf{x}) = \mathbf{x}^* A \mathbf{x}$, *in* n *variables can be written as*

$$H(\mathbf{x}) = |y_1|^2 + |y_2|^2 + \cdots + |y_p|^2 - |y_{p+1}|^2 - \cdots - |y_r|^2$$

where $y_1, y_2, \ldots, y_r$ *are linearly independent linear forms in* $x_1, x_2, \ldots, x_n$, *and the integers* p *and* r *satisfying* $0 \le p \le r \le n$, *depend only on* A.

Proof. From Equation (6.4.1) it is easily seen that $H(\mathbf{x})$ has the required form. We only need to show that p and r are uniquely determined by A. Hence, let us assume on the contrary that there exist positive integers p, q, r, s with $p > q$ such that

$$\begin{aligned} H(\mathbf{x}) &= |y_1|^2 + |y_2|^2 + \cdots + |y_p|^2 - |y_{p+1}|^2 - \cdots - |y_r|^2 \\ &= |z_1|^2 + |z_2|^2 + \cdots + |z_q|^2 - |z_{q+1}|^2 - \cdots - |z_s|^2, \end{aligned}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^* = M\mathbf{x}$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n)^* = N\mathbf{x}$ for some invertible matrices M and N. Hence, $\mathbf{z} = B\mathbf{y}$ for some invertible matrix B. Let us write $Y_1 = (y_1, \ldots, y_p)^*$, $Z_1 = (z_1, \ldots, z_q)^*$ and $B = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$, where B_1 is a $q \times p$ matrix. As $p > q$, the homogeneous linear system $B_1 Y_1 = \mathbf{0}$ has a non-zero solution. Let $\tilde{Y_1} = (\tilde{y_1}, \ldots, \tilde{y_p})^*$ be a non-zero solution and let $\tilde{\mathbf{y}}^* = (\tilde{Y_1}^*, \mathbf{0}^*)$. Then

$$H(\tilde{\mathbf{y}}) = |\tilde{y_1}|^2 + |\tilde{y_2}|^2 + \cdots + |\tilde{y_p}|^2 = -(|z_{q+1}|^2 + \cdots + |z_s|^2).$$

Now, this can hold only if $\tilde{y_1} = \tilde{y_2} = \cdots = \tilde{y_p} = 0$, which gives a contradiction. Hence $p = q$. Similarly, the case $r > s$ can be resolved. Thus, the proof of the lemma is over. $\square$

Remark 6.4.5. *The integer r is the rank of the matrix A and the number $r - 2p$ is sometimes called the inertial degree of A.*

We complete this chapter by understanding the graph of

$$ax^2 + 2hxy + by^2 + 2fx + 2gy + c = 0$$

for $a, b, c, f, g, h \in \mathbb{R}$. We first look at the following example.

Example 6.4.6. *Sketch the graph of $3x^2 + 4xy + 3y^2 = 5$.*

Solution: *Note that* $3x^2 + 4xy + 3y^2 = [x, \ \ y] \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ *and the eigen-pairs of the matrix* $\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ *are* $(5, (1,1)^t), \ (1, (1,-1)^t)$. *Thus,*

$$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Now, let $u = \frac{x+y}{\sqrt{2}}$ *and* $v = \frac{x-y}{\sqrt{2}}$. *Then*

$$\begin{aligned} 3x^2 + 4xy + 3y^2 &= [x, \ \ y] \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= [x, \ \ y] \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= [u, \ \ v] \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 5u^2 + v^2. \end{aligned}$$

Thus, the given graph reduces to $5u^2 + v^2 = 5$ *or equivalently to* $u^2 + \frac{v^2}{5} = 1$. *Therefore, the given graph represents an ellipse with the principal axes* $u = 0$ *and* $v = 0$ *(correspinding to the line* $x + y = 0$ *and* $x - y = 0$*, respectively). See Figure 6.4.6.*

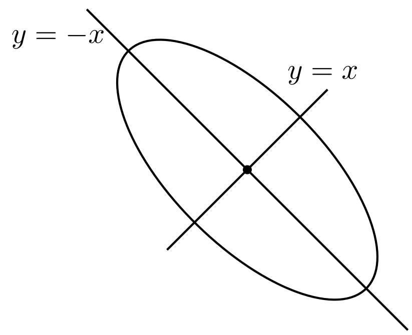

Figure 1: The ellipse $3x^2 + 4xy + 3y^2 = 5$.

We now consider the general conic. We obtain conditions on the eigenvalues of the associated quadratic form, defined below, to characterize conic sections in $\mathbb{R}^2$ (endowed with the standard inner product).

Definition 6.4.7 (Quadratic Form)**.** *Let* $ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ *be the equation of a general conic. The quadratic expression*

$$ax^2 + 2hxy + by^2 = \begin{bmatrix} x, & y \end{bmatrix} \begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

is called the quadratic form associated with the given conic.

Proposition 6.4.8. *For fixed real numbers* a, b, c, g, f *and* h*, consider the general conic*

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0.$$

Then prove that this conic represents

1. *an ellipse if* $ab - h^2 > 0$,
2. *a parabola if* $ab - h^2 = 0$, *and*
3. *a hyperbola if* $ab - h^2 < 0$.

Proof. Let $A = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$. Then $ax^2 + 2hxy + by^2 = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix}$ is the associated quadratic form. As A is a symmetric matrix, by Corollary 6.3.6, the eigenvalues λ_1, λ_2 of A are both real, the corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ are orthonormal and A is unitarily diagonalizable with $A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^t \\ \mathbf{u}_2^t \end{bmatrix}$. Let $\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^t \\ \mathbf{u}_2^t \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. Then $ax^2 + 2hxy + by^2 = \lambda_1 u^2 + \lambda_2 v^2$ and the equation of the conic section in the (u, v)-plane, reduces to

$$\lambda_1 u^2 + \lambda_2 v^2 + 2g_1 u + 2f_1 v + c = 0. \qquad (6.4.2)$$

Now, depending on the eigenvalues λ_1, λ_2, we consider different cases:

1. $\lambda_1 = 0 = \lambda_2$. Substituting $\lambda_1 = \lambda_2 = 0$ in Equation (6.4.2) gives the straight line $2g_1 u + 2f_1 v + c = 0$ in the (u, v)-plane.

2. $\lambda_1 = 0, \lambda_2 > 0$. As $\lambda_1 = 0$, $\det(A) = 0$. That is, $ab - h^2 = \det(A) = 0$. Also, in this case, Equation (6.4.2) reduces to

$$\lambda_2 (v + d_1)^2 = d_2 u + d_3 \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R}.$$

   To understand this case, we need to consider the following subcases:

   (a) Let $d_2 = d_3 = 0$. Then $v + d_1 = 0$ is a pair of coincident lines.

   (b) Let $d_2 = 0, \ d_3 \neq 0$.

      i. If $d_3 > 0$, then we get a pair of parallel lines given by $v = -d_1 \pm \sqrt{\frac{d_3}{\lambda_2}}$.

      ii. If $d_3 < 0$, the solution set of the corresponding conic is an empty set.

   (c) If $d_2 \neq 0$. Then the given equation is of the form $Y^2 = 4aX$ for some translates $X = x + \alpha$ and $Y = y + \beta$ and thus represents a parabola.

3. $\lambda_1 > 0$ and $\lambda_2 < 0$. In this case, $ab - h^2 = \det(A) = \lambda_1 \lambda_2 < 0$. Let $\lambda_2 = -\alpha_2$ with $\alpha_2 > 0$. Then Equation (6.4.2) can be rewritten as

$$\lambda_1 (u + d_1)^2 - \alpha_2 (v + d_2)^2 = d_3 \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R} \tag{6.4.3}$$

   whose understanding requires the following subcases:

   (a) Let $d_3 = 0$. Then Equation (6.4.3) reduces to

$$\left(\sqrt{\lambda_1}(u + d_1) + \sqrt{\alpha_2}(v + d_2)\right) \cdot \left(\sqrt{\lambda_1}(u + d_1) - \sqrt{\alpha_2}(v + d_2)\right) = 0$$

   or equivalently, a pair of intersecting straight lines in the (u, v)-plane.

   (b) Let $d_3 \neq 0$. In particular, let $d_3 > 0$. Then Equation (6.4.3) reduces to

$$\frac{\lambda_1 (u + d_1)^2}{d_3} - \frac{\alpha_2 (v + d_2)^2}{d_3} = 1$$

   or equivalently, a hyperbola in the (u, v)-plane, with principal axes $u + d_1 = 0$ and $v + d_2 = 0$.

4. $\lambda_1, \lambda_2 > 0$. In this case, $ab - h^2 = \det(A) = \lambda_1 \lambda_2 > 0$ and Equation (6.4.2) can be rewritten as

$$\lambda_1 (u + d_1)^2 + \lambda_2 (v + d_2)^2 = d_3 \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R}. \tag{6.4.4}$$

   We consider the following three subcases to understand this.

   (a) Let $d_3 = 0$. Then Equation (6.4.4) reduces to a pair of perpendicular lines $u + d_1 = 0$ and $v + d_2 = 0$ in the (u, v)-plane.

   (b) Let $d_3 < 0$. Then the solution set of Equation (6.4.4) is an empty set.

(c) Let $d_3 > 0$. Then Equation (6.4.4) reduces to the ellipse

$$\frac{\lambda_1(u+d_1)^2}{d_3} + \frac{\alpha_2(v+d_2)^2}{d_3} = 1$$

whose principal axes are $u + d_1 = 0$ and $v + d_2 = 0$.

□

Remark 6.4.9. *Observe that the condition* $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$ *implies that the principal axes of the conic are functions of the eigenvectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$.

Exercise 6.4.10. *Sketch the graph of the following surfaces:*

1. $x^2 + 2xy + y^2 - 6x - 10y = 3.$
2. $2x^2 + 6xy + 3y^2 - 12x - 6y = 5.$
3. $4x^2 - 4xy + 2y^2 + 12x - 8y = 10.$
4. $2x^2 - 6xy + 5y^2 - 10x + 4y = 7.$

As a last application, we consider the following problem that helps us in understanding the quadrics. Let

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + 2lx + 2my + 2nz + q = 0 \tag{6.4.5}$$

be a general quadric. Then to get the geometrical idea of this quadric, do the following:

1. Define $A = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2l \\ 2m \\ 2n \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Note that Equation (6.4.5) can be rewritten as $\mathbf{x}^t A\mathbf{x} + \mathbf{b}^t\mathbf{x} + q = 0.$
2. As A is symmetric, find an orthogonal matrix P such that $P^tAP = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$.
3. Let $\mathbf{y} = P^t\mathbf{x} = (y_1, y_2, y_3)^t$. Then Equation (6.4.5) reduces to
$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + 2l_1y_1 + 2l_2y_2 + 2l_3y_3 + q' = 0. \tag{6.4.6}$$
4. Depending on which $\lambda_i \neq 0$, rewrite Equation (6.4.6). That is, if $\lambda_1 \neq 0$ then rewrite $\lambda_1 y_1^2 + 2l_1y_1$ as $\lambda_1\left(y_1 + \frac{l_1}{\lambda_1}\right)^2 - \left(\frac{l_1}{\lambda_1}\right)^2$.
5. Use the condition $\mathbf{x} = P\mathbf{y}$ to determine the center and the planes of symmetry of the quadric in terms of the original system.

Example 6.4.11. *Determine the following quadrics*

1. $2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz + 4x + 2y + 4z + 2 = 0.$

2. $3x^2 - y^2 + z^2 + 10 = 0$.

Solution: *For Part 1, observe that* $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$ *and* $q = 2$. *Also, the orthonormal matrix* $P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$ *and* $P^t A P = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. *Hence, the quadric reduces to* $4y_1^2 + y_2^2 + y_3^2 + \frac{10}{\sqrt{3}} y_1 + \frac{2}{\sqrt{2}} y_2 - \frac{2}{\sqrt{6}} y_3 + 2 = 0$. *Or equivalently to*

$$4(y_1 + \frac{5}{4\sqrt{3}})^2 + (y_2 + \frac{1}{\sqrt{2}})^2 + (y_3 - \frac{1}{\sqrt{6}})^2 = \frac{9}{12}.$$

So, the standard form of the quadric is $4z_1^2 + z_2^2 + z_3^2 = \frac{9}{12}$, *where the center is given by* $(x, y, z)^t = P(\frac{-5}{4\sqrt{3}}, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{6}})^t = (\frac{-3}{4}, \frac{1}{4}, \frac{-3}{4})^t$.

For Part 2, observe that $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \mathbf{0}$ *and* $q = 10$. *In this case, we can rewrite the quadric as*

$$\frac{y^2}{10} - \frac{3x^2}{10} - \frac{z^2}{10} = 1$$

which is the equation of a hyperboloid consisting of two sheets.

The calculation of the planes of symmetry is left as an exercise to the reader.

Chapter 7

Appendix

7.1 Permutation/Symmetric Groups

In this section, S denotes the set $\{1, 2, \ldots, n\}$.

Definition 7.1.1. 1. *A function $\sigma : S \longrightarrow S$ is called a permutation on n elements if σ is both one to one and onto.*

2. *The set of all functions $\sigma : S \longrightarrow S$ that are both one to one and onto will be denoted by $\mathcal{S}_n$. That is, $\mathcal{S}_n$ is the set of all permutations of the set $\{1, 2, \ldots, n\}$.*

Example 7.1.2. 1. *In general, we represent a permutation σ by $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$. This representation of a permutation is called a* TWO ROW NOTATION *for σ.*

2. *For each positive integer n, $\mathcal{S}_n$ has a special permutation called the identity permutation, denoted Id_n, such that $Id_n(i) = i$ for $1 \leq i \leq n$. That is, $Id_n = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$.*

3. *Let $n = 3$. Then*

$$\begin{aligned} \mathcal{S}_3 \;=\; & \left\{ \tau_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \tau_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \tau_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \right. \\ & \left. \tau_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \tau_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \tau_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \right\} \end{aligned} \tag{7.1.1}$$

Remark 7.1.3. 1. *Let $\sigma \in \mathcal{S}_n$. Then σ is determined if $\sigma(i)$ is known for $i = 1, 2, \ldots, n$. As σ is both one to one and onto, $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\} = S$. So, there are n choices for $\sigma(1)$ (any element of S), $n-1$ choices for $\sigma(2)$ (any element of S different from $\sigma(1)$), and so on. Hence, there are $n(n-1)(n-2)\cdots 3\cdot 2\cdot 1 = n!$ possible permutations. Thus, the number of elements in $\mathcal{S}_n$ is $n!$. That is, $|\mathcal{S}_n| = n!$.*

2. *Suppose that $\sigma, \tau \in \mathcal{S}_n$. Then both σ and τ are one to one and onto. So, their composition map $\sigma \circ \tau$, defined by $(\sigma \circ \tau)(i) = \sigma\big(\tau(i)\big)$, is also both one to one and onto. Hence, $\sigma \circ \tau$ is also a permutation. That is, $\sigma \circ \tau \in \mathcal{S}_n$.*

3. *Suppose $\sigma \in \mathcal{S}_n$. Then σ is both one to one and onto. Hence, the function $\sigma^{-1} : S \longrightarrow S$ defined by $\sigma^{-1}(m) = \ell$ if and only if $\sigma(\ell) = m$ for $1 \leq m \leq n$, is well defined and indeed σ^{-1} is also both one to one and onto. Hence, for every element $\sigma \in \mathcal{S}_n$, $\sigma^{-1} \in \mathcal{S}_n$ and is the inverse of σ.*

4. *Observe that for any $\sigma \in \mathcal{S}_n$, the compositions $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = Id_n$.*

Proposition 7.1.4. *Consider the set of all permutations $\mathcal{S}_n$. Then the following holds:*

1. *Fix an element $\tau \in \mathcal{S}_n$. Then the sets $\{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}$ and $\{\tau \circ \sigma : \sigma \in \mathcal{S}_n\}$ have exactly $n!$ elements. Or equivalently,*

$$\mathcal{S}_n = \{\tau \circ \sigma : \sigma \in \mathcal{S}_n\} = \{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}.$$

2. *$\mathcal{S}_n = \{\sigma^{-1} : \sigma \in \mathcal{S}_n\}$.*

Proof. For the first part, we need to show that given any element $\alpha \in \mathcal{S}_n$, there exists elements $\beta, \gamma \in \mathcal{S}_n$ such that $\alpha = \tau \circ \beta = \gamma \circ \tau$. It can easily be verified that $\beta = \tau^{-1} \circ \alpha$ and $\gamma = \alpha \circ \tau^{-1}$.

For the second part, note that for any $\sigma \in \mathcal{S}_n$, $(\sigma^{-1})^{-1} = \sigma$. Hence the result holds. $\square$

Definition 7.1.5. *Let $\sigma \in \mathcal{S}_n$. Then the number of inversions of σ, denoted $n(\sigma)$, equals*

$$|\{(i,j) : \; i < j, \; \sigma(i) > \sigma(j) \; \}|.$$

Note that, for any $\sigma \in \mathcal{S}_n$, $n(\sigma)$ also equals

$$\sum_{i=1}^{n} |\{\sigma(j) < \sigma(i), \quad \text{for} \quad j = i+1, i+2, \ldots, n\}|.$$

Definition 7.1.6. *A permutation $\sigma \in \mathcal{S}_n$ is called a transposition if there exists two positive integers $m, r \in \{1, 2, \ldots, n\}$ such that $\sigma(m) = r$, $\sigma(r) = m$ and $\sigma(i) = i$ for $1 \leq i \neq m, r \leq n$.*

For the sake of convenience, a transposition σ for which $\sigma(m) = r$, $\sigma(r) = m$ and $\sigma(i) = i$ for $1 \leq i \neq m, r \leq n$ will be denoted simply by $\sigma = (m \; r)$ or $(r \; m)$. Also, note that for any transposition $\sigma \in \mathcal{S}_n$, $\sigma^{-1} = \sigma$. That is, $\sigma \circ \sigma = Id_n$.

Example 7.1.7. 1. *The permutation $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$ is a transposition as $\tau(1) = 3, \tau(3) = 1$, $\tau(2) = 2$ and $\tau(4) = 4$. Here note that $\tau = (1 \; 3) = (3 \; 1)$. Also, check that*

$$n(\tau) = |\{(1,2), (1,3), (2,3)\}| = 3.$$

2. *Let* $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 2 & 3 & 5 & 1 & 9 & 8 & 7 & 6 \end{pmatrix}$. *Then check that*

$$n(\tau) = 3 + 1 + 1 + 1 + 0 + 3 + 2 + 1 = 12.$$

3. *Let* ℓ, m *and* r *be distinct element from* $\{1, 2, \ldots, n\}$. *Suppose* $\tau = (m\ r)$ *and* $\sigma = (m\ \ell)$. *Then*

$$\begin{array}{rcl} (\tau \circ \sigma)(\ell) & = & \tau\big(\sigma(\ell)\big) = \tau(m) = r, \quad (\tau \circ \sigma)(m) = \tau\big(\sigma(m)\big) = \tau(\ell) = \ell \\ (\tau \circ \sigma)(r) & = & \tau\big(\sigma(r)\big) = \tau(r) = m, \quad \textit{and} \quad (\tau \circ \sigma)(i) = \tau\big(\sigma(i)\big) = \tau(i) = i \ \ \textit{if } i \neq \ell, m, r. \end{array}$$

Therefore,

$$\tau \circ \sigma = (m\ r) \circ (m\ \ell) = \begin{pmatrix} 1 & 2 & \cdots & \ell & \cdots & m & \cdots & r & \cdots & n \\ 1 & 2 & \cdots & r & \cdots & \ell & \cdots & m & \cdots & n \end{pmatrix} = (r\ l) \circ (r\ m).$$

Similarly check that $\sigma \circ \tau = \begin{pmatrix} 1 & 2 & \cdots & \ell & \cdots & m & \cdots & r & \cdots & n \\ 1 & 2 & \cdots & m & \cdots & r & \cdots & \ell & \cdots & n \end{pmatrix}$.

With the above definitions, we state and prove two important results.

Theorem 7.1.8. *For any* $\sigma \in \mathcal{S}_n$, σ *can be written as composition (product) of transpositions.*

Proof. We will prove the result by induction on $n(\sigma)$, the number of inversions of σ. If $n(\sigma) = 0$, then $\sigma = Id_n = (1\ 2) \circ (1\ 2)$. So, let the result be true for all $\sigma \in \mathcal{S}_n$ with $n(\sigma) \leq k$.

For the next step of the induction, suppose that $\tau \in \mathcal{S}_n$ with $n(\tau) = k + 1$. Choose the smallest positive number, say ℓ, such that

$$\tau(i) = i, \quad \text{for } i = 1, 2, \ldots, \ell - 1 \ \text{ and } \tau(\ell) \neq \ell.$$

As τ is a permutation, there exists a positive number, say m, such that $\tau(\ell) = m$. Also, note that $m > \ell$. Define a transposition σ by $\sigma = (\ell\ m)$. Then note that

$$(\sigma \circ \tau)(i) = i, \quad \text{for } i = 1, 2, \ldots, \ell.$$

So, the definition of "number of inversions" and $m > \ell$ implies that

$$\begin{aligned}
n(\sigma\circ\tau) &= \sum_{i=1}^{n} |\{(\sigma\circ\tau)(j) < (\sigma\circ\tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= \sum_{i=1}^{\ell} |\{(\sigma\circ\tau)(j) < (\sigma\circ\tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&\qquad\qquad + \sum_{i=\ell+1}^{n} |\{(\sigma\circ\tau)(j) < (\sigma\circ\tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= \sum_{i=\ell+1}^{n} |\{(\sigma\circ\tau)(j) < (\sigma\circ\tau)(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&\le \sum_{i=\ell+1}^{n} |\{\tau(j) < \tau(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \quad \text{as } \ m > \ell, \\
&< (m-\ell) + \sum_{i=\ell+1}^{n} |\{\tau(j) < \tau(i), \ \text{ for } \ j = i+1, i+2, \ldots, n\}| \\
&= n(\tau).
\end{aligned}$$

Thus, $n(\sigma\circ\tau) < k+1$. Hence, by the induction hypothesis, the permutation $\sigma\circ\tau$ is a composition of transpositions. That is, there exist transpositions, say $\alpha_i, \ 1 \le i \le t$ such that

$$\sigma\circ\tau = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t.$$

Hence, $\tau = \sigma\circ\alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t$ as $\sigma\circ\sigma = Id_n$ for any transposition $\sigma \in \mathcal{S}_n$. Therefore, by mathematical induction, the proof of the theorem is complete. $\square$

Before coming to our next important result, we state and prove the following lemma.

Lemma 7.1.9. *Suppose there exist transpositions $\alpha_i, \ 1 \le i \le t$ such that*

$$Id_n = \alpha_1 \circ \alpha_2 \circ \cdots \circ \alpha_t,$$

then t is even.

Proof. Observe that $t \neq 1$ as the identity permutation is not a transposition. Hence, $t \ge 2$. If $t = 2$, we are done. So, let us assume that $t \ge 3$. We will prove the result by the method of mathematical induction. The result clearly holds for $t = 2$. Let the result be true for all expressions in which the number of transpositions $t \le k$. Now, let $t = k+1$.

Suppose $\alpha_1 = (m \ r)$. Note that the possible choices for the composition $\alpha_1 \circ \alpha_2$ are $(m \ r)\circ(m \ r) = Id_n, \ (m \ r)\circ(m \ \ell) = (r \ \ell)\circ(r \ m), \ (m \ r)\circ(r \ \ell) = (\ell \ r)\circ(\ell \ m)$ and $(m \ r)\circ(\ell \ s) = (\ell \ s)\circ(m \ r)$, where ℓ and s are distinct elements of $\{1, 2, \ldots, n\}$ and are different from $m, \ r$. In the first case, we can remove $\alpha_1 \circ \alpha_2$ and obtain $Id_n = \alpha_3 \circ \alpha_4 \circ \cdots \circ \alpha_t$. In this expression for identity, the number of transpositions is $t-2 = k-1 < k$. So, by mathematical induction, $t-2$ is even and hence t is also even.

In the other three cases, we replace the original expression for $\alpha_1 \circ \alpha_2$ by their counterparts on the right to obtain another expression for identity in terms of $t = k+1$

transpositions. But note that in the new expression for identity, the positive integer m doesn't appear in the first transposition, but appears in the second transposition. We can continue the above process with the second and third transpositions. At this step, either the number of transpositions will reduce by 2 (giving us the result by mathematical induction) or the positive number m will get shifted to the third transposition. The continuation of this process will at some stage lead to an expression for identity in which the number of transpositions is $t - 2 = k - 1$ (which will give us the desired result by mathematical induction), or else we will have an expression in which the positive number m will get shifted to the right most transposition. In the later case, the positive integer m appears exactly once in the expression for identity and hence this expression does not fix m whereas for the identity permutation $Id_n(m) = m$. So the later case leads us to a contradiction.

Hence, the process will surely lead to an expression in which the number of transpositions at some stage is $t - 2 = k - 1$. Therefore, by mathematical induction, the proof of the lemma is complete.

□

Theorem 7.1.10. *Let $\alpha \in \mathcal{S}_n$. Suppose there exist transpositions $\tau_1, \tau_2, \ldots, \tau_k$ and $\sigma_1, \sigma_2, \ldots, \sigma_\ell$ such that*

$$\alpha = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_\ell$$

then either k and ℓ are both even or both odd.

Proof. Observe that the condition $\tau_1 \circ \tau_2 \circ \cdots \circ \tau_k = \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_\ell$ and $\sigma \circ \sigma = Id_n$ for any transposition $\sigma \in \mathcal{S}_n$, implies that

$$Id_n = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k \circ \sigma_\ell \circ \sigma_{\ell-1} \circ \cdots \circ \sigma_1.$$

Hence by Lemma 7.1.9, $k + \ell$ is even. Hence, either k and ℓ are both even or both odd. Thus the result follows. □

Definition 7.1.11. *A permutation $\sigma \in \mathcal{S}_n$ is called an even permutation if σ can be written as a composition (product) of an even number of transpositions. A permutation $\sigma \in \mathcal{S}_n$ is called an odd permutation if σ can be written as a composition (product) of an odd number of transpositions.*

Remark 7.1.12. *Observe that if σ and τ are both even or both odd permutations, then the permutations $\sigma \circ \tau$ and $\tau \circ \sigma$ are both even. Whereas if one of them is odd and the other even then the permutations $\sigma \circ \tau$ and $\tau \circ \sigma$ are both odd. We use this to define a function on $\mathcal{S}_n$, called the sign of a permutation, as follows:*

Definition 7.1.13. *Let $sgn : \mathcal{S}_n \longrightarrow \{1, -1\}$ be a function defined by*

$$sgn(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is an even permutation} \\ -1 & \text{if } \sigma \text{ is an odd permutation} \end{cases}.$$

Example 7.1.14. *1. The identity permutation, Id_n is an even permutation whereas every transposition is an odd permutation. Thus, $sgn(Id_n) = 1$ and for any transposition $\sigma \in \mathcal{S}_n$, $sgn(\sigma) = -1$.*

2. *Using Remark 7.1.12, $sgn(\sigma\circ\tau) = sgn(\sigma)\cdot sgn(\tau)$ for any two permutations $\sigma, \tau \in \mathcal{S}_n$.*

We are now ready to define determinant of a square matrix A.

Definition 7.1.15. *Let $A = [a_{ij}]$ be an $n\times n$ matrix with entries from $\mathbb{F}$. The determinant of A, denoted $\det(A)$, is defined as*

$$\det(A) = \sum_{\sigma\in\mathcal{S}_n} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} = \sum_{\sigma\in\mathcal{S}_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

Remark 7.1.16. 1. *Observe that $\det(A)$ is a scalar quantity. The expression for $\det(A)$ seems complicated at the first glance. But this expression is very helpful in proving the results related with "properties of determinant".*

2. *If $A = [a_{ij}]$ is a 3×3 matrix, then using (7.1.1),*

$$\begin{aligned}
\det(A) &= \sum_{\sigma\in\mathcal{S}_n} sgn(\sigma)\prod_{i=1}^{3} a_{i\sigma(i)} \\
&= sgn(\tau_1)\prod_{i=1}^{3} a_{i\tau_1(i)} + sgn(\tau_2)\prod_{i=1}^{3} a_{i\tau_2(i)} + sgn(\tau_3)\prod_{i=1}^{3} a_{i\tau_3(i)} + \\
&\qquad sgn(\tau_4)\prod_{i=1}^{3} a_{i\tau_4(i)} + sgn(\tau_5)\prod_{i=1}^{3} a_{i\tau_5(i)} + sgn(\tau_6)\prod_{i=1}^{3} a_{i\tau_6(i)} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.
\end{aligned}$$

Observe that this expression for $\det(A)$ for a 3×3 matrix A is same as that given in (2.5.1).

7.2 Properties of Determinant

Theorem 7.2.1 (Properties of Determinant)**.** *Let $A = [a_{ij}]$ be an $n\times n$ matrix. Then*

1. *if B is obtained from A by interchanging two rows, then* $\det(B) = -\det(A)$.

2. *if B is obtained from A by multiplying a row by c then* $\det(B) = c\det(A)$.

3. *if all the elements of one row is 0 then* $\det(A) = 0$.

4. *if A is a square matrix having two rows equal then* $\det(A) = 0$.

5. *Let $B = [b_{ij}]$ and $C = [c_{ij}]$ be two matrices which differ from the matrix $A = [a_{ij}]$ only in the m^{th} row for some m. If $c_{mj} = a_{mj} + b_{mj}$ for $1 \le j \le n$ then* $\det(C) = \det(A) + \det(B)$.

6. *if B is obtained from A by replacing the ℓth row by itself plus k times the mth row, for $\ell \ne m$ then* $\det(B) = \det(A)$.

7. *if A is a triangular matrix then* $\det(A) = a_{11}a_{22}\cdots a_{nn}$, *the product of the diagonal elements.*

8. *If E is an elementary matrix of order n then* $\det(EA) = \det(E)\det(A)$.

9. *A is invertible if and only if* $\det(A) \neq 0$.

10. *If B is an $n \times n$ matrix then* $\det(AB) = \det(A)\det(B)$.

11. $\det(A) = \det(A^t)$, *where recall that A^t is the transpose of the matrix A.*

Proof. **Proof of Part 1.** Suppose $B = [b_{ij}]$ is obtained from $A = [a_{ij}]$ by the interchange of the ℓ^{th} and m^{th} row. Then $b_{\ell j} = a_{mj},\ b_{mj} = a_{\ell j}$ for $1 \leq j \leq n$ and $b_{ij} = a_{ij}$ for $1 \leq i \neq \ell, m \leq n,\ 1 \leq j \leq n$.

Let $\tau = (\ell\ m)$ be a transposition. Then by Proposition 7.1.4, $\mathcal{S}_n = \{\sigma \circ \tau : \ \sigma \in \mathcal{S}_n\}$. Hence by the definition of determinant and Example 7.1.14.2, we have

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \sum_{\sigma\circ\tau \in \mathcal{S}_n} \operatorname{sgn}(\sigma \circ \tau) \prod_{i=1}^{n} b_{i(\sigma\circ\tau)(i)} \\
&= \sum_{\sigma\circ\tau \in \mathcal{S}_n} \operatorname{sgn}(\tau)\cdot\operatorname{sgn}(\sigma)\ b_{1(\sigma\circ\tau)(1)} b_{2(\sigma\circ\tau)(2)} \cdots b_{\ell(\sigma\circ\tau)(\ell)} \cdots b_{m(\sigma\circ\tau)(m)} \cdots b_{n(\sigma\circ\tau)(n)} \\
&= \operatorname{sgn}(\tau) \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\ b_{1\sigma(1)} \cdot b_{2\sigma(2)} \cdots b_{\ell\sigma(m)} \cdots b_{m\sigma(\ell)} \cdots b_{n\sigma(n)} \\
&= -\left(\sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\ a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{\ell\sigma(\ell)} \cdots a_{n\sigma(n)} \right) \quad \text{as } \operatorname{sgn}(\tau) = -1 \\
&= -\det(A).
\end{aligned}$$

Proof of Part 2. Suppose that $B = [b_{ij}]$ is obtained by multiplying the m^{th} row of A by $c \neq 0$. Then $b_{mj} = c\, a_{mj}$ and $b_{ij} = a_{ij}$ for $1 \leq i \neq m \leq n,\ 1 \leq j \leq n$. Then

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots c a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= c \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= c \det(A).
\end{aligned}$$

Proof of Part 3. Note that $\det(A) = \sum\limits_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}$. So, each term in the expression for determinant, contains one entry from each row. Hence, from the condition that A has a row consisting of all zeros, the value of each term is 0. Thus, $\det(A) = 0$.

Proof of Part 4. Suppose that the ℓ^{th} and m^{th} row of A are equal. Let B be the matrix obtained from A by interchanging the ℓ^{th} and m^{th} rows. Then by the first part,

$\det(B) = -\det(A)$. But the assumption implies that $B = A$. Hence, $\det(B) = \det(A)$. So, we have $\det(B) = -\det(A) = \det(A)$. Hence, $\det(A) = 0$.

Proof of Part 5. By definition and the given assumption, we have

$$\begin{aligned}
\det(C) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots c_{m\sigma(m)} \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) c_{1\sigma(1)} c_{2\sigma(2)} \cdots (b_{m\sigma(m)} + a_{m\sigma(m)}) \cdots c_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&\qquad + \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \det(B) + \det(A).
\end{aligned}$$

Proof of Part 6. Suppose that $B = [b_{ij}]$ is obtained from A by replacing the ℓth row by itself plus k times the mth row, for $\ell \neq m$. Then $b_{\ell j} = a_{\ell j} + k\ a_{mj}$ and $b_{ij} = a_{ij}$ for $1 \leq i \neq m \leq n$, $1 \leq j \leq n$. Then

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{\ell\sigma(\ell)} \cdots b_{m\sigma(m)} \cdots b_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots (a_{\ell\sigma(\ell)} + k a_{m\sigma(m)}) \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{\ell\sigma(\ell)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&\qquad + k \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{\ell\sigma(\ell)} \cdots a_{m\sigma(m)} \cdots a_{n\sigma(n)} \qquad \text{use Part } 4 \\
&= \det(A).
\end{aligned}$$

Proof of Part 7. First let us assume that A is an upper triangular matrix. Observe that if $\sigma \in \mathcal{S}_n$ is different from the identity permutation then $n(\sigma) \geq 1$. So, for every $\sigma \neq Id_n \in \mathcal{S}_n$, there exists a positive integer m, $1 \leq m \leq n-1$ (depending on σ) such that $m > \sigma(m)$. As A is an upper triangular matrix, $a_{m\sigma(m)} = 0$ for each $\sigma(\neq Id_n) \in \mathcal{S}_n$. Hence the result follows.

A similar reasoning holds true, in case A is a lower triangular matrix.

Proof of Part 8. Let I_n be the identity matrix of order n. Then using Part 7, $\det(I_n) = 1$. Also, recalling the notations for the elementary matrices given in Remark 2.2.2, we have $\det(E_{ij}) = -1$, (using Part 1) $\det(E_i(c)) = c$ (using Part 2) and $\det(E_{ij}(k) = 1$ (using Part 6). Again using Parts 1, 2 and 6, we get $\det(EA) = \det(E)\det(A)$.

Proof of Part 9. Suppose A is invertible. Then by Theorem 2.2.5, A is a product of elementary matrices. That is, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $A = E_1 E_2 \cdots E_k$. Now a repeated application of Part 8 implies that $\det(A) = \det(E_1)\det(E_2)\cdots\det(E_k)$. But $\det(E_i) \neq 0$ for $1 \leq i \leq k$. Hence, $\det(A) \neq 0$.

Now assume that $\det(A) \neq 0$. We show that A is invertible. On the contrary, assume that A is not invertible. Then by Theorem 2.2.5, the matrix A is not of full rank. That is there exists a positive integer $r < n$ such that $\text{rank}(A) = r$. So, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $E_1 E_2 \cdots E_k A = \begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}$. Therefore, by Part 3 and a repeated application of Part 8,

$$\det(E_1)\det(E_2)\cdots\det(E_k)\det(A) = \det(E_1 E_2 \cdots E_k A) = \det\left(\begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}\right) = 0.$$

But $\det(E_i) \neq 0$ for $1 \leq i \leq k$. Hence, $\det(A) = 0$. This contradicts our assumption that $\det(A) \neq 0$. Hence our assumption is false and therefore A is invertible.

Proof of Part 10. Suppose A is not invertible. Then by Part 9, $\det(A) = 0$. Also, the product matrix AB is also not invertible. So, again by Part 9, $\det(AB) = 0$. Thus, $\det(AB) = \det(A)\det(B)$.

Now suppose that A is invertible. Then by Theorem 2.2.5, A is a product of elementary matrices. That is, there exist elementary matrices $E_1, E_2, \ldots, E_k$ such that $A = E_1 E_2 \cdots E_k$. Now a repeated application of Part 8 implies that

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k B) = \det(E_1)\det(E_2)\cdots\det(E_k)\det(B) \\ &= \det(E_1 E_2 \cdots E_k)\det(B) = \det(A)\det(B). \end{aligned}$$

Proof of Part 11. Let $B = [b_{ij}] = A^t$. Then $b_{ij} = a_{ji}$ for $1 \leq i, j \leq n$. By Proposition 7.1.4, we know that $\mathcal{S}_n = \{\sigma^{-1} : \sigma \in \mathcal{S}_n\}$. Also $\text{sgn}(\sigma) = \text{sgn}(\sigma^{-1})$. Hence,

$$\begin{aligned} \det(B) &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} \\ &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma^{-1}) b_{\sigma^{-1}(1)\,1}\, b_{\sigma^{-1}(2)\,2} \cdots b_{\sigma^{-1}(n)\,n} \\ &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)} b_{2\sigma^{-1}(2)} \cdots b_{n\sigma^{-1}(n)} \\ &= \det(A). \end{aligned}$$

□

Remark 7.2.2. 1. *The result that* $\det(A) = \det(A^t)$ *implies that in the statements made in Theorem 7.2.1, where ever the word "row" appears it can be replaced by "column".*

2. *Let* $A = [a_{ij}]$ *be a matrix satisfying* $a_{11} = 1$ *and* $a_{1j} = 0$ *for* $2 \leq j \leq n$. *Let* B *be the submatrix of* A *obtained by removing the first row and the first column. Then it can be easily shown that* $\det(A) = \det(B)$. *The reason being is as follows:*
for every $\sigma \in \mathcal{S}_n$ *with* $\sigma(1) = 1$ *is equivalent to saying that* σ *is a permutation of the*

elements $\{2, 3, \ldots, n\}$. *That is,* $\sigma \in \mathcal{S}_{n-1}$. *Hence,*

$$\begin{aligned}\det(A) &= \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = \sum_{\sigma \in \mathcal{S}_n, \sigma(1)=1} sgn(\sigma) a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{\sigma \in \mathcal{S}_{n-1}} sgn(\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)} = \det(B).\end{aligned}$$

We are now ready to relate this definition of determinant with the one given in Definition 2.5.2.

Theorem 7.2.3. *Let A be an $n \times n$ matrix. Then* $\det(A) = \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1|j)\big)$, *where recall that $A(1|j)$ is the submatrix of A obtained by removing the 1^{st} row and the j^{th} column.*

Proof. For $1 \le j \le n$, define two matrices

$$B_j = \begin{bmatrix} 0 & 0 & \cdots & a_{1j} & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}_{n \times n} \quad \text{and} \quad C_j = \begin{bmatrix} a_{1j} & 0 & 0 & \cdots & 0 \\ a_{2j} & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{nj} & a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}_{n \times n}.$$

Then by Theorem 7.2.1.5,

$$\det(A) = \sum_{j=1}^{n} \det(B_j). \tag{7.2.2}$$

We now compute $\det(B_j)$ for $1 \le j \le n$. Note that the matrix B_j can be transformed into C_j by $j-1$ interchanges of columns done in the following manner:
first interchange the 1^{st} and 2^{nd} column, then interchange the 2^{nd} and 3^{rd} column and so on (the last process consists of interchanging the $(j-1)^{\text{th}}$ column with the j^{th} column. Then by Remark 7.2.2 and Parts 1 and 2 of Theorem 7.2.1, we have $\det(B_j) = a_{1j}(-1)^{j-1} \det(C_j)$. Therefore by (7.2.2),

$$\det(A) = \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det\big(A(1|j)\big) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det\big(A(1|j)\big).$$

□

7.3 Dimension of $M + N$

Theorem 7.3.1. *Let $V(\mathbb{F})$ be a finite dimensional vector space and let M and N be two subspaces of V. Then*

$$\dim(M) + \dim(N) = \dim(M + N) + \dim(M \cap N). \tag{7.3.3}$$

Proof. Since $M \cap N$ is a vector subspace of V, consider a basis $\mathcal{B}_1 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of $M \cap N$. As, $M \cap N$ is a subspace of the vector spaces M and N, we extend the basis $\mathcal{B}_1$ to form a basis $\mathcal{B}_M = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_r\}$ of M and also a basis $\mathcal{B}_N = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_s\}$ of N.

We now proceed to prove that the set $\mathcal{B}_2 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_s, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a basis of $M+N$.

To do this, we show that

1. the set $\mathcal{B}_2$ is linearly independent subset of V, and

2. $L(\mathcal{B}_2) = M+N$.

The second part can be easily verified. To prove the first part, we consider the linear system of equations

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta_1 \mathbf{w}_1 + \cdots + \beta_s \mathbf{w}_s + \gamma_1 \mathbf{v}_1 + \cdots + \gamma_r \mathbf{v}_r = \mathbf{0}. \tag{7.3.4}$$

This system can be rewritten as

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta_1 \mathbf{w}_1 + \cdots + \beta_s \mathbf{w}_s = -(\gamma_1 \mathbf{v}_1 + \cdots + \gamma_r \mathbf{v}_r).$$

The vector $\mathbf{v} = -(\gamma_1 \mathbf{v}_1 + \cdots + \gamma_r \mathbf{v}_r) \in M$, as $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathcal{B}_M$. But we also have $\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \beta_1 \mathbf{w}_1 + \cdots + \beta_s \mathbf{w}_s \in N$ as the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_s \in \mathcal{B}_N$. Hence, $\mathbf{v} \in M \cap N$ and therefore, there exists scalars $\delta_1, \ldots, \delta_k$ such that $\mathbf{v} = \delta_1 \mathbf{u}_1 + \delta_2 \mathbf{u}_2 + \cdots + \delta_k \mathbf{u}_k$.

Substituting this representation of $\mathbf{v}$ in Equation (7.3.4), we get

$$(\alpha_1 - \delta_1)\mathbf{u}_1 + \cdots + (\alpha_k - \delta_k)\mathbf{u}_k + \beta_1 \mathbf{w}_1 + \cdots + \beta_s \mathbf{w}_s = \mathbf{0}.$$

But then, the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{w}_1, \ldots, \mathbf{w}_s$ are linearly independent as they form a basis. Therefore, by the definition of linear independence, we get

$$\alpha_i - \delta_i = 0, \quad \text{for} \quad 1 \le i \le k \quad \text{and} \quad \beta_j = 0 \quad \text{for} \quad 1 \le j \le s.$$

Thus the linear system of Equations (7.3.4) reduces to

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k + \gamma_1 \mathbf{v}_1 + \cdots + \gamma_r \mathbf{v}_r = \mathbf{0}.$$

The only solution for this linear system is

$$\alpha_i = 0, \quad \text{for} \quad 1 \le i \le k \quad \text{and} \quad \gamma_j = 0 \quad \text{for} \quad 1 \le j \le r.$$

Thus we see that the linear system of Equations (7.3.4) has no non-zero solution. And therefore, the vectors are linearly independent.

Hence, the set $\mathcal{B}_2$ is a basis of $M+N$. We now count the vectors in the sets $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_M$ and $\mathcal{B}_N$ to get the required result. $\square$

Index