Lecture Notes on Linear Algebra

Arbind K Lal Sukant Pati

July 10, 2018

Contents

Chapter 1

Introduction to Matrices

1.1 Definition of a Matrix

Definition 1.1.1. 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**. Let A be a matrix having m rows and n columns. Then, A is said to have **order** $m \times n$ or is called a matrix of **size** $m \times n$ and can be represented in either of the following forms:

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

where a_{ij} is the entry at the intersection of the i^{th} row and j^{th} column. One writes $A \in \mathbb{M}_{m,n}(\mathbb{F})$ to mean that A is an $m \times n$ matrix with entries from the set $\mathbb{F}$, or in short $A = [a_{ij}]$. We write $A[i,:]$ to denote the i-th row of A, $A[:,j]$ to denote the j-th column of A and a_{ij} or $(A)_{ij}$, for the (i,j)-th entry of A.

For example, if $A = \begin{bmatrix} 1 & 3+\mathbf{i} & 7 \\ 4 & 5 & 6-5\mathbf{i} \end{bmatrix}$ then $A[1,:] = [1 \;\; 3+\mathbf{i} \;\; 7]$, $A[:,3] = \begin{bmatrix} 7 \\ 6-5\mathbf{i} \end{bmatrix}$ and $a_{22} = 5$. Sometimes commas are inserted to differentiate between entries of a row vector. Thus, $A[1,:]$ may also be written as $[1, \; 3+\mathbf{i}, \; 7]$. A matrix having only one column is called a **column vector** and a matrix with only one row is called a **row vector**. All our vectors will be column vectors and will be represented by bold letters.

Example 1.1.2. Consider a system of linear equations $2x + 5y = 7$ and $3x + 2y = 6$. Then, we identify it with the matrix $A = \left[\begin{array}{cc|c} 2 & 5 & 7 \\ 3 & 2 & 6 \end{array}\right]$. Here the variable/unknown x is associated with $A[:,1]$ and y is associated with $A[:,2]$.

Definition 1.1.3. Two matrices $A = [a_{ij}], B = [b_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$ are said to be **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.

1.1.1 Special Matrices

Definition 1.1.4. Let $A = [a_{ij}]$ be an $m \times n$ matrix with $a_{ij} \in \mathbb{F}$.

1. Then A is called a **zero-matrix**, denoted $\mathbf{0}$ (order is mostly clear from the context), if $a_{ij} = 0$ for all i and j. For example, $\mathbf{0}_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\mathbf{0}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

2. Then A is called a **square matrix** if $m = n$ and is denoted by $A \in \mathbb{M}_n(\mathbb{F})$.

3. Let $A \in \mathbb{M}_n(\mathbb{F})$.

   (a) Then, the entries $a_{11}, a_{22}, \ldots, a_{nn}$ are called the diagonal entries of A. They constitute the **principal diagonal** of A.

   (b) Then, A is said to be a **diagonal** matrix, , denoted $\text{diag}(a_{11}, \ldots, a_{nn})$, if $a_{ij} = 0$ for $i \neq j$. For example, the zero matrix $\mathbf{0}_n$ and $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ are diagonal matrices.

   (c) Then, $A = \text{diag}(1, \ldots, 1)$ is called the **identity matrix**, denoted I_n, or in short I. For example, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

   (d) If $A = \alpha I$, for some $\alpha \in \mathbb{F}$, then A is called a **scalar matrix**.

   (e) Then, A is said to be an **upper triangular** matrix if $a_{ij} = 0$ for $i > j$.

   (f) Then, A is said to be a **lower triangular** matrix if $a_{ij} = 0$ for $i < j$.

   (g) Then, 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 and the matrices $\mathbf{0}, I$ are upper as well as lower triangular matrices.

4. An $m \times n$ matrix $A = [a_{ij}]$ is said to have an **upper triangular form** if $a_{ij} = 0$ for all $i > j$. For example, the matrices $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$ have upper triangular forms.

5. For $1 \leq i \leq n$, define $\mathbf{e}_i = I_n[:, i]$, a matrix of order $n \times 1$. Then the column matrices $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are called the **standard unit vectors** or the **standard basis** of $\mathbb{M}_{n,1}(\mathbb{C})$ or $\mathbb{C}^n$. The dependence of n is omitted as it is understood from the context. For example, if $\mathbf{e}_1 \in \mathbb{C}^2$ then, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and if $\mathbf{e}_1 \in \mathbb{C}^3$ then $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

1.2 Operations on Matrices

Definition 1.2.1. Let $A = [a_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$. Then

1. the **transpose** of A, denoted A^T, is an $n \times m$ matrix with $(A^T)_{ij} = a_{ji}$, for all i, j.
2. the **conjugate transpose** of A, denoted A^*, is an $n \times m$ matrix with $(A^*)_{ij} = \overline{a_{ji}}$ (the complex-conjugate of a_{ji}), for all i, j.

If $A = \begin{bmatrix} 1 & 4+\mathbf{i} \\ 0 & 1-\mathbf{i} \end{bmatrix}$ then $A^T = \begin{bmatrix} 1 & 0 \\ 4+\mathbf{i} & 1-\mathbf{i} \end{bmatrix}$ and $A^* = \begin{bmatrix} 1 & 0 \\ 4-\mathbf{i} & 1+\mathbf{i} \end{bmatrix}$. Note that $A^* \neq A^T$.
Note that if $\mathbf{x}$ is a column vector then $\mathbf{x}^T$ and $\mathbf{x}^*$ are row vectors.

Theorem 1.2.2. *For any matrix A, $(A^*)^* = A$ and $(A^T)^T = A$.*

Proof. Let $A = [a_{ij}]$, $A^* = [b_{ij}]$ and $(A^*)^* = [c_{ij}]$. Clearly, the order of A and $(A^*)^*$ is the same. Also, by definition $c_{ij} = \overline{b_{ji}} = \overline{\overline{a_{ij}}} = a_{ij}$ for all i, j. ■

Definition 1.2.3. Let $A = [a_{ij}], B = [b_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$ and $k \in \mathbb{C}$.

1. . Then the **sum** of A and B, denoted $A + B$, is defined to be the matrix $C = [c_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$ with $c_{ij} = a_{ij} + b_{ij}$ for all i, j.
2. Then, the **product** of $k \in \mathbb{C}$ with A, denoted kA, equals $kA = [ka_{ij}] = [a_{ij}k] = Ak$.

   If $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -4 & 6 \\ 1 & 1 & 7 \end{bmatrix}$ then $A + B = \begin{bmatrix} 2 & 0 & 11 \\ 1 & 2 & 9 \end{bmatrix}$ and $5A = \begin{bmatrix} 5 & 20 & 25 \\ 0 & 5 & 10 \end{bmatrix}$.

Theorem 1.2.4. *Let $A, B, C \in \mathbb{M}_{m,n}(\mathbb{C})$ and let $k, \ell \in \mathbb{C}$. 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. (1). Let $A = [a_{ij}]$ and $B = [b_{ij}]$. Then by definition

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

as complex numbers commute. The other parts are left for the reader. ■

Definition 1.2.5. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then

1. the matrix $\mathbf{0}_{m\times n}$ satisfying $A + \mathbf{0} = \mathbf{0} + A = A$ is called the **additive identity**.
2. the matrix B with $A + B = \mathbf{0}$ is called the **additive inverse** of A, denoted $-A = (-1)A$.

EXERCISE **1.2.6.** *1. Find a few non zero, non-identity matrices A satisfying*

(a) $A^T = A$.

Ans: $A = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 3 & 5 \\ 2 & 5 & -1 \end{bmatrix} = A^T$

(b) $A^T = -A$.

Ans: $A = \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & 5 \\ -2 & -5 & 0 \end{bmatrix} = -A^T$.

2. *Find a few non zero, non-identity matrices A with complex entries satisfying*

(a) $A^* = A$.

(b) $A^* = -A$.

Ans: $A = \begin{bmatrix} 1 & -1+i & 2-i \\ -1-i & 3 & i \\ 2+i & -i & -1 \end{bmatrix} = A^*$, $A = \begin{bmatrix} 0 & -1+i & 2-i \\ 1+i & 0 & -i \\ -2-i & -i & 0 \end{bmatrix} = -A^*$.

3. *Suppose $A = [a_{ij}], B = [b_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$.*

(a) *If $A + B = \mathbf{0}$ then show that $B = (-1)A = [-a_{ij}]$.*

(b) *If $A + B = A$ then show that $B = \mathbf{0}$.*

---

4. *Let* $A = \begin{bmatrix} 1+\mathbf{i} & -1 \\ 2 & 3 \\ \mathbf{i} & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 1-\mathbf{i} & 2 \end{bmatrix}$. *Compute $A + B^*$ and $B + A^*$.*

5. *Write the 3×3 matrices $A = [a_{ij}]$ satisfying*

(a) *$a_{ij} = 1$ if $i \neq j$ and 2 otherwise.*

(b) *$a_{ij} = 1$ if $\mid i - j \mid \; \leq 1$ and 0 otherwise.*

(c) *$a_{ij} = i + j$.*

(d) *$a_{ij} = 2^{i+j}$.*

Ans: a) $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, b) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, c) $A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$, d) $A = \begin{bmatrix} 2^2 & 2^3 & 2^4 \\ 2^3 & 2^4 & 2^5 \\ 2^4 & 2^5 & 2^6 \end{bmatrix}$.

1.2.1 Multiplication of Matrices

Definition 1.2.7. Let $A = [a_{ij}] \in \mathbb{M}_{m,n}(\mathbb{C})$ and $B = [b_{ij}] \in \mathbb{M}_{n,r}(\mathbb{C})$. Then, the **product** of A and B, denoted AB, is a matrix $C = [c_{ij}] \in \mathbb{M}_{m,r}(\mathbb{C})$ with

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

Thus, AB is defined if and only if **the number of columns of A = the number of rows of B.** 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

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

Note that the rows of the matrix AB can be written directly as

$$\begin{aligned} (AB)[1,:] &= a\,[\alpha, \beta, \gamma, \delta] + b\,[x, y, z, t] + c\,[u, v, w, s] = aB[1,:] + bB[2,:] + cB[3,:] \\ &= a_{11}B[1,:] + a_{12}B[2,:] + a_{13}B[3,:] = \sum_{i=1}^{3} a_{1i}B[i,:] \quad (1.2.2) \\ (AB)[2,:] &= dB[1,:] + eB[2,:] + fB[3,:] = a_{21}B[1,:] + a_{22}B[2,:] + a_{23}B[3,:] \\ &= \sum_{i=1}^{3} a_{2i}B[i,:] \quad (1.2.3) \end{aligned}$$

and similarly, the columns of the matrix AB can be written directly as

$$(AB)[:,1] = \begin{bmatrix} a\alpha + bx + cu \\ d\alpha + ex + fu \end{bmatrix} = \alpha\ A[:,1] + x\ A[:,2] + u\ A[:,3] = \sum_{j=1}^{3} A[:,j]\ b_{j1}, \quad (1.2.4)$$

$(AB)[:,k] = \sum\limits_{j=1}^{3} A[:,j]\ b_{jk}$ for $k = 2, 3, 4$.

Remark 1.2.8. *Observe the following:*

1. *In the above 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 (adding or subtracting certain multiples) on the* rows *of B (see Equation (1.2.3)). This is called the* **row method** *for calculating the matrix product.*

3. *The product AB also corresponds to operating (adding or subtracting certain multiples) on the* columns *of A (see Equation (1.2.4)). This is called the* **column method** *for calculating the matrix product.*

4. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$ and $B \in \mathbb{M}_{n,p}(\mathbb{C})$. Then $(AB)[i,:] = A[i,:]B = a_{i1}B[1,:] + \cdots + a_{in}B[n,:]$ and $(AB)[:,j] = AB[:,j] = A[:,1]b_{1j} + \cdots + A[:,n]b_{nj}$.*

Example 1.2.9. 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

1. to find the second row of AB.
   Solution: $(AB)[2,:] = A[2,:]B = 1 \cdot [1, 0, -1] + 0 \cdot [0, 0, 1] + 1 \cdot [0, -1, 1] = [1, -1, 0]$.

2. to find the third column of AB.

   Solution: $(AB)[:,3] = A\,B[:,3] = -1 \cdot \begin{bmatrix}1\\1\\0\end{bmatrix} + 1 \cdot \begin{bmatrix}2\\0\\-1\end{bmatrix} + 1 \cdot \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\0\\0\end{bmatrix}$.

EXERCISE **1.2.10.** *1. Let $A \in \mathbb{M}_n(\mathbb{C})$, $D =$ diag$(d_1, d_2, \ldots, d_n)$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{M}_{n,1}(\mathbb{C})$ (see Definition 5). Then verify that*

(a) $A\mathbf{e}_1 = A[:,1], \ldots, A\mathbf{e}_n = A[:,n]$.

(b) $\mathbf{e}_1^T A = \mathbf{e}_1^* A = A[1,:], \ldots, \mathbf{e}_n^T A = \mathbf{e}_n^* A = A[n,:]$.

(c) $(DA)[i,:] = d_i\, A[i,:]$, *for* $1 \le i \le n$, *and*

(d) $(AD)[:,j] = d_j A[:,j]$, *for* $1 \le j \le n$. *In particular, if* $D = \alpha I$ *is a scalar matrix then* $DA = \alpha A = AD$.

Ans: Just use matrix multiplication to get the required results.

2. Let $\mathbf{x} = \begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}, \mathbf{y} = \begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix} \in \mathbb{M}_{n,1}(\mathbb{C})$. *Then* $\mathbf{y}^*\mathbf{x} = \sum\limits_{i=1}^{n} \overline{y_i} x_i$, $\mathbf{x}^*\mathbf{x} = \sum\limits_{i=1}^{n} |x_i|^2$,

$$\mathbf{x}\mathbf{y}^* = \begin{bmatrix} x_1\overline{y_1} & x_1\overline{y_2} & \cdots & x_1\overline{y_n} \\ \vdots & \ddots & \cdots & \vdots \\ x_n\overline{y_1} & x_n\overline{y_2} & \cdots & x_n\overline{y_n} \end{bmatrix} \text{ and } \mathbf{x}\mathbf{x}^* = \begin{bmatrix} |x_1|^2 & x_1\overline{x_2} & \cdots & x_1\overline{x_n} \\ x_2\overline{x_1} & |x_2|^2 & \cdots & x_2\overline{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_n\overline{x_1} & x_n\overline{x_2} & \cdots & |x_n|^2 \end{bmatrix}.$$

Ans: Just use matrix multiplication to get the required results.

3. 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.

Ans: Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$ be an upper triangular matrix. Then $(A^*A)_{11} = |a_{11}|^2$ and $(AA^*)_{11} = |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2$. Thus, $A^*A = AA^*$ implies $|a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2 = |a_{11}|^2$. Hence, $a_{12} = 0, \ldots, a_{1n} = 0$. Now, use $(A^*A)_{22} = (AA^*)_{22}$ to conclude $a_{23} = 0, \ldots, a_{2n} = 0$ and so on.

Definition 1.2.11. 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, verify that* $AB = \begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix} \ne \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} = BA$.

Theorem 1.2.13. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C}), B \in \mathbb{M}_{n,p}(\mathbb{C})$ *and* $C \in \mathbb{M}_{p,q}(\mathbb{C})$.

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

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

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

4. *If* $A \in \mathbb{M}_n(\mathbb{C})$ *then* $AI_n = I_nA = A$.

Proof. (1). Verify that $(BC)_{kj} = \sum\limits_{\ell=1}^{p} b_{k\ell}c_{\ell j}$ and $(AB)_{i\ell} = \sum\limits_{k=1}^{n} a_{ik}b_{k\ell}$. Therefore,

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

Using a similar argument, the next part follows. The other parts are left for the reader. ∎

EXERCISE **1.2.14.** 1. *Let* $L_1, L_2 \in \mathbb{M}_n(\mathbb{C})$ *be lower triangular matrices and* $U_1, U_2 \in \mathbb{M}_n(\mathbb{C})$ *be upper triangular matrices. If* $D \in \mathbb{M}_n(\mathbb{C})$ *is a diagonal matrix then*

   (a) L_1L_2 *is a lower triangular matrix.*

   (b) U_1U_2 *is an upper triangular matrix.*

   (c) DL_1 *and* L_1D *are lower triangular matrices.*

   (d) DU_1 *and* U_1D *are upper triangular matrices.*

   Ans: Just use matrix multiplication to get the required results.

2. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$. *If* $A\mathbf{x} = \mathbf{0}$ *for all* $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{C})$ *then* $A = \mathbf{0}$, *the zero matrix.*

   Ans: Take $\mathbf{x} = \mathbf{e}_i$. Then $\mathbf{0} = A\mathbf{x} = A\mathbf{e}_i = A[:,i]$. Hence the i-th column of A is the zero vector. Thus, as we vary i in $\{1, 2, \ldots, n\}$, we see that all the columns of A are zero.

3. *Let* $A, B \in \mathbb{M}_{m,n}(\mathbb{C})$. *If* $A\mathbf{x} = B\mathbf{x}$, *for all* $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{C})$ *then prove that* $A = B$.

   Ans: Take $C = A - B$. Now use (2) above to show that $C = \mathbf{0}$ and conclude that $A = B$.

4. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$ *and* $B \in \mathbb{M}_{n,p}(\mathbb{C})$.

   (a) *Prove that* $(AB)^* = B^*A^*$.

   Ans: By definition $(AB)^* = \overline{(AB)^T} = \overline{B^TA^T} = \overline{B^T}\ \overline{A^T} = B^*A^*$.

   (b) *If* $A[1,:] = \mathbf{0}^T$ *then* $(AB)[1,:] = \mathbf{0}^T$.

   Ans: By definition $(AB)[1,:] = A[1,:]B = \mathbf{0}^TB = \mathbf{0}^T$.

   (c) *If* $B[:,1] = \mathbf{0}$ *then* $(AB)[:,1] = \mathbf{0}$.

   Ans: By definition $(AB)[:,1] = AB[:,1] = A\mathbf{0} = \mathbf{0}$.

   (d) *If* $A[i,:] = A[j,:]$ *for some* i *and* j *then* $(AB)[i,:] = (AB)[j,:]$.

   Ans: By definition $(AB)[i,:] = A[i,:]B = A[j,:]B = (AB)[j,:]$.

   (e) *If* $B[:,i] = B[:,j]$ *for some* i *and* j *then* $(AB)[:,i] = (AB)[:,j]$.

   Ans: By definition $(AB)[:,i] = AB[:,i] = AB[:,j] = (AB)[:,j]$.

5. *Construct matrices A and B that satisfy the following statements.*

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

   Ans: Let A be a 2×3 matrix and B be a 3×1 matrix.

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

   Ans: Let A be a 2×3 matrix and B be a 3×2 matrix.

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

   Ans: Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$. Then $AB = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ whereas $BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

   (d) *Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Guess a formula for A^n and B^n and prove it?*

   Ans: $A^n = \mathbf{0}$ for $n \geq 2$ and $B^n = \mathbf{0}$ for $n \geq 3$.

   (e) *Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Is it true that $A^2 - 2A + I = \mathbf{0}$? What is $B^3 - 3B^2 + 3B - I$? Is $C^2 = 3C$?*

   Ans: Yes, all the three statements are TRUE.

---

6. *Let A and B be two $m \times n$ matrices. Then, prove that $(A+B)^* = A^* + B^*$.*

   Ans: $(A+B)^* = \overline{(A+B)^T} = \overline{A^T + B^T} = \overline{A^T} + \overline{B^T} = A^* + B^*$.

7. *Find $A \in \mathbb{M}_2(\mathbb{C})$ such that $A \neq \mathbf{0}$ but $A^2 = \mathbf{0}$.*

   Ans: See Exercise 5d.

8. *Find $A \in \mathbb{M}_2(\mathbb{C})$ such that $A \neq \mathbf{0}, I_2$ but $A^2 = A$.*

   Ans: Let $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

9. *Find $A, B, C \in \mathbb{M}_2(\mathbb{C})$ such that $AB = AC$ but $B \neq C$ (cancellation law **doesn't hold**).*

   Ans: Let $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & -3 \\ 2 & -3 \end{bmatrix}$. Then $AB = \mathbf{0} = AC$.

10. *Let $S = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$ and $T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Determine all $m, n \in \mathbb{N}$ such that $S^m = I$ and $T^n = I$.*

    Ans: Verify $S^{6m} = I$ and $T^{4m} = I$ for all $m \in \mathbb{N}$.

11. *Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Compute A^2 and A^3. Is $A^3 = I$? Determine $aA^3 + bA + cA^2$.*

Ans: $A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $A^3 = I$. So, $aA^3 + bA + cA^2 = \begin{bmatrix} a & b & c \\ c & b & a \\ a & b & c \end{bmatrix}$. Such matrices are called circulant matrices.

12. *Let* $A = \begin{bmatrix} 1 & 1+\mathbf{i} & -2 \\ 1 & -2 & \mathbf{i} \\ -\mathbf{i} & 1 & 1 \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1+\mathbf{i} & 1 \end{bmatrix}$. *Compute*

    (a) $A - A^*$, $A + A^*$, $(3AB)^* - 4B^*A$ *and* $3A - 2A^*$.

    (b) $(AB)[1,:]$, $(AB)[3,:]$, $(AB)[:,1]$ *and* $(AB)[:,2]$.

    (c) $(B^*A^*)[:,1]$, $(B^*A^*)[:,3]$, $(B^*A^*)[1,:]$ *and* $(B^*A^*)[2,:]$.

1.2.2 Inverse of a Matrix

Definition 1.2.15. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then

1. $B \in \mathbb{M}_n(\mathbb{C})$ is said to be a **left inverse** of A if $BA = I_n$.

2. $C \in \mathbb{M}_n(\mathbb{C})$ is called a **right inverse** of A if $AC = I_n$.

3. A is **invertible** (has an **inverse**) if there exists $B \in \mathbb{M}_n(\mathbb{C})$ such that $AB = BA = I_n$.

Lemma 1.2.16. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If there exist $B, C \in \mathbb{M}_n(\mathbb{C})$ such that $AB = I_n$ and $CA = I_n$ then $B = C$, i.e., If A has a left inverse and a right inverse then they are equal.*

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

Remark 1.2.17. *Lemma 1.2.16 implies that whenever A is invertible, the inverse is unique. Thus, we denote the inverse of A 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) In particular, the inverse of $\begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix}$ equals $\frac{1}{2}\begin{bmatrix} 7 & -3 \\ -4 & 2 \end{bmatrix}$.

    (c) If $ad - bc = 0$ then prove that either $A[1,:] = \mathbf{0}^*$ or $A[:,1] = \mathbf{0}$ or $A[2,:] = \alpha A[1,:]$ or $A[:,2] = \alpha A[:,1]$ for some $\alpha \in \mathbb{C}$. Hence, prove that A is not invertible.

    (d) 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. Justify your answer.

2. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix}$. Then $A^{-1} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$ (verify $AA^{-1} = A^{-1}A = I_3$).

3. Prove that the matrices $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ are not invertible.

   Solution: Suppose there exists C such that $CA = AC = I$. Then, using matrix product

   $$A[1,:]C = (AC)[1,:] = I[1,:] = [1,0,0] \text{ and } A[2,:]C = (AC)[2,:] = I[2,:] = [0,1,0].$$

   But $A[1,:] = A[2,:]$ and thus $[1,0,0] = [0,1,0]$, a contradiction.

   Similarly, if there exists D such that $BD = DB = I$ then

   $$DB[:,1] = (DB)[:,1] = I[:,1], \; DB[:,2] = (DB)[:,2] = I[:,2] \text{ and } DB[:,3] = I[:,3].$$

   But $B[:,3] = B[:,1] + B[:,2]$ and hence $I[:,3] = I[:,1] + I[:,2]$, a contradiction.

Theorem 1.2.19. *Let A and B be two invertible matrices. Then,*

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

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

3. $(A^*)^{-1} = (A^{-1})^*$.

Proof. (1). Let $B = A^{-1}$. Then $AB = BA = I$. Thus, by definition, B is invertible and $B^{-1} = A$. Or equivalently, $(A^{-1})^{-1} = A$.

(2). By associativity $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = I = (B^{-1}A^{-1})(AB)$.

(3). As $AA^{-1} = A^{-1}A = I$, we get $(AA^{-1})^* = (A^{-1}A)^* = I^*$. Or equivalently, $(A^{-1})^*A^* = A^*(A^{-1})^* = I$. Thus, by definition $(A^*)^{-1} = (A^{-1})^*$. ∎

We will again come back to the study of invertible matrices in Sections 2.2 and 2.5.1.

EXERCISE **1.2.20.**

1. *If A is an invertible matrix then $(A^{-1})^r = A^{-r}$, for all $r \in \mathbb{N}$.*

2. *If $A_1, \ldots, A_r$ are invertible matrices then $B = A_1A_2 \cdots A_r$ is also invertible.*

   Ans: Use Theorem 1.2.19.2 repeatedly.

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

   Ans: If $A = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$ then $A^{-1} = A$ and if $B = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ then $B^{-1} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$.

4. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be an invertible matrix. Then*

   (a) $A[i,:] \neq \mathbf{0}^T$*, for any i.*

   (b) $A[:,j] \neq \mathbf{0}$*, for any j.*

   (c) $A[i,:] \neq A[j,:]$*, for any i and j.*

(d) $A[:,i] \neq A[:,j]$, *for any* i *and* j.

(e) $A[3,:] \neq \alpha A[1,:] + \beta A[2,:]$, *for any* $\alpha, \beta \in \mathbb{C}$, *whenever* $n \geq 3$.

(f) $A[:,3] \neq \alpha A[:,1] + \beta A[:,2]$, *for any* $\alpha, \beta \in \mathbb{C}$, *whenever* $n \geq 3$.

Ans: As A is invertible, there exists $B \in \mathbb{M}_n(\mathbb{C})$ such that $AB = BA = I_n$. Therefore,

(a) if $A[i,:] = \mathbf{0}^T$ then $\mathbf{e}_i^T = I_n[i,:] = (AB)[i,:] = A[i,:]B = \mathbf{0}^T B = \mathbf{0}^T$.

(b) if $A[:,j] = \mathbf{0}$ then $\mathbf{e}_j = I_n[:,j] = (BA)[:,j] = BA[:,j] = B\mathbf{0} = \mathbf{0}$.

(c) if $A[i,:] = A[j,:]$ then
$\mathbf{e}_i^T = I_n[i,:] = (AB)[i,:] = A[i,:]B = A[j,:]B = (AB)[j,:] = I_n[j,:] = \mathbf{e}_j^T$.

(d) if $A[:,i] = A[:,j]$ then
$\mathbf{e}_i = I_n[:,i] = (BA)[:,i] = BA[:,i] = BA[:,j] = (BA)[:,j] = I_n[:,j] = \mathbf{e}_j$.

(e) if $A[3,:] = \alpha A[1,:] + \beta A[2,:]$ then

$$\begin{aligned}
\mathbf{e}_3^T &= I_n[3,:] = (AB)[3,:] = A[3,:]B = (\alpha A[1,:] + \beta A[2,:])\,B \\
&= \alpha A[1,:]B + \beta A[2,:]B = \alpha(AB)[1,:] + \beta(AB)[2,:] \\
&= \alpha I_n[1,:] + \beta I_n[2,:] = \alpha \mathbf{e}_1^T + \beta \mathbf{e}_2^T.
\end{aligned}$$

(f) if $A[:,3] = \alpha A[:,1] + \beta A[:,2]$ then

$$\begin{aligned}
\mathbf{e}_3 &= I_n[:,3] = (BA)[:,3] = BA[:,3] = B\,(\alpha A[:,1] + \beta A[:,2]) \\
&= \alpha BA[:,1] + \beta BA[:,2] = \alpha(BA)[:,1] + \beta(BA)[:,2] \\
&= \alpha I_n[:,1] + \beta I_n[:,2] = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2.
\end{aligned}$$

---

5. *Determine* A *that satisfies* $(I + 3A)^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

Ans: $A = \dfrac{-1}{9}\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$ as $(I + 3A) = \left((I + 3A)^{-1}\right)^{-1} = \left(\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\right)^{-1} = \dfrac{-1}{3}\begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix}$.

6. *Determine* A *that satisfies* $(I - A)^{-1} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$. *[See Example 1.2.18.2].*

Ans: Example 1.2.18.2 gives $I - A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix} \Rightarrow A = I - \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 0 & -2 & -3 \\ -2 & -2 & -4 \\ -3 & -4 & -5 \end{bmatrix}$.

7. *Let* A *be an invertible matrix satisfying* $A^3 + A - 2I = \mathbf{0}$. *Then* $A^{-1} = \dfrac{1}{2}\left(A^2 + I\right)$.

Ans: As A is invertible, multiplying by A^{-1} gives $A^2 + I - 2A^{-1} = \mathbf{0}$. Hence, the result.

8. *Let* $A = [a_{ij}]$ *be an invertible matrix and* $B = [p^{i-j} a_{ij}]$, *for some* $p \in \mathbb{C}$, $p \neq 0$. *Then* $B^{-1} = [p^{i-j} (A^{-1})_{ij}]$.

Ans: Note that $B = DAD^{-1}$, where $D =$ diag$(p, p^2, \ldots, p^n)$ is a diagonal matrix. As $p \neq 0$, D is invertible. Hence B^{-1} is invertible and $B^{-1} = (DAD^{-1})^{-1} = DA^{-1}D^{-1}$.

1.3 Some More Special Matrices

Definition 1.3.1. 1. For $1 \le k \le m$ and $1 \le \ell \le n$, define $\mathbf{e}_{k\ell} \in \mathbb{M}_{m,n}(\mathbb{C})$ by

$$(\mathbf{e}_{k\ell})_{ij} = \begin{cases} 1, & \text{if } (k,\ell) = (i,j) \\ 0, & \text{otherwise.} \end{cases}$$

Then, the matrices $\mathbf{e}_{k\ell}$ for $1 \le k \le m$ and $1 \le \ell \le n$ are called the **standard basis** elements for $\mathbb{M}_{m,n}(\mathbb{C})$.

So, if $\mathbf{e}_{k\ell} \in \mathbb{M}_{2,3}(\mathbb{C})$ then $\mathbf{e}_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\mathbf{e}_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ and $\mathbf{e}_{22} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$.

In particular, if $\mathbf{e}_{ij} \in \mathbb{M}_n(\mathbb{C})$ then $\mathbf{e}_{ij} = \mathbf{e}_i \mathbf{e}_j^T = \mathbf{e}_i \mathbf{e}_j^*$, for $1 \le i, j \le n$.

2. Let $A \in \mathbb{M}_n(\mathbb{R})$. Then

   (a) A is called **symmetric** if $A^T = A$. For example, $A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$.

   (b) A is called **skew-symmetric** if $A^T = -A$. For example, $A = \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix}$.

   (c) A is called **orthogonal** if $AA^T = A^TA = I$. For example, $A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

   (d) A is said to be a **permutation matrix** if A has exactly one non-zero entry, namely 1, in each row and column. For example, I_n for each positive integer n, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ are permutation matrices. Verify that permutation matrices are Orthogonal matrices.

3. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then

   (a) A is called **normal** if $A^*A = AA^*$. For example, $\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ is a normal matrix.

   (b) A is called **Hermitian** if $A^* = A$. For example, $A = \begin{bmatrix} 1 & 1+\mathbf{i} \\ 1-\mathbf{i} & 2 \end{bmatrix}$.

   (c) A is called **skew-Hermitian** if $A^* = -A$. For example, $A = \begin{bmatrix} 0 & 1+\mathbf{i} \\ -1+\mathbf{i} & 0 \end{bmatrix}$.

   (d) A is called **unitary** if $AA^* = A^*A = I$. For example, $A = \frac{1}{\sqrt{3}} \begin{bmatrix} 1+\mathbf{i} & 1 \\ -1 & 1-\mathbf{i} \end{bmatrix}$.

   Verify that Hermitian, skew-Hermitian and Unitary matrices are normal matrices.

4. A vector $\mathbf{u} \in \mathbb{M}_{n,1}(\mathbb{C})$ such that $\mathbf{u}^*\mathbf{u} = 1$ is called a **unit vector**.

5. A matrix A is called **idempotent** if $A^2 = A$. For example, $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ is idempotent.

6. An idempotent matrix which is also Hermitian is called a **projection** matrix. For example, if $\mathbf{u} \in \mathbb{M}_{n,1}(\mathbb{C})$ is a unit vector then $A = \mathbf{u}\mathbf{u}^*$ is a Hermitian, idempotent matrix. Thus A is a projection matrix.

   Verify that $\mathbf{u}^*(\mathbf{x} - A\mathbf{x}) = \mathbf{u}^*\mathbf{x} - \mathbf{u}^*A\mathbf{x} = \mathbf{u}^*\mathbf{x} - \mathbf{u}^*(\mathbf{u}\mathbf{u}^*)\mathbf{x} = 0$ (as $\mathbf{u}^*\mathbf{u} = 1$), for any $\mathbf{x} \in \mathbb{C}^3$. Thus, with respect to the dot product in $\mathbb{R}^3$, $A\mathbf{x}$ is the foot of the perpendicular from the point $\mathbf{x}$ on the vector $\mathbf{u}$. In particular, if $\mathbf{u} = \frac{1}{\sqrt{6}}[1, 2, -1]^T$ and $A = \mathbf{u}\mathbf{u}^T$. Then, for any vector $\mathbf{x} = [x_1, x_2, x_3]^T \in \mathbb{M}_{3,1}(\mathbb{R})$,

$$A\mathbf{x} = (\mathbf{u}\mathbf{u}^T)\mathbf{x} = \mathbf{u}(\mathbf{u}^T\mathbf{x}) = \frac{x_1 + 2x_2 - x_3}{\sqrt{6}}\mathbf{u} = \frac{x_1 + 2x_2 - x_3}{6}[1, 2, -1]^T.$$

7. Fix a unit vector $\mathbf{a} \in \mathbb{M}_{n,1}(\mathbb{R})$ and let $A = 2\mathbf{a}\mathbf{a}^T - I_n$. Then, verify that $A \in \mathbb{M}_n(\mathbb{R})$ and $A\mathbf{y} = 2(\mathbf{a}^T\mathbf{y})\mathbf{a} - \mathbf{y}$, for all $\mathbf{y} \in \mathbb{R}^n$. This matrix is called the **reflection** matrix about the line containing the points $\mathbf{0}$ and $\mathbf{a}$.

8. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is said to be **nilpotent** if there exists a positive integer n such that $A^n = \mathbf{0}$. The least positive integer k for which $A^k = \mathbf{0}$ is called the **order of nilpotency**. For example, if $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$ 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 $1 \le \ell \le n - 1$.

EXERCISE **1.3.2.** *1. Consider the matrices $\mathbf{e}_{ij} \in \mathbb{M}_n(\mathbb{C})$ for $1 \le i, j, \le n$. Is $\mathbf{e}_{12}\mathbf{e}_{11} = \mathbf{e}_{11}\mathbf{e}_{12}$? What about $\mathbf{e}_{12}\mathbf{e}_{22}$ and $\mathbf{e}_{22}\mathbf{e}_{12}$?*

Ans: Note $\mathbf{e}_{11} = \mathbf{e}_1\mathbf{e}_1^T$ and $\mathbf{e}_{12} = \mathbf{e}_1\mathbf{e}_2^T$. Thus $\mathbf{e}_{12}\mathbf{e}_{11} = (\mathbf{e}_1\mathbf{e}_2^T)(\mathbf{e}_1\mathbf{e}_1^T) = \mathbf{e}_1(\mathbf{e}_2^T\mathbf{e}_1)\mathbf{e}_1^T = \mathbf{0}$ as $\mathbf{e}_2^T\mathbf{e}_1 = 0$. Where as $\mathbf{e}_{11}\mathbf{e}_{12} = (\mathbf{e}_1\mathbf{e}_1^T)(\mathbf{e}_1\mathbf{e}_2^T) = \mathbf{e}_1(\mathbf{e}_1^T\mathbf{e}_1)\mathbf{e}_2^T = \mathbf{e}_1\mathbf{e}_2^T = \mathbf{e}_{12}$.

2. Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be three vectors in $\mathbb{R}^3$ such that $\mathbf{u}_i^*\mathbf{u}_i = 1$, for $1 \le i \le 3$, and $\mathbf{u}_i^*\mathbf{u}_j = 0$ whenever $i \neq j$. Prove the following.

(a) If $U = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$ then $U^*U = I$. What about $UU^* = \mathbf{u}_1\mathbf{u}_1^* + \mathbf{u}_2\mathbf{u}_2^* + \mathbf{u}_3\mathbf{u}_3^*$?

Ans: $U^*U = \begin{bmatrix} \mathbf{u}_1^* \\ \mathbf{u}_2^* \\ \mathbf{u}_3^* \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^*\mathbf{u}_1 & \mathbf{u}_1^*\mathbf{u}_2 & \mathbf{u}_1^*\mathbf{u}_3 \\ \mathbf{u}_2^*\mathbf{u}_1 & \mathbf{u}_2^*\mathbf{u}_2 & \mathbf{u}_2^*\mathbf{u}_3 \\ \mathbf{u}_3^*\mathbf{u}_1 & \mathbf{u}_3^*\mathbf{u}_2 & \mathbf{u}_3^*\mathbf{u}_3 \end{bmatrix} = I_3.$
Check $(UU^*)^2 = U(U^*U)U^* = UU^*$ and UU^* is Hermitian. So, UU^* is a projection matrix. It will be shown later that $UU^* = I_3$.

(b) If $A = \mathbf{u}_i\mathbf{u}_i^*$, for $1 \le i \le 3$ then $A^2 = A$. Is A Hermitian? Is A a projection matrix?

Ans: $A^2 = (\mathbf{u}_i\mathbf{u}_i^*)(\mathbf{u}_i\mathbf{u}_i^*) = \mathbf{u}_i(\mathbf{u}_i^*\mathbf{u}_i)\mathbf{u}_i^* = \mathbf{u}_i\mathbf{u}_i^* = A$. Clearly, A is Hermitian. Thus, A is a projection.

(c) If $A = \mathbf{u}_i\mathbf{u}_i^* + \mathbf{u}_j\mathbf{u}_j^*$, for $i \neq j$ then $A^2 = A$. Is A a projection matrix?

Ans: $A^2 = (\mathbf{u}_i\mathbf{u}_i^* + \mathbf{u}_j\mathbf{u}_j^*)(\mathbf{u}_i\mathbf{u}_i^* + \mathbf{u}_j\mathbf{u}_j^*) = \mathbf{u}_i\mathbf{u}_i^* + \mathbf{u}_j\mathbf{u}_j^* = A$ as $\mathbf{u}_i^*\mathbf{u}_j = 0 = \mathbf{u}_j^*\mathbf{u}_i$. Clearly, A is Hermitian. So, A is a projection matrix.

3. *Let $A, B \in \mathbb{M}_n(\mathbb{C})$ be two unitary matrices. Then both AB and BA are unitary matrices.*

4. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix.*

   (a) *Then the diagonal entries of A are necessarily real numbers.*

   Ans: Note that $a_{ii} = \mathbf{e}_i^* A\mathbf{e}_i = \mathbf{e}_i^* A^* \mathbf{e}_i = (\mathbf{e}_i^* A\mathbf{e}_i)^* = \overline{a_{ii}}$. Thus $a_{ii} = \overline{a_{ii}} \Rightarrow a_{ii} \in \mathbb{R}$.

   (b) *For each $B \in \mathbb{M}_n(\mathbb{C})$ prove that B^*AB is a Hermitian matrix.*

   Ans: $(B^*AB)^* = B^*A^*B = B^*AB$.

   (c) *Further if $A^2 = \mathbf{0}$ then show that $A = \mathbf{0}$.*

   Ans: $\mathbf{0} = A^2 = A^*A$. So, $0 = (A^*A)_{11} = |a_{11}|^2 + |a_{21}|^2 + \cdots + |a_{n1}|^2$ implies $a_{i1} = 0$ for $1 \le i \le n$. Similarly, use $0 = (A^*A)_{ii}$ for $i \ge 2$ to get other entries as zero.

   (d) *Then $\mathbf{x}^*A\mathbf{x}$ is a real number, for any $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{C})$.*

   Ans: As $\mathbf{x}^*A\mathbf{x}$ is a scalar, $\overline{\mathbf{x}^*A\mathbf{x}} = (\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A^*\mathbf{x} = \mathbf{x}^*A\mathbf{x} \Rightarrow \mathbf{x}^*A\mathbf{x} \in \mathbb{R}$.

5. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ for every $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{C})$ then A is a Hermitian matrix. [Hint: Use $\mathbf{e}_j, \mathbf{e}_j + \mathbf{e}_k$ and $\mathbf{e}_j + \mathbf{i}\mathbf{e}_k$ of $\mathbb{M}_{n,1}(\mathbb{C})$ for $\mathbf{x}$.]*

   Ans: Taking $\mathbf{x} = \mathbf{e}_i$ gives $a_{ii} = \mathbf{e}_i^* A\mathbf{e}_i = \mathbf{x}^*A\mathbf{x} \in \mathbb{R}$. So, $a_{ii} \in \mathbb{R}$.

   Taking $\mathbf{x} = \mathbf{e}_i + \mathbf{i}\mathbf{e}_j$, gives $\mathbf{x}^*A\mathbf{x} = a_{ii} - \mathbf{i}a_{ji} + \mathbf{i}a_{ij} + a_{jj}$, a real number. As $a_{ii}, a_{jj} \in \mathbb{R}$, $a_{ij} - a_{ji}$ is a purely imaginary number, *i.e.*, they have the same real part. Similarly, taking $\mathbf{x} = \mathbf{e}_i + \mathbf{e}_j$ gives $a_{ij} + a_{ji} \in \mathbb{R}$, *i.e.*, they have opposite imaginary parts. So $a_{ij} = \overline{a_{ji}}$.

6. *Let A and B be Hermitian matrices. Then, prove that AB is Hermitian if and only if $AB = BA$.*

7. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a skew-Hermitian matrix. Then prove that*

   (a) *the diagonal entries of A are either zero or purely imaginary.*

   (b) *for each $B \in \mathbb{M}_n(\mathbb{C})$ prove that B^*AB is a skew-Hermitian matrix.*

   Ans: Note that $-a_{ii} = \mathbf{e}_i^*(-A)\mathbf{e}_i = \mathbf{e}_i^* A^* \mathbf{e}_i = \overline{a_{ii}}$. Thus $-a_{ii} = \overline{a_{ii}}$ and hence a_{ii} is either zero or purely imaginary. $(B^*AB)^* = B^*A^*B = -(B^*AB)$.

8. *Let A be a complex square matrix. Then $S_1 = \frac{1}{2}(A + A^*)$ is Hermitian, $S_2 = \frac{1}{2}(A - A^*)$ is skew-Hermitian, and $A = S_1 + S_2$.*

9. *Let A, B be skew-Hermitian matrices with $AB = BA$. Is the matrix AB Hermitian or skew-Hermitian?*

   Ans: $(AB)^* = B^*A^* = (-B)(-A) = BA = AB$.

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

    Ans: Verify $(I + A)(I - A + \cdots + (-1)^{k-1}A^{k-1}) = (I - A + \cdots + (-1)^{k-1}A^{k-1})(I + A) = I$.

11. *Let* $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$ *and* $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \sin\theta & -\cos\theta \end{bmatrix}$, *for $\theta \in [-\pi, \ \pi)$. Are they orthogonal?*

Ans: Yes, as $AA^T = I = A^T A$ and $B^T B = I = BB^T$.

1.3.1 Submatrix of a Matrix

Definition 1.3.3. For $k \in \mathbb{N}$, let $[k] = \{1, \ldots, k\}$. Also, let $A \in \mathbb{M}_{m\times n}(\mathbb{C})$.

1. Then, a matrix obtained by deleting some of the rows and/or columns of A is said to be a **submatrix** of A.
2. If $S \subseteq [m]$ and $T \subseteq [n]$ then by $\mathbf{A(S|T)}$, we denote the submatrix obtained from A by deleting the rows with indices in S and columns with indices in T. By $A[S,T]$, we mean $A(S^c|T^c)$, where $S^c = [m] \setminus S$ and $T^c = [n] \setminus T$. Whenever, S or T consist of a single element, then we just write the element. If $S = [m]$, then $A[S,T] = A[:,T]$ and if $T = [n]$ then $A[S,T] = A[S,:]$ which matches with our notation in Definition 1.1.1.
3. If $m = n$, the submatrix $A[S,S]$ is called a **principal submatrix** of A.

Example 1.3.4. 1. Let $A = \begin{bmatrix}1&4&5\\0&1&2\end{bmatrix}$. Then, $A[\{1,2\},\{1,3\}] = A[:,\{1,3\}] = \begin{bmatrix}1&5\\0&2\end{bmatrix}$, $A[1,1] = [1], A[2,3] = [2], A[\{1,2\},1] = A[:,1] = \begin{bmatrix}1\\0\end{bmatrix}$, $A[1,\{1,3\}] = [1\ 5]$ and A are a few submatrices of 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.

2. Take $A = \begin{bmatrix}1&2&3\\5&6&7\\9&8&7\end{bmatrix}$, $S = \{1,3\}$ and $T = \{2,3\}$. Then, $A[S,S] = \begin{bmatrix}1&3\\9&7\end{bmatrix}$, $A[T,T] = \begin{bmatrix}6&7\\8&7\end{bmatrix}$, $A(S \mid S) = \begin{bmatrix}6\end{bmatrix}$ and $A(T \mid T) = \begin{bmatrix}1\end{bmatrix}$ are principal submatrices of A.

Let $A \in \mathbb{M}_{n,m}(\mathbb{C})$ and $B \in \mathbb{M}_{m,p}(\mathbb{C})$. Then the product AB is defined. Suppose $r < m$. Then A and B can be decomposed as $A = [P\ Q]$ and $B = \begin{bmatrix}H\\K\end{bmatrix}$, where $P \in \mathbb{M}_{n,r}(\mathbb{C})$ and $H \in \mathbb{M}_{r,p}(\mathbb{C})$ so that $AB = PH + QK$. This is proved next.

Theorem 1.3.5. *Let the matrices A, B, P, H, Q and K be defined as above. Then*

$$AB = PH + QK.$$

Proof. Verify that the matrix products PH and QK are valid. Further, their sum is defined as $PH, QK \in \mathbb{M}_{n,p}(\mathbb{C})$. Now, let $P = [P_{ij}]$, $Q = [Q_{ij}]$, $H = [H_{ij}]$, and $K = [K_{ij}]$. Then, for $1 \le i \le n$ and $1 \le j \le p$, we have

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

Thus, the required result follows. ■

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. *The matrices P, Q, H and K can be further partitioned so as to form blocks that are either identity or zero or matrices that have certain nice properties. So, such a partition may be quite useful during different matrix operations. Examples of such partitions appear throughout the notes.*
3. *Suppose one wants to prove a result for a square matrix A. If we want to prove it using induction then we can prove it for the 1×1 matrix (the initial step of induction). Then assume the result to hold for all $k \times k$ submatrices of A or just the first $k \times k$ principal submatrix of A. At the next step write $A = \begin{bmatrix} B & \mathbf{x} \\ \mathbf{x}^T & a \end{bmatrix}$, where B is a $k \times k$ matrix. Then the result holds for B and then one can proceed to prove it for A.*

EXERCISE **1.3.7.** 1. *Complete the proofs of Theorems 1.2.4 and 1.2.13.*

2. *Let* $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ *and* $B = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$.

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

   Ans: Note that A sends the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$ and the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix}$ which are counter-clockwise rotations by α of the respective vectors.

   (b) *Then $\mathbf{y} = B\mathbf{x}$ gives the reflection about the line $y = \tan(\theta)x$.*

   Ans: Let $y = \tan(\theta)x$ be the line ℓ_1. Then $\begin{bmatrix} a \\ a\tan\theta \end{bmatrix}$ is a general point on ℓ_1. Further, $B\begin{bmatrix} a \\ a\tan\theta \end{bmatrix} = \begin{bmatrix} a \\ a\tan\theta \end{bmatrix}$. So, B fixes every point on ℓ_1.

   Now let ℓ_2 be the line which passes through $\begin{bmatrix} a \\ a\tan\theta \end{bmatrix}$ and is perpendicular to ℓ_1. A general point on ℓ_2 is $\begin{bmatrix} a\sec^2\theta - y\tan\theta \\ y \end{bmatrix}$. Then

$$B\begin{bmatrix} a\sec^2\theta - y\tan\theta \\ y \end{bmatrix} = \begin{bmatrix} 2a - a\sec^2\theta + y\tan\theta \\ 2a\tan\theta - y \end{bmatrix}.$$

   Note that $\begin{bmatrix} 2a - a\sec^2\theta + y\tan\theta \\ 2a\tan\theta - y \end{bmatrix}$ lies on ℓ_2 and $\begin{bmatrix} a \\ a\tan\theta \end{bmatrix}$ is the mid-point of the two points $\begin{bmatrix} 2a - a\sec^2\theta + y\tan\theta \\ 2a\tan\theta - y \end{bmatrix}$ and $\begin{bmatrix} a\sec^2\theta - y\tan\theta \\ y \end{bmatrix}$. Thus, $\begin{bmatrix} 2a - a\sec^2\theta + y\tan\theta \\ 2a\tan\theta - y \end{bmatrix}$ is the reflection of $\begin{bmatrix} a\sec^2\theta - y\tan\theta \\ y \end{bmatrix}$ about the line ℓ_1.

(c) *Let $\alpha = \theta$ and compute $\mathbf{y} = (AB)\mathbf{x}$ and $\mathbf{y} = (BA)\mathbf{x}$. Do they correspond to reflection? If yes, then about which line(s)?*

Ans: Note $AB = \begin{bmatrix} \cos(3\theta) & \sin(3\theta) \\ \sin(3\theta) & -\cos(3\theta) \end{bmatrix}$ and $BA = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$. So, the lines are $y = \tan\left(\frac{3\theta}{2}\right) x$ and $y = \tan\left(\frac{\theta}{2}\right) x$.

(d) *Further, if $\mathbf{y} = C\mathbf{x}$ gives the counter-clockwise rotation through β and $\mathbf{y} = D\mathbf{x}$ gives the reflections about the line $y = \tan(\delta)\, x$. Then prove that*

i. *$AC = CA$ and $\mathbf{y} = (AC)\mathbf{x}$ gives the counter-clockwise rotation through $\alpha + \beta$.*

Ans: Verify that $AC = CA = \begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix}$

ii. *$\mathbf{y} = (BD)\mathbf{x}$ and $\mathbf{y} = (DB)\mathbf{x}$ give rotations. Which angles do they represent?*

Ans: $BD = \begin{bmatrix} \cos 2(\theta-\delta) & -\sin 2(\theta-\delta) \\ \sin 2(\theta-\delta) & \cos 2(\theta-\delta) \end{bmatrix}$, $DB = \begin{bmatrix} \cos 2(\delta-\theta) & -\sin 2(\delta-\theta) \\ \sin 2(\delta-\theta) & \cos 2(\delta-\theta) \end{bmatrix}$.

3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If $AB = BA$ for all $B \in \mathbb{M}_n(\mathbb{C})$ then A is a scalar matrix, i.e., $A = \alpha I$ for some $\alpha \in \mathbb{C}$ (use the matrices $\mathbf{e}_{ij}$ in Definition 1.3.1.1).*

Ans: Let $B = \mathbf{e}_{ij} = \mathbf{e}_i\mathbf{e}_j^T$ for $i \neq j$. Then $AB = A\mathbf{e}_i\mathbf{e}_j^T = A[:,i]\mathbf{e}_j^T$ and $BA = \mathbf{e}_i\mathbf{e}_j^T A = \mathbf{e}_i A[j,:]$. But,

$$A[:,i]\mathbf{e}_j^T = \begin{array}{ccc} & j\text{-th} & \\ & \downarrow & \\ [\mathbf{0},\cdots,\mathbf{0}, & A[:,i], & \mathbf{0},\cdots,\mathbf{0}] \end{array} \text{ and } \mathbf{e}_i A[j,:] = \begin{array}{cc} \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ A[j,:] \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} & \begin{array}{c} \\ \\ \\ \leftarrow i\text{-th} \\ \\ \\ \\ \end{array} \end{array}.$$

Hence $a_{ij} = 0$, if $i \neq j$ and $a_{jj} = a_{ii}$.

4. *Consider the two coordinate transformations*

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

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

Ans: Note $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = A \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = B \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$. Then $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = AB \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$.

(b) *Does the composition of two transformations obtained in the previous part correspond to multiplying two matrices? Give reasons for your answer.*

Ans: Yes, see the above solution.

5. *For $A_{n\times n} = [a_{ij}]$, the* **trace** *of A, denoted* $\mathsf{tr(A)}$, *is defined by* $\mathsf{tr(A)} = \mathsf{a}_{11} + \mathsf{a}_{22} + \cdots + \mathsf{a}_{nn}$.

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

Ans: $3 + 2 = 5$ and $4 + 1 = 5$.

(b) *Let* A *be a matrix with* $A \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *and* $A \begin{bmatrix} 1 \\ -2 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. *If* $B = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix}$ *then compute* $\mathsf{tr}(\mathsf{AB})$. *What about* $\mathsf{tr}(\mathsf{A})$*?*

Ans: Verify $AB = \begin{bmatrix} 2 & 3 \\ 4 & -6 \end{bmatrix}$. So $\mathsf{tr}(\mathsf{AB}) = -4$. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, the given conditions imply $a + 2b = 2, c + 2d = 4, a - 2b = 3$ and $c - 2d = -6$. Thus $\mathsf{tr}(\mathsf{A}) = \mathsf{a} + \mathsf{d} = \frac{5}{2} + \frac{5}{2} = 5$.

(c) *Let* A *and* B *be two square matrices of the same order. Then*

i. $\mathsf{tr}(\mathsf{A}+\mathsf{B}) = \mathsf{tr}(\mathsf{A}) + \mathsf{tr}(\mathsf{B})$.

Ans: $\mathsf{tr}(\mathsf{A}+\mathsf{B}) = \sum\limits_{i=1}^{n} (\mathsf{A}+\mathsf{B})_{ii} = \sum\limits_{i=1}^{n} (\mathsf{A})_{ii} + \sum\limits_{i=1}^{n} (\mathsf{B})_{ii} = \mathsf{tr}(\mathsf{A}) + \mathsf{tr}(\mathsf{B})$.

ii. $\mathsf{tr}(\mathsf{AB}) = \mathsf{tr}(\mathsf{BA})$.

Ans: $\mathsf{tr}(\mathsf{AB}) = \sum\limits_{i=1}^{n} (\mathsf{AB})_{ii} = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij} b_{ji} = \sum\limits_{j=1}^{n}\sum\limits_{i=1}^{n} b_{ji} a_{ij} = \sum\limits_{j=1}^{n} (\mathsf{BA})_{jj} = \mathsf{tr}(\mathsf{BA})$.

(d) *Does there exist matrices* $A, B \in \mathbb{M}_n(\mathbb{C})$ *such that* $AB - BA = cI$, *for some* $c \neq 0$*?*

Ans: No. Note that $\mathsf{tr}(\mathsf{AB} - \mathsf{BA}) = 0$, where as, for $c \neq 0$, $\mathsf{tr}(\mathsf{c\,I}) = \mathsf{nc} \neq 0$.

6. *Let* $J \in \mathbb{M}_n(\mathbb{R})$ *be a matrix having each entry* 1.

(a) *Verify that* $J = \mathbf{1}\mathbf{1}^T$, *where* $\mathbf{1}$ *is a column vector having all entries* 1.

(b) *Verify that* $J^2 = nJ$.

(c) *Also, for any* $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$, *verify that there exist* $\alpha_3, \beta_3 \in \mathbb{R}$ *such that*

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

(d) *Let* $\alpha, \beta \in \mathbb{R}$ *such that* $\alpha \neq 0$ *and* $\alpha + n\beta \neq 0$. *Now, define* $A = \alpha I_n + \beta J$. *Then, use the above to prove that* A *is invertible.*

Ans: $J^2 = (\mathbf{1}\mathbf{1}^T)(\mathbf{1}\mathbf{1}^T) = \mathbf{1}(\mathbf{1}^T\mathbf{1})\mathbf{1}^T = n\mathbf{1}\mathbf{1}^T = nJ$.

Note that in part (6c), $\alpha_3 = \alpha_1\alpha_2$ and $\beta_3 = \alpha_1\beta_2 + \alpha_2\beta_1 + n\beta_1\beta_2$. So, using the third part $B = \frac{1}{\alpha} I - \frac{\beta}{\alpha(\alpha + n\beta)} J$ is the inverse of A.

---

7. *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.*

Ans: Take $G = \frac{1}{3}\begin{bmatrix} -1+k & 2+z \\ 2-5k & -1-5z \\ 3k & 3z \end{bmatrix}$, for k, z arbitrary. Then $AG = I_2$. Does there exists a value of z for which $G = \begin{bmatrix} -8/35 & 3/5 \\ 1/7 & 0 \\ 11/35 & -1/5 \end{bmatrix}$? Note that for this choice of G, one has $AGA = A, GAG = G, (AG)^T = AG$ and $(GA)^T = GA$. The matrices G which satisfy the above are called **pseudo inverse** of A.

8. *Let* $A = \begin{bmatrix} P & Q \\ Q & R \end{bmatrix}$. *If* P, Q *and* R *are Hermitian, is the matrix* A *Hermitian?*

Ans: Yes, as $A^* = \begin{bmatrix} P^* & Q^*n \\ Q^* & R^* \end{bmatrix} = \begin{bmatrix} P & Q \\ Q & R \end{bmatrix}$.

9. *Let* $A = \begin{bmatrix} A_{11} & \mathbf{x} \\ \mathbf{y}^* & c \end{bmatrix}$, *where* $A_{11} \in \mathbb{M}_n(\mathbb{C})$ *is invertible and* $c \in \mathbb{C}$.

   (a) *If* $p = c - \mathbf{y}^* A_{11}^{-1}\mathbf{x}$ *is non zero, then verify that*

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

   is the inverse of A.

   Ans: Just multiply and verify.

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

   Ans: $A_{11}^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}, p = 1 - \begin{bmatrix} -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = 15$. So, the inverse equals $\left[\begin{array}{cc|c} 1 & 1 & 0 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] + \frac{1}{15}\begin{bmatrix} 6 \\ -2 \\ -1 \end{bmatrix}\begin{bmatrix} -3 & -2 & -1 \end{bmatrix} = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ -1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] + \frac{1}{15}\left[\begin{array}{cc|c} -18 & -12 & -6 \\ 6 & 2 & 2 \\ \hline 3 & 2 & 1 \end{array}\right]$ $= \begin{bmatrix} -1/5 & 1/5 & -2/5 \\ -3/5 & 4/15 & 2/15 \\ 1/5 & 2/15 & 1/15 \end{bmatrix}$. For the second matrix the inverse is $\begin{bmatrix} -23/33 & 7/33 & -2/11 \\ 1/33 & 4/33 & 2/11 \\ 17/33 & 2/33 & 1/11 \end{bmatrix}$.

10. *Let* $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{R})$ *be a unit vector.*

    (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**.

    Ans: $A^2 = (I_n - 2\mathbf{x}\mathbf{x}^T)(I_n - 2\mathbf{x}\mathbf{x}^T) = I_n - 4\mathbf{x}\mathbf{x}^T + 4\mathbf{x}\mathbf{x}^T = I_n$ as $\mathbf{x}^T\mathbf{x} = 1$.

    (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. [The inverse is also of the form* $I_n + \beta\mathbf{x}\mathbf{x}^T$, *for some* β.*]*

    Ans: Just multiply and verify that $\beta = \frac{\alpha}{\alpha - 1}$ as $\mathbf{x}^T\mathbf{x} = 1$.

11. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be an invertible matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{M}_{n,1}(\mathbb{R})$. Also, let $\beta \in \mathbb{R}$ such that $\alpha = 1 + \beta \mathbf{y}^T A^{-1}\mathbf{x} \neq 0$. Then, verify 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 (see Definition 2.3.1) one matrix.

Ans: Just multiply and verify.

12. *Suppose the matrices B and C are invertible and the involved partitioned products are defined, then verify that that*

$$\begin{bmatrix} A & B \\ C & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & C^{-1} \\ B^{-1} & -B^{-1}AC^{-1} \end{bmatrix}.$$

Ans: Just multiply and verify.

13. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, a matrix $G \in \mathbb{M}_{n,m}(\mathbb{C})$ is called a* **generalized inverse** *(for short, g-inverse) of A if $AGA = A$. For example, a generalized inverse of the matrix $A = [1, 2]$ is a matrix $G = \begin{bmatrix} 1 - 2\alpha \\ \alpha \end{bmatrix}$, for all $\alpha \in \mathbb{R}$. A generalized inverse G is called a* **pseudo inverse** *or a* **Moore-Penrose inverse** *if $GAG = G$ and the matrices AG and GA are symmetric. Check that for $\alpha = \frac{2}{5}$ the matrix G is a pseudo inverse of A. Further, among all the g-inverses, the inverse with the least euclidean norm also has $\alpha = \frac{2}{5}$.*

1.4 Summary

In this chapter, we started with the definition of a matrix and came across lots of examples. We recall these examples as they will be used in later chapters to relate different ideas:

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.
7. Idempotent matrices.
8. Nilpotent matrices.

We also learnt product of two matrices. Even though it seemed complicated, it basically tells that multiplying by a matrix on the

1. left of A is same as operating on (playing with) the rows of A.

2. right of A is same as operating on (playing with) the columns of A.

The matrix multiplication is not commutative. We also defined the inverse of a matrix. Further, there were exercises that informs us that the rows and columns of invertible matrices cannot have certain properties.

Chapter 2

System of Linear Equations

This chapter starts with understanding the effect of elementary row operations on the solution set of a system of linear equations. This helps us to conclusively give necessary and sufficient conditions for a system of linear equations to have either a unique solution, no solution or an infinite number of solutions.

2.1 Introduction

We start this section with our understanding of the system of linear equations in at most 2 variables/unknowns.

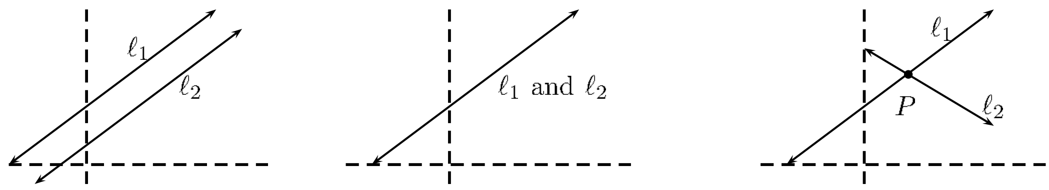
Example 2.1.1. Let us look at some examples of linear systems.

1. Suppose $a, b \in \mathbb{R}$. Consider the system $ax = b$ in the variable x. If
   (a) $a \neq 0$ then the system has a UNIQUE SOLUTION $x = \frac{b}{a}$.
   (b) $a = 0$ 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. Recall that the linear system $ax + by = c$ for $(a, b) \neq (0, 0)$, in the variables x and y, represents a line in $\mathbb{R}^2$. So, let us consider the points of intersection of the two lines

$$a_1x + b_1y = c_1,\ a_2x + b_2y = c_2, \tag{2.1.1}$$

   where $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{R}$ with $(a_1, b_1), (a_2, b_2) \neq (0, 0)$ (see Figure 2.1 for illustration of different cases).

   (a) UNIQUE SOLUTION $(a_1b_2 - a_2b_1 \neq 0)$: The linear system $x - y = 3$ and $2x + 3y = 11$ has $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ as the unique solution.
   (b) NO SOLUTION $(a_1b_2 - a_2b_1 = 0$ but $a_1c_2 - a_2c_1 \neq 0)$: The linear system $x + 2y = 1$ and $2x + 4y = 3$ represent a pair of parallel lines which have no point of intersection.



No Solution
Pair of Parallel lines

Infinite Number of Solutions
Coincident Lines

Unique Solution: Intersecting Lines
P: Point of Intersection

Figure 2.1: Examples in 2 dimension.

(c) Infinite Number of Solutions ($a_1b_2 - a_2b_1 = 0$ and $a_1c_2 - a_2c_1 = 0$): The linear system $x + 2y = 1$ and $2x + 4y = 2$ represent the same line. So, the solution set equals $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1-2y \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ with y arbitrary. Observe that the vector

i. $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ corresponds to the solution $x = 1, y = 0$ of the given system.

ii. $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ gives $x = -2, y = 1$ as the solution of $x + 2y = 0, 2x + 4y = 0$.

(d) If the linear system $ax + by = c$ has

i. $(a, b) = (0, 0)$ and $c \neq 0$ then $ax + by = c$ has no solution.

ii. $(a, b, c) = (0, 0, 0)$ then $ax+by = c$ has infinite number of solutions, namely whole of $\mathbb{R}^2$.

Let us now look at different interpretations of the solution concept.

Example 2.1.2. Observe the following of the linear system in Example 2.1.1.2a.

1. $\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ corresponds to the point of intersection of the corresponding two lines.

2. Using matrix multiplication, the given system equals $A\mathbf{x} = \mathbf{b}$, where $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 11 \end{bmatrix}$. So, the solution is $\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{5}\begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 11 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$.

3. Re-writing $A\mathbf{x} = \mathbf{b}$ as $\begin{bmatrix} 1 \\ 2 \end{bmatrix}x + \begin{bmatrix} -1 \\ 3 \end{bmatrix}y = \begin{bmatrix} 3 \\ 11 \end{bmatrix}$ gives us $4 \cdot (1, 2)^T + 1 \cdot (-1, 3)^T = (3, 11)^T$. This corresponds to addition of vectors in the Euclidean plane.

Thus, there are three ways of looking at the linear system $A\mathbf{x} = \mathbf{b}$, where, as the name suggests, one of the ways is looking at the point of intersection of planes, the other is the vector sum approach and the third is the matrix multiplication approach. We will see that all the three approaches are fundamental to the understanding of linear algebra.

Definition 2.1.3. A system of m linear equations in n variables $x_1, x_2, \ldots, x_n$ is a set of equations of the form

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad & \quad \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned} \tag{2.1.2}$$

where for $1 \le i \le m$ and $1 \le j \le n$; $a_{ij}, b_i \in \mathbb{R}$. Linear System (2.1.2) is called **homogeneous** if $b_1 = 0 = b_2 = \cdots = b_m$ and **non-homogeneous**, otherwise.

Definition 2.1.4. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$. Then, (2.1.2) can be re-written as $A\mathbf{x} = \mathbf{b}$. In this setup, the matrix A is called the **coefficient** matrix and the block matrix $[A \;\; \mathbf{b}]$ is called the **augmented** matrix of the linear system (2.1.2).

Remark 2.1.5. *Consider the linear system $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{M}_{m,n}(\mathbb{C})$, $\mathbf{b} \in \mathbb{M}_{m,1}(\mathbb{C})$ and $\mathbf{x} \in \mathbb{M}_{n,1}(\mathbb{C})$. If $[A\ \mathbf{b}]$ is the augmented matrix and $\mathbf{x}^T = [x_1, \ldots, x_n]$ then,*

1. *for $j = 1, 2, \ldots, n$, the variable x_j corresponds to the column $([A\ \mathbf{b}])[:, j]$.*
2. *the vector $\mathbf{b} = ([A\ \mathbf{b}])[:, n+1]$.*
3. *for $i = 1, 2, \ldots, m$, the i^{th} equation corresponds to the row $([A\ \mathbf{b}])[i, :]$.*

Definition 2.1.6. A **solution** of $A\mathbf{x} = \mathbf{b}$ is a vector $\mathbf{y}$ such that $A\mathbf{y}$ indeed equals $\mathbf{b}$. The set of all solutions is called the **solution set** of the system. For example, the solution set of $A\mathbf{x} = \mathbf{b}$, with $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 2 \\ 4 & 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ equals $\left\{ \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \right\}$.

Definition 2.1.7. Consider a linear system $A\mathbf{x} = \mathbf{b}$. Then, this linear system is called **consistent** if it admits a solution and is called **inconsistent** if it admits no solution. For example, the homogeneous system $A\mathbf{x} = \mathbf{0}$ is always consistent as $\mathbf{0}$ is a solution whereas, verify that the system $x + y = 2, 2x + 2y = 3$ is inconsistent.

Definition 2.1.8. Consider a linear system $A\mathbf{x} = \mathbf{b}$. Then, the corresponding linear system $A\mathbf{x} = \mathbf{0}$ is called the **associated homogeneous system**. $\mathbf{0}$ is always a solution of the associated homogeneous system.

The readers are advised to supply the proof of the next theorem that gives information about the solution set of a homogeneous system.

Theorem 2.1.9. *Consider a homogeneous linear system $A\mathbf{x} = \mathbf{0}$.*

1. *Then, $\mathbf{x} = \mathbf{0}$, the zero vector, is always a solution, called the* **trivial** *solution.*

2. *Let* $\mathbf{u} \neq \mathbf{0}$ *be a solution of* $A\mathbf{x} = \mathbf{0}$. *Then,* $\mathbf{y} = c\mathbf{u}$ *is also a solution, for all* $c \in \mathbb{C}$. *A nonzero solution is called a* **non-trivial** *solution. Note that, in this case, the system* $A\mathbf{x} = \mathbf{0}$ *has an infinite number of solutions.*

3. *Let* $\mathbf{u_1}, \ldots, \mathbf{u}_k$ *be solutions of* $A\mathbf{x} = \mathbf{0}$. *Then,* $\sum\limits_{i=1}^{k} a_i\mathbf{u}_i$ *is also a solution of* $A\mathbf{x} = \mathbf{0}$, *for each choice of* $a_i \in \mathbb{C}, 1 \leq i \leq k$.

Remark 2.1.10. *1. Let* $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. *Then,* $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ *is a non-trivial solution of* $A\mathbf{x} = \mathbf{0}$.

2. *Let* $\mathbf{u} \neq \mathbf{v}$ *be solutions of a non-homogeneous system* $A\mathbf{x} = \mathbf{b}$. *Then,* $\mathbf{x}_h = \mathbf{u} - \mathbf{v}$ *is a solution of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$. *That is, any two distinct solutions of* $A\mathbf{x} = \mathbf{b}$ *differ by a solution of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$. *Or equivalently, the solution set of* $A\mathbf{x} = \mathbf{b}$ *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* $A\mathbf{x} = \mathbf{0}$.

EXERCISE **2.1.11.** *1. Consider a system of* 2 *equations in* 3 *variables. If this system is consistent then how many solutions does it have?*

Ans: Since there are two intersecting (system is consistent) planes in $\mathbb{R}^3$ they will intersect in a line. So, infinite number of solutions.

2. *Give a linear system of* 3 *equations in* 2 *variables such that the system is inconsistent whereas it has 2 equations which form a consistent system.*

   Ans: $x + y = 2, x + 2y = 3, 2x + 3y = 4$.

3. *Give a linear system of* 4 *equations in* 3 *variables such that the system is inconsistent whereas it has three equations which form a consistent system.*

   Ans: $x + y + z = 3, x + 2y + 3z = 6, 2x + 3y + 4z = 4, 2x + 2y + z = 5$.

4. *Let* $A\mathbf{x} = \mathbf{b}$ *be a system of* m *equations in* n *variables, where* $A \in \mathbb{M}_{m,n}(\mathbb{C})$.

   (a) *Can the system,* $A\mathbf{x} = \mathbf{b}$ *have exactly two distinct solutions for any choice of* m *and* n*? Give reasons for your answer.*

   (b) *Can the system* $A\mathbf{x} = \mathbf{b}$ *have only a finitely many (greater than* 1*) solutions for any choice of* m *and* n*? Give reasons for your answer.*

   Ans: No. Let $\mathbf{x}_1, \mathbf{x}_2$ be two solutions. Define $\mathbf{z} = a\mathbf{x}_1 + (1-a)\mathbf{x}_2$ for $a \in \mathbb{R}$. Then $A\mathbf{z} = aA\mathbf{x}_1 + (1-a)A\mathbf{x}_2 = a\mathbf{b} + (1-a)\mathbf{b} = \mathbf{b}$.

2.1.1 Elementary Row Operations

A system of linear equations can be solved by people differently. But, the final solution remains the same. In this section, we use a systematic way to solve any linear system which is popularly known as the Guass Elimination method.

Example 2.1.12. Solve the linear system $y + z = 2,\ 2x + 3z = 5,\ x + y + z = 3$.

Solution: Let $B_0 = [A \;\; \mathbf{b}]$, the augmented matrix. Then, $B_0 = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$. We now systematically proceed to get the solution.

1. Interchange 1-st and 2-nd equations (interchange $B_0[1,:]$ and $B_0[2,:]$ to get B_1).

$$\begin{aligned} 2x + 3z &= 5 \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad B_1 = \begin{bmatrix} 2 & 0 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

2. In the new system, multiply 1-st equation by $\frac{1}{2}$ (multiply $B_1[1,:]$ by $\dfrac{1}{2}$ to get B_2).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ x + y + z &= 3 \end{aligned} \qquad B_2 = \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

3. In the new system, replace 3-rd equation by 3-rd equation minus 1-st equation (replace $B_2[3,:]$ by $B_2[3,:] - B_2[1,:]$ to get B_3).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ y - \tfrac{1}{2}z &= \tfrac{1}{2} \end{aligned} \qquad B_3 = \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

4. In the new system, replace 3-rd equation by 3-rd equation minus 2-nd equation (replace $B_3[3,:]$ by $B_3[3,:] - B_3[2,:]$ to get B_4).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ -\tfrac{3}{2}z &= -\tfrac{3}{2} \end{aligned} \qquad B_4 = \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

5. In the new system, multiply 3-rd equation by $\dfrac{-2}{3}$ (multiply $B_4[3,:]$ by $\dfrac{-2}{3}$ to get B_5).

$$\begin{aligned} x + \tfrac{3}{2}z &= \tfrac{5}{2} \\ y + z &= 2 \\ z &= 1 \end{aligned} \qquad B_5 = \begin{bmatrix} 1 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

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 $\{[x, y, z]^T \mid [x, y, z] = [1, 1, 1]\}$, A UNIQUE SOLUTION.

In Example 2.1.12, observe how each operation on the linear system corresponds to a similar operation on the rows of the augmented matrix. We use this idea to define elementary row operations and the equivalence of two linear systems.

Definition 2.1.13. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, the **elementary row operations** are

1. E_{ij}: Interchange the i-th and j-th rows, namely, interchange $A[i,:]$ and $A[j,:]$.
2. $E_k(c)$ for $c \neq 0$: Multiply the k-th row by c, namely, multiply $A[k,:]$ by c.
3. $E_{ij}(c)$ for $c \neq 0$: Replace the i-th row by i-th row plus c-times the j-th row, namely, replace $A[i,:]$ by $A[i,:] + cA[j,:]$.

Definition 2.1.14. Two matrices are said to be **row equivalent** if one can be obtained from the other by a finite number of elementary row operations.

Definition 2.1.15. The linear systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ are said to be **row equivalent** if their respective augmented matrices, $[A \;\; \mathbf{b}]$ and $[C \;\; \mathbf{d}]$, are row equivalent.

Thus, note that the linear systems at each step in Example 2.1.12 are row equivalent to each other. We now prove that the solution set of two row equivalent linear systems are same.

Lemma 2.1.16. *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 $E_{jk}(c)$ with $c \neq 0$. The reader is advised to prove the result for the other two elementary operations.

In this case, the systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ vary only in the j^{th} equation. So, we need to show that $\mathbf{y}$ satisfies the j^{th} equation of $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{y}$ satisfies the j^{th} equation of $C\mathbf{x} = \mathbf{d}$. So, let $\mathbf{y}^T = [\alpha_1, \ldots, \alpha_n]$. Then, the j^{th} and k^{th} equations of $A\mathbf{x} = \mathbf{b}$ are $a_{j1}\alpha_1 + \cdots + a_{jn}\alpha_n = b_j$ and $a_{k1}\alpha_1 + \cdots + a_{kn}\alpha_n = b_k$. Therefore, we see that α_i's satisfy

$$(a_{j1} + ca_{k1})\alpha_1 + \cdots + (a_{jn} + ca_{kn})\alpha_n = b_j + cb_k. \tag{2.1.3}$$

Also, by definition the j^{th} equation of $C\mathbf{x} = \mathbf{d}$ equals

$$(a_{j1} + ca_{k1})x_1 + \cdots + (a_{jn} + ca_{kn})x_n = b_j + cb_k. \tag{2.1.4}$$

Therefore, using Equation (2.1.3), we see that $\mathbf{y}^T = [\alpha_1, \ldots, \alpha_n]$ is also a solution for Equation (2.1.4). Now, use a similar argument to show that if $\mathbf{z}^T = [\beta_1, \ldots, \beta_n]$ is a solution of $C\mathbf{x} = \mathbf{d}$ then it is also a solution of $A\mathbf{x} = \mathbf{b}$. Hence, the required result follows. ■

The readers are advised to use Lemma 2.1.16 as an induction step to prove the next result.

Theorem 2.1.17. *Let $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ be two row equivalent linear systems. Then, they have the same solution set.*

The exercise below shows that every square matrix is row equivalent to an upper triangular matrix.

EXERCISE **2.1.18.** *Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{R})$. Then there exists an orthogonal matrix U such that UA is upper triangular. The proof uses the following ideas.*

1. *If $A[1,:] = \mathbf{0}$ then proceed to the next column. So, let $A[:,1] \neq \mathbf{0}$. If $a_{11} = 0$ then apply a permutation matrix P (an orthogonal matrix, see Definition 1.3.1.2d) to get $B = PA$ such that the $(1,1)$-th entry of B is non zero. Hence, without loss of generality, let $a_{11} \neq 0$.*

2. *Let* $[w_1, \ldots, w_n]^T = \mathbf{w} \in \mathbb{R}^n$ *with* $w_1 \neq 0$. *Then use the Householder matrix (see 1.3.7.10a)* H *such that* $H\mathbf{w} = \alpha\mathbf{e}_1$ *for some* $\alpha \in \mathbb{R}$, *i.e., find* $\mathbf{x} \in \mathbb{R}^n$ *such that* $(I_n - 2\mathbf{x}\mathbf{x}^T)\mathbf{w} = \alpha\mathbf{e}_1$.

   Ans: Given condition implies $\mathbf{w} - \alpha\mathbf{e}_1 = 2(\mathbf{x}^T\mathbf{w})\mathbf{x}$. So $\mathbf{x} = \dfrac{\mathbf{w} - \alpha\mathbf{e}_1}{2\mathbf{x}^T\mathbf{w}}$. As $\dfrac{1}{2\mathbf{x}^T\mathbf{w}}$ is scalar, use $\mathbf{x} = \mathbf{w} + \alpha\mathbf{e}_1$ to find a choice of α. Show that for $\alpha = \dfrac{1 - 2\mathbf{w}^T\mathbf{w}}{2w_1}$, $H\mathbf{w} = -\alpha\mathbf{e}_1$.

3. *So, Part 2 gives an orthogonal matrix* H_1 *with* $H_1 A = \begin{bmatrix} \alpha & * \\ \mathbf{0} & A_1 \end{bmatrix}$.

4. *Now, use induction to get* $H_2 \in \mathbb{M}_{n-1}(\mathbb{R})$ *to get* $H_2 A_1 = T_1$, *an upper triangular matrix.*

5. *Define* $H = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & H_2 \end{bmatrix} H_1$. *Then* H *is an orthogonal matrix and* $HA = \begin{bmatrix} \alpha & * \\ \mathbf{0} & T_1 \end{bmatrix}$, *an upper triangular matrix.*

2.2 Row-Reduced Echelon Form (RREF)

In the previous section, we saw that two row equivalent linear systems have the same solution set. Sometimes it helps to imagine an elementary row operation as left multiplication by a suitable matrix, known as an elementary matrix. In this section, we show that the product of such matrices can be used to obtain a matrix which has certain nice properties. This will also help us to understand the Gauss Elimination method and the Gauss-Jordan method. This understanding will be used to define the row-rank of a matrix in the next section and in subsequent sections we use them to obtain results for a system of linear equations.

Definition 2.2.1. A matrix $E \in \mathbb{M}_n(\mathbb{C})$ is called an **elementary matrix** if it is obtained by applying exactly one elementary row operation to the identity matrix I_n.

Remark 2.2.2. *The elementary matrices are of three types and they correspond to elementary row operations.*

1. $E_{ij} = I_n - \mathbf{e}_i\mathbf{e}_i^T - \mathbf{e}_j\mathbf{e}_j^T + \mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T$: *Matrix obtained by applying elementary row operation* E_{ij} *to* I_n.

2. $E_k(c) = I_n + (c-1)\mathbf{e}_k\mathbf{e}_k^T$ *for* $c \neq 0$: *Matrix obtained by applying elementary row operation* $E_k(c)$ *to* I_n.

3. $E_{ij}(c) = I_n + c\ \mathbf{e}_i\mathbf{e}_j^T$ *for* $c \neq 0$: *Matrix obtained by applying elementary row operation* $E_{ij}(c)$ *to* I_n.

Thus, when an elementary matrix is multiplied on the left of a matrix A, *it gives the same result as that of applying the corresponding elementary row operation on* A.

Example 2.2.3. 1. In particular, for $n = 3$ and $c \in \mathbb{C}, c \neq 0$, one has

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},\ E_1(c) = \begin{bmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ E_{31}(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & 0 & 1 \end{bmatrix} \text{ and } E_{23}(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}.$$

2. Verify that the transpose of an elementary matrix is again an elementary matrix of similar type (see the above examples).

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

   (a) If B_1 is obtained from A by applying the elementary row operation E_{23} then $B_1 = E_{23}A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 2 & 0 & 3 \end{bmatrix}$.

   (b) If B is obtained from A by applying the elementary row operation $E_{31}(-3)$ then $B = E_{31}(-3)A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \\ 0 & -2 & -4 \end{bmatrix}$.

   (c) If C is obtained from B by applying the elementary row operation $E_{21}(-2)$ then $C = E_{21}(-2)A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -3 \\ 0 & -2 & -4 \end{bmatrix}$.

   (d) Where as $AE_{23} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 3 & 0 \\ 3 & 5 & 4 \end{bmatrix}$, $AE_{31}(-3) = \begin{bmatrix} -8 & 2 & 3 \\ -7 & 0 & 3 \\ -12 & 4 & 5 \end{bmatrix}$.

Exercise **2.2.4.** *1. Which of the following matrices are elementary?*

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

2. Find some elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = I_2$.

3. Find some elementary matrices $F_1, \ldots, F_\ell$ such that $F_\ell \cdots F_1 \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix} = I_3$.

Exercise **2.2.5.** *Show that each elementary matrix is invertible. Further, the inverse is an elementary matrix of the same type.*

Ans: Verify that $(E_{ij})^{-1} = E_{ij}$ as $E_{ij}E_{ij} = I = E_{ij}E_{ij}$. If $c \neq 0$ then $(E_k(c))^{-1} = E_k(1/c)$ as $E_k(c)E_k(1/c) = I = E_k(1/c)E_k(c)$ and $(E_{ij}(c))^{-1} = E_{ij}(-c)$ as $E_{ij}(c)E_{ij}(-c) = I = E_{ij}(-c)E_{ij}(c)$.

Proposition 2.2.6. *Let A and B be two row equivalent matrices. Then, there exists elementary matrices $E_1, \ldots, E_k$ such that $B = E_1 \cdots E_k A$.*

Proof. By the definition of row equivalence, B can be obtained from A by a finite number of elementary row operations. But by Remark 2.2.2, each elementary row operation corresponds to left multiplication by an elementary matrix. Thus, the required result follows. ∎

We now give an alternate prove of Theorem 2.1.17.

Theorem 2.2.7. *Let* $A\mathbf{x} = \mathbf{b}$ *and* $C\mathbf{x} = \mathbf{d}$ *be two row equivalent linear systems. Then they have the same solution set.*

Proof. Let $E_1, \ldots, E_k$ be the elementary matrices such that $E_1 \cdots E_k[A \quad \mathbf{b}] = [C \quad \mathbf{d}]$. Put $E = E_1 \cdots E_k$. Then, by Exercise 2.2.5

$$EA = C, \quad E\mathbf{b} = \mathbf{d}, \quad A = E^{-1}C \text{ and } \mathbf{b} = E^{-1}\mathbf{d}. \tag{2.2.1}$$

Now assume that $A\mathbf{y} = \mathbf{b}$ holds. Then, by Equation (2.2.1)

$$C\mathbf{y} = EA\mathbf{y} = E\mathbf{b} = \mathbf{d}. \tag{2.2.2}$$

On the other hand if $C\mathbf{z} = \mathbf{d}$ holds then using Equation (2.2.1), we have

$$A\mathbf{z} = E^{-1}C\mathbf{z} = E^{-1}\mathbf{d} = \mathbf{b}. \tag{2.2.3}$$

Therefore, using Equations (2.2.2) and (2.2.3) the required result follows. ■

The following result is a particular case of Theorem 2.2.7.

Corollary 2.2.8. *Let* A *and* B *be two row equivalent matrices. Then, the systems* $A\mathbf{x} = \mathbf{0}$ *and* $B\mathbf{x} = \mathbf{0}$ *have the same solution set.*

Example 2.2.9. Are the matrices $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix}$ row equivalent?

Solution: No, as $\begin{bmatrix} a \\ b \\ -1 \end{bmatrix}$ is a solution of $B\mathbf{x} = \mathbf{0}$ but it isn't a solution of $A\mathbf{x} = \mathbf{0}$.

Definition 2.2.10. Let A be a nonzero matrix. Then, in each nonzero row of A, the left most nonzero entry is called a **pivot/leading entry**. The column containing the pivot is called a **pivotal column**. If a_{ij} is a pivot then we denote it by $\boxed{a_{ij}}$. For example, the entries a_{12} and a_{23} are pivots in $A = \begin{bmatrix} 0 & \boxed{3} & 4 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{2} & 1 \end{bmatrix}$. Thus, columns 2 and 3 are pivotal columns.

Definition 2.2.11. A matrix is in **row echelon form (REF)** (ladder like)

1. if the zero rows are at the bottom;

2. if the pivot of the $(i + 1)$-th row, if it exists, comes to the right of the pivot of the i-th row.

3. if the entries below the pivot in a pivotal column are 0.

Example 2.2.12. 1. The following matrices are in echelon form.

$$\begin{bmatrix} 0 & \boxed{2} & 4 & 2 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \boxed{1} & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & \boxed{3} & 4 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}, \begin{bmatrix} \boxed{1} & 2 & 0 & 5 \\ 0 & \boxed{2} & 0 & 6 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix}.$$

2. The following matrices are not in echelon form (determine the rule(s) that fail).

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

Definition 2.2.13. A matrix C is said to be in **row-reduced echelon form (RREF)**

1. if C is already in echelon form,
2. if the pivot of each nonzero row is 1,
3. if every other entry in each pivotal column is zero.

A matrix in RREF is also called a row-reduced echelon matrix.

Example 2.2.14. 1. The following matrices are in RREF.

$$\begin{bmatrix} 0 & \boxed{1} & 0 & -2 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \boxed{1} & 0 & 0 & 5 \\ 0 & \boxed{1} & 0 & 6 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} \boxed{1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}.$$

2. The following matrices are not in RREF (determine the rule(s) that fail).

$$\begin{bmatrix} 0 & \boxed{3} & 3 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}, \begin{bmatrix} 0 & \boxed{1} & 3 & 1 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. We now present an algorithm, commonly known as the Gauss-Jordan Elimination (GJE), to compute the RREF of A.

1. Input: A.
2. Output: a matrix B in RREF such that A is row equivalent to B.
3. **Step 1:** Put 'Region' $= A$.
4. **Step 2:** If all entries in the Region are 0, STOP. Else, in the Region, find the leftmost nonzero column and find its topmost nonzero entry. Suppose this nonzero entry is $a_{ij} = c$ (say). Box it. This is a pivot.
5. **Step 3:** Interchange the row containing the pivot with the top row of the region. Also, make the pivot entry 1 by dividing this top row by c. Use this pivot to make other entries in the pivotal column as 0.
6. **Step 4:** Put Region $=$ the submatrix below and to the right of the current pivot. Now, go to step 2.

   Important: The process will stop, as we can get at most $\min\{m, n\}$ pivots.

Example 2.2.15. Apply GJE to $\begin{bmatrix} 0 & 2 & 3 & 7 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 8 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

1. Region $= A$ as $A \neq \mathbf{0}$.

2. Then, $E_{12}A = \begin{bmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & 2 & 3 & 7 \\ 1 & 3 & 4 & 8 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Also, $E_{31}(-1)E_{12}A = \begin{bmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & 2 & 3 & 7 \\ 0 & 2 & 3 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix} = B$ (say).

3. Now, Region $= \begin{bmatrix} 2 & 3 & 7 \\ 2 & 3 & 7 \\ 0 & 0 & 1 \end{bmatrix} \neq \mathbf{0}$. Then, $E_2(\frac{1}{2})B = \begin{bmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 2 & 3 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix} = C$(say). Then,

   $$E_{12}(-1)E_{32}(-2)C = \begin{bmatrix} \boxed{1} & 0 & \frac{-1}{2} & \frac{-5}{2} \\ 0 & \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = D(\text{say}).$$

4. Now, Region $= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Then, $E_{34}D = \begin{bmatrix} \boxed{1} & 0 & \frac{-1}{2} & \frac{-5}{2} \\ 0 & \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Now, multiply on the left by $E_{13}(\frac{5}{2})$ and $E_{23}(\frac{-7}{2})$ to get $\begin{bmatrix} \boxed{1} & 0 & -\frac{1}{2} & 0 \\ 0 & \boxed{1} & \frac{3}{2} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$, a matrix in RREF. Thus, A is row equivalent to F, where $F = \text{RREF}(A) = \begin{bmatrix} \boxed{1} & 0 & -\frac{1}{2} & 0 \\ 0 & \boxed{1} & \frac{3}{2} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

EXERCISE **2.2.16.**

1. *Let $A\mathbf{x} = \mathbf{b}$ be a linear system of m equations in 2 variables. What are the possible choices for RREF($[A \ \ \mathbf{b}]$), if $m \geq 1$?*

2. *Let* $A = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ 2\mathbf{x}_1 - 5\mathbf{x}_2 + \pi\mathbf{x}_3 \end{bmatrix}$ *and* $B = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{0} \end{bmatrix}$ *be two matrices, where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are any three row vectors of the same size. Then, prove that RREF(A) = RREF(B).*

3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If A is not a scalar matrix, i.e., $A \neq \alpha I$, for any $\alpha \in \mathbb{C}$ then prove that there exists a non-singular matrix S such that $SAS^{-1} = B$ with $B = [b_{ij}]$ and $b_{11} = 0$.*

Ans: If A has a non-zero entry in the first row, say $a_{1i} \neq 0$, (the first column, say $a_{j1} \neq 0$) then take $S = I_n + \dfrac{a_{11}}{a_{1i}} \mathbf{e}_i \mathbf{e}_1^T$ ($S = I_n - \dfrac{a_{11}}{a_{j1}} \mathbf{e}_1 \mathbf{e}_j^T$).

---

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

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

The proof of the next result is beyond the scope of this book and hence is omitted.

Theorem 2.2.17. *Let A and B be two row equivalent matrices in RREF. Then $A = B$.*

As an immediate corollary, we obtain the following important result.

Corollary 2.2.18. *The RREF of a matrix A is unique.*

Proof. Suppose there exists a matrix A with two different RREFs, say B and C. As the RREFs are obtained by left multiplication of elementary matrices, there exist elementary matrices $E_1, \ldots, E_k$ and $F_1, \ldots, F_\ell$ such that $B = E_1 \cdots E_k A$ and $C = F_1 \cdots F_\ell A$. Let $E = E_1 \cdots E_k$ and $F = F_1 \cdots F_\ell$. Thus, $B = EA = EF^{-1}C$.

As inverse of an elementary matrix is an elementary matrix, F^{-1} is a product of elementary matrices and hence B and C are row equivalent. As B and C are in RREF, using Theorem 2.2.17, $B = C$. ■

Remark 2.2.19. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$.*

1. *Then, by Corollary 2.2.18, it's RREF is unique.*
2. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, the uniqueness of RREF implies that RREF(A) is independent of the choice of the row operations used to get the final matrix which is in RREF.*
3. *Let $B = EA$, for some elementary matrix E. Then, RREF(A) = RREF(B).*

   Proof. Let $E_1, \ldots, E_k$ and $F_1, \ldots, F_\ell$ be elementary matrices such that RREF$(A) = E_1 \cdots E_k A$ and RREF$(B) = F_1 \cdots F_\ell B$. Then,

   $$RREF(B) = F_1 \cdots F_\ell B = (F_1 \cdots F_\ell)EA = (F_1 \cdots F_\ell)E(E_k^{-1} \cdots E_1^{-1})RREF(A).$$

   Thus, the matrices RREF(A) and RREF(B) are row equivalent. Since they are also in RREF by Theorem 2.2.17, RREF(A) = RREF(B). ■

4. *Then, there exists an invertible matrix P, a product of elementary matrices, such that PA = RREF(A).*

   Proof. By definition, RREF$(A) = E_1 \cdots E_k A$, for certain elementary matrices $E_1, \ldots, E_k$. Take $P = E_1 \cdots E_k$. Then, P is invertible (product of invertible matrices is invertible) and PA = RREF(A). ■

5. *Let $F = RREF(A)$ and $B = [A[:,1], \ldots, A[:,s]]$, for some $s \leq n$. Then,*

$$RREF(B) = [F[:,1], \ldots, F[:,s]].$$

Proof. By Remark 2.2.19.4, there exist an invertible matrix P, such that

$$F = PA = [PA[:,1], \ldots, PA[:,n]] = [F[:,1], \ldots, F[:,n]].$$

Thus, $PB = [PA[:,1], \ldots, PA[:,s]] = [F[:,1], \ldots, F[:,s]]$. As F is in RREF, it's first s columns are also in RREF. Hence, by Corollary 2.2.18, $RREF(PB) = [F[:,1], \ldots, F[:,s]]$. Now, a repeated use of Remark 2.2.19.3 gives $RREF(B) = [F[:,1], \ldots, F[:,s]]$. Thus, the required result follows. ■

Example 2.2.20. Consider a linear system $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{M}_3(\mathbb{C})$ and $A[:,1] \neq \mathbf{0}$.Then, verify that the 7 different choices for $[C \;\; \mathbf{d}] = \text{RREF}([A \;\; \mathbf{b}])$ are

1. $\begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \end{bmatrix}$. Here, $A\mathbf{x} = \mathbf{b}$ is consistent. The UNIQUE SOLUTION equals $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$.

2. $\begin{bmatrix} 1 & 0 & \alpha & 0 \\ 0 & 1 & \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 1 & \alpha & \beta & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Here, $A\mathbf{x} = \mathbf{b}$ is inconsistent for any choice of α, β as $\text{RREF}([A \;\; \mathbf{b}])$ has a row of $[0 \;\; 0 \;\; 0 \;\; 1]$. This corresponds to solving $0 \cdot x + 0 \cdot y + 0 \cdot z = 1$, an equation which has no solution.

3. $\begin{bmatrix} 1 & 0 & \alpha & d_1 \\ 0 & 1 & \beta & d_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & \alpha & 0 & d_1 \\ 0 & 0 & 1 & d_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ or $\begin{bmatrix} 1 & \alpha & \beta & d_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Here, $A\mathbf{x} = \mathbf{b}$ is consistent and has INFINITE NUMBER OF SOLUTIONS for every choice of α, β as $\text{RREF}([A \; \mathbf{b}])$ has no row of the form $[0 \; 0 \; 0 \; 1]$.

Proposition 2.2.21. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is invertible if and only if $RREF(A) = I_n$, i.e., every invertible matrix is a product of elementary matrices.*

Proof. If $\text{RREF}(A) = I_n$ then $I_n = E_1 \cdots E_k A$, for some elementary matrices $E_1, \ldots, E_k$. As E_i's are invertible, $E_1^{-1} = E_2 \cdots E_k A$, $E_2^{-1} E_1^{-1} = E_3 \cdots E_k A$ and so on. Finally, one obtains $A = E_k^{-1} \cdots E_1^{-1}$. A similar calculation now gives $AE_1 \cdots E_k = I_n$. Hence, by definition of invertibility $A^{-1} = E_1 \cdots E_k$.

Now, let A be invertible with $B = \text{RREF}(A) = E_1 \cdots E_k A$, for some elementary matrices $E_1, \ldots, E_k$. As A and E_i's are invertible, the matrix B is invertible. Hence, B doesn't have any zero row. Thus, all the n rows of B have pivots. Therefore, B has n pivotal columns. As B has exactly n columns, each column is a pivotal column and hence $B = I_n$. Thus, the required result follows. ■

As a direct application of Proposition 2.2.21 and Remark 2.2.19.3 one obtains the following.

Theorem 2.2.22. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, for any invertible matrix S, RREF(SA) = RREF(A).*

Proposition 2.2.23. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be an invertible matrix. Then, for any matrix B, define $C = \begin{bmatrix} A & B \end{bmatrix}$ and $D = \begin{bmatrix} A \\ B \end{bmatrix}$. Then, RREF$(C) = \begin{bmatrix} I_n & A^{-1}B \end{bmatrix}$ and RREF$(D) = \begin{bmatrix} I_n \\ \mathbf{0} \end{bmatrix}$.*

Proof. Using matrix product,

$$A^{-1}C = \begin{bmatrix} A^{-1}A & A^{-1}B \end{bmatrix} = \begin{bmatrix} I_n & A^{-1}B \end{bmatrix}.$$

As $\begin{bmatrix} I_n & A^{-1}B \end{bmatrix}$ is in RREF, by Remark 2.2.19.1, RREF$(C) = \begin{bmatrix} I_n & A^{-1}B \end{bmatrix}$.

For the second part, note that the matrix $X = \begin{bmatrix} A^{-1} & \mathbf{0} \\ -BA^{-1} & I_n \end{bmatrix}$ is an invertible matrix. Thus, by Proposition 2.2.21, X is a product of elementary matrices. Now, verify that $XD = \begin{bmatrix} I_n \\ \mathbf{0} \end{bmatrix}$. As $\begin{bmatrix} I_n \\ \mathbf{0} \end{bmatrix}$ is in RREF, a repeated application of Remark 2.2.19.1 gives the required result. ∎

As an application of Proposition 2.2.23, we have the following observation.

Let $A \in \mathbb{M}_n(\mathbb{C})$. Suppose we start with $C = [A \;\; I_n]$ and compute RREF(C). If RREF(C) = $[G \;\; H]$ then, either $G = I_n$ or $G \neq I_n$. Thus, if $G = I_n$ then we must have $H = A^{-1}$. If $G \neq I_n$ then, A is not invertible. We explain this with an example.

Example 2.2.24. Use GJE to find the inverse of $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Solution: Applying GJE to $[A \mid I_3] = \left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$ gives

$$[A \mid I_3] \overset{E_{13}}{\to} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right] \overset{E_{13}(-1), E_{23}(-2)}{\to} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$$

$$\overset{E_{12}(-1)}{\to} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right].$$

Thus, $A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Exercise **2.2.25.** *Find the inverse of the following matrices using GJE.*

(*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*) $\begin{bmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$.

2.3 Rank of a Matrix

Definition 2.3.1. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, the **rank** of A, denoted $\mathsf{Rank}(A)$, is the number of pivots in the RREF(A). For example, $\mathsf{Rank}(I_n) = n$ and $\mathsf{Rank}(\mathbf{0}) = 0$.

Remark 2.3.2. *Before proceeding further, for $A \in \mathbb{M}_{m,n}(\mathbb{C})$, we observe the following.*

1. *The number of pivots in the RREF(A) is same as the number of pivots in REF of A. Hence, we need not compute the RREF(A) to determine the rank of A.*
2. *Since, the number of pivots cannot be more than the number of rows or the number of columns, one has $\mathsf{Rank}(A) \le \min\{m, n\}$.*
3. *If $B = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ then $\mathsf{Rank}(B) = \mathsf{Rank}(A)$ as RREF$(B) = \begin{bmatrix} RREF(A) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.*
4. *If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ then, by definition*

$$\mathsf{Rank}(A) \le \mathsf{Rank}\left(\begin{bmatrix} A_{11} & A_{12} \end{bmatrix}\right) + \mathsf{Rank}\left(\begin{bmatrix} A_{21} & A_{22} \end{bmatrix}\right).$$

   Further, using Remark 2.2.19,

   (a) $\mathsf{Rank}(A) \ge \mathsf{Rank}\left(\begin{bmatrix} A_{11} & A_{12} \end{bmatrix}\right)$.

   (b) $\mathsf{Rank}(A) \ge \mathsf{Rank}\left(\begin{bmatrix} A_{21} & A_{22} \end{bmatrix}\right)$.

   (c) $\mathsf{Rank}(A) \ge \mathsf{Rank}\left(\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}\right)$.

We now illustrate the calculation of the rank by giving a few examples.

Example 2.3.3. Determine the rank of the following matrices.

1. Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$. Then, $\mathsf{Rank}(A) = \mathsf{Rank}(B) = 1$. Also, verify that $AB = \mathbf{0}$ and $BA = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$. So, $\mathsf{Rank}(AB) = 0 \neq 1 = \mathsf{Rank}(BA)$.
2. Let $A =$ diag$(d_1, \ldots, d_n)$. Then, $\mathsf{Rank}(A)$ equals the number of nonzero d_i's.
3. Let $A = \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix}$. Then, $\mathsf{Rank}(A) = 2$ as it's REF has two pivots.

We now show that the rank doesn't change if a matrix is multiplied on the left by an invertible matrix.

Lemma 2.3.4. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If S is an invertible matrix then* $\mathsf{Rank}(SA) = \mathsf{Rank}(A)$.

Proof. By Theorem 2.2.22, RREF$(A) =$ RREF(SA). Hence, $\mathsf{Rank}(SA) = \mathsf{Rank}(A)$. ∎

We now have the following result.

Corollary 2.3.5. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$ and $B \in \mathbb{M}_{n,q}(\mathbb{C})$. Then,* $\mathsf{Rank}(AB) \le \mathsf{Rank}(A)$.
In particular, if $B \in \mathbb{M}_n(\mathbb{C})$ is invertible then $\mathsf{Rank}(AB) = \mathsf{Rank}(A)$.

Proof. Let $\mathsf{Rank}(A) = r$. Then, there exists an invertible matrix P and $A_1 \in \mathbb{M}_{r,n}(\mathbb{C})$ such that $PA = \mathrm{RREF}(A) = \begin{bmatrix} A_1 \\ \mathbf{0} \end{bmatrix}$. Then, $PAB = \begin{bmatrix} A_1 \\ \mathbf{0} \end{bmatrix} B = \begin{bmatrix} A_1 B \\ \mathbf{0} \end{bmatrix}$. So, using Lemma 2.3.4 and Remark 2.3.2.2, we get

$$\mathsf{Rank}(AB) = \mathsf{Rank}(PAB) = \mathsf{Rank}\left(\begin{bmatrix} A_1 B \\ \mathbf{0} \end{bmatrix}\right) = \mathsf{Rank}(A_1 B) \le r = \mathsf{Rank}(A). \qquad (2.3.4)$$

In particular, if B is invertible then, using Equation (2.3.4), we get

$$\mathsf{Rank}(A) = \mathsf{Rank}(ABB^{-1}) \le \mathsf{Rank}(AB)$$

and hence the required result follows. ∎

Theorem 2.3.6. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If $\mathsf{Rank}(A) = r$ then, there exist invertible matrices P and Q such that*

$$P\ A\ Q = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Proof. Let $C = \mathrm{RREF}(A)$. Then, by Remark 2.2.19.4 there exists as invertible matrix P such that $C = PA$. Note that C has r pivots and they appear in columns, say $i_1 < i_2 < \cdots < i_r$.

Now, let $D = CE_{1i_1}E_{2i_2}\cdots E_{ri_r}$. As E_{ji_j}'s are elementary matrices that interchange the columns of C, one has $D = \begin{bmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $B \in \mathbb{M}_{r,n-r}(\mathbb{C})$.

Put $Q_1 = E_{1i_1}E_{2i_2}\cdots E_{ri_r}$. Then, Q_1 is invertible. Let $Q_2 = \begin{bmatrix} I_r & -B \\ \mathbf{0} & I_{n-r} \end{bmatrix}$. Then, verify that Q_2 is invertible and

$$CQ_1Q_2 = DQ_2 = \begin{bmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} I_r & -B \\ 0 & I_{n-r} \end{bmatrix} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Thus, if we put $Q = Q_1Q_2$ then Q is invertible and $PAQ = CQ = CQ_1Q_2 = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and hence, the required result follows. ∎

We now prove the following result.

Proposition 2.3.7. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be an invertible matrix and let S be any subset of $\{1, 2, \ldots, n\}$. Then $\mathsf{Rank}(A[S,:]) = |S|$ and $\mathsf{Rank}(A[:,S]) = |S|$.*

Proof. Without loss of generality, let $S = \{1, 2, \ldots, r\}$ and $S^c = \{r+1, \ldots, n\}$. Let us write $A_1 = A[S,:]$ and $A_2 = A[S^c,:]$. Since A is invertible, $\mathrm{RREF}(A) = I_n$. Hence, by Remark 2.2.19.4, there exists an invertible matrix P such that $PA = I_n$. Thus,

$$\begin{bmatrix} PA_1 & PA_2 \end{bmatrix} = P\begin{bmatrix} A_1 & A_2 \end{bmatrix} = PA = I_n = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & I_{n-r} \end{bmatrix}.$$

Thus, $PA_1 = \begin{bmatrix} I_r \\ \mathbf{0} \end{bmatrix}$ and $PA_2 = \begin{bmatrix} \mathbf{0} \\ I_{n-r} \end{bmatrix}$. So, using Corollary 2.3.5, $\mathsf{Rank}(A_1) = r$.

For the second part, let $B_1 = A[:, S]$, $B_2 = A[:, S^c]$ and let $\mathsf{Rank}(B_1) = t < s$. Then, by Remark 2.2.19.4, there exists an invertible matrix Q and a matrix C in RREF which has exactly t pivots such that

$$QB_1 = \text{RREF}(B_1) = \begin{bmatrix} C \\ \mathbf{0} \end{bmatrix}. \tag{2.3.5}$$

As $t < s$, QB_1 has at least one zero row. As $PA = I_n$ by Proposition 2.2.21 $AP = I_n$. Hence, $\begin{bmatrix} B_1P \\ B_2P \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} P = AP = I_n = \begin{bmatrix} I_s & \mathbf{0} \\ \mathbf{0} & I_{n-s} \end{bmatrix}$. Thus,

$$B_1P = \begin{bmatrix} I_s & \mathbf{0} \end{bmatrix} \text{ and } B_2P = \begin{bmatrix} \mathbf{0} & I_{n-s} \end{bmatrix}. \tag{2.3.6}$$

Hence, using Equations (2.3.5) and (2.3.6), we see that

$$\begin{bmatrix} CP \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} C \\ \mathbf{0} \end{bmatrix} P = QB_1P = Q\begin{bmatrix} I_s & \mathbf{0} \end{bmatrix} = \begin{bmatrix} Q & \mathbf{0} \end{bmatrix}.$$

Thus, Q has a zero row, a contradiction to Q being invertible. Hence, $\mathsf{Rank}(B_1) = s$. ∎

As a direct corollary of Theorem 2.3.6 and Proposition 2.3.7, we have the following result which improves Corollary 2.3.5.

Corollary 2.3.8. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If $\mathsf{Rank}(A) = r < n$ then, there exists an invertible matrix Q and $B \in \mathbb{M}_{m,r}(\mathbb{C})$ such that $AQ = \begin{bmatrix} B & \mathbf{0} \end{bmatrix}$, where $\mathsf{Rank}(B) = r$.*

Proof. By Theorem 2.3.6, there exist invertible matrices P and Q such that $PAQ = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. If $P^{-1} = \begin{bmatrix} B & C \end{bmatrix}$, where $B \in \mathbb{M}_{m,r}(\mathbb{C})$ and $C \in \mathbb{M}_{m,m-r}(\mathbb{C})$ then,

$$AQ = P^{-1}\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} B & C \end{bmatrix}\begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} B & \mathbf{0} \end{bmatrix}.$$

Now, by Proposition 2.3.7, $\mathsf{Rank}(B) = r = \mathsf{Rank}(A)$ as the matrix $P^{-1} = \begin{bmatrix} B & C \end{bmatrix}$ is an invertible matrix. Thus, the required result follows. ∎

As an application of Corollary 2.3.8, we have the following result.

Corollary 2.3.9. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$ and $B \in \mathbb{M}_{n,p}(\mathbb{C})$. Then, $\mathsf{Rank}(AB) \leq \mathsf{Rank}(B)$.*

Proof. Let $\mathsf{Rank}(B) = r$. Then, by Corollary 2.3.8, there exists an invertible matrix Q and a matrix $C \in \mathbb{M}_{n,r}(\mathbb{C})$ such that $BQ = \begin{bmatrix} C & \mathbf{0} \end{bmatrix}$ and $\mathsf{Rank}(C) = r$. Hence, $ABQ = A\begin{bmatrix} C & \mathbf{0} \end{bmatrix} = \begin{bmatrix} AC & \mathbf{0} \end{bmatrix}$. Thus, using Corollary 2.3.5 and Remark 2.3.2.2, we get

$$\mathsf{Rank}(AB) = \mathsf{Rank}(ABQ) = \mathsf{Rank}\left(\begin{bmatrix} AC & \mathbf{0} \end{bmatrix}\right) = \mathsf{Rank}(AC) \leq r = \mathsf{Rank}(B). \qquad ∎$$

We end this section by relating the rank of the sum of two matrices with sum of their ranks.

Proposition 2.3.10. *Let $A, B \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, prove that* $\mathsf{Rank}(A+B) \le \mathsf{Rank}(A) + \mathsf{Rank}(B)$. *In particular, if $A = \sum\limits_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*$, for some $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{C}$, for $1 \le i \le k$, then* $\mathsf{Rank}(A) \le k$.

Proof. Let $\mathsf{Rank}(A) = r$. Then, there exists an invertible matrix P and a matrix $A_1 \in \mathbb{M}_{r,n}(\mathbb{C})$ such that $PA = \text{RREF}(A) = \begin{bmatrix} A_1 \\ \mathbf{0} \end{bmatrix}$. Then,

$$P(A+B) = PA + PB = \begin{bmatrix} A_1 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_1 + B_1 \\ B_2 \end{bmatrix}.$$

Now using Corollary 2.3.5, Remark 2.3.2.4 and the condition $\mathsf{Rank}(A) = \mathsf{Rank}(A_1) = r$, the number of rows of A_1, we have

$$\mathsf{Rank}(A+B) = \mathsf{Rank}(P(A+B)) \le r + \mathsf{Rank}(B_2) \le r + \mathsf{Rank}(B) = \mathsf{Rank}(A) + \mathsf{Rank}(B).$$

Thus, the required result follows. The other part follows, as $\mathsf{Rank}(\mathbf{x}_i \mathbf{y}_i^*) = 1$, for $1 \le i \le k$. ∎

EXERCISE **2.3.11.**

1. *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 $B = PAQ$.*

2. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If* $\mathsf{Rank}(A) = r$ *then, prove that $A = BC$, where $B \in \mathbb{M}_{m,r}(\mathbb{C})$ and $C \in \mathbb{M}_{r,n}(\mathbb{C})$ and* $\mathsf{Rank}(B) = \mathsf{Rank}(C) = r$. *Now, use matrix product to give the existence of $\mathbf{x}_i \in \mathbb{C}^m$ and $\mathbf{y}_i \in \mathbb{C}^n$ such that $A = \sum\limits_{i=1}^{r} \mathbf{x}_i \mathbf{y}_i^*$.*

3. *If* $\mathsf{Rank}(A) = r$ *then prove that there exist invertible matrices B_i, C_i such that $B_1 A = \begin{bmatrix} R_1 & R_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $AC_1 = \begin{bmatrix} S_1 & \mathbf{0} \\ S_3 & \mathbf{0} \end{bmatrix}$ and $B_2 A C_2 = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where the $(1,1)$ block of each matrix has size $r \times r$. Also, prove that A_1 is an invertible matrix.*

4. *Prove that if* $\mathsf{Rank}(A) = \mathsf{Rank}(AB)$ *then $A = ABX$, for some matrix X. Similarly, if* $\mathsf{Rank}(A) = \mathsf{Rank}(BA)$ *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 an invertible matrix R such that $P(AB)R = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Use the above result to show that C is invertible. Then $X = R \begin{bmatrix} C^{-1}A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1}$ gives the required result.]*

   Ans: $PAQ = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \Rightarrow AQ = \begin{bmatrix} A_{11} & \mathbf{0} \\ A_{21} & \mathbf{0} \end{bmatrix} \Rightarrow BAQ = \begin{bmatrix} B_{11}A_{11} + B_{12}A_{21} & \mathbf{0} \\ B_{21}A_{11} + B_{22}A_{21} & \mathbf{0} \end{bmatrix}$. Thus, there exists an invertible matrix P_1 such that $P_1 BAQ = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ for some invertible matrix C. Define $Y = P^{-1} \begin{bmatrix} A_1 C^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} P_1$ and compute YBA.

5. *Let M and N be invertible matrices. Then prove that* $\mathsf{Rank}(MAN) = \mathsf{Rank}(A)$.

6. *Let A be an $m \times n$ matrix with $\mathsf{Rank}(A) = m$. Then prove the following:*
   (a) *There exists an invertible matrix P and a permutation matrix Q such that $PAQ = \begin{bmatrix} I_m & \mathbf{0} \end{bmatrix}$.*
   (b) *As Q is a permutation matrix Q is an orthogonal matrix, i.e., $QQ^T = I = Q^TQ$.*
   (c) *$P(AA^T)P^T = (PAQ)(Q^TA^TP^T) = (PAQ)(PAQ)^T = I_m$. Hence $\mathsf{Rank}(AA^T) = m$.*

2.4 Solution set of a Linear System

Definition 2.4.1. Consider the linear system $A\mathbf{x} = \mathbf{b}$. If $\text{RREF}([A \;\; \mathbf{b}]) = [C \;\; \mathbf{d}]$. Then, the variables corresponding to the pivotal columns of C are called the **basic** variables and the variables that are not basic are called **free** variables.

Example 2.4.2. 1. If the system $A\mathbf{x} = \mathbf{b}$ in n variables is consistent and $\text{RREF}(A)$ has r nonzero rows then, $A\mathbf{x} = \mathbf{b}$ has r basic variables and $n - r$ free variables.

2. Let $\text{RREF}([A \;\; \mathbf{b}]) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Hence, x and y are basic variables and z is the free variable. Thus, the solution set of $A\mathbf{x} = \mathbf{b}$ is given by

$$\left\{ [x, y, z]^T \mid [x, y, z] = [1, 2 - z, z] = [1, 2, 0] + z[0, -1, 1], \text{ with } z \text{ arbitrary.} \right\}$$

3. Let $\text{RREF}([A \quad \mathbf{b}]) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Then, the system $A\mathbf{x} = \mathbf{b}$ has no solution as $(\text{RREF}([A \;\; \mathbf{b}]))[3, :] = [0\,0\,0\,1]$.

We now prove the main result in the theory of linear systems. Before doing so, we look at the following example.

Example 2.4.3. Consider a linear system $A\mathbf{x} = \mathbf{b}$. Suppose $\text{RREF}([A\,\mathbf{b}]) = [C\,\mathbf{d}]$, where

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

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

1. C has 4 pivotal columns, namely, the columns $1, 2, 5$ and 6. Thus, x_1, x_2, x_5 and x_6 are basic variables.

2. Hence, the remaining variables x_3, x_4 and x_7 are free variables.

Therefore, the solution set is given by

$$\begin{bmatrix} x_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_6 \\ \mathbf{x}_7 \end{bmatrix} = \begin{bmatrix} 8 - 2x_3 + x_4 - 2x_7 \\ 1 - x_3 - 3x_4 - 5x_7 \\ x_3 \\ x_4 \\ 2 + x_7 \\ 4 - x_7 \\ x_7 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \\ 0 \\ 0 \\ 2 \\ 4 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix},$$

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 $\mathbf{u}_3 = \begin{bmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$. In this example, verify that $C\mathbf{x}_0 = \mathbf{d}$, and for $1 \le i \le 3$, $C\mathbf{u}_i = \mathbf{0}$. Hence, it follows that $A\mathbf{x}_0 = \mathbf{d}$, and for $1 \le i \le 3$, $A\mathbf{u}_i = \mathbf{0}$.

Theorem 2.4.4. *Let $A\mathbf{x} = \mathbf{b}$ be a linear system in n variables with RREF$([A \;\; \mathbf{b}]) = [C \;\; \mathbf{d}]$ with $\mathsf{Rank}(A) = r$ and $\mathsf{Rank}([A\ \mathbf{b}]) = r_a$.*

1. *Then, the system $A\mathbf{x} = \mathbf{b}$ is inconsistent if $r < r_a$*

2. *Then, the system $A\mathbf{x} = \mathbf{b}$ is consistent if $r = r_a$.*

   (a) *Further, $A\mathbf{x} = \mathbf{b}$ has* A UNIQUE SOLUTION *if $r = n$.*

   (b) *Further, $A\mathbf{x} = \mathbf{b}$ has* INFINITE NUMBER OF SOLUTIONS *if $r < n$. In this case, there exist vectors $\mathbf{x}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r} \in \mathbb{R}^n$ with $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$, for $1 \le i \le n - r$. Furthermore, the solution set is given by*

$$\{\mathbf{x}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_{n-r}\mathbf{u}_{n-r} \mid k_i \in \mathbb{C},\ 1 \le i \le n - r\}.$$

Proof. PART 1: As $r < r_a$, by Remark 2.2.19.5 $([C \;\; \mathbf{d}])[r+1, :] = [\mathbf{0}^T \; 1]$. Note that this row corresponds to the linear equation

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

which clearly has no solution. Thus, by definition and Theorem 2.1.17, $A\mathbf{x} = \mathbf{b}$ is inconsistent.

PART 2: As $r = r_a$, by Remark 2.2.19.5, $[C \;\; \mathbf{d}]$ doesn't have a row of the form $[\mathbf{0}^T \; 1]$. Further, the number of pivots in $[C \;\; \mathbf{d}]$ and that in C is same, namely, r pivots. Suppose the pivots appear in columns $i_1, \ldots, i_r$ with $1 \le i_1 < \cdots < i_r \le n$. Thus, the variables x_{i_j}, for $1 \le j \le r$, are basic variables and the remaining $n - r$ variables, say $x_{t_1}, \ldots, x_{t_{n-r}}$, are free variables with $t_1 < \cdots < t_{n-r}$. Since C is in RREF, in terms of the free variables and basic variables, the ℓ-th row of $[C \;\; \mathbf{d}]$, for $1 \le \ell \le r$, corresponds to the equation

$$x_{i_\ell} + \sum_{k=1}^{n-r} c_{\ell t_k} x_{t_k} = d_\ell \Leftrightarrow x_{i_\ell} = d_\ell - \sum_{k=1}^{n-r} c_{\ell t_k} x_{t_k}.$$

Thus, the system $C\mathbf{x} = \mathbf{d}$ is consistent. Hence, by Theorem 2.1.17 the system $A\mathbf{x} = \mathbf{b}$ is consistent and the solution set of the system $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ are the same. Therefore, the solution set of the system $C\mathbf{x} = \mathbf{d}$ (or equivalently $A\mathbf{x} = \mathbf{b}$) is given by

$$\begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_r} \\ x_{t_1} \\ x_{t_2} \\ \vdots \\ x_{t_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 - \sum\limits_{k=1}^{n-r} c_{1t_k} x_{t_k} \\ \vdots \\ d_r - \sum\limits_{k=1}^{n-r} c_{rt_k} x_{t_k} \\ x_{t_1} \\ x_{t_2} \\ \vdots \\ x_{t_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_{t_1} \begin{bmatrix} c_{1t_1} \\ \vdots \\ c_{rt_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_{t_2} \begin{bmatrix} c_{1t_2} \\ \vdots \\ c_{rt_2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_{t_{n-r}} \begin{bmatrix} c_{1t_{n-r}} \\ \vdots \\ c_{rt_{n-r}} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{2.4.7}$$

PART 2A: As $r = n$, there are no free variables. Hence, $x_i = d_i$, for $1 \le i \le n$, is the unique solution.

PART 2B: Define $\mathbf{x}_0 = \begin{bmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and $\mathbf{u}_1 = \begin{bmatrix} c_{1t_1} \\ \vdots \\ c_{rt_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{u}_{n-r} = \begin{bmatrix} c_{1t_{n-r}} \\ \vdots \\ c_{rt_{n-r}} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$. Then, it can be easily verified that $A\mathbf{x}_0 = \mathbf{b}$ and, for $1 \le i \le n-r$, $A\mathbf{u}_i = \mathbf{0}$. Also, by Equation (2.4.7) the solution set has indeed the required form, where k_i corresponds to the free variable x_{t_i}. As there is at least one free variable the system has infinite number of solutions. Thus, the proof of the theorem is complete. ■

EXERCISE **2.4.5.** *Consider the linear system given below. Use GJE to find the RREF of it's augmented matrix. Now, use the technique used in the previous theorem to find the solution of the linear system*

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

Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, $\mathsf{Rank}(A) \le m$. Thus, using Theorem 2.4.4 the next result follows.

Corollary 2.4.6. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If $\mathsf{Rank}(A) = r < \min\{m, n\}$ then $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions. In particular, if $m < n$, then $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions. Hence, in either case, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has at least one non-trivial solution.*

Remark 2.4.7. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, Theorem 2.4.4 implies that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathsf{Rank}(A) = \mathsf{Rank}([A\ \mathbf{b}])$. Further, the vectors associated to the free variables in Equation (2.4.7) are solutions to the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.*

We end this subsection with some applications.

Example 2.4.8. 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 Euclidean plane is given by $a(x^2 + y^2) + bx + cy + d = 0$, where a, b, c and d are variables. Since this curve passes through the given points, we get a homogeneous system in 3 equations and 4 variables, namely $\begin{bmatrix} (-1)^2+4^2 & -1 & 4 & 1 \\ (0)^2+1^2 & 0 & 1 & 1 \\ 1^2+4^2 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \mathbf{0}$. Solving this system, we get $[a, b, c, d] = [\frac{3}{13}d, 0, -\frac{16}{13}d, d]$. Hence, choosing $d = 13$, the required circle is given by $3(x^2 + y^2) - 16y + 13 = 0$.

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 space is given by $ax + by + cz + d = 0$, where a, b, c and d are variables. Since this plane passes through the 3 given points, we get a homogeneous system in 3 equations and 4 variables. So, it has a non-trivial solution, namely $[a, b, c, d] = [-\frac{4}{3}d, -\frac{d}{3}, -\frac{2}{3}d, d]$. Hence, choosing $d = 3$, the required plane is given by $-4x - y + 2z + 3 = 0$.

3. Let $A = \begin{bmatrix} 2 & 3 & 4 \\ 0 & -1 & 0 \\ 0 & -3 & 4 \end{bmatrix}$. Then, find a non-trivial solution of $A\mathbf{x} = 2\mathbf{x}$. Does there exist a nonzero vector $\mathbf{y} \in \mathbb{R}^3$ such that $A\mathbf{y} = 4\mathbf{y}$?

Solution: Solving for $A\mathbf{x} = 2\mathbf{x}$ is equivalent to solving $(A - 2I)\mathbf{x} = \mathbf{0}$. The augmented matrix of this system equals $\begin{bmatrix} 0 & 3 & 4 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 4 & 2 & 0 \end{bmatrix}$. Verify that $\mathbf{x}^T = [1, 0, 0]$ is a nonzero solution. For the other part, the augmented matrix for solving $(A - 4I)\mathbf{y} = \mathbf{0}$ equals $\begin{bmatrix} -2 & 3 & 4 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & -3 & 0 & 0 \end{bmatrix}$. Thus, verify that $\mathbf{y}^T = [2, 0, 1]$ is a nonzero solution.

Exercise **2.4.9.** *1. Let $A \in \mathbb{M}_n(\mathbb{C})$. If $A^2\mathbf{x} = \mathbf{0}$ has a non trivial solution then show that $A\mathbf{x} = \mathbf{0}$ also has a non trivial solution.*

2. Prove that 5 distinct points are needed to specify a general conic, namely, $ax^2 + by^2 + cxy + dx + ey + f = 0$, in the Euclidean plane.

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

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

5. *Find the condition(s) on x,y,z so that the systems given below (in the variables a,b and c) is consistent?*

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

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

---

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

7. *Solve the linear systems*
   $x+y+z+w=0,\ x-y+z+w=0$ *and* $-x+y+3z+3w=0,$ *and*
   $x+y+z=3,\ x+y-z=1,\ x+y+4z=6$ *and* $x+y-4z=-1.$

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

9. *Consider the linear system $A\mathbf{x}=\mathbf{b}$ in m equations and 3 variables. Then, for each of the given solution set, determine the possible choices of m? Further, for each choice of m, determine a choice of A and $\mathbf{b}$.*

   (a) $(1,1,1)^T$ *is the only solution.*

   (b) $\{(1,1,1)^T+c(1,2,1)^T | c\in\mathbb{R}\}$ *as the solution set.*

   (c) $\{c(1,2,1)^T | c\in\mathbb{R}\}$ *as the solution set.*

   (d) $\{(1,1,1)^T+c(1,2,1)^T+d(2,2,-1)^T | c,d\in\mathbb{R}\}$ *as the solution set.*

   (e) $\{c(1,2,1)^T+d(2,2,-1)^T | c,d\in\mathbb{R}\}$ *as the solution set.*

2.5 Square Matrices and Linear Systems

In this section the coefficient matrix of the linear system $A\mathbf{x}=\mathbf{b}$ will be a square matrix. We start with proving a few equivalent conditions that relate different ideas.

Theorem 2.5.1. *Let $A\in\mathbb{M}_n(\mathbb{C})$. Then, the following statements are equivalent.*

1. *A is invertible.*

2. *$RREF(A)=I_n$.*

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

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

5. $\mathsf{Rank}(A) = n$.

Proof. $1 \Leftrightarrow 2$ Already done in Proposition 2.2.21.

$2 \Leftrightarrow 3$ Again, done in Proposition 2.2.21.

$3 \Longrightarrow 4$ Let $A = E_1 \cdots E_k$, for some elementary matrices $E_1, \ldots, E_k$. Then, by previous equivalence A is invertible. So, A^{-1} exists and $A^{-1}A = I_n$. Hence, if $\mathbf{x}_0$ is any solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$ then,

$$\mathbf{x}_0 = I_n \cdot \mathbf{x}_0 = (A^{-1}A)\mathbf{x}_0 = A^{-1}(A\mathbf{x}_0) = A^{-1}\,\mathbf{0} = \mathbf{0}.$$

Thus, $\mathbf{0}$ is the only solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

$4 \Longrightarrow 5$ Let if possible $\mathsf{Rank}(A) = r < n$. Then, by Corollary 2.4.6, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has infinitely many solution. A contradiction. Thus, A has full rank.

$5 \Longrightarrow 2$ Suppose $\mathsf{Rank}(A) = n$. So, RREF(A) has n pivotal columns. But, RREF(A) has exactly n columns and hence each column is a pivotal column. Thus, RREF$(A) = I_n$. ∎

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

Theorem 2.5.2. *The following statements are equivalent for* $A \in \mathbb{M}_n(\mathbb{C})$.

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. $1 \Longrightarrow 2$ Note that $\mathbf{x}_0 = A^{-1}\mathbf{b}$ is the unique solution of $A\mathbf{x} = \mathbf{b}$.

$2 \Longrightarrow 3$ The system is consistent as $A\mathbf{x} = \mathbf{b}$ has a solution.

$3 \Longrightarrow 1$ For $1 \le i \le n$, define $\mathbf{e}_i^T = I_n[i, :]$. By assumption, the linear system $A\mathbf{x} = \mathbf{e}_i$ has a solution, say $\mathbf{x}_i$, for $1 \le i \le n$. Define a matrix $B = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$. 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, $n = \mathsf{Rank}(I_n) = \mathsf{Rank}(AB) \le \mathsf{Rank}(A)$ and hence $\mathsf{Rank}(A) = n$. Thus, by Theorem 2.5.1, A is invertible. ∎

We now give an immediate application of Theorem 2.5.2 and Theorem 2.5.1 without proof.

Theorem 2.5.3. *The following two statements cannot hold together for* $A \in \mathbb{M}_n(\mathbb{C})$.

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

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

As an immediate consequence of Theorem 2.5.1, the readers should prove that one needs to compute either the left or the right inverse to prove invertibility of $A \in \mathbb{M}_n(\mathbb{C})$.

Corollary 2.5.4. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then the following holds.*

1. *If there exists C such that $CA = I_n$ then A^{-1} exists.*

2. *If there exists B such that $AB = I_n$ then A^{-1} exists.*

Exercise **2.5.5.**

1. *Let A be a square matrix. Then, prove that A is invertible $\Leftrightarrow$ A^T is invertible $\Leftrightarrow$ A^TA is invertible $\Leftrightarrow$ AA^T is invertible.*

2. [**Theorem of the Alternative**] *The following two statements cannot hold together for $A \in \mathbb{M}_n(\mathbb{C})$ and $\mathbf{b} \in \mathbb{R}^n$.*

   (a) *The system $A\mathbf{x} = \mathbf{b}$ has a solution.*

   (b) *The system $\mathbf{y}^TA = \mathbf{0}^T, \mathbf{y}^T\mathbf{b} \neq 0$ has a solution.*

3. *Let A and B be two matrices having positive entries and of orders $1 \times n$ and $n \times 1$, respectively. Which of BA or AB is invertible? Give reasons.*

4. *Let $A \in \mathbb{M}_{n,m}(\mathbb{C})$ and $B \in \mathbb{M}_{m,n}(\mathbb{C})$.*

   (a) *Then, prove that $I - BA$ is invertible if and only if $I - AB$ is invertible [use Theorem 2.5.1.4].*

   (b) *If $I - AB$ is invertible then, prove that $(I - BA)^{-1} = I + B(I - AB)^{-1}A$.*

   (c) *If $I - AB$ is invertible then, prove that $(I - BA)^{-1}B = B(I - AB)^{-1}$.*

   (d) *If A, B and $A + B$ are invertible then, prove that $(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B$.*

5. *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 $B = RREF(A)$. Then,*

      i. $B[4, :] = [0, 0, 0, 0]$.

      ii. $B[4, :] = [0, 0, 0, 1]$.

      iii. $B[3, :] = [0, 0, 0, 0]$.

      iv. $B[3, :] = [0, 0, 0, 1]$.

      v. $B[3, :] = [0, 0, 1, \alpha]$, *where α is any real number.*

2.5.1 Determinant

Recall the notations used in Section 1.3.1 on Page 19 . If $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 7 \end{bmatrix}$ then $A(1 \mid 2) = \begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix}$ and $A(\{1, 2\} \mid \{1, 3\}) = [4]$. We are ready to give an inductive definition of the determinant of a square matrix. The advanced students can find an alternate definition of the determinant in Appendix 7.2.22, where it is proved that the definition given below corresponds to the expansion of determinant along the first row.

Definition 2.5.6. Let A be a square matrix of order n. Then, the determinant of A, denoted $\det(A)$ (or $\mid A \mid$) is defined by

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

Example 2.5.7. 1. Let $A = [-2]$. Then, $\det(A) = \mid A \mid = -2$.

2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, $\det(A) = \mid A \mid = a \ \det(A(1 \mid 1)) - b \ \det(A(1 \mid 2)) = ad - bc$.

   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 $A = [a_{ij}]$ be a 3×3 matrix. Then,

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

   For $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$, $\mid A \mid = 1 \cdot \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} + 3 \cdot \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 4 - 2(3) + 3(1) = 1.$

EXERCISE **2.5.8.** *Find the determinant of the following matrices.*

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

Definition 2.5.9. A matrix A is said to be a SINGULAR if $\det(A) = 0$ and is called NON-SINGULAR if $\det(A) \neq 0$.

The next result relates the determinant with row operations. For proof, see Appendix 7.3.

Theorem 2.5.10. *Let A be an $n \times n$ matrix.*

1. *If $B = E_{ij}A$, for $1 \leq i \neq j \leq n$, then* $\det(B) = -\det(A)$.

2. *If $B = E_i(c)A$, for $c \neq 0, 1 \leq i \leq n$, then* $\det(B) = c\det(A)$.

3. *If $B = E_{ij}(c)A$, for $c \neq 0$ and $1 \leq i \neq j \leq n$, then* $\det(B) = \det(A)$.

4. *If $A[i,:]^T = \mathbf{0}$, for $1 \leq i, j \leq n$ then* $\det(A) = 0$.

5. *If* $A[i,:] = A[j,:]$ *for* $1 \leq i \neq j \leq n$ *then* $\det(A) = 0$.

6. *If* A *is a triangular matrix with* $d_1, \ldots, d_n$ *on the diagonal then* $\det(A) = d_1 \cdots d_n$.

As $\det(I_n) = 1$, we have the following result.

Corollary 2.5.11. *Fix a positive integer* n.

1. *Then* $\det(E_{ij}) = -1$.

2. *If* $c \neq 0$ *then* $\det(E_k(c)) = c$.

3. *If* $c \neq 0$ *then* $\det(E_{ij}(c)) = 1$.

Example 2.5.12. Let $A = \begin{bmatrix} 2 & 2 & 6 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{bmatrix}$. Then $A \overset{E_1(\frac{1}{2})}{\to} \begin{vmatrix} 1 & 1 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{vmatrix} \overset{E_{21}(-1)E_{31}(-1)}{\to} \begin{vmatrix} 1 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{vmatrix}$.
Thus, using Theorem 2.5.10, $\det(A) = 2 \cdot (1 \cdot 2 \cdot (-1)) = -4$, where the first 2 appears from the elementary matrix $E_1(\frac{1}{2})$.

EXERCISE **2.5.13.** *Prove the following without computing the determinant (use Theorem 2.5.10).*

1. *Let* $A = \begin{bmatrix} \mathbf{u} & \mathbf{v} & 2\mathbf{u} + 3\mathbf{v} \end{bmatrix}$, *where* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^3$. *Then,* $\det(A) = 0$.

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

By Theorem 2.5.10.6 $\det(I_n) = 1$. The next result about the determinant of elementary matrices is an immediate consequence of Theorem 2.5.10 and hence the proof is omitted.

Remark 2.5.14. *Theorem 2.5.10.1 implies that the determinant can be calculated by expanding along any row. Hence, the readers are advised to verify that*

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(A(k \mid j)), \quad \text{for } 1 \leq k \leq n.$$

Example 2.5.15. Using Remark 2.5.14, one has

$$\begin{vmatrix} 2 & 2 & 6 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 2 & 0 \\ 1 & 2 & 1 & 1 \end{vmatrix} = (-1)^{2+3} \cdot 2 \cdot \begin{vmatrix} 2 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{vmatrix} + (-1)^{2+4} \cdot \begin{vmatrix} 2 & 2 & 6 \\ 0 & 1 & 2 \\ 1 & 2 & 1 \end{vmatrix} = -2 \cdot 1 + (-8) = -10.$$

2.5.2 Adjugate (classical Adjoint) of a Matrix

Definition 2.5.16. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the **cofactor** matrix, denoted $\mathsf{Cof}(A)$, is an $\mathbb{M}_n(\mathbb{C})$ matrix with $\mathsf{Cof}(A) = [C_{ij}]$, where

$$C_{ij} = (-1)^{i+j} \det\left(A(i \mid j)\right), \text{ for } 1 \leq i, j \leq n.$$

And, the **Adjugate** (classical Adjoint) of A, denoted $Adj(A)$, equals $\mathsf{Cof}^T(A)$.

Example 2.5.17. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 4 \end{bmatrix}$.

1. Then,

$$\begin{aligned} Adj(A) &= \mathsf{Cof}^T(A) = \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} \\ &= \begin{bmatrix} (-1)^{1+1}\det(A(1|1)) & (-1)^{2+1}\det(A(2|1)) & (-1)^{3+1}\det(A(3|1)) \\ (-1)^{1+2}\det(A(1|2)) & (-1)^{2+2}\det(A(2|2)) & (-1)^{3+2}\det(A(3|2)) \\ (-1)^{1+3}\det(A(1|3)) & (-1)^{2+3}\det(A(2|3)) & (-1)^{3+3}\det(A(3|3)) \end{bmatrix} \\ &= \begin{bmatrix} 10 & -2 & -7 \\ -7 & 1 & 5 \\ 1 & 0 & -1 \end{bmatrix}. \end{aligned}$$

Now, verify that $AAdj(A) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \det(A) & 0 & 0 \\ 0 & \det(A) & 0 \\ 0 & 0 & \det(A) \end{bmatrix} = Adj(A)A.$

2. Consider $xI_3 - A = \begin{bmatrix} x-1 & -2 & -3 \\ -2 & x-3 & -1 \\ -1 & -2 & x-4 \end{bmatrix}$. Then,

$$\begin{aligned} Adj(xI - A) &= \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} x^2-7x+10 & 2x-2 & 3x-7 \\ 2x-7 & x^2-5x+1 & x+5 \\ x+1 & 2x & x^2-4x-1 \end{bmatrix} \\ &= x^2 I + x\begin{bmatrix} -7 & 2 & 3 \\ 2 & -5 & 1 \\ 1 & 2 & -4 \end{bmatrix} + Adj(A) = x^2 I + Bx + C(\text{say}). \end{aligned}$$

Hence, we observe that $Adj(xI - A) = x^2 I + Bx + C$ is a polynomial in x with coefficients as matrices. Also, note that $(xI - A)Adj(xI - A) = (x^3 - 8x^2 + 10x - \det(A))I_3$. Thus, we see that

$$(xI - A)(x^2 I + Bx + C) = (x^3 - 8x^2 + 10x - \det(A))I_3.$$

That is, we have obtained a matrix equality and hence, replacing x by A makes sense. But, then the LHS is $\mathbf{0}$. So, for the RHS to be zero, we must have $A^3 - 8A^2 + 10A - \det(A)I = \mathbf{0}$ (this equality is famously known as the Cayley-Hamilton Theorem).

The next result relates adjugate matrix with the inverse, in case $\det(A) \neq 0$.

Theorem 2.5.18. *Let $A \in \mathbb{M}_n(\mathbb{C})$.*

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

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

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

$$\textit{whenever } \det(A) \neq 0 \ \textit{ one has } A^{-1} = \frac{1}{\det(A)}Adj(A). \tag{2.5.1}$$

Proof. Part 1: It follows directly from Remark 2.5.14 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 with $B[\ell, :] = A[i, :]$ and $B[t, :] = A[t, :]$, for $t \neq \ell$. As $\ell \neq i$, $B[\ell, :] = B[i, :]$ and thus, by Theorem 2.5.10.5, $\det(B) = 0$. As $A(\ell \mid j) = B(\ell \mid j)$, for $1 \leq j \leq n$, using Remark 2.5.14

$$\begin{aligned} 0 = \det(B) &= \sum_{j=1}^{n}(-1)^{\ell+j} b_{\ell j} \det\big(B(\ell \mid j)\big) = \sum_{j=1}^{n}(-1)^{\ell+j} a_{ij} \det\big(B(\ell \mid j)\big) \\ &= \sum_{j=1}^{n}(-1)^{\ell+j} a_{ij} \det\big(A(\ell \mid j)\big) = \sum_{j=1}^{n} a_{ij} C_{\ell j}. \end{aligned} \tag{2.5.2}$$

This completes the proof of Part 2.

Part 3: Using Equation (2.5.2) and Remark 2.5.14, observe that

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

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 Proposition 2.2.21, $A^{-1} = \frac{1}{\det(A)}Adj(A)$. ∎

Example 2.5.19. 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 2.5.18.3, $A^{-1} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 \\ 1/2 & 3/2 & -1/2 \end{bmatrix}$.

Let A be a non-singular matrix. Then, by Theorem 2.5.18.3, $A^{-1} = \frac{1}{\det(A)}Adj(A)$. Thus $A\big(Adj(A)\big) = \big(Adj(A)\big)A = \det(A)\ I_n$ and this completes the proof of the next result

Corollary 2.5.20. *Let A be a non-singular matrix. Then,*

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

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

Theorem 2.5.21. *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 $A^{-1} = \frac{1}{\det(A)}Adj(A)$.

Now, let us assume that A is invertible. Then, using Theorem 2.5.1, $A = E_1 \cdots E_k$, a product of elementary matrices. Also, by Corollary 2.5.11, $\det(E_i) \neq 0$, for $1 \leq i \leq k$. Thus, a repeated application of Parts 1, 2 and 3 of Theorem 2.5.10 gives $\det(A) \neq 0$. ∎

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

Theorem 2.5.22. *Let A be a square matrix. Then,* $\det(A) = \det(A^T)$.

Proof. If A is a non-singular, Corollary 2.5.20 gives $\det(A) = \det(A^T)$.

If A is singular then, by Theorem 2.5.21, A is not invertible. So, A^T is also not invertible and hence by Theorem 2.5.21, $\det(A^T) = 0 = \det(A)$. ■

The next result relates the determinant of product of two matrices with their determinants.

Theorem 2.5.23. *Let A and B be square matrices of order n. Then,*

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

Proof. Case 1: Let A be non-singular. Then, by Theorem 2.5.18.3, A is invertible and by Theorem 2.5.1, $A = E_1 \cdots E_k$, a product of elementary matrices. Thus, a repeated application of Parts $1, 2$ and 3 of Theorem 2.5.10 gives the desired result as

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

Case 2: Let A be singular. Then, by Theorem 2.5.21 A is not invertible. So, by Proposition 2.2.21 there exists an invertible matrix P such that $PA = \begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}$. So, $A = P^{-1}\begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}$. As P is invertible, using Part 1, we have

$$\begin{aligned}
\det(AB) &= \det\left(\left(P^{-1}\begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}\right) B\right) = \det\left(P^{-1}\begin{bmatrix} C_1 B \\ \mathbf{0} \end{bmatrix}\right) = \det(P^{-1}) \cdot \det\left(\begin{bmatrix} C_1 B \\ \mathbf{0} \end{bmatrix}\right) \\
&= \det(P) \cdot 0 = 0 = 0 \cdot \det(B) = \det(A)\det(B).
\end{aligned}$$

Thus, the proof of the theorem is complete. ■

Example 2.5.24. Let A be an orthogonal matrix then, by definition, $AA^T = I$. Thus, by Theorems 2.5.23 and 2.5.22

$$1 = \det(I) = \det(AA^T) = \det(A)\det(A^T) = \det(A)\det(A) = (\det(A))^2.$$

Hence $\det A = \pm 1$. In particular, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $I = AA^T = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix}$.

1. Thus, $a^2 + b^2 = 1$ and hence there exists $\theta \in [-pi, \ \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$.
2. As $ac + bd = 0$, we get $c = r\sin\theta$ and $d = -r\cos\theta$, for some $r \in \mathbb{R}$. But, $c^2 + d^2 = 1$ implies that either $c = \sin\theta$ and $d = -\cos\theta$ or $c = -\sin\theta$ and $d = \cos\theta$.
3. Thus, $A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$ or $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$.
4. For $A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$, $\det(A) = -1$. Then A represents a reflection about the line $y = m\mathbf{x}$. Determine m? (see Exercise 2.2b).

5. For $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, $\det(A) = 1$. Then A represents a rotation through the angle α. Determine α? (see Exercise 2.2a).

EXERCISE **2.5.25.** *1. Let $A \in \mathbb{M}_n(\mathbb{C})$ be an upper triangular matrix with nonzero entries on the diagonal. Then, prove that A^{-1} is also an upper triangular matrix.*

2. [LU **decomposition of an invertible matrix**] *Let $A \in \mathbb{M}_n(\mathbb{R})$ such that $\det(A[S|S]) \neq 0$ for all $S \subseteq \{1, 2, \ldots, n\}$. Then there exists an invertible lower triangular matrix L such that LA is an invertible upper triangular matrix. The proof uses the following ideas.*

   (a) *Let $\mathbf{u} \in \mathbb{R}^n$ with $\mathbf{u}^T = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}$ and $u_1 \neq \mathbf{0}$. Then there exists an invertible lower triangular matrix L such that $L\mathbf{u} = u_1\mathbf{e}_1$.*

   Ans: Define $L = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{x} & I_{n-1} \end{bmatrix}$, where $\mathbf{x} = -\frac{1}{u_1}\begin{bmatrix} u_2 \\ \vdots \\ u_n \end{bmatrix}$. Then verify that $L\mathbf{u} = u_1\mathbf{e}_1$.

   (b) *As $a_{11} = \det(A[S|S]) \neq 0$ for $S = \{1\}$, Part 2a gives an invertible lower triangular matrix L_1 with $L_1 A = \begin{bmatrix} a_{11} & * \\ \mathbf{0} & A_1 \end{bmatrix}$.*

   (c) *Deduce that $\det(A) = a_{11}\det(A_1)$. So $\det(A_1[S|S]) \neq 0$ for all $S \subseteq \{1, 2, \ldots, n-1\}$.*

   (d) *Now, use induction to get $L_2 \in \mathbb{M}_{n-1}(\mathbb{R})$, an invertible lower triangular matrix, such that $L_2 A_1 = T_1$, an invertible upper triangular matrix.*

   (e) *Define $L = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & L_2 \end{bmatrix} L_1$. Then $LA = \begin{bmatrix} \alpha & * \\ \mathbf{0} & T_1 \end{bmatrix}$, is an upper triangular matrix with L as an invertible lower triangular matrix.*

   (f) *Since L^{-1} is also a lower triangular matrix, $A = L^{-1}\begin{bmatrix} \alpha & * \\ \mathbf{0} & T_1 \end{bmatrix}$. Thus, A is a product of a lower triangular invertible matrix and an upper triangular invertible matrix $U = \begin{bmatrix} \alpha & * \\ \mathbf{0} & T_1 \end{bmatrix}$.*

---

3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, $\det(A) = 0$ if*

   (a) *either $A[i,:]^T = \mathbf{0}^T$ or $A[:,i] = \mathbf{0}$, for some $i, 1 \leq i \leq n$,*

   (b) *or $A[i,:] = cA[j,:]$, for some $c \in \mathbb{C}$ and for some $i \neq j$,*

   (c) *or $A[:,i] = cA[:,j]$, for some $c \in \mathbb{C}$ and for some $i \neq j$,*

   (d) *or $A[i,:] = c_1 A[j_1,:] + c_2 A[j_2,:] + \cdots + c_k A[j_k,:]$, for some rows $i, j_1, \ldots, j_k$ of A and some c_i's in $\mathbb{C}$,*

   (e) *or $A[:,i] = c_1 A[:,j_1] + c_2 A[:,j_2] + \cdots + c_k A[:,j_k]$, for some columns $i, j_1, \ldots, j_k$ of A and some c_i's in $\mathbb{C}$.*

4. *Let $A = \begin{bmatrix} a & b & c \\ e & f & g \\ h & j & \ell \end{bmatrix}$ and $B = \begin{bmatrix} a & e & 10^2 a + 10e + h \\ b & f & 10^2 b + 10f + j \\ c & g & 10^2 c + 10g + \ell \end{bmatrix}$, where $a, b \ldots, \ell \in \mathbb{C}$. Without computing deduce that $\det(A) = \det(B)$. Hence, conclude that 17 divides $\begin{vmatrix} 3 & 1 & 1 \\ 4 & 8 & 1 \\ 0 & 7 & 9 \end{vmatrix}$.*

2.5.3 Cramer's Rule

We start with a corollary which is a direct application of Theorems 2.5.2 and 2.5.21.

Corollary 2.5.26. *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 2.5.27 (Cramer's Rule)**.** *Let A be an $n \times n$ non-singular matrix. Then, the unique solution of the linear system $A\mathbf{x} = \mathbf{b}$ with $\mathbf{x}^T = [x_1, \ldots, x_n]$ is given by*

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

where A_j is the matrix obtained from A by replacing $A[:, j]$ by $\mathbf{b}$.

Proof. Since $\det(A) \neq 0$, A is invertible. Thus, there exists an invertible matrix P such that $PA = I_n$ and $P[A \mid \mathbf{b}] = [I \mid P\mathbf{b}]$. Then $A^{-1} = P$. Let $\mathbf{d} = P\mathbf{b} = A^{-1}\mathbf{b}$. Then, $A\mathbf{x} = \mathbf{b}$ has the unique solution $x_j = \mathbf{d}_j$, for $1 \leq j \leq n$. Also, $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = I = PA = [PA[:, 1], \ldots, PA[:, n]]$. Thus,

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

Thus, $\det(PA_j) = \mathbf{d}_j$, for $1 \leq j \leq n$. Also, $\mathbf{d}_j = \dfrac{\mathbf{d}_j}{1} = \dfrac{\det(PA_j)}{\det(PA)} = \dfrac{\det(P)\det(A_j)}{\det(P)\det(A)} = \dfrac{\det(A_j)}{\det(A)}$. Hence, $x_j = \dfrac{\det(A_j)}{\det(A)}$ and the required result follows. ∎

Example 2.5.28. 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

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

2.6 Miscellaneous Exercises

EXERCISE **2.6.1.** 1. *Let A be a unitary matrix then what can you say about $\mid \det(A) \mid$?*

2. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Prove that the following statements are equivalent:*

   (a) *A is not invertible.*

   (b) $\mathsf{Rank}(A) \neq n$.

   (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 has an infinite number of solutions.*

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

3. *Let A be a Hermitian matrix. Prove that $\det A$ is a real number.*

4. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is invertible if and only if Adj(A) is invertible.*

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

6. *Let A be an $n \times n$ invertible matrix and let $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$. Then, show that $\mathsf{Rank}(P) = n$ if and only if $D = CA^{-1}B$.*

7. *Let A be a 2×2 matrix with $\mathsf{tr(A) = 0}$ and $\det(A) = 0$. Then, A is a nilpotent matrix.*

8. *Determine necessary and sufficient condition for a triangular matrix to be invertible.*

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) $\begin{bmatrix} I & \mathbf{0} \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$,

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

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

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

12. *Let $a_1, \ldots, a_n \in \mathbb{C}$ and define $A = [a_{ij}]_{n \times n}$ with $a_{ij} = a_i^{j-1}$. Prove that* $\det(A) = \prod\limits_{1 \le i < j \le n} (a_j - a_i)$. *This matrix is usually called the van der monde matrix.*

13. *Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$ with $a_{ij} = \max\{i, j\}$. Prove that $\det A = (-1)^{n-1}n$.*

14. *Solve the following linear system 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) $x-y+z-w=1,\ x+y-z+w=2,\ 2x+y-z-w=7,\ x-y-z+w=3.$

15. *Let* $p \in \mathbb{C}, p \neq 0$. *Let* $A=[a_{ij}], B=[b_{ij}] \in \mathbb{M}_n(\mathbb{C})$ *with* $b_{ij} = p^{i-j}a_{ij}$, *for* $1 \leq i,j \leq n$. *Then, compute* $\det(B)$ *in terms of* $\det(A)$.

16. *The position of an element* a_{ij} *of a determinant is called even or odd according as* $i+j$ *is even or odd. Prove that if all the entries in*
    (a) *odd positions are multiplied with* -1 *then the value of determinant doesn't change.*
    (b) *even positions are multiplied with* -1 *then the value of determinant*
        i. *does not change if the matrix is of even order.*
        ii. *is multiplied by* -1 *if the matrix is of odd order.*

2.7 Summary

In this chapter, we started with a system of m linear equations in n variables and formally wrote it as $A\mathbf{x} = \mathbf{b}$ and in turn to the augmented matrix $[A \mid \mathbf{b}]$. Then, the basic operations on equations led to multiplication by elementary matrices on the right of $[A \mid \mathbf{b}]$. These elementary matrices are invertible and applying the GJE on a matrix A, resulted in getting the RREF of A. We used the pivots in RREF matrix to define the rank of a matrix. So, if $\mathsf{Rank}(A) = r$ and $\mathsf{Rank}([A \mid \mathbf{b}]) = r_a$

1. then, $r < r_a$ implied the linear system $A\mathbf{x} = \mathbf{b}$ is inconsistent.
2. then, $r = r_a$ implied the linear system $A\mathbf{x} = \mathbf{b}$ is consistent. Further,
    (a) if $r = n$ then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution.
    (b) if $r < n$ then the system $A\mathbf{x} = \mathbf{b}$ has an infinite number of solutions.

We have also seen that the following conditions are equivalent for $A \in \mathbb{M}_n(\mathbb{C})$.

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. $\mathsf{Rank}(A) = n$.
8. $\det(A) \neq 0$.

So, overall we have learnt to solve the following type of problems:

1. Solving the linear system $A\mathbf{x} = \mathbf{b}$. This idea will lead to the question "is the vector $\mathbf{b}$ a linear combination of the columns of A"?

2. Solving the linear system $A\mathbf{x} = \mathbf{0}$. This will lead to the question "are the columns of A linearly independent/dependent"? In particular, we will see that

   (a) if $A\mathbf{x} = \mathbf{0}$ has a unique solution then the columns of A are linear independent.

   (b) if $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution then the columns of A are linearly dependent.

Chapter 3

Vector Spaces

In this chapter, we will mainly be concerned with finite dimensional vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Please note that the real and complex numbers have the property that any pair of elements can be added, subtracted or multiplied. Also, division is allowed by a nonzero element. Such sets in mathematics are called field. So, $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are examples of field and they have infinite number of elements. But, in mathematics, we do have fields that have only finitely many elements. For example, consider the set $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$. In $\mathbb{Z}_5$, we define addition and multiplication, respectively, as

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

and

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

.

Then, we see that the elements of $\mathbb{Z}_5$ can be added, subtracted and multiplied. Note that 4 behaves as -1 and 3 behaves as -2. Thus, 1 behaves as -4 and 2 behaves as -3. Also, we see that in this multiplication $2 \cdot 3 = 1$ and $4 \cdot 4 = 1$. Hence,

1. the division by 2 is similar to multiplying by 3,
2. the division by 3 is similar to multiplying by 2, and
3. the division by 4 is similar to multiplying by 4.

Thus, $\mathbb{Z}_5$ indeed behaves like a field. So, in this chapter, $\mathbb{F}$ will represent a field.

3.1 Vector Spaces: Definition and Examples

Let $A \in \mathbb{M}_{m,n}(\mathbb{F})$ and let $\mathbb{V}$ denote the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then, by Theorem 2.1.9, $\mathbb{V}$ satisfies:

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

2. if $\mathbf{x} \in \mathbb{V}$ then $\alpha\mathbf{x} \in \mathbb{V}$, for all $\alpha \in \mathbb{F}$. In particular, for $\alpha = -1$, $-\mathbf{x} \in \mathbb{V}$.

3. if $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ then, for any $\alpha, \beta \in \mathbb{F}$, $\alpha\mathbf{x} + \beta\mathbf{y} \in \mathbb{V}$.

We see that the solution set of a homogeneous linear system satisfies certain properties which are also satisfied by the Euclidean plane, $\mathbb{R}^2$, or the Euclidean space, $\mathbb{R}^3$. In this chapter, our aim is to understand sets that satisfy such properties. We start with the formal definition.

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

1. **Vector Addition:** To every pair $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ there corresponds a unique element $\mathbf{u} \oplus \mathbf{v} \in \mathbb{V}$ (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) $\mathbb{V}$ has a unique element, denoted $\mathbf{0}$, called **the zero vector** that satisfies $\mathbf{u} \oplus \mathbf{0} = \mathbf{u}$, for every $\mathbf{u} \in \mathbb{V}$ (called **the additive identity**).

   (d) for every $\mathbf{u} \in \mathbb{V}$ there is an element $\mathbf{w} \in \mathbb{V}$ that satisfies $\mathbf{u} \oplus \mathbf{w} = \mathbf{0}$.

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

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

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

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

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

   (b) $(\alpha + \beta) \odot \mathbf{u} = (\alpha \odot \mathbf{u}) \ \oplus \ (\beta \odot \mathbf{u})$ ($+$ is addition in $\mathbb{F}$).

Remark 3.1.2. **[Real / Complex Vector Space]**

1. *The elements of* $\mathbb{F}$ *are called* **scalars**.

2. *The elements of* $\mathbb{V}$ *are called* **vectors**.

3. *We denote the zero element of* $\mathbb{F}$ *by* 0*, whereas the zero element of* $\mathbb{V}$ *will be denoted by* $\mathbf{0}$.

4. *Observe that Condition 3.1.1.1d implies that for every* $\mathbf{u} \in \mathbb{V}$*, the vector* $\mathbf{w} \in \mathbb{V}$ *such that* $\mathbf{u} \oplus \mathbf{w} = \mathbf{0}$ *holds, is unique. For if,* $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{V}$ *with* $\mathbf{u} \oplus \mathbf{w}_i = \mathbf{0}$*, for* $i = 1, 2$ *then by commutativity of vector addition, we see that*

$$\mathbf{w}_1 = \mathbf{w}_1 \oplus \mathbf{0} = \mathbf{w}_1 + (\mathbf{u} \oplus \mathbf{w}_2) = (\mathbf{w}_1 \oplus \mathbf{u}) \oplus \mathbf{w}_2 = \mathbf{0} \oplus \mathbf{w}_2 = \mathbf{w}_2.$$

   Hence, we represent this unique vector by $-\mathbf{u}$ *and call it* **the additive inverse**.

5. *If* $\mathbb{V}$ *is a vector space over* $\mathbb{R}$ *then* $\mathbb{V}$ *is called a* **real vector space**.

6. *If* $\mathbb{V}$ *is a vector space over* $\mathbb{C}$ *then* $\mathbb{V}$ *is called a* **complex vector space**.

7. *In general, a vector space over* $\mathbb{R}$ *or* $\mathbb{C}$ *is called a* **linear space**.

Some interesting consequences of Definition 3.1.1 is stated next. Intuitively, they seem obvious but for better understanding of the given conditions, it is desirable to go through the proof.

Theorem 3.1.3. *Let* $\mathbb{V}$ *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* $\mathbf{u} = \mathbf{0}$ *or* $\alpha = 0$.

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

Proof. Part 1: By Condition 3.1.1.1d, for each $\mathbf{u} \in \mathbb{V}$ there exists $-\mathbf{u} \in \mathbb{V}$ 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 Condition 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, using Part 1, $\alpha \odot \mathbf{0} = \mathbf{0}$ for any $\alpha \in \mathbb{F}$. In the same way, using Condition 3.1.1.3b,

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

Hence, using Part 1, one has $0 \odot \mathbf{u} = \mathbf{0}$ for any $\mathbf{u} \in \mathbb{V}$.

Now suppose $\alpha \odot \mathbf{u} = \mathbf{0}$. If $\alpha = 0$ then the proof is over. Therefore, assume that $\alpha \neq 0, \alpha \in \mathbb{F}$. Then, $(\alpha)^{-1} \in \mathbb{F}$ and

$$\mathbf{0} = (\alpha)^{-1} \odot \mathbf{0} = (\alpha)^{-1} \odot (\alpha \odot \mathbf{u}) = ((\alpha)^{-1} \cdot \alpha) \odot \mathbf{u} = 1 \odot \mathbf{u} = \mathbf{u}$$

as $1 \odot \mathbf{u} = \mathbf{u}$ for every vector $\mathbf{u} \in \mathbb{V}$ (see Condition 2.2b). Thus, if $\alpha \neq 0$ and $\alpha \odot \mathbf{u} = \mathbf{0}$ then $\mathbf{u} = \mathbf{0}$.

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

Example 3.1.4. The readers are advised to justify the statements given below.

1. Let $\mathbb{V} = \{\mathbf{0}\}$. Then, $\mathbb{V}$ is a real as well as a complex vector space.

2. Let $A \in \mathbb{M}_{m,n}(\mathbb{F})$ with $\text{Rank}(A) = r \leq n$. Then, using Theorem 2.4.4, the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$ is a vector space over $\mathbb{F}$.

3. Consider $\mathbb{R}$ with the usual addition and multiplication. That is, $a \oplus b = a + b$ and $a \odot b = a \cdot b$. Then, $\mathbb{R}$ forms a real vector space.

4. Let $\mathbb{R}^2 = \{(x_1, x_2)^T \mid x_1, x_2 \in \mathbb{R}\}$ Then, for $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $\alpha \in \mathbb{R}$, define

   $$(x_1, x_2)^T \oplus (y_1, y_2)^T = (x_1 + y_1, x_2 + y_2)^T \ \text{ and } \ \alpha \odot (x_1, x_2)^T = (\alpha x_1, \alpha x_2)^T.$$

   Verify that $\mathbb{R}^2$ is a real vector space.

5. Let $\mathbb{R}^n = \{(a_1, \ldots, a_n)^T \mid a_i \in \mathbb{R}, 1 \le i \le n\}$. For $\mathbf{u} = (a_1, \ldots, a_n)^T$, $\mathbf{v} = (b_1, \ldots, b_n)^T \in \mathbb{V}$ and $\alpha \in \mathbb{R}$, define

$$\mathbf{u} \oplus \mathbf{v} = (a_1 + b_1, \ldots, a_n + b_n)^T \text{ and } \alpha \odot \mathbf{u} = (\alpha a_1, \ldots, \alpha a_n)^T$$

(CALLED COMPONENT WISE OPERATIONS). Then, $\mathbb{V}$ is a real vector space. The vector space $\mathbb{R}^n$ is called **the real vector space of n-tuples**.

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

6. Consider $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$, the set of complex numbers. Let $\mathbf{z}_1 = x_1 + iy_1$ and $\mathbf{z}_2 = x_2 + iy_2$ and define $\mathbf{z}_1 \oplus \mathbf{z}_2 = (x_1 + x_2) + i(y_1 + y_2)$. For scalar multiplication,

   (a) let $\alpha \in \mathbb{R}$ and define, $\alpha \odot \mathbf{z}_1 = (\alpha x_1) + i(\alpha y_1)$. Then, $\mathbb{C}$ is a vector space over $\mathbb{R}$ (called the **real** vector space).

   (b) let $\alpha + i\beta \in \mathbb{C}$ and define, $(\alpha + i\beta) \odot (x_1 + iy_1) = (\alpha x_1 - \beta y_1) + i(\alpha y_1 + \beta x_1)$. Then, $\mathbb{C}$ forms a vector space over $\mathbb{C}$ (called the **complex** vector space).

7. Let $\mathbb{C}^n = \{(z_1, \ldots, z_n)^T \mid z_i \in \mathbb{C}, 1 \le i \le n\}$. For $\mathbf{z} = (z_1, \ldots, z_n), \mathbf{w} = (w_1, \ldots, w_n)^T \in \mathbb{C}^n$ and $\alpha \in \mathbb{F}$, define

$$\mathbf{z} + \mathbf{w} = (z_1 + w_1, \ldots, z_n + w_n)^T, \text{ and } \alpha \odot \mathbf{z} = (\alpha z_1, \ldots, \alpha z_n)^T.$$

Then, verify that $\mathbb{C}^n$ forms a vector space over $\mathbb{C}$ (called the **complex** vector space) as well as over $\mathbb{R}$ (called the **real** vector space). Unless specified otherwise, $\mathbb{C}^n$ will be considered a complex vector space.

Remark 3.1.5. *If $\mathbb{F} = \mathbb{C}$ then $i(1,0) = (i,0)$ is allowed. Whereas, if $\mathbb{F} = \mathbb{R}$ then $i(1,0)$ doesn't make sense as $i \notin \mathbb{R}$.*

8. Fix $m, n \in \mathbb{N}$ and let $\mathbb{M}_{m,n}(\mathbb{C}) = \{A_{m \times n} = [a_{ij}] \mid a_{ij} \in \mathbb{C}\}$. For $A, B \in \mathbb{M}_{m,n}(\mathbb{C})$ and $\alpha \in \mathbb{C}$, define $(A + \alpha B)_{ij} = a_{ij} + \alpha b_{ij}$. Then, $\mathbb{M}_{m,n}(\mathbb{C})$ is a complex vector space. If $m = n$, the vector space $\mathbb{M}_{m,n}(\mathbb{C})$ is denoted by $\mathbb{M}_n(\mathbb{C})$.

9. Let S be a non-empty set and let $\mathbb{R}^S = \{f \mid f \text{ is a function from } S \text{ to } \mathbb{R}\}$. For $f, g \in \mathbb{R}^S$ and $\alpha \in \mathbb{R}$, define $(f + \alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in S$. Then, $\mathbb{R}^S$ is a real vector space. In particular, for $S = \mathbb{N}$, observe that $\mathbb{R}^{\mathbb{N}}$ consists of all real sequences and forms a real vector space.

10. Fix $a, b \in \mathbb{R}$ with $a < b$ and let $\mathcal{C}([a,b], \mathbb{R}) = \{f : [a,b] \to \mathbb{R} \mid f \text{ is continuous}\}$. Then, $\mathcal{C}([a,b], \mathbb{R})$ with $(f + \alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in [a,b]$, is a real vector space.

11. Let $\mathcal{C}(\mathbb{R}, \mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$. Then, $\mathcal{C}(\mathbb{R}, \mathbb{R})$ is a real vector space, where $(f + \alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in \mathbb{R}$.

12. Fix $a < b \in \mathbb{R}$ and let $\mathcal{C}^2((a,b), \mathbb{R}) = \{f : (a,b) \to \mathbb{R} \mid f'' \text{ is continuous}\}$. Then, $\mathcal{C}^2((a,b), \mathbb{R})$ with $(f + \alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in (a,b)$, is a real vector space.

13. Fix $a < b \in \mathbb{R}$ and let $\mathcal{C}^{\infty}((a,b),\mathbb{R}) = \{f : (a,b) \to \mathbb{R} \mid f \text{ is infinitely differentiable}\}$. Then, $\mathcal{C}^{\infty}((a,b),\mathbb{R})$ with $(f+\alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in (a,b)$ is a real vector space.

14. Fix $a < b \in \mathbb{R}$. Then, $\mathbb{V} = \{f : (a,b) \to \mathbb{R} \mid f'' + f' + 2f = 0\}$ is a real vector space.

15. Let $\mathbb{R}[x] = \{a_0 + a_1 x + \cdots + a_n x^n \mid a_i \in \mathbb{R}, \text{ for } 0 \le i \le n\}$. Now, let $p(x), q(x) \in \mathbb{R}[x]$. Then, we can choose m such that $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ and $q(x) = b_0 + b_1 x + \cdots + b_m x^m$, where some of the a_i's or b_j's may be zero. Then, we define

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_m + b_m)x^m$$

and $\alpha p(x) = (\alpha a_0) + (\alpha a_1)x + \cdots + (\alpha a_m)x^m$, for $\alpha \in \mathbb{R}$. With these operations "componentwise addition and multiplication", it can be easily verified that $\mathbb{R}[x]$ forms a real vector space.

16. Fix $n \in \mathbb{N}$ and let $\mathbb{R}[x;n] = \{p(x) \in \mathbb{R}[x] \mid p(x) \text{ has degree } \le n\}$. Then, with componentwise addition and multiplication, the set $\mathbb{R}[x;n]$ forms a real vector space.

17. Let $\mathbb{C}[x] = \{a_0 + a_1 x + \cdots + a_n x^n \mid a_i \in \mathbb{C}, \text{ for } 0 \le i \le n\}$. Then, under componentwise addition and multiplication the set $\mathbb{C}[x]$ forms a real/complex vector space. Further $\mathbb{C}[x;n]$, the set of complex polynomials of degree less than or equal to n also forms a real/complex vector space.

18. Let $\mathbb{V} = \{A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C}) \mid a_{11} = 0\}$. Then, $\mathbb{V}$ is a complex vector space.

19. Let $\mathbb{V} = \{A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C}) \mid A = A^*\}$. Then, verify that $\mathbb{V}$ is a real vector space but not a complex vector space.

20. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$, with operations $(+, \bullet)$ and $(\oplus, \odot)$, respectively. Let $\mathbb{V} \times \mathbb{W} = \{(\mathbf{v}, \mathbf{w}) \mid \mathbf{v} \in \mathbb{V}, \mathbf{w} \in \mathbb{W}\}$. Then, $\mathbb{V} \times \mathbb{W}$ forms a vector space over $\mathbb{F}$, if for every $(\mathbf{v}_1, \mathbf{w}_1), (\mathbf{v}_2, \mathbf{w}_2) \in \mathbb{V} \times \mathbb{W}$ and $\alpha \in \mathbb{R}$, we define

$$\begin{aligned}(\mathbf{v}_1, \mathbf{w}_1) \oplus' (\mathbf{v}_2, \mathbf{w}_2) &= (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 \oplus \mathbf{w}_2), \quad \text{and} \\ \alpha \circ (\mathbf{v}_1, \mathbf{w}_1) &= (\alpha \bullet \mathbf{v}_1, \alpha \odot \mathbf{w}_1).\end{aligned}$$

$\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{w}_1 \oplus \mathbf{w}_2$ on the right hand side mean vector addition in $\mathbb{V}$ and $\mathbb{W}$, respectively. Similarly, $\alpha \bullet \mathbf{v}_1$ and $\alpha \odot \mathbf{w}_1$ correspond to scalar multiplication in $\mathbb{V}$ and $\mathbb{W}$, respectively.

21. Let $\mathbb{Q}$ be the set of scalars. Then,
   (a) $\mathbb{R}$ is a vector space over $\mathbb{Q}$. In this space, all the irrational numbers are vectors but not scalars.
   (b) $\mathbb{V} = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ is a vector space.
   (c) $\mathbb{V} = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} : a, b, c, d \in \mathbb{Q}\}$ is a vector space.
   (d) $\mathbb{V} = \{a + b\sqrt{-3} : a, b \in \mathbb{Q}\}$ is a vector space.

22. Let $\mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$. Then,

(a) $\mathbb{R}^+$ is not a vector space under usual operations of addition and scalar multiplication.

(b) $\mathbb{R}^+$ is a real vector space with 1 as the additive identity if we define

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

23. For any $\alpha \in \mathbb{R}$ and $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$, define

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

Then, $\mathbb{R}^2$ is a real vector space with $(-1, 3)^T$ as the additive identity.

24. Recall the field $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ given on the first page of this chapter. Then, $\mathbb{V} = \{(a, b) \mid a, b \in \mathbb{Z}_5\}$ is a vector space over $\mathbb{Z}_5$ having 25 elements/vectors.

Note that all our vector spaces, except the last two , are linear spaces.

From now on, we will use '$\mathbf{u} + \mathbf{v}$' for '$\mathbf{u} \oplus \mathbf{v}$' and '$\alpha\mathbf{u}$ or $\alpha \cdot \mathbf{u}$' for '$\alpha \odot \mathbf{u}$'.

EXERCISE **3.1.6.** *1. Verify that the vectors spaces mentioned in Example 3.1.4 do satisfy all the conditions for vector spaces.*

2. Does $\mathbb{R}$ with $x \oplus y = x - y$ and $\alpha \odot x = -\alpha x$, for all $x, y, \alpha \in \mathbb{R}$ form a vector space?

3. Let $\mathbb{V} = \mathbb{R}^2$. For $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$, define

(a) $(x_1, y_1)^T \oplus (x_2, y_2)^T = (x_1 + x_2, 0)^T$ and $\alpha \odot (x_1, y_1)^T = (\alpha x_1, 0)^T$.

(b) $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2)^T$ and $\alpha\mathbf{x} = (\alpha x_1, 0)^T$.

Then, does $\mathbb{V}$ form a vector space under any of the two operations?

4. Does the set $\mathbb{V}$ given below form a real/complex or both real and complex vector space? Give reasons for your answer.

(a) Let $\mathbb{V} = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a, b, c, d \in \mathbb{C}, a + c = 0 \right\}$.

(b) Let $\mathbb{V} = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a = \bar{b},\ a, b, c, d \in \mathbb{C} \right\}$.

(c) Let $\mathbb{V} = \{(x, y, z)^T \mid x + y + z = 1\}$.

(d) Let $\mathbb{V} = \{(x, y)^T \in \mathbb{R}^2 \mid x \cdot y = 0\}$.

(e) Let $\mathbb{V} = \{(x, y)^T \in \mathbb{R}^2 \mid x = y^2\}$.

(f) Let $\mathbb{V} = \{\alpha(1, 1, 1)^T + \beta(1, 1, -1)^T \mid \alpha, \beta \in \mathbb{R}\}$.

3.1.1 Subspaces

Definition 3.1.7. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, a non-empty subset S of $\mathbb{V}$ is called a **subspace** of $\mathbb{V}$ if S is also a vector space with vector addition and scalar multiplication inherited from $\mathbb{V}$.

Example 3.1.8. 1. The vector space $\mathbb{R}[x; n]$ is a subspace of $\mathbb{R}[x]$.

2. Is $\mathbb{V} = \{xp(x) \mid p(x) \in \mathbb{R}[x]\}$ a subspace of $\mathbb{R}[x]$?

3. Let $\mathbb{V}$ be a vector space. Then $\mathbb{V}$ and $\{\mathbf{0}\}$ are subspaces, called **trivial subspaces**.

4. The real vector space $\mathbb{R}$ has no non-trivial subspace. To check this, let $\mathbb{V} \neq \{\mathbf{0}\}$ be a vector subspace of $\mathbb{R}$. Then, there exists $x \in \mathbb{R}, x \neq \mathbf{0}$ such that $x \in \mathbb{V}$. Now, using scalar multiplication, we see that $\{\alpha x \mid \alpha \in \mathbb{R}\} \subseteq \mathbb{V}$. As, $x \neq \mathbf{0}$, the set $\{\alpha x \mid \alpha \in \mathbb{R}\} = \mathbb{R}$. This in turn implies that $\mathbb{V} = \mathbb{R}$.

5. $\mathbb{W} = \{\mathbf{x} \in \mathbb{R}^3 \mid \ [1, 2, -1]\mathbf{x} = 0\}$ is a plane in $\mathbb{R}^3$ containing $\mathbf{0}$ (hence a subspace).

6. $\mathbb{W} = \{\mathbf{x} \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0}\}$ is a line in $\mathbb{R}^3$ containing $\mathbf{0}$ (hence a subspace).

7. Verify that $\mathcal{C}^2(a, b)$ is a subspace of $\mathcal{C}(a, b)$.

8. Verify that $\mathbb{W} = \{(x, 0)^T \in \mathbb{R}^2 \ \mid \ x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

9. Is the set of sequences converging to 0 a subspace of the set of all bounded sequences?

10. Let $\mathbb{V}$ be the vector space of Example 3.1.4.23. Then,
    (a) $S = \{(x, 0)^T \mid \ x \in \mathbb{R}\}$ is not a subspace of $\mathbb{V}$ as $(x, 0)^T \oplus (y, 0)^T = (x+y+1, -3)^T \notin S$.
    (b) Verify that $\mathbb{W} = \{(x, 3)^T \ \mid \ x \in \mathbb{R}\}$ is a subspace of $\mathbb{V}$.

11. The vector space $\mathbb{R}^+$ defined in Example 3.1.4.22 is not a subspace of $\mathbb{R}$.

Let $\mathbb{V}(\mathbb{F})$ be a vector space and $\mathbb{W} \subseteq \mathbb{V}, \mathbb{W} \neq \emptyset$. We now prove a result which implies that to check $\mathbb{W}$ to be a subspace, we need to verify only one condition.

Theorem 3.1.9. *Let $\mathbb{V}(\mathbb{F})$ be a vector space and $\mathbb{W} \subseteq \mathbb{V}, \mathbb{W} \neq \emptyset$. Then, $\mathbb{W}$ is a subspace of $\mathbb{V}$ if and only if $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathbb{W}$ whenever $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{W}$. Note that the vector addition and scalar multiplication are inherited from that in $\mathbb{V}(\mathbb{F})$.*

Proof. Let $\mathbb{W}$ be a subspace of $\mathbb{V}$ and let $\mathbf{u}, \mathbf{v} \in \mathbb{W}$. Then, for every $\alpha, \beta \in \mathbb{F}$, $\alpha\mathbf{u}, \beta\mathbf{v} \in \mathbb{W}$ and hence $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathbb{W}$.

Now, we assume that $\alpha\mathbf{u} + \beta\mathbf{v} \in \mathbb{W}$, whenever $\alpha, \beta \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{W}$. To show, $\mathbb{W}$ is a subspace of $\mathbb{V}$:

1. Taking $\alpha = 1$ and $\beta = 1$, we see that $\mathbf{u} + \mathbf{v} \in \mathbb{W}$, for every $\mathbf{u}, \mathbf{v} \in \mathbb{W}$.

2. Taking $\alpha = 0$ and $\beta = 0$, we see that $\mathbf{0} \in \mathbb{W}$.

3. Taking $\beta = 0$, we see that $\alpha\mathbf{u} \in \mathbb{W}$, for every $\alpha \in \mathbb{F}$ and $\mathbf{u} \in \mathbb{W}$. Hence, using Theorem 3.1.3.3, $-\mathbf{u} = (-1)\mathbf{u} \in \mathbb{W}$ as well.

4. The commutative and associative laws of vector addition hold as they hold in $\mathbb{V}$.

5. The conditions related with scalar multiplication and the distributive laws also hold as they hold in $\mathbb{V}$.

Thus, one obtains the required result. ■

EXERCISE **3.1.10.** *1. Determine all the subspaces of* $\mathbb{R}$ *and* $\mathbb{R}^2$*.*

2. *Prove that a line in* $\mathbb{R}^2$ *is a subspace if and only if it passes through* $(0,0)^T \in \mathbb{R}^2$*.*

3. *Fix* $n \in \mathbb{N}$*. In the exampless given below, is* $\mathbb{W}$ *a subspace of* $M_n(\mathbb{R})$*, where*

   (a) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A \text{ is upper triangular}\}$*?*

   (b) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A \text{ is symmetric}\}$*?*

   (c) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A \text{ is skew-symmetric}\}$*?*

   (d) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A \text{ is a diagonal matrix}\}$*?*

   (e) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid \text{trace}(A) = 0\}$*?*

   (f) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A^T = 2A\}$*?*

4. *Fix* $n \in \mathbb{N}$*. Then, is* $\mathbb{W} = \{A = [a_{ij}] \in \mathbb{M}_n(\mathbb{R} \mid a_{11} + \overline{a_{22}} = 0\}$ *a subspace of the complex vector space* $M_n(\mathbb{C})$*? What if* $M_n(\mathbb{C})$ *is a real vector space?*

5. *Are all the sets given below subspaces of* $\mathcal{C}([-1,1])$*?*

   (a) $\mathbb{W} = \{f \in C([-1,1]) \mid f(1/2) = 0\}$.

   (b) $\mathbb{W} = \{f \in C([-1,1]) \mid f(-1/2) = 0, f(1/2) = 0\}$.

6. *Are all the sets given below subspaces of* $\mathbb{R}[x]$*? Recall that the degree of the zero polynomial is assumed to be* $-\infty$*.*

   (a) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid \deg(f(x)) = 3\}$.

   (b) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid \deg(f(x)) \le 0\}$.

   (c) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid f(0) = 0\}$.

7. *Which of the following are subspaces of* $\mathbb{R}^n(\mathbb{R})$*?*

   (a) $\{(x_1, x_2, \ldots, x_n)^T \mid x_1 \ge 0\}$.

   (b) $\{(x_1, x_2, \ldots, x_n)^T \mid x_1 \text{ is rational}\}$.

   (c) $\{(x_1, x_2, \ldots, x_n)^T \mid \mid x_1 \mid \le 1\}$.

8. *Among the following, determine the subspaces of the complex vector space* $\mathbb{C}^n$*?*

   (a) $\{(z_1, z_2, \ldots, z_n)^T \mid z_1 \text{ is real }\}$.

   (b) $\{(z_1, z_2, \ldots, z_n)^T \mid z_1 + z_2 = \overline{z_3}\}$.

   (c) $\{(z_1, z_2, \ldots, z_n)^T \mid \mid z_1 \mid = \mid z_2 \mid\}$.

9. *Prove that the following sets are not subspaces of* $M_n(\mathbb{R})$*.*

   (a) $G = \{A \in M_n(\mathbb{R}) \mid \det(A) = 0\}$.

   (b) $G = \{A \in M_n(\mathbb{R}) \mid \det(A) = 1\}$.

3.1.2 Linear Span

Definition 3.1.11. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, for any $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{V}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, the vector $\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n = \sum\limits_{i=1}^{n} \alpha_i\mathbf{u}_i$ is said to be a **linear combination** of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

Example 3.1.12. 1. $(3,4,3)$ is a linear combination of $(1,1,1)$ and $(1,2,1)$ as $(3,4,3) = 2(1,1,1) + (1,2,1)$.

2. $(3,4,5)$ is not a linear combination of $(1,1,1)$ and $(1,2,1)$ as the linear system $(3,4,5) = a(1,1,1) + b(1,2,1)$, in the variables a and b has no solution.

3. Is $(4,5,5)$ a linear combination of $\mathbf{e}_1^T = (1,0,0)$, $\mathbf{e}_2^T = (0,1,0)$ and $\mathbf{e}_3^T = (3,3,1)$?
Solution: $(4,5,5)$ is a linear combination as $(4,5,5) = 4\mathbf{e}_1^T + 5\mathbf{e}_2^T + 5\mathbf{e}_3^T$.

4. Is $(4,5,5)$ a linear combination of $(1,0,0)$, $(2,1,0)$ and $(3,3,1)$?
Solution: $(4,5,5)$ 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 variables $a, b, c \in \mathbb{R}$ has a solution. Clearly, Equation (3.1.1) has solution $a = 9, b = -10$ and $c = 5$.

5. Is $4 + 5x + 5x^2 + x^3$ a linear combination of the polynomials $p_1(x) = 1$, $p_2(x) = 2 + x^2$ and $p_3(x) = 3 + 3x + x^2 + x^3$?
Solution: The polynomial $4 + 5x + 5x^2 + x^3$ is a linear combination if the linear system
$$ap_1(x) + bp_2(x) + cp_3(x) = 4 + 5x + 5x^2 + x^3 \tag{3.1.2}$$
in the variables $a, b, c \in \mathbb{R}$ has a solution. Verify that the system has no solution. Thus, $4 + 5x + 5x^2 + x^3$ is not a linear combination of the given set of polynomials.

6. Is $\begin{bmatrix} 1 & 3 & 4 \\ 3 & 3 & 6 \\ 4 & 6 & 5 \end{bmatrix}$ a linear combination of the vectors I_3, $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{bmatrix}$?
Solution: Verify that $\begin{bmatrix} 1 & 3 & 4 \\ 3 & 3 & 6 \\ 4 & 6 & 5 \end{bmatrix} = I_3 + 2\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{bmatrix}$. Hence, it is indeed a linear combination of given vectors of $\mathbb{M}_3(\mathbb{R})$.

EXERCISE **3.1.13.** *1. Let* $\mathbf{x} \in \mathbb{R}^3$. *Prove that* $\mathbf{x}^T$ *is a linear combination of* $(1,0,0)$, $(2,1,0)$ *and* $(3,3,1)$. *Is this linear combination unique? That is, does there exist* $(a,b,c) \neq (e,f,g)$ *with* $\mathbf{x}^T = a(1,0,0) + b(2,1,0) + c(3,3,1) = e(1,0,0) + f(2,1,0) + g(3,3,1)$*?*

2. Find condition(s) on $x, y, z \in \mathbb{R}$ *such that*

(a) (x,y,z) *is a linear combination of* $(1,2,3), (-1,1,4)$ *and* $(3,3,2)$.

(b) (x,y,z) *is a linear combination of* $(1,2,1), (1,0,-1)$ *and* $(1,1,0)$.

(c) (x, y, z) *is a linear combination of* $(1,1,1),(1,1,0)$ *and* $(1,-1,0)$.

Definition 3.1.14. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $S \subseteq \mathbb{V}$. Then, the **linear span** of S, denoted $LS(S)$, is defined as

$$LS(S) \quad = \quad \{\alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n \mid \alpha_i \in \mathbb{F}, \mathbf{u}_i \in S, \text{ for } 1 \le i \le n\}.$$

That is, $LS(S)$ is the set of all possible linear combinations of finitely many vectors of S. If S is an empty set, we define $LS(S) = \{\mathbf{0}\}$.

Example 3.1.15. For the set S given below, determine $LS(S)$.

1. $S = \{(1,0)^T, (0,1)^T\} \subseteq \mathbb{R}^2$.
   Solution: $LS(S) = \{a(1,0)^T + b(0,1)^T \mid a, b \in \mathbb{R}\} = \{(a,b)^T \mid a, b \in \mathbb{R}\} = \mathbb{R}^2$.

2. $S = \{(1,1,1)^T, (2,1,3)^T\}$. What does $LS(S)$ represent in $\mathbb{R}^3$?
   Solution: $LS(S) = \{a(1,1,1)^T + b(2,1,3)^T \mid a, b \in \mathbb{R}\} = \{(a+2b, a+b, a+3b)^T \mid a, b \in \mathbb{R}\}$. Note that $LS(S)$ represents a plane passing through the points $(0,0,0)^T, (1,1,1)^T$ and $(2,1,3)^T$. To get he equation of the plane, we proceed as follows:
   Find conditions on x, y and z such that $(a+2b, a+b, a+3b) = (x, y, z)$. Or equivalently, find conditions on x, y and z such that $a+2b = x, a+b = y$ and $a+3b = z$ has a solution for all $a, b \in \mathbb{R}$. The RREF of the augmented matrix equals $\begin{bmatrix} 1 & 0 & 2y-x \\ 0 & 1 & x-y \\ 0 & 0 & z+y-2x \end{bmatrix}$. Thus, the required condition on x, y and z is given by $z+y-2x = 0$. Hence,

   $$LS(S) = \{a(1,1,1)^T + b(2,1,3)^T \mid a, b \in \mathbb{R}\} = \{(x,y,z)^T \in \mathbb{R}^3 \mid 2x - y - z = 0\}.$$

3. $S = \{1 + 2x + 3x^2, 1 + x + 2x^2, 1 + 2x + x^3\}$.
   Solution: To understand $LS(S)$, we need to find condition(s) on $\alpha, \beta, \gamma, \delta$ such that the linear system

   $$a(1 + 2x + 3x^2) + b(1 + x + 2x^2) + c(1 + 2x + x^3) = \alpha + \beta x + \gamma x^2 + \delta x^3$$

   in the unknowns a, b, c is always consistent. An application of GJE method gives $\alpha + \beta - \gamma - 3\delta = 0$ as the required condition. Thus,

   $$LS(S) = \{\alpha + \beta x + \gamma x^2 + \delta x^3 \in \mathbb{R}[x] \mid \alpha + \beta - \gamma - 3\delta = 0\}.$$

4. $S = \left\{ I_3, \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{bmatrix} \right\} \subseteq \mathbb{M}_3(\mathbb{R})$.
   Solution: To get the equation, we need to find conditions on a_{ij}'s such that the system

   $$\begin{bmatrix} \alpha & \beta+\gamma & \beta+2\gamma \\ \beta+\gamma & \alpha+\beta & 2\beta+2\gamma \\ \beta+2\gamma & 2\beta+2\gamma & \alpha+2\gamma \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

in the unknowns α, β, γ is always consistent. Now, verify that the required condition equals

$$LS(S) = \{A = [a_{ij}] \in \mathbb{M}_3(\mathbb{R}) \mid A = A^T, a_{11} = \frac{a_{22} + a_{33} - a_{13}}{2}, \\ a_{12} = \frac{a_{22} - a_{33} + 3a_{13}}{4}, a_{23} = \frac{a_{22} - a_{33} + 3a_{13}}{2}\}.$$

Exercise **3.1.16.** *Determine the equation of the geometrical object represented by $LS(S)$.*

1. $S = \{\pi\} \subseteq \mathbb{R}$.
2. $S = \{(x, y)^T : x, y < 0\} \subseteq \mathbb{R}^2$.
3. $S = \{(x, y)^T :$ *either* $x \neq 0$ *or* $y \neq 0\} \subseteq \mathbb{R}^2$.
4. $S = \{(1, 0, 1)^T, (0, 1, 0)^T, (2, 0, 2)^T\} \subseteq \mathbb{R}^3$. *Give two examples of vectors* $\mathbf{u}, \mathbf{v}$ *different from the given set such that* $LS(S) = LS(\mathbf{u}, \mathbf{v})$.
5. $S = \{(x, y, z)^T : x, y, z > 0\} \subseteq \mathbb{R}^3$.
6. $S = \left\{ \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \right\} \subseteq \mathbb{M}_3(\mathbb{R})$.
7. $S = \{(1, 2, 3, 4)^T, (-1, 1, 4, 5)^T, (3, 3, 2, 3)^T\} \subseteq \mathbb{R}^4$.
8. $S = \{1 + 2x + x^2, x, 1 + x^2\} \subseteq \mathbb{C}[x; 2]$. *Give two examples of polynomial* $p(x), q(x)$ *different from the given set such that* $LS(S) = LS(p(x), q(x))$.
9. $S = \{1 + 2x + 3x^2, -1 + x + 4x^2, 3 + 3x + 2x^2\} \subseteq \mathbb{C}[x; 2]$.
10. $S = \{1, x, x^2, \ldots\} \subseteq \mathbb{C}[x]$.

Definition 3.1.17. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, $\mathbb{V}$ is called **finite dimensional** if there exists $S \subseteq \mathbb{V}$, such that S has finite number of elements and $\mathbb{V} = LS(S)$. If such an S does not exist then $\mathbb{V}$ is called **infinite dimensional**.

Example 3.1.18.

1. $\{(1, 2)^T, (2, 1)^T\}$ spans $\mathbb{R}^2$. Thus, $\mathbb{R}^2$ is finite dimensional.
2. $\{1, 1 + x, 1 - x + x^2, x^3, x^4, x^5\}$ spans $\mathbb{C}[x; 5]$. Thus, $\mathbb{C}[x; 5]$ is finite dimensional.
3. Fix $n \in \mathbb{N}$. Then, $\mathbb{C}[x; n]$ is finite dimensional as $\mathbb{C}[x; n] = LS(\{1, x, x^2, \ldots, x^n\})$.
4. $\mathbb{C}[x]$ is not finite dimensional as the degree of a polynomial can be any large positive integer. Indeed, verify that $\mathbb{C}[x] = LS(\{1, x, x^2, \ldots, x^n, \ldots\})$.
5. The vector space $\mathbb{R}$ over $\mathbb{Q}$ is infinite dimensional. An argument to justify it will be given later. The same argument also implies that the vector space $\mathbb{C}$ over $\mathbb{Q}$ is infinite dimensional.

Lemma 3.1.19 (Linear Span is a Subspace)**.** *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $S \subseteq \mathbb{V}$. Then, $LS(S)$ is a subspace of $\mathbb{V}$.*

Proof. By definition, $\mathbf{0} \in LS(S)$. So, $LS(S)$ is non-empty. Let $\mathbf{u}, \mathbf{v} \in LS(S)$. To show, $a\mathbf{u} + b\mathbf{v} \in LS(S)$ for all $a, b \in \mathbb{F}$. As $\mathbf{u}, \mathbf{v} \in LS(S)$, there exist $n \in \mathbb{N}$, vectors $\mathbf{w}_i \in S$ and scalars $\alpha_i, \beta_i \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1\mathbf{w}_1 + \cdots + \alpha_n\mathbf{w}_n$ and $\mathbf{v} = \beta_1\mathbf{w}_1 + \cdots + \beta_n\mathbf{w}_n$. Hence,

$$a\mathbf{u} + b\mathbf{v} = (a\alpha_1 + b\beta_1)\mathbf{w}_1 + \cdots + (a\alpha_n + b\beta_n)\mathbf{w}_n \in LS(S)$$

as $a\alpha_i + b\beta_i \in \mathbb{F}$ for $1 \le i \le n$. Thus, by Theorem 3.1.9, $LS(S)$ is a vector subspace. ∎

EXERCISE **3.1.20.** *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $W \subseteq \mathbb{V}$.*

1. *Then, $LS(W) = W$ if and only if W is a subspace of $\mathbb{V}$.*
2. *If W is a subspace of $\mathbb{V}$ and $S \subseteq W$ then $LS(S)$ is a subspace of W as well.*

Theorem 3.1.21. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $S \subseteq \mathbb{V}$. Then, $LS(S)$ is the smallest subspace of $\mathbb{V}$ containing S.*

Proof. For every $\mathbf{u} \in S$, $\mathbf{u} = 1 \cdot \mathbf{u} \in LS(S)$. Thus, $S \subseteq LS(S)$. Need to show that $LS(S)$ is the smallest subspace of $\mathbb{V}$ containing S. So, let $\mathbb{W}$ be any subspace of $\mathbb{V}$ containing S. Then, by Exercise 3.1.20, $LS(S) \subseteq \mathbb{W}$ and hence the result follows. ∎

Definition 3.1.22. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$.

1. Let S and T be two subsets of $\mathbb{V}$. Then, the **sum** of S and T, denoted $S + T$ equals $\{s + t | s \in S, t \in T\}$. For example,
   (a) if $\mathbb{V} = \mathbb{R}$, $S = \{0, 1, 2, 3, 4, 5, 6\}$ and $T = \{5, 10, 15\}$ then $S + T = \{5, 6, \ldots, 21\}$.
   (b) if $\mathbb{V} = \mathbb{R}^2$, $S = \left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$ and $T = \left\{\begin{bmatrix}-1\\1\end{bmatrix}\right\}$ then $S + T = \left\{\begin{bmatrix}0\\2\end{bmatrix}\right\}$.
   (c) if $\mathbb{V} = \mathbb{R}^2$, $S = \left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$ and $T = LS\left(\begin{bmatrix}-1\\1\end{bmatrix}\right)$ then $S + T = \left\{\begin{bmatrix}1\\1\end{bmatrix} + c\begin{bmatrix}-1\\1\end{bmatrix} | c \in \mathbb{R}\right\}$.
2. Let P and Q be two subspaces of $\mathbb{R}^2$. Then, $P + Q = \mathbb{R}^2$, if
   (a) $P = \{(x, 0)^T \mid x \in \mathbb{R}\}$ and $Q = \{(0, x)^T \mid x \in \mathbb{R}\}$ as $(x, y) = (x, 0) + (0, y)$.
   (b) $P = \{(x, 0)^T \mid x \in \mathbb{R}\}$ and $Q = \{(x, x)^T \mid x \in \mathbb{R}\}$ as $(x, y) = (x - y, 0) + (y, y)$.
   (c) $P = LS((1, 2)^T)$ and $Q = LS((2, 1)^T)$ as $(x, y) = \frac{2y - x}{3}(1, 2) + \frac{2x - y}{3}(2, 1)$.

We leave the proof of the next result for readers.

Lemma 3.1.23. *Let P and Q be two subspaces of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, $P + Q$ is a subspace of $\mathbb{V}$. Furthermore, $P + Q$ is the smallest subspace of $\mathbb{V}$ containing both P and Q.*

EXERCISE **3.1.24.**
1. *Let $\mathbf{a} \in \mathbb{R}^2, \mathbf{a} \neq \mathbf{0}$. Then, show that $\{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{a}^T\mathbf{x} = 0\}$ is a non-trivial subspace of $\mathbb{R}^2$. Geometrically, what does this set represent in $\mathbb{R}^2$?*
2. *Find all subspaces of $\mathbb{R}^3$.*

3. *Let* $\mathbb{U} = \left\{ \begin{bmatrix} a & b \\ c & 0 \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}$ *and* $\mathbb{W} = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}$ *be subspaces of* $\mathbb{M}_2(\mathbb{R})$. *Determine* $\mathbb{U} \cap \mathbb{W}$. *Is* $\mathbb{M}_2(\mathbb{R}) = \mathbb{U} \cup \mathbb{W}$? *What is* $\mathbb{U} + \mathbb{W}$?

4. *Let* $\mathbb{W}$ *and* $\mathbb{U}$ *be two subspaces of a vector space* $\mathbb{V}$ *over* $\mathbb{F}$.

   (a) *Prove that* $\mathbb{W} \cap \mathbb{U}$ *is a subspace of* $\mathbb{V}$.

   (b) *Give examples of* $\mathbb{W}$ *and* $\mathbb{U}$ *such that* $\mathbb{W} \cup \mathbb{U}$ *is not a subspace of* $\mathbb{V}$.

   (c) *Determine conditions on* $\mathbb{W}$ *and* $\mathbb{U}$ *such that* $\mathbb{W} \cup \mathbb{U}$ *a subspace of* $\mathbb{V}$?

   (d) *Prove that* $LS(\mathbb{W} \cup \mathbb{U}) = \mathbb{W} + \mathbb{U}$.

---

5. *Prove that* $\{(x, y, z)^T \in \mathbb{R}^3 \mid ax + by + cz = d\}$ *is a subspace of* $\mathbb{R}^3$ *if and only if* $d = 0$.

6. *Determine all subspaces of the vector space in Example 3.1.4.23.*

7. *Let* $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$, *where* $\mathbf{x}_1 = (1, 0, 0)^T, \mathbf{x}_2 = (1, 1, 0)^T, \mathbf{x}_3 = (1, 2, 0)^T$ *and* $\mathbf{x}_4 = (1, 1, 1)^T$. *Then, determine all* $\mathbf{x}_i$ *such that* $LS(S) = LS(S \setminus \{\mathbf{x}_i\})$.

8. *Let* $\mathbb{W} = LS((1, 0, 0)^T, (1, 1, 0)^T)$ *and* $\mathbb{U} = LS((1, 1, 1)^T)$. *Prove that* $\mathbb{W} + \mathbb{U} = \mathbb{R}^3$ *and* $\mathbb{W} \cap \mathbb{U} = \{\mathbf{0}\}$. *If* $\mathbf{v} \in \mathbb{R}^3$, *determine* $\mathbf{w} \in \mathbb{W}$ *and* $\mathbf{u} \in \mathbb{U}$ *such that* $\mathbf{v} = \mathbf{w} + \mathbf{u}$. *Is it necessary that* $\mathbf{w}$ *and* $\mathbf{u}$ *are unique?*

9. *Let* $\mathbb{W} = LS((1, -1, 0), (1, 1, 0))$ *and* $\mathbb{U} = LS((1, 1, 1), (1, 2, 1))$. *Prove that* $\mathbb{W} + \mathbb{U} = \mathbb{R}^3$ *and* $\mathbb{W} \cap \mathbb{U} \neq \{\mathbf{0}\}$. *Find* $\mathbf{v} \in \mathbb{R}^3$ *such that* $\mathbf{v} = \mathbf{w} + \mathbf{u}$, *for* 2 *different choices of* $\mathbf{w} \in \mathbb{W}$ *and* $\mathbf{u} \in \mathbb{U}$. *That is, the choice of vectors* $\mathbf{w}$ *and* $\mathbf{u}$ *is not unique.*

Let $\mathbb{V}$ be a vector space over either $\mathbb{R}$ or $\mathbb{C}$. Then, we have learnt the following:

1. for any $S \subseteq \mathbb{V}$, $LS(S)$ is again a vector space. Moreover, $LS(S)$ is the smallest subspace containing S.
2. if $S = \emptyset$ then $LS(S) = \{\mathbf{0}\}$.
3. if S has at least one non zero vector then $LS(S)$ contains infinite number of vectors.

Therefore, the following questions arise:

1. Are there conditions under which $LS(S_1) = LS(S_2)$, for $S_1 \neq S_2$?
2. Is it always possible to find S so that $LS(S) = \mathbb{V}$?
3. Suppose we have found $S \subseteq \mathbb{V}$ such that $LS(S) = \mathbb{V}$. Can we find S such that no proper subset of S spans $\mathbb{V}$?

We try to answer these questions in the subsequent sections.

3.2 Linear Independence

Definition 3.2.1. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a non-empty subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, S is said to be **linearly independent** if the linear system

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}, \tag{3.2.1}$$

in the variables α_i's, $1 \le i \le m$, has only the trivial solution. If Equation (3.2.1) has a non-trivial solution then S is said to be **linearly dependent**.

If S has infinitely many vectors then S is said to be **linearly independent** if for every finite subset T of S, T is linearly independent.

Observe that we are solving a linear system over $\mathbb{F}$. Hence, linear independence and dependence depend on $\mathbb{F}$, the set of scalars.

Example 3.2.2. 1. Is the set S a linear independent set? Give reasons.

(a) Let $S = \{1 + 2x + x^2, 2 + x + 4x^2, 3 + 3x + 5x^2\} \subseteq \mathbb{R}[x; 2]$.

Solution: Consider the system $\begin{bmatrix} 1 + 2x + x^2 & 2 + x + 4x^2 & 3 + 3x + 5x^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$, or equivalently $a(1 + 2x + x^2) + b(2 + x + 4x^2) + c(3 + 3x + 5x^2) = 0$, in the variables a, b and c. As two polynomials are equal if and only if their coefficients are equal, the above system reduces to the homogeneous system $a + 2b + 3c = 0, 2a + b + 3c = 0, a + 4b + 5c = 0$. The corresponding coefficient matrix has rank $2 < 3$, the number of variables. Hence, the system has a non-trivial solution. Thus, S is a linearly dependent subset of $\mathbb{R}[x; 2]$.

(b) $S = \{1, \sin(x), \cos(x)\}$ is a linearly independent subset of $\mathcal{C}([-\pi, \ \pi], \mathbb{R})$ over $\mathbb{R}$ as the system

$$\begin{bmatrix} 1 & \sin(x) & \cos(x) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \Leftrightarrow a \cdot 1 + b \cdot \sin(x) + c \cdot \cos(x) = 0, \tag{3.2.2}$$

in the variables a, b and c has only the trivial solution. To verify this, evaluate Equation (3.2.2) at $-\frac{\pi}{2}, 0$ and $\frac{\pi}{2}$ to get the homogeneous system $a - b = 0, a + c = 0, a + b = 0$. Clearly, this system has only the trivial solution.

(c) Let $S = \{(0, 1, 1)^T, (1, 1, 0)^T, (1, 0, 1)^T\}$.

Solution: Consider the system $\begin{bmatrix} (0, 1, 1) & (1, 1, 0) & (1, 0, 1) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = (0, 0, 0)$ in the variables a, b and c. As rank of coefficient matrix is $3 =$ the number of variables, the system has only the trivial solution. Hence, S is a linearly independent subset of $\mathbb{R}^3$.

(d) Consider $\mathbb{C}$ as a complex vector space and let $S = \{1, i\}$.

Solution: Since $\mathbb{C}$ is a complex vector space, $i \cdot 1 + (-1)i = i - i = 0$. So, S is a linear dependent subset of the complex vector space $\mathbb{C}$.

(e) Consider $\mathbb{C}$ as a real vector space and let $S = \{1, i\}$.
Solution: Consider the linear system $a \cdot 1 + b \cdot i = 0$, in the variables $a, b \in \mathbb{R}$. Since $a, b \in \mathbb{R}$, equating real and imaginary parts, we get $a = b = 0$. So, S is a linear independent subset of the real vector space $\mathbb{C}$.

2. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If Rank$(A) < m$ then, the rows of A are linearly dependent.
Solution: As $\mathsf{Rank}(A) < m$, there exists an invertible matrix P such that $PA = \begin{bmatrix} C \\ \mathbf{0} \end{bmatrix}$. Thus, $\mathbf{0}^T = (PA)[m,:] = \sum\limits_{i=1}^{m} p_{mi} A[i,:]$. As P is invertible, at least one $p_{mi} \neq 0$. Thus, the required result follows.

3. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If Rank$(A) < n$ then, the columns of A are linearly dependent.
Solution: As $\mathsf{Rank}(A) < n$, by Corollary 2.3.8, there exists an invertible matrix Q such that $AQ = \begin{bmatrix} B & \mathbf{0} \end{bmatrix}$. Thus, $\mathbf{0} = (AQ)[:,n] = \sum\limits_{i=1}^{n} q_{in} A[:,i]$. As Q is invertible, at least one $q_{in} \neq 0$. Thus, the required result follows.

3.2.1 Basic Results on Linear Independence

The reader is expected to supply the proof of parts that are not given.

Proposition 3.2.3. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$.*

1. *Then, $\mathbf{0}$, the zero-vector, cannot belong to a linearly independent set.*

2. *Then, every subset of a linearly independent set in $\mathbb{V}$ is also linearly independent.*

3. *Then, a set containing a linearly dependent set of $\mathbb{V}$ is also linearly dependent.*

Proof. Let $\mathbf{0} \in S$. Then, $1 \cdot \mathbf{0} = \mathbf{0}$. That is, a non-trivial linear combination of some vectors in S is $\mathbf{0}$. Thus, the set S is linearly dependent. ∎

We now prove a couple of results which will be very useful in the next section.

Proposition 3.2.4. *Let S be a linearly independent subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. If T_1, T_2 are two subsets of S such that $T_1 \cap T_2 = \emptyset$ then, $LS(T_1) \cap LS(T_2) = \{\mathbf{0}\}$. That is, if $\mathbf{v} \in LS(T_1) \cap LS(T_2)$ then $\mathbf{v} = \mathbf{0}$.*

Proof. Let $\mathbf{v} \in LS(T_1) \cap LS(T_2)$. Then, there exist vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in T_1$, $\mathbf{w}_1, \ldots, \mathbf{w}_\ell \in T_2$ and scalars α_i's and β_j's (not all zero) such that $\mathbf{v} = \sum\limits_{i=1}^{k} \alpha_i \mathbf{u}_i$ and $\mathbf{v} = \sum\limits_{j=1}^{\ell} \beta_j \mathbf{w}_j$. Thus, we see that $\sum\limits_{i=1}^{k} \alpha_i \mathbf{u}_i + \sum\limits_{j=1}^{\ell} (-\beta_j) \mathbf{w}_j = \mathbf{0}$. As the scalars α_i's and β_j's are not all zero, we see that a non-trivial linear combination of some vectors in $T_1 \cup T_2 \subseteq S$ is $\mathbf{0}$. This contradicts the assumption that S is a linearly independent subset of $\mathbb{V}$. Hence, each of α's and β_j's is zero. That is $\mathbf{v} = \mathbf{0}$. ∎

We now prove another useful result.

Theorem 3.2.5. *Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a non-empty subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. If $T \subseteq LS(S)$ having more than k vectors then, T is a linearly dependent subset in $\mathbb{V}$.*

Proof. Let $T = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. As $\mathbf{w}_i \in LS(S)$, there exist $a_{ij} \in \mathbb{F}$ such that

$$\mathbf{w}_i = a_{i1}\mathbf{u}_1 + \cdots + a_{ik}\mathbf{u}_k, \text{ for } 1 \le i \le m.$$

So,

$$\begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} = \begin{bmatrix} a_{11}\mathbf{u}_1 + \cdots + a_{1k}\mathbf{u}_k \\ \vdots \\ a_{m1}\mathbf{u}_1 + \cdots + a_{mk}\mathbf{u}_k \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mk} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix}.$$

As $m > k$, using Corollary 2.4.6, the linear system $\mathbf{x}^T A = \mathbf{0}^T$ has a non-trivial solution, say $\mathbf{y} \neq \mathbf{0}$, *i.e.*, $\mathbf{y}^T A = \mathbf{0}^T$. Thus,

$$\mathbf{y}^T \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} = \mathbf{y}^T \left(A \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} \right) = (\mathbf{y}^T A) \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{0}^T \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{0}^T.$$

As $\mathbf{y} \neq \mathbf{0}$, a non-trivial linear combination of vectors in T is $\mathbf{0}$. Thus, the set T is linearly dependent subset of $\mathbb{V}$. ■

Corollary 3.2.6. *Fix $n \in \mathbb{N}$. Then, any subset S of $\mathbb{R}^n$ with $\mid S \mid \ge n+1$ is linearly dependent.*

Proof. Observe that $\mathbb{R}^n = LS(\{\mathbf{e}_1, \ldots, \mathbf{e}_n\})$, where $\mathbf{e}_i = I_n[:, i]$, is the i-th column of I_n. Hence, using Theorem 3.2.5, the required result follows. ■

Theorem 3.2.7. *Let S be a linearly independent subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, for any $\mathbf{v} \in \mathbb{V}$ the set $S \cup \{\mathbf{v}\}$ is linearly dependent if and only if $\mathbf{v} \in LS(S)$.*

Proof. Let us assume that $S \cup \{\mathbf{v}\}$ is linearly dependent. Then, there exist $\mathbf{v}_i$'s in S such that the linear system

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p + \alpha_{p+1}\mathbf{v} = \mathbf{0} \tag{3.2.3}$$

in the variables α_i's has a non-trivial solution, say $\alpha_i = c_i$, for $1 \le i \le p+1$. We claim that $c_{p+1} \neq 0$.

For, if $c_{p+1} = 0$ then, Equation (3.2.3) has a non-trivial solution corresponds to having a non-trivial solution of the linear system $\alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p = \mathbf{0}$ in the variables $\alpha_1, \ldots, \alpha_p$. This contradicts Proposition 3.2.3.2 as $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subseteq S$, a linearly independent set. Thus, $c_{p+1} \neq 0$ and we get

$$\mathbf{v} = -\frac{1}{c_{p+1}}(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \in LS(\mathbf{v}_1, \ldots, \mathbf{v}_p)$$

as $-\frac{c_i}{c_{p+1}} \in \mathbb{F}$, for $1 \le i \le p$. That is, $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

Now, assume that $\mathbf{v} \in LS(S)$. Then, there exists $\mathbf{v}_i \in S$ and $c_i \in \mathbb{F}$, not all zero, such that $\mathbf{v} = \sum\limits_{i=1}^{p} c_i\mathbf{v}_i$. Thus, the linear system $\alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p + \alpha_{p+1}\mathbf{v} = \mathbf{0}$ in the variables α_i's has a non-trivial solution $[c_1, \ldots, c_p, -1]$. Hence, $S \cup \{\mathbf{v}\}$ is linearly dependent. ■

We now state a very important corollary of Theorem 3.2.7 without proof. This result can also be used as an alternative definition of linear independence and dependence.

Corollary 3.2.8. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let S be a subset of $\mathbb{V}$ containing a non-zero vector $\mathbf{u}_1$.*

1. *If S is linearly dependent then, there exists k such that $LS(\mathbf{u}_1,\ldots,\mathbf{u}_k) = LS(\mathbf{u}_1,\ldots,\mathbf{u}_{k-1})$. Or equivalently, if S is a linearly dependent set then there exists a vector $\mathbf{u}_k$, for $k \geq 2$, which is a linear combination of the previous vectors.*

2. *If S linearly independent then, $\mathbf{v} \in \mathbb{V} \setminus LS(S)$ if and only if $S \cup \{\mathbf{v}\}$ is also a linearly independent subset of $\mathbb{V}$.*

3. *If S is linearly independent then, $LS(S) = \mathbb{V}$ if and only if each proper superset of S is linearly dependent.*

3.2.2 Application to Matrices

We start with our understanding of the RREF.

Theorem 3.2.9. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, the rows of A corresponding to the pivotal rows of RREF(A) are linearly independent. Also, the columns of A corresponding to the pivotal columns of RREF(A) are linearly independent.*

Proof. Let RREF$(A) = B$. Then, the pivotal rows of B are linearly independent due to the pivotal 1's. Now, let B_1 be the submatrix of B consisting of the pivotal rows of B. Also, let A_1 be the submatrix of A whose rows corresponds to the rows of B_1. As the RREF of a matrix is unique (see Corollary 2.2.18) there exists an invertible matrix Q such that $QA_1 = B_1$. So, if there exists $\mathbf{c} \neq \mathbf{0}$ such that $\mathbf{c}^T A_1 = \mathbf{0}^T$ then

$$\mathbf{0}^T = \mathbf{c}^T A_1 = \mathbf{c}^T(Q^{-1}B_1) = (\mathbf{c}^T Q^{-1})B_1 = \mathbf{d}^T B_1,$$

with $\mathbf{d}^T = \mathbf{c}^T Q^{-1} \neq \mathbf{0}^T$ as Q is an invertible matrix (see Theorem 2.5.1). This contradicts the linear independence of the rows of B_1.

Let $B[:,i_1],\ldots,B[:,i_r]$ be the pivotal columns of B. Then, they are linearly independent due to pivotal 1's. As $B = $ RREF(A), there exists an invertible matrix P such that $B = PA$. Then, the corresponding columns of A satisfy

$$[A[:,i_1],\ldots,A[:,i_r]] = [P^{-1}B[:,i_1],\ldots,P^{-1}B[:,i_r]] = P^{-1}[B[:,i_1],\ldots,B[:,i_r]].$$

As P is invertible, the systems $[A[:,i_1],\ldots,A[:,i_r]]\begin{bmatrix}x_1\\ \vdots\\ x_r\end{bmatrix} = \mathbf{0}$ and $[B[:,i_1],\ldots,B[:,i_r]]\begin{bmatrix}x_1\\ \vdots\\ x_r\end{bmatrix} = \mathbf{0}$ are row-equivalent. Thus, they have the same solution set. Hence, $\{A[:,i_1],\ldots,A[:,i_r]\}$ is linearly independent if and only if $\{B[:,i_1],\ldots,B[:,i_r]\}$ is linear independent. Thus, the required result follows. ∎

The next result follows directly from Theorem 3.2.9 and hence the proof is left to readers.

Corollary 3.2.10. *The following statements are equivalent for $A \in \mathbb{M}_n(\mathbb{C})$.*

1. *A is invertible.*

2. *The columns of A are linearly independent.*

3. *The rows of A are linearly independent.*

We give an example for better understanding.

Example 3.2.11. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$ with $\text{RREF}(A) = B = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

1. Then, $B[:,3] = -B[:,1] + 2B[:,2]$. Thus, $A[:,3] = -A[:,1] + 2A[:,2]$.
2. As the 1-st, 2-nd and 4-th columns of B are linearly independent, the set $\{A[:,1], A[:,2], A[:,4]\}$ is linearly independent.
3. Also, note that during the application of GJE, the 3-rd and 4-th rows were interchanged. Hence, the rows $A[1,:], A[2,:]$ and $A[4,:]$ are linearly independent.

3.2.3 Linear Independence and Uniqueness of Linear Combination

We end this section with a result that states that linear combination with respect to linearly independent set is unique.

Lemma 3.2.12. *Let S be a linearly independent subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, each $\mathbf{v} \in LS(S)$ is a unique linear combination of vectors from S.*

Proof. Suppose there exists $\mathbf{v} \in LS(S)$ with $\mathbf{v} \in LS(T_1), LS(T_2)$ with $T_1, T_2 \subseteq S$. Let $T_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $T_2 = \{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$, for some $\mathbf{v}_i$'s and $\mathbf{w}_j$'s in S. Define $T = T_1 \cup T_2$. Then, T is a subset of S. Hence, using Proposition 3.2.3, the set T is linearly independent. Let $T = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Then, there exist α_i's and β_j's in $\mathbb{F}$, not all zero, such that $\mathbf{v} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_p\mathbf{u}_p$ as well as $\mathbf{v} = \beta_1\mathbf{u}_1 + \cdots + \beta_p\mathbf{u}_p$. Equating the two expressions for $\mathbf{v}$ gives

$$(\alpha_1 - \beta_1)\mathbf{u}_1 + \cdots + (\alpha_p - \beta_p)\mathbf{u}_p = \mathbf{0}. \tag{3.2.4}$$

As T is a linearly independent subset of $\mathbb{V}$, the system $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$, in the variables $c_1, \ldots, c_p$, has only the trivial solution. Thus, in Equation (3.2.4), $\alpha_i - \beta_i = 0$, for $1 \le i \le p$. Thus, for $1 \le i \le p$, $\alpha_i = \beta_i$ and the required result follows. ■

Exercise **3.2.13.**

1. *Suppose $\mathbb{V}$ is a vector space over $\mathbb{R}$ as well as over $\mathbb{C}$. Then, prove that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linearly independent subset of $\mathbb{V}$ over $\mathbb{C}$ if and only if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, i\mathbf{u}_1, \ldots, i\mathbf{u}_k\}$ is a linearly independent subset of $\mathbb{V}$ over $\mathbb{R}$.*
2. *Is the set $\{1, x, x^2, \ldots\}$ a linearly independent subset of the vector space $\mathbb{C}[x]$ over $\mathbb{C}$?*
3. *Is the set $\{\mathbf{e}_{ij} | \ 1 \le i \le m, 1 \le j \le n\}$ a linearly independent subset of the vector space $\mathbb{M}_{m,n}(\mathbb{C})$ over $\mathbb{C}$ (see Definition 1.3.1.1)?*
4. *Let $\mathbb{W}$ be a subspace of a vector space $\mathbb{V}$ over $\mathbb{F}$. For $\mathbf{u}, \mathbf{v} \in \mathbb{V} \setminus \mathbb{W}$, define $K = LS(\mathbb{W}, \mathbf{u})$ and $M = LS(\mathbb{W}, \mathbf{v})$. Then, prove that $\mathbf{v} \in K$ if and only if $\mathbf{u} \in M$.*
5. *Prove that*

(a) *the rows/columns of $A \in M_n(\mathbb{C})$ are linearly independent if and only if* $\det(A) \neq 0$.

(b) *the rows/columns of $A \in M_n(\mathbb{C})$ span $\mathbb{C}^n$ if and only if A is an invertible matrix.*

(c) *the rows/columns of a skew-symmetric matrix A of odd order are linearly dependent.*

6. *Let $\mathbb{V}$ and $\mathbb{W}$ be subspaces of $\mathbb{R}^n$ such that $\mathbb{V} + \mathbb{W} = \mathbb{R}^n$ and $\mathbb{V} \cap \mathbb{W} = \{\mathbf{0}\}$. Prove that each $\mathbf{u} \in \mathbb{R}^n$ is uniquely expressible as $\mathbf{u} = \mathbf{v} + \mathbf{w}$, where $\mathbf{v} \in \mathbb{V}$ and $\mathbf{w} \in \mathbb{W}$.*

7. *Let $S_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $S_2 = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be subsets of a complex vector space $\mathbb{V}$. Also, let $\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} A$ for some matrix $A \in \mathbb{M}_n(\mathbb{C})$.*

   (a) *If $A = [a_{ij}]$ is invertible then S_1 is a linearly independent if and only if S_2 is linearly independent.*

   Hint: Suppose S_2 is linearly independent and consider the linear system $\sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i = \mathbf{0}$ in the variables α_i's. Then $\mathbf{0} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i = \sum\limits_{i=1}^{n} \alpha_i \left(\sum\limits_{j=1}^{n} (A^{-1})_{ji} \mathbf{w}_j\right) = \sum\limits_{j=1}^{n} \left(\sum\limits_{i=1}^{n} (A^{-1})_{ji} \alpha_i\right) \mathbf{w}_j$. As S_2 is linearly independent, $\sum\limits_{i=1}^{n} (A^{-1})_{ji} \alpha_i = 0$, for $1 \le j \le n$. Or equivalently $A^{-1} \begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_n\end{bmatrix} = \mathbf{0}$. Thus $\alpha_i = 0$ for all i and hence the set S_1 is linearly independent.

   (b) *If S_2 is linearly independent then prove that A is invertible. Further, in this case, the set S_1 is necessarily linearly independent.*

   Hint: Suppose A is not invertible. Then there exists $\mathbf{x}_0 = \begin{bmatrix}\mathbf{x}_{01} & \cdots & \mathbf{x}_{0n}\end{bmatrix}^T \neq \mathbf{0}$ such that $A\mathbf{x}_0 = \mathbf{0}$. Thus, we have obtained $\mathbf{x}_0 \neq \mathbf{0}$ such that

$$\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n\end{bmatrix} \begin{bmatrix}\mathbf{x}_{01} \\ \vdots \\ \mathbf{x}_{0n}\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} A \begin{bmatrix}\mathbf{x}_{01} \\ \vdots \\ \mathbf{x}_{0n}\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} \mathbf{0} = \mathbf{0},$$

   a contradiction to S_2 being a linearly independent set.

8. *Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{C}^n$ and $T = \{A\mathbf{u}_1, \ldots, A\mathbf{u}_n\}$, for some matrix $A \in \mathbb{M}_n(\mathbf{C})$.*

   (a) *If S is linearly dependent then prove that T is linear dependent.*

   (b) *If S is linearly independent then prove that T is linearly independent for every invertible matrix A.*

   (c) *If T is linearly independent then S is linearly independent. Further, in this case, the matrix A is necessarily invertible.*

---

9. *Let $S = \{(1,1,1,1)^T, (1,-1,1,2)^T, (1,1,-1,1)^T\} \subseteq \mathbb{R}^4$. Does $(1,1,2,1)^T \in LS(S)$? Furthermore, determine conditions on x, y, z and u such that $(x,y,z,u)^T \in LS(S)$.*

10. *Show that $S = \{(1,2,3)^T, (-2,1,1)^T, (8,6,10)^T\} \subseteq \mathbb{R}^3$ is linearly dependent.*

11. *Find $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^4$ 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.*

12. *Let $A \in M_n(\mathbb{R})$. Suppose $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $A\mathbf{x} = 3\mathbf{x}$ and $A\mathbf{y} = 2\mathbf{y}$. Then, 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* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ *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 above to prove the following.*

    (a) $A^2\mathbf{v} = 4\mathbf{v}$, *where* $\mathbf{v} = c\mathbf{y} + d\mathbf{z}$ *for any* $c, d \in \mathbb{R}$.

    (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* $AP = PD$.

3.3 Basis of a Vector Space

Definition 3.3.1. Let S be a subset of a set T. Then, S is said to be a **maximal subset** of T having property P if

1. S has property P and
2. no proper superset of S in T has property P.

Example 3.3.2. Let $T = \{2, 3, 4, 7, 8, 10, 12, 13, 14, 15\}$. Then, a maximal subset of T of consecutive integers is $S = \{2, 3, 4\}$. Other maximal subsets are $\{7, 8\}, \{10\}$ and $\{12, 13, 14, 15\}$. Note that $\{12, 13\}$ is not maximal. Why?

Definition 3.3.3. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, S is called a **maximal linearly independent** subset of $\mathbb{V}$ if

1. S is linearly independent and
2. no proper superset of S in $\mathbb{V}$ is linearly independent.

Example 3.3.4.

1. In $\mathbb{R}^3$, the set $S = \{\mathbf{e}_1, \mathbf{e}_2\}$ is linearly independent but not maximal as $S \cup \{(1, 1, 1)^T\}$ is a linearly independent set containing S.
2. In $\mathbb{R}^3$, $S = \{(1, 0, 0)^T, (1, 1, 0)^T, (1, 1, -1)^T\}$ is a maximal linearly independent set as S is linearly independent and any collection of 4 or more vectors from $\mathbb{R}^3$ is linearly dependent (see Corollary 3.2.6).
3. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$. Now, form the matrix $A = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ and let $B = \text{RREF}(A)$. Then, using Theorem 3.2.9, we see that if $B[:, i_1], \ldots, B[:, i_r]$ are the pivotal columns of B then $\{\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_r}\}$ is a maximal linearly independent subset of S.
4. Is the set $\{1, x, x^2, \ldots\}$ a maximal linearly independent subset of $\mathbb{C}[x]$ over $\mathbb{C}$?
5. Is the set $\{\mathbf{e}_{ij} |\ 1 \le i \le m, 1 \le j \le n\}$ a maximal linearly independent subset of $\mathbb{M}_{m,n}(\mathbb{C})$ over $\mathbb{C}$?

Theorem 3.3.5. *Let* $\mathbb{V}$ *be a vector space over* $\mathbb{F}$ *and* S *a linearly independent set in* $\mathbb{V}$. *Then,* S *is maximal linearly independent if and only if* $LS(S) = \mathbb{V}$.

Proof. Let $\mathbf{v} \in \mathbb{V}$. As S is linearly independent, using Corollary 3.2.8.2, the set $S \cup \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \in \mathbb{V} \setminus LS(S)$. Thus, the required result follows. ∎

Let $\mathbb{V} = LS(S)$ for some set S with $\mid S \mid = k$. Then, using Theorem 3.2.5, we see that if $T \subseteq \mathbb{V}$ is linearly independent then $\mid T \mid \leq k$. Hence, a maximal linearly independent subset of $\mathbb{V}$ can have at most k vectors. Thus, we arrive at the following important result.

Theorem 3.3.6. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let S and T be two finite maximal linearly independent subsets of $\mathbb{V}$. Then, $\mid S \mid = \mid T \mid$.*

Proof. By Theorem 3.3.5, S and T are maximal linearly independent if and only if $LS(S) = \mathbb{V} = LS(T)$. Now, use the previous paragraph to get the required result. ∎

Let $\mathbb{V}$ be a finite dimensional vector space. Then, by Theorem 3.3.6, the number of vectors in any two maximal linearly independent set is the same. We use this number to define the dimension of a vector space. We do so now.

Definition 3.3.7. Let $\mathbb{V}$ be a finite dimensional vector space over $\mathbb{F}$. Then, the number of vectors in any maximal linearly independent set is called the **dimension** of $\mathbb{V}$, denoted $\dim(\mathbb{V})$. By convention, $\dim(\{\mathbf{0}\}) = 0$.

Example 3.3.8.
1. As $\{1\}$ is a maximal linearly independent subset of $\mathbb{R}$, $\dim(\mathbb{R}) = 1$.
2. As $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} \subseteq \mathbb{R}^3$ is maximal linearly independent, $\dim(\mathbb{R}^3) = 3$.
3. As $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a maximal linearly independent subset in $\mathbb{R}^n$, $\dim(\mathbb{R}^n) = n$.
4. As $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a maximal linearly independent subset in $\mathbb{C}^n$ over $\mathbb{C}$, $\dim(\mathbb{C}^n) = n$.
5. Using Exercise 3.2.13.1, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n, i\mathbf{e}_1, \ldots, i\mathbf{e}_n\}$ is a maximal linearly independent subset in $\mathbb{C}^n$ over $\mathbb{R}$. Thus, as a real vector space, $\dim(\mathbb{C}^n) = 2n$.
6. As $\{\mathbf{e}_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ is a maximal linearly independent subset of $\mathbb{M}_{m,n}(\mathbb{C})$ over $\mathbb{C}$, $\dim(\mathbb{M}_{m,n}(\mathbb{C})) = mn$.

Definition 3.3.9. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, a maximal linearly independent subset of $\mathbb{V}$ is called a **basis/Hamel basis** of $\mathbb{V}$. The vectors in a basis are called **basis** vectors. By convention, a basis of $\{\mathbf{0}\}$ is the empty set.

Existence of Hamel basis

Definition 3.3.10. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, a subset S of $\mathbb{V}$ is called **minimal spanning** if $LS(S) = \mathbb{V}$ and no proper subset of S spans $\mathbb{V}$.

Remark 3.3.11 (Standard Basis)**.** *The readers should verify the statements given below.*

1. *All the maximal linearly independent set given in Example 3.3.8 form the standard basis of the respective vector space.*
2. $\{1, x, x^2, \ldots\}$ *is the standard basis of* $\mathbb{R}[x]$ *over* $\mathbb{R}$.
3. *Fix a positive integer n. Then,* $\{1, x, x^2, \ldots, x^n\}$ *is the standard basis of* $\mathbb{R}[x; n]$ *over* $\mathbb{R}$.

4. *Let $\mathbb{V} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A = A^T\}$. Then, $\mathbb{V}$ is a vector space over $\mathbb{R}$ with standard basis $\{\mathbf{e}_{ii}, \mathbf{e}_{ij} + \mathbf{e}_{ji} \mid 1 \le i < j \le n\}$.*

5. *Let $\mathbb{V} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A^T = -A\}$. Then, $\mathbb{V}$ is a vector space over $\mathbb{R}$ with standard basis $\{\mathbf{e}_{ij} - \mathbf{e}_{ji} \mid 1 \le i < j \le n\}$.*

Example 3.3.12. 1. Note that $\{-2\}$ is a basis and a minimal spanning subset in $\mathbb{R}$.

2. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \in \mathbb{R}^2$. Then, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ can neither be a basis nor a minimal spanning subset of $\mathbb{R}^2$.

3. $\{(1,1,-1)^T, (1,-1,1)^T, (-1,1,1)^T\}$ is a basis and a minimal spanning subset of $\mathbb{R}^3$.

4. Let $\mathbb{V} = \{(x,y,0)^T \mid \ x, y \in \mathbb{R}\} \subseteq \mathbb{R}^3$. Then, $\mathcal{B} = \{(1,0,0)^T, (1,3,0)^T\}$ is a basis of $\mathbb{V}$.

5. Let $\mathbb{V} = \{(x,y,z)^T \in \mathbb{R}^3 \mid \ x + y - z = 0\} \subseteq \mathbb{R}^3$. As each element $(x,y,z)^T \in \mathbb{V}$ satisfies $x + y - z = 0$. Or equivalently $z = x + y$, we see that

$$(x,y,z) = (x,y,x+y) = (x,0,x) + (0,y,y) = x(1,0,1) + y(0,1,1).$$

   Hence, $\{(1,0,1)^T, (0,1,1)^T\}$ forms a basis of $\mathbb{V}$.

6. Let $S = \{a_1, \ldots, a_n\}$. Then, $\mathbb{R}^S$ is a real vector space (see Example 3.1.4.9). For $1 \le i \le n$, define the functions

$$\mathbf{e}_i(a_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}.$$

   Then, prove that $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a linearly independent subset of $\mathbb{R}^S$ over $\mathbb{R}$. Is it a basis of $\mathbb{R}^S$ over $\mathbb{R}$? What can you say if S is a countable set?

7. Let $S = \mathbb{R}^n$ and consider the vector space $\mathbb{R}^S$ (see Example 3.1.4.9). For $1 \le i \le n$, define the functions $\mathbf{e}_i(\mathbf{x}) = \mathbf{e}_i\big((x_1, \ldots, x_n)\big) = x_i$. Then, verify that $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a linearly independent subset of $\mathbb{R}^S$ over $\mathbb{R}$. Is it a basis of $\mathbb{R}^S$ over $\mathbb{R}$?

8. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$. Define $A = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$. Then, using Example 3.3.4.3, we see that $\dim(LS(S)) = \text{Rank}(A)$. Further, using Theorem 3.2.9, the columns of A corresponding to the pivotal columns in $\text{RREF}(A)$ form a basis of $LS(S)$.

9. Recall the vector space $\mathcal{C}[a,b]$, where $a < b \in \mathbb{R}$. For each $\alpha \in [a,b]$, define

$$\mathbf{f}_\alpha(x) = x - \alpha, \text{ for all } x \in [a,b].$$

   Suppose α, β and γ are three distinct real numbers in $[a,b]$. Then prove that $\{\mathbf{f}_\alpha, \mathbf{f}_\beta, \mathbf{f})_\gamma\}$ is linearly dependent subset of $\mathcal{C}[a,b]$.

   Ans: Choose α, β and γ such that $\beta = a\alpha + (1-a)\gamma$, *i.e.*, β is a convex combination of α and γ. Then $a\mathbf{f}_\alpha + (1-a)\mathbf{f}_\gamma = \mathbf{f}_\beta$.

3.3.1 Main Results associated with Bases

Theorem 3.3.13. *Let $\mathbb{V}$ be a non-zero vector space over $\mathbb{F}$. Then, the following statements are equivalent.*

1. *$\mathcal{B}$ is a basis (maximal linearly independent subset) of $\mathbb{V}$.*
2. *$\mathcal{B}$ is linearly independent and spans $\mathbb{V}$.*
3. *$\mathcal{B}$ is a minimal spanning set in $\mathbb{V}$.*

Proof. $1 \Rightarrow 2$ By definition, every basis is a maximal linearly independent subset of $\mathbb{V}$. Thus, using Corollary 3.2.8.2, we see that $\mathcal{B}$ spans $\mathbb{V}$.

$2 \Rightarrow 3$ Let S be a linearly independent set that spans $\mathbb{V}$. As S is linearly independent, for any $\mathbf{x} \in S$, $\mathbf{x} \notin LS(S - \{\mathbf{x}\})$. Hence $LS(S - \{\mathbf{x}\}) \subsetneq LS(S) = \mathbb{V}$.

$3 \Rightarrow 1$ If $\mathcal{B}$ is linearly dependent then using Corollary 3.2.8.1, $\mathcal{B}$ is not minimal spanning. A contradiction. Hence, $\mathcal{B}$ is linearly independent.

We now need to show that $\mathcal{B}$ is a maximal linearly independent set. Since $LS(\mathcal{B}) = \mathbb{V}$, for any $\mathbf{x} \in \mathbb{V} \setminus \mathcal{B}$, using Corollary 3.2.8.2, the set $\mathcal{B} \cup \{\mathbf{x}\}$ is linearly dependent. That is, every proper superset of $\mathcal{B}$ is linearly dependent. Hence, the required result follows. ∎

Now, using Lemma 3.2.12, we get the following result.

Remark 3.3.14. *Let $\mathcal{B}$ be a basis of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, for each $\mathbf{v} \in \mathbb{V}$, there exist unique $\mathbf{u}_i \in \mathcal{B}$ and unique $\alpha_i \in \mathbb{F}$, for $1 \le i \le n$, such that $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i$.*

The next result is generally known as "every linearly independent set can be extended to form a basis of a finite dimensional vector space".

Theorem 3.3.15. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. If S is a linearly independent subset of $\mathbb{V}$ then there exists a basis T of $\mathbb{V}$ such that $S \subseteq T$.*

Proof. If $LS(S) = \mathbb{V}$, done. Else, choose $\mathbf{u}_1 \in \mathbb{V} \setminus LS(S)$. Thus, by Corollary 3.2.8.2, the set $S \cup \{\mathbf{u}_1\}$ is linearly independent. We repeat this process till we get n vectors in T as $\dim(\mathbb{V}) = n$. By Theorem 3.3.13, this T is indeed a required basis. ∎

3.3.2 Constructing a Basis of a Finite Dimensional Vector Space

We end this section with an algorithm which is based on the proof of the previous theorem.

Step 1: Let $\mathbf{v}_1 \in \mathbb{V}$ with $\mathbf{v}_1 \neq \mathbf{0}$. Then, $\{\mathbf{v}_1\}$ is linearly independent.

Step 2: If $\mathbb{V} = LS(\mathbf{v}_1)$, we have got a basis of $\mathbb{V}$. Else, pick $\mathbf{v}_2 \in \mathbb{V} \setminus LS(\mathbf{v}_1)$. Then, by Corollary 3.2.8.2, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

Step i: Either $\mathbb{V} = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$ or $LS(\mathbf{v}_1, \ldots, \mathbf{v}_i) \subsetneq \mathbb{V}$. In the first case, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is a basis of $\mathbb{V}$. Else, pick $\mathbf{v}_{i+1} \in \mathbb{V} \setminus LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$. Then, by Corollary 3.2.8.2, the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i+1}\}$ is linearly independent.

This process will finally end as $\mathbb{V}$ is a finite dimensional vector space.

EXERCISE **3.3.16.** 1. *Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, does the condition $\sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i = \mathbf{0}$ in α_i's imply that $\alpha_i = 0$, for $1 \le i \le n$?*

2. *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. Suppose $LS(S) = \mathbb{V}$ but S is not a linearly independent set. Then, does this imply that each $\mathbf{v} \in \mathbb{V}$ is expressible in more than one way as a linear combination of vectors from S? Is it possible to get a subset T of S such that T is a basis of $\mathbb{V}$ over $\mathbb{F}$? Give reasons for your answer.*

3. *Let $\mathbb{V}$ be a vector space of dimension n. Then,*

   (a) *prove that any set consisting of n linearly independent vectors forms a basis of $\mathbb{V}$.*

   (b) *prove that if S is a subset of $\mathbb{V}$ having n vectors with $LS(S) = \mathbb{V}$ then, S forms a basis of $\mathbb{V}$.*

4. *Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbb{C}^n$. Then, prove that the two matrices $B = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $C = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$ are invertible.*

5. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be an invertible matrix. Then, prove that the rows/columns of A form a basis of $\mathbb{C}^n$ over $\mathbb{C}$.*

6. *Let $\mathbb{W}_1$ and $\mathbb{W}_2$ be two subspaces of a finite dimensional vector space $\mathbb{V}$ such that $\mathbb{W}_1 \subseteq \mathbb{W}_2$. Then, prove that $\mathbb{W}_1 = \mathbb{W}_2$ if and only if $\dim(\mathbb{W}_1) = \dim(\mathbb{W}_2)$.*

7. *Let $\mathbb{W}_1$ be a non-trivial subspace of a finite dimensional vector space $\mathbb{V}$ over $\mathbb{F}$. Then, prove that there exists a subspace $\mathbb{W}_2$ of $\mathbb{V}$ such that*

$$\mathbb{W}_1 \cap \mathbb{W}_2 = \{\mathbf{0}\}, \mathbb{W}_1 + \mathbb{W}_2 = \mathbb{V} \text{ and } \dim(\mathbb{W}_2) = \dim(\mathbb{V}) - \dim(\mathbb{W}_1).$$

   Also, prove that for each $\mathbf{v} \in \mathbb{V}$ there exist unique vectors $\mathbf{w}_1 \in \mathbb{W}_1$ and $\mathbf{w}_2 \in \mathbb{W}_2$ with $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$. The subspace $\mathbb{W}_2$ is called the **complementary subspace** *of $\mathbb{W}_1$ in $\mathbb{V}$.*

8. *Let $\mathbb{V}$ be a finite dimensional vector space over $\mathbb{F}$. If $\mathbb{W}_1$ and $\mathbb{W}_2$ are two subspaces of $\mathbb{V}$ such that $\mathbb{W}_1 \cap \mathbb{W}_2 = \{\mathbf{0}\}$ and $\dim(\mathbb{W}_1) + \dim(\mathbb{W}_2) = \dim(\mathbb{V})$ then prove that $\mathbb{W}_1 + \mathbb{W}_2 = \mathbb{V}$.*

9. *Consider the vector space $\mathcal{C}([-\pi, \pi])$ over $\mathbb{R}$. For each $n \in \mathbb{N}$, define $\mathbf{e}_n(x) = \sin(nx)$. Then, prove that $S = \{\mathbf{e}_n \mid n \in \mathbb{N}\}$ is linearly independent. [Hint: Need to show that every finite subset of S is linearly independent. So, on the contrary assume that there exists $\ell \in \mathbb{N}$ and functions $\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_\ell}$ such that $\alpha_1 \mathbf{e}_{k_1} + \cdots + \alpha_\ell \mathbf{e}_{k_\ell} = \mathbf{0}$, for some $\alpha_t \neq 0$ with $1 \leq t \leq \ell$. But, the above system is equivalent to looking at $\alpha_1 \sin(k_1 x) + \cdots + \alpha_\ell \sin(k_\ell x) = \mathbf{0}$ for all $x \in [-\pi, \pi]$. Now in the integral*

$$\int_{-\pi}^{\pi} \sin(mx) \left(\alpha_1 \sin(k_1 x) + \cdots + \alpha_\ell \sin(k_\ell x)\right) \, dx = \int_{-\pi}^{\pi} \sin(mx)\mathbf{0} \, dx = 0$$

   replace m with k_i's to show that $\alpha_i = 0$, for all $i, 1 \leq i \leq \ell$. This gives the required contradiction.]

10. *Is the set $\{1, \sin(x), \cos(x), \sin(2x), \cos(2x), \sin(3x), \cos(3x), \ldots\}$ a linearly subset of the vector space $\mathcal{C}([-\pi, \pi], \mathbb{R})$ over $\mathbb{R}$?*

---

11. *Find a basis of $\mathbb{R}^3$ containing the vector $(1, 1, -2)^T$.*

12. *Find a basis of* $\mathbb{R}^3$ *containing the vector* $(1,1,-2)^T$ *and* $(1,2,-1)^T$.

13. *Is it possible to find a basis of* $\mathbb{R}^4$ *containing the vectors* $(1,1,1,-2)^T$, $(1,2,-1,1)^T$ *and* $(1,-2,7,-11)^T$?

14. *Show that* $\mathcal{B} = \{(1,0,1)^T, (1,i,0)^T, (1,1,1-i)^T\}$ *is a basis of* $\mathbb{C}^3$ *over* $\mathbb{C}$.

15. *Find a basis of* $\mathbb{C}^3$ *over* $\mathbb{R}$ *containing the basis* $\mathcal{B}$ *given in Example 3.3.16.14.*

16. *Determine a basis and dimension of* $W = \{(x,y,z,w)^T \in \mathbb{R}^4 \mid x+y-z+w=0\}$.

17. *Find a basis of* $\mathbb{V} = \{(x,y,z,u) \in \mathbb{R}^4 \mid x-y-z=0, x+z-u=0\}$.

18. *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* $\mathbb{V} = \{\mathbf{x} \in \mathbb{R}^5 \mid A\mathbf{x} = \mathbf{0}\}$.

19. *Let* $\mathbf{u}^T = (1,1,-2), \mathbf{v}^T = (-1,2,3)$ *and* $\mathbf{w}^T = (1,10,1)$. *Find a basis of* $LS(\mathbf{u},\mathbf{v},\mathbf{w})$. *Determine a geometrical representation of* $LS(\mathbf{u},\mathbf{v},\mathbf{w})$.

20. *Is the set* $\mathbb{W} = \{p(x) \in \mathbb{R}[x;4] \mid p(-1) = p(1) = 0\}$ *a subspace of* $\mathbb{R}[x;4]$? *If yes, find its dimension.*

3.4 Fundamental Subspaces Associated with a Matrix

In this section, we will study results that are intrinsic to the understanding of linear algebra from the point of view of matrices. So, we start with defining the four fundamental subspaces associated with a matrix.

Definition 3.4.1. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, we define the four fundamental subspaces associated with A as

1. $\text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n\}$ is a subspace of $\mathbb{C}^m$, called the **Column space**, and is the linear span of the columns of A.
2. $\text{Row}(A) = \text{Col}(A^T) = \{A^T\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^m\}$ is a subspace of $\mathbb{C}^n$, called the **row space** of A and is the linear span of the rows of A.
3. $\text{Col}(A^*) = \{A^*\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^m\}$.
4. $\text{Null}(A) = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \mathbf{0}\}$, called the **Null space** of A.
5. $\text{Null}(A^*) = \{\mathbf{x} \in \mathbb{C}^m \mid A^*\mathbf{x} = \mathbf{0}\}$.

Remark 3.4.2. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$.

1. *Then,* $\text{Col}(A)$ *is a subspace of* $\mathbb{C}^m$ *and* $\text{Col}(A^*)$ *is a subspace of* $\mathbb{C}^n$.
2. *Then,* $\text{Null}(A)$ *is a subspace of* $\mathbb{C}^n$ *and* $\text{Null}(A^*)$ *is a subspace of* $\mathbb{C}^m$.

Example 3.4.3. 1. Compute the fundamental subspaces for $A = \begin{bmatrix} 1 & 1 & 1 & -2 \\ 1 & 2 & -1 & 1 \\ 1 & -2 & 7 & -11 \end{bmatrix}$.

Solution: Verify the following

(a) $\text{Row}(A) = \{(x,y,z,u)^T \in \mathbb{C}^4 \mid 3x-2y=z, 5x-3y+u=0\}$.

(b) $\text{Col}(A) = \{(x,y,z)^T \in \mathbb{C}^3 \mid 4x-3y-z=0\}$.

(c) $\text{Null}(A) = \{(x,y,z,u)^T \in \mathbb{C}^4 \mid x+3z-5u=0, y-2z+3u=0\}$.

(d) $\text{Null}(A^T) = \{(x,y,z)^T \in \mathbb{C}^3 \mid x+4z=0, y-3z=0\}$.

2. Let $A = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 2 \\ 2 & 0 & 1 & 1 \end{bmatrix}$. Then, verify that

(a) $\text{Col}(A) = \{\mathbf{x} = (x_1,x_2,x_3)^T \in \mathbb{R}^3 \mid x_1+x_2-x_3=0\}$.

(b) $\text{Row}(A) = \{\mathbf{x} = (x_1,x_2,x_3,x_4)^T \in \mathbb{R}^4 \mid x_1-x_2-2x_3=0, x_1-3x_2-2x_4=0\}$.

(c) $\text{Null}(A) = LS(\{(1,-1,-2,0)^T, (1,-3,0,-2)^T\})$.

(d) $\text{Null}(A^T) = LS((1,1,-1)^T)$.

3. Let $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 & 3 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$. Find a basis and dimension of $\text{Null}(A)$.

Solution: Writing the basic vairables x_1, x_3 and x_6 in terms of the free variables x_2, x_4, x_5 and x_7, we get $x_1 = x_7 - x_2 - x_4 - x_5$, $x_3 = 2x_7 - 2x_4 - 3x_5$ and $x_6 = -x_7$. Hence, the solution set has the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} x_7 - x_2 - x_4 - x_5 \\ x_2 \\ 2x_7 - 2x_4 - 3x_5 \\ x_4 \\ x_5 \\ -x_7 \\ x_7 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}. \quad (3.4.1)$$

Now, let $\mathbf{u}_1^T = \begin{bmatrix} -1,1,0,0,0,0,0 \end{bmatrix}$, $\mathbf{u}_2^T = \begin{bmatrix} -1,0,-2,1,0,0,0 \end{bmatrix}$, $\mathbf{u}_3^T = \begin{bmatrix} -1,0,-3,0,1,0,0 \end{bmatrix}$ and $\mathbf{u}_4^T = \begin{bmatrix} 1,0,2,0,0,-1,1 \end{bmatrix}$. Then, $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is a basis of $\text{Null}(A)$. The reasons for S to be a basis are as follows:

(a) By Equation (3.4.1) $\text{Null}(A) = LS(S)$.

(b) For Linear independence, the homogeneous system $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 = \mathbf{0}$ in the variables c_1, c_2, c_3 and c_4 has only the trivial solution as

i. $\mathbf{u}_4$ is the only vector with a nonzero entry at the 7-th place ($\mathbf{u}_4$ corresponds to x_7) and hence $c_4 = 0$.

ii. $\mathbf{u}_3$ is the only vector with a nonzero entry at the 5-th place ($\mathbf{u}_3$ corresponds to x_5) and hence $c_3 = 0$.

iii. Similar arguments hold for the variables c_2 and c_1.

Remark 3.4.4. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$. Then, in Example* **??***, observe that the direction ratios of normal vectors of* $\text{Col}(A)$ *matches with vector in* $\text{Null}(A^T)$*. Similarly, the direction ratios of normal vectors of* $\text{Row}(A)$ *matches with vectors in* $\text{Null}(A)$*. Are these true in the general setting? Do similar relations hold if $A \in \mathbb{M}_{m,n}(\mathbb{C})$? We will come back to these spaces again and again.*

Exercise **3.4.5.**

1. *For the matrices given below, determine* $\text{Col}(A)$, $\text{Row}(A)$, $\text{Null}(A)$, $\text{Null}(A^T)$*. Further, find the dimensions of all the vector subspaces so obtained.*

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 2 & 2 & 2 & 4 \\ 2 & -2 & 4 & 0 & 8 \\ 4 & 2 & 5 & 6 & 10 \end{bmatrix}, \; B = \begin{bmatrix} 2 & 4 & 0 & 6 \\ -1 & 0 & -2 & 5 \\ -3 & -5 & 1 & -4 \\ -1 & -1 & 1 & 2 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & i & 2i \\ i & -2 & -3 \\ 1 & 1 & 1+i \end{bmatrix}.$$

   Ans: Verify that $\text{Col}(C) = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid (2+i)x_1 - (1-i)x_2 - x_3 = 0\}$.
   $\text{Col}(C^*) = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid ix_1 - x_2 + x_3 = 0\}$. $\text{Null}(C) = LS((i, 1, -1)^T)$.
   $\text{Null}(C^*) = LS((-2+i, 1+i, 1)^T)$.

2. *Let $A = [X \; Y]$. Then, determine the condition under which* $\text{Col}(X) = \text{Col}(Y)$.

The next result is a re-writing of the results on system of linear equations. We give the proof for the sake of completeness.

Lemma 3.4.6. *Let $A \in M_{m\times n}(\mathbb{C})$ and let E be an elementary matrix. If*

1. *$B = EA$ then*
   (a) $\text{Null}(A) = \text{Null}(B)$, $\text{Row}(A) = \text{Row}(B)$. *Thus, the dimensions of the corresponding spaces are equal.*
   (b) $\text{Null}(\overline{A}) = \text{Null}(\overline{B})$, $\text{Row}(\overline{A}) = \text{Row}(\overline{B})$. *Thus, the dimensions of the corresponding spaces are equal.*

2. *$B = AE$ then*
   (a) $\text{Null}(A^*) = \text{Null}(B^*)$, $\text{Col}(\overline{A}) = \text{Col}(\overline{B})$. *Thus, the dimensions of the corresponding spaces are equal.*
   (b) $\text{Null}(A^T) = \text{Null}(B^T)$, $\text{Col}(A) = \text{Col}(B)$. *Thus, the dimensions of the corresponding spaces are equal.*

Proof. Part 1a: Let $\mathbf{x} \in \text{Null}(A)$. Then, $B\mathbf{x} = EA\mathbf{x} = E\mathbf{0} = \mathbf{0}$. So, $\text{Null}(A) \subseteq \text{Null}(B)$. Further, if $\mathbf{x} \in \text{Null}(B)$, then $A\mathbf{x} = (E^{-1}E)A\mathbf{x} = E^{-1}(EA)\mathbf{x} = E^{-1}B\mathbf{x} = E^{-1}\mathbf{0} = \mathbf{0}$. Hence, $\text{Null}(B) \subseteq \text{Null}(A)$. Thus, $\text{Null}(A) = \text{Null}(B)$.

Let us now prove $\text{Row}(A) = \text{Row}(B)$. So, let $\mathbf{x}^T \in \text{Row}(A)$. Then, there exists $\mathbf{y} \in \mathbb{C}^m$ such that $\mathbf{x}^T = \mathbf{y}^T A$. Thus, $\mathbf{x}^T = \left(\mathbf{y}^T E^{-1}\right) EA = \left(\mathbf{y}^T E^{-1}\right) B$ and hence $\mathbf{x}^G \in \text{Row}(B)$. That is, $\text{Row}(A) \subseteq \text{Row}(B)$. A similar argument gives $\text{Row}(B) \subseteq \text{Row}(A)$ and hence the required result follows.

Part 1b: E is invertible implies $\overline{E}$ is invertible and $\overline{B} = \overline{EA}$. Thus, an argument similar to the previous part gives us the required result.

For Part 2, note that $B^* = E^*A^*$ and E^* is invertible. Hence, an argument similar to the first part gives the required result. ∎

Let $\mathbb{W}_1$ and $\mathbb{W}_1$ be two subspaces of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then, recall that (see Exercise 3.1.24.4d) $\mathbb{W}_1 + \mathbb{W}_2 = \{\mathbf{u}+\mathbf{v} \mid \mathbf{u} \in \mathbb{W}_1, \ \mathbf{v} \in \mathbb{W}_2\} = LS(\mathbb{W}_1 \cup \mathbb{W}_2)$ is the smallest subspace of $\mathbb{V}$ containing both $\mathbb{W}_1$ and $\mathbb{W}_2$. We now state a result similar to a result in Venn diagram that states $\mid A \mid + \mid B \mid = \mid A \cup B \mid + \mid A \cap B \mid$, whenever the sets A and B are finite (for a proof, see Appendix 7.4.1).

Theorem 3.4.7. *Let V be a finite dimensional vector space over $\mathbb{F}$. If $\mathbb{W}_1$ and $\mathbb{W}_2$ are two subspaces of V then*

$$\dim(\mathbb{W}_1) + \dim(\mathbb{W}_2) = \dim(\mathbb{W}_1 + \mathbb{W}_2) + \dim(\mathbb{W}_1 \cap \mathbb{W}_2). \tag{3.4.2}$$

For better understanding, we give an example for finite subsets of $\mathbb{R}^n$. The example uses Theorem 3.2.9 to obtain bases of $LS(S)$, for different choices S. The readers are advised to see Example 3.2.9 before proceeding further.

Example 3.4.8. Let $\mathbb{V} = \{(v,w,x,y,z)^T \in \mathbb{R}^5 \mid v+x+z = 3y\}$ and $\mathbb{W} = \{(v,w,x,y,z)^T \in \mathbb{R}^5 \mid w-x=z, v=y\}$. Find bases of $\mathbb{V}$ and $\mathbb{W}$ containing a basis of $\mathbb{V}\cap\mathbb{W}$.

Solution: One can first find a basis of $\mathbb{V}\cap\mathbb{W}$ and then heuristically add a few vectors to get bases for $\mathbb{V}$ and $\mathbb{W}$, separately.

Alternatively, First find bases of $\mathbb{V}, \mathbb{W}$ and $\mathbb{V}\cap\mathbb{W}$, say $\mathcal{B}_V, \mathcal{B}_W$ and $\mathcal{B}$. Now, consider $S = \mathcal{B}\cup\mathcal{B}_V$. This set is linearly dependent. So, obtain a linearly independent subset of S that contains all the elements of $\mathcal{B}$. Similarly, do for $T = \mathcal{B}\cup\mathcal{B}_W$.

So, we first find a basis of $\mathbb{V}\cap\mathbb{W}$. Note that $(v,w,x,y,z)^T \in \mathbb{V}\cap\mathbb{W}$ if v,w,x,y and z satisfy $v+x-3y+z=0$, $w-x-z=0$ and $v=y$. The solution of the system is given by

$$(v,w,x,y,z)^T = (y,2y,x,y,2y-x)^T = y(1,2,0,1,2)^T + x(0,0,1,0,-1)^T.$$

Thus, $\mathcal{B} = \{(1,2,0,1,2)^T, (0,0,1,0,-1)^T\}$ is a basis of $\mathbb{V}\cap\mathbb{W}$. Similarly, a basis of $\mathbb{V}$ is given by $\mathcal{C} = \{(-1,0,1,0,0)^T, (0,1,0,0,0)^T, (3,0,0,1,0)^T, (-1,0,0,0,1)^T\}$ and that of W is given by $\mathcal{D} = \{(1,0,0,1,0)^T, (0,1,1,0,0)^T, (0,1,0,0,1)^T\}$. To find the required basis form a matrix whose rows are the vectors in $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$ (see Equation(3.4.3)) and apply row operations other than E_{ij}. Then, after a few row operations, we get

$$\left[\begin{array}{ccccc} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ \hline -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccccc} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{3.4.3}$$

Thus, a required basis of $\mathbb{V}$ is $\{(1,2,0,1,2)^T, (0,0,1,0,-1)^T, (0,1,0,0,0)^T, (0,0,0,1,3)^T\}$. Similarly, a required basis of W is $\{(1,2,0,1,2)^T, (0,0,1,0,-1)^T, (0,1,0,0,1)^T\}$.

Exercise **3.4.9.**

1. *Give an example to show that if A and B are equivalent then* $\text{Col}(A)$ *need not equal* $\text{Col}(B)$.
2. *Let* $\mathbb{V} = \{(x,y,z,w)^T \in \mathbb{R}^4 \mid x+y-z+w=0, x+y+z+w=0, x+2y=0\}$ *and* $W = \{(x,y,z,w)^T \in \mathbb{R}^4 \mid x-y-z+w=0, x+2y-w=0\}$ *be two subspaces of* $\mathbb{R}^4$. *Think of a method to find bases and dimensions of* $\mathbb{V}$, $\mathbb{W}$, $\mathbb{V} \cap W$ *and* $\mathbb{V}+\mathbb{W}$.
3. *Let* $\mathbb{W}_1$ *and* $\mathbb{W}_2$ *be two subspaces of a vector space* $\mathbb{V}$. *If* $\dim(W_1)+\dim(W_2) > \dim(\mathbb{V})$, *then prove that* $\dim(\mathbb{W}_1 \cap \mathbb{W}_2) \geq 1$.
4. *Let* $A \in M_{m\times n}(\mathbb{C})$ *with* $m < n$. *Prove that the columns of* A *are linearly dependent.*

We now prove the rank-nullity theorem and give some of it's consequences.

Theorem 3.4.10 (Rank-Nullity Theorem)**.** *Let* $A \in M_{m\times n}(\mathbb{C})$. *Then,*

$$\dim(\text{Col}(A)) + \dim(\text{Null}(A)) = n. \tag{3.4.4}$$

Proof. Let $\dim(\text{Null}(A)) = r \leq n$ and let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis of $\text{Null}(A)$. Since $\mathcal{B}$ is a linearly independent set in $\mathbb{R}^n$, extend it to get $\mathcal{C} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ as a basis of $\mathbb{R}^n$. Then,

$$\begin{aligned} \text{Col}(A) &= LS(A\mathcal{B}) = LS(A\mathbf{u}_1, \ldots, A\mathbf{u}_n) \\ &= LS(\mathbf{0}, \ldots, \mathbf{0}, A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n) = LS(A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n). \end{aligned}$$

So, $\mathcal{D} = \{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ spans $\text{Col}(A)$. We further need to show that $\mathcal{D}$ is linearly independent. So, consider the linear system

$$\alpha_1 A\mathbf{u}_{r+1} + \cdots + \alpha_{n-r} A\mathbf{u}_n = \mathbf{0} \Leftrightarrow A(\alpha_1\mathbf{u}_{r+1} + \cdots + \alpha_{n-r}\mathbf{u}_n) = \mathbf{0} \tag{3.4.5}$$

in the variables $\alpha_1, \ldots, \alpha_{n-r}$. Thus, $\alpha_1\mathbf{u}_{r+1} + \cdots + \alpha_{n-r}\mathbf{u}_n \in \text{Null}(A) = LS(\mathcal{B})$. Therefore, there exist scalars β_i, $1 \leq i \leq r$, such that $\sum\limits_{i=1}^{n-r} \alpha_i \mathbf{u}_{r+i} = \sum\limits_{j=1}^{r} \beta_j \mathbf{u}_j$. Or equivalently,

$$\beta_1\mathbf{u}_1 + \cdots + \beta_r\mathbf{u}_r - \alpha_1\mathbf{u}_{r+1} - \cdots - \alpha_{n-r}\mathbf{u}_n = \mathbf{0}. \tag{3.4.6}$$

Equation (3.4.6) is a linear system in vectors from $\mathcal{C}$ with α_i's and β_j's as unknowns. As $\mathcal{C}$ is a linearly independent set, the only solution of Equation (3.4.6) is

$$\alpha_i = 0, \text{ for } 1 \leq i \leq n-r \text{ and } \beta_j = 0, \text{ for } 1 \leq j \leq r.$$

In other words, we have shown that the only solution of Equation (3.4.5) is the trivial solution. Hence, $\{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ is a basis of $\text{Col}(A)$. Thus, the required result follows. ∎

Theorem 3.4.10 is part of what is known as the fundamental theorem of linear algebra (see Theorem 3.4.13). The following are some of the consequences of the rank-nullity theorem. The proofs are left as an exercise for the reader.

Exercise **3.4.11.** *1. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$.*

(a) If $n > m$ then the system $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions,

(b) If $n < m$ then there exists $\mathbf{b} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ such that $A\mathbf{x} = \mathbf{b}$ is inconsistent.

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) $\dim(\textsc{Col}(A)) = k$.

(e) There exists a $k \times k$ submatrix B of A with $\det(B) \neq 0$. Further, the determinant of every $(k+1) \times (k+1)$ submatrix of A is zero.

(f) There exists a linearly independent subset $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ of $\mathbb{R}^m$ such that the system $A\mathbf{x} = \mathbf{b}_i$, for $1 \leq i \leq k$, is consistent.

(g) $\dim(\textsc{Null}(A)) = n - k$.

3. Let $A = \sum\limits_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*$, for some $\mathbf{x}_i \in \mathbb{C}^m$ and $\mathbf{y}_i \in \mathbb{C}^n$. Then does it imply that $\mathsf{Rank}(A) \leq k$*?*

We end this section by proving the fundamental theorem of linear algebra. We start with the following result.

Lemma 3.4.12. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$. Then,* $\textsc{Null}(A) = \textsc{Null}(A^T A)$.

Proof. Let $\mathbf{x} \in \textsc{Null}(A)$. Then, $A\mathbf{x} = \mathbf{0}$. So, $(A^T A)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}$. Thus, $\mathbf{x} \in \textsc{Null}(A^T A)$. That is, $\textsc{Null}(A) \subseteq \textsc{Null}(A^T A)$.

Suppose that $\mathbf{x} \in \textsc{Null}(A^T A)$. Then, $(A^T A)\mathbf{x} = \mathbf{0}$ and $0 = \mathbf{x}^T\mathbf{0} = \mathbf{x}^T(A^T A)\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2$. Thus, $A\mathbf{x} = \mathbf{0}$ and the required result follows. ■

Let $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$. Then $\mathbf{u}$ is said to be **orthogonal** to $\mathbf{v}$ if $\mathbf{u}^*\mathbf{v} = 0$ (dot product in case the vectors are from $\mathbb{R}^n$). For a subset S of $\mathbb{C}^n$, the **orthogonal complement** of S, denoted $S^\perp$ is defined as

$$S^\perp = \{\mathbf{x} \in \mathbb{C}^n : \mathbf{x}^*\mathbf{s} = 0 \text{ for all } \mathbf{s} \in S\}.$$

As an exercise prove the following:

1. $S^\perp$ is a subspace of $\mathbb{C}^n$.
2. $S^\perp = (LS(S))^\perp$

Theorem 3.4.13 (Fundamental Theorem of Linear Algebra)**.** *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,*

1. $\dim(\textsc{Null}(A)) + \dim(\textsc{Col}(A)) = n$.

2. $\textsc{Null}(A) = \big(\textsc{Col}(A^*)\big)^\perp$ *and* $\textsc{Null}(A^*) = \big(\textsc{Col}(A)\big)^\perp$.

3. $\dim(\textsc{Col}(A)) = \dim(\textsc{Col}(A^*))$. *Or equivalently, Row-rank(A) = Column-rank(A).*

Proof. Part 1: Proved in Theorem 3.4.10.

Part 2: We first prove that $\mathrm{Null}(A) \subseteq \mathrm{Col}(A^*)^{\perp}$. Let $\mathbf{x} \in \mathrm{Null}(A)$. Then, $A\mathbf{x} = \mathbf{0}$ and

$$0 = \mathbf{u}^*\mathbf{0} = \mathbf{u}^*(A\mathbf{x}) = \mathbf{u}^*A\mathbf{x} = (A^*\mathbf{u})^*\mathbf{x}, \text{ for all } \mathbf{u} \in \mathbb{C}^n.$$

But $\mathrm{Col}(A^*) = \{A^*\mathbf{u} \mid \mathbf{u} \in \mathbb{C}^n\}$. Thus, $\mathbf{x} \in \mathrm{Col}(A^*)^{\perp}$ and $\mathrm{Null}(A) \subseteq \mathrm{Col}(A^*)^{\perp}$.

We now prove that $\mathrm{Col}(A^*)^{\perp} \subseteq \mathrm{Null}(A)$. Let $\mathbf{x} \in \mathrm{Col}(A^*)^{\perp}$. Then, for every $\mathbf{y} \in \mathbb{C}^n$, $A^*\mathbf{y} \in \mathrm{Col}(A^*)$ and hence

$$0 = (A^*\mathbf{y})^*\mathbf{x} = \mathbf{y}^*(A^*)^*\mathbf{x} = \mathbf{y}^*A\mathbf{x}.$$

In particular, for $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^n$, we get $\|A\mathbf{x}\|^2 = 0$. Hence $A\mathbf{x} = \mathbf{0}$. That is, $\mathbf{x} \in \mathrm{Null}(A)$. 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 required result.

Hence the proof of the fundamental theorem is complete. ■

Remark 3.4.14. *Theorem 3.4.13.2 implies that* $\mathrm{Null}(A) = \left(\mathrm{Col}(A^*)\right)^{\perp}$. *This statement is just stating the usual fact that if* $\mathbf{x} \in \mathrm{Null}(A)$ *then* $A\mathbf{x} = \mathbf{0}$ *and hence the usual dot product of every row of* A *with* $\mathbf{x}$ *equals* 0.

As an implication of Theorem 3.4.13.2 and Theorem 3.4.13.3, we show the existence of an invertible linear map $T : \mathrm{Col}(A^*) \to \mathrm{Col}(A)$.

Corollary 3.4.15. *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Then, the function* $T : \mathrm{Col}(A^*) \to \mathrm{Col}(A)$ *defined by* $T(\mathbf{x}) = A\mathbf{x}$ *is invertible.*

Proof. In view of Theorem 3.4.13.3 and the rank-nullity theorem, we just need to show that the map is one-one. So, suppose that there exist $\mathbf{x}, \mathbf{y} \in \mathrm{Col}(A^*)$ such that $T(\mathbf{x}) = T(\mathbf{y})$. Or equivalently, $A\mathbf{x} = A\mathbf{y}$. Thus, $\mathbf{x} - \mathbf{y} \in \mathrm{Null}(A) = (\mathrm{Col}(A^*))^{\perp}$ (by Theorem 3.4.13.2). Therefore, $\mathbf{x} - \mathbf{y} \in (\mathrm{Col}(A^*))^{\perp} \cap \mathrm{Col}(A^*) = \{\mathbf{0}\}$. Thus, $\mathbf{x} = \mathbf{y}$ and hence the map is one-one. Thus, the required result follows. ■

The readers should look at Example 3.4.3 and Remark 3.4.4. We give one more example.

Example 3.4.16. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}$. Then, verify that

1. $\{(0,1,1)^T, (1,1,2)^T\}$ is a basis of $\mathrm{Col}(A)$.
2. $\{(1,1,-1)^T\}$ is a basis of $\mathrm{Null}(A^T)$.
3. $\mathrm{Null}(A^T) = (\mathrm{Col}(A))^{\perp}$.

Exercise **3.4.17.** 1. *Find distinct subspaces* $\mathbb{W}_1$ *and* $\mathbb{W}_2$

   (a) *in* $\mathbb{R}^2$ *such that* $\mathbb{W}_1$ *and* $\mathbb{W}_2$ *are orthogonal but not orthogonal complement.*

   (b) *in* $\mathbb{R}^3$ *such that* $\mathbb{W}_1 \neq \{\mathbf{0}\}$ *and* $\mathbb{W}_2 \neq \{\mathbf{0}\}$ *are orthogonal, but not orthogonal complement.*

2. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$. *Then,* $\textsc{Null}(A) = \textsc{Null}(A^*A)$.

3. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$. *Then,* $\textsc{Col}(A) = \textsc{Col}(A^TA)$.

4. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$. *Then,* $Rank(A) = n$ *if and only if* $Rank(A^TA) = n$.

5. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$. *Then, for every*

   (a) $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \mathbf{u} + \mathbf{v}$, *where* $\mathbf{u} \in \textsc{Col}(A^T)$ *and* $\mathbf{v} \in \textsc{Null}(A)$ *are unique.*

   (b) $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{y} = \mathbf{w} + \mathbf{z}$, *where* $\mathbf{w} \in \textsc{Col}(A)$ *and* $\mathbf{z} \in \textsc{Null}(A^T)$ *are unique.*

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."

3.5 Ordered Bases

Let $\mathbb{V}$ be a vector space of dimension n over $\mathbb{F}$. Then, in the previous class, we learnt that $\mathbb{V}$ is isomorphic to $\mathbb{F}^n$. So, if it should be possible to write the elements of $\mathbb{V}$ as an n-tuple. Further, our problem may require us to look at a subspace $\mathbb{W}$ of $\mathbb{V}$ whose dimension is very small as compared to the dimension of $\mathbb{V}$. It may also be possible that a basis of $\mathbb{W}$ may not look like a standard basis where the coefficient of $\mathbf{e}_i$ gave the i-th component of the vector. In the above cases, it is helpful to fix the vectors in a particular order and then concentrate only on the coefficients of the vectors as was done for the system of linear equations where we didn't worry about the unknowns. We start with the following example. Note that we will be using 'small brackets' in place of 'braces' to represent a basis.

Example 3.5.1. 1. Let $f(x) = 1 - x^2 \in \mathbb{R}[x;2]$. If $\mathcal{B} = (1, x, x^2)$ be a basis of $\mathbb{R}[x;2]$ then,

$$f(x) = \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

2. Consider the first three Legendre polynomials

$$\mathbf{P}_0(x) = 1, \ \mathbf{P}_1(x) = x \text{ and } \mathbf{P}_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \text{ for all } x \in [-1, \ 1].$$

   Then $\mathcal{B} = (\mathbf{P}_0(x), \mathbf{P}_1(x), \mathbf{P}_2(x))$ forms a basis of $\mathbb{R}[x;2]$. These polynomials have been defined for all positive integers n and are very helpful in numerical computations due to the following properties:

   (a) $\deg(\mathbf{P}_i(x)) = i$, for $i \geq 0$;

   (b) $\mathbf{P}_i(1) = 1$, for $i \geq 0$;

   (c) $\int\limits_{-1}^{1} \mathbf{P}_i(x)\mathbf{P}_j(x)dx = 0$, whenever $i \neq j$; and

   (d) $\int\limits_{-1}^{1} (\mathbf{P}_i(x))^2 dx = \dfrac{2}{2i+1}$, for $i \geq 0$.

Here, $a_0 + a_1x + a_2x^2 = \left(a_0 + \frac{a_2}{3}\right)\mathbf{P}_0(x) + a_1\mathbf{P}_1(x) + \frac{2a_2}{3}\mathbf{P}_2(x) = \begin{bmatrix} P_0(x) \\ P_1(x) \\ P_2(x) \end{bmatrix} \begin{bmatrix} a_0 + \frac{a_2}{3} \\ a_1 \\ \frac{2a_2}{3} \end{bmatrix}$.

3. Let $\mathbb{V} = \{(u,v,w,x,y)^T \in \mathbb{R}^5 \mid w - x = u, v = y, u + v + x = 3y\}$. Then, verify that $\mathcal{B} = \left((-1,0,0,1,0)^T, (2,1,2,0,1)^T\right) = (\mathbf{u}_1, \mathbf{u}_2)$, say, can be taken as a basis of $\mathbb{V}$. In this case, $(7,5,10,3,5) = [\mathbf{u}_1, \mathbf{u}_2]\begin{bmatrix} 3 \\ 5 \end{bmatrix} = [\mathbf{u}_2, \mathbf{u}_1]\begin{bmatrix} 5 \\ 3 \end{bmatrix}$.

So, from the above, we conclude the following: Let $\mathbb{V}$ be a vector space of dimension n over $\mathbb{F}$. If we fix a basis, say, $\mathcal{B} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ of $\mathbb{V}$ and if then $\mathbf{v} \in \mathbb{V}$ with $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i \Rightarrow$

$$\mathbf{v} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = [\mathbf{u}_2, \mathbf{u}_1, \ldots, \mathbf{u}_n]\begin{bmatrix} \alpha_2 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Note the change in the first two components of the column vectors which are elements of $\mathbb{F}^n$. So, a change in the position of the vectors $\mathbf{u}_i$'s gives a change in the column vector. Hence, if we fix the order of the basis vectors $\mathbf{u}_i$'s then then with respect to this order all vectors can be thought of as elements of $\mathbb{F}^n$. To clarify, we have the following definition.

Definition 3.5.2. Let $\mathbb{W}$ be a vector space over $\mathbb{F}$ with a basis $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. Then, an **ordered basis** for $\mathbb{W}$ is a basis $\mathcal{B}$ together with a one-to-one correspondence between $\mathcal{B}$ and $\{1, 2, \ldots, m\}$. Since there is an order among the elements of $\mathcal{B}$, we write $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$. The matrix $B = [\mathbf{u}_1, \ldots, \mathbf{u}_m]$ is an element of $\mathbb{W}^m$ and is generally called the **basis matrix**.

Example 3.5.3. Note that in Example 3.5.5 the matrices $[1, x, x^2]$, $[\mathbf{P}_0(x), \mathbf{P}_1(x), \mathbf{P}_2(x)]$ and $[\mathbf{u}_1, \mathbf{u}_2]$ were basis matrices corresponding to different vector spaces.

Definition 3.5.4. Let $B = [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ be the basis matrix corresponding to an ordered basis $\mathcal{B}$ of $\mathbb{W}$. Since $\mathcal{B}$ is a basis of $\mathbb{W}$, for each $\mathbf{v} \in \mathbb{W}$, there exist $\beta_i, 1 \le i \le m$, such that $\mathbf{v} = \sum\limits_{i=1}^{m} \beta_i \mathbf{v}_i = B\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$. The vector $\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$, denoted $[\mathbf{v}]_{\mathcal{B}}$, is called **the coordinate vector of v with respect to $\mathcal{B}$**. Thus,

$$\mathbf{v} = B[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}_1, \ldots, \mathbf{v}_m][\mathbf{v}]_{\mathcal{B}}, \text{ or equivalently, } \mathbf{v} = [\mathbf{v}]_{\mathcal{B}}^T \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{bmatrix}. \tag{3.5.1}$$

The last expression is generally viewed as a symbolic expression.

Example 3.5.5. Consider Example 3.5.5. Then

1. for $f(x) = 1 - x^2 \in \mathbb{R}[x; 2]$ with $\mathcal{B} = (1, x, x^2)$ as an ordered basis $[f(x)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$.

2. for $(7,5,10,3,5) \in \mathbb{V} = \{(u,v,w,x,y)^T \in \mathbb{R}^5 \mid w - x = u, v = y, u+v+x = 3y\}$ with $\mathcal{B} = \left((-1,0,0,1,0)^T, (2,1,2,0,1)^T\right)$ as an ordered basis of $\mathbb{V}$ $[(7,5,10,3,5)]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

Remark 3.5.6. *1. Let $\mathcal{B}$ be an ordered basis of a vector space $\mathbb{V}$ over $\mathbb{F}$ of dimension n.*

(a) Then $[\alpha\mathbf{v} + \mathbf{w}]_{\mathcal{B}} = \alpha[\mathbf{v}]_{\mathcal{B}} + [\mathbf{w}]_{\mathcal{B}}$, for all $\alpha \in \mathbb{F}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{V}$.

(b) Further, let $S = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \subseteq \mathbb{V}$. Then, observe that S is linearly independent if and only if $\{[\mathbf{w}_1]_{\mathcal{B}}, \ldots, [\mathbf{w}_m]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{F}^n$.

2. Suppose $V = \mathbb{F}^n$ in Definition 3.5.4. Then, $B = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is an $n \times n$ invertible matrix (see Exercise 3.3.16.4). Thus, $\mathbf{v} = B[\mathbf{v}]_{\mathcal{B}}$ implies

$$[\mathbf{v}]_{\mathcal{B}} = B^{-1}\mathbf{v} \text{ for every } \mathbf{v} \in \mathbb{V}. \tag{3.5.2}$$

Example 3.5.7. Let $\mathcal{B} = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)$ be an ordered basis of $\mathbb{R}^2$. Then, the matrix $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is invertible and using Equation (3.5.2), $\begin{bmatrix} \pi \\ e \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} \pi \\ e \end{bmatrix}$.

EXERCISE **3.5.8.** *Recall that any real square matrix can be written as a sum of Hermitian and skew-Hermitian matrices. Thus $\mathbb{M}_3(\mathbb{R}) = \mathbb{U} + \mathbb{W}$, where $\mathbb{U} = \{A \in \mathbb{M}_3(\mathbb{R}) | A^T = A\}$ and $\mathbb{W} = \{A \in \mathbb{M}_3(\mathbb{R}) | A^T = -A\}$. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$. Then $A = X + Y$ for some $X \in \mathbb{U}$ and $Y \in \mathbb{W}$.*

1. If $\mathcal{B} = (\mathbf{e}_{11}, \mathbf{e}_{12}, \mathbf{e}_{13}, \ldots, \mathbf{e}_{33})$ is an ordered basis of $\mathbb{M}_3(\mathbb{R})$ then

$$[A]_{\mathcal{B}}^T = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 & 3 & 3 & 1 & 4 \end{bmatrix}.$$

2. If $\mathcal{C} = (\mathbf{e}_{11}, \mathbf{e}_{12} + \mathbf{e}_{21}, \mathbf{e}_{13} + \mathbf{e}_{31}, \mathbf{e}_{22}, \mathbf{e}_{23} + \mathbf{e}_{32}, \mathbf{e}_{33})$ is an ordered basis of $\mathbb{U}$ then $[X]_{\mathcal{C}}^T = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 4 \end{bmatrix}$.

3. If $\mathcal{D} = (\mathbf{e}_{12} - \mathbf{e}_{21}, \mathbf{e}_{13} - \mathbf{e}_{31}, \mathbf{e}_{23} - \mathbf{e}_{32})$ is an ordered basis of $\mathbb{W}$ then $[Y]_{\mathcal{D}}^T = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}$.

Definition 3.5.9. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Let $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $B = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ be basis matrices corresponding to the ordered bases $\mathcal{A}$ and $\mathcal{B}$, respectively, of $\mathbb{V}$. Thus, using Equation (3.5.1), we have

$$A = [\mathbf{v}_1, \ldots, \mathbf{v}_n] = [B[\mathbf{v}_1]_{\mathcal{B}}, \ldots, B[\mathbf{v}_n]_{\mathcal{B}}] = B\left[[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_n]_{\mathcal{B}}\right] = B[\mathcal{A}]_{\mathcal{B}}, \tag{3.5.3}$$

where $[\mathcal{A}]_{\mathcal{B}} = [[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_n]_{\mathcal{B}}]$. The matrix $[\mathcal{A}]_{\mathcal{B}}$ is called the matrix of $\mathcal{A}$ **with respect to the ordered basis** $\mathcal{B}$ or the **change of basis matrix** from $\mathcal{A}$ to $\mathcal{B}$.

We now summarize the above discussion which helps us to understand the name 'change of basis matrix' for the matrix $[\mathcal{A}]_{\mathcal{B}}$.

Theorem 3.5.10. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Further, let $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ be two ordered bases of $\mathbb{V}$*

1. *Then the matrix* $[\mathcal{A}]_{\mathcal{B}}$ *is invertible.*
2. *Similarly, the matrix* $[\mathcal{B}]_{\mathcal{A}}$ *is invertible.*
3. *Moreover,* $[\mathbf{x}]_{\mathcal{B}} = [\mathcal{A}]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{A}}$*, for all* $\mathbf{x} \in \mathbb{V}$*, i.e.,* $[\mathcal{A}]_{\mathcal{B}}$ *takes coordinate vector of* $\mathbf{x}$ *with respect to* $\mathcal{A}$ *to the coordinate vector of* $\mathbf{x}$ *with respect to* $\mathcal{B}$*.*
4. *Similarly,* $[\mathbf{x}]_{\mathcal{A}} = [\mathcal{B}]_{\mathcal{A}}[\mathbf{x}]_{\mathcal{B}}$*, for all* $\mathbf{x} \in \mathbb{V}$*.*
5. *Furthermore,* $([\mathcal{A}]_{\mathcal{B}})^{-1} = [\mathcal{B}]_{\mathcal{A}}$*.*

Proof. Part 1: Note that using Equation (3.5.3), we have $[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [\mathbf{u}_1, \ldots, \mathbf{u}_n][\mathcal{A}]_{\mathcal{B}}$. Hence, by Exercise 3.2.13.7, the matrix $[\mathcal{A}]_{\mathcal{B}}$ is invertible, which proves Part 1. A similar argument gives Part 2.

Part 3: Note that using Equation (3.5.1), $B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x} = A[\mathbf{x}]_{\mathcal{A}}$ for all $\mathbf{x} \in \mathbb{V}$. Therefore, using Equation (3.5.3), we get $B[\mathbf{x}]_{\mathcal{B}} = (B[\mathcal{A}]_{\mathcal{B}})\,[\mathbf{x}]_{\mathcal{A}}$. As B is invertible, $[\mathbf{x}]_{\mathcal{B}} = [\mathcal{A}]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{A}}$. This completes the proof of Part 3. We leave the proof of other parts to the reader. ■

Example 3.5.11.
1. Let $\mathbb{V} = \mathbb{C}^n$, $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{B} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be the standard ordered basis. Then $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n] = [[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_n]_{\mathcal{B}}] = [A]_{\mathcal{B}}$.
2. Suppose $\mathcal{A} = \left((1,0,0)^T, (1,1,0)^T, (1,1,1)^T\right)$ and $\mathcal{B} = \left((1,1,1)^T, (1,-1,1)^T, (1,1,0)^T\right)$ are two ordered bases of $\mathbb{R}^3$. Then, we verify the statements in the previous result.

   (a) Using Equation (3.5.2), $\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathcal{A}} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}\right)^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x-y \\ y-z \\ z \end{bmatrix}$.

   (b) Similarly, $\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 & 1 & 2 \\ 1 & -1 & 0 \\ 2 & 0 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -x+y+2z \\ x-y \\ 2x-2z \end{bmatrix}$.

   (c) $[\mathcal{A}]_{\mathcal{B}} = \begin{bmatrix} -1/2 & 0 & 1 \\ 1/2 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$, $[\mathcal{B}]_{\mathcal{A}} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ and $[\mathcal{A}]_{\mathcal{B}}[\mathcal{B}]_{\mathcal{A}} = I_3$.

Remark 3.5.12. *Let* $\mathbb{V}$ *be a vector space over* $\mathbb{F}$ *with* $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *as an ordered basis. Then, by Theorem 3.5.10,* $[\mathbf{v}]_{\mathcal{A}}$ *is an element of* $\mathbb{F}^n$*, for each* $\mathbf{v} \in \mathbb{V}$*. Therefore,*

1. *if* $\mathbb{F} = \mathbb{R}$ *then, the elements of* $\mathbb{V}$ *correspond to vectors in* $\mathbb{R}^n$*.*
2. *if* $\mathbb{F} = \mathbb{C}$ *then, the elements of* $\mathbb{V}$ *correspond to vectors in* $\mathbb{C}^n$*.*

EXERCISE **3.5.13.** *Let* $\mathcal{A} = \left((1,2,0)^T, (1,3,2)^T, (0,1,3)^T\right)$ *and* $\mathcal{B} = \left((1,2,1)^T, (0,1,2)^T, (1,4,6)^T\right)$ *be two ordered bases of* $\mathbb{R}^3$*. Then, determine* $[\mathcal{A}]_{\mathcal{B}}$*,* $[\mathcal{B}]_{\mathcal{A}}$ *and verify that* $[\mathcal{A}]_{\mathcal{B}}[\mathcal{B}]_{\mathcal{A}} = I_3$*.*

3.6 Summary

In this chapter, we defined vector spaces over $\mathbb{F}$. The set $\mathbb{F}$ was either $\mathbb{R}$ or $\mathbb{C}$. To define a vector space, we start with a non-empty set $\mathbb{V}$ of vectors and $\mathbb{F}$ the set of scalars. We also needed to do the following:

1. first define vector addition and scalar multiplication and

2. then verify the conditions in Definition 3.1.1.

If all conditions in Definition 3.1.1 are satisfied then $\mathbb{V}$ is a vector space over $\mathbb{F}$. If $\mathbb{W}$ was a non-empty subset of a vector space $\mathbb{V}$ over $\mathbb{F}$ then for $\mathbb{W}$ to be a space, we only need to check whether the vector addition and scalar multiplication inherited from that in $\mathbb{V}$ hold in $\mathbb{W}$.

We then learnt 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 $\mathbb{V}$ is the smallest subspace of $\mathbb{V}$ containing S. Also, to check whether a given vector $\mathbf{v}$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_n$, we needed to solve the linear system $c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \mathbf{v}$ in the variables $c_1, \ldots, c_n$. Or equivalently, the system $A\mathbf{x} = \mathbf{b}$, where in some sense $A[:, i] = \mathbf{u}_i$, $1 \le i \le n$, $\mathbf{x}^T = [c_1, \ldots, c_n]$ and $\mathbf{b} = \mathbf{v}$. It was also shown that the geometrical representation of the linear span of $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is equivalent to finding conditions in the entries of $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ was always consistent.

Then, we learnt linear independence and dependence. A set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is linearly independent set in the vector space $\mathbb{V}$ over $\mathbb{F}$ if the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution in $\mathbb{F}$. Else S is linearly dependent, whereas before the columns of A correspond to the vectors $\mathbf{u}_i$'s.

We then talked about the maximal linearly independent set (coming from the homogeneous system) and the minimal spanning set (coming from the non-homogeneous system) and culminating in the notion of the basis of a finite dimensional vector space $\mathbb{V}$ over $\mathbb{F}$. The following important results were proved.

1. A linearly independent set can be extended to form a basis of $\mathbb{V}$.

2. Any two bases of $\mathbb{V}$ have the same number of elements.

This number was defined as the dimension of $\mathbb{V}$, denoted $\dim(\mathbb{V})$.

Now let $A \in \mathbb{M}_n(\mathbb{R})$. Then, combining a few results from the previous chapter, we have the following equivalent conditions.

1. A is invertible.

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

3. $\text{RREF}(A) = I_n$.

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. $\text{Col}(A^T) = \text{Row}(A) = \mathbb{R}^n$.

10. Rows of A form a basis of $\mathbb{R}^n$.

11. $\textsc{Col}(A) = \mathbb{R}^n$.

12. Columns of A form a basis of $\mathbb{R}^n$.

13. $\textsc{Null}(A) = \{\mathbf{0}\}$.

Chapter 4

Linear Transformations

4.1 Definitions and Basic Properties

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Also, let $\mathcal{B}$ be an ordered basis of $\mathbb{V}$. Then, in the last section of the previous chapter, it was shown that for each $\mathbf{x} \in \mathbb{V}$, the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ is a column vector of size n and has entries from $\mathbb{F}$. So, in some sense, each element of $\mathbb{V}$ looks like elements of $\mathbb{F}^n$. In this chapter, we concretize this idea. We also show that matrices give rise to functions between two finite dimensional vector spaces. 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. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. A function (map) $f : \mathbb{V} \to \mathbb{W}$ is called a **linear transformation** if for all $\alpha \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ the function f satisfies

$$f(\alpha \cdot \mathbf{u}) = \alpha \odot f(\mathbf{u}) \quad \text{and} \quad f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) \oplus f(\mathbf{v}),$$

where $+, \cdot$ are binary operations in $\mathbb{V}$ and $\oplus, \odot$ are the binary operations in $\mathbb{W}$. By $\mathcal{L}(\mathbb{V}, \mathbb{W})$, we denote the set of all linear transformations from $\mathbb{V}$ to $\mathbb{W}$. In particular, if $\mathbb{W} = \mathbb{V}$ then the linear transformation f is called a **linear operator** and the corresponding set of linear operators is denoted by $\mathcal{L}(\mathbb{V})$.

Definition 4.1.2. Let $g, h \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then g and h are said to be **equal** if $g(\mathbf{x}) = h(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{V}$.

We now give examples of linear transformations.

Example 4.1.3.

1. Let $\mathbb{V}$ be a vector space. Then, the maps $\mathrm{Id}, \mathbf{0} \in \mathcal{L}(\mathbb{V})$, where
   (a) $\mathrm{Id}(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathbb{V}$, is commonly called the **identity operator**.
   (b) $\mathbf{0}(\mathbf{v}) = \mathbf{0}$, for all $\mathbf{v} \in \mathbb{V}$, is commonly called the **zero operator**.
2. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. Then, $\mathbf{0} \in \mathcal{L}(\mathbb{V}, \mathbb{W})$, where $\mathbf{0}(\mathbf{v}) = \mathbf{0}$, for all $\mathbf{v} \in \mathbb{V}$, is commonly called the **zero transformation**.
3. The map $f(\mathbf{x}) = \mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}$, is an element of $\mathcal{L}(\mathbb{R})$ as $f(a\mathbf{x}) = a\mathbf{x} = af(\mathbf{x})$ and $f(\mathbf{x} + \mathbf{y}) = \mathbf{x} + \mathbf{y} = f(\mathbf{x}) + f(\mathbf{y})$.

4. The map $f(x) = (x, 3x)^T$, for all $x \in \mathbb{R}$, is an element of $\mathcal{L}(\mathbb{R}, \mathbb{R}^2)$ as $f(\lambda x) = (\lambda x, 3\lambda x)^T = \lambda(x, 3x)^T = \lambda f(x)$ and $f(x + y) = (x + y, 3(x + y))^T = (x, 3x)^T + (y, 3y)^T = f(x) + f(y)$.

5. Let $\mathbb{V}, \mathbb{W}$ and $\mathbb{Z}$ be vector spaces over $\mathbb{F}$. Then, for any $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $S \in \mathcal{L}(\mathbb{W}, \mathbb{Z})$, the map $S \circ T \in \mathcal{L}(\mathbb{V}, \mathbb{Z})$, defined by $(S \circ T)(\mathbf{v}) = S\big(T(\mathbf{v})\big)$ for all $\mathbf{v} \in \mathbb{V}$, is called the **composition** of maps. Observe that for each $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ and $\alpha, \beta \in \mathbb{R}$,

$$\begin{aligned}(S \circ T)(\alpha\mathbf{v} + \beta\mathbf{u}) &= S\big(T(\alpha\mathbf{v} + \beta\mathbf{u})\big) = S\big(\alpha T(\mathbf{v}) + \beta f(\mathbf{u})\big) \\ &= \alpha S\big(T(\mathbf{v})\big) + \beta S\big(T(\mathbf{u})\big) = \alpha(S \circ T)(\mathbf{v}) + \beta(S \circ T)(\mathbf{u}).\end{aligned}$$

Hence $S \circ T$, in short ST, is an element of $\mathcal{L}(\mathbb{V}, \mathbb{Z})$.

6. Fix $\mathbf{a} \in \mathbb{R}^n$ and define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^n$. Then $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. In particular, if $\mathbf{x} = [x_1, \ldots, x_n]^T$ then, for all $\mathbf{x} \in \mathbb{R}^n$,

   (a) $f(\mathbf{x}) = \sum\limits_{i=1}^{n} \mathbf{x}_i = \mathbf{1}^T\mathbf{x}$ is a linear transformation.

   (b) $f_1(\mathbf{x}) = x_1 = \mathbf{e}_1^T\mathbf{x}$ is a linear transformation.

   (c) $f_i(\mathbf{x}) = x_i = \mathbf{e}_i^T\mathbf{x}$ is a linear transformation, for $1 \le i \le n$.

7. Define $f : \mathbb{R}^2 \to \mathbb{R}^3$ by $f\left((x, y)^T\right) = (x + y, 2x - y, x + 3y)^T$. Then $f \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$. Here $f(\mathbf{e}_1) = (1, 2, 1)^T$ and $f(\mathbf{e}_2) = (1, -1, 3)^T$.

8. Let $A \in M_{m\times n}(\mathbb{C})$. Define $f_A(\mathbf{x}) = A\mathbf{x}$, for every $\mathbf{x} \in \mathbb{C}^n$. Then, $f_A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$. Thus, for each $A \in \mathbb{M}_{m,n}(\mathbb{C})$, there exists a linear transformation $f_A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$.

9. Define $f : \mathbb{R}^{n+1} \to \mathbb{R}[x; n]$ by $f\left((a_1, \ldots, a_{n+1})^T\right) = a_1 + a_2x + \cdots + a_{n+1}x^n$, for each $(a_1, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. Then f is a linear transformation.

10. Fix $A \in M_n(\mathbb{C})$. Now, define $f_A : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ and $g_A : M_n(\mathbb{C}) \to \mathbb{C}$ by $f_A(B) = AB$ and $g_A(B) = \text{Tr}(AB)$, for every $B \in M_n(\mathbb{C})$. Then, f_A and g_A are both linear transformations.

    Are the maps $f(B) = A^*B$, $g(B) = BA$, $h(B) = \mathsf{tr}(\mathsf{A^*B})$ and $t(B) = \mathsf{tr}(\mathsf{BA})$, for every $B \in M_n(\mathbb{C})$ linear?

11. Is the map $T : \mathbb{R}[x; n] \to \mathbb{R}[x; n + 1]$ defined by $T(f(x)) = xf(x)$, for all $f(x) \in \mathbb{R}[x; n]$ a linear transformation?

12. The maps $T, S : \mathbb{R}[x] \to \mathbb{R}[x]$ defined by $T(f(x)) = \frac{d}{dx}f(x)$ and $S(f(x)) = \int\limits_0^x f(t)dt$, for all $f(x) \in \mathbb{R}[x]$ are linear transformations. Is it true that $TS = \text{Id}$? What about ST?

13. Recall the vector space $\mathbb{R}^{\mathbb{N}}$ in Example 3.1.4.9. Now, define maps $T, S : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ by $T(\{a_1, a_2, \ldots\}) = \{0, a_1, a_2, \ldots\}$ and $S(\{a_1, a_2, \ldots\}) = \{a_2, a_3, \ldots\}$. Then, T and S, commonly called the **shift operators**, are linear operators with exactly one of ST or TS as the Id map.

14. Recall the vector space $\mathcal{C}(\mathbb{R},\mathbb{R})$ (see Example 3.1.4.11). Define $T : \mathcal{C}(\mathbb{R},\mathbb{R}) \to \mathcal{C}(\mathbb{R},\mathbb{R})$ by $T(f(x)) = \int\limits_0^x f(t)dt$. For example, $T(\sin(x)) = \int\limits_0^x \sin(t)dt = 1 - \cos(x)$, for all $x \in \mathbb{R}$. Then, verify that T is a linear transformation.

Remark 4.1.4. *Let $A \in M_n(\mathbb{C})$ and define $T_A : \mathbb{C}^n \to \mathbb{C}^n$ by $T_A(\mathbf{x}) = A\mathbf{x}$, for every $\mathbf{x} \in \mathbb{C}^n$. Then, verify that $T_A^k(\mathbf{x}) = \underbrace{(T_A \circ T_A \circ \cdots \circ T_A)}_{k \text{ times}}(\mathbf{x}) = A^k\mathbf{x}$, for any positive integer k.*

We now prove that any linear transformation sends the zero vector to a zero vector.

Proposition 4.1.5. *Let $T \in \mathcal{L}(\mathbb{V},\mathbb{W})$. Suppose that $\mathbf{0}_{\mathbb{V}}$ is the zero vector in $\mathbb{V}$ and $\mathbf{0}_{\mathbb{W}}$ is the zero vector of $\mathbb{W}$. Then $T(\mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{W}}$.*

Proof. Since $\mathbf{0}_{\mathbb{V}} = \mathbf{0}_{\mathbb{V}} + \mathbf{0}_{\mathbb{V}}$, we get $T(\mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}} + \mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}}) + T(\mathbf{0}_{\mathbb{V}})$. As $T(\mathbf{0}_{\mathbb{V}}) \in \mathbb{W}$,

$$\mathbf{0}_{\mathbb{W}} + T(\mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}}) + T(\mathbf{0}_{\mathbb{V}}).$$

Hence, $T(\mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{W}}$. ■

From now on $\mathbf{0}$ will be used as the zero vector of the domain and codomain. We now consider a few more examples.

Example 4.1.6. 1. Does there exist a linear transformation $T : \mathbb{V} \to \mathbb{W}$ such that $T(\mathbf{v}) \neq \mathbf{0}$, for all $\mathbf{v} \in \mathbb{V}$?
Solution: No, as $T(\mathbf{0}) = \mathbf{0}$ (see Proposition 4.1.5).

2. Does there exist a linear transformation $T : \mathbb{R} \to \mathbb{R}$ such that $T(x) = x^2$, for all $x \in \mathbb{R}$?
Solution: No, as $T(ax) = (ax)^2 = a^2x^2 = a^2T(x) \neq aT(x)$, unless $a = 0, 1$.

3. Does there exist a linear transformation $T : \mathbb{R} \to \mathbb{R}$ such that $T(x) = \sqrt{x}$, for all $x \in \mathbb{R}$?
Solution: No, as $T(ax) = \sqrt{ax} = \sqrt{a}\sqrt{x} \neq a\sqrt{x} = aT(x)$, unless $a = 0, 1$.

4. Does there exist a linear transformation $T : \mathbb{R} \to \mathbb{R}$ such that $T(x) = \sin(x)$, for all $x \in \mathbb{R}$?
Solution: No, as $T(ax) \neq aT(x)$.

5. Does there exist a linear transformation $T : \mathbb{R} \to \mathbb{R}$ such that $T(5) = 10$ and $T(10) = 5$?
Solution: No, as $T(10) = T(5+5) = T(5) + t(5) = 10 + 10 = 20 \neq 5$.

6. Does there exist a linear transformation $T : \mathbb{R} \to \mathbb{R}$ such that $T(5) = \pi$ and $T(e) = \pi$?
Solution: No, as $5T(1) = T(5) = \pi$ implies that $T(1) = \frac{\pi}{5}$. So, $T(e) = eT(1) = \frac{e\pi}{5}$.

7. Does there exist a linear transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f((x,y)^T) = (x+y, 2)^T$?
Solution: No, as $f(\mathbf{0}) \neq \mathbf{0}$.

8. Does there exist a linear transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f((x,y)^T) = (x+y, xy)^T$?
Solution: No, as $f((2,2)^T) = (4,4)^T \neq 2(2,1)^T = 2f((1,1)^T)$.

9. Define a map $T : \mathbb{C} \to \mathbb{C}$ by $T(\mathbf{z}) = \overline{\mathbf{z}}$, the complex conjugate of z. Is T a linear operator over the real vector space $\mathbb{R}$?
Solution: Yes, as for any $\alpha \in \mathbb{R}$, $T(\alpha\mathbf{z}) = \overline{\alpha\mathbf{z}} = \alpha\overline{\mathbf{z}} = \alpha T(\mathbf{z})$.

The next result states that a linear transformation is known if we know its image on a basis of the domain space.

Lemma 4.1.7. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then T is determined if the image of T on basis vectors of $\mathbb{V}$ are known.*

In particular, if $\mathbb{V}$ is finite dimensional and $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is an ordered basis of $\mathbb{V}$ over $\mathbb{F}$ then $T(\mathbf{v}) = \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} [\mathbf{v}]_{\mathcal{B}}$.

Proof. Let $\mathcal{B}$ be a basis of $\mathbb{V}$ over $\mathbb{F}$. Then, for each $\mathbf{v} \in \mathbb{V}$, there exist vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ in $\mathcal{B}$ and scalars $c_1, \ldots, c_k \in \mathbb{F}$ such that $\mathbf{v} = \sum\limits_{i=1}^{k} c_i \mathbf{u}_i$. Thus

$$T(\mathbf{v}) = T\left(\sum_{i=1}^{k} c_i \mathbf{u}_i\right) = \sum_{i=1}^{k} T(c_i \mathbf{u}_i) = \sum_{i=1}^{k} c_i T(\mathbf{u}_i).$$

Or equivalently, whenever

$$\mathbf{v} = [\mathbf{u}_1, \ldots, \mathbf{u}_k] \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} \text{ then } T(\mathbf{v}) = \begin{bmatrix} T(\mathbf{u}_1) & \cdots & T(\mathbf{u}_k) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}. \tag{4.1.1}$$

Thus, the image of T on $\mathbf{v}$ just depends on where the basis vectors are mapped. This completes the first part.

For the second part, let $\mathbf{v} = \sum\limits_{i=1}^{n} c_i \mathbf{v}_i$. Then $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. Hence, using Equation (4.1.1), we have $T(\mathbf{v}) = \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} [\mathbf{v}]_{\mathcal{B}}$. Thus, the required result follows. ∎

As another application of Lemma 4.1.7, we have the following result.

Corollary 4.1.8. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T : \mathbb{V} \to \mathbb{W}$ be a linear transformation. If $\mathcal{B}$ is a basis of $\mathbb{V}$ then, $\text{RNG}(T) = LS(T(\mathbf{x}) | \mathbf{x} \in \mathcal{B})$.*

Recall that by Example 4.1.3.6, for each $\mathbf{a} \in \mathbb{R}^n$, the map $T(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, for each $\mathbf{x} \in \mathbb{R}^n$, is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$. We now show that these are the only ones.

Corollary 4.1.9. [Reisz Representation Theorem] *Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. Then, there exists $\mathbf{a} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$.*

Proof. By Lemma 4.1.7, T is known if we know the image of T on $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, the standard basis of $\mathbb{R}^n$. So, for $1 \le i \le n$, let $T(\mathbf{e}_i) = a_i$, for some $a_i \in \mathbb{R}$. Now define $\mathbf{a} = [a_1, \ldots, a_n]^T$ and $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$. Then

$$T(\mathbf{x}) = T\left(\sum_{i=1}^{n} x_i \mathbf{e}_i\right) = \sum_{i=1}^{n} x_i T(\mathbf{e}_i) = \sum_{i=1}^{n} x_i a_i = \mathbf{a}^T\mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Thus, the required result follows. ∎

Example 4.1.10. In each of the examples given below, state whether a linear transformation exists or not. If yes, give at least one linear transformation. If not, then give the condition due to which a linear transformation doesn't exist.

1. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((1,1)^T) = (1,2)^T$ and $T((1,-1)^T) = (5,10)^T$?

   Solution: Yes, as the set $\{(1,1),(1,-1)\}$ is a basis of $\mathbb{R}^2$. Write $\mathcal{B} = \left[\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right]$. Then, using Equation (4.1.1) and $[\mathbf{v}]_{\mathcal{B}} = B^{-1}\mathbf{v}$, we get

$$\begin{aligned} T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) &= \left[T\left(\begin{bmatrix}1\\1\end{bmatrix}\right),\ T\left(\begin{bmatrix}1\\-1\end{bmatrix}\right)\right]\left[\begin{bmatrix}x\\y\end{bmatrix}\right]_{\mathcal{B}} \\ &= \begin{bmatrix}1 & 5\\2 & 10\end{bmatrix}\left(\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}^{-1}\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}1 & 5\\2 & 10\end{bmatrix}\begin{bmatrix}\frac{x+y}{2}\\\frac{x-y}{2}\end{bmatrix} = \begin{bmatrix}3x-2y\\6x-4y\end{bmatrix}. \end{aligned}$$

2. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((1,1)^T) = (1,2)^T$ and $T((5,5)^T) = (5,10)^T$?

   Solution: Yes, as $(5,10)^T = T((5,5)^T) = 5T((1,1)^T) = 5(1,2)^T = (5,10)^T$.

   To construct one such linear transformation, note that $\mathcal{B} = ((1,1)^T,(1,0)^T)$ is a basis of $\mathbb{R}^2$. Pick $\mathbf{v} \in \mathbb{R}^2$ and define $T((1,0)^T) = \mathbf{v} = (v_1,v_2)^T$. Then, as above, we get

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \left[T\left(\begin{bmatrix}1\\1\end{bmatrix}\right), T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\right]\left[\begin{bmatrix}x\\y\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}1 & v_1\\2 & v_2\end{bmatrix}\begin{bmatrix}y\\x-y\end{bmatrix} = y\begin{bmatrix}1\\2\end{bmatrix} + (x-y)\mathbf{v}.$$

3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((1,1)^T) = (1,2)^T$ and $T((5,5)^T) = (5,11)^T$?

   Solution: No, as $(5,11)^T = T((5,5)^T) = 5T((1,1)^T)5(1,2)^T = (5,10)^T$, a contradiction.

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Rng}(T) = \{T(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^2\} = LS\{(1,\pi)^T\}$?

   Solution: Yes. Define $T(\mathbf{e}_1) = (1,\pi)^T$ and $T(\mathbf{e}_2) = a(1,\pi)^T$, for some $a \in \mathbb{R}$.

5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Rng}(T) = \mathbb{R}^2$?

   Solution: Yes. Let $\{\mathbf{u},\mathbf{v}\}$ be a basis of $\mathbb{R}^2$ and define $T(\mathbf{e}_1) = \mathbf{u}$ and $T(\mathbf{e}_2) = \mathbf{v}$.

6. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Rng}(T) = \{T(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^2\} = \{\mathbf{0}\}$?

   Solution: Yes. Define $T(\mathbf{e}_1) = \mathbf{0}$ and $T(\mathbf{e}_2) = \mathbf{0}$.

7. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Ker}(T) = \{\mathbf{x} \in \mathbb{R}^2 \mid T(\mathbf{x}) = \mathbf{0}\} = LS\{(1,\pi)^T\}$?

   Solution: Yes. Take $\{(1,\pi)^T, \mathbf{u}\}$ as a basis of $\mathbb{R}^2$ and define $T((1,\pi)^T) = \mathbf{0}$ and $T(\mathbf{u}) = \mathbf{u}$.

Exercise **4.1.11.** *1. Use matrices to construct linear operators $T, S : \mathbb{R}^3 \to \mathbb{R}^3$ that satisfy:*

   (a) $T \neq \mathbf{0}$, $T \circ T = T^2 \neq \mathbf{0}$, $T \circ T \circ T = T^3 = \mathbf{0}$.

   (b) $T \neq \mathbf{0}$, $S \neq \mathbf{0}$, $S \circ T = ST \neq \mathbf{0}$, $T \circ S = TS = \mathbf{0}$.

   (c) $S \circ S = S^2 = T^2 = T \circ T$, $S \neq T$.

   (d) $T \circ T = T^2 = Id$, $T \neq Id$.

2. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ 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.*

3. *Fix a positive integer p and let $T : \mathbb{R}^n \to \mathbb{R}^n$ 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.*

4. *Fix $\mathbf{x}_0 \in \mathbb{R}^n$ with $\mathbf{x}_0 \neq \mathbf{0}$. Now, define $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ by $T(\mathbf{x}_0) = \mathbf{y}_0$, for some $\mathbf{y}_0 \in \mathbb{R}^m$. Define $T^{-1}(\mathbf{y}_0) = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{y}_0\}$. Then prove that $\mathbf{x} \in T^{-1}(\mathbf{y}_0)$ if and only if $\mathbf{x} - \mathbf{x}_0 \in T^{-1}(\mathbf{0})$. Further, $T^{-1}(\mathbf{y}_0)$ is a subspace of $\mathbb{R}^n$ if and only if $\mathbf{y}_0 = \mathbf{0}$.*

5. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{V}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \subseteq \mathbb{W}$ then prove that there exists a unique $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ with $T(\mathbf{v}_i) = \mathbf{w}_i$, for $i = 1, \ldots, n$.*

6. *Prove that there exists infinitely many linear transformations $T : \mathbb{R}^3 \to \mathbb{R}^2$ such that $T((1,-1,1)^T) = (1,2)^T$ and $T((-1,1,2)^T) = (1,0)^T$?*

7. *Let $\mathbb{V}$ be a vector space and let $\mathbf{a} \in \mathbb{V}$. Then the map $T_{\mathbf{a}} : \mathbb{V} \to \mathbb{V}$ defined by $T_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$, for all $\mathbf{x} \in \mathbb{V}$ is called the* **translation** *map. Prove that $T_{\mathbf{a}} \in \mathcal{L}(\mathbb{V})$ if and only if $\mathbf{a} = \mathbf{0}$.*

8. *Does there exist a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ such that*
   (a) *$T((1,0,1)^T) = (1,2)^T$, $T((0,1,1)^T) = (1,0)^T$ and $T((1,1,1)^T) = (2,3)^T$?*
   (b) *$T((1,0,1)^T) = (1,2)^T$, $T((0,1,1)^T) = (1,0)^T$ and $T((1,1,2)^T) = (2,3)^T$?*

9. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T((x,y,z)^T) = (2x+3y+4z, x+y+z, x+y+3z)^T$. Find the value of k for which there exists a vector $\mathbf{x} \in \mathbb{R}^3$ such that $T(\mathbf{x}) = (9,3,k)^T$.*

10. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T((x,y,z)^T) = (2x-2y+2z, -2x+5y+2z, 8x+y+4z)^T$. Find $\mathbf{x} \in \mathbb{R}^3$ such that $T(\mathbf{x}) = (1,1,-1)^T$.*

11. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T((x,y,z)^T) = (2x+y+3z, 4x-y+3z, 3x-2y+5z)^T$. Determine $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ such that $T(\mathbf{x}) = 6\mathbf{x}$, $T(\mathbf{y}) = 2\mathbf{y}$ and $T(\mathbf{z}) = -2\mathbf{z}$. Is the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ linearly independent?*

12. *Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T((x,y,z)^T) = (2x+3y+4z, -y, -3y+4z)^T$. Determine $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ such that $T(\mathbf{x}) = 2\mathbf{x}$, $T(\mathbf{y}) = 4\mathbf{y}$ and $T(\mathbf{z}) = -\mathbf{z}$. Is the set $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ linearly independent?*

13. *Let $n \in \mathbb{N}$. Does there exist a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^n$ such that $T((1,1,-2)^T) = \mathbf{x}$, $T((-1,2,3)^T) = \mathbf{y}$ and $T((1,10,1)^T) = \mathbf{z}$*
    (a) *with $\mathbf{z} = \mathbf{x} + \mathbf{y}$?*
    (b) *with $\mathbf{z} = c\mathbf{x} + d\mathbf{y}$, for some $c, d \in \mathbb{R}$?*

14. *For each matrix A given below, define $T \in \mathcal{L}(\mathbb{R}^2)$ by $T(\mathbf{x}) = A\mathbf{x}$. What do these linear operators signify geometrically?*
    (a) $A \in \left\{ \frac{1}{2}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & -\sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & \cos\left(\frac{2\pi}{3}\right) \end{bmatrix} \right\}$.
    (b) $A \in \left\{ \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}$.
    (c) $A \in \left\{ \frac{1}{2}\begin{bmatrix} \sqrt{3} & 1 \\ 1 & -\sqrt{3} \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & \sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & -\cos\left(\frac{2\pi}{3}\right) \end{bmatrix} \right\}$.

15. *Find all functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ 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 $(x, y) \in \mathbb{R}^2$.*

16. *Consider the space $\mathbb{C}^3$ over $\mathbb{C}$. If $f \in \mathcal{L}(\mathbb{C}^3)$ with $f(\mathbf{x}) = \mathbf{x}, f(\mathbf{y}) = (1+i)\mathbf{y}$ and $f(\mathbf{z}) = (2+3i)\mathbf{z}$, for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^3 \setminus \{0\}$ then prove that $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ forms a basis of $\mathbb{C}^3$.*

4.2 Rank-Nullity Theorem

The readers are advised to see Exercise 3.2.13.8 and Theorem 3.4.10 for clarity and similarity with the results in this section. To start with, we define two spaces related with a linear transformation.

Definition 4.2.1. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T : \mathbb{V} \to \mathbb{W}$ be a linear transformation. Then the set

1. $\{T(\mathbf{v}) | \mathbf{v} \in \mathbb{V}\}$ is called the **range space** of T, denoted $\text{Rng}(T)$.
2. $\{\mathbf{v} \in \mathbb{V} | T(\mathbf{v}) = \mathbf{0}\}$ is called the **kernel** of T, denoted $\text{Ker}(T)$. In certain books, it is also called the **null space** of T.

Example 4.2.2. Determine $\text{Rng}(T)$ and $\text{Ker}(T)$ of the following linear transformations.

1. $f \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^4)$, where $f((x, y, z)^T) = (x - y + z, y - z, x, 2x - 5y + 5z)^T$.
   Solution: Consider the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of $\mathbb{R}^3$. Then

$$\begin{aligned}
\text{Rng}(f) &= LS(f(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)) = LS\big((1,0,1,2)^T, (-1,1,0,-5)^T, (1,-1,0,5)^T\big) \\
&= LS\big((1,0,1,2)^T, (1,-1,0,5)^T\big) = \{\lambda(1,0,1,2)^T + \beta(1,-1,0,5)^T \mid \lambda, \beta \in \mathbb{R}\} \\
&= \{(\lambda + \beta, -\beta, \lambda, 2\lambda + 5\beta) \ : \lambda, \beta \in \mathbb{R}\} \\
&= \{(x, y, z, w)^T \in \mathbb{R}^4 \mid x + y - z = 0, 5y - 2z + w = 0\},
\end{aligned}$$

$$\begin{aligned}
\text{Ker}(f) &= \{(x, y, z)^T \in \mathbb{R}^3 \ : f((x, y, z)^T) = \mathbf{0}\} \\
&= \{(x, y, z)^T \in \mathbb{R}^3 \ : (x - y + z, y - z, x, 2x - 5y + 5z)^T = \mathbf{0}\} \\
&= \{(x, y, z)^T \in \mathbb{R}^3 \ : x - y + z = 0, y - z = 0, x = 0, 2x - 5y + 5z = 0\} \\
&= \{(x, y, z)^T \in \mathbb{R}^3 \ : y - z = 0, x = 0\} \\
&= \{(0, z, z)^T \in \mathbb{R}^3 \ : z \in \mathbb{R}\} = LS((0, 1, 1)^T)
\end{aligned}$$

2. Let $B \in \mathbb{M}_2(\mathbb{R})$. Now, define a map $T : \mathbb{M}_2(\mathbb{R}) \to \mathbb{M}_2(\mathbb{R})$ by $T(A) = BA - AB$, for all $A \in \mathbb{M}_2(\mathbb{R})$. Determine $\text{Rng}(T)$ and $\text{Ker}(T)$.
   Solution: Note that $A \in \text{Ker}(T)$ if and only if A commutes with B. In particular, $\{I, B, B^2, \ldots\} \subseteq \text{Ker}(T)$. For example, if B is a scalar matrix then, $\text{Ker}(T) = \mathbb{M}_2(\mathbb{R})$.

   For computing, $\text{Rng}(T)$, recall that $\{\mathbf{e}_{ij} | 1 \leq i, j \leq 2\}$ is a basis of $\mathbb{M}_2(\mathbb{R})$. So,

(a) if $B = cI_2$ then $\text{Rng}(T) = \{\mathbf{0}\}$.

(b) if $B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ then $T(\mathbf{e}_{11}) = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$, $T(\mathbf{e}_{12}) = \begin{bmatrix} -2 & -3 \\ 0 & 2 \end{bmatrix}$, $T(\mathbf{e}_{21}) = \begin{bmatrix} 2 & 0 \\ 3 & -2 \end{bmatrix}$ and $T(\mathbf{e}_{22}) = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$. Thus, $\text{Rng}(T) = LS\left(\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ -3 & 2 \end{bmatrix}\right)$.

(c) for $B = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$, verify that $\text{Rng}(T) = LS\left(\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ -2 & 2 \end{bmatrix}\right)$.

Exercise **4.2.3.** *1. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then*

(a) $\text{Rng}(T)$ *is a subspace of* $\mathbb{W}$.

(b) $\text{Ker}(T)$ *is a subspace of* $\mathbb{V}$.

Furthermore, if $\mathbb{V}$ is finite dimensional then

(a) $\dim(\text{Ker}(T)) \le \dim(\mathbb{V})$.

(b) $\dim(\text{Rng}(T))$ *is finite and whenever* $\dim(\mathbb{W})$ *is finite* $\dim(\text{Rng}(T)) \le \dim(\mathbb{W})$.

2. *Which of the following maps are linear transformations? In case, the map is a linear transformation, determine its range space and the null space.*

   (a) Let $\mathbb{V} = \mathbb{R}^2$ and $\mathbb{W} = \mathbb{R}^3$ with $T((x,y)^T) = (x+y+1, 2x-y, x+3y)^T$.

   (b) Let $\mathbb{V} = \mathbb{W} = \mathbb{R}^2$ with $T((x,y)^T) = (x-y, x^2-y^2)^T$.

   (c) Let $\mathbb{V} = \mathbb{W} = \mathbb{R}^2$ with $T((x,y)^T) = (x-y, \mid x \mid)^T$.

   (d) Let $\mathbb{V} = \mathbb{R}^2$ and $\mathbb{W} = \mathbb{R}^4$ with $T((x,y)^T) = (x+y, x-y, 2x+y, 3x-4y)^T$.

   (e) Let $\mathbb{V} = \mathbb{W} = \mathbb{R}^4$ with $T((x,y,z,w)^T) = (z,x,w,y)^T$.

3. *Which of the following maps $T : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ are linear operators? In case, T is a linear operator, determine* $\text{Rng}(T)$ *and* $\text{Ker}(T)$.

   (a) $T(A) = A^T$.

   (b) $T(A) = I + A$.

   (c) $T(A) = A^2$.

   (d) $T(A) = BAB^{-1}$, for some fixed $B \in \mathbb{M}_2(\mathbb{R})$.

4. *Describe* $\text{Ker}(D)$ *and* $\text{Rng}(D)$, *where $D \in \mathcal{L}(\mathbb{R}[x;n])$ and is defined by $D(f(x)) = f'(x)$, the differentiation with respect to x. Note that* $\text{Rng}(D) \subseteq \mathbb{R}[x;n-1]$.

5. *Define $T \in \mathcal{L}(\mathbb{R}[x])$ by $T(f(x)) = xf(x)$, for all $f(x) \in \mathcal{L}(\mathbb{R}[x])$. What can you say about* $\text{Ker}(T)$ *and* $\text{Rng}(T)$*?*

6. *For T in Example 4.2.2, compute* $\dim(\text{Ker}(T))$ *and* $\dim(\text{Rng}(T))$.

7. *Define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T(\mathbf{e}_1) = \mathbf{e}_1 + \mathbf{e}_3$, $T(\mathbf{e}_2) = \mathbf{e}_2 + \mathbf{e}_3$ and $T(\mathbf{e}_3) = -\mathbf{e}_3$. Then*

   (a) determine $T((x,y,z)^T)$, for $x,y,z \in \mathbb{R}$.

   (b) determine $\text{Null}(T)$ *and* $\text{Rng}(T)$.

(c) *is it true that* $T^2 = Id$?

8. *Find* $T \in \mathcal{L}(\mathbb{R}^3)$ *for which* $\text{RNG}(T) = LS\big((1,2,0)^T, (0,1,1)^T, (1,3,1)^T\big)$.

Theorem 4.2.4. *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be vector spaces over* $\mathbb{F}$ *and let* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$.

1. *If* $S \subseteq \mathbb{V}$ *is linearly dependent then* $T(S) = \{T(\mathbf{v}) \mid \mathbf{v} \in \mathbb{V}\}$ *is linearly dependent.*
2. *Suppose* $S \subseteq \mathbb{V}$ *such that* $T(S)$ *is linearly independent then* S *is linearly independent.*

Proof. As S is linearly dependent, there exist $k \in \mathbb{N}$ and $\mathbf{v}_i \in S$, for $1 \le i \le k$, such that the system $\sum\limits_{i=1}^{k} x_i\mathbf{v}_i = \mathbf{0}$, in the unknowns x_i's, has a non-trivial solution, say $x_i = a_i \in \mathbb{F}, 1 \le i \le k$. Thus $\sum\limits_{i=1}^{k} a_i\mathbf{v}_i = \mathbf{0}$. Then a_i's also give a non-trivial solution to the system $\sum\limits_{i=1}^{k} y_iT(\mathbf{v}_i) = \mathbf{0}$, where y_i's are unknown, as $\sum\limits_{i=1}^{k} a_iT(\mathbf{v}_i) = \sum\limits_{i=1}^{k} T(a_i\mathbf{v}_i) = T\left(\sum\limits_{i=1}^{k} a_i\mathbf{v}_i\right) = T(\mathbf{0}) = \mathbf{0}$. Hence the required result follows.

The second part is left as an exercise for the reader. ∎

Definition 4.2.5. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $\dim(\mathbb{V})$ is finite then we define $\text{RANK}(T) = \dim(\text{RNG}(T))$ and $\text{NULLITY}(T) = \dim(\text{KER}(T))$.

We now prove the rank-nullity Theorem. The proof of this result is similar to the proof of Theorem 3.4.10. We give it again for the sake of completeness.

Theorem 4.2.6 (Rank-Nullity Theorem)**.** *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be vector spaces over* $\mathbb{F}$. *If* $\dim(\mathbb{V})$ *is finite and* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ *then*

$$\text{RANK}(T) + \text{NULLITY}(T) = \dim(\text{RNG}(T)) + \dim(\text{KER}(T)) = \dim(\mathbb{V}).$$

Proof. By Exercise 4.2.3.1.1a, $\dim(\text{KER}(T)) \le \dim(\mathbb{V})$. Let $\mathcal{B}$ be a basis of $\text{KER}(T)$. We extend it to form a basis $\mathcal{C}$ of $\mathbb{V}$. As $T(\mathbf{v}) = \mathbf{0}$, for all $\mathbf{v} \in \mathcal{B}$, using Corollary 4.1.8, we get

$$\text{RNG}(T) = LS(\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C}\}) = LS(\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C} \setminus \mathcal{B}\}).$$

We claim that $\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C} \setminus \mathcal{B}\}$ is linearly independent subset of $\mathbb{W}$.

On the contrary, assume that there exists $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathcal{C} \setminus \mathcal{B}$ and $\mathbf{a} = [a_1, \ldots, a_k]^T$ such that $\mathbf{a} \neq \mathbf{0}$ and $\sum\limits_{i=1}^{k} a_iT(\mathbf{v}_i) = \mathbf{0}$. Thus $T\big(\sum\limits_{i=1}^{k} a_i\mathbf{v}_i\big) = \sum\limits_{i=1}^{k} a_iT(\mathbf{v}_i) = \mathbf{0}$, *i.e.*, $\sum\limits_{i=1}^{k} a_i\mathbf{v}_i \in \text{KER}(T)$. Hence, there exists $b_1, \ldots, b_\ell \in \mathbb{F}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_\ell \in \mathcal{B}$ such that $\sum\limits_{i=1}^{k} a_i\mathbf{v}_i = \sum\limits_{j=1}^{\ell} b_j\mathbf{u}_j$. Or equivalently, the system $\sum\limits_{i=1}^{k} x_i\mathbf{v}_i + \sum\limits_{j=1}^{\ell} y_j\mathbf{u}_j = \mathbf{0}$, in the unknowns x_i's and y_j's, has a non-trivial solution $[a_1, \ldots, a_k, -b_1, \ldots, -b_\ell]^T$ (non-trivial as $\mathbf{a} \neq \mathbf{0}$). Hence, $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is linearly dependent subset in $\mathbb{V}$. A contradiction to $S \subseteq \mathcal{C}$. Thus,

$$\dim(\text{RNG}(T)) + \dim(\text{KER}(T)) = |\mathcal{C} \setminus \mathcal{B}| + |\mathcal{B}| = |\mathcal{C}| = \dim(\mathbb{V}).$$

Thus, we have proved the required result. ∎

As an immediate corollary, we have the following result. The proof is left for the reader.

Corollary 4.2.7. *Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If $\dim(\mathbb{V}) = \dim(\mathbb{W})$ then the following statements are equivalent.*

1. *T is one-one.*
2. $\textsc{Ker}(T) = \{\mathbf{0}\}$.
3. *T is onto.*
4. $\dim(\textsc{Rng}(T)) = \dim(\mathbb{V})$.

Corollary 4.2.8. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. If $S, T \in \mathcal{L}(\mathbb{V})$ then*

1. $\textsc{Nullity}(T) + \textsc{Nullity}(S) \geq \textsc{Nullity}(ST) \geq \max\{\textsc{Nullity}(T), \textsc{Nullity}(S)\}$.
2. $\min\{\mathsf{Rank}(S), \mathsf{Rank}(T)\} \geq \mathsf{Rank}(ST) \geq n - \mathsf{Rank}(S) - \mathsf{Rank}(T)$.

Proof. The prove of Part 2 is omitted as it directly follows from Part 1 and Theorem 4.2.6. Part 1: We first prove the second inequality. Suppose $\mathbf{v} \in \textsc{Ker}(T)$. Then

$$(ST)(\mathbf{v}) = S(T(\mathbf{v})) = S(\mathbf{0}) = \mathbf{0}$$

implies $\textsc{Ker}(T) \subseteq \textsc{Ker}(ST)$. Thus $\textsc{Nullity}(T) \leq \textsc{Nullity}(ST)$.

By Theorem 4.2.6, $\textsc{Nullity}(S) \leq \textsc{Nullity}(ST) \Leftrightarrow \mathsf{Rank}(S) \geq \mathsf{Rank}(ST)$. This holds as $\textsc{Rng}(T) \subseteq \mathbb{V}$ implies $\textsc{Rng}(ST) = S(\textsc{Rng}(T)) \subseteq S(\mathbb{V}) = \textsc{Rng}(S)$.

To prove the first inequality, let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\textsc{Ker}(T)$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \textsc{Ker}(ST)$. So, let us extend it to get a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ of $\textsc{Ker}(ST)$.

Now, proceeding as in the proof of the rank-nullity theorem, implies that $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_\ell)\}$ is a linearly independent subset of $\textsc{Ker}(S)$. Hence, $\textsc{Nullity}(S) \geq \ell$ and therefore, we get $\textsc{Nullity}(ST) = k + \ell \leq \textsc{Nullity}(T) + \textsc{Nullity}(S)$. ■

Exercise **4.2.9.**

1. *Let $A \in M_n(\mathbb{R})$ with $A^2 = A$. Define $T \in \mathcal{L}(\mathbb{R}^n)$ by $T(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. Then prove that*
   (a) *$T^2 = T$, or equivalently, $(T(Id - T))(\mathbf{x}) = \mathbf{0}$, for all $\mathbf{x} \in \mathbb{R}^n$.*
   (b) $\textsc{Null}(T) \cap \textsc{Rng}(T) = \{\mathbf{0}\}$.
   (c) $\mathbb{R}^n = \textsc{Rng}(T) + \textsc{Null}(T)$.
2. *Let $z_1, z_2, \ldots, z_k$ be k distinct complex numbers. Define $T \in \mathcal{L}(\mathbb{C}[x; n], \mathbb{C}^k)$ by $T\big(P(z)\big) = \big(P(z_1), \ldots, P(z_k)\big)^T$, for all $P(z) \in \mathbb{C}[x; n]$. Determine $\textsc{Rank}(T)$.*

4.2.1 Algebra of Linear Transformations

We start with the following definition.

Definition 4.2.10. Let $\mathbb{V}, \mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $S, T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, we define the point-wise

1. **sum** of S and T, denoted $S + T$, by $(S + T)(\mathbf{v}) = S(\mathbf{v}) + T(\mathbf{v})$, for all $\mathbf{v} \in \mathbb{V}$.
2. **scalar multiplication**, denoted cT for $c \in \mathbb{F}$, by $(cT)(\mathbf{v}) = c\,(T(\mathbf{v}))$, for all $\mathbf{v} \in \mathbb{V}$.

Theorem 4.2.11. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. Then $\mathcal{L}(\mathbb{V},\mathbb{W})$ is a vector space over $\mathbb{F}$. Furthermore, if $\dim \mathbb{V} = n$ and $\dim \mathbb{W} = m$, then $\dim \mathcal{L}(\mathbb{V},\mathbb{W}) = mn$.*

Proof. It can be easily verified that under point-wise addition and scalar multiplication, defined above, $\mathcal{L}(\mathbb{V},\mathbb{W})$ is indeed a vector space over $\mathbb{F}$. We now prove the other part. So, let us assume that $\mathcal{B} = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ and $\mathcal{C} = \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ are bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. For $1 \le i \le n, 1 \le j \le m$, we define the functions $\mathbf{f}_{ij}$ on the basis vectors of $\mathbb{V}$ by

$$\mathbf{f}_{ij}(\mathbf{v}_k) = \begin{cases} \mathbf{w}_j, & \text{if } k = i \\ \mathbf{0}, & \text{if } k \ne i. \end{cases}$$

For other vectors of $\mathbb{V}$, we extend the definition by linearity, *i.e.*, if $\mathbf{v} = \sum\limits_{s=1}^{n} \alpha_s \mathbf{v}_s$ then

$$\mathbf{f}_{ij}(\mathbf{v}) = \mathbf{f}_{ij}\left(\sum_{s=1}^{n} \alpha_s \mathbf{v}_s\right) = \sum_{s=1}^{n} \alpha_s \mathbf{f}_{ij}(\mathbf{v}_s) = \alpha_i \mathbf{f}_{ij}(\mathbf{v}_i) = \alpha_i \mathbf{w}_j. \tag{4.2.1}$$

Thus $\mathbf{f}_{ij} \in \mathcal{L}(\mathbb{V},\mathbb{W})$. We now show that $\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}$ is a basis of $\mathcal{L}(\mathbb{V},\mathbb{W})$.

As a first step, we show that $\mathbf{f}_{ij}$'s are linearly independent. So, consider the linear system $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} c_{ij}\mathbf{f}_{ij} = \mathbf{0}$, in the unknowns c_{ij}'s, for $1 \le i \le n, 1 \le j \le m$. Using the point-wise addition and scalar multiplication, we get

$$\mathbf{0} = \mathbf{0}(\mathbf{v}_k) = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\mathbf{f}_{ij}\right)(\mathbf{v}_k) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\mathbf{f}_{ij}(\mathbf{v}_k) = \sum_{j=1}^{m} c_{kj}\mathbf{w}_j.$$

But, the set $\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ is linearly independent and hence the only solution equals $c_{kj} = 0$, for $1 \le j \le m$. Now, as we vary $\mathbf{v}_k$ from $\mathbf{v}_1$ to $\mathbf{v}_n$, we see that $c_{ij} = 0$, for $1 \le j \le m$ and $1 \le i \le n$. Thus, we have proved the linear independence.

Now, let us prove that $LS\left(\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}\right) = \mathcal{L}(\mathbb{V},\mathbb{W})$. So, let $f \in \mathcal{L}(\mathbb{V},\mathbb{W})$. Then, for $1 \le s \le n$, $f(\mathbf{v}_s) \in \mathbb{W}$ and hence there exists β_{st}'s such that $f(\mathbf{v}_s) = \sum\limits_{t=1}^{m} \beta_{st}\mathbf{w}_t$. So, if $\mathbf{v} = \sum\limits_{s=1}^{n} \alpha_s \mathbf{v}_s \in \mathbb{V}$ then, using Equation (4.2.1), we get

$$\begin{aligned} f(\mathbf{v}) &= f\left(\sum_{s=1}^{n} \alpha_s \mathbf{v}_s\right) = \sum_{s=1}^{n} \alpha_s f(\mathbf{v}_s) = \sum_{s=1}^{n} \alpha_s \left(\sum_{t=1}^{m} \beta_{st}\mathbf{w}_t\right) = \sum_{s=1}^{n}\sum_{t=1}^{m} \beta_{st}(\alpha_s \mathbf{w}_t) \\ &= \sum_{s=1}^{n}\sum_{t=1}^{m} \beta_{st}\mathbf{f}_{st}(\mathbf{v}) = \left(\sum_{s=1}^{n}\sum_{t=1}^{m} \beta_{st}\mathbf{f}_{st}\right)(\mathbf{v}). \end{aligned}$$

Since the above is true for every $\mathbf{v} \in \mathbb{V}$, $LS\left(\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}\right) = \mathcal{L}(\mathbb{V},\mathbb{W})$ and thus the required result follows. ∎

Before proceeding further, recall the following definition about a function.

Definition 4.2.12. Let $f : S \to T$ be any function.

1. Then, a function $g : T \to S$ is called a **left inverse** of f if $(g \circ f)(x) = x$, for all $x \in S$. That is, $g \circ f = \mathrm{Id}$, the identity function on S.

2. Then, a function $h : T \to S$ is called a **right inverse** of f if $(f \circ h)(y) = y$, for all $y \in T$. That is, $f \circ h = \mathrm{Id}$, the identity function on T.

3. Then f is said to be **invertible** if it has a right inverse and a left inverse.

Remark 4.2.13. *Let $f : S \to T$ be invertible. Then, it can be easily shown that any right inverse and any left inverse are the same. Thus, the inverse function is unique and is denoted by f^{-1}. It is well known that f is invertible if and only if f is both one-one and onto.*

Lemma 4.2.14. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If T is one-one and onto then, the map $T^{-1} : \mathbb{W} \to \mathbb{V}$ is also a linear transformation. The map T^{-1} is called the* **inverse linear transform** *of T and is defined by $T^{-1}(\mathbf{w}) = \mathbf{v}$, whenever $T(\mathbf{v}) = \mathbf{w}$.*

Proof. PART 1: As T is one-one and onto, by Theorem 4.2.6, $\dim(\mathbb{V}) = \dim(\mathbb{W})$. So, by Corollary 4.2.7, for each $\mathbf{w} \in \mathbb{W}$ there exists a unique $\mathbf{v} \in \mathbb{V}$ such that $T(\mathbf{v}) = \mathbf{w}$. Thus, one defines $T^{-1}(\mathbf{w}) = \mathbf{v}$.

We need to show that $T^{-1}(\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2) = \alpha_1 T^{-1}(\mathbf{w}_1) + \alpha_2 T^{-1}(\mathbf{w}_2)$, for all $\alpha_1, \alpha_2 \in \mathbb{F}$ and $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{W}$. Note that by previous paragraph, there exist unique vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$ 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, $T(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$, for all $\alpha_1, \alpha_2 \in \mathbb{F}$. Hence, for all $\alpha_1, \alpha_2 \in \mathbb{F}$, we get

$$T^{-1}(\alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2) = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \alpha_1 T^{-1}(\mathbf{w}_1) + \alpha_2 T^{-1}(\mathbf{w}_2).$$

Thus, the required result follows. ■

Example 4.2.15.

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $(x, y) \rightsquigarrow (x + y, x - y)$. Then, verify that T^{-1} is given by $\rightsquigarrow \left(\frac{x+y}{2}, \frac{x-y}{2}\right)$.

2. Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}[x; n-1])$ be given by $(a_1, \ldots, a_n) \rightsquigarrow \sum\limits_{i=1}^{n} a_i x^{i-1}$, for $(a_1, \ldots, a_n) \in \mathbb{R}^n$. Then, T^{-1} maps $\sum\limits_{i=1}^{n} a_i x^{i-1} \rightsquigarrow (a_1, \ldots, a_n)$, for each polynomial $\sum\limits_{i=1}^{n} a_i x^{i-1} \in \mathbb{R}[x; n-1]$. Verify that $T^{-1} \in \mathcal{L}(\mathbb{R}[x; n-1], \mathbb{R}^n)$.

Definition 4.2.16. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, T is said to be **singular** if $\{\mathbf{0}\} \subsetneq \mathrm{KER}(T)$, *i.e.*, $\mathrm{KER}(T)$ contains a non-zero vector. If $\mathrm{KER}(T) = \{\mathbf{0}\}$ then, T is called **non-singular**.

Example 4.2.17. Let $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ be defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$. Then, verify that T is non-singular. Is T invertible?

We now prove a result that relates non-singularity with linear independence.

Theorem 4.2.18. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then the following statements are equivalent.*

1. T is one-one.

2. *T is non-singular.*

3. *Whenever $S \subseteq \mathbb{V}$ is linearly independent then $T(S)$ is necessarily linearly independent.*

Proof. 1⇒2 Let T be singular. Then, there exists $\mathbf{v} \neq \mathbf{0}$ such that $T(\mathbf{v}) = \mathbf{0} = T(\mathbf{0})$. This implies that T is not one-one, a contradiction.

2⇒3 Let $S \subseteq \mathbb{V}$ be linearly independent. Let if possible $T(S)$ be linearly dependent. Then, there exists $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$ and $\alpha = (\alpha_1, \ldots, \alpha_k)^T \neq \mathbf{0}$ such that $\sum\limits_{i=1}^{k} \alpha_i T(\mathbf{v}_i) = \mathbf{0}$. Thus, $T\left(\sum\limits_{i=1}^{k} \alpha_i \mathbf{v}_i\right) = \mathbf{0}$. But T is nonsingular and hence we get $\sum\limits_{i=1}^{k} \alpha_i \mathbf{v}_i = \mathbf{0}$ with $\alpha \neq \mathbf{0}$, a contradiction to S being a linearly independent set.

3⇒1 Suppose that T is not one-one. Then, there exists $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ such that $\mathbf{x} \neq \mathbf{y}$ but $T(\mathbf{x}) = T(\mathbf{y})$. Thus, we have obtained $S = \{\mathbf{x} - \mathbf{y}\}$, a linearly independent subset of $\mathbb{V}$ with $T(S) = \{\mathbf{0}\}$, a linearly dependent set. A contradiction to our assumption. Thus, the required result follows. ∎

Definition 4.2.19. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, T is said to be an **isomorphism** if T is one-one and onto. The vector spaces $\mathbb{V}$ and $\mathbb{W}$ are said to be **isomorphic**, denoted $\mathbb{V} \cong \mathbb{W}$, if there is an isomorphism from $\mathbb{V}$ to $\mathbb{W}$.

We now give a formal proof of the statement in Remark 3.5.12.

Theorem 4.2.20. *Let $\mathbb{V}$ be an n-dimensional vector space over $\mathbb{F}$. Then $\mathbb{V} \cong \mathbb{F}^n$.*

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbb{V}$ and $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, the standard basis of $\mathbb{F}^n$. Now define $T(\mathbf{v}_i) = \mathbf{e}_i$, for $1 \leq i \leq n$ and $T\left(\sum\limits_{i=1}^{n} \alpha_i \mathbf{v}_i\right) = \sum\limits_{i=1}^{n} \alpha_i \mathbf{e}_i$, for $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Then, it is easy to observe that $T \in \mathcal{L}(\mathbb{V}, \mathbb{F}^n)$, T is one-one and onto. Hence, T is an isomorphism. ∎

As a direct application using the countability argument, one obtains the following result

Corollary 4.2.21. *The vector space $\mathbb{R}$ over $\mathbb{Q}$ is not finite dimensional. Similarly, the vector space $\mathbb{C}$ over $\mathbb{Q}$ is not finite dimensional.*

We now summarize the different definitions related with a linear operator on a finite dimensional vector space. The proof basically uses the rank-nullity theorem and they appear in some form in previous results. Hence, we leave the proof for the reader.

Theorem 4.2.22. *Let $\mathbb{V}$ be a finite dimensional vector space over $\mathbb{F}$ with $\dim \mathbb{V} = n$. Then the following statements are equivalent for $T \in \mathcal{L}(\mathbb{V})$.*

1. *T is one-one.*

2. $\textsc{Ker}(T) = \{\mathbf{0}\}$.

3. $\mathsf{Rank}(T) = n$.

4. *T is onto.*

5. *T is an isomorphism.*

6. *If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{V}$ then so is $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$.*

7. *T is non-singular.*

8. *T is invertible.*

Exercise **4.2.23.** 1. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If $\dim(\mathbb{V})$ is finite then prove that*

   (a) *T cannot be onto if $\dim(\mathbb{V}) < \dim(\mathbb{W})$.*

   (b) *T cannot be one-one if $\dim(\mathbb{V}) > \dim(\mathbb{W})$.*

2. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the function $T : \mathrm{Col}(A^*) \to \mathrm{Col}(A)$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is invertible. [Hint: Use Theorem 3.4.13.3 and the rank-nullity theorem]*

   Ans: In view of Theorem 3.4.13.3 and the rank-nullity theorem, we just need to show that the map is one-one. So, suppose that there exist $\mathbf{x}, \mathbf{y} \in \mathrm{Col}(A^*)$ such that $T(\mathbf{x}) = T(\mathbf{y})$. Or equivalently, $A\mathbf{x} = A\mathbf{y}$. Thus, $\mathbf{x} - \mathbf{y} \in \mathrm{Null}(A) = (\mathrm{Col}(A^*))^{\perp}$ (by Theorem 3.4.13.2). Therefore, $\mathbf{x} - \mathbf{y} \in (\mathrm{Col}(A^*))^{\perp} \cap \mathrm{Col}(A^*) = \{\mathbf{0}\}$. Thus, $\mathbf{x} = \mathbf{y}$ and hence the map is one-one. Thus, the required result follows.

4.3 Matrix of a linear transformation

In Example 4.1.3.8, we saw that for each $A \in M_{m\times n}(\mathbb{C})$ there exists a linear transformation $T \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ given by $T(\mathbf{x}) = A\mathbf{x}$, for each $\mathbf{x} \in \mathbb{C}^n$. In this section, we prove that if $\mathbb{V}$ and $\mathbb{W}$ are vector spaces over $\mathbb{F}$ with dimensions n and m, respectively, then any $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ corresponds to a set of $m \times n$ matrices. Before proceeding further, the readers should recall the results on ordered basis (see Section 3.5).

So, let $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. Also, let $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $B = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ be the basis matrix of $\mathcal{A}$ and $\mathcal{B}$, respectively. Then, using Equation (3.5.1), $\mathbf{v} = A[\mathbf{v}]_{\mathcal{A}}$ and $\mathbf{w} = B[\mathbf{w}]_{\mathcal{B}}$, for all $\mathbf{v} \in \mathbb{V}$ and $\mathbf{w} \in \mathbb{W}$. Thus, we see that for $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $\mathbf{x} \in \mathbb{V}$ (T is determined by its image on basis vectors),

$$\begin{aligned} B[\mathbf{T}(\mathbf{x})]_{\mathcal{B}} &= T(\mathbf{x}) = T([\mathbf{v}_1, \ldots, \mathbf{v}_n][\mathbf{x}]_{\mathcal{A}}) = \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} [\mathbf{x}]_{\mathcal{A}} \\ &= \begin{bmatrix} B[T(\mathbf{v}_1)]_{\mathcal{B}} & \cdots & B[T(\mathbf{v}_n)]_{\mathcal{B}} \end{bmatrix} [\mathbf{x}]_{\mathcal{A}} = B\begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{B}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{B}} \end{bmatrix} [\mathbf{x}]_{\mathcal{A}}. \end{aligned}$$

As B is an invertible matrix, we cancel it to get $[\mathbf{T}(\mathbf{x})]_{\mathcal{B}} = [[T(\mathbf{v}_1)]_{\mathcal{B}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}}]\,[\mathbf{x}]_{\mathcal{A}}$, for each $\mathbf{x} \in \mathbb{V}$. Note that the matrix $\begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{B}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{B}} \end{bmatrix}$, denoted $T[\mathcal{A}, \mathcal{B}]$, is an $m \times n$ matrix and is unique with respect to the ordered basis $\mathcal{B}$ as the i-th column equals $[T(\mathbf{v}_i)]_{\mathcal{B}}$, for $1 \le i \le n$. So, we immediately have the following definition and result.

Definition 4.3.1. Let $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then the matrix $T[\mathcal{A}, \mathcal{B}]$ is called the **coordinate matrix** of T or the **matrix of the linear transformation** T with respect to the basis $\mathcal{A}$ and $\mathcal{B}$, respectively.

When there is no mention of bases, we take it to be the standard ordered bases and denote the corresponding matrix by $[T]$. Also, note that for each $\mathbf{x} \in \mathbb{V}$, the matrix $T[\mathcal{A}, \mathcal{B}][\mathbf{x}]_{\mathcal{A}}$ is the coordinate vector of $T(\mathbf{x})$. Thus, the matrix $T[\mathcal{A}, \mathcal{B}]$ takes coordinate vector of the domain points to the coordinate vector of its images. The above discussion is stated as the next result.

Theorem 4.3.2. *Let $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then there exists a matrix $S \in M_{m\times n}(\mathbb{F})$ with*

$$S = T[\mathcal{A}, \mathcal{B}] = \Big[[T(\mathbf{v}_1)]_{\mathcal{B}} \quad \cdots \quad [T(\mathbf{v}_n)]_{\mathcal{B}}\Big] \text{ and } [T(\mathbf{x})]_{\mathcal{B}} = S\,[\mathbf{x}]_{\mathcal{A}}, \text{ for all } \mathbf{x} \in \mathbb{V}.$$

Remark 4.3.3. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{A}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B}_1 = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$, respectively. Also, for $\alpha \in \mathbb{F}$ with $\alpha \neq 0$, let $\mathcal{A}_2 = (\alpha\mathbf{v}_1, \ldots, \alpha\mathbf{v}_n)$ and $\mathcal{B}_1 = (\alpha\mathbf{w}_1, \ldots, \alpha\mathbf{w}_m)$ be another set of ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. Then, for any $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$*

$$T[\mathcal{A}_2, \mathcal{B}_2] = \Big[[T(\alpha\mathbf{v}_1)]_{\mathcal{B}_2} \quad \cdots \quad [T(\alpha\mathbf{v}_1)]_{\mathcal{B}_2}\Big] = \Big[[T(\mathbf{v}_n)]_{\mathcal{B}_1} \quad \cdots \quad [T(\mathbf{v}_1)]_{\mathcal{B}_1}\Big] = T[\mathcal{A}_1, \mathcal{B}_1].$$

Thus, we see that the same matrix can be the matrix representation of T for two different pairs of bases.

We now give a few examples to understand the above discussion and Theorem 4.3.2.

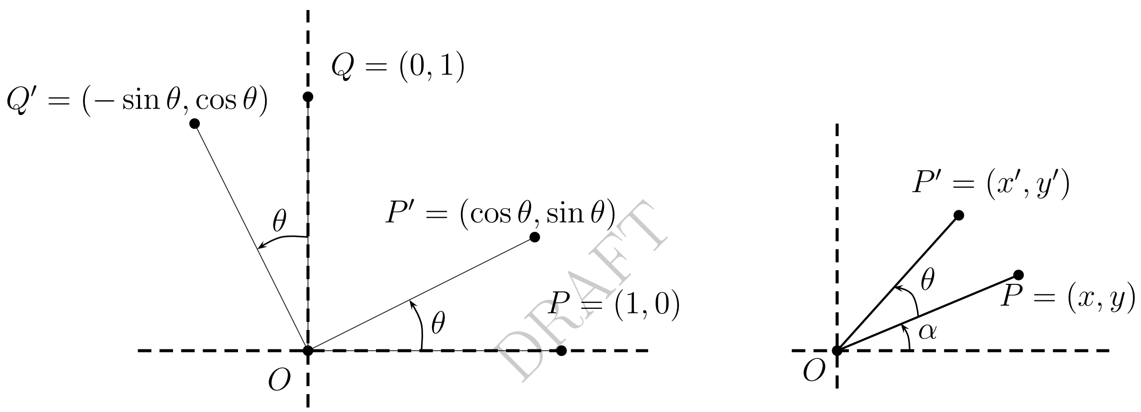


Figure 4.1: Counter-clockwise Rotation by an angle θ

Example 4.3.4. 1. Let $T \in \mathcal{L}(\mathbb{R}^2)$ represent a counterclockwise rotation by an angle $\theta, 0 \leq \theta < 2\pi$. Then, using Figure 4.1, $x = OP\cos\alpha$ and $y = OP\sin\alpha$, verify that

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} OP'\cos(\alpha+\theta) \\ OP'\sin(\alpha+\theta) \end{bmatrix} = \begin{bmatrix} OP\big(\cos\alpha\cos\theta - \sin\alpha\sin\theta\big) \\ OP\big(\sin\alpha\cos\theta + \cos\alpha\sin\theta\big) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Or equivalently, the matrix in the standard ordered basis of $\mathbb{R}^2$ equals

$$[T] = \Big[T(\mathbf{e}_1), T(\mathbf{e}_2)\Big] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{4.3.1}$$

2. Let $T \in \mathcal{L}(\mathbb{R}^2)$ with $T((x,y)^T) = (x+y, x-y)^T$.

   (a) Then $[T] = \Big[[T(\mathbf{e}_1)] \quad [T(\mathbf{e}_2)]\Big] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

   (b) On the image space take the ordered basis $\mathcal{B} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$. Then

$$[T] = \Big[[T(\mathbf{e}_1)]_{\mathcal{B}} \quad [T(\mathbf{e}_2)]_{\mathcal{B}}\Big] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{B}} \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}.$$

(c) In the above, let the ordered basis of the domain space be $\mathcal{A} = \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}\right)$. Then

$$T[\mathcal{A},\mathcal{B}] = \left[\left[T\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right]_{\mathcal{B}} \left[T\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right]_{\mathcal{B}}\right] = \left[\begin{bmatrix} 0 \\ -2 \end{bmatrix}_{\mathcal{B}} \begin{bmatrix} 4 \\ 2 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}.$$

3. Let $\mathcal{A} = (\mathbf{e}_1, \mathbf{e}_2)$ and $\mathcal{B} = (\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_2)$ be two ordered bases of $\mathbb{R}^2$. Then Compute $T[\mathcal{A},\mathcal{A}]$ and $T[\mathcal{B},\mathcal{B}]$, where $T((x,y)^T) = (x+y, x-2y)^T$.

Solution: Let $A = \mathrm{Id}_2$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then, $A^{-1} = \mathrm{Id}_2$ and $B^{-1} = \dfrac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. So,

$$T[\mathcal{A},\mathcal{A}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{A}}, \left[T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{A}}\right] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{A}}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}_{\mathcal{A}}\right] = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \text{ and}$$

$$T[\mathcal{B},\mathcal{B}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}}, \left[T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\right]_{\mathcal{B}}\right] = \left[\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{bmatrix}$$

as $\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{B}} = B^{-1}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{B}} = B^{-1}\begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Also, verify that $T[\mathcal{B},\mathcal{B}] = B^{-1}T[\mathcal{A},\mathcal{A}]B$.

4. Let $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ be defined by $T((x,y,z)^T) = (x+y-z, x+z)^T$. Determine $[T]$. By definition

$$[T] = [[T(\mathbf{e}_1)], [T(\mathbf{e}_2)], [T(\mathbf{e}_3)]] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right] = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

5. Define $T \in \mathcal{L}(\mathbb{C}^3)$ by $T(\mathbf{x}) = \mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^3$. Determine the coordinate matrix with respect to the ordered basis $\mathcal{A} = \big(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\big)$ and $\mathcal{B} = \big((1,0,0), (1,1,0), (1,1,1)\big)$. By definition, verify that

$$T[\mathcal{A},\mathcal{B}] = [[T(\mathbf{e}_1)]_{\mathcal{B}}, [T(\mathbf{e}_2)]_{\mathcal{B}}, [T(\mathbf{e}_3)]_{\mathcal{B}}] = \left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$T[\mathcal{B},\mathcal{A}] = \left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{A}}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{A}}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{A}}\right] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, verify that $T[\mathcal{B},\mathcal{A}]^{-1} = T[\mathcal{A},\mathcal{B}]$ and $T[\mathcal{A},\mathcal{A}] = T[\mathcal{B},\mathcal{B}] = I_3$ as the given map is indeed the identity map.

6. Fix $S \in \mathbb{M}_n(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n)$ by $T(\mathbf{x}) = S\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. If $\mathcal{A}$ is the standard basis of $\mathbb{C}^n$ then $[T] = S$ as

$$[T][:,i] = [T(\mathbf{e}_i)]_{\mathcal{A}} = [S(\mathbf{e}_i)]_{\mathcal{A}} = [S[:,i]]_{\mathcal{A}} = S[:,i], \quad \text{for } 1 \le i \le n.$$

7. Fix $S \in \mathbb{M}_{m,n}(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ by $T(\mathbf{x}) = S\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. Let $\mathcal{A}$ and $\mathcal{B}$ be the standard ordered bases of $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. Then $T[\mathcal{A},\mathcal{B}] = S$ as

$$(T[\mathcal{A},\mathcal{B}])[:,i] = [T(\mathbf{e}_i)]_{\mathcal{B}} = [S(\mathbf{e}_i)]_{\mathcal{B}} = [S[:,i]]_{\mathcal{B}} = S[:,i], \quad \text{for } 1 \le i \le n.$$

8. Fix $S \in \mathbb{M}_n(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n)$ by $T(\mathbf{x}) = S\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. Let $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ be two ordered basses of $\mathbb{C}^n$ with respective basis matrices A and B. Then

$$\begin{aligned} T[\mathcal{A},\mathcal{B}] &= \begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{B}} & \cdots & [T(\mathbf{v}_1)]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} B^{-1}T(\mathbf{v}_1) & \cdots & B^{-1}T(\mathbf{v}_1) \end{bmatrix} \\ &= \begin{bmatrix} B^{-1}S\mathbf{v}_1 & \cdots & B^{-1}S\mathbf{v}_1 \end{bmatrix} = B^{-1}S \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = B^{-1}SA. \end{aligned}$$

In particular, if

(a) $\mathcal{A} = \mathcal{B}$ then $T[\mathcal{A},\mathcal{A}] = A^{-1}SA$. Thus, if $S = I_n$ so that $T = \mathrm{Id}$ then $\mathrm{Id}[\mathcal{A},\mathcal{A}] = I_n$.

(b) $S = I_n$ so that $T = \mathrm{Id}$ then $\mathrm{Id}[\mathcal{A},\mathcal{B}] = B^{-1}A$, an invertible matrix. Similarly, $\mathrm{Id}[\mathcal{B},\mathcal{A}] = A^{-1}B$. So, $\mathrm{Id}[\mathcal{B},\mathcal{A}] \cdot \mathrm{Id}[\mathcal{A},\mathcal{B}] = (A^{-1}B)(B^{-1}A) = I_n$.

9. Let $T\left((x,y)^T\right) = (x+y, x-y)^T$ and $\mathcal{A} = (\mathbf{e}_1, \mathbf{e}_1 + \mathbf{e}_2)$ be the ordered basis of $\mathbb{R}^2$. Then, using Example 4.3.4.8a we obtain

$$T[\mathcal{A},\mathcal{A}] = \left(\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_1 + \mathbf{e}_2 \end{bmatrix}\right)^{-1} \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_1 + \mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}.$$

Example 4.3.5. [Finding T from $T[\mathcal{A},\mathcal{B}]$]

1. Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{A}$ and $\mathcal{B}$, respectively. Suppose we are given the matrix $S = T[\mathcal{A},\mathcal{B}]$. Then determine the corresponding $T \in \mathcal{L}(\mathbb{V},\mathbb{W})$.

   Solution: Let B be the basis matrix corresponding to the ordered basis $\mathcal{B}$. Then, using Equation (3.5.1) and Theorem 4.3.2, we see that

$$T(\mathbf{v}) = B[T(\mathbf{v})]_{\mathcal{B}} = BT[\mathcal{A},\mathcal{B}][\mathbf{v}]_{\mathcal{A}} = BS[\mathbf{v}]_{\mathcal{A}}.$$

2. In particular, if $\mathbb{V} = \mathbb{W} = \mathbb{F}^n$ and $\mathcal{A} = \mathcal{B}$ then we see that

$$T(\mathbf{v}) = BSB^{-1}\mathbf{v}. \tag{4.3.2}$$

EXERCISE **4.3.6.** *1. Let $T \in \mathcal{L}(\mathbb{R}^2)$ represent the reflection about the line $y = mx$. Find $[T]$.*

2. Let $T \in \mathcal{L}(\mathbb{R}^3)$ represent the reflection about the X-axis. Find $[T]$.

3. Let $T \in \mathcal{L}(\mathbb{R}^3)$ represent the counterclockwise rotation around the positive Z-axis by an angle $\theta, 0 \le \theta < 2\pi$. 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 $D \in \mathcal{L}(\mathbb{R}[x;n])$ by $D(f(x)) = f'(x)$. Find the matrix of D with respect to the standard ordered basis of $\mathbb{R}[x;n]$. Observe that $\mathrm{RNG}(D) \subseteq \mathbb{R}[x;n-1]$.

$$\begin{tikzcd}[column sep=large]
(\mathbb{V}, \mathcal{B}, n) \arrow[r, "T[\mathcal{B},\mathcal{C}]_{m\times n}"] \arrow[rr, bend right, "(ST)[\mathcal{B},\mathcal{D}]_{p\times n} = S[\mathcal{C},\mathcal{D}]\cdot T[\mathcal{B},\mathcal{C}]"'] & (\mathbb{W}, \mathcal{C}, m) \arrow[r, "S[\mathcal{C},\mathcal{D}]_{p\times m}"] & (\mathbb{Z}, \mathcal{D}, p)
\end{tikzcd}$$

Figure 4.2: Composition of Linear Transformations

4.4 Similarity of Matrices

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$ and ordered basis $\mathcal{B}$. Then any $T \in \mathcal{L}(\mathbb{V})$ corresponds to a matrix in $\mathbb{M}_n(\mathbb{F})$. What happens if the ordered basis needs to change? We answer this in this subsection.

Theorem 4.4.1 (Composition of Linear Transformations)**.** *Let $\mathbb{V}$, $\mathbb{W}$ and $\mathbb{Z}$ be finite dimensional vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{B}, \mathcal{C}$ and $\mathcal{D}$, respectively. Also, let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $S \in \mathcal{L}(\mathbb{W}, \mathbb{Z})$. Then $S \circ T = ST \in \mathcal{L}(\mathbb{V}, \mathbb{Z})$ (see Figure 4.2). Then*

$$(ST)\ [\mathcal{B}, \mathcal{D}] = S[\mathcal{C}, \mathcal{D}]\ \cdot\ T[\mathcal{B}, \mathcal{C}].$$

Proof. Let $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, $\mathcal{C} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ and $\mathcal{D} = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$ be the ordered bases of $\mathbb{V}, \mathbb{W}$ and $\mathbb{Z}$, respectively. Then using Theorem 4.3.2, we have

$$\begin{aligned}
(ST)[\mathcal{B}, \mathcal{D}] &= [[ST(\mathbf{u}_1)]_{\mathcal{D}}, \ldots, [ST(\mathbf{u}_n)]_{\mathcal{D}}] = [[S(T(\mathbf{u}_1))]_{\mathcal{D}}, \ldots, [S(T(\mathbf{u}_n))]_{\mathcal{D}}] \\
&= [S[\mathcal{C}, \mathcal{D}]\,[T(\mathbf{u}_1)]_{\mathcal{C}}, \ldots, S[\mathcal{C}, \mathcal{D}]\,[T(\mathbf{u}_n)]_{\mathcal{C}}] \\
&= S[\mathcal{C}, \mathcal{D}]\ \ [[T(\mathbf{u}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{u}_n)]_{\mathcal{C}}] = S[\mathcal{C}, \mathcal{D}]\ \cdot\ T[\mathcal{B}, \mathcal{C}].
\end{aligned}$$

Hence, the proof of the theorem is complete. ∎

As an immediate corollary of Theorem 4.4.1 we have the following result.

Theorem 4.4.2 (Inverse of a Linear Transformation)**.** *Let $\mathbb{V}$ is a vector space with $\dim(\mathbb{V}) = n$. If $T \in \mathcal{L}(\mathbb{V})$ is invertible then for any ordered basis $\mathcal{B}$ and $\mathcal{C}$ of the domain and co-domain, respectively, one has $(T[\mathcal{C}, \mathcal{B}])^{-1} = T^{-1}[\mathcal{B}, \mathcal{C}]$. That is, the inverse of the coordinate matrix of T is the coordinate matrix of the inverse linear transform.*

Proof. As T is invertible, $TT^{-1} = \mathrm{Id}$. Thus, Example 4.3.4.8a and Theorem 4.4.1 imply

$$I_n = \mathrm{Id}[\mathcal{B}, \mathcal{B}] = (TT^{-1})[\mathcal{B}, \mathcal{B}] = T[\mathcal{C}, \mathcal{B}]\ \cdot T^{-1}[\mathcal{B}, \mathcal{C}].$$

Hence, by definition of inverse, $T^{-1}[\mathcal{B}, \mathcal{C}] = (T[\mathcal{C}, \mathcal{B}])^{-1}$ and the required result follows. ∎

Exercise **4.4.3.** *Find the matrix of the linear transformations given below.*

1. *Let $\mathcal{B} = \big(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\big)$ be an ordered basis of $\mathbb{R}^3$. Now, define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T(\mathbf{x}_1) = \mathbf{x}_2$, $T(\mathbf{x}_2) = \mathbf{x}_3$ and $T(\mathbf{x}_3) = \mathbf{x}_1$. Determine $T[\mathcal{B}, \mathcal{B}]$. Is T invertible?*

2. *Let $\mathcal{B} = \big(1, x, x^2, x^3\big)$ be an ordered basis of $\mathbb{R}[x; 3]$ and define $T \in \mathcal{L}(\mathbb{R}[x; 3])$ by $T(1) = 1$, $T(x) = 1 + x$, $T(x^2) = (1 + x)^2$ and $T(x^3) = (1 + x)^3$. Prove that T is invertible. Also, find $T[\mathcal{B}, \mathcal{B}]$ and $T^{-1}[\mathcal{B}, \mathcal{B}]$.*

$$\begin{tikzcd}
(\mathbb{V},\mathcal{B}) \arrow[r, "T[\mathcal{B},\mathcal{B}]"] & (\mathbb{V},\mathcal{B}) \arrow[d, "\mathrm{Id}[\mathcal{B},\mathcal{C}]"] \\
(\mathbb{V},\mathcal{C}) \arrow[u, "\mathrm{Id}[\mathcal{C},\mathcal{B}]"] \arrow[r, "T[\mathcal{C},\mathcal{C}]"'] & (\mathbb{V},\mathcal{C})
\end{tikzcd}$$

Figure 4.3: $T[\mathcal{C},\mathcal{C}] = \mathrm{Id}[\mathcal{B},\mathcal{C}] \cdot T[\mathcal{B},\mathcal{B}] \cdot \mathrm{Id}[\mathcal{C},\mathcal{B}]$ - Similarity of Matrices

Let $\mathbb{V}$ be a finite dimensional vector space. Then, the next result answers the question "what happens to the matrix $T[\mathcal{B},\mathcal{B}]$ if the ordered basis $\mathcal{B}$ changes to $\mathcal{C}$?"

Theorem 4.4.4. *Let $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and $\mathcal{C} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ be two ordered bases of $\mathbb{V}$ and Id the identity operator. Then, for any linear operator $T \in \mathcal{L}(\mathbb{V})$*

$$T[\mathcal{C},\mathcal{C}] = Id[\mathcal{B},\mathcal{C}] \cdot T[\mathcal{B},\mathcal{B}] \cdot Id[\mathcal{C},\mathcal{B}] = (Id[\mathcal{C},\mathcal{B}])^{-1} \cdot T[\mathcal{B},\mathcal{B}] \cdot Id[\mathcal{C},\mathcal{B}]. \tag{4.4.1}$$

Proof. As Id is an identity operator, $T[\mathcal{B},\mathcal{C}]$ as $(\mathrm{Id} \circ T \circ \mathrm{Id})[\mathcal{B},\mathcal{C}]$ (see Figure 4.3 for clarity). Thus, using Theorem 4.4.1, we get

$$T[\mathcal{B},\mathcal{C}] = (\mathrm{Id} \circ T \circ \mathrm{Id})[\mathcal{B},\mathcal{C}] = \mathrm{Id}[\mathcal{B},\mathcal{C}] \cdot T[\mathcal{B},\mathcal{B}] \cdot \mathrm{Id}[\mathcal{C},\mathcal{B}].$$

Hence, using Theorem 4.4.2, the required result follows. ∎

Let $\mathbb{V}$ be a vector space and let $T \in \mathcal{L}(\mathbb{V})$. If $\dim(\mathbb{V}) = n$ then every ordered basis $\mathcal{B}$ of $\mathbb{V}$ gives an $n \times n$ matrix $T[\mathcal{B},\mathcal{B}]$. So, as we change the ordered basis, the coordinate matrix of T changes. Theorem 4.4.4 tells us that all these matrices are related by an invertible matrix. Thus, we are led to the following definitions.

Definition 4.4.5. Let $\mathbb{V}$ be a vector space with ordered bases $\mathcal{B}$ and $\mathcal{C}$. If $T \in \mathcal{L}(\mathbb{V})$ then, $T[\mathcal{C},\mathcal{C}] = \mathrm{Id}[\mathcal{B},\mathcal{C}] \cdot T[\mathcal{B},\mathcal{B}] \cdot \mathrm{Id}[\mathcal{C},\mathcal{B}]$. The matrix $\mathrm{Id}[\mathcal{B},\mathcal{C}]$ is called the **change of basis matrix** (also, see Theorem 3.5.10) from $\mathcal{B}$ to $\mathcal{C}$.

Definition 4.4.6. Let $X, Y \in \mathbb{M}_n(\mathbb{C})$. Then, X and Y are said to be **similar** if there exists a non-singular matrix P such that $P^{-1}XP = Y \Leftrightarrow XP = PY$.

Example 4.4.7. Let $\mathcal{B} = \left(1+x, 1+2x+x^2, 2+x\right)$ and $\mathcal{C} = \left(1, 1+x, 1+x+x^2\right)$ be ordered bases of $\mathbb{R}[x;2]$. Then, verify that $\mathrm{Id}[\mathcal{B},\mathcal{C}]^{-1} = \mathrm{Id}[\mathcal{C},\mathcal{B}]$, as

$$\begin{aligned}
\mathrm{Id}[\mathcal{C},\mathcal{B}] &= [[1]_{\mathcal{B}}, [1+x]_{\mathcal{B}}, [1+x+x^2]_{\mathcal{B}}] = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and} \\
\mathrm{Id}[\mathcal{B},\mathcal{C}] &= [[1+x]_{\mathcal{C}}, [1+2x+x^2]_{\mathcal{C}}, [2+x]_{\mathcal{C}}] = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.
\end{aligned}$$

EXERCISE **4.4.8.** *1. Let $\mathbb{V}$ be a vector space with $\dim(\mathbb{V}) = n$. Let $T \in \mathcal{L}(\mathbb{V})$ satisfy $T^{n-1} \neq \mathbf{0}$ but $T^n = \mathbf{0}$. Then, use Exercise 4.1.11.3 to get an ordered basis $\mathcal{B} = \big(\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\big)$ of $\mathbb{V}$.*

(a) Now, prove that $T[\mathcal{B},\mathcal{B}] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$.

(b) Let $A \in \mathbb{M}_n(\mathbb{C})$ satisfy $A^{n-1} \neq \mathbf{0}$ but $A^n = \mathbf{0}$. Then, prove that A is similar to the matrix given in Part 1a.

2. Let $\mathcal{A}$ be an ordered basis of a vector space $\mathbb{V}$ over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Then prove that the set of all possible matrix representations of T is given by (also see Definition 4.4.5)

$$\{S \cdot T[\mathcal{A},\mathcal{A}] \cdot S^{-1} \mid S \in \mathbb{M}_n(\mathbb{F}) \text{ is an invertible matrix}\}.$$

3. Let $B_1(\alpha,\beta) = \{(x,y)^T \in \mathbb{R}^2 : (x-\alpha)^2 + (y-\beta)^2 \leq 1\}$. Then, can we get a linear transformation $T \in \mathcal{L}(\mathbb{R}^2)$ such that $T(S) = W$, where S and W are given below?

(a) $S = B_1(0,0)$ and $W = B_1(1,1)$.

(b) $S = B_1(0,0)$ and $W = B_1(.3,0)$.

(c) $S = B_1(0,0)$ and $W = hull(\pm e_1, \pm e_2)$, where hull means the **convex hull**.

(d) $S = B_1(0,0)$ and $W = \{(x,y)^T \in \mathbb{R}^2 : x^2 + y^2/4 = 1\}$.

(e) $S = hull(\pm e_1, \pm e_2)$ and W is the interior of a pentagon.

4. Let $\mathbb{V}$, $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$ and $\dim(\mathbb{W}) = m$ and ordered bases $\mathcal{B}$ and $\mathcal{C}$, respectively. Define $\mathcal{I}_{\mathcal{B},\mathcal{C}} : \mathcal{L}(\mathbb{V},\mathbb{W}) \to \mathbb{M}_{m,n}(\mathbb{F})$ by $\mathcal{I}_{\mathcal{B},\mathcal{C}}(T) = T[\mathcal{B},\mathcal{C}]$. Show that $\mathcal{I}_{\mathcal{B},\mathcal{C}}$ is an isomorphism. Thus, when bases are fixed, the number of $m \times n$ matrices is same as the number of linear transformations.

5. Define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T((x,y,z)^T) = (x+y+2z, x-y-3z, 2x+3y+z)^T$. Let $\mathcal{B}$ be the standard ordered basis and $\mathcal{C} = \big((1,1,1),(1,-1,1),(1,1,2)\big)$ be another ordered basis of $\mathbb{R}^3$. Then find

(a) matrices $T[\mathcal{B},\mathcal{B}]$ and $T[\mathcal{C},\mathcal{C}]$.

(b) the matrix P such that $P^{-1}T[\mathcal{B},\mathcal{B}]\, P = T[\mathcal{C},\mathcal{C}]$.

4.5 Dual Space*

Definition 4.5.1. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then a map $T \in \mathcal{L}(\mathbb{V},\mathbb{F})$ is called a **linear functional** on $\mathbb{V}$.

Example 4.5.2. 1. Let $\mathbf{a} \in \mathbb{C}^n$ be fixed. Then, $T(\mathbf{x}) = \mathbf{a}^*\mathbf{x}$ is a linear function from $\mathbb{C}^n$ to $\mathbb{C}$.

2. Define $T(A) = \mathsf{tr}(\mathsf{A})$, for all $A \in \mathbb{M}_n(\mathbb{R})$. Then, T is a linear functional from $\mathbb{M}_n(\mathbb{R})$ to $\mathbb{R}$.
3. Define $T(f) = \int\limits_a^b f(t)dt$, for all $f \in \mathcal{C}([a,b],\mathbb{R})$. Then, T is a linear functional from $\mathcal{L}(\mathcal{C}([a,b],\mathbb{R})$ to $\mathbb{R}$.
4. Define $T(f) = \int\limits_a^b t^2 f(t)dt$, , for all $f \in \mathcal{C}([a,b],\mathbb{R})$. Then, T is a linear functional from $\mathcal{L}(\mathcal{C}([a,b],\mathbb{R})$ to $\mathbb{R}$.
5. Define $T : \mathbb{C}^3 \to \mathbb{C}$ by $T((x,y,z)^T) = x$. Is it a linear functional?
6. Let $\mathcal{B}$ be a basis of a vector space $\mathbb{V}$ over $\mathbb{F}$. For a fixed element $\mathbf{u} \in \mathcal{B}$, define

$$T(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{u} \\ 0 & \text{if } \mathbf{x} \in \mathcal{B} \setminus \mathbf{u}. \end{cases}$$

   Now, extend T linearly to all of $\mathbb{V}$. Does, T give rise to a linear functional?

Definition 4.5.3. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then $\mathcal{L}(\mathbb{V},\mathbb{F})$ is called the **dual space** of $\mathbb{V}$ and is denoted by $\mathbb{V}^*$. The **double dual space** of $\mathbb{V}$, denoted $\mathbb{V}^{**}$, is the dual space of $\mathbb{V}^*$.

We first give an immediate corollary of Theorem 4.2.20.

Corollary 4.5.4. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with $\dim \mathbb{V} = n$ and $\dim \mathbb{W} = m$.*

1. *Then $\mathcal{L}(\mathbb{V},\mathbb{W}) \cong \mathbb{F}^{mn}$. Moreover, $\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}$ is a basis of $\mathcal{L}(\mathbb{V},\mathbb{W})$.*
2. *In particular, if $\mathbb{W} = \mathbb{F}$ then $\mathcal{L}(\mathbb{V},\mathbb{F}) = \mathbb{V}^* \cong \mathbb{F}^n$. Moreover, if $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ is a basis of $\mathbb{V}$ then the set $\{\mathbf{f}_i | 1 \le i \le n\}$ is a basis of $\mathbb{V}^*$, where $\mathbf{f}_i(\mathbf{v}_k) = \begin{cases} 1, & \text{if } k = i \\ \mathbf{0}, & k \ne i. \end{cases}$ The basis $\{\mathbf{f}_i | 1 \le i \le n\}$ is called the* **dual basis** *of $\mathbb{F}^n$.*

Exercise **4.5.5.** *Let $\mathbb{V}$ be a vector space. Suppose there exists $\mathbf{v} \in \mathbb{V}$ such that $\mathbf{f}(\mathbf{v}) = 0$, for all $\mathbf{f} \in \mathbb{V}^*$. Then prove that $\mathbf{v} = \mathbf{0}$.*

So, we see that $\mathbb{V}^*$ can be understood through a basis of $\mathbb{V}$. Thus, one can understand $\mathbb{V}^{**}$ again via a basis of $\mathbb{V}^*$. But, the question arises "can we understand it directly via the vector space $\mathbb{V}$ itself?" We answer this in affirmative by giving a canonical isomorphism from $\mathbb{V}$ to $\mathbb{V}^{**}$. To do so, for each $\mathbf{v} \in \mathbb{V}$, we define a map $L_{\mathbf{v}} : \mathbb{V}^* \to \mathbb{F}$ by $L_{\mathbf{v}}(\mathbf{f}) = \mathbf{f}(\mathbf{v})$, for each $\mathbf{f} \in \mathbb{V}^*$. Then $L_{\mathbf{v}}$ is a linear functional as

$$L_{\mathbf{v}}(\alpha\mathbf{f} + \mathbf{g}) = (\alpha\mathbf{f} + \mathbf{g})(\mathbf{v}) = \alpha\mathbf{f}(\mathbf{v}) + \mathbf{g}(\mathbf{v}) = \alpha L_{\mathbf{v}}(\mathbf{f}) + L_{\mathbf{v}}(\mathbf{g}).$$

So, for each $\mathbf{v} \in \mathbb{V}$, we have obtained a linear functional $L_{\mathbf{v}} \in \mathbb{V}^{**}$. Note that, if $\mathbf{v} \ne \mathbf{w}$ then, $L_{\mathbf{v}} \ne L_{\mathbf{w}}$. Indeed, if $L_{\mathbf{v}} = L_{\mathbf{w}}$ then, $L_{\mathbf{v}}(f) = L_{\mathbf{w}}(f)$, for all $f \in \mathbb{V}^*$. Thus, $f(\mathbf{v}) = f(\mathbf{w})$, for all $f \in \mathbb{V}^*$. That is, $f(\mathbf{v} - \mathbf{w}) = 0$, for each $f \in \mathbb{V}^*$. Hence, using Exercise 4.5.5, we get $\mathbf{v} - \mathbf{w} = \mathbf{0}$, or equivalently, $\mathbf{v} = \mathbf{w}$.

We use the above argument to give the required canonical isomorphism.

Theorem 4.5.6. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. If $\dim(\mathbb{V}) = n$ then the canonical map $T : \mathbb{V} \to \mathbb{V}^{**}$ defined by $T(\mathbf{v}) = L_{\mathbf{v}}$ is an isomorphism.*

Proof. Note that for each $\mathbf{f} \in \mathbb{V}^*$,

$$L_{\alpha\mathbf{v}+\mathbf{u}}(\mathbf{f}) = \mathbf{f}(\alpha\mathbf{v}+\mathbf{u}) = \alpha\mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{u}) = \alpha L_{\mathbf{v}}(\mathbf{f}) + L_{\mathbf{u}}(\mathbf{f}) = (\alpha L_{\mathbf{v}} + L_{\mathbf{u}})(\mathbf{f}).$$

Thus, $L_{\alpha\mathbf{v}+\mathbf{u}} = \alpha L_{\mathbf{v}} + L_{\mathbf{u}}$. Hence, $T(\alpha\mathbf{v}+\mathbf{u}) = \alpha T(\mathbf{v}) + T(\mathbf{u})$. Thus, T is a linear transformation. For verifying T is one-one, assume that $T(\mathbf{v}) = T(\mathbf{u})$, for some $\mathbf{u}, \mathbf{v} \in \mathbb{V}$. Then, $L_{\mathbf{v}} = L_{\mathbf{u}}$. Now, use the argument just before this theorem to get $\mathbf{v} = \mathbf{u}$. Therefore, T is one-one.

Thus, T gives an inclusion (one-one) map from $\mathbb{V}$ to $\mathbb{V}^{**}$. Further, applying Corollary 4.5.4.2 to $\mathbb{V}^*$, gives $\dim(\mathbb{V}^{**}) = \dim(\mathbb{V}^*) = n$. Hence, the required result follows. ∎

We now give a few immediate consequences of Theorem 4.5.6.

Corollary 4.5.7. *Let $\mathbb{V}$ be a vector space of dimension n with basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.*

1. *Then, a basis of $\mathbb{V}^{**}$, the double dual of $\mathbb{V}$, equals $\mathcal{D} = \{L_{\mathbf{v}_1}, \ldots, L_{\mathbf{v}_n}\}$. Thus, for each $T \in \mathbb{V}^{**}$ there exists $\mathbf{x} \in \mathbb{V}$ such that $T(\mathbf{f}) = \mathbf{f}(\mathbf{x})$, for all $\mathbf{f} \in \mathbb{V}^*$. Or equivalently, there exists $\mathbf{x} \in \mathbb{V}$ such that $T = T_{\mathbf{x}}$.*
2. *If $\mathcal{C} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ is the dual basis of $\mathbb{V}^*$ defined using the basis $\mathcal{B}$ (see Corollary 4.5.4.2) then $\mathcal{D}$ is indeed the dual basis of $\mathbb{V}^{**}$ obtained using the basis $\mathcal{C}$ of $\mathbb{V}^*$. Thus, each basis of $\mathbb{V}^*$ is the dual basis of some basis of $\mathbb{V}$.*

Proof. Part 1 is direct as $T : \mathbb{V} \to \mathbb{V}^{**}$ was a canonical inclusion map. For Part 2, we need to show that

$$L_{\mathbf{v}_i}(\mathbf{f}_j) = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases} \quad \text{or equivalently } \mathbf{f}_j(\mathbf{v}_i) = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i \end{cases}$$

which indeed holds true using Corollary 4.5.4.2. ∎

Let $\mathbb{V}$ be a finite dimensional vector space. Then Corollary 4.5.7 implies that the spaces $\mathbb{V}$ and $\mathbb{V}^*$ are naturally dual to each other.

We are now ready to prove the main result of this subsection. To start with, let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. Then, for each $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$, we want to define a map $\widehat{T} : \mathbb{W}^* \to \mathbb{V}^*$. So, if $g \in \mathbb{W}^*$ then, $\widehat{T}(\mathbf{g})$ a linear functional from $\mathbb{V}$ to $\mathbb{F}$. So, we need to be evaluate $\widehat{T}(\mathbf{g})$ at an element of $\mathbb{V}$. Thus, we define $\left(\widehat{T}(\mathbf{g})\right)(\mathbf{v}) = g\,(T(\mathbf{v}))$, for all $\mathbf{v} \in \mathbb{V}$. Now, we note that $\widehat{T} \in \mathcal{L}(\mathbb{W}^*, \mathbb{V}^*)$, as for every $g, h \in \mathbb{W}^*$,

$$\left(\widehat{T}(\alpha\mathbf{g} + \mathbf{h})\right)(\mathbf{v}) = (\alpha\mathbf{g} + \mathbf{h})\,(T(\mathbf{v})) = \alpha\mathbf{g}\,(T(\mathbf{v})) + \mathbf{h}\,(T(\mathbf{v})) = \left(\alpha\widehat{T}(\mathbf{g}) + \widehat{T}(\mathbf{h})\right)(\mathbf{v}),$$

for all $\mathbf{v} \in \mathbb{V}$ implies that $\widehat{T}(\alpha\mathbf{g} + \mathbf{h}) = \alpha\widehat{T}(\mathbf{g}) + \widehat{T}(\mathbf{h})$.

Theorem 4.5.8. *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{B} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$, respectively. Also, let $\mathcal{A}^* = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ and $\mathcal{B}^* = (\mathbf{g}_1, \ldots, \mathbf{g}_m)$ be the corresponding ordered bases of the dual spaces $\mathbb{V}^*$ and $\mathbb{W}^*$, respectively. Then,*

$$\widehat{T}[\mathcal{B}^*, \mathcal{A}^*] = (T[\mathcal{A}, \mathcal{B}])^T,$$

the transpose of the coordinate matrix T.

Proof. Note that we need to compute $\widehat{T}[\mathcal{B}^*, \mathcal{A}^*] = \left[\left[\widehat{T}(\mathbf{g}_1)\right]_{\mathcal{A}^*}, \ldots, \left[\widehat{T}(\mathbf{g}_m)\right]_{\mathcal{A}^*}\right]$ and prove that it equals the transpose of the matrix $T[\mathcal{A}, \mathcal{B}]$. So, let

$$T[\mathcal{A}, \mathcal{B}] = [[T(\mathbf{v}_1)]_{\mathcal{B}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{B}}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Thus, to prove the required result, we need to show that

$$\left[\widehat{T}(\mathbf{g}_j)\right]_{\mathcal{A}^*} = [\mathbf{f}_1, \ldots, \mathbf{f}_n] \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{bmatrix} = \sum_{k=1}^{n} a_{jk}\mathbf{f}_k, \text{ for } 1 \le j \le m. \tag{4.5.1}$$

Now, recall that the functionals $\mathbf{f}_i$'s and $\mathbf{g}_j$'s satisfy $\left(\sum\limits_{k=1}^{n} \alpha_k \mathbf{f}_k\right)(\mathbf{v}_t) = \sum\limits_{k=1}^{n} \alpha_k \left(\mathbf{f}_k(\mathbf{v}_t)\right) = \alpha_t$, for $1 \le t \le n$ and $[\mathbf{g}_j(\mathbf{w}_1), \ldots, \mathbf{g}_j(\mathbf{w}_m)] = \mathbf{e}_j^T$, a row vector with 1 at the j-th place and 0, elsewhere. So, let $B = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ and evaluate $\widehat{T}(\mathbf{g}_j)$ at $\mathbf{v}_t$'s, the elements of $\mathcal{A}$.

$$\begin{aligned} \left(\widehat{T}(\mathbf{g}_j)\right)(\mathbf{v}_t) &= \mathbf{g}_j\left(T(\mathbf{v}_t)\right) = \mathbf{g}_j\left(B\left[T(\mathbf{v}_t)\right]_{\mathcal{B}}\right) = \left[\mathbf{g}_j(\mathbf{w}_1), \ldots, \mathbf{g}_j(\mathbf{w}_m)\right]\left[T(\mathbf{v}_t)\right]_{\mathcal{B}} \\ &= \mathbf{e}_j^T \begin{bmatrix} a_{1t} \\ a_{2t} \\ \vdots \\ a_{mt} \end{bmatrix} = a_{jt} = \left(\sum_{k=1}^{n} a_{jk}\mathbf{f}_k\right)(\mathbf{v}_t). \end{aligned}$$

Thus, the linear functional $\widehat{T}(\mathbf{g}_j)$ and $\sum\limits_{k=1}^{n} a_{jk}\mathbf{f}_k$ are equal at $\mathbf{v}_t$, for $1 \le t \le n$, the basis vectors of $\mathbb{V}$. Hence $\widehat{T}(\mathbf{g}_j) = \sum\limits_{k=1}^{n} a_{jk}\mathbf{f}_k$ which gives Equation (4.5.1). ■

Remark 4.5.9. *The proof of Theorem 4.5.8 also shows the following.*

1. *For each $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ there exists a unique map $\widehat{T} \in \mathcal{L}(\mathbb{W}^*, \mathbb{V}^*)$ such that*

$$\left(\widehat{T}(\mathbf{g})\right)(\mathbf{v}) = \mathbf{g}\left(T(\mathbf{v})\right), \text{ for each } \mathbf{g} \in \mathbb{W}^*.$$

2. *The coordinate matrices $T[\mathcal{A}, \mathcal{B}]$ and $\widehat{T}[\mathcal{B}^*, \mathcal{A}^*]$ are transpose of each other, where the ordered bases $\mathcal{A}^*$ of $\mathbb{V}^*$ and $\mathcal{B}^*$ of $\mathbb{W}^*$ correspond, respectively, to the ordered bases $\mathcal{A}$ of $\mathbb{V}$ and $\mathcal{B}$ of $\mathbb{W}$.*

3. *Thus, the results on matrices and its transpose can be re-written in the language a vector space and its dual space.*

4.6 Summary

Chapter 5

Inner Product Spaces

5.1 Definition and Basic Properties

Recall the dot product in $\mathbb{R}^2$ and $\mathbb{R}^3$. Dot product helped us to compute the length of vectors and angle between vectors. This enabled us to rephrase geometrical problems in $\mathbb{R}^2$ and $\mathbb{R}^3$ in the language of vectors. We generalize the idea of dot product to achieve similar goal for a general vector space over $\mathbb{R}$ or $\mathbb{C}$. So, in this chapter $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$.

Definition 5.1.1. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. An **inner product** over $\mathbb{V}$, denoted by $\langle\ ,\ \rangle$, is a map from $\mathbb{V} \times \mathbb{V}$ 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 $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$ 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 $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ and
3. $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ for all $\mathbf{u} \in \mathbb{V}$. Furthermore, equality holds if and only if $\mathbf{u} = \mathbf{0}$.

Remark 5.1.2. *Using the definition of inner product, we immediately observe that*

1. $\langle \mathbf{v}, \alpha\mathbf{w}\rangle = \overline{\langle \alpha\mathbf{w}, \mathbf{v}\rangle} = \overline{\alpha}\overline{\langle \mathbf{w}, \mathbf{v}\rangle} = \overline{\alpha}\langle \mathbf{v}, \mathbf{w}\rangle$, *for all* $\alpha \in \mathbb{F}$ *and* $\mathbf{v}, \mathbf{w} \in \mathbb{V}$.
2. *If* $\langle \mathbf{u}, \mathbf{v}\rangle = 0$ *for all* $\mathbf{v} \in \mathbb{V}$ *then in particular* $\langle \mathbf{u}, \mathbf{u}\rangle = 0$. *Hence,* $\mathbf{u} = \mathbf{0}$.

Definition 5.1.3. Let $\mathbb{V}$ be a vector space with an inner product $\langle\ ,\ \rangle$. Then, $(\mathbb{V}, \langle\ ,\ \rangle)$ is called an **inner product space** (in short, IPS).

Example 5.1.4. Examples 1 and 2 that appear below are called the **standard inner product** or the **dot product** on $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively. Whenever an inner product is not clearly mentioned, it will be assumed to be the standard inner product.

1. For $\mathbf{u} = (u_1, \ldots, u_n)^T, \mathbf{v} = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ define $\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + \cdots + u_nv_n = \mathbf{v}^T\mathbf{u}$. Then, $\langle\ ,\ \rangle$ is indeed an inner product and hence $\big(\mathbb{R}^n, \langle\ ,\ \rangle\big)$ is an IPS.
2. For $\mathbf{u} = (u_1, \ldots, u_n)^*, \mathbf{v} = (v_1, \ldots, v_n)^* \in \mathbb{C}^n$ define $\langle \mathbf{u}, \mathbf{v}\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n} = \mathbf{v}^*\mathbf{u}$. Then, $\big(\mathbb{C}^n, \langle\ ,\ \rangle\big)$ is an IPS.
3. For $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$ and $A = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$, define $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^TA\mathbf{x}$. Then, $\langle\ ,\ \rangle$ is an inner product as $\langle \mathbf{x}, \mathbf{x}\rangle = (x_1 - x_2)^2 + 3x_1^2 + x_2^2$.

4. Fix $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ with $a, c > 0$ and $ac > b^2$. Then, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T A\mathbf{x}$ is an inner product on $\mathbb{R}^2$ as $\langle \mathbf{x}, \mathbf{x} \rangle = ax_1^2 + 2bx_1x_2 + cx_2^2 = a\left[x_1 + \frac{bx_2}{a}\right]^2 + \frac{1}{a}\left[ac - b^2\right]x_2^2$.

5. Verify that for $\mathbf{x} = (x_1, x_2, x_3)^T, \mathbf{y} = (y_1, y_2, y_3)^T \in \mathbb{R}^3$, $\langle \mathbf{x}, \mathbf{y} \rangle = 10x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2 + x_2y_3 + x_3y_2 + x_3y_3$ defines an inner product.

6. For $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$, 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) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1^3 + x_2y_2^3$.

7. Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then, for $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, define $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* A\mathbf{x}$. Then, $\langle\ ,\ \rangle$ satisfies $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ and $\langle \mathbf{x} + \alpha\mathbf{z}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \alpha\langle \mathbf{z}, \mathbf{y} \rangle$, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$. Does there exist conditions on A such that $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x} \in \mathbb{C}$? This will be answered in affirmative in the chapter on eigenvalues and eigenvectors.

8. For $A, B \in M_n(\mathbb{R})$, define $\langle A, B \rangle = \mathsf{tr}(\mathsf{B}^\mathsf{T}\mathsf{A})$. Then,
$$\begin{aligned} \langle A + B, C \rangle &= \mathsf{tr}\big(\mathsf{C}^\mathsf{T}(\mathsf{A}+\mathsf{B})\big) = \mathsf{tr}(\mathsf{C}^\mathsf{T}\mathsf{A}) + \mathsf{tr}(\mathsf{C}^\mathsf{T}\mathsf{B}) = \langle \mathsf{A}, \mathsf{C} \rangle + \langle \mathsf{B}, \mathsf{C} \rangle \text{ and} \\ \langle A, B \rangle &= \mathsf{tr}(\mathsf{B}^\mathsf{T}\mathsf{A}) = \mathsf{tr}(\ (\mathsf{B}^\mathsf{T}\mathsf{A})^\mathsf{T}\ ) = \mathsf{tr}(\mathsf{A}^\mathsf{T}\mathsf{B}) = \langle \mathsf{B}, \mathsf{A} \rangle. \end{aligned}$$
If $A = [a_{ij}]$ then $\langle A, A \rangle = \mathsf{tr}(\mathsf{A}^\mathsf{T}\mathsf{A}) = \sum\limits_{\mathsf{i}=1}^{\mathsf{n}} (\mathsf{A}^\mathsf{T}\mathsf{A})_{\mathsf{ii}} = \sum\limits_{\mathsf{i,j}=1}^{\mathsf{n}} \mathsf{a_{ij}a_{ij}} = \sum\limits_{\mathsf{i,j}=1}^{\mathsf{n}} \mathsf{a_{ij}^2}$ and therefore, $\langle A, A \rangle > 0$ for all nonzero matrix A.

9. Consider the complex vector space $\mathcal{C}[-1,\ 1]$ and define $\langle f, g \rangle = \int\limits_{-1}^{1} f(x)\overline{g(x)}dx$. Then,

   (a) $\langle \mathbf{f}, \mathbf{f} \rangle = \int\limits_{-1}^{1} \mid \mathbf{f}(x) \mid^2 dx \geq 0$ as $\mid \mathbf{f}(x) \mid^2 \geq 0$ and this integral is 0 if and only if $\mathbf{f} \equiv \mathbf{0}$ as f is continuous.

   (b) $\overline{\langle \mathbf{g}, \mathbf{f} \rangle} = \overline{\int\limits_{-1}^{1} \mathbf{g}(x)\overline{\mathbf{f}(x)}dx} = \int\limits_{-1}^{1} \overline{\mathbf{g}(x)\overline{\mathbf{f}(x)}}dx = \int\limits_{-1}^{1} \mathbf{f}(x)\overline{\mathbf{g}(x)}dx = \langle \mathbf{f}, \mathbf{g} \rangle$.

   (c) $\langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle = \int\limits_{-1}^{1} (\mathbf{f} + \mathbf{g})(x)\overline{\mathbf{h}(x)}dx = \int\limits_{-1}^{1} [\mathbf{f}(x)\overline{\mathbf{h}(x)} + \mathbf{g}(x)\overline{\mathbf{h}(x)}]dx = \langle \mathbf{f}, \mathbf{h} \rangle + \langle \mathbf{g}, \mathbf{h} \rangle$.

   (d) $\langle \alpha\mathbf{f}, \mathbf{g} \rangle = \int\limits_{-1}^{1} (\alpha\mathbf{f}(x))\overline{\mathbf{g}(x)}dx = \alpha \int\limits_{-1}^{1} \mathbf{f}(x)\overline{\mathbf{g}(x)}dx = \alpha\langle \mathbf{f}, \mathbf{g} \rangle$.

   (e) Fix an ordered basis $\mathcal{B} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ of a complex vector space $\mathbb{V}$. Then, for any $\mathbf{u}, \mathbf{v} \in \mathbb{V}$, with $[\mathbf{u}]_\mathcal{B} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $[\mathbf{v}]_\mathcal{B} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$, define $\langle \mathbf{u}, \mathbf{v} \rangle = \sum\limits_{i=1}^{n} a_i\overline{b_i}$. Then, $\langle\ ,\ \rangle$ is indeed an inner product in $\mathbb{V}$. So, any finite dimensional vector space can be endowed with an inner product.

5.1.1 Cauchy Schwartz Inequality

As $\langle \mathbf{u}, \mathbf{u} \rangle > 0$, for all $\mathbf{u} \neq \mathbf{0}$, we use inner product to define length of a vector.

Definition 5.1.5. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then, for any vector $\mathbf{u} \in \mathbb{V}$, we define the **length (norm)** of $\mathbf{u}$, denoted $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$, the positive square root. A vector of norm 1 is called a **unit vector**. Thus, $\dfrac{\mathbf{u}}{\|\mathbf{u}\|}$ is called the **unit vector in the direction** of $\mathbf{u}$.

Example 5.1.6. 1. Let $\mathbb{V}$ be an IPS and $\mathbf{u} \in \mathbb{V}$. Then, for any scalar α, $\|\alpha \mathbf{u}\| = \big|\alpha\big| \cdot \|\mathbf{u}\|$.

2. Let $\mathbf{u} = (1, -1, 2, -3)^T \in \mathbb{R}^4$. 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. Moreover $\frac{1}{\sqrt{15}}\mathbf{u}$ is a unit vector in the direction of $\mathbf{u}$.

EXERCISE **5.1.7.** *1. Let* $\mathbf{u} = (-1, 1, 2, 3, 7)^T \in \mathbb{R}^5$. *Find all* $\alpha \in \mathbb{R}$ *such that* $\|\alpha \mathbf{u}\| = 1$.

2. *Let* $\mathbf{u} = (-1, 1, 2, 3, 7)^T \in \mathbb{C}^5$. *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* $\mathbf{x}^T, \mathbf{y}^T \in \mathbb{R}^n$. *This equality is called the* PARALLELOGRAM LAW *as in a parallelogram the sum of square of the lengths of the diagonals is equal to twice the sum of squares of the lengths of the sides.*

4. **Apollonius' Identity:** *Let the length of the sides of a triangle be* $a, b, c \in \mathbb{R}$ *and that of the median be* $d \in \mathbb{R}$. *If the median is drawn on the side with length* a *then prove that* $b^2 + c^2 = 2\left(d^2 + \left(\frac{a}{2}\right)^2\right)$.

5. *Let* $\mathbf{u} = (1, 2)^T, \mathbf{v} = (2, -1)^T \in \mathbb{R}^2$. *Then, does there exist an inner product in* $\mathbb{R}^2$ *such that* $\|\mathbf{u}\| = 1, \|\mathbf{v}\| = 1$ *and* $\langle \mathbf{u}, \mathbf{v} \rangle = 0$? *[Hint: Let* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *and define* $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T A \mathbf{x}$. *Use given conditions to get a linear system of* 3 *equations in the variables* a, b, c.*]*

6. *Let* $\mathbf{x} = (x_1, x_2)^T,\ \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$. *Then,* $\langle \mathbf{x}, \mathbf{y} \rangle = 3x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2$ *defines an inner product. Use this inner product to find*

   (a) the angle between $\mathbf{e}_1 = (1, 0)^T$ *and* $\mathbf{e}_2 = (0, 1)^T$.

   (b) $\mathbf{v} \in \mathbb{R}^2$ *such that* $\langle \mathbf{v}, \mathbf{e}_1 \rangle = 0$.

   (c) $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *such that* $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ *and* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

7. *Under the standard inner product in* $\mathbb{M}_{m,n}(\mathbb{R}), \mathbb{R}^m$ *and* $\mathbb{R}^n$, *prove that*

   (a) for $A \in \mathbb{M}_{m,n}(\mathbb{R})$, $\|A\|^2 = \mathrm{tr}(A^T A) = \sum\limits_{k=1}^{m} \|A[k,:]\|^2 = \sum\limits_{\ell=1}^{n} \|A[:,\ell]\|^2$.

   (b) for $A \in \mathbb{M}_{m,n}(\mathbb{R})$ *and* $\mathbf{x} \in \mathbb{R}^n$, $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$.

   Ans: $\|A\mathbf{x}\|^2 = \sum\limits_{k=1}^{m} |\,(A\mathbf{x})_k\,|^2 = \sum\limits_{k=1}^{m} |\,\big(A[k,:]^T \mathbf{x}\big)\,|^2 = \sum\limits_{k=1}^{n} |\langle \mathbf{x}, A[k,:] \rangle|^2$
   $\leq \sum\limits_{k=1}^{m} \|\mathbf{x}\|^2 \cdot \|A[k,:]\|^2 = \|\mathbf{x}\|^2 \sum\limits_{k=1}^{m} \|A[k,:]\|^2 = \|\mathbf{x}\|^2 \|A\|^2$.

A very useful and a fundamental inequality, commonly called the Cauchy-Schwartz inequality, concerning the inner product is proved next.

Theorem 5.1.8 (Cauchy-Bunyakovskii-Schwartz inequality)**.** *Let $\mathbb{V}$ be an inner product space over $\mathbb{F}$. Then, for any $\mathbf{u}, \mathbf{v} \in \mathbb{V}$*

$$|\ \langle \mathbf{u}, \mathbf{v} \rangle\ | \ \leq \|\mathbf{u}\|\ \|\mathbf{v}\|. \tag{5.1.1}$$

Moreover, equality holds in Inequality (5.1.1) if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. Furthermore, if $\mathbf{u} \neq \mathbf{0}$ then $\mathbf{v} = \left\langle \mathbf{v}, \dfrac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$.

Proof. If $\mathbf{u} = \mathbf{0}$ then Inequality (5.1.1) holds. Hence, let $\mathbf{u} \neq \mathbf{0}$. Then, by Definition 5.1.1.3, $\langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v} \rangle \geq 0$ for all $\lambda \in \mathbb{F}$ and $\mathbf{v} \in \mathbb{V}$. 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.

Now, note that equality holds in Inequality (5.1.1) if and only if $\langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v} \rangle = 0$, or equivalently, $\lambda\mathbf{u} + \mathbf{v} = \mathbf{0}$. Hence, $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. Moreover,

$$0 = \langle \mathbf{0}, \mathbf{u} \rangle = \langle \lambda\mathbf{u} + \mathbf{v}, \mathbf{u} \rangle = \lambda\langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle$$

implies that $\mathbf{v} = -\lambda\mathbf{u} = -\dfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}\mathbf{u} = \left\langle \mathbf{v}, \dfrac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$. ∎

Corollary 5.1.9. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then* $\left(\sum\limits_{i=1}^{n} \mathbf{x}_i\mathbf{y}_i\right)^2 \leq \left(\sum\limits_{i=1}^{n} \mathbf{x}_i^2\right)\left(\sum\limits_{i=1}^{n} \mathbf{y}_i^2\right)$.

5.1.2 Angle between two Vectors

Let $\mathbb{V}$ be a real vector space. Then, for $\mathbf{u}, \mathbf{v} \in \mathbb{V}$, the Cauchy-Schwartz inequality implies that $-1 \leq \frac{\langle \mathbf{u},\mathbf{v} \rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|} \leq 1$. We use this together with the properties of the cosine function to define the angle between two vectors in an inner product space.

Definition 5.1.10. Let $\mathbb{V}$ be a real vector space. If $\theta \in [0,\ \pi]$ is the angle between $\mathbf{u}, \mathbf{v} \in \mathbb{V}\backslash\{\mathbf{0}\}$ then we define

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|}.$$

Example 5.1.11.

1. Take $(1, 0)^T, (1, 1)^T \in \mathbb{R}^2$. Then, $\cos\theta = \frac{1}{\sqrt{2}}$. So $\theta = \pi/4$.
2. Take $(1, 1, 0)^T, (1, 1, 1)^T \in \mathbb{R}^3$. Then, angle between them, say $\beta = \cos^{-1}\frac{2}{\sqrt{6}}$.
3. Angle depends on the IP. Take $\langle \mathbf{x}, \mathbf{y} \rangle = 2\mathbf{x}_1\mathbf{y}_1 + \mathbf{x}_1\mathbf{y}_2 + \mathbf{x}_2\mathbf{y}_1 + \mathbf{x}_2\mathbf{y}_2$ on $\mathbb{R}^2$. Then, angle between $(1, 0)^T, (1, 1)^T \in \mathbb{R}^2$ equals $\cos^{-1}\frac{3}{\sqrt{10}}$.
4. As $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for any real vector space, the angle between $\mathbf{x}$ and $\mathbf{y}$ is same as the angle between $\mathbf{y}$ and $\mathbf{x}$.
5. Let $a, b \in \mathbb{R}$ with $a, b > 0$. Then, prove that $(a + b)\left(\dfrac{1}{a} + \dfrac{1}{b}\right) \geq 4$.
6. For $1 \leq i \leq n$, let $a_i \in \mathbb{R}$ with $a_i > 0$. Then, use Corollary 5.1.9 to show that $\left(\sum\limits_{i=1}^{n} a_i\right)\left(\sum\limits_{i=1}^{n} \dfrac{1}{a_i}\right) \geq n^2$.

7. Prove that $\mid z_1 + \cdots + z_n \mid \leq \sqrt{n(\mid z_1 \mid^2 + \cdots + \mid z_n \mid^2)}$, for $z_1, \ldots, z_n \in \mathbb{C}$. When does the equality hold?

8. Let $\mathbb{V}$ be an IPS. If $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ with $\|\mathbf{u}\| = 1, \|\mathbf{v}\| = 1$ and $\langle \mathbf{u}, \mathbf{v} \rangle = 1$ then prove that $\mathbf{u} = \alpha \mathbf{v}$ for some $\alpha \in \mathbb{F}$. Is $\alpha = 1$?

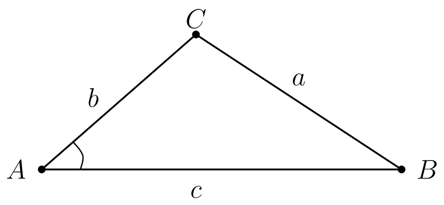


Figure 5.1: Triangle with vertices A, B and C

We will now prove that if A, B and C are the vertices of a triangle (see Figure 5.1) and a, b and c, respectively, are the lengths of the corresponding sides then $\cos(A) = \frac{b^2+c^2-a^2}{2bc}$. This in turn implies that the angle between vectors has been rightly defined.

Lemma 5.1.12. *Let A, B and C be the vertices of a triangle (see Figure 5.1) with corresponding side lengths a, b and c, respectively, in a real inner product space $\mathbb{V}$ then*

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc}.$$

Proof. Let $\mathbf{0}$, $\mathbf{u}$ and $\mathbf{v}$ be the coordinates of the vertices A, B and C, respectively, of the triangle ABC. Then, $\vec{AB} = \mathbf{u}, \vec{AC} = \mathbf{v}$ and $\vec{BC} = \mathbf{v} - \mathbf{u}$. Thus, we need to prove that

$$\cos(A) = \frac{\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2}{2\|\mathbf{v}\|\|\mathbf{u}\|} \Leftrightarrow \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2 = 2\ \|\mathbf{v}\|\ \|\mathbf{u}\|\ \cos(A).$$

Now, by definition $\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\langle \mathbf{v}, \mathbf{u} \rangle$ and hence $\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2 = 2\ \langle \mathbf{u}, \mathbf{v} \rangle$. As $\langle \mathbf{v}, \mathbf{u} \rangle = \|\mathbf{v}\|\ \|\mathbf{u}\|\ \cos(A)$, the required result follows. ■

Definition 5.1.13. Let $\mathbb{V}$ be an inner product space over $\mathbb{R}$. Then,

1. the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ are called **orthogonal/perpendicular** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

2. Let $S \subseteq \mathbb{V}$. Then, the **orthogonal complement** of S in $\mathbb{V}$, denoted $S^\perp$, equals

$$S^\perp = \{\mathbf{v} \in \mathbb{V} : \ \langle \mathbf{v}, \mathbf{w} \rangle = 0, \ \text{ for all } \ \mathbf{w} \in S\}.$$

Example 5.1.14. 1. $\mathbf{0}$ is orthogonal to every vector as $\langle \mathbf{0}, \mathbf{x} \rangle = 0$ for all $\mathbf{x} \in \mathbb{V}$.

2. If $\mathbb{V}$ is a vector space over $\mathbb{R}$ or $\mathbb{C}$ then $\mathbf{0}$ is the only vector that is orthogonal to itself.

3. Let $\mathbb{V} = \mathbb{R}$.

   (a) $S = \{0\}$. Then, $S^\perp = \mathbb{R}$.

   (b) $S = \mathbb{R}$, Then, $S^\perp = \{0\}$.

   (c) Let S be any subset of $\mathbb{R}$ containing a nonzero real number. Then, $S^\perp = \{0\}$.

4. Let $\mathbf{u} = (1,2)^T$. What is $\mathbf{u}^\perp$ in $\mathbb{R}^2$?

   Solution: $\{(x,y)^T \in \mathbb{R}^2 \mid x+2y=0\}$. Is this $\textsc{Null}(\mathbf{u})$? Note that $(2,-1)^T$ is a basis of $\mathbf{u}^\perp$ and for any vector $\mathbf{x} \in \mathbb{R}^2$,

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} + \left(\mathbf{x} - \langle \mathbf{x}, \mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2}\right) = \frac{x_1+2x_2}{5}(1,2)^T + \frac{2x_1-x_2}{5}(2,-1)^T$$

   is a decomposition of $\mathbf{x}$ into two vectors, one parallel to $\mathbf{u}$ and the other parallel to $\mathbf{u}^\perp$.

5. Fix $\mathbf{u} = (1,1,1,1)^T, \mathbf{v} = (1,1,-1,0)^T \in \mathbb{R}^4$. Determine $\mathbf{z}, \mathbf{w} \in \mathbb{R}^4$ such that $\mathbf{u} = \mathbf{z} + \mathbf{w}$ with the condition that $\mathbf{z}$ is parallel to $\mathbf{v}$ and $\mathbf{w}$ is orthogonal to $\mathbf{v}$.

   Solution: As $\mathbf{z}$ is parallel to $\mathbf{v}$, $\mathbf{z} = k\mathbf{v} = (k,k,-k,0)^T$, for some $k \in \mathbb{R}$. Since $\mathbf{w}$ is orthogonal to $\mathbf{v}$ the vector $\mathbf{w} = (a,b,c,d)^T$ satisfies $a+b-c=0$. Thus, $c = a+b$ and

$$(1,1,1,1)^T = \mathbf{u} = \mathbf{z} + \mathbf{w} = (k,k,-k,0)^T + (a,b,a+b,d)^T.$$

   Comparing the corresponding coordinates, gives the linear system $d=1$, $a+k=1$, $b+k=1$ and $a+b-k=1$ in the variables a,b,d and k. Thus, solving for a,b,d and k gives $\mathbf{z} = \frac{1}{3}(1,1,-1,0)^T$ and $\mathbf{w} = \frac{1}{3}(2,2,4,3)^T$.

6. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then prove that

   (a) $\langle \mathbf{x}, \mathbf{y}\rangle = 0 \Longleftrightarrow \|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ (Pythagoras Theorem).
   Solution: Use $\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y}\rangle$ to get the required result follows.

   (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).
   Solution: Use $\langle \mathbf{x}+\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle = \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2$ to get the required result follows.

   (c) $4\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2$ (polarization identity in $\mathbb{R}^n$).
   Solution: Just expand the right hand side to get the required result follows.

   (d) $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$ (parallelogram law: the sum of squares of the diagonals of a parallelogram equals twice the sum of squares of its sides).
   Solution: Just expand the left hand side to get the required result follows.

7. Let $P = (1,1,1)^T$, $Q = (2,1,3)^T$ and $R = (-1,1,2)^T$ be three vertices of a triangle in $\mathbb{R}^3$. Compute the angle between the sides PQ and PR.

   Solution: Method 1: Note that $\vec{PQ} = (2,1,3)^T - (1,1,1)^T = (1,0,2)^T$, $\vec{PR} = (-2,0,1)^T$ and $\vec{RQ} = (-3,0,-1)^T$. As $\langle \vec{PQ}, \vec{PR}\rangle = 0$, the angle between the sides PQ and PR is $\frac{\pi}{2}$.

   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 $\frac{\pi}{2}$.

Exercise **5.1.15.** *1. Let $\mathbb{V}$ be an IPS.*

(a) If $S \subseteq \mathbb{V}$ then $S^\perp$ is a subspace of $\mathbb{V}$ and $S^\perp = (LS(S))^\perp$.

(b) Furthermore, if $\mathbb{V}$ is finite dimensional then $S^\perp$ and $LS(S)$ are complementary. That is, $\mathbb{V} = LS(S) + S^\perp$. Equivalently, $\langle \mathbf{u}, \mathbf{w}\rangle = 0$, for all $\mathbf{u} \in LS(S)$ and $\mathbf{w} \in S^\perp$.

2. *Consider* $\mathbb{R}^3$ *with the standard inner product. Find*
   (a) $S^\perp$ *for* $S = \{(1,1,1)^T, (0,1,-1)^T\}$ *and* $S = LS((1,1,1)^T, (0,1,-1)^T)$.
   (b) *vectors* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ *such that* $\mathbf{v}, \mathbf{w}, \mathbf{u} = (1,1,1)^T$ *are mutually orthogonal.*
   (c) *the line passing through* $(1,1,-1)^T$ *and parallel to* $(a,b,c) \neq \mathbf{0}$.
   (d) *the plane containing* $(1,1-1)$ *with* $(a,b,c) \neq \mathbf{0}$ *as the normal vector.*
   (e) *the area of the parallelogram with three vertices* $\mathbf{0}^T$, $(1,2,-2)^T$ *and* $(2,3,0)^T$.
   (f) *the area of the parallelogram when* $\|\mathbf{x}\| = 5, \|\mathbf{x}-\mathbf{y}\| = 8$ *and* $\|\mathbf{x}+\mathbf{y}\| = 14$.
   (g) *the plane containing* $(2,-2,1)^T$ *and perpendicular to the line with parametric equation* $x = t-1, y = 3t+2, z = t+1$.
   (h) *the plane containing the lines* $(1,2,-2) + t(1,1,0)$ *and* $(1,2,-2) + t(0,1,2)$.
   (i) k *such that* $\cos^{-1}(\langle \mathbf{u}, \mathbf{v} \rangle) = \pi/3$, *where* $\mathbf{u} = (1,-1,1)^T$ *and* $\mathbf{v} = (1,k,1)^T$.
   (j) *the plane containing* $(1,1,2)^T$ *and orthogonal to the line with parametric equation* $x = 2+t, y = 3$ *and* $z = 1-t$.
   (k) *a parametric equation of a line containing* $(1,-2,1)^T$ *and orthogonal to* $x+3y+2z = 1$.

3. *Let* $P = (3,0,2)^T, Q = (1,2,-1)^T$ *and* $R = (2,-1,1)^T$ *be three points in* $\mathbb{R}^3$. *Then,*
   (a) *find the area of the triangle with vertices* P, Q *and* R.
   (b) *find the area of the parallelogram built on vectors* $\vec{PQ}$ *and* $\vec{QR}$.
   (c) *find a nonzero vector orthogonal to the plane of the above triangle.*
   (d) *find all vectors* $\mathbf{x}$ *orthogonal to* $\vec{PQ}$ *and* $\vec{QR}$ *with* $\|\mathbf{x}\| = \sqrt{2}$.
   (e) *the volume of the parallelepiped built on vectors* $\vec{PQ}$ *and* $\vec{QR}$ *and* $\mathbf{x}$, *where* $\mathbf{x}$ *is one of the vectors found in Part 3d. Do you think the volume would be different if you choose the other vector* $\mathbf{x}$*?*

4. *Let* p_1 *be a plane containing* $A = (1,2,3)^T$ *and* $(2,-1,1)^T$ *as its normal vector. Then,*
   (a) *find the equation of the plane* p_2 *that is parallel to* p_1 *and contains* $(-1,2,-3)^T$.
   (b) *calculate the distance between the planes* p_1 *and* p_2.

5. *In the parallelogram* $ABCD$, $AB \| DC$ *and* $AD \| BC$ *and* $A = (-2,1,3)^T, B = (-1,2,2)^T$ *and* $C = (-3,1,5)^T$. *Find the*
   (a) *coordinates of the point* D,
   (b) *cosine of the angle* BCD.
   (c) *area of the triangle* ABC
   (d) *volume of the parallelepiped determined by* AB, AD *and* $(0,0,-7)^T$.

6. *Let* $\mathbb{W} = \{(x,y,z,w)^T \in \mathbb{R}^4 : x+y+z-w = 0\}$. *Find a basis of* $\mathbb{W}^\perp$.

7. *Recall the* IPS $\mathbb{M}_n(\mathbb{R})$ *(see Example 5.1.4.8). If* $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A^T = A\}$ *then* $\mathbb{W}^\perp$*?*

5.1.3 Normed Linear Space

To proceed further, recall that a vector space over $\mathbb{R}$ or $\mathbb{C}$ was a linear space.

Definition 5.1.16. Let $\mathbb{V}$ be a linear space.

1. Then, a **norm** on $\mathbb{V}$ is a function $f(\mathbf{x}) = \|\mathbf{x}\|$ from $\mathbb{V}$ to $\mathbb{R}$ such that
   (a) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in \mathbb{V}$ and if $\|\mathbf{x}\| = 0$ then $\mathbf{x} = \mathbf{0}$.
   (b) $\|\alpha\mathbf{x}\| = \mid \alpha \mid \ \|\mathbf{x}\|$ for all $\alpha \in \mathbb{F}$ and $\mathbf{x} \in \mathbb{V}$.
   (c) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ (triangle inequality).
2. A linear space with a norm on it is called a **normed linear space** (NLS).

Theorem 5.1.17. *Let $\mathbb{V}$ be a normed linear space and $\mathbf{x}, \mathbf{y} \in \mathbb{V}$. Then, $\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \leq \|\mathbf{x}-\mathbf{y}\|$.*

Proof. As $\|\mathbf{x}\| = \|\mathbf{x}-\mathbf{y}+\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}\|$ one has $\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}\|$. Similarly, one obtains $\|\mathbf{y}\| - \|\mathbf{x}\| \leq \|\mathbf{y}-\mathbf{x}\| = \|\mathbf{x}-\mathbf{y}\|$. Combining the two, the required result follows. ∎

Example 5.1.18.
1. On $\mathbb{R}^3$, $\|\mathbf{x}\| = \sqrt{\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2}$ is a norm. Also, observe that this norm corresponds to $\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$, where $\langle , \rangle$ is the standard inner product.
2. Let $\mathbb{V}$ be an IPS. Is it true that $f(\mathbf{x}) = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ is a norm?
   Solution: Yes. The readers should verify the first two conditions. For the third condition, recalling the Cauchy-Schwartz inequality, we get
   $$\begin{aligned} f(\mathbf{x}+\mathbf{y})^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{x}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (f(\mathbf{x})+f(\mathbf{y}))^2. \end{aligned}$$
   Thus, $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ is a norm, called the norm **induced** by the inner product $\langle \cdot, \cdot \rangle$.

EXERCISE **5.1.19.**
1. *Let $\mathbb{V}$ be an IPS. Then,*
   $$4\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 + i\|\mathbf{x}+i\mathbf{y}\|^2 - i\|\mathbf{x}-i\mathbf{y}\|^2 \quad (\textbf{Polarization Identity}).$$
2. *Consider the complex vector space $\mathbb{C}^n$. If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ then prove that*
   (a) If $\mathbf{x} \neq \mathbf{0}$ then $\|\mathbf{x}+i\mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{x}\|^2$, even though $\langle \mathbf{x}, i\mathbf{x}\rangle \neq 0$.
   (b) $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ whenever $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ and $\|\mathbf{x}+i\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{y}\|^2$.
3. *Let $A \in \mathbb{M}_n(\mathbb{C})$ satisfy $\|A\mathbf{x}\| \leq \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$. Then, prove that if $\alpha \in \mathbb{C}$ with $\mid \alpha \mid > 1$ then $A - \alpha I$ is invertible.*

The next result is stated without proof as the proof is beyond the scope of this book.

Theorem 5.1.20. *Let $\|\cdot\|$ be a norm on a NLS $\mathbb{V}$. Then, $\|\cdot\|$ is induced by some inner product if and only if $\|\cdot\|$ satisfies the PARALLELOGRAM LAW: $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$.*

Example 5.1.21.
1. For $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$, we define $\|\mathbf{x}\|_1 = |\mathbf{x}_1| + |\mathbf{x}_2|$. Verify that $\|\mathbf{x}\|_1$ is indeed a norm. But, for $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = \mathbf{e}_2$, $2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) = 4$ whereas
   $$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = \|(1,1)\|^2 + \|(1,-1)\|^2 = (|1| + |1|)^2 + (|1| + |-1|)^2 = 8.$$
   So, the parallelogram law fails. Thus, $\|\mathbf{x}\|_1$ is not induced by any inner product in $\mathbb{R}^2$.

2. Does there exist an inner product in $\mathbb{R}^2$ such that $\|\mathbf{x}\| = \max\{|x_1|, |x_2|\}$?
3. If $\|\cdot\|$ is a norm in $\mathbb{V}$ then $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, for $\mathbf{x}, \mathbf{y} \in \mathbb{V}$, defines a distance function as
   (a) $d(\mathbf{x}, \mathbf{x}) = 0$, for each $\mathbf{x} \in \mathbb{V}$.
   (b) using the triangle inequality, for any $\mathbf{z} \in \mathbb{V}$, we have
   $$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \|(\mathbf{x} - \mathbf{z}) - (\mathbf{y} - \mathbf{z})\| \le \|(\mathbf{x} - \mathbf{z})\| + \|(\mathbf{y} - \mathbf{z})\| = d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}).$$

5.2 Gram-Schmidt Orthonormalization Process

We start with the definition of an orthonormal set.

Definition 5.2.1. Let $\mathbb{V}$ be an IPS. Then, a non-empty set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq \mathbb{V}$ is called an **orthogonal set** if $\mathbf{v}_i$ and $\mathbf{v}_j$ are **mutually orthogonal**, for $1 \le i \ne j \le n$, *i.e.*,
$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0, \text{ for } 1 \le i < j \le n.$$

Further, if $\|\mathbf{v}_i\| = 1$, for $1 \le i \le n$, Then S is called an **orthonormal set**. If S is also a basis of $\mathbb{V}$ then S is called an **orthonormal basis** of $\mathbb{V}$.

Example 5.2.2. 1. A few orthonormal sets in $\mathbb{R}^2$ are
$$\{(1,0)^T, (0,1)^T\}, \{\frac{1}{\sqrt{2}}(1,1)^T, \frac{1}{\sqrt{2}}(1,-1)^T\} \text{ and } \{\frac{1}{\sqrt{5}}(2,1)^T, \frac{1}{\sqrt{5}}(1,-2)^T\}.$$

2. Let $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$. Then, S is an orthonormal set as
   (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. The set $\left\{\left[\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]^T, \left[0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T, \left[\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right]^T\right\}$ is an orthonormal set in $\mathbb{R}^3$.
4. Recall that $\langle f(x), g(x) \rangle = \int\limits_{-\pi}^{\pi} f(x)\overline{g(x)}dx$ defines the standard inner product in $\mathcal{C}[-\pi, \pi]$. Consider $S = \{\mathbf{1}\} \cup \{\mathbf{e}_m \mid m \ge 1\} \cup \{\mathbf{f}_n \mid n \ge 1\}$, where $\mathbf{1}(x) = 1$, $\mathbf{e}_m(x) = \cos(mx)$ and $\mathbf{f}_n(x) = \sin(nx)$, for all $m, n \ge 1$ and for all $x \in [-\pi, \pi]$. Then,
   (a) S is a linearly independent set.
   (b) $\|\mathbf{1}\|^2 = 2\pi$, $\|\mathbf{e}_m\|^2 = \pi$ and $\|\mathbf{f}_n\|^2 = \pi$.
   (c) the functions in S are orthogonal.

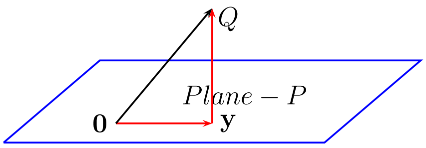
   Hence, $\left\{\frac{\mathbf{1}}{\sqrt{2\pi}}\right\} \cup \left\{\frac{1}{\sqrt{\pi}}\mathbf{e}_m \mid m \ge 1\right\} \cup \left\{\frac{1}{\sqrt{\pi}}\mathbf{f}_n \mid n \ge 1\right\}$ is an orthonormal set in $\mathcal{C}[-\pi, \pi]$.

To proceed further, we consider a few examples for better understanding.

Example 5.2.3. Which point on the plane P is closest to the point, say Q?

Solution: Let $\mathbf{y}$ be the foot of the perpendicular from Q on P. Thus, by Pythagoras Theorem, this point is unique. So, the question arises: how do we find $\mathbf{y}$?

Note that $\overrightarrow{\mathbf{y}Q}$ gives a normal vector of the plane P. Hence, $\overrightarrow{Q} = \mathbf{y} + \overrightarrow{\mathbf{y}Q}$. So, need to decompose $\overrightarrow{Q}$ into two vectors such that one of them lies on the plane P and the other is orthogonal to the plane.



Thus, we see that given $\mathbf{u}, \mathbf{v} \in \mathbb{V} \setminus \{\mathbf{0}\}$, we need to find two vectors, say $\mathbf{y}$ and $\mathbf{z}$, such that $\mathbf{y}$ is parallel to $\mathbf{u}$ and $\mathbf{z}$ is perpendicular to $\mathbf{u}$. Thus, $\mathbf{y} = \mathbf{u}\cos(\theta)$ and $\mathbf{z} = \mathbf{u}\sin(\theta)$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

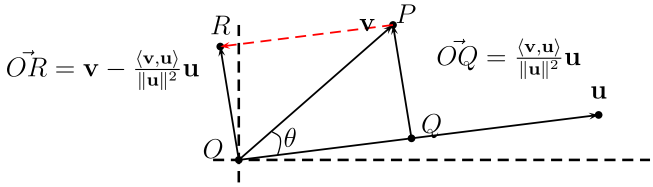


Figure 5.2: Decomposition of vector $\mathbf{v}$

We do this as follows (see Figure 5.2). Let $\hat{\mathbf{u}} = \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$ be the unit vector in the direction of $\mathbf{u}$. Then, using trigonometry, $\cos(\theta) = \frac{\|\vec{OQ}\|}{\|\vec{OP}\|}$. Hence $\|\vec{OQ}\| = \|\vec{OP}\| \cos(\theta)$. Now using Definition 5.1.10, $\|\vec{OQ}\| = \|\mathbf{v}\| \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{v}\|\,\|\mathbf{u}\|}\right| = \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{u}\|}\right|$, where the absolute value is taken as the length/norm is a positive quantity. Thus,

$$\vec{OQ} = \|\vec{OQ}\|\,\hat{\mathbf{u}} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Hence, $\mathbf{y} = \vec{OQ} = \left\langle \mathbf{v}, \dfrac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$ and $\mathbf{z} = \mathbf{v} - \left\langle \mathbf{v}, \dfrac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \dfrac{\mathbf{u}}{\|\mathbf{u}\|}$. In literature, the vector $\mathbf{y} = \vec{OQ}$ is called the **orthogonal projection** of $\mathbf{v}$ on $\mathbf{u}$, denoted $\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})$. Thus,

$$\mathrm{Proj}_{\mathbf{u}}(\mathbf{v}) = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|} \quad \text{and} \quad \|\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})\| = \|\vec{OQ}\| = \left|\frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\|\mathbf{u}\|}\right|. \tag{5.2.1}$$

Moreover, the distance of $\mathbf{u}$ from the point P equals $\|\vec{OR}\| = \|\vec{PQ}\| = \left\|\mathbf{v} - \langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\|$.

Example 5.2.4.

1. Determine the foot of the perpendicular from the point $(1, 2, 3)$ on the XY-plane.

   Solution: Verify that the required point is $(1, 2, 0)$?

2. Determine the foot of the perpendicular from the point $Q = (1, 2, 3, 4)$ on the plane generated by $(1, 1, 0, 0), (1, 0, 1, 0)$ and $(0, 1, 1, 1)$.

   Answer: (x, y, z, w) lies on the plane $x-y-z+2w = 0 \Leftrightarrow \langle (1, -1, -1, 2), (x, y, z, w)\rangle = 0$.

   So, the required point equals

$$\begin{aligned} &(1, 2, 3, 4) - \langle (1, 2, 3, 4), \frac{1}{\sqrt{7}}(1, -1, -1, 2)\rangle \frac{1}{\sqrt{7}}(1, -1, -1, 2) \\ =\ &(1, 2, 3, 4) - \frac{4}{7}(1, -1, -1, 2) = \frac{1}{7}(3, 18, 25, 20). \end{aligned}$$

3. Determine the projection of $\mathbf{v} = (1,1,1,1)^T$ on $\mathbf{u} = (1,1,-1,0)^T$.
   Solution: By Equation (5.2.1), we have $\text{Proj}_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle \dfrac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{3}(1,1,-1,0)^T$ and $\mathbf{w} = (1,1,1,1)^T - \text{Proj}_{\mathbf{v}}(\mathbf{u}) = \frac{1}{3}(2,2,4,3)^T$ is orthogonal to $\mathbf{u}$.

4. Let $\mathbf{u} = (1,1,1,1)^T, \mathbf{v} = (1,1,-1,0)^T, \mathbf{w} = (1,1,0,-1)^T \in \mathbb{R}^4$. 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) $\mathbf{v}_1 = \text{Proj}_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^T$ is parallel to $\mathbf{u}$.
   (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 $\text{Proj}_{\mathbf{u}}(\mathbf{w})$ is parallel to $\mathbf{u}$ and $\text{Proj}_{\mathbf{v}_2}(\mathbf{w})$ is parallel to $\mathbf{v}_2$. Hence, we have

   (a) $\mathbf{w}_1 = \text{Proj}_{\mathbf{u}}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{u} \rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^T$ is parallel to $\mathbf{u}$,
   (b) $\mathbf{w}_2 = \text{Proj}_{\mathbf{v}_2}(\mathbf{w}) = \langle \mathbf{w}, \mathbf{v}_2 \rangle \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|^2} = \frac{7}{44}(3,3,-5,-1)^T$ 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 $\mathbf{v}_2$.

We now prove the most important initial result of this section.

Theorem 5.2.5. *Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal subset of an* IPS $\mathbb{V}(\mathbb{F})$.

1. *Then, S is a linearly independent subset of $\mathbb{V}$.*
2. *Suppose $\mathbf{v} \in LS(S)$ with $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i$, for some α_i's in $\mathbb{F}$. Then,*
   (a) $\alpha_i = \langle \mathbf{v}, \mathbf{u}_i \rangle$.
   (b) $\|\mathbf{v}\|^2 = \|\sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 = \sum\limits_{i=1}^{n} \mid \alpha_i \mid^2$.
3. *Let $\mathbf{z} \in \mathbb{V}$ and $\mathbf{w} = \sum\limits_{i=1}^{n} \langle \mathbf{z}, \mathbf{u}_i \rangle \mathbf{u}_i$. Then, $\mathbf{z} = \mathbf{w} + (\mathbf{z} - \mathbf{w})$ with $\langle \mathbf{z} - \mathbf{w}, \mathbf{w} \rangle = 0$, i.e., $\mathbf{z} - \mathbf{w} \in LS(S)^{\perp}$. Further, $\|\mathbf{z}\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{z} - \mathbf{w}\|^2 \geq \|\mathbf{w}\|^2$.*
4. *Let $\dim(\mathbb{V}) = n$. 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. Part 1: Consider the linear system $c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \mathbf{0}$ in the variables $c_1, \ldots, c_n$. As $\langle \mathbf{0}, \mathbf{u} \rangle = 0$ and $\langle \mathbf{u}_j, \mathbf{u}_i \rangle = 0$, for all $j \neq i$, we have

$$0 = \langle \mathbf{0}, \mathbf{u}_i \rangle = \langle c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_i \rangle = \sum_{j=1}^{n} c_j \langle \mathbf{u}_j, \mathbf{u}_i \rangle = c_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle = c_i.$$

Hence, $c_i = 0$, for $1 \leq i \leq n$. Thus, the above linear system has only the trivial solution. So, the set S is linearly independent.

Part 2: Note that $\langle \mathbf{v}, \mathbf{u}_i \rangle = \langle \sum\limits_{j=1}^{n} \alpha_j \mathbf{u}_j, \mathbf{u}_i \rangle = \sum_{j=1}^{n} \alpha_j \langle \mathbf{u}_j, \mathbf{u}_i \rangle = \alpha_i \langle \mathbf{u}_i, \mathbf{u}_i \rangle = \alpha_i$. This completes the first sub-part. For the second sub-part, we have

$$\begin{aligned} \|\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 &= \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \sum_{i=1}^{n} \alpha_i \mathbf{u}_i \right\rangle = \sum_{i=1}^{n} \alpha_i \left\langle \mathbf{u}_i, \sum_{j=1}^{n} \alpha_j \mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \overline{\alpha_j} \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i} \langle \mathbf{u}_i, \mathbf{u}_i \rangle = \sum_{i=1}^{n} \mid \alpha_i \mid^2. \end{aligned}$$

Part 3: Note that for $1 \le i \le n$,

$$\begin{aligned}\langle \mathbf{z}-\mathbf{w}, \mathbf{u}_i\rangle &= \langle \mathbf{z}, \mathbf{u}_i\rangle - \langle \mathbf{w}, \mathbf{u}_i\rangle = \langle \mathbf{z}, \mathbf{u}_i\rangle - \left\langle \sum_{j=1}^{n} \langle \mathbf{z}, \mathbf{u}_j\rangle \mathbf{u}_j, \mathbf{u}_i \right\rangle \\ &= \langle \mathbf{z}, \mathbf{u}_i\rangle - \sum_{j=1}^{n} \langle \mathbf{z}, \mathbf{u}_j\rangle \langle \mathbf{u}_j, \mathbf{u}_i\rangle = \langle \mathbf{z}, \mathbf{u}_i\rangle - \langle \mathbf{z}, \mathbf{u}_i\rangle = 0.\end{aligned}$$

So, $\mathbf{z}-\mathbf{w} \in \mathrm{LS}(S)^{\perp}$. Hence, $\langle \mathbf{z}-\mathbf{w}, \mathbf{w}\rangle = 0$ as $\mathbf{w} \in \mathrm{LS}(S)$. Further,

$$\|\mathbf{z}\|^2 = \|\mathbf{w} + (\mathbf{z}-\mathbf{w})\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{z}-\mathbf{w}\|^2 \ge \|\mathbf{w}\|^2.$$

Part 4: Follows directly using Part 2b as $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of $\mathbb{V}$. ∎

A rephrasing of Theorem 5.2.5.2b gives a generalization of the pythagoras theorem, popularly known as the Parseval's formula. The proof is left as an exercise for the reader.

Theorem 5.2.6. *Let $\mathbb{V}$ be a finite dimensional* IPS *with an orthonormal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. Then, for each $\mathbf{x}, \mathbf{y} \in \mathbb{V}$,*

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i\rangle \overline{\langle \mathbf{y}, \mathbf{v}_i\rangle}.$$

Furthermore, if $\mathbf{x} = \mathbf{y}$ then $\|\mathbf{x}\|^2 = \sum\limits_{i=1}^{n} |\langle \mathbf{x}, \mathbf{v}_i\rangle|^2$ (generalizing the **Pythagoras Theorem***).*

As a corollary to Theorem 5.2.5, we have the following result.

Theorem 5.2.7 (Bessel's Inequality)**.** *Let $\mathbb{V}$ be an* IPS *with $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ as an orthogonal set. Then, $\sum\limits_{k=1}^{n} \dfrac{|\langle \mathbf{z}, \mathbf{v}_k\rangle|^2}{\|\mathbf{v}_k\|^2} \le \|\mathbf{z}\|^2$, for each $\mathbf{z} \in \mathbb{V}$. Equality holds if and only if $\mathbf{z} = \sum\limits_{k=1}^{n} \dfrac{\langle \mathbf{z}, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|^2}\mathbf{v}_k$.*

Proof. For $1 \le k \le n$, define $\mathbf{u}_k = \dfrac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$ and use Theorem 5.2.5.4 to get the required result. ∎

Remark 5.2.8. *Using Theorem 5.2.5, we see that if $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is an ordered orthonormal basis of an* IPS *$\mathbb{V}$ then for each $\mathbf{u} \in \mathbb{V}$, $[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} \langle \mathbf{u}, \mathbf{v}_1\rangle \\ \vdots \\ \langle \mathbf{u}, \mathbf{v}_n\rangle \end{bmatrix}$. Thus, in place of solving a linear system to get the coordinates of a vector, we just need to compute the inner product with basis vectors.*

EXERCISE **5.2.9.**

1. *Find $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ such that $\mathbf{v}, \mathbf{w}, (1, -1, -2)^T$ are mutually orthogonal.*

2. *Let $\mathcal{B} = \left[\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}\right]$ be an ordered basis of $\mathbb{R}^2$. Then, $\left[\begin{bmatrix}x\\y\end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix}\frac{x+y}{\sqrt{2}}\\ \frac{x-y}{\sqrt{2}}\end{bmatrix}$.*

3. *For the ordered basis $\mathcal{B} = \left[\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}\right]$ of $\mathbb{R}^3$, $[(2,3,1)^T]_{\mathcal{B}} = \begin{bmatrix}2\sqrt{3}\\ \frac{-1}{\sqrt{2}}\\ \frac{3}{\sqrt{6}}\end{bmatrix}$.*

In view of the importance of Theorem 5.2.5, we inquire into the question of extracting an orthonormal basis from a given basis. The process of extracting an orthonormal basis from a finite linearly independent set is called the **Gram-Schmidt Orthonormalization process**. We first consider a few examples. Note that Theorem 5.2.5 also gives us an algorithm for doing so, *i.e.*, from the given vector subtract all the orthogonal projections/components. If the new vector is nonzero then this vector is orthogonal to the previous ones. The proof follows directly from Theorem 5.2.5 but we give it again for the sake of completeness.

Theorem 5.2.10 (Gram-Schmidt Orthogonalization Process)**.** *Let $\mathbb{V}$ be an* IPS*. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of linearly independent vectors in $\mathbb{V}$ then there exists an orthonormal set $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ in $\mathbb{V}$. Furthermore, $LS(\mathbf{w}_1, \ldots, \mathbf{w}_i) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$, for $1 \le i \le n$.*

Proof. Note that for orthonormality, we need $\|\mathbf{w}_i\| = 1$, for $1 \le i \le n$ and $\langle \mathbf{w}_i, \mathbf{w}_j\rangle = 0$, for $1 \le i \ne j \le n$. Also, by Corollary 3.2.8.2, $\mathbf{v}_i \notin LS(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$, for $2 \le i \le n$, as $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a linearly independent set. We are now ready to prove the result by induction.

Step 1: Define $\mathbf{w}_1 = \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ then $LS(\mathbf{v}_1) = LS(\mathbf{w}_1)$.

Step 2: Define $\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{w}_1\rangle \mathbf{w}_1$. Then, $\mathbf{u}_2 \ne \mathbf{0}$ as $\mathbf{v}_2 \notin LS(\mathbf{v}_1)$. So, let $\mathbf{w}_2 = \dfrac{\mathbf{u}_2}{\|\mathbf{u}_2\|}$.

Note that $\{\mathbf{w}_1, \mathbf{w}_2\}$ is orthonormal and $LS(\mathbf{w}_1, \mathbf{w}_2) = LS(\mathbf{v}_1, \mathbf{v}_2)$.

Step 3: For induction, assume that we have obtained an orthonormal set $\{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}$ such that $LS(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}) = LS(\mathbf{w}_1, \ldots, \mathbf{w}_{k-1})$. Now, note that $\mathbf{u}_k = \mathbf{v}_k - \sum\limits_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i\rangle \mathbf{w}_i = \mathbf{v}_k - \sum\limits_{i=1}^{k-1} \text{Proj}_{\mathbf{w}_i}(\mathbf{v}_k) \ne \mathbf{0}$ as $\mathbf{v}_k \notin LS(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$. So, let us put $\mathbf{w}_k = \dfrac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$. Then, $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is orthonormal as $\|\mathbf{w}_k\| = 1$ and

$$\begin{aligned}
\|\mathbf{u}_k\|\langle \mathbf{w}_k, \mathbf{w}_1\rangle &= \langle \mathbf{u}_k, \mathbf{w}_1\rangle = \langle \mathbf{v}_k - \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i\rangle \mathbf{w}_i, \mathbf{w}_1\rangle = \langle \mathbf{v}_k, \mathbf{w}_1\rangle - \langle \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i\rangle \mathbf{w}_i,\ \mathbf{w}_1\rangle \\
&= \langle \mathbf{v}_k, \mathbf{w}_1\rangle - \sum_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i\rangle \langle \mathbf{w}_i,\ \mathbf{w}_1\rangle = \langle \mathbf{v}_k, \mathbf{w}_1\rangle - \langle \mathbf{v}_k, \mathbf{w}_1\rangle = \mathbf{0}.
\end{aligned}$$

Similarly, $\langle \mathbf{w}_k, \mathbf{w}_i\rangle = 0$, for $2 \le i \le k-1$. Clearly, $\mathbf{w}_k = \mathbf{u}_k/\|\mathbf{u}_k\| \in LS(\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}, \mathbf{v}_k)$. So, $\mathbf{w}_k \in LS(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

As $\mathbf{v}_k = \|\mathbf{u}_k\|\mathbf{w}_k + \sum\limits_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i\rangle \mathbf{w}_i$, we get $\mathbf{v}_k \in LS(\mathbf{w}_1, \ldots, \mathbf{w}_k)$. Hence, by the principle of mathematical induction $LS(\mathbf{w}_1, \ldots, \mathbf{w}_k) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ and the required result follows. ∎

We now illustrate the Gram-Schmidt process with a few examples.

Example 5.2.11. 1. Let $S = \{(1, -1, 1, 1), (1, 0, 1, 0), (0, 1, 0, 1)\} \subseteq \mathbb{R}^4$. Find an orthonormal set T such that $LS(S) = LS(T)$.

Solution: Let $\mathbf{v}_1 = (1, 0, 1, 0)^T, \mathbf{v}_2 = (0, 1, 0, 1)^T$ and $\mathbf{v}_3 = (1, -1, 1, 1)^T$. Then, $\mathbf{w}_1 = \frac{1}{\sqrt{2}}(1, 0, 1, 0)^T$. As $\langle \mathbf{v}_2, \mathbf{w}_1\rangle = 0$, we get $\mathbf{w}_2 = \frac{1}{\sqrt{2}}(0, 1, 0, 1)^T$. For the third vector, let $\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{w}_1\rangle \mathbf{w}_1 - \langle \mathbf{v}_3, \mathbf{w}_2\rangle \mathbf{w}_2 = (0, -1, 0, 1)^T$. Thus, $\mathbf{w}_3 = \frac{1}{\sqrt{2}}(0, -1, 0, 1)^T$.

2. Let $S = \left\{\mathbf{v}_1 = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}^T, \mathbf{v}_2 = \begin{bmatrix} \frac{3}{2} & 2 & 0 \end{bmatrix}^T, \mathbf{v}_3 = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} & 0 \end{bmatrix}^T, \mathbf{v}_4 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T\right\}$. Find an orthonormal set T such that $LS(S) = LS(T)$.

Solution: Take $\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^T = \mathbf{e}_1$. For the second vector, consider $\mathbf{u}_2 = \mathbf{v}_2 - \frac{3}{2}\mathbf{w}_1 = \begin{bmatrix}0 & 2 & 0\end{bmatrix}^T$. So, put $\mathbf{w}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \begin{bmatrix}0 & 1 & 0\end{bmatrix}^T = \mathbf{e}_2$.

For the third vector, let $\mathbf{u}_3 = \mathbf{v}_3 - \sum\limits_{i=1}^{2}\langle \mathbf{v}_3, \mathbf{w}_i\rangle \mathbf{w}_i = (0,0,0)^T$. So, $\mathbf{v}_3 \in LS((\mathbf{w}_1, \mathbf{w}_2))$. Or equivalently, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

So, for again computing the third vector, define $\mathbf{u}_4 = \mathbf{v}_4 - \sum\limits_{i=1}^{2}\langle \mathbf{v}_4, \mathbf{w}_i\rangle \mathbf{w}_i$. Then, $\mathbf{u}_4 = \mathbf{v}_4 - \mathbf{w}_1 - \mathbf{w}_2 = \mathbf{e}_3$. So $\mathbf{w}_4 = \mathbf{e}_3$. Hence, $T = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4\} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

3. Find an orthonormal set in $\mathbb{R}^3$ containing $(1,2,1)^T$.
   Solution: Let $(x,y,z)^T \in \mathbb{R}^3$ with $\big\langle (1,2,1), (x,y,z)\big\rangle = 0$. Thus,

$$(x,y,z) = (-2y - z, y, z) = y(-2,1,0) + z(-1,0,1).$$

   Observe that $(-2,1,0)$ and $(-1,0,1)$ are orthogonal to $(1,2,1)$ but are themselves not orthogonal.

   METHOD 1: Apply Gram-Schmidt process to $\{\frac{1}{\sqrt{6}}(1,2,1)^T, (-2,1,0)^T, (-1,0,1)^T\} \subseteq \mathbb{R}^3$.

   METHOD 2: Valid only in $\mathbb{R}^3$ using the cross product of two vectors.

   In either case, verify that $\{\frac{1}{\sqrt{6}}(1,2,1), \frac{-1}{\sqrt{5}}(2,-1,0), \frac{-1}{\sqrt{30}}(1,2,-5)\}$ is the required set.

We now state two immediate corollaries without proof.

Corollary 5.2.12. *Let $\mathbb{V} \neq \{\mathbf{0}\}$ be an* IPS. *If*

1. *$\mathbb{V}$ is finite dimensional then $\mathbb{V}$ has an orthonormal basis.*
2. *S is a non-empty orthonormal set and* $\dim(\mathbb{V})$ *is finite then S can be extended to form an orthonormal basis of $\mathbb{V}$.*

Remark 5.2.13. *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \neq \{\mathbf{0}\}$ be a non-empty subset of a finite dimensional vector space $\mathbb{V}$. If we apply Gram-Schmidt process to*

1. *S then we obtain an orthonormal basis of $LS(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.*
2. *a re-arrangement of elements of S then we may obtain another orthonormal basis of $LS(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. But, observe that the size of the two bases will be the same.*

EXERCISE **5.2.14.**

1. *Let $\mathbb{V}$ be an* IPS *with $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ as a basis. Then, prove that $\mathcal{B}$ is orthonormal if and only if for each $x \in \mathbb{V}$, $x = \sum\limits_{i=1}^{n}\langle \mathbf{x}, \mathbf{v}_i\rangle \mathbf{v}_i$. [Hint: Since $\mathcal{B}$ is a basis, each $\mathbf{x} \in \mathbb{V}$ has a unique linear combination in terms of $\mathbf{v}_i$'s.]*
2. *Let S be a subset of $\mathbb{V}$ having* 101 *elements. Suppose that the application of the Gram-Schmidt process yields $\mathbf{u}_5 = \mathbf{0}$. Does it imply that $LS(\mathbf{v}_1, \ldots, \mathbf{v}_5) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_4)$? Give reasons for your answer.*
3. *Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal set in $\mathbb{R}^n$. For $1 \le k \le n$, define $A_k = \sum\limits_{i=1}^{k}\mathbf{v}_i\mathbf{v}_i^T$. Then, prove that $A_k^T = A_k$ and $A_k^2 = A_k$. Thus, A_k's are projection matrices.*
4. *Determine an orthonormal basis of $\mathbb{R}^4$ containing $(1,-2,1,3)^T$ and $(2,1,-3,1)^T$.*

5. *Let* $\mathbf{x} \in \mathbb{R}^n$ *with* $\|\mathbf{x}\| = 1$.

   (a) *Then, prove that* $\{\mathbf{x}\}$ *can be extended to form an orthonormal basis of* $\mathbb{R}^n$.

   (b) *Let the extended basis be* $\{\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ *and* $\mathcal{B} = [\mathbf{e}_1, \ldots, \mathbf{e}_n]$ *the standard ordered basis of* $\mathbb{R}^n$. *Prove that* $A = \begin{bmatrix} [\mathbf{x}]_{\mathcal{B}}, & [\mathbf{x}_2]_{\mathcal{B}}, & \ldots, & [\mathbf{x}_n]_{\mathcal{B}} \end{bmatrix}$ *is an orthogonal matrix.*

6. *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* $\det(A) = 1$.

7. *Let* $(\mathbb{V}, \langle\ ,\ \rangle)$ *be an* n*-dimensional* IPS. *If* $\mathbf{u} \in \mathbb{V}$ *with* $\|\mathbf{u}\| = 1$ *then give reasons for the following statements.*

   (a) *Let* $S^{\perp} = \{\mathbf{v} \in \mathbb{V} \mid \langle \mathbf{v}, \mathbf{u} \rangle = 0\}$. *Then,* $\dim(S^{\perp}) = n - 1$.

   (b) *Let* $0 \neq \beta \in \mathbb{F}$. *Then* $S = \{\mathbf{v} \in \mathbb{V} : \langle \mathbf{v}, \mathbf{u} \rangle = \beta\}$ *is not a subspace of* $\mathbb{V}$.

   (c) *Let* $\mathbf{v} \in \mathbb{V}$. *Then* $\mathbf{v} = \mathbf{v}_0 + \langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u}$ *for a vector* $\mathbf{v}_0 \in S^{\perp}$. *Thus* $\mathbb{V} = LS(\mathbf{u}, S^{\perp})$.

5.2.1 QR Decomposition*

The next result gives the proof of the QR decomposition for real matrices. The readers are advised to prove similar results for matrices with complex entries. This decomposition and its generalizations are helpful in the numerical calculations related with eigenvalue problems (see Chapter 6).

Theorem 5.2.1 (QR Decomposition). *Let* $A \in \mathbb{M}_n(\mathbb{R})$ *be invertible. Then, there exist matrices* Q *and* R *such that* Q *is orthogonal and* R *is upper triangular with* $A = QR$. *Furthermore, if* $\det(A) \neq 0$ *then the diagonal entries of* R *can be chosen to be positive. Also, in this case, the decomposition is unique.*

Proof. As A is invertible, it's columns form a basis of $\mathbb{R}^n$. So, an application of the Gram-Schmidt orthonormalization process to $\{A[:,1], \ldots, A[:,n]\}$ gives an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ satisfying

$$LS(A[:,1], \ldots, A[:,i]) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i), \quad \text{for } 1 \leq i \leq n.$$

Since $A[:,i] \in LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$, for $1 \leq i \leq n$, there exist $\alpha_{ji} \in \mathbb{R}, 1 \leq j \leq i$, such that $A[:,i] = [\mathbf{v}_1, \ldots, \mathbf{v}_i] \begin{bmatrix} \alpha_{1i} \\ \vdots \\ \alpha_{ii} \end{bmatrix}$. Thus, if $Q = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ 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

1. Q is an orthogonal matrix (see Exercise 5.4.8.5),
2. R is an upper triangular matrix, and
3. $A = QR$.

Thus, this completes the proof of the first part. Note that

1. $\alpha_{ii} \neq 0$, for $1 \leq i \leq n$, as $A[:,1] \neq \mathbf{0}$ and $A[:,i] \notin LS(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$.

2. if $\alpha_{ii} < 0$, for some $i, 1 \le i \le n$ then we can replace $\mathbf{v}_i$ in Q by $-\mathbf{v}_i$ to get a new Q ad R in which the diagonal entries of R are positive.

Uniqueness: Suppose $Q_1R_1 = Q_2R_2$ for some orthogonal matrices Q_i's and upper triangular matrices R_i's with positive diagonal entries. As Q_i's and R_i's are invertible, we get $Q_2^{-1}Q_1 = R_2R_1^{-1}$. Now, using

1. Exercises 2.5.25.1, 1.2.14.1, the matrix $R_2R_1^{-1}$ is an upper triangular matrix.
2. Exercises 1.3.2.3, $Q_2^{-1}Q_1$ is an orthogonal matrix.

So, the matrix $R_2R_1^{-1}$ is an orthogonal upper triangular matrix and hence, by Exercise 1.2.10.3, $R_2R_1^{-1} = I_n$. So, $R_2 = R_1$ and therefore $Q_2 = Q_1$. ■

Let A be an $n \times k$ matrix with $\text{Rank}(A) = r$. Then, by Remark 5.2.13, an application of the Gram-Schmidt orthonormalization process to columns of A yields an orthonormal set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subseteq \mathbb{R}^n$ such that

$$LS(A[:,1], \ldots, A[:,j]) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i), \text{ for } 1 \le i \le j \le k.$$

Hence, proceeding on the lines of the above theorem, we have the following result.

Theorem 5.2.2 (Generalized QR Decomposition). *Let A be an $n \times k$ matrix of rank r. Then, $A = QR$, where*

1. $Q = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ *is an $n \times r$ matrix with $Q^TQ = I_r$,*
2. $LS(A[:,1], \ldots, A[:,j]) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$*, for $1 \le i \le j \le k$ and*
3. *R is an $r \times k$ matrix with $\text{Rank}(R) = r$.*

Example 5.2.3. 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 5.2.11, we know that $\mathbf{w}_1 = \frac{1}{\sqrt{2}}(1,0,1,0)^T$, $\mathbf{w}_2 = \frac{1}{\sqrt{2}}(0,1,0,1)^T$ and $\mathbf{w}_3 = \frac{1}{\sqrt{2}}(0,-1,0,1)^T$. We now compute $\mathbf{w}_4$. If $\mathbf{v}_4 = (2,1,1,1)^T$ then

$$\mathbf{u}_4 = \mathbf{v}_4 - \langle \mathbf{v}_4, \mathbf{w}_1\rangle \mathbf{w}_1 - \langle \mathbf{v}_4, \mathbf{w}_2\rangle \mathbf{w}_2 - \langle \mathbf{v}_4, \mathbf{w}_3\rangle \mathbf{w}_3 = \frac{1}{2}(1,0,-1,0)^T.$$

Thus, $\mathbf{w}_4 = \frac{1}{\sqrt{2}}(-1,0,1,0)^T$. Hence, we see that $A = QR$ with

$$Q = \left[\mathbf{w}_1, \ldots, \mathbf{w}_4\right] = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \text{ and } R = \begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & -\frac{3}{\sqrt{2}} \\ 0 & \sqrt{2} & 0 & -\sqrt{2} \\ 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

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 $Q^TQ = I_3$ and an upper triangular matrix R such that $A = QR$.

   Solution: Let us apply the Gram-Schmidt orthonormalization process to the columns of A. As $\mathbf{v}_1 = (1, -1, 1, 1)^T$, we get $\mathbf{w}_1 = \frac{1}{2}\mathbf{v}_1$. Let $\mathbf{v}_2 = (1, 0, 1, 0)^T$. Then,

$$\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{w}_1 \rangle \mathbf{w}_1 = (1, 0, 1, 0)^T - \mathbf{w}_1 = \frac{1}{2}(1, 1, 1, -1)^T.$$

   Hence, $\mathbf{w}_2 = \frac{1}{2}(1, 1, 1, -1)^T$. Let $\mathbf{v}_3 = (1, -2, 1, 2)^T$. Then,

$$\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{v}_3, \mathbf{w}_2 \rangle \mathbf{w}_2 = \mathbf{v}_3 - 3\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{0}.$$

   So, we again take $\mathbf{v}_3 = (0, 1, 0, 1)^T$. Then,

$$\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{v}_3, \mathbf{w}_2 \rangle \mathbf{w}_2 = \mathbf{v}_3 - 0\mathbf{w}_1 - 0\mathbf{w}_2 = \mathbf{v}_3.$$

   So, $\mathbf{w}_3 = \frac{1}{\sqrt{2}}(0, 1, 0, 1)^T$. Hence,

$$Q = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{-1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{-1}{2} & \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } R = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & \sqrt{2} \end{bmatrix}.$$

   The readers are advised to check the following:

   (a) Rank$(A) = 3$,

   (b) $A = QR$ with $Q^TQ = I_3$, and

   (c) R is a 3×4 upper triangular matrix with Rank$(R) = 3$.

Remark 5.2.4. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$.*

1. *If $A = QR$ with $Q = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ then* $R = \begin{bmatrix} \langle \mathbf{v}_1, A[:,1] \rangle & \langle \mathbf{v}_1, A[:,2] \rangle & \langle \mathbf{v}_1, A[:,3] \rangle & \cdots \\ 0 & \langle \mathbf{v}_2, A[:,2] \rangle & \langle \mathbf{v}_2, A[:,3] \rangle & \cdots \\ 0 & 0 & \langle \mathbf{v}_3, A[:,3] \rangle & \cdots \\ \vdots & \vdots & & \ddots \end{bmatrix}$.

   In case Rank$(A) < n$ then a slight modification gives the matrix R.

2. *Further, let $m = n$ and Rank$(A) = n$.*

   (a) *Then, A^TA is invertible (see Exercise 3.4.17.4).*

   (b) *By Theorem 5.2.2, $A = QR$ with Q a matrix of size $m \times n$ and R an upper triangular matrix of size $n \times n$. Also, $Q^TQ = I_n$ and Rank$(R) = n$.*

   (c) *Thus, $A^TA = R^TQ^TQR = R^TR$. As A^TA is invertible, the matrix R^TR is invertible. Since R is a square matrix, by Exercise 2.5.5.1, the matrix R itself is invertible. Hence, $(R^TR)^{-1} = R^{-1}(R^T)^{-1}$.*

(d) *So, if* $Q = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ *then*

$$A(A^TA)^{-1}A^T = QR(R^TR)^{-1}R^TQ^T = (QR)(R^{-1}(R^T)^{-1})R^TQ^T = QQ^T.$$

(e) *Hence, using Theorem 5.3.7, we see that the matrix*

$$P = A(A^TA)^{-1}A^T = QQ^T = [\mathbf{v}_1, \ldots, \mathbf{v}_n]\begin{bmatrix}\mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T\end{bmatrix} = \sum_{i=1}^{n} \mathbf{v}_i\mathbf{v}_i^T$$

is the orthogonal projection matrix on $\text{Col}(A)$.

3. *Further, let* $Rank(A) = r < n$. *If* $j_1, \ldots, j_r$ *are the pivot columns of* A *then* $\text{Col}(A) = \text{Col}(B)$, *where* $B = [A[:, j_1], \ldots, A[:, j_r]]$ *is an* $m \times r$ *matrix with* $Rank(B) = r$. *So, using Part 2e we see that* $B(B^TB)^{-1}B^T$ *is the orthogonal projection matrix on* $\text{Col}(A)$. *So, compute RREF of* A *and choose columns of* A *corresponding to the pivot columns.*

5.3 Orthogonal Projections and Applications

Till now, our main interest was to understand the linear system $A\mathbf{x} = \mathbf{b}$, for $A \in \mathbb{M}_{m,n}(\mathbb{C}), \mathbf{x} \in \mathbb{C}^n$ and $\mathbf{b} \in \mathbb{C}^m$, from different view points. But, in most practical situations the system has no solution. So, we are interested in finding a point $\mathbf{x}_0 \in \mathbb{R}^n$ such that the **err** $= \mathbf{b} - A\mathbf{x}_0$ is the least. Thus, we consider the problem of finding $\mathbf{x}_0 \in \mathbb{R}^n$ such that

$$\|\mathbf{err}\| = \|\mathbf{b} - A\mathbf{x}_0\| = \min\{\|\mathbf{b} - A\mathbf{x}\| : \mathbf{x} \in \mathbb{R}^n\},$$

i.e., we try to find the vector $\mathbf{x}_0 \in \mathbb{R}^n$ which is nearest to $\mathbf{b}$.

To begin with, recall the following result from Page 135.

Theorem 5.3.1 (Decomposition)**.** *Let* $\mathbb{V}$ *be an* IPS *having* $\mathbb{W}$ *as a finite dimensional subspace. Suppose* $\{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$ *is an orthonormal basis of* $\mathbb{W}$. *Then, for each* $\mathbf{b} \in \mathbb{V}$, $\mathbf{y} = \sum\limits_{i=1}^{k} \langle \mathbf{b}, \mathbf{f}_i \rangle \mathbf{f}_i$ *is the closest point in* $\mathbb{W}$ *from* $\mathbf{b}$. *Furthermore,* $\mathbf{b} - \mathbf{y} \in \mathbb{W}^\perp$.

We now give a definition and then an implication of Theorem 5.3.1.

Definition 5.3.2. Let $\mathbb{W}$ be a finite dimensional subspace of an IPS $\mathbb{V}$. Then, by Theorem 5.3.1, for each $\mathbf{v} \in \mathbb{V}$ there exist unique vectors $\mathbf{w} \in \mathbb{W}$ and $\mathbf{u} \in \mathbb{W}^\perp$ with $\mathbf{v} = \mathbf{w} + \mathbf{u}$. We thus define the **orthogonal projection** of $\mathbb{V}$ onto $\mathbb{W}$, denoted $P_{\mathbb{W}}$, by

$$P_{\mathbb{W}} : \mathbb{V} \to \mathbb{W} \text{ by } P_{\mathbb{W}}(\mathbf{v}) = \mathbf{w}.$$

The vector $\mathbf{w}$ is called the **projection** of $\mathbf{v}$ on $\mathbb{W}$.

Remark 5.3.3. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$ *and* $\mathbb{W} = \text{Col}(A)$. *Then, to find the orthogonal projection* $P_{\mathbb{W}}(\mathbf{b})$, *we can use either of the following ideas:*

1. *Determine an orthonormal basis* $\{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$ *of* $\text{Col}(A)$ *and get* $P_{\mathbb{W}}(\mathbf{b}) = \sum\limits_{i=1}^{k} \langle \mathbf{b}, \mathbf{f}_i \rangle \mathbf{f}_i$.

2. *By Theorem 3.4.13.2,* $\text{COL}(A) = \text{NULL}(A^T)^\perp$. *Hence, for* $\mathbf{b} \in \mathbb{R}^m$ *there exists unique* $\mathbf{u} \in \text{COL}(A)$ *and* $\mathbf{v} \in \text{NULL}(A^T)$ *such that* $\mathbf{b} = \mathbf{u} + \mathbf{v}$. *Thus, using Definition 5.3.2 and Theorem 5.3.1,* $P_{\mathbb{W}}(\mathbf{b}) = \mathbf{u}$.

Before proceeding to projections, we give an application of Theorem 5.3.1 to a linear system.

Corollary 5.3.4. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$ *and* $\mathbf{b} \in \mathbb{R}^m$. *Then, every least square solution of* $A\mathbf{x} = \mathbf{b}$ *is a solution of the system* $A^TA\mathbf{x} = A^T\mathbf{b}$. *Conversely, every solution of* $A^TA\mathbf{x} = A^T\mathbf{b}$ *is a least square solution of* $A\mathbf{x} = \mathbf{b}$.

Proof. As $\mathbf{b} \in \mathbb{R}^m$, by Remark 5.3.3, there exists $\mathbf{y} \in \text{COL}(A)$ and $\mathbf{v} \in \text{NULL}(A^T)$ such that $\mathbf{b} = \mathbf{y} + \mathbf{v}$ and $\min\{\|\mathbf{b} - \mathbf{w}\| \mid \mathbf{w} \in \text{COL}(A)\} = \|\mathbf{b} - \mathbf{y}\|$. As $\mathbf{y} \in \text{COL}(A)$, there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $A\mathbf{x}_0 = \mathbf{y}$, *i.e.*, $\mathbf{x}_0$ is the least square solution of $A\mathbf{x} = \mathbf{b}$. Hence,

$$(A^TA)\mathbf{x}_0 = A^T(A\mathbf{x}_0) = A^T\mathbf{y} = A^T(\mathbf{b} - \mathbf{v}) = A^T\mathbf{b} - \mathbf{0} = A^T\mathbf{b}.$$

Conversely, let $\mathbf{x}_1 \in \mathbb{R}^n$ be a solution of $A^TA\mathbf{x} = A^T\mathbf{b}$, *i.e.*, $A^T(A\mathbf{x}_1 - \mathbf{b}) = \mathbf{0}$. To show

$$\min\{\|\mathbf{b} - A\mathbf{x}\| \mid \mathbf{x} \in \mathbb{R}^n\} = \|\mathbf{b} - A\mathbf{x}_1\|.$$

Note that $A^T(A\mathbf{x}_1 - \mathbf{b}) = \mathbf{0}$ implies

$$0 = (\mathbf{x} - \mathbf{x}_1)^TA^T(A\mathbf{x}_1 - \mathbf{b}) = (A\mathbf{x} - A\mathbf{x}_1)^T(A\mathbf{x}_1 - \mathbf{b}) = \langle A\mathbf{x}_1 - \mathbf{b}, A\mathbf{x} - A\mathbf{x}_1\rangle.$$

Thus, the vectors $\mathbf{b} - A\mathbf{x}_1$ and $A\mathbf{x}_1 - A\mathbf{x}$ are orthogonal and hence

$$\|\mathbf{b} - A\mathbf{x}\|^2 = \|\mathbf{b} - A\mathbf{x}_1 + A\mathbf{x}_1 - A\mathbf{x}\|^2 = \|\mathbf{b} - A\mathbf{x}_1\|^2 + \|A\mathbf{x}_1 - A\mathbf{x}\|^2 \geq \|\mathbf{b} - A\mathbf{x}_1\|^2.$$

Hence, the required result follows. ■

The above corollary gives the following result.

Corollary 5.3.5. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$ *and* $\mathbf{b} \in \mathbb{R}^m$. *If*

1. A^TA *is invertible then the least square solution of* $A\mathbf{x} = \mathbf{b}$ *equals* $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$.
2. A^TA *is not invertible then the least square solution of* $A\mathbf{x} = \mathbf{b}$ *equals* $\mathbf{x} = (A^TA)^-A^T\mathbf{b}$, *where* $(A^TA)^-$ *is the pseudo-inverse of* A^TA.

Proof. Part 1 directly follows from Corollary 5.3.5. For Part 2, let $\mathbf{b} = \mathbf{y} + \mathbf{v}$, for $\mathbf{y} \in \text{COL}(A)$ and $\mathbf{v} \in \text{NULL}(A^T)$. As $\mathbf{y} \in \text{COL}(A)$, there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $A\mathbf{x}_0 = \mathbf{y}$. Thus, by Remark 5.3.3, $A^T\mathbf{b} = A^T(\mathbf{y} + \mathbf{v}) = A^T\mathbf{y} = A^TA\mathbf{x}_0$. Now, using the definition of pseudo-inverse (see Exercise 1.3.7.13), we see that

$$(A^AA)\left((A^TA)^-A^T\mathbf{b}\right) = (A^TA)(A^TA)^-(A^TA)\mathbf{x}_0 = (A^TA)\mathbf{x}_0 = A^T\mathbf{b}.$$

Thus, we see that $(A^TA)^-A^T\mathbf{b}$ is a solution of the system $A^TA\mathbf{x} = A^T\mathbf{b}$. Hence, by Corollary 5.3.4, the required result follows. ■

We now give a few examples to understand projections.

Example 5.3.6. Use the fundamental theorem of linear algebra to compute the vector of the orthogonal projection.

1. Determine the projection of $(1,1,1,1,1)^T$ on $\textsc{Null}\left([1,-1,1,-1,1]\right)$.
   Solution: Here $A=[1,-1,1,-1,1]$. So, a basis of $\textsc{Col}(A^T)$ equals $\{(1,-1,1,-1,1)^T\}$ and that of $\textsc{Null}(A)$ equals $\{(1,1,0,0,0)^T,(1,0,-1,0,0)^T,(1,0,0,1,0)^T,(1,0,0,0,-1)^T\}$. Note that $\textsc{Null}(A)$ and $\textsc{Col}(A^T)$ are orthogonal and hence $(1,1,1,1,1)^T=\mathbf{y}+\mathbf{z}$, where $\mathbf{y}\in\textsc{Null}(A)$ and $\mathbf{z}\in\textsc{Col}(A^T)=LS([1,-1,1,-1,1]^T)$.
   So, taking $B=\begin{bmatrix}1&1&1&1&1\\1&0&0&0&-1\\0&-1&0&0&1\\0&0&1&0&-1\\0&0&0&-1&1\end{bmatrix}$ and solving $B\mathbf{x}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$ gives $\mathbf{x}=\frac{1}{5}\begin{bmatrix}6\\-4\\6\\-4\\1\end{bmatrix}$.
   Thus, $\mathbf{z}=\mathbf{x}_5(1,-1,1,-1,1)^T=\frac{1}{5}(1,-1,1,-1,1)^T$ and the projection vector $\mathbf{y}$ equals $(1,1,1,1,1)^T-\mathbf{z}=\frac{1}{5}(4,6,4,6,4)^T$, which is also equal to $\begin{bmatrix}1&1&1&1\\1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix}\begin{bmatrix}6/5\\-4/5\\6/5\\-4/5\end{bmatrix}$.

2. Determine the projection of $(1,1,1)^T$ on $\textsc{Null}\left([1,1,-1]\right)$.
   Solution: Here $A=[1,1,-1]$. So, a basis of $\textsc{Null}(A)$ equals $\{(1,-1,0)^T,(1,0,1)^T\}$ and that of $\textsc{Col}(A^T)$ equals $\{(1,1,-1)^T\}$. Then, the solution of the linear system $B\mathbf{x}=\begin{bmatrix}1\\1\\1\end{bmatrix}$, where $B=\begin{bmatrix}1&1&1\\-1&0&1\\0&1&-1\end{bmatrix}$ equals $\mathbf{x}=\frac{1}{3}\begin{bmatrix}-2\\4\\1\end{bmatrix}$. Thus, the projection is $\frac{1}{3}\left((-2)(1,-1,0)^T+4(1,0,1)^T\right)=\frac{2}{3}(1,1,2)^T$.

3. Determine the projection of $(1,1,1)^T$ on $\textsc{Col}\left([1,2,1]^T\right)$.
   Solution: Here, $A^T=[1,2,1]$, a basis of $\textsc{Col}(A)$ equals $\{(1,2,1)^T\}$ and that of $\textsc{Null}(A^T)$ equals $\{(1,0,-1)^T,(2,-1,0)^T\}$. Then, using the solution of the linear system $B\mathbf{x}=\begin{bmatrix}1\\1\\1\end{bmatrix}$, where $B=\begin{bmatrix}1&2&1\\0&-1&2\\-1&0&1\end{bmatrix}$ gives $\frac{2}{3}(1,2,1)^T$ as the required vector.

To use the first idea in Remark 5.3.3, we prove the following result which helps us to get the matrix of the orthogonal projection from an orthonormal basis.

Theorem 5.3.7. *Let $\{\mathbf{f}_1,\ldots,\mathbf{f}_k\}$ be an orthonormal basis of a finite dimensional subspace $\mathbb{W}$ of an* IPS *$\mathbb{V}$. Then $P_{\mathbb{W}}=\sum\limits_{i=1}^{k}\mathbf{f}_i\mathbf{f}_i^*$.*

Proof. Let $\mathbf{v}\in\mathbb{V}$. Then,

$$P_{\mathbb{W}}\mathbf{v}=\left(\sum_{i=1}^{k}\mathbf{f}_i\mathbf{f}_i^*\right)\mathbf{v}=\sum_{i=1}^{k}\mathbf{f}_i\left(\mathbf{f}_i^*\mathbf{v}\right)=\sum_{i=1}^{k}\langle\mathbf{v},\mathbf{f}_i\rangle\mathbf{f}_i.$$

As $P_{\mathbb{W}}\mathbf{v}$ is the only closet point (see Theorem 5.3.1), the required result follows. ∎

Example 5.3.8. In each of the following, determine the matrix of the orthogonal projection. Also, verify that $P_{\mathbb{W}} + P_{\mathbb{W}^\perp} = I$. What can you say about $\text{Rank}(P_{\mathbb{W}^\perp})$ and $\text{Rank}(P_{\mathbb{W}})$? Also, verify the orthogonal projection vectors obtained in Example 5.3.6.

1. $\mathbb{W} = \{(x_1, \ldots, x_5)^T \in \mathbb{R}^5 \mid x_1 - x_2 + x_3 - x_4 + x_5 = 0\} = \text{NULL}\,([1,-1,1,-1,1])$.

   Solution: An orthonormal basis of $\mathbb{W}$ is $\left\{ \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0\\0\\1\\1\\0\end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-1\\0\\0\\-2\end{bmatrix}, \frac{1}{\sqrt{30}}\begin{bmatrix}-2\\2\\3\\-3\\-2\end{bmatrix} \right\}$. Thus,

$$P_{\mathbb{W}} = \sum_{i=1}^{4} \mathbf{f}_i\mathbf{f}_i^T = \frac{1}{5}\begin{bmatrix} 4 & 1 & -1 & 1 & -1\\ 1 & 4 & 1 & -1 & 1\\ -1 & 1 & 4 & 1 & -1\\ 1 & -1 & 1 & 4 & 1\\ -1 & 1 & -1 & 1 & 4\end{bmatrix} \text{ and } P_{\mathbb{W}^\perp} = \frac{1}{5}\begin{bmatrix} 1 & -1 & 1 & -1 & 1\\ -1 & 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1 & 1\\ -1 & 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1 & 1\end{bmatrix}.$$

2. $\mathbb{W} = \{(x,y,z)^T \in \mathbb{R}^3 \mid x + y - z = 0\} = \text{NULL}\,([1,1,-1])$.

   Solution: Note $\{(1,1,-1)\}$ is a basis of $\mathbb{W}^\perp$ and $\left\{\frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,2)\right\}$ an orthonormal basis of $\mathbb{W}$. So,

$$P_{\mathbb{W}^\perp} = \frac{1}{3}\begin{bmatrix} 1 & 1 & -1\\ 1 & 1 & -1\\ -1 & -1 & 1\end{bmatrix} \text{ and } P_{\mathbb{W}} = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1\\ -1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix}.$$

   Verify that $P_{\mathbb{W}} + P_{\mathbb{W}^\perp} = I_3$, $\text{Rank}(P_{\mathbb{W}^\perp}) = 2$ and $\text{Rank}(P_{\mathbb{W}}) = 1$.

3. $\mathbb{W} = LS(\ (1,2,1)\ ) = \text{COL}\left([1,2,1]^T\right) \subseteq \mathbb{R}^3$.

   Solution: Using Example 5.2.11.3 and Equation (5.2.1)

$$\mathbb{W}^\perp = LS(\{(-2,1,0),(-1,0,1)\}) = LS(\{(-2,1,0),(1,2,-5)\}).$$

   So, $P_{\mathbb{W}} = \frac{1}{6}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}$ and $P_{\mathbb{W}^\perp} = \frac{1}{6}\begin{bmatrix}5 & -2 & -1\\ -2 & 2 & -2\\ -1 & -2 & 5\end{bmatrix}$.

We advise the readers to give a proof of the next result.

Theorem 5.3.9. *Let $\{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$ be an orthonormal basis of a subspace $\mathbb{W}$ of $\mathbb{R}^n$. If $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ is an extended orthonormal basis of $\mathbb{R}^n$, $P_{\mathbb{W}} = \sum\limits_{i=1}^{k} \mathbf{f}_i\mathbf{f}_i^T$ and $P_{\mathbb{W}^\perp} = \sum\limits_{i=k+1}^{n} \mathbf{f}_i\mathbf{f}_i^T$ then prove that*

1. *$I_n - P_{\mathbb{W}} = P_{\mathbb{W}^\perp}$.*
2. *$(P_{\mathbb{W}})^T = P_{\mathbb{W}}$ and $(P_{\mathbb{W}^\perp})^T = P_{\mathbb{W}^\perp}$. That is, $P_{\mathbb{W}}$ and $P_{\mathbb{W}^\perp}$ are symmetric.*
3. *$(P_{\mathbb{W}})^2 = P_{\mathbb{W}}$ and $(P_{\mathbb{W}^\perp})^2 = P_{\mathbb{W}^\perp}$. That is, $P_{\mathbb{W}}$ and $P_{\mathbb{W}^\perp}$ are idempotent.*
4. *$P_{\mathbb{W}} \circ P_{\mathbb{W}^\perp} = P_{\mathbb{W}^\perp} \circ P_{\mathbb{W}} = \mathbf{0}$.*

EXERCISE **5.3.10.** 1. *Let $\mathbb{W} = \{(x,y,z,w) \in \mathbb{R}^4 : x = y, z = w\}$ be a subspace of $\mathbb{R}^4$. Determine the matrix of the orthogonal projection.*

2. *Let $P_{\mathbb{W}_1}$ and $P_{\mathbb{W}_2}$ be the orthogonal projections of $\mathbb{R}^2$ onto $\mathbb{W}_1 = \{(x, 0) : x \in \mathbb{R}\}$ and $\mathbb{W}_2 = \{(x, x) : x \in \mathbb{R}\}$, respectively. Note that $P_{\mathbb{W}_1} \circ P_{\mathbb{W}_2}$ is a projection onto $\mathbb{W}_1$. But, it is not an orthogonal projection. Hence or otherwise, conclude that the composition of two orthogonal projections need not be an orthogonal projection?*

3. *Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. Then, A is idempotent but not symmetric. Now, define $P : \mathbb{R}^2 \to \mathbb{R}^2$ by $P(\mathbf{v}) = A\mathbf{v}$, for all $\mathbf{v} \in \mathbb{R}^2$. Then,*

   (a) *P is idempotent.*

   (b) $\text{NULL}(P) \cap \text{RNG}(P) = \text{NULL}(A) \cap \text{COL}(A) = \{\mathbf{0}\}$.

   (c) $\mathbb{R}^2 = \text{NULL}(P) + \text{RNG}(P)$. *But,* $(\text{RNG}(P))^\perp = (\text{COL}(A))^\perp \neq \text{NULL}(A)$.

   (d) *Since $(\text{COL}(A))^\perp \neq \text{NULL}(A)$, the map P is not an orthogonal projector. In this case, P is called a projection of $\mathbb{R}^2$ onto $\text{RNG}(P)$ along $\text{NULL}(P)$.*

4. *Find all 2×2 real matrices A such that $A^2 = A$. Hence, or otherwise, determine all projection operators of $\mathbb{R}^2$.*

5. *Let $\mathbb{W}$ be an $(n-1)$-dimensional subspace of $\mathbb{R}^n$ with ordered basis $\mathcal{B}_{\mathbb{W}} = [\mathbf{f}_1, \ldots, \mathbf{f}_{n-1}]$. Suppose $\mathcal{B} = [\mathbf{f}_1, \ldots, \mathbf{f}_{n-1}, \mathbf{f}_n]$ is an orthogonal ordered basis of $\mathbb{R}^n$ obtained by extending $\mathcal{B}_{\mathbb{W}}$. Now, define a function $Q : \mathbb{R}^n \to \mathbb{R}^n$ by $Q(\mathbf{v}) = \langle \mathbf{v}, \mathbf{f}_n \rangle \mathbf{f}_n - \sum\limits_{i=1}^{n-1} \langle \mathbf{v}, \mathbf{f}_i \rangle \mathbf{f}_i$. Then,*

   (a) *Q fixes every vector in $\mathbb{W}^\perp$.*

   (b) *Q sends every vector $\mathbf{w} \in \mathbb{W}$ to $-\mathbf{w}$.*

   (c) *$Q \circ Q = I_n$.*

   The function Q is called the **reflection operator** *with respect to $\mathbb{W}^\perp$.*

5.3.1 Orthogonal Projections as Self-Adjoint Operators*

Theorem 5.3.9 implies that the matrix of the projection operator is symmetric. We use this idea to proceed further.

Definition 5.3.11. Let $\mathbb{V}$ be an IPS with inner product $\langle\ ,\ \rangle$. A linear operator $P : \mathbb{V} \to \mathbb{V}$ is called **self-adjoint** if $\langle P(\mathbf{v}), \mathbf{u} \rangle = \langle \mathbf{v}, P(\mathbf{u}) \rangle$, for every $\mathbf{u}, \mathbf{v} \in \mathbb{V}$.

A careful understanding of the examples given below shows that 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 column space of Hermitian matrices. They also follow directly from the fundamental theorem of linear algebra.

Example 5.3.12. 1. Let A be an $n \times n$ real symmetric matrix. If $P : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $P(\mathbf{x}) = A\mathbf{x}$, for every $\mathbf{x} \in \mathbb{R}^n$ then

   (a) P is a self adjoint operator as $A = A^T$, for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, implies

$$\langle P(\mathbf{x}), \mathbf{y} \rangle = (\mathbf{y}^T) A\mathbf{x} = (\mathbf{y}^T) A^T \mathbf{x} = (A\mathbf{y})^T \mathbf{x} = \langle \mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, P(\mathbf{y}) \rangle.$$

(b) $\text{NULL}(P) = (\text{RNG}(P))^\perp$ as $A = A^T$. Thus, $\mathbb{R}^n = \text{NULL}(P) \oplus \text{RNG}(P)$.

2. Let A be an $n \times n$ Hermitian matrix. If $P : \mathbb{C}^n \to \mathbb{C}^n$ is defined by $P(\mathbf{z}) = A\mathbf{z}$, for all $\mathbf{z} \in \mathbb{C}^n$ then using similar arguments (see Example 5.3.12.1) prove the following:

   (a) P is a self-adjoint operator.

   (b) $\text{NULL}(P) = (\text{RNG}(P))^\perp$ as $A = A^*$. Thus, $\mathbb{C}^n = \text{NULL}(P) \oplus \text{RNG}(P)$.

We now state and prove the main result related with orthogonal projection operators.

Theorem 5.3.13. *Let $\mathbb{V}$ be a finite dimensional* IPS. *If $\mathbb{V} = \mathbb{W} \oplus \mathbb{W}^\perp$ then the orthogonal projectors $P_{\mathbb{W}} : \mathbb{V} \to \mathbb{V}$ on $\mathbb{W}$ and $P_{\mathbb{W}^\perp} : \mathbb{V} \to \mathbb{V}$ on $\mathbb{W}^\perp$ satisfy*

1. $\text{NULL}(P_{\mathbb{W}}) = \{\mathbf{v} \in V : P_{\mathbb{W}}(\mathbf{v}) = \mathbf{0}\} = \mathbb{W}^\perp = \text{RNG}(P_{\mathbb{W}^\perp})$.

2. $\text{RNG}(P_{\mathbb{W}}) = \{P_{\mathbb{W}}(\mathbf{v}) : \mathbf{v} \in \mathbb{V}\} = \mathbb{W} = \text{NULL}(P_{\mathbb{W}^\perp})$.

3. $P_{\mathbb{W}} \circ P_{\mathbb{W}} = P_{\mathbb{W}}$, $P_{\mathbb{W}^\perp} \circ P_{\mathbb{W}^\perp} = P_{\mathbb{W}^\perp}$ *(*IDEMPOTENT*)*.

4. $P_{\mathbb{W}^\perp} \circ P_{\mathbb{W}} = \mathbf{0}_{\mathbb{V}}$ *and* $P_{\mathbb{W}} \circ P_{\mathbb{W}^\perp} = \mathbf{0}_{\mathbb{V}}$, *where* $\mathbf{0}_{\mathbb{V}}(\mathbf{v}) = \mathbf{0}$, *for all* $\mathbf{v} \in \mathbb{V}$

5. $P_{\mathbb{W}} + P_{\mathbb{W}^\perp} = I_{\mathbb{V}}$, *where* $I_{\mathbb{V}}(\mathbf{v}) = \mathbf{v}$, *for all* $\mathbf{v} \in \mathbb{V}$.

6. *The operators $P_{\mathbb{W}}$ and $P_{\mathbb{W}^\perp}$ are self-adjoint.*

Proof. PART 1: As $\mathbb{V} = \mathbb{W} \oplus \mathbb{W}^\perp$, for each $\mathbf{u} \in \mathbb{W}^\perp$, one uniquely writes $\mathbf{u} = \mathbf{0} + \mathbf{u}$, where $\mathbf{0} \in \mathbb{W}$ and $\mathbf{u} \in \mathbb{W}^\perp$. Hence, by definition, $P_{\mathbb{W}}(\mathbf{u}) = \mathbf{0}$ and $P_{\mathbb{W}^\perp}(\mathbf{u}) = \mathbf{u}$. Thus, $\mathbb{W}^\perp \subseteq \text{NULL}(P_{\mathbb{W}})$ and $\mathbb{W}^\perp \subseteq \text{RNG}(P_{\mathbb{W}^\perp})$.

Now suppose that $\mathbf{v} \in \text{NULL}(P_{\mathbb{W}})$. So, $P_{\mathbb{W}}(\mathbf{v}) = \mathbf{0}$. As $\mathbb{V} = \mathbb{W} \oplus \mathbb{W}^\perp$, $\mathbf{v} = \mathbf{w} + \mathbf{u}$, for unique $\mathbf{w} \in W$ and unique $\mathbf{u} \in \mathbb{W}^\perp$. So, by definition, $P_{\mathbb{W}}(\mathbf{v}) = \mathbf{w}$. Thus, $\mathbf{w} = P_{\mathbb{W}}(\mathbf{v}) = \mathbf{0}$. That is, $\mathbf{v} = \mathbf{0} + \mathbf{u} = \mathbf{u} \in \mathbb{W}^\perp$. Thus, $\text{NULL}(P_{\mathbb{W}}) \subseteq \mathbb{W}^\perp$.

A similar argument implies $\text{RNG}(P_{\mathbb{W}^\perp}) \subseteq W^\perp$ and thus completing the proof of the first part.

PART 2: Use an argument similar to the proof of Part 1.

PART 3, PART 4 AND PART 5: Let $\mathbf{v} \in \mathbb{V}$. Then, $\mathbf{v} = \mathbf{w} + \mathbf{u}$, for unique $\mathbf{w} \in \mathbb{W}$ and unique $\mathbf{u} \in \mathbb{W}^\perp$. Thus, by definition,

$$\begin{aligned}
(P_{\mathbb{W}} \circ P_{\mathbb{W}})(\mathbf{v}) &= P_{\mathbb{W}}\big(P_{\mathbb{W}}(\mathbf{v})\big) = P_{\mathbb{W}}(\mathbf{w}) = \mathbf{w} \text{ and } P_{\mathbb{W}}(\mathbf{v}) = \mathbf{w} \\
(P_{\mathbb{W}^\perp} \circ P_{\mathbb{W}})(\mathbf{v}) &= P_{\mathbb{W}^\perp}\big(P_{\mathbb{W}}(\mathbf{v})\big) = P_{\mathbb{W}^\perp}(\mathbf{w}) = \mathbf{0} \text{ and} \\
(P_{\mathbb{W}} \oplus P_{\mathbb{W}^\perp})(\mathbf{v}) &= P_{\mathbb{W}}(\mathbf{v}) + P_{\mathbb{W}^\perp}(\mathbf{v}) = \mathbf{w} + \mathbf{u} = \mathbf{v} = I_{\mathbb{V}}(\mathbf{v}).
\end{aligned}$$

Hence, $P_{\mathbb{W}} \circ P_{\mathbb{W}} = P_{\mathbb{W}}$, $P_{\mathbb{W}^\perp} \circ P_{\mathbb{W}} = \mathbf{0}_{\mathbb{V}}$ and $I_{\mathbb{V}} = P_{\mathbb{W}} \oplus P_{\mathbb{W}^\perp}$.

PART 6: Let $\mathbf{u} = \mathbf{w}_1 + \mathbf{x}_1$ and $\mathbf{v} = \mathbf{w}_2 + \mathbf{x}_2$, for unique $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{W}$ and unique $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{W}^\perp$. Then, by definition, $\langle \mathbf{w}_i, \mathbf{x}_j \rangle = 0$ for $1 \le i, j \le 2$. Thus,

$$\langle P_{\mathbb{W}}(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{w}_1, \mathbf{v} \rangle = \langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \langle \mathbf{u}, \mathbf{w}_2 \rangle = \langle \mathbf{u}, P_{\mathbb{W}}(\mathbf{v}) \rangle$$

and the proof of the theorem is complete. ∎

Remark 5.3.14. *Theorem 5.3.13 gives us the following:*

1. *The orthogonal projectors $P_{\mathbb{W}}$ and $P_{\mathbb{W}^\perp}$ are idempotent and self-adjoint.*

2. *Let $\mathbf{v} \in \mathbb{V}$. Then, $\mathbf{v} - P_{\mathbb{W}}(\mathbf{v}) = (I_{\mathbb{V}} - P_{\mathbb{W}})(\mathbf{v}) = P_{\mathbb{W}^\perp}(\mathbf{v}) \in \mathbb{W}^\perp$. Thus, $\langle \mathbf{v} - P_{\mathbb{W}}(\mathbf{v}), \mathbf{w}\rangle = 0$, for every $\mathbf{v} \in \mathbb{V}$ and $\mathbf{w} \in \mathbb{W}$.*

3. *As $P_{\mathbb{W}}(\mathbf{v}) - \mathbf{w} \in \mathbb{W}$, for each $\mathbf{v} \in \mathbb{V}$ and $\mathbf{w} \in \mathbb{W}$, we have*

$$\begin{aligned}
\|\mathbf{v} - \mathbf{w}\|^2 &= \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v}) + P_{\mathbb{W}}(\mathbf{v}) - \mathbf{w}\|^2 \\
&= \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|^2 + \|P_{\mathbb{W}}(\mathbf{v}) - \mathbf{w}\|^2 + 2\langle \mathbf{v} - P_{\mathbb{W}}(\mathbf{v}), P_{\mathbb{W}}(\mathbf{v}) - \mathbf{w}\rangle \\
&= \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|^2 + \|P_{\mathbb{W}}(\mathbf{v}) - \mathbf{w}\|^2.
\end{aligned}$$

Therefore, $\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|$ and equality holds if and only if $\mathbf{w} = P_{\mathbb{W}}(\mathbf{v})$. Since $P_{\mathbb{W}}(\mathbf{v}) \in \mathbb{W}$, we see that

$$d(\mathbf{v}, \mathbb{W}) = \inf\left\{\|\mathbf{v} - \mathbf{w}\| \ : \mathbf{w} \in W\right\} = \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|.$$

That is, $P_{\mathbb{W}}(\mathbf{v})$ is the vector nearest to $\mathbf{v} \in \mathbb{W}$. This can also be stated as: the vector $P_{\mathbb{W}}(\mathbf{v})$ solves the following minimization problem:

$$\inf_{\mathbf{w} \in \mathbb{W}} \|\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|.$$

The next theorem is a generalization of Theorem 5.3.13. We omit the proof as the arguments are similar and uses the following:
Let $\mathbb{V}$ be a finite dimensional IPS with $\mathbb{V} = \mathbb{W}_1 \oplus \cdots \oplus \mathbb{W}_k$, for certain subspaces $\mathbb{W}_i$'s of $\mathbb{V}$. Then, for each $\mathbf{v} \in \mathbb{V}$ there exist unique vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ such that

1. $\mathbf{v}_i \in \mathbb{W}_i$, for $1 \leq i \leq k$,

2. $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ for each $\mathbf{v}_i \in \mathbb{W}_i, \mathbf{v}_j \in \mathbb{W}_j, 1 \leq i \neq j \leq k$ and

3. $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_k$.

Theorem 5.3.15. *Let $\mathbb{V}$ be a finite dimensional* IPS *with subspaces $\mathbb{W}_1, \ldots, \mathbb{W}_k$ of $\mathbb{V}$ such that $\mathbb{V} = W_1 \oplus \cdots \oplus \mathbb{W}_k$. Then, for each i, j, $1 \leq i \neq j \leq k$, there exist orthogonal projectors $P_{\mathbb{W}_i} : \mathbb{V} \to \mathbb{V}$ of $\mathbb{V}$ onto $\mathbb{W}_i$ satisfying the following:*

1. $\textsc{Null}(P_{\mathbb{W}_i}) = W_i^\perp = \mathbb{W}_1 \oplus \mathbb{W}_2 \oplus \cdots \oplus W_{i-1} \oplus \mathbb{W}_{i+1} \oplus \cdots \oplus W_k$.

2. $\textsc{Rng}(P_{\mathbb{W}_i}) = \mathbb{W}_i$.

3. $P_{\mathbb{W}_i} \circ P_{\mathbb{W}_i} = P_{\mathbb{W}_i}$.

4. $P_{\mathbb{W}_i} \circ P_{\mathbb{W}_j} = \mathbf{0}_{\mathbb{V}}$.

5. *$P_{\mathbb{W}_i}$ is a self-adjoint operator, and*

6. $I_{\mathbb{V}} = P_{\mathbb{W}_1} \oplus P_{\mathbb{W}_2} \oplus \cdots \oplus P_{\mathbb{W}_k}$.

5.4 Orthogonal Operator and Rigid Motion*

We now give the definition and a few properties of an orthogonal operator.

Definition 5.4.1. Let $\mathbb{V}$ be a vector space. Then, a linear operator $T : \mathbb{V} \to \mathbb{V}$ is said to be an **orthogonal operator** if $\|T(\mathbf{x})\| = \|\mathbf{x}\|$, for all $\mathbf{x} \in \mathbb{V}$.

Example 5.4.2. Each $T \in \mathcal{L}(\mathbb{V})$ given below is an orthogonal operator.

1. Fix a unit vector $\mathbf{a} \in \mathbb{V}$ and define $T(\mathbf{x}) = 2\langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a} - \mathbf{x}$, for all $\mathbf{x} \in \mathbb{V}$.
   Solution: Note that $\mathrm{Proj}_{\mathbf{a}}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a}$. So, $\big\langle \langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a},\ \mathbf{x} - \langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a} \big\rangle = 0$. Also, by Pythagoras theorem $\|\mathbf{x} - \langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a}\|^2 = \|\mathbf{x}\|^2 - (\langle \mathbf{x}, \mathbf{a}\rangle)^2$. Thus,
   $$\|T(\mathbf{x})\|^2 = \|(\langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a}) + (\langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a} - \mathbf{x})\|^2 = \|\langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a}\|^2 + \|\mathbf{x} - \langle \mathbf{x}, \mathbf{a}\rangle \mathbf{a}\|^2 = \|\mathbf{x}\|^2.$$
2. Let $n = 2, \mathbb{V} = \mathbb{R}^2$ and $0 \le \theta < 2\pi$. Now define $T(\mathbf{x}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$.

We now show that an operator is orthogonal if and only if it preserves the angle.

Theorem 5.4.3. *Let $T \in \mathcal{L}(\mathbb{V})$. Then, the following statements are equivalent.*

1. *T is an orthogonal operator.*
2. *$\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$. That is, T preserves inner product.*

Proof. $1 \Rightarrow 2$ Let T be an orthogonal operator. Then, $\|T(\mathbf{x}+\mathbf{y})\|^2 = \|\mathbf{x}+\mathbf{y}\|^2$. So, $\|T(\mathbf{x})\|^2 + \|T(\mathbf{y})\|^2 + 2\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \|T(\mathbf{x}) + T(\mathbf{y})\|^2 = \|T(\mathbf{x}+\mathbf{y})\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle$. Thus, using definition again $\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.

$2 \Rightarrow 1$ If $\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ then T is an orthogonal operator as $\|T(\mathbf{x})\|^2 = \langle T(\mathbf{x}), T(\mathbf{x})\rangle = \langle \mathbf{x}, \mathbf{x}\rangle = \|\mathbf{x}\|^2$. ∎

As an immediate corollary, we obtain the following result.

Corollary 5.4.4. *Let $T \in \mathcal{L}(\mathbb{V})$. Then, T is an orthogonal operator if and only if "for every orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathbb{V}$, $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_n)\}$ is an orthonormal basis of $\mathbb{V}$". Thus, if $\mathcal{B}$ is an orthonormal ordered basis of $\mathbb{V}$ then $T[\mathcal{B}, \mathcal{B}]$ is an orthogonal matrix.*

Definition 5.4.5. Let $\mathbb{V}$ be a vector space. Then, a map $T : \mathbb{V} \to \mathbb{V}$ is said to be an **isometry or a rigid motion** if $\|T(\mathbf{x}) - T(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$. That is, an isometry is distance preserving.

Observe that if T and S are two rigid motions then ST is also a rigid motion. Furthermore, it is clear from the definition that every rigid motion is invertible.

Example 5.4.6. The maps given below are rigid motions/isometry.

1. Let $\mathbb{V}$ be a linear space with norm $\|\cdot\|$. If $\mathbf{a} \in \mathbb{V}$ then the translation map $T_{\mathbf{a}} : \mathbb{V} \to \mathbb{V}$ (see Exercise 7), defined by $T_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ for all $\mathbf{x} \in \mathbb{V}$, is an isometry/rigid motion as
   $$\|T_{\mathbf{a}}(\mathbf{x}) - T_{\mathbf{a}}(\mathbf{y})\| = \|\,(\mathbf{x}+\mathbf{a}) - (\mathbf{y}+\mathbf{a})\,\| = \|\mathbf{x} - \mathbf{y}\|.$$

2. Let $\mathbb{V}$ be an ips. Then, using Theorem 5.4.3, we see that every orthogonal operator is an isometry.

We now prove that every rigid motion that fixes origin is an orthogonal operator.

Theorem 5.4.7. *Let $\mathbb{V}$ be a real ips. Then, the following statements are equivalent for any map $T : \mathbb{V} \to \mathbb{V}$.*

1. *T is a rigid motion that fixes origin.*
2. *T is linear and $\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ (preserves inner product).*
3. *T is an orthogonal operator.*

Proof. We have already seen the equivalence of Part 2 and Part 3 in Theorem 5.4.3. Let us now prove the equivalence of Part 1 and Part 2/Part 3.

If T is an orthogonal operator then $T(\mathbf{0}) = \mathbf{0}$ and $\|T(\mathbf{x}) - T(\mathbf{y})\| = \|T(\mathbf{x} - \mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$. This proves Part 3 implies Part 1.

We now prove Part 1 implies Part 2. So, let T be a rigid motion that fixes $\mathbf{0}$. Thus, $T(\mathbf{0}) = \mathbf{0}$ and $\|T(\mathbf{x}) - T(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$. Hence, in particular for $\mathbf{y} = \mathbf{0}$, we have $\|T(\mathbf{x})\| = \|\mathbf{x}\|$, for all $\mathbf{x} \in \mathbb{V}$. So,

$$\begin{aligned}
\|T(\mathbf{x})\|^2 + \|T(\mathbf{y})\|^2 - 2\langle T(\mathbf{x}), T(\mathbf{y})\rangle &= \langle T(\mathbf{x}) - T(\mathbf{y}), T(\mathbf{x}) - T(\mathbf{y})\rangle = \|T(\mathbf{x}) - T(\mathbf{y})\|^2 \\
&= \|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y}\rangle \\
&= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y}\rangle.
\end{aligned}$$

Thus, using $\|T(\mathbf{x})\| = \|\mathbf{x}\|$, for all $\mathbf{x} \in \mathbb{V}$, we get $\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$. Now, to prove T is linear, we use $\langle T(\mathbf{x}), T(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ in 3-rd and 4-th line to get

$$\begin{aligned}
\|T(\mathbf{x} + \mathbf{y}) - (T(\mathbf{x}) + T(\mathbf{y}))\|^2 &= \langle T(\mathbf{x} + \mathbf{y}) - (T(\mathbf{x}) + T(\mathbf{y})), T(\mathbf{x} + \mathbf{y}) - (T(\mathbf{x}) + T(\mathbf{y}))\rangle \\
&= \langle T(\mathbf{x} + \mathbf{y}), T(\mathbf{x} + \mathbf{y})\rangle - 2\langle T(\mathbf{x} + \mathbf{y}), T(\mathbf{x})\rangle \\
&\quad -2\langle T(\mathbf{x} + \mathbf{y}), T(\mathbf{y})\rangle + \langle T(\mathbf{x}) + T(\mathbf{y}), T(\mathbf{x}) + T(\mathbf{y})\rangle \\
&= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}\rangle - 2\langle \mathbf{x} + \mathbf{y}, \mathbf{x}\rangle - 2\langle \mathbf{x} + \mathbf{y}, \mathbf{y}\rangle \\
&\quad +\langle T(\mathbf{x}), T(\mathbf{x})\rangle + 2\langle T(\mathbf{x}), T(\mathbf{y})\rangle + \langle T(\mathbf{y}), T(\mathbf{y})\rangle \\
&= -\langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle + 2\langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle = 0.
\end{aligned}$$

Thus, $T(\mathbf{x}+\mathbf{y}) - (T(\mathbf{x}) + T(\mathbf{y})) = \mathbf{0}$ and hence $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$. A similar calculation gives $T(\alpha\mathbf{x}) = \alpha T(\mathbf{x})$ and hence T is linear. ∎

EXERCISE **5.4.8.** *1. Let $A, B \in \mathbb{M}_n(\mathbb{C})$. Then, A and B are said to be*

(a) **Orthogonally Congruent** *if $B = S^T AS$, for some invertible matrix S.*

(b) **Unitarily Congruent** *if $B = S^* AS$, for some invertible matrix S.*

Prove that Orthogonal and Unitary congruences are equivalence relations on $\mathbb{M}_n(\mathbb{R})$ and $\mathbb{M}_n(\mathbb{C})$, respectively.

2. *Let* $\mathbf{x} \in \mathbb{C}^2$. *Identify it with the complex number* $\mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_2$. *If we rotate* $\mathbf{x}$ *by a counterclockwise rotation* $\theta, 0 \le \theta < 2\pi$ *then, we have*

$$\mathbf{x}e^{i\theta} = (\mathbf{x}_1 + i\mathbf{x}_2)(\cos\theta + i\sin\theta) = \mathbf{x}_1\cos\theta - \mathbf{x}_2\sin\theta + i[\mathbf{x}_1\sin\theta + \mathbf{x}_2\cos\theta].$$

Thus, the corresponding vector in $\mathbb{R}^2$ *is*

$$\begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Is the matrix, $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, *the matrix of the corresponding rotation? Justify.*

3. *Let* $A \in M_2(\mathbb{R})$ *and* $T(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, *for* $\theta \in \mathbb{R}$. *Then,* A *is an orthogonal matrix if and only if* $A = T(\theta)$ *or* $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} T(\theta)$, *for some* $\theta \in \mathbb{R}$.

Ans: To see this assume that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is orthogonal. Thus $a^2 + b^2 = c^2 + d^2 = 1$ and $ac + bd = ab + cd = 0$. Note that $(b - c)(d - a) = ac + bd - ab - cd = 0$ and so either $b = c$ or $a = d$.

Without loss we assume $a = d$, otherwise we consider $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} A$. If $a \neq 0$, then from $0 = ac + bd = a(c + b)$, we get $c = -b$. So $A = \begin{bmatrix} \cos\theta & \pm\sin\theta \\ \mp\sin\theta & \cos\theta \end{bmatrix}$. If $a = d = 0$, then $b, c \in \{-1, 1\}$. In both the cases $A = T(\theta)$, for some θ.

4. *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Then, the following statements are equivalent.*

   (a) A *is an orthogonal matrix.*

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

   (c) A^T *is orthogonal.*

   (d) *the columns of* A *form an orthonormal basis of the real vector space* $\mathbb{R}^n$.

   (e) *the rows of* A *form an orthonormal basis of the real vector space* $\mathbb{R}^n$.

   (f) *for any two vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ Orthogonal matrices preserve angle.

   (g) *for any vector* $\mathbf{x} \in \mathbb{C}^n$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$ Orthogonal matrices preserve length.

5. *Let* U *be an* $n \times n$ *matrix. Then, prove that the following statements are equivalent.*

   (a) U *is a unitary matrix.*

   (b) $U^{-1} = U^*$.

   (c) U^* *is unitary.*

   (d) *the columns of* U *form an orthonormal basis of the complex vector space* $\mathbb{C}^n$.

(e) *the rows of U form an orthonormal basis of the complex vector space $\mathbb{C}^n$.*

(f) *for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ UNITARY MATRICES PRESERVE ANGLE.*

(g) *for any vector $\mathbf{x} \in \mathbb{C}^n$, $\|U\mathbf{x}\| = \|\mathbf{x}\|$ UNITARY MATRICES PRESERVE LENGTH.*

Ans: Part 5a$\Leftrightarrow$ Part 5g. If U is unitary, then $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x} = \mathbf{x}^*U^*U\mathbf{x} = \|U\mathbf{x}\|^2$. Conversely, we have

$$\langle U^*U\mathbf{x}, \mathbf{x}\rangle = \langle U\mathbf{x}, U\mathbf{x}\rangle = \|U\mathbf{x}\|^2 = \|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x}\rangle, \text{ for all } \mathbf{x}.$$

That is $\langle (U^*U - I)\mathbf{x}, \mathbf{x}\rangle = 0$, for all $\mathbf{x}$. Put $B = U^*U - I$. Now, taking $\mathbf{x} = \mathbf{e}_i$, we see that $B(i,i) = 0$. For $i \neq j$, taking $\mathbf{x} = \mathbf{e}_i + \mathbf{e}_j$, we get

$$\mathbf{x}^*B\mathbf{x} = B(i,i) + B(i,j) + B(j,i) + B(j,j) = 0,$$

so that $B(i,j) + B(j,i) = 0$. Taking $\mathbf{x} = \mathbf{e}_i + \mathbf{i}\mathbf{e}_j$ (here $\mathbf{i}^2 = -1$), we get

$$\mathbf{x}^*B\mathbf{x} = B(i,i) + \mathbf{i}B(i,j) - \mathbf{i}B(j,i) + B(j,j) = 0,$$

so that $B(i,j) - B(j,i) = 0$. Thus $B = 0$ and so $U^*U = I$. The rest is exercise.

6. *Let A be an $n \times n$ orthogonal matrix. Then, prove that $\det(A) = \pm 1$.*

7. *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.*

8. *Prove that in $M_5(\mathbb{R})$, there are infinitely many orthogonal matrices of which only finitely many are diagonal (in fact, there number is just 32).*

9. *Prove that permutation matrices are real orthogonal.*

10. *Let $A, B \in \mathbb{M}_n(\mathbb{C})$ be two unitary matrices. Then, prove that AB and BA are unitary matrices.*

11. *If $A = [a_{ij}]$ and $B = [b_{ij}]$ are unitarily equivalent then prove that $\sum\limits_{ij} |a_{ij}|^2 = \sum\limits_{ij} |b_{ij}|^2$.*

Ans: Notice that

$$\begin{aligned}\sum |b_{ij}|^2 &= \sum_i \|B[:,i]\|^2 = \sum_i \|(UB)[:,i]\|^2 = \sum_i \|(AU)[:,i]\|^2 \qquad (5.4.1)\\ &= \sum_i \|(AU)[i,:]\|^2 = \sum_i \|A[i,:]\|^2 = \sum |a_{ij}|^2. \qquad (5.4.2)\end{aligned}$$

We have used that U gives an isometry. ∎

Alternate. An alternate proof for the above is the following:

$$\sum |a_{ij}|^2 = \mathrm{tr}(A^*A) = \mathrm{tr}(UB^*U^*UBU^*) = \mathrm{tr}(U[B^*BU^*]) = \mathrm{tr}([B^*BU^*]U) = \mathrm{tr}(B^*B) = \sum |b_{ij}|^2.$$

12. *Let U be a unitary matrix and for every $\mathbf{x} \in \mathbb{C}^n$, define*

$$\|\mathbf{x}\|_1 = \max\{|\mathbf{x}_i| : \mathbf{x}^T = [\mathbf{x}_1, \ldots, \mathbf{x}_n]\}.$$

Then, is it necessary that $\|U\mathbf{x}\|_1 = \|\mathbf{x}\|_1$?

Ans: No. You may use rotation matrices to see this.

5.5 Summary

In the previous chapter, we learnt that if $\mathbb{V}$ is vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$ then $\mathbb{V}$ basically looks like $\mathbb{F}^n$. Also, any subspace of $\mathbb{F}^n$ is either $\textsc{Col}(A)$ or $\textsc{Null}(A)$ or both, for some matrix A with entries from $\mathbb{F}$.

So, we started this chapter with inner product, a generalization of the dot product in $\mathbb{R}^3$ or $\mathbb{R}^n$. We used the inner product to define the length/norm of a vector. The norm has the property that "the norm of a vector is zero if and only if the vector itself is the zero vector". We then proved the Cauchy-Bunyakovskii-Schwartz Inequality which helped us in defining the angle between two vector. Thus, one can talk of geometrical problems in $\mathbb{R}^n$ and proved some geometrical results.

We then independently defined the notion of a norm in $\mathbb{R}^n$ and showed that a norm is induced by an inner product if and only if the norm satisfies the parallelogram law (sum of squares of the diagonal equals twice the sum of square of the two non-parallel sides).

The next subsection dealt with the fundamental theorem of linear algebra where we showed that if $A \in \mathbb{M}_{m,n}(\mathbb{C})$ then

1. $\dim(\textsc{Null}(A)) + \dim(\textsc{Col}(A)) = n$.
2. $\textsc{Null}(A) = \big(\textsc{Col}(A^*)\big)^{\perp}$ and $\textsc{Null}(A^*) = \big(\textsc{Col}(A)\big)^{\perp}$.
3. $\dim(\textsc{Col}(A)) = \dim(\textsc{Col}(A^*))$.

We then saw that having an orthonormal basis is an asset as determining the

1. coordinates of a vector boils down to computing the inner product.
2. projection of a vector on a subspace boils down to finding an orthonormal basis of the subspace and then summing the corresponding rank 1 matrices.

So, the question arises, how do we compute an orthonormal basis? This is where we came across the Gram-Schmidt Orthonormalization process. This algorithm helps us to determine an orthonormal basis of $LS(S)$ for any finite subset S of a vector space. This also lead to the QR-decomposition of a matrix.

Thus, we observe the following about the linear system $A\mathbf{x} = \mathbf{b}$. If

1. $\mathbf{b} \in \textsc{Col}(A)$ then we can use the Gauss-Jordan method to get a solution.
2. $\mathbf{b} \notin \textsc{Col}(A)$ then in most cases we need a vector $\mathbf{x}$ such that the least square error between $\mathbf{b}$ and $A\mathbf{x}$ is minimum. We saw that this minimum is attained by the projection of $\mathbf{b}$ on $\textsc{Col}(A)$. Also, this vector can be obtained either using the fundamental theorem of linear algebra or by computing the matrix $B(B^TB)^{-1}B^T$, where the columns of B are either the pivot columns of A or a basis of $\textsc{Col}(A)$.

Chapter 6

Eigenvalues, Eigenvectors and Diagonalizability

6.1 Introduction and Definitions

In this chapter, every matrix is an element of $\mathbb{M}_n(\mathbb{C})$ and $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$, for some $n \in \mathbb{N}$. We start with a few examples to motivate this chapter.

Example 6.1.1. 1. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$.

(a) Then A magnifies the nonzero vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ three times as $A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and behaves by changing the direction of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ as $A\begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Further, the vectors $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ are orthogonal.

(b) B magnifies both the vectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ as $B\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 5\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $B\begin{bmatrix} 2 \\ -1 \end{bmatrix} = 10\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Here again, the vectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ are orthogonal.

(c) $\mathbf{x}^T A\mathbf{x} = 3\dfrac{(x+y)^2}{2} - \dfrac{(x-y)^2}{2}$. Here, the displacements occur along perpendicular lines $x+y=0$ and $x-y=0$, where $x+y = (x,y)\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $x-y = (x,y)\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Whereas $\mathbf{x}^T B\mathbf{x} = 5\dfrac{(x+2y)^2}{5} + 10\dfrac{(2x-y)^2}{5}$. Here also the maximum/minimum displacements occur along the orthogonal lines $x+2y=0$ and $2x-y=0$, where $x+2y = (x,y)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $2x-y = (x,y)\begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

(d) the curve $\mathbf{x}^T A\mathbf{x} = 10$ represents a hyperbola, where as the curve $\mathbf{x}^T B\mathbf{x} = 10$ represents an ellipse (see Figure 6.1 drawn using the package "Sagemath").

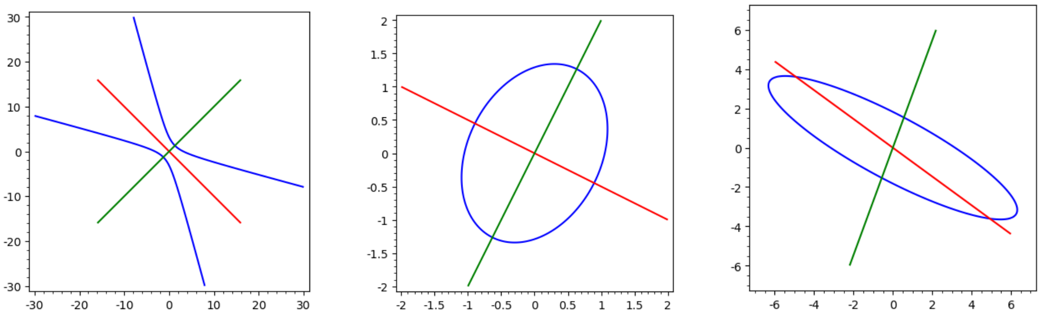


Figure 6.1: A Hyperbola and two Ellipses (first one has orthogonal axes)

.

2. Let $C = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$, a non-symmetric matrix. Then, does there exist a nonzero $\mathbf{x} \in \mathbb{C}^2$ which gets magnified by C?

   So, we need $\mathbf{x} \neq \mathbf{0}$ and $\alpha \in \mathbb{C}$ such that $C\mathbf{x} = \alpha\mathbf{x} \Leftrightarrow [\alpha I_2 - C]\mathbf{x} = \mathbf{0}$. As $\mathbf{x} \neq 0$, $[\alpha I_2 - C]\mathbf{x} = \mathbf{0}$ has a solution if and only if $\det[\alpha I - A] = 0$. But,

$$\det[\alpha I - A] = \det\left(\begin{bmatrix} \alpha - 1 & -2 \\ -1 & \alpha - 3 \end{bmatrix}\right) = \alpha^2 - 4\alpha + 1.$$

   So, $\alpha = 2 \pm \sqrt{3}$. For $\alpha = 2 + \sqrt{3}$, verify that the $\mathbf{x} \neq \mathbf{0}$ that satisfies $\begin{bmatrix} 1+\sqrt{3} & -2 \\ -1 & \sqrt{3}-1 \end{bmatrix}\mathbf{x} = \mathbf{0}$ equals $\mathbf{x} = \begin{bmatrix} \sqrt{3}-1 \\ 1 \end{bmatrix}$. Similarly, for $\alpha = 2 - \sqrt{3}$, the vector $\mathbf{x} = \begin{bmatrix} \sqrt{3}+1 \\ -1 \end{bmatrix}$ satisfies $\begin{bmatrix} 1-\sqrt{3} & -2 \\ -1 & -\sqrt{3}-1 \end{bmatrix}\mathbf{x} = \mathbf{0}$. In this example,

   (a) we still have magnifications in the directions $\begin{bmatrix} \sqrt{3}-1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} \sqrt{3}+1 \\ -1 \end{bmatrix}$.

   (b) the maximum/minimum displacements do not occur along the lines $(\sqrt{3}-1)x+y = 0$ and $(\sqrt{3}+1)x - y = 0$ (see the third curve in Figure 6.1).

   (c) the lines $(\sqrt{3}-1)x + y = 0$ and $(\sqrt{3}+1)x - y = 0$ are not orthogonal.

3. 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

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T A\mathbf{x} - \lambda(\mathbf{x}^T\mathbf{x} - 1) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j - \lambda\left(\sum_{i=1}^{n} x_i^2 - 1\right).$$

Partially differentiating $L(\mathbf{x}, \lambda)$ with respect to x_i for $1 \le i \le n$, we get

$$\begin{aligned}\frac{\partial L}{\partial x_1} &= 2a_{11}x_1 + 2a_{12}x_2 + \cdots + 2a_{1n}x_n - 2\lambda x_1,\\ \vdots &= \vdots \\ \frac{\partial L}{\partial x_n} &= 2a_{n1}x_1 + 2a_{n2}x_2 + \cdots + 2a_{nn}x_n - 2\lambda x_n.\end{aligned}$$

Therefore, to get the points of extremum, we solve for

$$\mathbf{0}^T = \left(\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2}, \ldots, \frac{\partial L}{\partial x_n}\right)^T = \frac{\partial L}{\partial \mathbf{x}} = 2(A\mathbf{x} - \lambda\mathbf{x}).$$

Thus, to solve the extremal problem, we need $\lambda \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$ and $A\mathbf{x} = \lambda\mathbf{x}$.

We observe the following about the matrices A, B and C that appear in Example 6.1.1.

1. $\det(A) = -3 = 3 \times -1$, $\det(B) = 50 = 5 \times 10$ and $\det(C) = 1 = (2+\sqrt{3}) \times (2-\sqrt{3})$.
2. $\mathsf{tr(A)} = 2 = 3 - 1$, $\mathsf{tr(B)} = 15 = 5 + 10$ and $\det(C) = 4 = (2+\sqrt{3}) + (2-\sqrt{3})$.
3. The sets $\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}\right\}$, $\left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\-1\end{bmatrix}\right\}$ and $\left\{\begin{bmatrix}\sqrt{3}-1\\1\end{bmatrix}, \begin{bmatrix}\sqrt{3}+1\\-1\end{bmatrix}\right\}$ are linearly independent.
4. If $\mathbf{v}_1 = \begin{bmatrix}1\\1\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}1\\-1\end{bmatrix}$ and $S = [\mathbf{v}_1, \mathbf{v}_2]$ then
   (a) $AS = [A\mathbf{v}_1, A\mathbf{v}_2] = [3\mathbf{v}_1, -\mathbf{v}_2] = S\begin{bmatrix}3 & 0\\0 & -1\end{bmatrix} \Leftrightarrow S^{-1}AS = \begin{bmatrix}3 & 0\\0 & -1\end{bmatrix} = \mathrm{diag}(3, -1)$.
   (b) Let $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\mathbf{v}_1$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\mathbf{v}_2$. Then, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal unit vectors, *i.e.*, if $U = [\mathbf{u}_1, \mathbf{u}_2]$ then $I = UU^* = \mathbf{u}_1\mathbf{u}_1^* + \mathbf{u}_2\mathbf{u}_2^*$ and $A = 3\mathbf{u}_1\mathbf{u}_1^* - \mathbf{u}_2\mathbf{u}_2^*$.
5. If $\mathbf{v}_1 = \begin{bmatrix}1\\2\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}2\\-1\end{bmatrix}$ and $S = [\mathbf{v}_1, \mathbf{v}_2]$ then
   (a) $AS = [A\mathbf{v}_1, A\mathbf{v}_2] = [5\mathbf{v}_1, 10\mathbf{v}_2] = S\begin{bmatrix}5 & 0\\0 & 10\end{bmatrix} \Leftrightarrow S^{-1}AS = \begin{bmatrix}5 & 0\\0 & 10\end{bmatrix} = \mathrm{diag}(3, -1)$.
   (b) Let $\mathbf{u}_1 = \frac{1}{\sqrt{5}}\mathbf{v}_1$ and $\mathbf{u}_2 = \frac{1}{\sqrt{5}}\mathbf{v}_2$. Then, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal unit vectors, *i.e.*, if $U = [\mathbf{u}_1, \mathbf{u}_2]$ then $I = UU^* = \mathbf{u}_1\mathbf{u}_1^* + \mathbf{u}_2\mathbf{u}_2^*$ and $A = 5\mathbf{u}_1\mathbf{u}_1^* + 10\mathbf{u}_2\mathbf{u}_2^*$.
6. If $\mathbf{v}_1 = \begin{bmatrix}\sqrt{3}-1\\1\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}\sqrt{3}+1\\-1\end{bmatrix}$ and $S = [\mathbf{v}_1, \mathbf{v}_2]$ then

$$S^{-1}CS = \begin{bmatrix}2+\sqrt{3} & 0\\0 & 2-\sqrt{3}\end{bmatrix} = \mathrm{diag}(2+\sqrt{3}, 2-\sqrt{3}).$$

Thus, we see that given $A \in \mathbb{M}_n(\mathbb{C})$, the number $\lambda \in \mathbb{C}$ and $\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}$ satisfying $A\mathbf{x} = \lambda\mathbf{x}$ have certain nice properties. For example, there exists a basis of $\mathbb{C}^2$ in which the matrices A, B and C behave like diagonal matrices. To understand the ideas better, we start with the following definitions.

Definition 6.1.2. **[Eigenvalues, Eigenvectors and Eigenspace]** Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,

1. the equation
$$A\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow (A - \lambda I_n)\mathbf{x} = \mathbf{0} \tag{6.1.1}$$
is called the **eigen-condition**.
2. an $\alpha \in \mathbb{C}$ is called a **characteristic value/root** or **eigenvalue** or **latent root** of A if there exists $\mathbf{x} \neq \mathbf{0}$ satisfying $A\mathbf{x} = \alpha\mathbf{x}$.
3. an $\mathbf{x} \neq \mathbf{0}$ satisfying Equation (6.1.1) is called a **characteristic vector** or **eigenvector** or **invariant/latent vector** of A corresponding to λ.
4. the tuple $(\alpha, \mathbf{x})$ with $\mathbf{x} \neq \mathbf{0}$ and $A\mathbf{x} = \alpha\mathbf{x}$ is called an **eigen-pair** or **characteristic-pair**.
5. for an eigenvalue $\alpha \in \mathbb{C}$, $\text{NULL}(A - \alpha I) = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \alpha\mathbf{x}\}$ is called the **eigenspace** or **characteristic vector space** of A corresponding to α.

Theorem 6.1.3. *Let $A \in \mathbb{M}_n(\mathbb{C})$ and $\alpha \in \mathbb{C}$. Then, the following statements are equivalent.*

1. *α is an eigenvalue of A.*
2. $\det(A - \alpha I_n) = 0$.

Proof. We know that α is an eigenvalue of A if any only if the system $(A - \alpha I_n)\mathbf{x} = \mathbf{0}$ has a non-trivial solution. By Theorem 2.4.4 this holds if and only if $\det(A - \alpha I) = 0$. ∎

Definition 6.1.4. **[Characteristic Polynomial / Equation, Spectrum and Spectral Radius]** Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,

1. $\det(A - \lambda I)$ is a polynomial of degree n in λ and is called the **characteristic polynomial** of A, denoted $P_A(\lambda)$, or in short $P(\lambda)$.
2. the equation $P_A(\lambda) = 0$ is called the **characteristic equation** of A.
3. The multi-set (collection with multiplicities) $\{\alpha \in \mathbb{C} : P_A(\alpha) = 0\}$ is called the **spectrum** of A, denoted $\sigma(A)$. Hence, $\sigma(A)$ contains all the eigenvalues of A.
4. The **Spectral Radius**, denoted $\rho(A)$ of $A \in \mathbb{M}_n(\mathbb{C})$, equals $\max\{|\alpha| : \alpha \in \sigma(A)\}$.

We thus observe the following.

Remark 6.1.5. *Let $A \in \mathbb{M}_n(\mathbb{C})$.*

1. *Then, A is singular if and only if $0 \in \sigma(A)$.*
2. *Further, if $\alpha \in \sigma(A)$ then the following statements hold.*
   (a) $\{\mathbf{0}\} \subsetneq \text{NULL}(A - \alpha I)$. *Therefore, if* $\text{RANK}(A - \alpha I) = r$ *then* $r < n$. *Hence, by Theorem 2.4.4, the system* $(A - \alpha I)\mathbf{x} = \mathbf{0}$ *has* $n - r$ *linearly independent solutions.*
   (b) $\mathbf{x} \in \text{NULL}(A - \alpha I)$ *if and only if* $c\mathbf{x} \in \text{NULL}(A - \alpha I)$, *for* $c \neq 0$.
   (c) *If* $\mathbf{x}_1, \ldots, \mathbf{x}_r \in \text{NULL}(A - \alpha I)$ *are linearly independent then* $\sum\limits_{i=1}^{r} c_i\mathbf{x}_i \in \text{NULL}(A - \alpha I)$, *for all* $c_i \in \mathbb{C}$. *Hence, if S is a collection of eigenvectors then, we necessarily want the set S to be* LINEARLY INDEPENDENT.

(d) *Thus, an eigenvector* $\mathbf{v}$ *of* A *is in some sense a line* $\ell = Span(\{\mathbf{v}\})$ *that passes through* $\mathbf{0}$ *and* $\mathbf{v}$ *and has the property that the image of* ℓ *is either* ℓ *itself or* $\mathbf{0}$.

3. *Since the eigenvalues of* A *are roots of the characteristic equation,* A *has exactly* n *eigenvalues, including multiplicities.*

4. *If the entries of* A *are real and* $\alpha \in \sigma(A)$ *is also real then the corresponding eigenvector has real entries.*

5. *Further, if* $(\alpha, \mathbf{x})$ *is an eigenpair for* A *and* $f(A) = b_0 I + b_1 A + \cdots + b_k A^k$ *is a polynomial in* A *then* $(f(\alpha), \mathbf{x})$ *is an eigenpair for* $f(A)$.

Almost all books in mathematics differentiate between characteristic value and eigenvalue as the ideas change when one moves from complex numbers to any other scalar field. We give the following example for clarity.

Remark 6.1.6. *Let* $A \in \mathbb{M}_2(\mathbb{F})$. *Then,* A *induces a map* $T \in \mathcal{L}(\mathbb{F}^2)$ *defined by* $T(\mathbf{x}) = A\mathbf{x}$, *for all* $\mathbf{x} \in \mathbb{F}^2$. *We use this idea to understand the difference.*

1. *Let* $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. *Then,* $p_A(\lambda) = \lambda^2 + 1$. *So,* $\pm i$ *are the roots of* $P(\lambda) = 0$ *in* $\mathbb{C}$. *Hence,*

   (a) A *has* $(i, (1, i)^T)$ *and* $(-i, (i, 1)^T)$ *as eigen-pairs or characteristic-pairs.*

   (b) A *has no characteristic value over* $\mathbb{R}$.

2. *Let* $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$. *Then,* $2 \pm \sqrt{3}$ *are the roots of the characteristic equation. Hence,*

   (a) A *has characteristic values or eigenvalues over* $\mathbb{R}$.

   (b) A *has no characteristic value over* $\mathbb{Q}$.

Let us look at some more examples.

Example 6.1.7. 1. Let $A = \text{diag}(d_1, \ldots, d_n)$ with $d_i \in \mathbb{C}, 1 \le i \le n$. Then, $p(\lambda) = \prod\limits_{i=1}^{n} (\lambda - d_i)$ and thus verify that $(d_1, \mathbf{e}_1), \ldots, (d_n, \mathbf{e}_n)$ are the eigen-pairs.

2. Let $A = (a_{ij})$ be an $n \times n$ triangular matrix. Then, $p(\lambda) = \prod\limits_{i=1}^{n} (\lambda - a_{ii})$ and thus verify that $\sigma(A) = \{a_{11}, a_{22}, \ldots, a_{nn}\}$. What can you say about the eigen-vectors of an upper triangular matrix if the diagonal entries are all distinct?

3. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then, $p(\lambda) = (1 - \lambda)^2$. Hence, $\sigma(A) = \{1, 1\}$. But the complete solution of the system $(A - I_2)\mathbf{x} = \mathbf{0}$ equals $\mathbf{x} = c\mathbf{e}_1$, for $c \in \mathbb{C}$. Hence using Remark 6.1.5.2, $\mathbf{e}_1$ is an eigenvector. Therefore, 1 IS A REPEATED EIGENVALUE WHEREAS THERE IS ONLY ONE EIGENVECTOR.

4. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then, 1 is a repeated eigenvalue of A. In this case, $(A - I_2)\mathbf{x} = \mathbf{0}$ has a solution for every $\mathbf{x} \in \mathbb{C}^2$. Hence, any two LINEARLY INDEPENDENT vectors $\mathbf{x}^T, \mathbf{y}^T \in \mathbb{C}^2$

gives $(1,\mathbf{x})$ and $(1,\mathbf{y})$ as the two eigen-pairs for A. In general, if $S = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ is a basis of $\mathbb{C}^n$ then $(1,\mathbf{x}_1),\ldots,(1,\mathbf{x}_n)$ are eigen-pairs of I_n, the identity matrix.

5. Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Then, $\left(1+i, \begin{bmatrix} i \\ 1 \end{bmatrix}\right)$ and $\left(1-i, \begin{bmatrix} 1 \\ i \end{bmatrix}\right)$ are the eigen-pairs of A.

6. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Then, $\sigma(A) = \{0,0,0\}$ with $\mathbf{e}_1$ as the only eigenvector.

7. Let $A = \left[\begin{array}{ccc|cc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$. Then, $\sigma(A) = \{0,0,0,0,0\}$. Note that $A\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{0}$ implies $x_2 = 0 = x_3 = x_5$. Thus, $\mathbf{e}_1$ and $\mathbf{e}_4$ are the only eigenvectors. Note that the diagonal blocks of A are nilpotent matrices.

EXERCISE **6.1.8.** *1. Let $A \in \mathbb{M}_n(\mathbb{R})$. Then, prove that*

(a) if $\alpha \in \sigma(A)$ then $\alpha^k \in \sigma(A^k)$, for all $k \in \mathbb{N}$.

(b) if A is invertible and $\alpha \in \sigma(A)$ then $\alpha^k \in \sigma(A^k)$, for all $k \in \mathbb{Z}$.

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 $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$ with $\sum\limits_{j=1}^{n} a_{ij} = a$, for all $1 \le i \le n$. Then, prove that a is an eigenvalue of A with corresponding eigenvector $\mathbf{1} = [1,1,\ldots,1]^T$.

4. 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.

Ans: $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$. Then $0 \in \sigma(A)$. Verify that $\mathbf{1}^T A = 0\mathbf{1}^T$ and $A\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

5. Prove that $\lambda \in \mathbb{C}$ is an eigenvalue of A if and only if $\overline{\lambda} \in \mathbb{C}$ is an eigenvalue of A^*.

6. Let A be an idempotent matrix. Then, prove that its eigenvalues are either 0 *or* 1 *or both.*

7. Let A be a nilpotent matrix. Then, prove that its eigenvalues are all 0.

8. Let $J = \mathbf{1}\mathbf{1}^T \in \mathbb{M}_n(\mathbb{C})$. Then, J is a matrix with each entry 1. *Show that*

(a) $(n,\mathbf{1})$ is an eigenpair for J.

(b) $0 \in \sigma(J)$ with multiplicity $n-1$. Find a set of $n-1$ linearly independent eigenvectors for $0 \in \sigma(J)$.

9. *Let $B \in \mathbb{M}_n(\mathbb{C})$ and $C \in \mathbb{M}_m(\mathbb{C})$. Now, define the* **Direct Sum** $B \oplus C = \begin{bmatrix} B & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix}$. *Then, prove that*

   (a) *if $(\alpha, \mathbf{x})$ is an eigen-pair for B then $\left(\alpha, \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix}\right)$ is an eigen-pair for $B \oplus C$.*

   (b) *if $(\beta, \mathbf{y})$ is an eigen-pair for C then $\left(\beta, \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}\right)$ is an eigen-pair for $B \oplus C$.*

Definition 6.1.9. Let $A \in \mathcal{L}(\mathbb{C}^n)$. Then, a vector $\mathbf{y} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ satisfying $\mathbf{y}^* A = \lambda \mathbf{y}^*$ is called a **left eigenvector** of A for λ.

Example 6.1.10.

1. Let $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$. Then, $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a left eigenvector of A corresponding to the eigenvalue 0 and $\left(0, \mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$ is a (right) eigenpair of A.

2. Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$. Then, $\left(0, \mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$ and $\left(3, \mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)$ are (right) eigen-pairs of A. Also, $\left(3, \mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ and $\left(0, \mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)$ are left eigen-pairs of A. Note that $\mathbf{x}$ is orthogonal to $\mathbf{u}$ and $\mathbf{y}$ is orthogonal to $\mathbf{v}$. This is true in general and is proved next.

3. Let S be a nonsingular matrix such that its columns are left eigenvectors of A. Then, prove that the columns of $(S^*)^{-1}$ are right eigenvectors of A.

   Ans: We have $S^* A = \Lambda S^*$ and hence $A(S^*)^{-1} = (S^*)^{-1}\Lambda$.

Theorem 6.1.11. **[Principle of bi-orthogonality]** *Let $(\lambda, \mathbf{x})$ be a (right) eigenpair and $(\mu, \mathbf{y})$ be a left eigenpair of A, where $\lambda \neq \mu$. Then, $\mathbf{y}$ is orthogonal to $\mathbf{x}$.*

Proof. Verify that $\mu \mathbf{y}^* \mathbf{x} = (\mathbf{y}^* A)\mathbf{x} = \mathbf{y}^*(\lambda \mathbf{x}) = \lambda \mathbf{y}^* \mathbf{x}$. Thus, $\mathbf{y}^* \mathbf{x} = 0$. ∎

EXERCISE **6.1.12.** *Let $A\mathbf{x} = \lambda \mathbf{x}$ and $\mathbf{x}^* A = \mu \mathbf{x}^*$. Then $\mu = \lambda$.*

Ans: Note $\lambda(\mathbf{x}^*\mathbf{x}) = \mathbf{x}^*(\lambda \mathbf{x}) = \mathbf{x}^*(A\mathbf{x}) = (\mathbf{x}^* A)\mathbf{x} = (\mu \mathbf{x}^*)\mathbf{x} = \mu(\mathbf{x}^*\mathbf{x})$.

Definition 6.1.13. **[Eigenvalues of a linear Operator]** Let $T \in \mathcal{L}(\mathbb{C}^n)$. Then, $\alpha \in \mathcal{C}$ is called an **eigenvalue** of T if there exists $\mathbf{v} \in \mathbb{C}^n$ with $\mathbf{v} \neq \mathbf{0}$ such that $T(\mathbf{v}) = \alpha \mathbf{v}$.

Proposition 6.1.14. *Let $T \in L(\mathbb{C}^n)$ and let $\mathcal{B}$ be an ordered basis in $\mathbb{C}^n$. Then, $(\alpha, \mathbf{v})$ is an eigenpair for T if and only if $(\alpha, [\mathbf{v}]_{\mathcal{B}})$ is an eigenpair of $A = T[\mathcal{B}, \mathcal{B}]$.*

Proof. Note that, by definition, $T(\mathbf{v}) = \alpha \mathbf{v}$ if and only if $[Tv]_{\mathcal{B}} = [\alpha \mathbf{v}]_{\mathcal{B}}$. Or equivalently, $\alpha \in \sigma(T)$ if and only if $A[\mathbf{v}]_{\mathcal{B}} = \alpha[\mathbf{v}]_{\mathcal{B}}$. Thus, the required result follows. ∎

Remark 6.1.15. **[A linear operator on an infinite dimensional space may not have any eigenvalue]** *Let $\mathbb{V}$ be the space of all real sequences (see Example 3.1.4.9). Now, define a linear operator $T \in \mathcal{L}(\mathbb{V})$ by*

$$T(a_0, a_1, \ldots) = (0, a_1, a_2, \ldots).$$

We now show that T doesn't have any eigenvalue.

Solution: *Let if possible α be an eigenvalue of T with corresponding eigenvector $\mathbf{x} = (x_1, x_2, \ldots)$. Then, the eigen-condition $T(\mathbf{x}) = \alpha\mathbf{x}$ implies that*

$$(0, x_1, x_2, \ldots) = \alpha(x_1, x_2, \ldots) = (\alpha x_1, \alpha x_2, \ldots).$$

So, if $\alpha \neq 0$ then $x_1 = 0$ and this in turn implies that $\mathbf{x} = \mathbf{0}$, a contradiction. If $\alpha = 0$ then $(0, x_1, x_2, \ldots) = (0, 0, \ldots)$ and we again get $\mathbf{x} = \mathbf{0}$, a contradiction. Hence, the required result follows.

Theorem 6.1.16. *Let $\lambda_1, \ldots, \lambda_n$, not necessarily distinct, be the $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$. Then, $\det(A) = \prod\limits_{i=1}^{n} \lambda_i$ and $\mathsf{tr(A)} = \sum\limits_{\mathsf{i=1}}^{\mathsf{n}} \mathsf{a_{ii}} = \sum\limits_{\mathsf{i=1}}^{\mathsf{n}} \lambda_{\mathsf{i}}$.*

Proof. Since $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, by definition,

$$\det(A - xI_n) = (-1)^n \prod_{i=1}^{n} (x - \lambda_i) \tag{6.1.2}$$

is an identity in x as polynomials. Therefore, by substituting $x = 0$ in Equation (6.1.2), we get $\det(A) = (-1)^n (-1)^n \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i$. Also,

$$\begin{aligned} \det(A - xI_n) &= \begin{bmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{bmatrix} && (6.1.3) \\ &= a_0 - xa_1 + \cdots + (-1)^{n-1} x^{n-1} a_{n-1} + (-1)^n x^n && (6.1.4) \end{aligned}$$

for some $a_0, a_1, \ldots, a_{n-1} \in \mathbb{C}$. Then, a_{n-1}, the coefficient of $(-1)^{n-1} x^{n-1}$, comes from the term

$$(a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x).$$

So, $a_{n-1} = \sum\limits_{i=1}^{n} a_{ii} = \mathsf{tr(A)}$, the trace of A. Also, from Equation (6.1.2) and (6.1.4), we have

$$a_0 - xa_1 + \cdots + (-1)^{n-1} x^{n-1} a_{n-1} + (-1)^n x^n = (-1)^n \prod_{i=1}^{n} (x - \lambda_i).$$

Therefore, comparing the coefficient of $(-1)^{n-1} x^{n-1}$, we have

$$\mathsf{tr(A)} = \mathsf{a_{n-1}} = (-1) \left\{ (-1) \sum_{\mathsf{i=1}}^{\mathsf{n}} \lambda_{\mathsf{i}} \right\} = \sum_{\mathsf{i=1}}^{\mathsf{n}} \lambda_{\mathsf{i}}.$$

Hence, we get the required result. ■

EXERCISE **6.1.17.**

1. *Let A be a 3×3 orthogonal matrix ($AA^T = I$). If $\det(A) = 1$, then prove that there exists $\mathbf{v} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ such that $A\mathbf{v} = \mathbf{v}$.*

2. *Let $A \in \mathbb{M}_{2n+1}(\mathbb{R})$ with $A^T = -A$. Then, prove that 0 is an eigenvalue of A.*

3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is invertible if and only if 0 is not an eigenvalue of A.*

4. *Let $A \in \mathbb{M}_n(\mathbb{C})$ satisfy $\|A\mathbf{x}\| \leq \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$. Then, prove that if $\alpha \in \mathbb{C}$ with $\mid \alpha \mid > 1$ then $A - \alpha I$ is invertible.*

6.1.1 Spectrum of a Matrix

Definition 6.1.18. [**Algebraic, Geometric Multiplicity**] Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,

1. the multiplicity of $\alpha \in \sigma(A)$ is called the **algebraic multiplicity** of A, denoted $\text{ALG.MUL}_\alpha(A)$.
2. for $\alpha \in \sigma(A)$, $\dim(\text{NULL}(A-\alpha I))$ is called the **geometric multiplicity** of A, $\text{GEO.MUL}_\alpha(A)$.

We now state the following observations.

Remark 6.1.19. *Let $A \in \mathbb{M}_n(\mathbb{C})$.*

1. *Then, for each $\alpha \in \sigma(A)$, using Theorem 2.4.4 $\dim(\text{NULL}(A-\alpha I)) \geq 1$. So, we have at least one eigenvector.*
2. *If the algebraic multiplicity of $\alpha \in \sigma(A)$ is $r \geq 2$ then the Example 6.1.7.7 implies that we need not have r linearly independent eigenvectors.*

Theorem 6.1.20. *Let A and B be two similar matrices. Then,*

1. *$\alpha \in \sigma(A)$ if and only if $\alpha \in \sigma(B)$.*
2. *for each $\alpha \in \sigma(A)$, $\text{ALG.MUL}_\alpha(A) = \text{ALG.MUL}_\alpha(B)$ and $\text{GEO.MUL}_\alpha(A) = \text{GEO.MUL}_\alpha(B)$.*

Proof. Since A and B are similar, there exists an invertible matrix S such that $A = SBS^{-1}$. So, $\alpha \in \sigma(A)$ if and only if $\alpha \in \sigma(B)$ as

$$\begin{aligned}\det(A - xI) &= \det(SBS^{-1} - xI) = \det\left(S(B - xI)S^{-1}\right) \\ &= \det(S)\det(B - xI)\det(A^{-1}) = \det(B - xI).\end{aligned} \tag{6.1.5}$$

Note that Equation (6.1.5) also implies that $\text{ALG.MUL}_\alpha(A) = \text{ALG.MUL}_\alpha(B)$. We will now show that $\text{GEO.MUL}_\alpha(A) = \text{GEO.MUL}_\alpha(B)$.

So, let $Q_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\text{NULL}(A - \alpha I)$. Then, $B = SAS^{-1}$ implies that $Q_2 = \{S\mathbf{v}_1, \ldots, S\mathbf{v}_k\} \subseteq \text{NULL}(B - \alpha I)$. Since Q_1 is linearly independent and S is invertible, we get Q_2 is linearly independent. So, $\text{GEO.MUL}_\alpha(A) \leq \text{GEO.MUL}_\alpha(B)$. Now, we can start with eigenvectors of B and use similar arguments to get $\text{GEO.MUL}_\alpha(B) \leq \text{GEO.MUL}_\alpha(A)$ and hence the required result follows. ∎

Remark 6.1.21.

1. *Let $A = S^{-1}BS$. Then, from the proof of Theorem 6.1.20, we see that $\mathbf{x}$ is an eigenvector of A for λ if and only if $S\mathbf{x}$ is an eigenvector of B for λ.*
2. *Let A and B be two similar matrices then $\sigma(A) = \sigma(B)$. But, the converse is not true. For example, take $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.*
3. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, for any invertible matrix B, the matrices AB and $BA = B(AB)B^{-1}$ are similar. Hence, in this case the matrices AB and BA have*
   (a) *the same set of eigenvalues.*
   (b) *$\text{ALG.MUL}_\alpha(AB) = \text{ALG.MUL}_\alpha(BA)$, for each $\alpha \in \sigma(A)$.*
   (c) *$\text{GEO.MUL}_\alpha(AB) = \text{GEO.MUL}_\alpha(BA)$, for each $\alpha \in \sigma(A)$.*

We will now give a relation between the geometric multiplicity and the algebraic multiplicity.

Theorem 6.1.22. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, for $\alpha \in \sigma(A)$, $\text{Geo.Mul}_\alpha(A) \le \text{Alg.Mul}_\alpha(A)$.*

Proof. Let $\text{Geo.Mul}_\alpha(A) = k$. Suppose $Q_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthonormal basis of $\text{Null}(A - \alpha I)$. Extend Q_1 to get $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ as an orthonormal basis of $\mathbb{C}^n$. Put $P = [\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n]$. Then, $P^* = P^{-1}$ and

$$
\begin{aligned}
P^*AP &= P^*[A\mathbf{v}_1, \ldots, A\mathbf{v}_k, A\mathbf{v}_{k+1}, \ldots, A\mathbf{v}_n] \\
&= \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_k^* \\ \mathbf{v}_{k+1}^* \\ \vdots \\ \mathbf{v}_n^* \end{bmatrix} [\alpha\mathbf{v}_1, \ldots, \alpha\mathbf{v}_k, *, \ldots, *] = \left[\begin{array}{ccc|ccc} \alpha & \cdots & 0 & * & \cdots & * \\ 0 & \ddots & 0 & * & \cdots & * \\ 0 & \cdots & \alpha & * & \cdots & * \\ \hline 0 & \cdots & 0 & * & \cdots & * \\ \vdots & & & & & \\ 0 & \cdots & 0 & * & \cdots & * \end{array}\right].
\end{aligned}
$$

Now, if we denote the lower diagonal submatrix as D then

$$
P_A(x) = \det(A - xI) = \det(P^*AP - xI) = (\alpha - x)^k \det(D - xI). \tag{6.1.6}
$$

So, $\text{Alg.Mul}_\alpha(A) = \text{Alg.Mul}_\alpha(P^*AP) \ge k = \text{Geo.Mul}_\alpha(A)$. ∎

Remark 6.1.23. *Note that in the proof of Theorem 6.1.22, the remaining eigenvalues of A are the eigenvalues of D (see Equation (6.1.6)). This technique is called* **deflation**.

Exercise **6.1.24.** 1. *Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix}$. Notice that $\mathbf{x}_1 = \frac{1}{\sqrt{3}}\mathbf{1}$ is an eigenvector for A. Find an ordered basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ of $\mathbb{C}^3$. Put $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$. Compute $X^{-1}AX$ to get a block-triangular matrix. Can you now find the remaining eigenvalues of A?*

Ans: $X^{-1}AX = \begin{bmatrix} 6 & -\frac{1}{\sqrt{2}} & -\frac{3}{\sqrt{6}} \\ 0 & -1 & -\sqrt{3} \\ 0 & \frac{1}{\sqrt{3}} & -1 \end{bmatrix}$. The other eigenvalues are $-1 \pm i$.

2. *Let $A \in \mathbb{M}_{m\times n}(\mathbb{R})$ and $B \in \mathbb{M}_{n\times m}(\mathbb{R})$.*
   (a) *If $\alpha \in \sigma(AB)$ and $\alpha \ne 0$ then*
      i. *$\alpha \in \sigma(BA)$.*
      ii. *$\text{Alg.Mul}_\alpha(AB) = \text{Alg.Mul}_\alpha(BA)$.*
      iii. *$\text{Geo.Mul}_\alpha(AB) = \text{Geo.Mul}_\alpha(BA)$.*

      Ans: Verify that $\begin{bmatrix} I & \mathbf{0} \\ -A & I \end{bmatrix} \begin{bmatrix} \mathbf{0} & B \\ \mathbf{0} & AB \end{bmatrix} = \begin{bmatrix} BA & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ -A & I \end{bmatrix} = \begin{bmatrix} \mathbf{0} & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$

      Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be k linearly independent eigenvalues of AB corresponding to α. Then, $\{B\mathbf{u}_1, \ldots, B\mathbf{u}_k\}$ are k linearly independent eigenvalues of BA corresponding to α as $\alpha \ne 0$ and $\sum\limits_{j=1}^{k} c_j B\mathbf{u}_j = \mathbf{0}$ implies $\sum\limits_{j=1}^{k} c_j \mathbf{u}_j = \mathbf{0}$, a contradiction.

   (b) *If $0 \in \sigma(AB)$ and $n = m$ then $\text{Alg.Mul}_0(AB) = \text{Alg.Mul}_0(BA)$ as there are n eigenvalues, counted with multiplicity.*

(c) *Give an example to show that* $\text{Geo.Mul}_0(AB)$ *need not equal* $\text{Geo.Mul}_0(BA)$ *even when* $n = m$.

3. *Let* $A \in \mathbb{M}_n(\mathbb{R})$ *be an invertible matrix and let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *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 6.1.24.2a].*

   (c) $1 + \alpha\lambda_0$ *is an eigenvalue of* $I + \alpha B$ *of multiplicity* 1, *for any* $\alpha \in \mathbb{R}$.

   (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.*

4. *Let* $A, B \in \mathbb{M}_2(\mathbb{R})$ *such that* $\det(A) = \det(B)$ *and* $\mathsf{tr}(\mathsf{A}) = \mathsf{tr}(\mathsf{B})$.

   (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.*

5. *Let* $A, B \in \mathbb{M}_n(\mathbb{R})$. *Also, let* $(\lambda_1, \mathbf{u})$ *and* $(\lambda_2, \mathbf{v})$ *are eigen-pairs of* A *and* B, *respectively.*

   (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* $A + B$.

6. *Let* $A \in \mathbb{M}_n(\mathbb{R})$ *be an invertible matrix with eigen-pairs* $(\lambda_1, \mathbf{u}_1), \ldots, (\lambda_n, \mathbf{u}_n)$. *Then, prove that* $\mathcal{B} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ *forms a basis of* $\mathbb{R}^n$. *If* $[\mathbf{b}]_{\mathcal{B}} = (c_1, \ldots, c_n)^T$ *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 $A \in \mathbb{M}_n(\mathbb{C})$ and let $T \in \mathcal{L}(\mathbb{C}^n)$ be defined by $T(\mathbf{x}) = A\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. In this section, we first find conditions under which one can obtain a basis $\mathcal{B}$ of $\mathbb{C}^n$ such that $T[\mathcal{B}, \mathcal{B}]$ (see Theorem 4.4.4) is a diagonal matrix. And, then it is shown that normal matrices satisfy the above conditions. To start with, we have the following definition.

Definition 6.2.1. [**Matrix Diagonalizability**] A matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix. Or equivalently, $P^{-1}AP = D \Leftrightarrow AP = PD$, for some diagonal matrix D and invertible matrix P.

Example 6.2.2. 1. Let A be an $n \times n$ diagonalizable matrix. Then, by definition, A is similar to a diagonal matrix, say $D = \text{diag}(d_1, \ldots, d_n)$. Thus, by Remark 6.1.21, $\sigma(A) = \sigma(D) = \{d_1, \ldots, d_n\}$.

2. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then, A cannot be diagonalized.

   Solution: Suppose A is diagonalizable. Then, A is similar to $D = \text{diag}(d_1, d_2)$. Thus, by Theorem 6.1.20, $\{d_1, d_2\} = \sigma(D) = \sigma(A) = \{0, 0\}$. Hence, $D = \mathbf{0}$ and therefore, $A = SDS^{-1} = \mathbf{0}$, a contradiction.

3. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$. Then, A cannot be diagonalized.

   Solution: Suppose A is diagonalizable. Then, A is similar to $D = \text{diag}(d_1, d_2, d_3)$. Thus, by Theorem 6.1.20, $\{d_1, d_2, d_3\} = \sigma(D) = \sigma(A) = \{2, 2, 2\}$. Hence, $D = 2I_3$ and therefore, $A = SDS^{-1} = 2I_3$, a contradiction.

4. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then, $\left(i, \begin{bmatrix} i \\ 1 \end{bmatrix}\right)$ and $\left(-i, \begin{bmatrix} -i \\ 1 \end{bmatrix}\right)$ are two eigen-pairs of A. Define $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$. Then, $U^*U = I_2 = UU^*$ and $U^*AU = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}$.

Theorem 6.2.3. *Let $A \in \mathbb{M}_n(\mathbb{R})$.*

1. *Let S be an invertible matrix such that $S^{-1}AS = \text{diag}(d_1, \ldots, d_n)$. Then, for $1 \le i \le n$, the i-th column of S is an eigenvector of A corresponding to d_i.*

2. *Then, A is diagonalizable if and only if A has n linearly independent eigenvectors.*

Proof. Let $S = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. Then, $AS = SD$ gives

$$[A\mathbf{u}_1, \ldots, A\mathbf{u}_n] = A[\mathbf{u}_1, \ldots, \mathbf{u}_n] = AS = SD = S\ \text{diag}(d_1, \ldots, d_n) = [d_1\mathbf{u}_1, \ldots, d_n\mathbf{u}_n].$$

Or equivalently, $A\mathbf{u}_i = d_i\mathbf{u}_i$, for $1 \le i \le n$. As S is invertible, $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are linearly independent. Hence, $(d_i, \mathbf{u}_i)$, for $1 \le i \le n$, are eigen-pairs of A. This proves Part 1 and "only if" part of Part 2.

Conversely, let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be n linearly independent eigenvectors of A corresponding to eigenvalues $\alpha_1, \ldots, \alpha_n$. Then, by Corollary 3.2.10, $S = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is non-singular and

$$AS = [A\mathbf{u}_1, \ldots, A\mathbf{u}_n] = [\alpha_1\mathbf{u}_1, \ldots, \lambda_n\mathbf{u}_n] = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \alpha_n \end{bmatrix} = SD,$$

where $D = \text{diag}(\alpha_1, \ldots, \alpha_n)$. Therefore, $S^{-1}AS = D$ and hence A is diagonalizable. ∎

Definition 6.2.4.
1. A matrix $A \in \mathbb{M}_n(\mathbb{C})$ is called **defective** if for some $\alpha \in \sigma(A)$, $\text{Geo.Mul}_\alpha(A) < \text{Alg.Mul}_\alpha(A)$.

2. A matrix $A \in \mathbb{M}_n(\mathbb{C})$ is called **non-derogatory** if $\text{Geo.Mul}_\alpha(A) = 1$, for each $\alpha \in \sigma(A)$.

As a direct consequence of Theorem 6.2.3, we obtain the following result.

Corollary 6.2.5. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,*

1. *A is non-defective if and only if A is diagonalizable.*
2. *A has distinct eigenvalues if and only if A is non-derogatory and non-defective.*

Theorem 6.2.6. *Let $(\alpha_1, \mathbf{v}_1), \ldots, (\alpha_k, \mathbf{v}_k)$ be k eigen-pairs of $A \in \mathbb{M}_n(\mathbb{C})$ with α_i's distinct. Then, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent.*

Proof. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent. Then, there exists a smallest $\ell \in \{1, \ldots, k-1\}$ and $\beta \neq \mathbf{0}$ such that $\mathbf{v}_{\ell+1} = \beta_1 \mathbf{v}_1 + \cdots + \beta_\ell \mathbf{v}_\ell$. So,

$$\alpha_{\ell+1}\mathbf{v}_{\ell+1} = \alpha_{\ell+1}\beta_1\mathbf{v}_1 + \cdots + \alpha_{\ell+1}\beta_\ell\mathbf{v}_\ell. \tag{6.2.1}$$

and

$$\alpha_{\ell+1}\mathbf{v}_{\ell+1} \;=\; A\mathbf{v}_{\ell+1} = A\left(\beta_1\mathbf{v}_1 + \cdots + \beta_\ell\mathbf{v}_\ell\right) = \alpha_1\beta_1\mathbf{v}_1 + \cdots + \alpha_\ell\beta_\ell\mathbf{v}_\ell. \tag{6.2.2}$$

Now, subtracting Equation (6.2.2) from Equation (6.2.1), we get

$$\mathbf{0} = \left(\alpha_{\ell+1} - \alpha_1\right)\beta_1\mathbf{v}_1 + \cdots + \left(\alpha_{\ell+1} - \alpha_\ell\right)\beta_\ell\mathbf{v}_\ell.$$

So, $\mathbf{v}_\ell \in LS(\mathbf{v}_1, \ldots, \mathbf{v}_{\ell-1})$, a contradiction to the choice of ℓ. Thus, the required result follows. ■

An immediate corollary of Theorem 6.2.3 and Theorem 6.2.6 is stated next without proof.

Corollary 6.2.7. *Let $A \in \mathbb{M}_n(\mathbb{C})$ have n distinct eigenvalues. Then, A is diagonalizable.*

The converse of Theorem 6.2.6 is not true as I_n has n linearly independent eigenvectors corresponding to the eigenvalue 1, repeated n times.

Corollary 6.2.8. *Let $\alpha_1, \ldots, \alpha_k$ be k distinct eigenvalues $A \in \mathbb{M}_n(\mathbb{C})$. Also, for $1 \leq i \leq k$, let $\dim(\textsc{Null}(A - \alpha_i I_n)) = n_i$. Then, A has $\sum\limits_{i=1}^{k} n_i$ linearly independent eigenvectors.*

Proof. For $1 \leq i \leq k$, let $S_i = \{\mathbf{u}_{i1}, \ldots, \mathbf{u}_{in_i}\}$ be a basis of $\textsc{Null}(A - \alpha_i I_n)$. Then, we need to prove that $\bigcup\limits_{i=1}^{k} S_i$ is linearly independent. To do so, denote $p_j(A) = \left(\prod\limits_{i=1}^{k}(A - \alpha_i I_n)\right) / (A - \alpha_j I_n)$, for $1 \leq j \leq k$. Then, note that $p_j(A)$ is a polynomial in A of degree $k-1$ and

$$p_j(A)\mathbf{u} = \begin{cases} \mathbf{0}, & \text{if } \mathbf{u} \in \textsc{Null}(A - \alpha_i I_n), \text{ for some } i \neq j \\ \prod\limits_{i \neq j}(\alpha_j - \alpha_i)\mathbf{u} & \text{if } \mathbf{u} \in \textsc{Null}(A - \alpha_j I_n) \end{cases} \tag{6.2.3}$$

So, to prove that $\bigcup\limits_{i=1}^{k} S_i$ is linearly independent, consider the linear system

$$c_{11}\mathbf{u}_{11} + \cdots + c_{1n_1}\mathbf{u}_{1n_1} + \cdots + c_{k1}\mathbf{u}_{k1} + \cdots + c_{kn_k}\mathbf{u}_{kn_k} = \mathbf{0}$$

in the variables c_{ij}'s. Now, applying the matrix $p_j(A)$ and using Equation (6.2.3), we get

$$\prod_{i \neq j}(\alpha_j - \alpha_i)\left(c_{j1}\mathbf{u}_{j1} + \cdots + c_{jn_j}\mathbf{u}_{jn_j}\right) = \mathbf{0}.$$

But $\prod\limits_{i \neq j}(\alpha_j - \alpha_i) \neq 0$ as α_i's are distinct. Hence, $c_{j1}\mathbf{u}_{j1} + \cdots + c_{jn_j}\mathbf{u}_{jn_j} = \mathbf{0}$. As S_j is a basis of $\textsc{Null}(A - \alpha_j I_n)$, we get $c_{jt} = 0$, for $1 \leq t \leq n_j$. Thus, the required result follows. ■

Corollary 6.2.9. *Let $A \in \mathbb{M}_n(\mathbb{C})$ with distinct eigenvalues $\alpha_1, \ldots, \alpha_k$. Then, A is diagonalizable if and only if* $\text{Geo.Mul}_{\alpha_i}(A) = \text{Alg.Mul}_{\alpha_i}(A)$, *for each $1 \le i \le k$.*

Proof. Let $\text{Alg.Mul}_{\alpha_i}(A) = m_i$. Then, $\sum\limits_{i=1}^{k} m_i = n$. Let $\text{Geo.Mul}_{\alpha_i}(A) = n_i$, for $1 \le i \le k$. Then, by Corollary 6.2.8 A has $\sum\limits_{i=1}^{k} n_i$ linearly independent eigenvectors. Also, by Theorem 6.1.22, $n_i \le m_i$, for $1 \le i \le m_i$.

Now, let A be diagonalizable. Then, by Theorem 6.2.3, A has n linearly independent eigenvectors. So, $n = \sum\limits_{i=1}^{k} n_i$. As $n_i \le m_i$ and $\sum\limits_{i=1}^{k} m_i = n$, we get $n_i = m_i$.

Now, assume that $\text{Geo.Mul}_{\alpha_i}(A) = \text{Alg.Mul}_{\alpha_i}(A)$, for $1 \le i \le k$. Then, for each $i, 1 \le i \le n$, A has $n_i = m_i$ linearly independent eigenvectors. Thus, A has $\sum\limits_{i=1}^{k} n_i = \sum\limits_{i=1}^{k} m_i = n$ linearly independent eigenvectors. Hence by Theorem 6.2.3, A is diagonalizable. ∎

Example 6.2.10. Let $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. Then, $\left(1, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right)$ and $\left(2, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right)$ are the only eigen-pairs. Hence, by Theorem 6.2.3, A is not diagonalizable.

Exercise **6.2.11.**

1. *Let A be diagonalizable. Then, prove that $A + \alpha I$ is diagonalizable for every $\alpha \in \mathbb{C}$.*
2. *Let A be an strictly upper triangular matrix. Then, prove that A is not diagonalizable.*
3. *Let A be an $n \times n$ matrix with $\lambda \in \sigma(A)$ with* $\text{alg.mul}_\lambda(A) = m$. *If* $\text{Rank}[A - \lambda I] \neq n - m$ *then prove that A is not diagonalizable.*
4. *If $\sigma(A) = \sigma(B)$ and both A and B are diagonalizable then prove that A is similar to B. That is, they are two basis representation of the same linear transformation.*
5. *Let A and B be two similar matrices such that A is diagonalizable. Prove that B is diagonalizable.*
6. *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.*
7. *Is the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ diagonalizable?*
8. *Let J_n be an $n \times n$ matrix with all entries 1. Then,* $\text{Geo.Mul}_1(J_n) = \text{Alg.Mul}_1(J_n) = 1$ *and* $\text{Geo.Mul}_0(J_n) = \text{Alg.Mul}_0(J_n) = n - 1$.
9. *Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{R})$, where $a_{ij} = a$, if $i = j$ and b, otherwise. Then, verify that $A = (a - b)I_n + bJ_n$. Hence, or otherwise determine the eigenvalues and eigenvectors of J_n. Is A diagonalizable?*

10. *Let $T : \mathbb{R}^5 \longrightarrow \mathbb{R}^5$ be a linear operator with* $\text{Rank}(T - I) = 3$ *and*

$$\text{Null}(T) = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 \mid x_1 + x_4 + x_5 = 0,\ x_2 + x_3 = 0\}.$$

   (a) *Determine the eigenvalues of T?*

   (b) *For each distinct eigenvalue α of T, determine* $\text{Geo.Mul}_\alpha(T)$.

   (c) *Is T diagonalizable? Justify your answer.*

11. *Let $A \in \mathbb{M}_n(\mathbb{R})$ with $A \neq \mathbf{0}$ but $A^2 = \mathbf{0}$. Prove that A cannot be diagonalized.*

12. *Are the following matrices diagonalizable?*

$$i) \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 2 & 3 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix},\quad ii) \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix},\quad iii) \begin{bmatrix} 1 & -3 & 3 \\ 0 & -5 & 6 \\ 0 & -3 & 4 \end{bmatrix} \text{ and } iv) \begin{bmatrix} 2 & i \\ i & 0 \end{bmatrix}.$$

13. *Let $A \in \mathbb{M}_n(\mathbb{C})$.*

   (a) *Then, prove that* $\text{Rank}(A) = 1$ *if and only if $A = \mathbf{x}\mathbf{y}^*$, for some non-zero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.*

   (b) *If* $\text{Rank}(A) = 1$ *then*

      i. *A has at most one nonzero eigenvalue of algebraic multiplicity* 1.

      ii. *find this eigenvalue and its geometric multiplicity.*

      iii. *when is A diagonalizable?*

Ans: (a) A has a nonzero row, call it y^*. Other rows are scalar multiples of this row. So $A = \mathbf{x}\mathbf{y}^*$.
(b.i) Note that $A\mathbf{x} = (\mathbf{x}\mathbf{y}^*)\mathbf{x} = (\mathbf{y}^*\mathbf{x})\mathbf{x}$. Thus, $\alpha = \mathbf{y}^*\mathbf{x} \in \sigma(A)$.
(b.ii) Since $\mathbf{y} \neq \mathbf{0}$, let $\{\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}\}$ be an orthonormal basis of $\mathbf{y}^\perp$. Then, $A\mathbf{z}_i = \mathbf{x}\mathbf{y}^*\mathbf{z}_i = \mathbf{0}$, hence the geometric multiplicity of 0 is at least $n-1$. So, if $\mathbf{y}^*\mathbf{x} \neq 0$, then the geometric multiplicity of $\mathbf{y}^*\mathbf{x}$ is 1. If $\mathbf{y}^*\mathbf{x} = 0$, then the geometric multiplicity of 0 could be $n-1$ or n.
(b.iii) A is not diagonalizable if and only if $\mathbf{y}^*\mathbf{x} = 0$ and the geometric multiplicity of the eigenvalue 0 is $n-1$. Or equivalently, A is diagonalizable if and only if tr(A) $\neq$ 0.

14. *Let $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ such that $\{\mathbf{u}, \mathbf{v}\}$ is a linearly independent set. Define $A = \mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T$.*

   (a) *Then prove that A is a symmetric matrix.*

   (b) *Then prove that* $\dim(\text{Ker}(A)) = n - 2$.

   (c) *Then $0 \in \sigma(A)$ and has multiplicity $n-2$.*

   (d) *Determine the other eigenvalues of A.*

   Ans: (a) $A^T = (\mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T)^T = \mathbf{v}\mathbf{u}^T + \mathbf{u}\mathbf{v}^T = A$. Also, $A = [\mathbf{u}, \mathbf{v}] \begin{bmatrix} \mathbf{v}^T \\ \mathbf{u}^T \end{bmatrix}$.
   (b) Let $\mathbf{w} \in \{\mathbf{u}, \mathbf{v}\}^\perp$. Then $A\mathbf{w} = \mathbf{0}$ and $\dim\left(\{\mathbf{u}, \mathbf{v}\}^\perp\right) = n - 2$.
   (c) Hence, 0 is an eigenvalue with multiplicity $n-2$.
   (d) As the eigenvalues of AB and BA are same (except for the multiplicity of the

eigenvalue 0), consider the 2×2 matrix $\begin{bmatrix} \mathbf{v}^T\mathbf{u} & \mathbf{v}^T\mathbf{v} \\ \mathbf{u}^T\mathbf{u} & \mathbf{u}^T\mathbf{v} \end{bmatrix}$. The eigenvalue of this 2×2 matrix gives the other eigenvalues.

15. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If $\mathsf{Rank}(A) = k$ then there exists $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{C}^n$ such that $A = \sum\limits_{i=1}^{k} \mathbf{x}_i \mathbf{y}_i^*$. Is the converse true?*

6.2.1 Schur's Unitary Triangularization

We now prove one of the most important results in diagonalization, called the Schur's Lemma or the Schur's unitary triangularization.

Lemma 6.2.12 (Schur's unitary triangularization (SUT))**.** *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, there exists a unitary matrix U such that A is similar to an upper triangular matrix. Further, if $A \in \mathbb{M}_n(\mathbb{R})$ and $\sigma(A)$ have real entries then U is a real orthogonal matrix.*

Proof. We prove the result by induction on n. The result is clearly true for $n = 1$. So, let $n > 1$ and assume the result to be true for $k < n$ and prove it for n.

Let $(\lambda_1, \mathbf{x}_1)$ be an eigen-pair of A with $\|\mathbf{x}_1\| = 1$. Now, extend it to form an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{u}_n\}$ of $\mathbb{C}^n$ and define $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{u}_n]$. Then, X is a unitary matrix and

$$X^*AX = X^*[A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n] = \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \\ \vdots \\ \mathbf{x}_n^* \end{bmatrix} [\lambda_1\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n] = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & B \end{bmatrix}, \tag{6.2.4}$$

where $B \in \mathbb{M}_{n-1}(\mathbb{C})$. Now, by induction hypothesis there exists a unitary matrix $U \in \mathbb{M}_{n-1}(\mathbb{C})$ such that $U^*BU = T$ is an upper triangular matrix. Define $\widehat{U} = X \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{bmatrix}$. Then, using Exercise 5.4.8.10, the matrix $\widehat{U}$ is unitary and

$$\begin{aligned} \left(\widehat{U}\right)^* A\widehat{U} &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U^* \end{bmatrix} X^*AX \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U^* \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & B \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & U^*B \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & U^*BU \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & T \end{bmatrix}. \end{aligned}$$

Since T is upper triangular, $\begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & T \end{bmatrix}$ is upper triangular.

Further, if $A \in \mathbb{M}_n(\mathbb{R})$ and $\sigma(A)$ has real entries then $\mathbf{x}_1 \in \mathbb{R}^n$ with $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$. Now, one uses induction once again to get the required result. ∎

Remark 6.2.13. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, by Schur's Lemma there exists a unitary matrix U such that $U^*AU = T = [t_{ij}]$, a triangular matrix. Thus,*

$$\{\alpha_1, \ldots, \alpha_n\} = \sigma(A) = \sigma(U^*AU) = \{t_{11}, \ldots, t_{nn}\}. \tag{6.2.5}$$

Furthermore, we can get the α_i's in the diagonal of T in any prescribed order.

Definition 6.2.14. **[Unitary Equivalence]** Let $A, B \in \mathbb{M}_n(\mathbb{C})$. Then, A and B are said to be **unitarily equivalent/similar** if there exists a unitary matrix U such that $A = U^*BU$.

Remark 6.2.15. *We know that if two matrices are unitarily equivalent then they are necessarily similar as $U^* = U^{-1}$, for every unitary matrix U. But, similarity doesn't imply unitary equivalence (see Exercise 6.2.17.6). In numerical calculations, unitary transformations are preferred as compared to similarity transformations due to the following main reasons:*

1. *Exercise 5.4.8.5g implies that $\|A\mathbf{x}\| = \|\mathbf{x}\|$, whenever A is a normal matrix. This need not be true under a similarity change of basis.*
2. *As $U^{-1} = U^*$, for a unitary matrix, unitary equivalence is computationally simpler.*
3. *Also, computation of "conjugate transpose" doesn't create round-off error in calculation.*

Example 6.2.16. Consider the two matrices $A = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Then, we show that they are similar but not unitarily similar.

Solution: Note that $\sigma(A) = \sigma(B) = \{1, 2\}$. As the eigenvalues are distinct, by Theorem 6.2.7, the matrices A and B are diagonalizable and hence there exists invertible matrices S and T such that $A = S\Lambda S^{-1}$, $B = T\Lambda T^{-1}$, where $\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Thus, $A = ST^{-1}B(ST^{-1})^{-1}$. That is, A and B are similar. But, $\sum |a_{ij}|^2 \neq \sum |b_{ij}|^2$ and hence by Exercise 5.4.8.11, they cannot be unitarily similar.

EXERCISE **6.2.17.**

1. *If A is unitarily similar to an upper triangular matrix $T = [t_{ij}]$ then prove that $\sum\limits_{i<j} |t_{ij}|^2 = \mathsf{tr}(\mathsf{A}^*\mathsf{A}) - \sum |\lambda_i|^2$.*
2. *Use the exercises given below to conclude that the upper triangular matrix obtained in the "Schur's Lemma" need not be unique.*
   (a) *Prove that $B = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 1 & 3\sqrt{2} \\ 0 & 1 & -\sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}$ are unitarily equivalent.*
   (b) *Prove that $D = \begin{bmatrix} 2 & 0 & 3\sqrt{2} \\ 1 & 1 & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$ and $E = \begin{bmatrix} 2 & 0 & 3\sqrt{2} \\ -1 & 1 & -\sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$ are unitarily equivalent.*
   (c) *Let $A_1 = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. Then, prove that*
      i. *A_1 and D are unitarily equivalent.*
      ii. *A_2 and B are unitarily equivalent.*
      iii. *Do the above results contradict Exercise 5.4.8.5c? Give reasons for your answer.*
3. *Prove that $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 & \sqrt{2} \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ are unitarily equivalent.*

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 $A \in \mathbb{M}_n(\mathbb{C})$. Then, Prove that if $\mathbf{x}^* A \mathbf{x} = 0$, for all $\mathbf{x} \in \mathbb{C}^n$, then $A = \mathbf{0}$. Do these results hold for arbitrary matrices?*

6. *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 $A = U^* B U$?*

   Ans: Take $S = \begin{bmatrix} 3 & 2 \\ -2 & 0 \end{bmatrix}$. Then $S^{-1}BS = A$. There doesn't exist an unitary matrix as the sum of the squares of the matrix entries are NOT equal.

We now use Lemma 6.2.12 to give another proof of Theorem 6.1.16.

Corollary 6.2.18. *Let $A \in \mathbb{M}_n(\mathbb{C})$. If $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\}$ then $\det(A) = \prod\limits_{i=1}^{n} \alpha_i$ and $\mathsf{tr(A)} = \sum\limits_{i=1}^{n} \alpha_i$.*

Proof. By Schur's Lemma there exists a unitary matrix U such that $U^*AU = T = [t_{ij}]$, a triangular matrix. By Remark 6.2.13, $\sigma(A) = \sigma(T)$. Hence, $\det(A) = \det(T) = \prod\limits_{i=1}^{n} t_{ii} = \prod\limits_{i=1}^{n} \alpha_i$ and $\mathsf{tr(A)} = \mathsf{tr(A(UU^*))} = \mathsf{tr(U^*(AU))} = \mathsf{tr(T)} = \sum\limits_{i=1}^{n} t_{ii} = \sum\limits_{i=1}^{n} \alpha_i$. ■

6.2.2 Diagonalizability of some Special Matrices

We now use Schur's unitary triangularization Lemma to state the main theorem of this subsection. Also, recall that A is said to be a normal matrix if $AA^* = A^*A$.

Theorem 6.2.19 (Spectral Theorem for Normal Matrices)**.** *Let $A \in \mathbb{M}_n(\mathbb{C})$. If A is a normal matrix then there exists a unitary matrix U such that $U^*AU = \text{diag}(\alpha_1, \ldots, \alpha_n)$.*

Proof. By Schur's Lemma there exists a unitary matrix U such that $U^*AU = T = [t_{ij}]$, a triangular matrix. Since A is a normal

$$T^*T = (U^*AU)^*(U^*AU) = U^*A^*AU = U^*AA^*U = (U^*AU)(U^*AU)^* = TT^*.$$

Thus, we see that T is an upper triangular matrix with $T^*T = TT^*$. Thus, by Exercise 1.2.10.3, T is a diagonal matrix and this completes the proof. ■

EXERCISE **6.2.20.** *Let $A \in \mathbb{M}_n(\mathbb{C})$. If A is either a Hermitian, skew-Hermitian or Unitary matrix then A is a normal matrix.*

We re-write Theorem 6.2.19 in another form to indicate that A can be decomposed into linear combination of orthogonal projectors onto eigen-spaces. Thus, it is independent of the choice of eigenvectors.

Remark 6.2.21. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a normal matrix with eigenvalues $\alpha_1, \ldots, \alpha_n$.*

1. *Then, there exists a unitary matrix $U = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ such that*

(a) $I_n = \mathbf{u}_1\mathbf{u}_1^* + \cdots + \mathbf{u}_n\mathbf{u}_n^*$.

(b) *the columns of U form a set of orthonormal eigenvectors for A (use Theorem 6.2.3).*

(c) $A = A \cdot I_n = A\left(\mathbf{u}_1\mathbf{u}_1^* + \cdots + \mathbf{u}_n\mathbf{u}_n^*\right) = \alpha_1\mathbf{u}_1\mathbf{u}_1^* + \cdots + \alpha_n\mathbf{u}_n\mathbf{u}_n^*$.

2. *Let $\alpha_1, \ldots, \alpha_k$ be the distinct eigenvalues of A. Also, let $W_i = \textsc{Null}(A - \alpha_i I_n)$, for $1 \leq i \leq k$, be the corresponding eigen-spaces.*

   (a) *Then, we can group the $\mathbf{u}_i$'s such that they form an orthonormal basis of W_i, for $1 \leq i \leq k$. Hence, $\mathbb{C}^n = W_1 \oplus \cdots \oplus W_k$.*

   (b) *If P_{α_i} is the orthogonal projector onto W_i, for $1 \leq i \leq k$ then $A = \alpha_1 P_1 + \cdots + \alpha_k P_k$. Thus, A depends only on eigen-spaces and not on the computed eigenvectors.*

We now give the spectral theorem for Hermitian matrices.

Theorem 6.2.22. **[Spectral Theorem for Hermitian Matrices]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then,*

1. *the eigenvalues α_i, for $1 \leq i \leq n$, of A are real.*

2. *there exists a unitary matrix U, say $U = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ such that*

   (a) $I_n = \mathbf{u}_1\mathbf{u}_1^* + \cdots + \mathbf{u}_n\mathbf{u}_n^*$.

   (b) $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *forms a set of orthonormal eigenvectors for A.*

   (c) $A = \alpha_1\mathbf{u}_1\mathbf{u}_1^* + \cdots + \alpha_n\mathbf{u}_n\mathbf{u}_n^*$, *or equivalently, $U^*AU = D$, where $D = \ diag(\alpha_1, \ldots, \alpha_n)$.*

Proof. The second part is immediate from Theorem 6.2.19 as Hermitian matrices are also normal matrices. For Part 1, let $(\alpha, \mathbf{x})$ be an eigen-pair. Then, $A\mathbf{x} = \alpha\mathbf{x}$. As A is Hermitian $A^* = A$. Thus, $\mathbf{x}^*A = \mathbf{x}^*A^* = (A\mathbf{x})^* = (\alpha\mathbf{x})^* = \overline{\alpha}\mathbf{x}^*$. Hence, using $\mathbf{x}^*A = \overline{\alpha}\mathbf{x}^*$, we get

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

As $\mathbf{x}$ is an eigenvector, $\mathbf{x} \neq \mathbf{0}$. Hence, $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x} \neq 0$. Thus $\alpha = \overline{\alpha}$, *i.e.*, $\alpha \in \mathbb{R}$. ∎

As an immediate corollary of Theorem 6.2.22 and the second part of Lemma 6.2.12, we give the following result without proof.

Corollary 6.2.23. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be symmetric. Then, $A = U\ diag(\alpha_1, \ldots, \alpha_n)\ U^*$, where*

1. *the α_i's are all real,*

2. *the columns of U can be chosen to have real entries,*

3. *the eigenvectors that correspond to the columns of U form an orthonormal basis of $\mathbb{R}^n$.*

Exercise **6.2.24.**

1. *Let A be a skew-symmetric matrix. Then, the eigenvalues of A are either zero or purely imaginary and A is unitarily diagonalizable.*

2. *Let A be a skew-Hermitian matrix. Then, A is unitarily diagonalizable.*

3. *Characterize all normal matrices in $\mathbb{M}_2(\mathbb{R})$.*

   Ans: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix}$ and $\begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix}$. From $b^2 = c^2$, we have either $b = c$, in which case $A^T = A$ or $b = -c \neq 0$, in which case $a = d$. If $a = d = 0$, we get $A^T = -A$. We could have other matrices like $\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$.

4. *Let $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$. Then, prove that the following statements are equivalent.*

   (a) *A is normal.*

   (b) *A is unitarily diagonalizable.*

   (c) $\sum\limits_{i,j} |a_{ij}|^2 = \sum\limits_{i} |\lambda_i|^2$.

   (d) *A has n orthonormal eigenvectors.*

   Ans: In view of earlier results, we only prove $c) \Rightarrow b)$. By Schur' theorem, there exists a unitary matrix U such that $U^*AU = T$ is upper triangular. As $U^*AU = T$, we have $\sum\limits_{i,j} |a_{ij}|^2 = \sum\limits_{i,j} |t_{ij}|^2 = \sum\limits_{i} |t_{ii}|^2$. So $t_{ij} = 0$, for all $i < j$.

5. *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 $\|A^*\mathbf{x} - \overline{\lambda}\mathbf{x}\|^2 = \|A\mathbf{x} - \lambda\mathbf{x}\|^2$.]*

6. *Let A be an $n \times n$ unitary matrix. Then,*

   (a) *$|\lambda| = 1$ for any eigenvalue λ of A.*

   (b) *the eigenvectors $\mathbf{x}, \mathbf{y}$ corresponding to distinct eigenvalues are orthogonal.*

7. *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 $\theta, 0 \le \theta < 2\pi$. 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 $\theta, 0 \le \theta < 2\pi$. 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 $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.*

8. *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.*

9. *Let A be a normal matrix. Then, prove that* $\text{Rank}(A)$ *equals the number of nonzero eigenvalues of A.*

10. [**Equivalent characterizations of Hermitian matrices**] *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the following statements are equivalent.*

    (a) *The matrix A is Hermitian.*

    (b) *The number x^*Ax is real for each $x \in \mathbb{C}^n$.*

    (c) *The matrix A is normal and has real eigenvalues.*

    (d) *The matrix S^*AS is Hermitian for each $S \in \mathbb{M}_n(\mathbb{C})$.*

    Ans: i)⇒ii),iii),iv) can be shown easily.

    ii)⇒i). Taking $\mathbf{x} = \mathbf{e}_i + \imath \mathbf{e}_j$, we have $\mathbf{x}^*A\mathbf{x} = a_{ii} - \imath a_{ji} + \imath a_{ij} + a_{jj} \in \mathbb{R}$. As $a_{ii}, a_{jj} \in \mathbb{R}$, we see that $a_{ij} - a_{ji}$ is a purely imaginary number, *i.e.*, they have the same real part. Similarly, taking $\mathbf{x} = \mathbf{e}_i + \mathbf{e}_j$, we see that $a_{ij} + a_{ji} \in \mathbb{R}$, that is, they have opposite imaginary parts. So $a_{ij} = \overline{a_{ji}}$.

    iii)⇒i). Suppose that $A^*A = AA^*$ and $\lambda(A) \in \mathbb{R}$. By Spectral theorem $A = U^*\Lambda U$, for some unitary matrix, where Λ is a real matrix. Taking conjugate transpose, we see that $A^* = A$.

    iv)⇒i). Follows by taking $S = I$.

6.2.3 Cayley Hamilton Theorem

Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, in Theorem 6.1.16, we saw that

$$P_A(x) = \det(A - xI) = (-1)^n \left(x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + (-1)^{n-1}a_1 x + (-1)^n a_0\right) \tag{6.2.6}$$

for certain $a_i \in \mathbb{C}$, $0 \le i \le n-1$. Also, if α is an eigenvalue of A then $P_A(\alpha) = 0$. So, $x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + (-1)^{n-1}a_1 x + (-1)^n a_0 = 0$ is satisfied by n complex numbers. It turns out that the expression

$$A^n - a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \cdots + (-1)^{n-1}a_1 A + (-1)^n a_0 I = \mathbf{0}$$

holds true as a matrix identity. This is a celebrated theorem called the **Cayley Hamilton Theorem**. We give a proof using Schur's unitary triangularization. To do so, we look at multiplication of certain upper triangular matrices.

Lemma 6.2.25. *Let $A_1, \ldots, A_n \in \mathbb{M}_n(\mathbb{C})$ be upper triangular matrices such that the (i,i)-th entry of A_i equals 0, for $1 \le i \le n$. Then, $A_1 A_2 \cdots A_n = \mathbf{0}$.*

Proof. We use induction to prove that the first k columns of $A_1 A_2 \cdots A_k$ is $\mathbf{0}$, for $1 \le k \le n$. The result is clearly true for $k = 1$ as the first column of A_1 is $\mathbf{0}$. For clarity, we show that the first two columns of $A_1 A_2$ is $\mathbf{0}$. Let $B = A_1 A_2$. Then, by matrix multiplication

$$B[:, i] = A_1[:, 1](A_2)_{1i} + A_1[:, 2](A_2)_{2i} + \cdots + A_1[:, n](A_2)_{ni} = \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$$

as $A_1[:,1] = \mathbf{0}$ and $(A_2)_{ji} = 0$, for $i = 1,2$ and $j \geq 2$. So, assume that the first $n-1$ columns of $C = A_1 \cdots A_{n-1}$ is $\mathbf{0}$ and let $B = CA_n$. Then, for $1 \leq i \leq n$, we see that

$$B[:,i] = C[:,1](A_n)_{1i} + C[:,2](A_n)_{2i} + \cdots + C[:,n](A_n)_{ni} = \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$$

as $C[:,j] = \mathbf{0}$, for $1 \leq j \leq n-1$ and $(A_n)_{ni} = 0$, for $i = n-1, n$. Thus, by induction hypothesis the required result follows. ■

EXERCISE **6.2.26.** *Let $A, B \in \mathbb{M}_n(\mathbb{C})$ be upper triangular matrices with the top leading principal submatrix of A of size k being $\mathbf{0}$. If $B[k+1,k+1] = 0$ then prove that the leading principal submatrix of size $k+1$ of AB is $\mathbf{0}$.*

We now prove the Cayley Hamilton Theorem using Schur's unitary triangularization.

Theorem 6.2.27 (Cayley Hamilton Theorem)**.** *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A satisfies its characteristic equation. That is, if $P_A(x) = \det(A - xI_n) = a_0 - xa_1 + \cdots + (-1)^{n-1}a_{n-1}x^{n-1} + (-1)^n x^n$ then*

$$A^n - a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \cdots + (-1)^{n-1}a_1 A + (-1)^n a_0 I = \mathbf{0}$$

holds true as a matrix identity.

Proof. Let $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\}$ then $P_A(x) = \prod\limits_{i=1}^{n}(x - \alpha_i)$. And, by Schur's unitary triangularization there exists a unitary matrix U such that $U^*AU = T$, an upper triangular matrix with $t_{ii} = \alpha_i$, for $1 \leq i \leq n$. Now, observe that if $A_i = T - \alpha_i I$ then the A_i's satisfy the conditions of Lemma 6.2.25. Hence,

$$(T - \alpha_1 I) \cdots (T - \alpha_n I) = \mathbf{0}.$$

Therefore,

$$P_A(A) = \prod_{i=1}^{n}(A - \alpha_i I) = \prod_{i=1}^{n}(UTU^* - \alpha_i UIU^*) = U\Big[(T - \alpha_1 I) \cdots (T - \alpha_n I)\Big]U^* = U\mathbf{0}U^* = \mathbf{0}.$$

Thus, the required result follows. ■

We now give some examples and then implications of the Cayley Hamilton Theorem.

Remark 6.2.28. *1. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix}$. Then, $P_A(x) = x^2 + 2x - 5$. Hence, verify that*

$$A^2 + 2A - 5I_2 = \begin{bmatrix} 3 & -4 \\ -2 & 11 \end{bmatrix} + 2\begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix} - 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{0}.$$

Further, verify that $A^{-1} = \dfrac{1}{5}(A + 2I_2) = \dfrac{1}{5}\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$. Furthermore, $A^2 = -2A + 5I$ implies that

$$A^3 = A(A^2) = A(-2A+5I) = -2A^2+5I = -2(-2A+5I)+5I = 4A-10I+5I = 4A-5I.$$

We can keep using the above technique to get A^m as a linear combination of A and I, for all $m \geq 1$.

2. *Let* $A = \begin{bmatrix} 3 & 1 \\ 2 & 0 \end{bmatrix}$. *Then,* $P_A(t) = t(t-3) - 2 = t^2 - 3t - 2$. *So, using* $P_A(A) = \mathbf{0}$, *we have* $A^{-1} = \frac{A-3I}{2}$. *Further,* $A^2 = 3A + 2I$ *implies that* $A^3 = 3A^2 + 2A = 3(3A + 2I) + 2A = 11A + 6I$. *So, as above,* A^m *is a combination of* A *and* I, *for all* $m \geq 1$.

3. *Let* $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. *Then,* $P_A(x) = x^2$. *So, even though* $A \neq \mathbf{0}$, $A^2 = \mathbf{0}$.

4. *For* $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $P_A(x) = x^3$. *Thus, by the Cayley Hamilton Theorem* $A^3 = \mathbf{0}$. *But, it turns out that* $A^2 = \mathbf{0}$.

5. *For* $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, *note that* $P_A(t) = (t-1)^3$. *So* $P_A(A) = \mathbf{0}$. *But, observe that if* $q(t) = (t-1)^2$ *then* $q(A)$ *is also* $\mathbf{0}$.

6. *Let* $A \in \mathbb{M}_n(\mathbb{C})$ *with* $P_A(x) = a_0 - xa_1 + \cdots + (-1)^{n-1}a_{n-1}x^{n-1} + (-1)^n x^n$.

   (a) *Then, for any* $\ell \in \mathbb{N}$, *the division algorithm gives* $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{C}$ *and a polynomial* $f(x)$ *with coefficients from* $\mathbb{C}$ *such that*

   $$x^\ell = f(x)P_A(x) + \alpha_0 + x\alpha_1 + \cdots + x^{n-1}\alpha_{n-1}.$$

   Hence, by the Cayley Hamilton Theorem, $A^\ell = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1}$.

   i. *Thus, to compute any power of* A, *one needs to apply the division algorithm to get* α_i*'s and know* A^i, *for* $1 \leq i \leq n-1$. *This is quite helpful in numerical computation as computing powers takes much more time than division.*

   ii. *Note that* $LS\left\{I, A, A^2, \ldots\right\}$ *is a subspace of* $\mathbb{M}_n(\mathbb{C})$. *Also,* $\dim\left(\mathbb{M}_n(\mathbb{C})\right) = n^2$. *But, the above argument implies that* $\dim\left(LS\left\{I, A, A^2, \ldots\right\}\right) \leq n$.

   iii. *In the language of graph theory, it says the following: "Let* G *be a graph on* n *vertices and* A *its adjacency matrix. Suppose there is no path of length* $n-1$ *or less from a vertex* v *to a vertex* u *in* G. *Then,* G *doesn't have a path from* v *to* u *of any length. That is, the graph* G *is disconnected and* v *and* u *are in different components of* G."

   (b) *Suppose* A *is non-singular. Then, by definition* $a_0 = \det(A) \neq 0$. *Hence,*

   $$A^{-1} = \frac{1}{a_0}\left[a_1 I - a_2 A + \cdots + (-1)^{n-2}a_{n-1}A^{n-2} + (-1)^{n-1}A^{n-1}\right].$$

   This matrix identity can be used to calculate the inverse.

   (c) *The above also implies that if* A *is invertible then* $A^{-1} \in LS\left\{I, A, A^2, \ldots\right\}$. *That is,* A^{-1} *is a linear combination of the vectors* $I, A, \ldots, A^{n-1}$.

EXERCISE **6.2.29.** *Find the inverse of* $\begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 1 & 1 & 2 \end{bmatrix}$, $\begin{bmatrix} -1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ *and* $\begin{bmatrix} 1 & -2 & -1 \\ -2 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ *by the Cayley Hamilton Theorem.*

EXERCISE **6.2.30.** *Miscellaneous Exercises:*

1. *Let* $A, B \in \mathbb{M}_2(\mathbb{C})$ *such that* $A = AB - BA$. *Then, prove that* $A^2 = \mathbf{0}$.
2. *Let* B *be an* $m \times n$ *matrix and* $A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}$. *Then, prove that* $\left(\lambda, \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\right)$ *is an eigen-pair if and only if* $\left(-\lambda, \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix}\right)$ *is an eigen-pair.*
3. *Let* $B, C \in \mathbb{M}_n(\mathbb{R})$. *Define* $A = \begin{bmatrix} B & C \\ -C & B \end{bmatrix}$. *Then, prove the following:*
   (a) *if* s *is a real eigenvalue of* A *with corresponding eigenvector* $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ *then* s *is also an eigenvalue corresponding to the eigenvector* $\begin{bmatrix} -\mathbf{y} \\ \mathbf{x} \end{bmatrix}$.
   (b) *if* $s + it$ *is a complex eigenvalue of* A *with corresponding eigenvector* $\begin{bmatrix} \mathbf{x} + i\mathbf{y} \\ -\mathbf{y} + i\mathbf{x} \end{bmatrix}$ *then* $s - it$ *is also an eigenvalue of* A *with corresponding eigenvector* $\begin{bmatrix} \mathbf{x} - i\mathbf{y} \\ -\mathbf{y} - i\mathbf{x} \end{bmatrix}$.
   (c) $(s + it, \mathbf{x} + i\mathbf{y})$ *is an eigen-pair of* $B+iC$ *if and only if* $(s - it, \mathbf{x} - i\mathbf{y})$ *is an eigen-pair of* $B - iC$.
   (d) $\left(s + it, \begin{bmatrix} \mathbf{x} + i\mathbf{y} \\ -\mathbf{y} + i\mathbf{x} \end{bmatrix}\right)$ *is an eigen-pair of* A *if and only if* $(s + it, \mathbf{x} + i\mathbf{y})$ *is an eigen-pair of* $B + iC$.
   (e) $\det(A) = |\det(B + iC)|^2$.

The next section deals with quadratic forms which helps us in better understanding of conic sections in analytic geometry.

6.3 Quadratic Forms

Definition 6.3.1. [Positive, Semi-positive and Negative definite matrices] Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, A is said to be

1. **positive semi-definite** (psd) if $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ and $\mathbf{x}^*A\mathbf{x} \geq 0$, for all $\mathbf{x} \in \mathbb{C}^n$.
2. **positive definite** (pd) if $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ and $\mathbf{x}^*A\mathbf{x} > 0$, for all $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.
3. **negative semi-definite** (nsd) if $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ and $\mathbf{x}^*A\mathbf{x} \leq 0$, for all $\mathbf{x} \in \mathbb{C}^n$.
4. **negative definite** (nd) if $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ and $\mathbf{x}^*A\mathbf{x} < 0$, for all $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.
5. **indefinite** if $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ and there exist $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ such that $\mathbf{x}^*A\mathbf{x} < 0 < \mathbf{y}^*A\mathbf{y}$.

Lemma 6.3.2. *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Then* A *is Hermitian if and only if at least one of the following statements hold:*

1. S^*AS *is Hermitian for all* $S \in \mathbb{M}_n$.
2. A *is normal and has real eigenvalues.*

3. *$\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}^n$.*

Proof. Let $S \in \mathbb{M}_n$, $(S^*AS)^* = S^*A^*S = S^*AS$. Thus S^*AS is Hermitian.

Suppose $A = A^*$. Then, A is clearly normal as $AA^* = A^2 = A^*A$. Further, if $(\lambda, \mathbf{x})$ is an eigenpair then $\lambda\mathbf{x}^*\mathbf{x} = \mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ implies $\lambda \in \mathbb{R}$.

For the last part, note that $\mathbf{x}^*A\mathbf{x} \in \mathbb{C}$. Thus $\overline{\mathbf{x}^*A\mathbf{x}} = (\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A^*\mathbf{x} = \mathbf{x}^*A\mathbf{x}$, we get $\text{Im}(\mathbf{x}^*A\mathbf{x}) = 0$. Thus, $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$.

If S^*AS is Hermitian for all $S \in \mathbb{M}_n$ then taking $S = I_n$ gives A is Hermitian.

If A is normal then $A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U$ for some unitary matrix U. Since $\lambda_i \in \mathbb{R}$, $A^* = (U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U)^* = U^* \operatorname{diag}(\overline{\lambda_1}, \ldots, \overline{\lambda_n})U = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U = A$. So, A is Hermitian.

If $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}^n$ then $a_{ii} = \mathbf{e}_i^*A\mathbf{e}_i \in \mathbb{R}$. Also, $a_{ii} + a_{jj} + a_{ij} + a_{ji} = (\mathbf{e}_i + \mathbf{e}_j)^*A(\mathbf{e}_i + \mathbf{e}_j) \in \mathbb{R}$. So, $\text{Im}(a_{ij}) = -\text{Im}(a_{ji})$. Similarly, $a_{ii} + a_{jj} + ia_{ij} - ia_{ji} = (\mathbf{e}_i + i\mathbf{e}_j)^*A(\mathbf{e}_i + i\mathbf{e}_j) \in \mathbb{R}$ implies that $\text{Re}(a_{ij}) = \text{Re}(a_{ji})$. Thus, $A = A^*$. ∎

Remark 6.3.3. *Let $A \in \mathbb{M}_n(\mathbb{R})$. Then the condition $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$ in Definition 6.3.9 is always true and hence doesn't put any restriction on the matrix A. So, in Definition 6.3.9, we assume that $A^T = A$, i.e., A is a symmetric matrix.*

Example 6.3.4.
1. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ or $A = \begin{bmatrix} 3 & 1+i \\ 1-i & 4 \end{bmatrix}$. Then, A is positive definite.

2. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ or $A = \begin{bmatrix} \sqrt{2} & 1+i \\ 1-i & \sqrt{2} \end{bmatrix}$. Then, A is positive semi-definite but not positive definite.

3. Let $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$ or $A = \begin{bmatrix} -2 & 1-i \\ 1+i & -2 \end{bmatrix}$. Then, A is negative definite.

4. Let $A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$ or $A = \begin{bmatrix} -2 & 1-i \\ 1+i & -1 \end{bmatrix}$. Then, A is negative semi-definite.

5. Let $A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$ or $A = \begin{bmatrix} 1 & 1+i \\ 1-i & 1 \end{bmatrix}$. Then, A is indefinite.

Theorem 6.3.5. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the following statements are equivalent.*

1. *A is positive semi-definite.*
2. *$A^* = A$ and each eigenvalue of A is non-negative.*
3. *$A = B^*B$ for some $B \in \mathbb{M}_n(\mathbb{C})$.*

Proof. $1 \Rightarrow 2$: Let A be positive semi-definite. Then, by Lemma 6.3.2 A is Hermitian. If $(\alpha, \mathbf{v})$ is an eigen-pair of A then $\alpha\|\mathbf{v}\|^2 = \mathbf{v}^*A\mathbf{v} \geq 0$. So, $\alpha \geq 0$.

$2 \Rightarrow 3$: Let $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\}$. Then, by spectral theorem, there exists a unitary matrix U such that $U^*AU = D$ with $D = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$. As $\alpha_i \geq 0$, for $1 \leq i \leq n$, define $D^{\frac{1}{2}} = \operatorname{diag}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n})$. Then, $A = UD^{\frac{1}{2}}[D^{\frac{1}{2}}U^*] = B^*B$.

$3 \Rightarrow 1$: Let $A = B^*B$. Then, for $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*A\mathbf{x} = \mathbf{x}^*B^*B\mathbf{x} = \|B\mathbf{x}\|^2 \geq 0$. Thus, the required result follows. ∎

A similar argument gives the next result and hence the proof is omitted.

Theorem 6.3.6. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, the following statements are equivalent.*

1. *A is positive definite.*
2. *$A^* = A$ and each eigenvalue of A is positive.*
3. *$A = B^*B$ for a non-singular matrix $B \in \mathbb{M}_n(\mathbb{C})$.*

Remark 6.3.7. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then, there exists a unitary matrix $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ and a diagonal matrix $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ such that $A = UDU^*$. Now, for $1 \leq i \leq n$, define $\alpha_i = \max\{\lambda_i, 0\}$ and $\beta_i = \min\{\lambda_i, 0\}$. Then*

1. *for $D_1 = diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$, the matrix $A_1 = UD_1U^*$ is positive semi-definite.*
2. *for $D_2 = diag(\beta_1, \beta_2, \ldots, \beta_n)$, the matrix $A_2 = UD_2U^*$ is positive semi-definite.*
3. *$A = A_1 - A_2$. The matrix A_1 is generally called the positive semi-definite part of A.*

Definition 6.3.8. **[Multilinear Function]** Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then,

1. for a fixed $m \in \mathbb{N}$, a function $f : \mathbb{V}^m \to \mathbb{F}$ is called an m**-multilinear** function if f is linear in each component. That is,

$$\begin{aligned} f(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, (\mathbf{v}_i + \alpha\mathbf{u}), \mathbf{v}_{i+1} \ldots, \mathbf{v}_m) &= f(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1} \ldots, \mathbf{v}_m) \\ &\quad + \alpha f(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{u}, \mathbf{v}_{i+1} \ldots, \mathbf{v}_m) \end{aligned}$$

   for $\alpha \in \mathbb{F}$, $\mathbf{u} \in \mathbb{V}$ and $\mathbf{v}_i \in \mathbb{V}$, for $1 \leq i \leq m$.
2. An m-multilinear form is also called an m**-form**.
3. A 2-form is called a **bilinear form**.

Definition 6.3.9. **[Sesquilinear, Hermitian and Quadratic Forms]** Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Then, a **sesquilinear form** in $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ is defined as $H(\mathbf{x}, \mathbf{y}) = \mathbf{y}^*A\mathbf{x}$. In particular, $H(\mathbf{x}, \mathbf{x})$, denoted $H(\mathbf{x})$, is called a **Hermitian form**. In case $A \in \mathbb{M}_n(\mathbb{R})$, $H(\mathbf{x})$ is called a **quadratic** form.

Remark 6.3.10. *Observe that*

1. *if $A = I_n$ then the bilinear/sesquilinear form reduces to the standard inner product.*
2. *$H(\mathbf{x}, \mathbf{y})$ is 'linear' in the first component and 'conjugate linear' in the second component.*
3. *the quadratic form $H(\mathbf{x})$ is a real number. Hence, for $\alpha \in \mathbb{R}$, the equation $H(\mathbf{x}) = \alpha$, represents a conic in $\mathbb{R}^n$.*

Example 6.3.11.

1. Let $\mathbf{v}_i \in \mathbb{C}^n$, for $1 \leq i \leq n$. Then, $f(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \det([\mathbf{v}_1, \ldots, \mathbf{v}_n])$ is an n-form on $\mathbb{C}^n$.
2. Let $A \in \mathbb{M}_n(\mathbb{R})$. Then, $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^TA\mathbf{x}$, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, is a bilinear form on $\mathbb{R}^n$.
3. Let $A = \begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix}$. Then, $A^* = A$ and for $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, verify that

$$H(\mathbf{x}) = \mathbf{x}^*A\mathbf{x} = |x|^2 + 2|y|^2 + 2\mathrm{Re}\left((2-i)\overline{x}y\right)$$

   where 'Re' denotes the real part of a complex number, is a sesquilinear form.

6.3.1 Sylvester's law of inertia

The main idea of this section is to express $H(\mathbf{x})$ as sum or difference of squares. Since $H(\mathbf{x})$ is a quadratic in $\mathbf{x}$, replacing $\mathbf{x}$ by $c\mathbf{x}$, for $c \in \mathbb{C}$, just gives a multiplication factor by $|c|^2$. Hence, one needs to study only the normalized vectors. Let us consider Example 6.1.1 again. There we see that

$$\mathbf{x}^T A\mathbf{x} = 3\frac{(x+y)^2}{2} - \frac{(x-y)^2}{2} = (x+2y)^2 - 3y^2, \text{ and} \tag{6.3.1}$$

$$\mathbf{x}^T B\mathbf{x} = 5\frac{(x+2y)^2}{5} + 10\frac{(2x-y)^2}{5} = (3x - \frac{2y}{3})^2 + \frac{50y^2}{9}. \tag{6.3.2}$$

Note that both the expressions in Equation (6.3.1) is the difference of two non-negative terms. Whereas, both the expressions in Equation (6.3.2) consists of sum of two non-negative terms. Is this just a coincidence?

In general, let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then, by Theorem 6.2.22, $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{R}$ and there exists a unitary matrix U such that $U^*AU = D = \text{diag}(\alpha_1, \ldots, \alpha_n)$. Let $\mathbf{x} = U\mathbf{z}$. Then, $\|\mathbf{x}\| = 1$ and U is unitary implies that $\|\mathbf{z}\| = 1$. If $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^*$ then

$$H(\mathbf{x}) = \mathbf{z}^*U^*AU\mathbf{z} = \mathbf{z}^*D\mathbf{z} = \sum_{i=1}^{n} \alpha_i |\mathbf{z}_i|^2 = \sum_{i=1}^{p} |\sqrt{\alpha_i}\ \mathbf{z}_i|^2 - \sum_{i=p+1}^{r} \left|\sqrt{|\alpha_i|}\ \mathbf{z}_i\right|^2, \tag{6.3.3}$$

where $\alpha_1, \ldots, \alpha_p > 0$, $\alpha_{p+1}, \ldots, \alpha_r < 0$ and $\alpha_{r+1}, \ldots, \alpha_n = 0$. Thus, we see that the possible values of $H(\mathbf{x})$ seem to depend only on the eigenvalues of A. Since U is an invertible matrix, the components $\mathbf{z}_i$'s of $\mathbf{z} = U^{-1}\mathbf{x} = U^*\mathbf{x}$ are commonly known as the **linearly independent linear forms**. Note that each $\mathbf{z}_i$ is a linear expression in the components of $\mathbf{x}$. Also, note that in Equation (6.3.3), p corresponds to the number of positive eigenvalues and $r - p$ to the number of negative eigenvalues. For a better understanding, we define the following numbers.

Definition 6.3.12. **[Inertia and Signature of a Matrix]** Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. The **inertia** of A, denoted $i(A)$, is the triplet $(i_+(A), i_-(A), i_0(A))$, where $i_+(A)$ is the number of positive eigenvalues of A, $i_-(A)$ is the number of negative eigenvalues of A and $i_0(A)$ is the nullity of A. The difference $i_+(A) - i_-(A)$ is called the **signature** of A.

EXERCISE **6.3.13.** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. If the signature and the rank of A is known then prove that one can find out the inertia of A.*

As a next result, we show that in any expression of $H(\mathbf{x})$ as a sum or difference 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 popularly known as the 'Sylvester's law of inertia'.

Lemma 6.3.14. **[Sylvester's Law of Inertia]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and let $\mathbf{x} \in \mathbb{C}^n$. Then, every Hermitian form $H(\mathbf{x}) = \mathbf{x}^*A\mathbf{x}$, in n variables can be written as*

$$H(\mathbf{x}) = |\mathbf{y}_1|^2 + \cdots + |\mathbf{y}_p|^2 - |\mathbf{y}_{p+1}|^2 - \cdots - |\mathbf{y}_r|^2$$

where $\mathbf{y}_1, \ldots, \mathbf{y}_r$ are linearly independent linear forms in the components of $\mathbf{x}$ and the integers p and r satisfying $0 \le p \le r \le n$, depend only on A.

Proof. Equation (6.3.3) implies 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 $p, q, r, s \in \mathbb{N}$ with $p > q$ such that

$$\begin{aligned} H(\mathbf{x}) &= |\mathbf{y}_1|^2 + \cdots + |\mathbf{y}_p|^2 - |\mathbf{y}_{p+1}|^2 - \cdots - |\mathbf{y}_r|^2 && (6.3.4)\\ &= |\mathbf{z}_1|^2 + \cdots + |\mathbf{z}_q|^2 - |\mathbf{z}_{q+1}|^2 - \cdots - |\mathbf{z}_s|^2, && (6.3.5)\end{aligned}$$

where $\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = M\mathbf{x}$, $\mathbf{z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = N\mathbf{x}$ with $Y_1 = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_p \end{bmatrix}$ and $Z_1 = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_q \end{bmatrix}$ for some invertible matrices M and N. Now the invertibility of M and N implies $\mathbf{z} = B\mathbf{y}$, for some invertible matrix B. Decompose $B = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$, where B_1 is a $q \times p$ matrix. Then $\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$. As $p > q$, the homogeneous linear system $B_1 Y_1 = \mathbf{0}$ has a nontrivial solution, say $\widetilde{Y_1} = \begin{bmatrix} \tilde{y_1} \\ \vdots \\ \tilde{y_p} \end{bmatrix}$ and consider $\widetilde{\mathbf{y}} = \begin{bmatrix} \widetilde{Y_1} \\ \mathbf{0} \end{bmatrix}$. Then for this choice of $\widetilde{\mathbf{y}}$, $Z_1 = \mathbf{0}$ and thus, using Equations (6.3.4) and (6.3.5), we have

$$H(\tilde{\mathbf{y}}) = |\tilde{y_1}|^2 + |\tilde{y_2}|^2 + \cdots + |\tilde{y_p}|^2 - 0 = 0 - (|z_{q+1}|^2 + \cdots + |z_s|^2).$$

Now, this can hold only if $\widetilde{Y_1} = \mathbf{0}$, a contradiction to $\widetilde{Y_1}$ being a non-trivial solution. Hence $p = q$. Similarly, the case $r > s$ can be resolved. This completes the proof of the lemma. $\square$

Remark 6.3.15. *Since A is Hermitian,* $\text{Rank}(A)$ *equals the number of nonzero eigenvalues. Hence,* $\text{Rank}(A) = r$. *The number r is called the* **rank** *and the number $r - 2p$ is called the* **inertial degree** *of the Hermitian form $H(\mathbf{x})$.*

We now look at another form of the Sylvester's law of inertia. We start with the following definition.

Definition 6.3.16. **[Star Congruence]** Let $A, B \in \mathbb{M}_n(\mathbb{C})$. Then, A is said to be **$*$-congruent** (read star-congruent) to B if there exists an invertible matrix S such that $A = S^*BS$.

Theorem 6.3.17. **[Second Version: Sylvester's Law of Inertia]** *Let $A, B \in \mathbb{M}_n(\mathbb{C})$ be Hermitian. Then, A is $*$-congruent to B if and only if $i(A) = i(B)$.*

Proof. By spectral theorem $U^*AU = \Lambda_A$ and $V^*BV = \Lambda_B$, for some unitary matrices U, V and diagonal matrices Λ_A, Λ_B of the form $\operatorname{diag}(+, \cdots, +, -, \cdots, -, 0, \cdots, 0)$. Thus, there exist invertible matrices S, T such that $S^*AS = D_A$ and $T^*BT = D_B$, where D_A, D_B are diagonal matrices of the form $\operatorname{diag}(1, \cdots, 1, -1, \cdots, -1, 0, \cdots, 0)$.

If $i(A) = i(B)$, then it follows that $D_A = D_B$, *i.e.*, $S^*AS = T^*BT$ and hence $A = (TS^{-1})^*B(TS^{-1})$.

Conversely, suppose that $A = P^*BP$, for some invertible matrix P, and $i(B) = (k, l, m)$. As $T^*BT = D_B$, we have, $A = P^*(T^*)^{-1}D_BT^{-1}P = (T^{-1}P)^*D_B(T^{-1}P)$. Now, let $X = (T^{-1}P)^{-1}$. Then, $A = (X^{-1})^*D_BX^{-1}$ and we have the following observations.

1. As rank and nullity do not change under similarity transformation, $i_0(A) = i_0(D_B) = m$ as $i(B) = (k, l, m)$.
2. Using $i(B) = (k, l, m)$, we also have

   $$X[:, k+1]^* A X[:, k+1] = X[:, k+1]^* \left((X^{-1})^* D_B (X^{-1})\right) X[:, k+1] = \mathbf{e}_{k+1}^* D_B \mathbf{e}_{k+1} = -1.$$

   Similarly, $X[:, k+2]^* A X[:, k+2] = \cdots = X[:, k+l]^* A X[:, k+l] = -1$. As the vectors $X[:, k+1], \ldots, X[:, k+l]$ are linearly independent, using 7.7.10, we see that A has at least l negative eigenvalues.
3. Similarly, $X[:, 1]^* A X[:, 1] = \cdots = X[:, k]^* A X[:, k] = 1$. As $X[:, 1], \ldots, X[:, k]$ are linearly independent, using 7.7.10 again, we see that A has at least k positive eigenvalues.

Thus, it now follows that $i(A) = (k, l, m)$. ∎

6.3.2 Applications in Eculidean Plane and Space

We now obtain conditions on the eigenvalues of A, corresponding to the associated quadratic form, to characterize conic sections in $\mathbb{R}^2$, with respect to the standard inner product.

Definition 6.3.18. **[Associated Quadratic Form]** Let $f(x, y) = ax^2 + 2hxy + by^2 + 2fx + 2gy + c$ be a general quadratic in x and y, with coefficients from $\mathbb{R}$. Then,

$$H(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x, & y \end{bmatrix} \begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2hxy + by^2$$

is called the **associated quadratic form** of the conic $f(x, y) = 0$.

Proposition 6.3.19. *Consider the quadratic $f(x, y) = ax^2 + 2hxy + by^2 + 2gx + 2fy + c$, for $a, b, c, g, f, h \in \mathbb{R}$. If $(a, b, h) \neq (0, 0, 0)$ then $f(x, y) = 0$ represents*

1. *a parabola or a pair of parallel lines if $ab - h^2 = 0$,*
2. *a hyperbola or a pair of perpendicular lines if $ab - h^2 < 0$,*
3. *an ellipse or a circle or a point (point of intersection of a pair of perpendicular lines) if $ab - h^2 > 0$.*

Proof. Consider the associated quadratic $ax^2 + 2hxy + by^2$ with $A = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$ as the associated symmetric matrix. Then, by Corollary 6.2.23, $A = U \operatorname{diag}(\alpha_1, \alpha_2) U^T$, where $U = [\mathbf{u}_1, \mathbf{u}_2]$ is an orthogonal matrix, with $(\alpha_1, \mathbf{u}_1)$ and $(\alpha_2, \mathbf{u}_2)$ as eigen-pairs of A. As $(a, b, h) \neq (0, 0, 0)$ at least one of $\alpha_1, \alpha_2 \neq 0$. Also,

$$\mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x, & y \end{bmatrix} U \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} U^T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \alpha_1 u^2 + \alpha_2 v^2,$$

where $\begin{bmatrix} u \\ v \end{bmatrix} = U^T \mathbf{x}$. The lines $u = 0, v = 0$ are the two linearly independent linear forms, which correspond to two perpendicular lines passing through the origin in the (x, y)-plane. In terms of u, v, $f(x, y)$ reduces to $f(u, v) = \alpha_1 u^2 + \alpha_2 v^2 + d_1 u + d_2 v + c$, for some choice of $d_1, d_2 \in \mathbb{R}$. We now look at different cases:

1. if $\alpha_1 = 0$ and $\alpha_2 \neq 0$ then $ab - h^2 = \det(A) = \alpha_1\alpha_2 = 0$. In this case,

$$f(u,v) = 0 \Leftrightarrow \alpha_2 \left(v + \frac{d_2}{2\alpha_2}\right)^2 = c_1 - d_1 u,$$

for some $c_1 \in \mathbb{R}$.

   (a) If $d_1 = 0$, the quadratic corresponds to either the same line $v + \dfrac{d_2}{2\alpha_2} = 0$, two parallel lines or two imaginary lines, depending on whether $c_1 = 0, c_1\alpha_2 > 0$ and $c_1\alpha_2 < 0$, respectively.

   (b) If $d_1 \neq 0$, the quadratic corresponds to a parabola of the form $V^2 = 4aU$, for some translate $U = u + \alpha$ and $V = v + \beta$.

2. If $\alpha_1\alpha_2 < 0$ then $ab - h^2 = \det(A) = \lambda_1\lambda_2 < 0$. If $\alpha_2 = -\beta_2 < 0$, for $\beta_2 > 0$ then the quadratic reduces to $\alpha_1(u + d_1)^2 - \beta_2(v + d_2)^2 = d_3$, or equivalently, to $\left(\sqrt{\alpha_1}(u+d_1) + \sqrt{\beta_2}(v+d_2)\right)\cdot\left(\sqrt{\alpha_1}(u+d_1) - \sqrt{\beta_2}(v+d_2)\right) = d_3$, for some $d_1, d_2, d_3 \in \mathbb{R}$. Thus, the quadratic corresponds to

   (a) a pair of perpendicular lines $u + d_1 = 0$ and $v + d_2 = 0$ whenever $d_3 = 0$.

   (b) a hyperbola with orthogonal principal axes $u + d_1 = 0$ and $v + d_2 = 0$ whenever $d_3 \neq 0$. In particular, if $d_3 > 0$ then the corresponding equation equals

$$\frac{\alpha_1(u+d_1)^2}{d_3} - \frac{\alpha_2(v+d_2)^2}{d_3} = 1.$$

3. If $\alpha_1\alpha_2 > 0$ then $ab - h^2 = \det(A) = \alpha_1\alpha_2 > 0$. Here, the quadratic reduces to $\alpha_1(u + d_1)^2 + \alpha_2(v + d_2)^2 = d_3$, for some $d_1, d_2, d_3 \in \mathbb{R}$. Thus, the quadratic corresponds to

   (a) a point which is the point of intersection of the pair of orthogonal lines $u + d_1 = 0$ and $v + d_2 = 0$ if $d_3 = 0$.

   (b) an empty set if $\alpha_1 d_3 < 0$.

   (c) an ellipse or circle with $u + d_1 = 0$ and $v + d_2 = 0$ as the orthogonal principal axes if $\alpha_1 d_3 > 0$ with the corresponding equation

$$\frac{\alpha_1(u+d_1)^2}{d_3} + \frac{\alpha_2(v+d_2)^2}{d_3} = 1.$$

Thus, we have considered all the possible cases and the required result follows. ■

Remark 6.3.20. *Observe that the linearly independent forms* $\begin{bmatrix} u \\ v \end{bmatrix} = U^T \begin{bmatrix} x \\ y \end{bmatrix}$ *are functions of the eigenvectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$*. Further, the linearly independent forms together with the shifting of the origin give us the principal axes of the corresponding conic.*

Example 6.3.21. 1. Let $H(\mathbf{x}) = x^2 + y^2 + 2xy$ be the associated quadratic form for a class of curves. Then, $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and the eigen-pairs are $\left(2, \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}\right)$ and $\left(0, \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}\right)$. In particular, for

(a) $f(x,y) = x^2+2xy+y^2-8x-8y+16$, we have $f(x,y)=0 \Leftrightarrow (x+y-4)^2=0$, just one line.

(b) $f(x,y) = x^2+2xy+y^2-8x-8y$, we have $f(x,y)=0 \Leftrightarrow (x+y-8)\cdot(x+y)=0$, a pair of parallel lines.

(c) $f(x,y) = x^2+2xy+y^2-6x-10y-3$, we have

$$\begin{aligned} f(x,y)=0 \quad &\Leftrightarrow \quad 2\left(\frac{x+y}{\sqrt{2}}\right)^2 + 0\left(\frac{x-y}{\sqrt{2}}\right)^2 = 8\sqrt{2}\left(\frac{x+y}{\sqrt{2}}\right) - 2\sqrt{2}\left(\frac{x-y}{\sqrt{2}}\right) + 3 \\ &\Leftrightarrow \quad \left(\frac{x+y-4}{\sqrt{2}}\right)^2 = -\sqrt{2}\left(\frac{x-y-19/2}{\sqrt{2}}\right), \end{aligned}$$

a parabola with principal axes $x+y=4$, $2x-2y=19$ and directrix $x-y=10$.

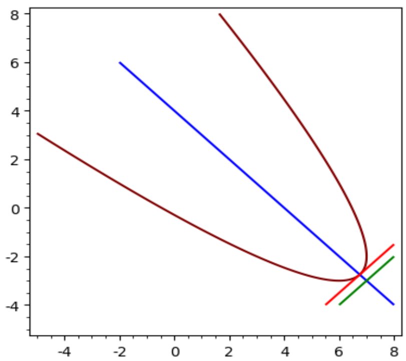


Figure 6.2: Conic $x^2+2xy+y^2-6x-10y=3$

2. Let $H(\mathbf{x}) = 10x^2-5y^2+20xy$ be the associated quadratic form for a class of curves. Then $A = \begin{bmatrix} 10 & 10 \\ 10 & -5 \end{bmatrix}$ and the eigen-pairs are $\left(15, \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}\right)$ and $\left(-10, \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}\right)$. So, for

(a) $f(x,y) = 10x^2-5y^2+20xy+16x-2y+1$, we have $f(x,y)=0 \Leftrightarrow 3(2x+y+1)^2 - 2(x-2y-1)^2 = 0$, a pair of perpendicular lines.

(b) $f(x,y) = 10x^2-5y^2+20xy+16x-2y+19$, we have

$$f(x,y)=0 \Leftrightarrow \left(\frac{x-2y-1}{3}\right)^2 - \left(\frac{2x+y+1}{\sqrt{6}}\right)^2 = 1,$$

a hyperbola.

(c) $f(x,y) = 10x^2-5y^2+20xy+16x-2y-17$, we have

$$f(x,y)=0 \Leftrightarrow \left(\frac{2x+y+1}{\sqrt{6}}\right)^2 - \left(\frac{x-2y-1}{3}\right)^2 = 1,$$

a hyperbola.

3. Let $H(\mathbf{x}) = 6x^2+9y^2+4xy$ be the associated quadratic form for a class of curves. Then, $A = \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix}$, and the eigen-pairs are $\left(10, \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}\right)$ and $\left(5, \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}\right)$. So, for

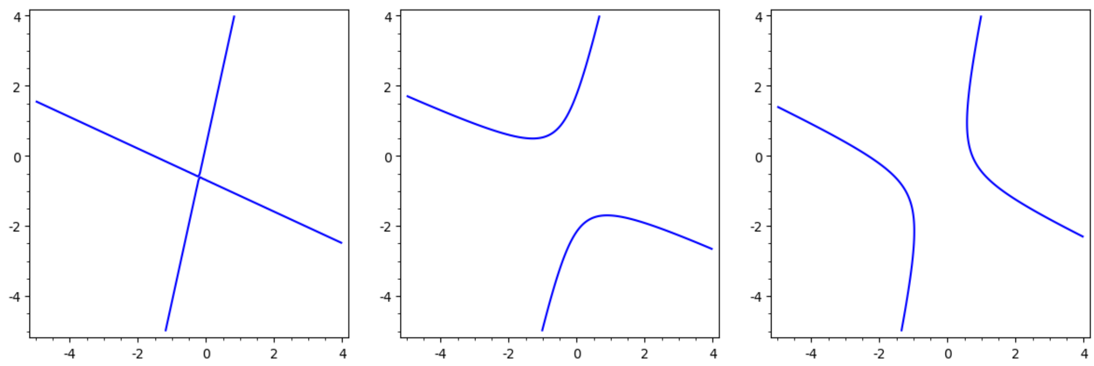


Figure 6.3: Conic $10x^2 - 5y^2 + 20xy + 16x - 2y = c$, $c = -1$, $c = -19$ and $c = 17$

(a) $f(x,y) = 6x^2 + 9y^2 + 4xy + 10y - 53$, we have

$$f(x,y) = 0 \Leftrightarrow \left(\frac{x+2y+1}{5}\right)^2 + \left(\frac{2x-y-1}{5\sqrt{2}}\right)^2 = 1,$$

an ellipse.

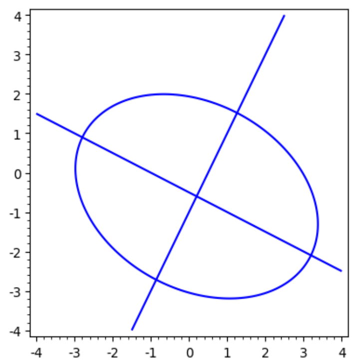


Figure 6.4: Conic $6x^2 + 9y^2 + 4xy + 10y = 53$

.

EXERCISE **6.3.22.** *Sketch the graph of the following surfaces:*

1. $x^2 + 2xy + y^2 + 6x + 10y = 3$.

   Ans: a parabola.

2. $2x^2 + 6xy + 3y^2 - 12x - 6y = 5$.

   Ans: a hyperbola.

3. $4x^2 - 4xy + 2y^2 + 12x - 8y = 10$.

   Ans: an ellipse.

4. $2x^2 - 6xy + 5y^2 - 10x + 4y = 7$.

   Ans: an ellipse.

As a last application, we consider a quadratic in 3 variables, namely x_1, x_2 and x_3. To do so, let $A = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} l \\ m \\ n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ with

$$\begin{aligned} f(x_1, x_2, x_3) &= \mathbf{x}^T A\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + q \\ &= ax_1^2 + bx_2^2 + cx_3^2 + 2hx_1x_2 + 2gx_1x_3 + 2fx_2x_3 \\ &\qquad +2lx_1 + 2mx_2 + 2nx_3 + q \end{aligned} \tag{6.3.6}$$

Then, we observe the following:

1. As A is symmetric, $P^TAP = \operatorname{diag}(\alpha_1, \alpha_2, \alpha_3)$, where $P = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ is an orthogonal matrix and $(\alpha_i, \mathbf{u}_i)$, for $i = 1, 2, 3$ are eigen-pairs of A.
2. Let $\mathbf{y} = P^T\mathbf{x}$. Then, $f(x_1, x_2, x_3)$ reduces to
$$g(y_1, y_2, y_3) = \alpha_1y_1^2 + \alpha_2y_2^2 + \alpha_3y_3^2 + 2l_1y_1 + 2l_2y_2 + 2l_3y_3 + q. \tag{6.3.7}$$
3. Depending on the values of α_i's, rewrite $g(y_1, y_2, y_3)$ to determine the center and the planes of symmetry of $f(x_1, x_2, x_3) = 0$.

Example 6.3.23. Determine the following quadrics $f(x, y, z) = 0$, where

1. $f(x, y, z) = 2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz + 4x + 2y + 4z + 2$.
2. $f(x, y, z) = 3x^2 - y^2 + z^2 + 10$.
3. $f(x, y, z) = 3x^2 - y^2 + z^2 - 10$.
4. $f(x, y, z) = 3x^2 - y^2 + z - 10$.

Solution: (1) Here $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ and $q = 2$. So, verify $P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$ and $P^TAP = \operatorname{diag}(4, 1, 1)$. Hence, $f(x, y, z) = 0$ reduces to

$$4\left(\frac{x + y + z}{\sqrt{3}}\right)^2 + \left(\frac{x - y}{\sqrt{2}}\right)^2 + \left(\frac{x + y - 2z}{\sqrt{6}}\right)^2 = -(4x + 2y + 4z + 2).$$

Or equivalently to $4\left(\frac{4(x + y + z) + 5}{4\sqrt{3}}\right)^2 + \left(\frac{x - y + 1}{\sqrt{2}}\right)^2 + \left(\frac{x + y - 2z - 1}{\sqrt{6}}\right)^2 = \frac{9}{12}$. So, the principal axes of the quadric (an ellipsoid) are $4(x + y + z) = -5, x - y = 1$ and $x + y - 2z = 1$.

Part 2 Here $f(x, y, z) = 0$ reduces to $\frac{y^2}{10} - \frac{3x^2}{10} - \frac{z^2}{10} = 1$ which is the equation of a hyperboloid consisting of two sheets with center $\mathbf{0}$ and the axes x, y and z as the principal axes.

Part 3 Here $f(x, y, z) = 0$ reduces to $\frac{3x^2}{10} - \frac{y^2}{10} + \frac{z^2}{10} = 1$ which is the equation of a hyperboloid consisting of one sheet with center $\mathbf{0}$ and the axes x, y and z as the principal axes.

Part 4 Here $f(x, y, z) = 0$ reduces to $z = y^2 - 3x^2 + 10$ which is the equation of a hyperbolic paraboloid.

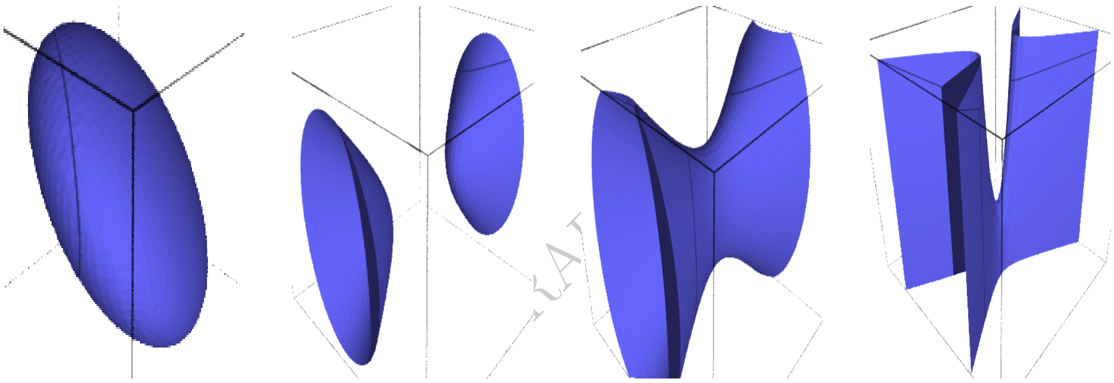


Figure 6.5: Ellipsoid, hyperboloid of two sheets and one sheet, hyperbolic paraboloid

.

Chapter 7

Appendix

7.1 Uniqueness of RREF

Definition 7.1.1. Fix $n \in \mathbb{N}$. Then, for each $f \in \mathcal{S}_n$, we associate an $n \times n$ matrix, denoted $P^f = [p_{ij}]$, such that $p_{ij} = 1$, whenver $f(j) = i$ and 0, otherwise. The matrix P^f is called the **Permutation matrix** corresponding to the permutation f. For example, I_2, corresponding to Id_2, and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = E_{12}$, corresponding to the permutation $(1, 2)$, are the two permutation matrices of order 2×2.

Remark 7.1.2. *Recall that in Remark 7.2.16.1, it was observed that each permutation is a product of n transpositions, $(1, 2), \ldots, (1, n)$.*

1. *Verify that the elementary matrix E_{ij} is the permutation matrix corresponding to the transposition (i, j) .*
2. *Thus, every permutation matrix is a product of elementary matrices E_{1j}, $1 \le j \le n$.*
3. *For $n = 3$, the permutation matrices are I_3, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = E_{23} = E_{12}E_{13}E_{12}$, $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_{12}$, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = E_{12}E_{13}$, $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = E_{13}E_{12}$ and $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = E_{13}$.*
4. *Let $f \in \mathcal{S}_n$ and $P^f = [p_{ij}]$ be the corresponding permutation matrix. Since $p_{ij} = \delta_{i,j}$ and $\{f(1), \ldots, f(n)\} = [n]$, each entry of P^f is either 0 or 1. Furthermore, every row and column of P^f has exactly one nonzero entry. This nonzero entry is a 1 and appears at the position $p_{i,f(i)}$.*
5. *By the previous paragraph, we see that when a permutation matrix is multiplied to A*
   (a) *from left then it permutes the rows of A.*
   (b) *from right then it permutes the columns of A.*
6. *P is a permutation matrix if and only if P has exactly one 1 in each row and column.*
   Solution: *If P has exactly one 1 in each row and column, then P is a square matrix, say*

$n \times n$. Now, apply GJE to P. The occurrence of exactly one 1 in each row and column implies that these 1's are the pivots in each column. We just need to interchange rows to get it in RREF. So, we need to multiply by E_{ij}. Thus, GJE of P is I_n and P is indeed a product of E_{ij}'s. The other part has already been explained earlier.

We are now ready to prove Theorem 2.2.17.

Theorem 7.1.3. *Let A and B be two matrices in RREF. If they are row equivalent then $A = B$.*

Proof. Note that the matrix $A = \mathbf{0}$ if and only if $B = \mathbf{0}$. So, let us assume that the matrices $A, B \neq \mathbf{0}$. Also, the row-equivalence of A and B implies that there exists an invertible matrix C such that $A = CB$, where C is product of elementary matrices.

Since B is in RREF, either $B[:,1] = \mathbf{0}^T$ or $B[:,1] = (1,0,\ldots,0)^T$. If $B[:,1] = \mathbf{0}^T$ then $A[:,1] = CB[:,1] = C\mathbf{0} = \mathbf{0}$. If $B[:,1] = (1,0,\ldots,0)^T$ then $A[:,1] = CB[:,1] = C[:,1]$. As C is invertible, the first column of C cannot be the zero vector. So, $A[:,1]$ cannot be the zero vector. Further, A is in RREF implies that $A[:,1] = (1,0,\ldots,0)^T$. So, we have shown that if A and B are row-equivalent then their first columns must be the same.

Now, let us assume that the first $k-1$ columns of A and B are equal and it contains r pivotal columns. We will now show that the k-th column is also the same.

Define $A_k = [A[:,1],\ldots,A[:,k]]$ and $B_k = [B[:,1],\ldots,B[:,k]]$. Then, our assumption implies that $A[:,i] = B[:,i]$, for $1 \le i \le k-1$. Since, the first $k-1$ columns contain r pivotal columns, there exists a permutation matrix P such that

$$A_kP = \left[\begin{array}{cc|c} I_r & W & A[:,k] \\ \mathbf{0} & \mathbf{0} & \end{array}\right] \text{ and } B_kP = \left[\begin{array}{cc|c} I_r & W & B[:,k] \\ \mathbf{0} & \mathbf{0} & \end{array}\right].$$

If the k-th columns of A and B are pivotal columns then by definition of RREF, $A[:,k] = \begin{bmatrix} \mathbf{0} \\ \mathbf{e}_1 \end{bmatrix} = B[:,k]$, where $\mathbf{0}$ is a vector of size r and $\mathbf{e}_1 = (1,0,\ldots,0)^T$. So, we need to consider two cases depending on whether both are non-pivotal or one is pivotal and the other is not.

As $A = CB$, we get $A_k = CB_k$ and

$$\left[\begin{array}{cc|c} I_r & W & A[:,k] \\ \mathbf{0} & \mathbf{0} & \end{array}\right] = A_kP = CB_kP = \left[\begin{array}{c|c} C_1 & C_2 \\ C_3 & C_4 \end{array}\right]\left[\begin{array}{cc|c} I_r & W & B[:,k] \\ \mathbf{0} & \mathbf{0} & \end{array}\right] = \left[\begin{array}{cc|c} C_1 & C_1W & CB[:,k] \\ C_3 & C_3W & \end{array}\right].$$

So, we see that $C_1 = I_r$, $C_3 = \mathbf{0}$ and $A[:,k] = \left[\begin{array}{c|c} I_r & C_2 \\ \mathbf{0} & C_4 \end{array}\right] B[:,k]$.

Case 1: Neither $A[:,k]$ nor $B[:,k]$ are pivotal. Then

$$\begin{bmatrix} X \\ \mathbf{0} \end{bmatrix} = A[:,k] = \left[\begin{array}{c|c} I_r & C_2 \\ \mathbf{0} & C_4 \end{array}\right] B[:,k] = \left[\begin{array}{c|c} I_r & C_2 \\ \mathbf{0} & C_4 \end{array}\right] \begin{bmatrix} Y \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} Y \\ \mathbf{0} \end{bmatrix}.$$

Thus, $X = Y$ and in this case the k-th columns are equal.

Case 2: $A[:,k]$ is pivotal but $B[:,k]$ in non-pivotal. Then

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_1 \end{bmatrix} = A[:,k] = \left[\begin{array}{c|c} I_r & C_2 \\ \mathbf{0} & C_4 \end{array}\right] B[:,k] = \left[\begin{array}{c|c} I_r & C_2 \\ \mathbf{0} & C_4 \end{array}\right] \begin{bmatrix} Y \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} Y \\ \mathbf{0} \end{bmatrix},$$

a contradiction as $\mathbf{e}_1 \neq \mathbf{0}$. Thus, this case cannot arise.

Therefore, combining both the cases, we get the required result. ∎

7.2 Permutation/Symmetric Groups

Definition 7.2.1. For a positive integer n, denote $[n] = \{1, 2, \ldots, n\}$. A function $f : A \to B$ is called

1. **one-one/injective** if $f(x) = f(y)$ for some $x, y \in A$ necessarily implies that $x = y$.
2. **onto/surjective** if for each $b \in B$ there exists $a \in A$ such that $f(a) = b$.
3. a **bijection** if f is both one-one and onto.

Example 7.2.2. Let $A = \{1, 2, 3\}$, $B = \{a, b, c, d\}$ and $C = \{\alpha, \beta, \gamma\}$. Then, the function

1. $j : A \to B$ defined by $j(1) = a, j(2) = c$ and $j(3) = c$ is neither one-one nor onto.
2. $f : A \to B$ defined by $f(1) = a, f(2) = c$ and $f(3) = d$ is one-one but not onto.
3. $g : B \to C$ defined by $g(a) = \alpha, g(b) = \beta, g(c) = \alpha$ and $g(d) = \gamma$ is onto but not one-one.
4. $h : B \to A$ defined by $h(a) = 2, h(b) = 2, h(c) = 3$ and $h(d) = 1$ is onto.
5. $h \circ f : A \to A$ is a bijection.
6. $g \circ f : A \to C$ is neither one-one not onto.

Remark 7.2.3. *Let $f : A \to B$ and $g : B \to C$ be functions. Then, the* **composition** *of functions, denoted $g \circ f$, is a function from A to C defined by $(g \circ f)(a) = g(f(a))$. Also, if*

1. *f and g are one-one then $g \circ f$ is one-one.*
2. *f and g are onto then $g \circ f$ is onto.*

Thus, if f and g are bijections then so is $g \circ f$.

Definition 7.2.4. A function $f : [n] \to [n]$ is called a **permutation** on n elements if f is a bijection. For example, $f, g : [2] \to [2]$ defined by $f(1) = 1, f(2) = 2$ and $g(1) = 2, g(2) = 1$ are permutations.

EXERCISE **7.2.5.** *Let $\mathcal{S}_3$ be the set consisting of all permutation on 3 elements. Then, prove that $\mathcal{S}_3$ has 6 elements. Moreover, they are one of the 6 functions given below.*

1. $f_1(1) = 1, f_1(2) = 2$ *and* $f_1(3) = 3$.
2. $f_2(1) = 1, f_2(2) = 3$ *and* $f_2(3) = 2$.
3. $f_3(1) = 2, f_3(2) = 1$ *and* $f_3(3) = 3$.
4. $f_4(1) = 2, f_4(2) = 3$ *and* $f_4(3) = 1$.
5. $f_5(1) = 3, f_5(2) = 1$ *and* $f_5(3) = 2$.
6. $f_6(1) = 3, f_6(2) = 2$ *and* $f_6(3) = 1$.

Remark 7.2.6. *Let $f : [n] \to [n]$ be a bijection. Then, the* **inverse** *of f, denote f^{-1}, is defined by $f^{-1}(m) = \ell$ whenever $f(\ell) = m$ for $m \in [n]$ is well defined and f^{-1} is a bijection. For example, in Exercise 7.2.5, note that $f_i^{-1} = f_i$, for $i = 1, 2, 3, 6$ and $f_4^{-1} = f_5$.*

Remark 7.2.7. *Let $\mathcal{S}_n = \{f : [n] \to [n] \ : \ \sigma \text{ is a permutation}\}$. Then, $\mathcal{S}_n$ has $n!$ elements and forms a* **group** *with respect to composition of functions, called* **product***, due to the following.*

1. *Let $f \in \mathbf{S}_n$. Then,*
   (a) *f can be written as* $f = \begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$, *called a* **two row notation**.
   (b) *f is one-one. Hence, $\{f(1), f(2), \ldots, f(n)\} = [n]$ and thus, $f(1) \in [n], f(2) \in [n] \setminus \{f(1)\}, \ldots$ and finally $f(n) = [n] \setminus \{f(1), \ldots, f(n-1)\}$. Therefore, there are n choices for $f(1)$, $n-1$ choices for $f(2)$ and so on. Hence, the number of elements in $\mathcal{S}_n$ equals $n(n-1)\cdots 2 \cdot 1 = n!$.*

2. *By Remark 7.2.3, $f \circ g \in \mathcal{S}_n$, for any $f, g \in \mathbf{S}_n$.*

3. *Also associativity holds as $f \circ (g \circ h) = (f \circ g) \circ h$ for all functions f, g and h.*

4. *$\mathcal{S}_n$ has a special permutation called the* **identity** *permutation, denoted Id_n, such that $Id_n(i) = i$, for $1 \le i \le n$.*

5. *For each $f \in \mathcal{S}_n$, $f^{-1} \in \mathcal{S}_n$ and is called the* **inverse** *of f as $f \circ f^{-1} = f^{-1} \circ f = Id_n$.*

Lemma 7.2.8. *Fix a positive integer n. Then, the group $\mathcal{S}_n$ satisfies the following:*

1. *Fix an element $f \in \mathcal{S}_n$. Then, $\mathcal{S}_n = \{f \circ g : g \in \mathcal{S}_n\} = \{g \circ f : g \in \mathcal{S}_n\}$.*

2. *$\mathcal{S}_n = \{g^{-1} : g \in \mathcal{S}_n\}$.*

Proof. Part 1: Note that for each $\alpha \in \mathcal{S}_n$ the functions $f^{-1} \circ \alpha, \alpha \circ f^{-1} \in \mathcal{S}_n$ and $\alpha = f \circ (f^{-1} \circ \alpha)$ as well as $\alpha = (\alpha \circ f^{-1}) \circ f$.

Part 2: Note that for each $f \in \mathcal{S}_n$, by definition, $(f^{-1})^{-1} = f$. Hence the result holds. □

Definition 7.2.9. Let $f \in \mathcal{S}_n$. Then, the number of inversions of f, denoted $n(f)$, equals

$$\begin{aligned} n(f) &= |\{(i,j) : i < j, f(i) > f(j)\}| \\ &= |\{j : i+1 \le j \le n, f(j) < f(i)\}| \text{ using two row notation.} \end{aligned} \tag{7.2.1}$$

Example 7.2.10. 1. For $f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$, $n(f) = |\{(1,2), (1,3), (2,3)\}| = 3$.

2. In Exercise 7.2.5, $n(f_5) = 2 + 0 = 2$.

3. Let $f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 2 & 3 & 5 & 1 & 9 & 8 & 7 & 6 \end{pmatrix}$. Then, $n(f) = 3+1+1+1+0+3+2+1 = 12$.

Definition 7.2.11. [Cycle Notation] Let $f \in \mathcal{S}_n$. Suppose there exist $r, 2 \le r \le n$ and $i_1, \ldots, i_r \in [n]$ such that $f(i_1) = i_2, f(i_2) = i_3, \ldots, f(i_r) = i_1$ and $f(j) = j$ for all $j \ne i_1, \ldots, i_r$. Then, we represent such a permutation by $f = (i_1, i_2, \ldots, i_r)$ and call it an **r-cycle**. For example, $f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix} = (1, 4, 5)$ and $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 4 & 5 \end{pmatrix} = (2, 3)$.

Remark 7.2.12. 1. *One also write the r-cycle $(i_1, i_2, \ldots, i_r)$ as $(i_2, i_3, \ldots, i_r, i_1)$ and so on. For example, $(1, 4, 5) = (4, 5, 1) = (5, 1, 4)$.*

2. *The permutation* $f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \end{pmatrix}$ *is not a cycle.*

3. *Let* $f = (1, 3, 5, 4)$ *and* $g = (2, 4, 1)$ *be two cycles. Then, their product, denoted* $f \circ g$ *or* $(1, 3, 5, 4)(2, 4, 1)$ *equals* $(1, 2)(3, 5, 4)$*. The calculation proceeds as (the arrows indicate the images):*
$1 \to 2$*. Note* $(f \circ g)(1) = f(g(1)) = f(2) = 2$.
$2 \to 4 \to 1$ *as* $(f \circ g)(2) = f(g(2)) = f(4) = 1$*. So,* $(1, 2)$ *forms a cycle.*
$3 \to 5$ *as* $(f \circ g)(3) = f(g(3)) = f(3) = 5$.
$5 \to 4$ *as* $(f \circ g)(5) = f(g(5)) = f(5) = 4$.
$4 \to 1 \to 3$ *as* $(f \circ g)(4) = f(g(4)) = f(1) = 3$*. So, the other cycle is* $(3, 5, 4)$.
4. *Let* $f = (1, 4, 5)$ *and* $g = (2, 4, 1)$ *be two permutations. Then,* $(1, 4, 5)(2, 4, 1) = (1, 2, 5)(4) = (1, 2, 5)$ *as* $1 \to 2, 2 \to 4 \to 5, 5 \to 1, 4 \to 1 \to 4$ *and*
$(2, 4, 1)(1, 4, 5) = (1)(2, 4, 5) = (2, 4, 5)$ *as* $1 \to 4 \to 1, 2 \to 4, 4 \to 5, 5 \to 1 \to 2$.
5. *Even though* $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 2 & 5 & 1 \end{pmatrix}$ *is not a cycle, verify that it is a product of the cycles* $(1, 4, 5)$ *and* $(2, 3)$.

Definition 7.2.13. A permutation $f \in \mathcal{S}_n$ is called a **transposition** if there exist $m, r \in [n]$ such that $f = (m, r)$.

Remark 7.2.14. *Verify that*

1. $(2, 4, 5) = (2, 5)(2, 4) = (4, 2)(4, 5) = (5, 4)(5, 2) = (1, 2)(1, 5)(1, 4)(1, 2)$.
2. *in general, the* r*-cycle* $(i_1, \ldots, i_r) = (1, i_1)(1, i_r)(1, i_{r-1}) \cdots (1, i_2)(1, i_1)$.
3. *So, every* r*-cycle can be written as product of transpositions. Furthermore, they can be written using the* n *transpositions* $(1, 2), (1, 3), \ldots, (1, n)$.

With the above definitions, we state and prove two important results.

Theorem 7.2.15. *Let* $f \in \mathcal{S}_n$*. Then,* f *can be written as product of transpositions.*

Proof. Note that using use Remark 7.2.14, we just need to show that f can be written as product of disjoint cycles.

Consider the set $S = \{1, f(1), f^{(2)}(1) = (f \circ f)(1), f^{(3)}(1) = (f \circ (f \circ f))(1), \ldots\}$. As S is an infinite set and each $f^{(i)}(1) \in [n]$, there exist i, j with $0 \le i < j \le n$ such that $f^{(i)}(1) = f^{(j)}(1)$. Now, let j_1 be the least positive integer such that $f^{(i)}(1) = f^{(j_1)}(1)$, for some i, $0 \le i < j_1$. Then, we claim that $i = 0$.

For if, $i - 1 \ge 0$ then $j_1 - 1 \ge 1$ and the condition that f is one-one gives

$$f^{(i-1)}(1) = (f^{-1} \circ f^{(i)})(1) = f^{-1}\left(f^{(i)}(1)\right) = f^{-1}\left(f^{(j_1)}(1)\right) = (f^{-1} \circ f^{(j_1)})(1) = f^{(j_1-1)}(1).$$

Thus, we see that the repetition has occurred at the $(j_1 - 1)$-th instant, contradicting the assumption that j_1 was the least such positive integer. Hence, we conclude that $i = 0$. Thus, $(1, f(1), f^{(2)}(1), \ldots, f^{(j_1-1)}(1))$ is one of the cycles in f.

Now, choose $i_1 \in [n] \setminus \{1, f(1), f^{(2)}(1), \ldots, f^{(j_1-1)}(1)\}$ and proceed as above to get another cycle. Let the new cycle by $(i_1, f(i_1), \ldots, f^{(j_2-1)}(i_1))$. Then, using f is one-one follows that

$$\left\{1, f(1), f^{(2)}(1), \ldots, f^{(j_1-1)}(1)\right\} \cap \left\{i_1, f(i_1), \ldots, f^{(j_2-1)}(i_1)\right\} = \emptyset.$$

So, the above process needs to be repeated at most n times to get all the disjoint cycles. Thus, the required result follows. $\square$

Remark 7.2.16. *Note that when one writes a permutation as product of disjoint cycles, cycles of length* 1 *are suppressed so as to match Definition 7.2.11. For example, the algorithm in the proof of Theorem 7.2.15 implies*

1. *Using Remark 7.2.14.3, we see that every permutation can be written as product of the* n *transpositions* $(1,2),(1,3),\ldots,(1,n)$.
2. $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 5 & 2 \end{pmatrix} = (1)(2,4,5)(3) = (2,4,5)$.
3. $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 2 & 3 & 5 & 1 & 9 & 8 & 7 & 6 \end{pmatrix} = (1,4,5)(2)(3)(6,9)(7,8) = (1,4,5)(6,9)(7,8)$.

Note that $Id_3 = (1,2)(1,2) = (1,2)(2,3)(1,2)(1,3)$, as well. The question arises, is it possible to write Id_n as a product of odd number of transpositions? The next lemma answers this question in negative.

Lemma 7.2.17. *Suppose there exist transpositions* f_i, $1 \le i \le t$, *such that*

$$Id_n = f_1 \circ f_2 \circ \cdots \circ f_t,$$

then t *is even.*

Proof. We will prove the result by mathematical induction. Observe that $t \neq 1$ as Id_n is not a transposition. Hence, $t \ge 2$. If $t = 2$, we are done. So, let us assume that the result holds for all expressions in which the number of transpositions $t \le k$. Now, let $t = k+1$.

Suppose $f_1 = (m,r)$ and let $\ell, s \in [n] \setminus \{m,r\}$. Then, the possible choices for the composition $f_1 \circ f_2$ are $(m,r)(m,r) = Id_n$, $(m,r)(m,\ell) = (r,\ell)(r,m)$, $(m,r)(r,\ell) = (\ell,r)(\ell,m)$ and $(m,r)(\ell,s) = (\ell,s)(m,r)$. In the first case, f_1 and f_2 can be removed to obtain $Id_n = f_3 \circ f_4 \circ \cdots \circ f_t$, where 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 remaining cases, the expression for $f_1 \circ f_2$ is replaced by their counterparts to obtain another expression for Id_n. But in the new expression for Id_n, m doesn't appear in the first transposition, but appears in the second transposition. The shifting of m to the right can continue till the number of transpositions reduces by 2 (which in turn gives the result by mathematical induction). For if, the shifting of m to the right doesn't reduce the number of transpositions then m will get shifted to the right and will appear only in the right most transposition. Then, this expression for Id_n does not fix m whereas $Id_n(m) = m$. So, the later case leads us to a contradiction. Hence, the shifting of m to the right will surely lead to an expression in which the number of transpositions at some stage is $t-2 = k-1$. At this stage, one applies mathematical induction to get the required result. $\square$

Theorem 7.2.18. *Let* $f \in \mathcal{S}_n$. *If there exist transpositions* $g_1, \ldots, g_k$ *and* $h_1, \ldots, h_\ell$ *with*

$$f = g_1 \circ g_2 \circ \cdots \circ g_k = h_1 \circ h_2 \circ \cdots \circ h_\ell$$

then, either k *and* ℓ *are both even or both odd.*

Proof. As $g_1 \circ \cdots \circ g_k = h_1 \circ \cdots \circ h_\ell$ and $h^{-1} = h$ for any transposition $h \in \mathcal{S}_n$, we have

$$Id_n = g_1 \circ g_2 \circ \cdots \circ g_k \circ h_\ell \circ h_{\ell-1} \circ \cdots \circ h_1.$$

Hence by Lemma 7.2.17, $k + \ell$ is even. Thus, either k and ℓ are both even or both odd. □

Definition 7.2.19. [Even and Odd Permutation] A permutation $f \in \mathcal{S}_n$ is called an

1. **even permutation** if f can be written as product of even number of transpositions.
2. **odd permutation** if f can be written as a product of odd number of transpositions.

Definition 7.2.20. Observe that if f and g are both even or both odd permutations, then $f \circ g$ and $g \circ f$ are both even. Whereas, if one of them is odd and the other even then $f \circ g$ and $g \circ f$ are both odd. We use this to define a function sgn : $\mathcal{S}_n \to \{1, -1\}$, called the **signature** of a permutation, by

$$\text{sgn}(f) = \begin{cases} 1 & \text{if } f \text{ is an even permutation} \\ -1 & \text{if } f \text{ is an odd permutation} \end{cases}.$$

Example 7.2.21. Consider the set $\mathcal{S}_n$. Then,

1. by Lemma 7.2.17, Id_n is an even permutation and $\text{sgn}(Id_n) = 1$.
2. a transposition, say f, is an odd permutation and hence $\text{sgn}(f) = -1$
3. using Remark 7.2.20, $\text{sgn}(f \circ g) = \text{sgn}(f) \cdot \text{sgn}(g)$ for any two permutations $f, g \in \mathcal{S}_n$.

We are now ready to define determinant of a square matrix A.

Definition 7.2.22. Let $A = [a_{ij}]$ be an $n \times n$ matrix with complex entries. Then, the **determinant** of A, denoted $\det(A)$, is defined as

$$\det(A) = \sum_{g \in \mathcal{S}_n} \text{sgn}(g) a_{1g(1)} a_{2g(2)} \ldots a_{ng(n)} = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(g) \prod_{i=1}^{n} a_{ig(i)}. \tag{7.2.2}$$

For example, if $\mathcal{S}_2 = \{Id, f = (1,2)\}$ then for $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, $\det(A) = \text{sgn}(Id) \cdot a_{1Id(1)} a_{2Id(2)} + \text{sgn}(f) \cdot a_{1f(1)} a_{2f(2)} = 1 \cdot a_{11} a_{22} + (-1) a_{12} a_{21} = 1 - 4 = -3$.

Observe that $\det(A)$ is a scalar quantity. Even though the expression for $\det(A)$ seems complicated at first glance, it is very helpful in proving the results related with "properties of determinant". We will do so in the next section. As another examples, we verify that this definition also matches for 3×3 matrices. So, let $A = [a_{ij}]$ be a 3×3 matrix. Then, using Equation (7.2.2),

$$\begin{aligned} \det(A) &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^{3} a_{i\sigma(i)} \\ &= \text{sgn}(f_1) \prod_{i=1}^{3} a_{if_1(i)} + \text{sgn}(f_2) \prod_{i=1}^{3} a_{if_2(i)} + \text{sgn}(f_3) \prod_{i=1}^{3} a_{if_3(i)} + \\ &\qquad \text{sgn}(f_4) \prod_{i=1}^{3} a_{if_4(i)} + \text{sgn}(f_5) \prod_{i=1}^{3} a_{if_5(i)} + \text{sgn}(f_6) \prod_{i=1}^{3} a_{if_6(i)} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

7.3 Properties of Determinant

Theorem 7.3.1 (Properties of Determinant)**.** *Let $A = [a_{ij}]$ be an $n \times n$ matrix.*

1. *If $A[i,:] = \mathbf{0}^T$ for some i then $\det(A) = 0$.*
2. *If $B = E_i(c)A$, for some $c \neq 0$ and $i \in [n]$ then $\det(B) = c\det(A)$.*
3. *If $B = E_{ij}A$, for some $i \neq j$ then $\det(B) = -\det(A)$.*
4. *If $A[i,:] = A[j,:]$ for some $i \neq j$ then $\det(A) = 0$.*
5. *Let B and C be two $n \times n$ matrices. If there exists $m \in [n]$ such that $B[i,:] = C[i,:] = A[i,:]$ for all $i \neq m$ and $C[m,:] = A[m,:] + B[m,:]$ then $\det(C) = \det(A) + \det(B)$.*
6. *If $B = E_{ij}(c)$, for $c \neq 0$ then $\det(B) = \det(A)$.*
7. *If A is a triangular matrix then $\det(A) = a_{11} \cdots a_{nn}$, the product of the diagonal entries.*
8. *If E is an $n \times n$ elementary matrix 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. *If A^T denotes the transpose of the matrix A then $\det(A) = \det(A^T)$.*

Proof. Part 1: Note that each sum in $\det(A)$ contains one entry from each row. So, each sum has an entry from $A[i,:] = \mathbf{0}^T$. Hence, each sum in itself is zero. Thus, $\det(A) = 0$.
Part 2: By assumption, $B[k,:] = A[k,:]$ for $k \neq i$ and $B[i,:] = cA[i,:]$. So,

$$\begin{aligned}\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} b_{k\sigma(k)} \right) b_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) c a_{i\sigma(i)} \\ &= c \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} a_{k\sigma(k)} = c\det(A).\end{aligned}$$

Part 3: Let $\tau = (i,j)$. Then, $\operatorname{sgn}(\tau) = -1$, by Lemma 7.2.8, $\mathcal{S}_n = \{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}$ and

$$\begin{aligned}\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \sum_{\sigma \circ \tau \in \mathcal{S}_n} \operatorname{sgn}(\sigma \circ \tau) \prod_{i=1}^{n} b_{i,(\sigma \circ \tau)(i)} \\ &= \sum_{\sigma \circ \tau \in \mathcal{S}_n} \operatorname{sgn}(\tau) \cdot \operatorname{sgn}(\sigma) \left(\prod_{k \neq i,j} b_{k\sigma(k)} \right) b_{i(\sigma \circ \tau)(i)} b_{j(\sigma \circ \tau)(j)} \\ &= \operatorname{sgn}(\tau) \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i,j} b_{k\sigma(k)} \right) b_{i\sigma(j)} b_{j\sigma(i)} = - \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} a_{k\sigma(k)} \\ &= -\det(A).\end{aligned}$$

Part 4: As $A[i,:] = A[j,:]$, $A = E_{ij}A$. Hence, by Part 3, $\det(A) = -\det(A)$. Thus, $\det(A) = 0$.

Part 5: By assumption, $C[i,:] = B[i,:] = A[i,:]$ for $i \neq m$ and $C[m,:] = B[m,:] + A[m,:]$. So,

$$
\begin{aligned}
\det(C) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} c_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{i \neq m} c_{i\sigma(i)} \right) c_{m\sigma(m)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{i \neq m} c_{i\sigma(i)} \right) (a_{m\sigma(m)} + b_{m\sigma(m)}) \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} + \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \det(A) + \det(B).
\end{aligned}
$$

Part 6: By assumption, $B[k,:] = A[k,:]$ for $k \neq i$ and $B[i,:] = A[i,:] + cA[j,:]$. So,

$$
\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} b_{k\sigma(k)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} b_{k\sigma(k)} \right) b_{i\sigma(i)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) (a_{i\sigma(i)} + c a_{j\sigma(j)}) \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) a_{i\sigma(i)} + c \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) a_{j\sigma(j)}) \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} a_{k\sigma(k)} + c \cdot 0 = \det(A). \qquad\qquad UsePart\ 4
\end{aligned}
$$

Part 7: Observe that if $\sigma \in \mathcal{S}_n$ and $\sigma \neq Id_n$ then $n(\sigma) \geq 1$. Thus, for every $\sigma \neq Id_n$, there exists $m \in [n]$ (depending on σ) such that $m > \sigma(m)$ or $m < \sigma(m)$. So, if A is triangular, $a_{m\sigma(m)} = 0$. So, for each $\sigma \neq Id_n$, $\prod_{i=1}^{n} a_{i\sigma(i)} = 0$. Hence, $\det(A) = \prod_{i=1}^{n} a_{ii}$. the result follows.
Part 8: Using Part 7, $\det(I_n) = 1$. By definition $E_{ij} = E_{ij} I_n$ and $E_i(c) = E_i(c) I_n$ and $E_{ij}(c) = E_{ij}(c) I_n$, for $c \neq 0$. Thus, using Parts 2, 3 and 6, we get $\det(E_i(c)) = c, \det(E_{ij}) = -1$ and $\det(E_{ij}(k)) = 1$. Also, again using Parts 2, 3 and 6, we get $\det(EA) = \det(E)\det(A)$.
Part 9: Suppose A is invertible. Then, by Theorem 2.5.1, $A = E_1 \cdots E_k$, for some elementary matrices $E_1, \ldots, E_k$. So, a repeated application of Part 8 implies $\det(A) = \det(E_1) \cdots \det(E_k) \neq 0$ as $\det(E_i) \neq 0$ for $1 \leq i \leq k$.

Now, suppose that $\det(A) \neq 0$. We need to show that A is invertible. On the contrary, assume that A is not invertible. Then, by Theorem 2.5.1, $\operatorname{Rank}(A) < n$. So, by Proposition 2.2.21, there exist elementary matrices $E_1, \ldots, E_k$ such that $E_1 \cdots E_k A = \begin{bmatrix} B \\ \mathbf{0} \end{bmatrix}$. Therefore, by Part 1 and a repeated application of Part 8 gives

$$
\det(E_1) \cdots \det(E_k) \det(A) = \det(E_1 \cdots E_k A) = \det\left(\begin{bmatrix} B \\ \mathbf{0} \end{bmatrix} \right) = 0.
$$

As $\det(E_i) \neq 0$, for $1 \leq i \leq k$, we have $\det(A) = 0$, a contradiction. Thus, A is invertible.
Part 10: Let A be invertible. Then, by Theorem 2.5.1, $A = E_1 \cdots E_k$, for some elementary matrices $E_1, \ldots, E_k$. So, applying Part 8 repeatedly gives $\det(A) = \det(E_1) \cdots \det(E_k)$ and

$$
\det(AB) = \det(E_1 \cdots E_k B) = \det(E_1) \cdots \det(E_k) \det(B) = \det(A)\det(B).
$$

In case A is not invertible, by Part 9, $\det(A) = 0$. Also, AB is not invertible (AB is invertible will imply A is invertible using the rank argument). So, again by Part 9, $\det(AB) = 0$. Thus, $\det(AB) = \det(A)\det(B)$.
Part 11: Let $B = [b_{ij}] = A^T$. Then, $b_{ij} = a_{ji}$, for $1 \le i, j \le n$. By Lemma 7.2.8, we know that $\mathcal{S}_n = \{\sigma^{-1} : \sigma \in \mathcal{S}_n\}$. As $\sigma \circ \sigma^{-1} = Id_n$, $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$. Hence,

$$\begin{aligned}\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i} = \sum_{\sigma^{-1} \in \mathcal{S}_n} \operatorname{sgn}(\sigma^{-1}) \prod_{i=1}^{n} a_{i\sigma^{-1}(i)} \\ &= \det(A).\end{aligned}$$

□

Remark 7.3.2. 1. *As* $\det(A) = \det(A^T)$, *we observe that in Theorem 7.3.1, the condition on "row" can be replaced by the condition on "column".*

2. *Let* $A = [a_{ij}]$ *be a matrix satisfying* $a_{1j} = 0$, *for* $2 \le j \le n$. *Let* $B = A(1 \mid 1)$, *the submatrix of* A *obtained by removing the first row and the first column. Then* $\det(A) = a_{11}\det(B)$.
Proof: *Let* $\sigma \in \mathcal{S}_n$ *with* $\sigma(1) = 1$. *Then,* σ *has a cycle* (1). *So, a disjoint cycle representation of* σ *only has numbers* $\{2, 3, \ldots, n\}$. *That is, we can think of* σ *as an element of* $\mathcal{S}_{n-1}$. *Hence,*

$$\begin{aligned}\det(A) &= \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n, \sigma(1)=1} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} \\ &= a_{11} \sum_{\sigma \in \mathcal{S}_n, \sigma(1)=1} sgn(\sigma) \prod_{i=2}^{n} a_{i\sigma(i)} = a_{11} \sum_{\sigma \in \mathcal{S}_{n-1}} sgn(\sigma) \prod_{i=1}^{n-1} b_{i\sigma(i)} = a_{11}\det(B).\end{aligned}$$

We now relate this definition of determinant with the one given in Definition 2.5.6.

Theorem 7.3.3. *Let* A *be an* $n \times n$ *matrix. Then,* $\det(A) = \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1 \mid j)\big)$, *where recall that* $A(1 \mid j)$ *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 an $n \times n$ matrix $B_j = \begin{bmatrix} 0 & 0 & \cdots & a_{1j} & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}$. Also, for each matrix B_j, we define the $n \times n$ matrix C_j by

1. $C_j[:, 1] = B_j[:, j]$,
2. $C_j[:, i] = B_j[:, i-1]$, for $2 \le i \le j$ and
3. $C_j[:, k] = B_j[:, k]$ for $k \ge j + 1$.

Also, observe that B_j's have been defined to satisfy $B_1[1, :] + \cdots + B_n[1, :] = A[1, :]$ and $B_j[i, :] = A[i, :]$ for all $i \ge 2$ and $1 \le j \le n$. Thus, by Theorem 7.3.1.5,

$$\det(A) = \sum_{j=1}^{n} \det(B_j). \tag{7.3.1}$$

Let us now compute $\det(B_j)$, for $1 \le j \le n$. Note that $C_j = E_{12}E_{23}\cdots E_{j-1,j}B_j$, for $1 \le j \le n$. Then, by Theorem 7.3.1.3, we get $\det(B_j) = (-1)^{j-1}\det(C_j)$. So, using Remark 7.3.2.2 and Theorem 7.3.1.2 and Equation (7.3.1), we have

$$\det(A) = \sum_{j=1}^{n}(-1)^{j-1}\det(C_j) = \sum_{j=1}^{n}(-1)^{j+1}a_{1j}\det\big(A(1 \mid j)\big).$$

Thus, we have shown that the determinant defined in Definition 2.5.6 is valid. $\square$

7.4 Dimension of $\mathbb{W}_1 + \mathbb{W}_2$

Theorem 7.4.1. *Let $\mathbb{V}$ be a finite dimensional vector space over $\mathbb{F}$ and let $\mathbb{W}_1$ and $\mathbb{W}_2$ be two subspaces of $\mathbb{V}$. Then,*

$$\dim(\mathbb{W}_1) + \dim(\mathbb{W}_2) = \dim(\mathbb{W}_1 + \mathbb{W}_2) + \dim(\mathbb{W}_1 \cap \mathbb{W}_2). \tag{7.4.1}$$

Proof. Since $\mathbb{W}_1 \cap \mathbb{W}_2$ is a vector subspace of V, let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis of $\mathbb{W}_1 \cap \mathbb{W}_2$. As, $\mathbb{W}_1 \cap \mathbb{W}_2$ is a subspace of both $\mathbb{W}_1$ and $\mathbb{W}_2$, let us extend the basis $\mathcal{B}$ to form a basis $\mathcal{B}_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{v}_1, \ldots, \mathbf{v}_s\}$ of $\mathbb{W}_1$ and a basis $\mathcal{B}_2 = \{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{w}_1, \ldots, \mathbf{w}_t\}$ of $\mathbb{W}_2$.

We now prove that $\mathcal{D} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{w}_1, \ldots, \mathbf{w}_s, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t\}$ is a basis of $\mathbb{W}_1 + \mathbb{W}_2$. To do this, we show that

1. $\mathcal{D}$ is linearly independent subset of $\mathbb{V}$ and
2. $LS(\mathcal{D}) = \mathbb{W}_1 + \mathbb{W}_2$.

The second part can be easily verified. For the first part, consider the linear system

$$\alpha_1\mathbf{u}_1 + \cdots + \alpha_r\mathbf{u}_r + \beta_1\mathbf{w}_1 + \cdots + \beta_s\mathbf{w}_s + \gamma_1\mathbf{v}_1 + \cdots + \gamma_t\mathbf{v}_t = \mathbf{0} \tag{7.4.2}$$

in the variables α_i's, β_j's and γ_k's. We re-write the system as

$$\alpha_1\mathbf{u}_1 + \cdots + \alpha_r\mathbf{u}_r + \beta_1\mathbf{w}_1 + \cdots + \beta_s\mathbf{w}_s = -(\gamma_1\mathbf{v}_1 + \cdots + \gamma_t\mathbf{v}_t).$$

Then, $\mathbf{v} = -\sum\limits_{i=1}^{s}\gamma_i\mathbf{v}_i \in LS(\mathcal{B}_1) = \mathbb{W}_1$. Also, $\mathbf{v} = \sum\limits_{j=1}^{r}\alpha_r\mathbf{u}_r + \sum\limits_{k=1}^{T}\beta_k\mathbf{w}_k$. So, $\mathbf{v} \in LS(\mathcal{B}_2) = \mathbb{W}_2$. Hence, $\mathbf{v} \in \mathbb{W}_1 \cap \mathbb{W}_2$ and therefore, there exists scalars $\delta_1, \ldots, \delta_k$ such that $\mathbf{v} = \sum\limits_{j=1}^{r}\delta_j\mathbf{u}_j$.

Substituting this representation of $\mathbf{v}$ in Equation (7.4.2), we get

$$(\alpha_1 - \delta_1)\mathbf{u}_1 + \cdots + (\alpha_r - \delta_r)\mathbf{u}_r + \beta_1\mathbf{w}_1 + \cdots + \beta_t\mathbf{w}_t = \mathbf{0}.$$

So, using Exercise 3.3.16.1, $\alpha_i - \delta_i = 0$, for $1 \le i \le r$ and $\beta_j = 0$, for $1 \le j \le t$. Thus, the system (7.4.2) reduces to

$$\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k + \gamma_1\mathbf{v}_1 + \cdots + \gamma_r\mathbf{v}_r = \mathbf{0}$$

which has $\alpha_i = 0$ for $1 \le i \le r$ and $\gamma_j = 0$ for $1 \le j \le s$ as the only solution. Hence, we see that the linear system of Equations (7.4.2) has no nonzero solution. Therefore, the set $\mathcal{D}$ is linearly independent and the set $\mathcal{D}$ is indeed a basis of $\mathbb{W}_1 + \mathbb{W}_2$. We now count the vectors in the sets $\mathcal{B}, \mathcal{B}_1, \mathcal{B}_2$ and $\mathcal{D}$ to get the required result. $\square$

7.5 When does Norm imply Inner Product

In this section, we prove the following result. A generalization of this result to complex vector space is left as an exercise for the reader as it requires similar ideas.

Theorem 7.5.1. *Let $\mathbb{V}$ be a real vector space. A norm $\|\cdot\|$ is induced by an inner product if and only if, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$, the norm satisfies*

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \qquad \text{(PARALLELOGRAM LAW)}. \tag{7.5.1}$$

Proof. Suppose that $\|\cdot\|$ is indeed induced by an inner product. Then, by Exercise 5.1.7.3 the result follows.

So, let us assume that $\|\cdot\|$ satisfies the parallelogram law. So, we need to define an inner product. We claim that the function $f : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ defined by

$$f(\mathbf{x}, \mathbf{y}) = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right), \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{V}$$

satisfies the required conditions for an inner product. So, let us proceed to do so.

STEP 1: Clearly, for each $\mathbf{x} \in \mathbb{V}$, $f(\mathbf{x}, \mathbf{0}) = 0$ and $f(\mathbf{x}, \mathbf{x}) = \frac{1}{4}\|\mathbf{x}+\mathbf{x}\|^2 = \|\mathbf{x}\|^2$. Thus, $f(\mathbf{x}, \mathbf{x}) \geq 0$. Further, $f(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

STEP 2: By definition $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$.

STEP 3: Now note that $\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 = 2\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2\right)$. Or equivalently,

$$2f(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2, \text{ for } \mathbf{x}, \mathbf{y} \in \mathbb{V}. \tag{7.5.2}$$

Thus, for $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{V}$, we have

$$\begin{aligned}
4\left(f(\mathbf{x}, \mathbf{y}) + f(\mathbf{z}, \mathbf{y})\right) &= \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 + \|\mathbf{z}+\mathbf{y}\|^2 - \|\mathbf{z}-\mathbf{y}\|^2 \\
&= 2\left(\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{z}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{z}\|^2 - 2\|\mathbf{y}\|^2\right) \\
&= \|\mathbf{x}+\mathbf{z}+2\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{z}\|^2 - \left(\|\mathbf{x}+\mathbf{z}\|^2 + \|\mathbf{x}-\mathbf{z}\|^2\right) - 4\|\mathbf{y}\|^2 \\
&= \|\mathbf{x}+\mathbf{z}+2\mathbf{y}\|^2 - \|\mathbf{x}+\mathbf{z}\|^2 - \|2\mathbf{y}\|^2 \\
&= 2f(\mathbf{x}+\mathbf{z}, 2\mathbf{y}) \text{ using Equation (7.5.2).}
\end{aligned} \tag{7.5.3}$$

Now, substituting $\mathbf{z} = \mathbf{0}$ in Equation (7.5.3) and using Equation (7.5.2), we get $2f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, 2\mathbf{y})$ and hence $4f(\mathbf{x}+\mathbf{z}, \mathbf{y}) = 2f(\mathbf{x}+\mathbf{z}, 2\mathbf{y}) = 4\left(f(\mathbf{x}, \mathbf{y}) + f(\mathbf{z}, \mathbf{y})\right)$. Thus,

$$f(\mathbf{x}+\mathbf{z}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) + f(\mathbf{z}, \mathbf{y}), \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{V}. \tag{7.5.4}$$

STEP 4: Using Equation (7.5.4), $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x})$ and the principle of mathematical induction, it follows that $nf(\mathbf{x}, \mathbf{y}) = f(n\mathbf{x}, \mathbf{y})$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ and $n \in \mathbb{N}$. Another application of Equation (7.5.4) with $f(\mathbf{0}, \mathbf{y}) = 0$ implies that $nf(\mathbf{x}, \mathbf{y}) = f(n\mathbf{x}, \mathbf{y})$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ and $n \in \mathbb{Z}$. Also, for $m \neq 0$,

$$mf\left(\frac{n}{m}\,\mathbf{x}, \mathbf{y}\right) = f(m\frac{n}{m}\,\mathbf{x}, \mathbf{y}) = f(n\mathbf{x}, \mathbf{y}) = nf(\mathbf{x}, \mathbf{y}).$$

Hence, we see that for all $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ and $a \in \mathbb{Q}$, $f\left(a\mathbf{x}, \mathbf{y}\right) = af(\mathbf{x}, \mathbf{y})$.

STEP 5: Fix $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ and define a function $g : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} g(x) &= f(x\mathbf{u}, \mathbf{v}) - xf(\mathbf{u}, \mathbf{v}) \\ &= \frac{1}{2}\left(\|x\mathbf{u} + \mathbf{v}\|^2 - \|x\mathbf{u}\|^2 - \|\mathbf{v}\|^2\right) - \frac{x}{2}\left(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2\right). \end{aligned}$$

Then, by previous step $g(x) = 0$, for all $x \in \mathbb{Q}$. So, if g is a continuous function then continuity implies $g(x) = 0$, for all $x \in \mathbb{R}$. Hence, $f(x\mathbf{u}, \mathbf{v}) = xf(\mathbf{u}, \mathbf{v})$, for all $x \in \mathbb{R}$.

Note that the second term of $g(x)$ is a constant multiple of x and hence continuous. Using a similar reason, it is enough to show that $g_1(x) = \|x\mathbf{u} + \mathbf{v}\|$, for certain fixed vectors $\mathbf{u}, \mathbf{v} \in \mathbb{V}$, is continuous. To do so, note that

$$\|x_1\mathbf{u} + \mathbf{v}\| = \|(x_1 - x_2)\mathbf{u} + x_2\mathbf{u} + \mathbf{v}\| \le \|(x_1 - x_2)\mathbf{u}\| + \|x_2\mathbf{u} + \mathbf{v}\|.$$

Thus, $\Big|\|x_1\mathbf{u} + \mathbf{v}\| - \|x_2\mathbf{u} + \mathbf{v}\|\Big| \le \|(x_1 - x_2)\mathbf{u}\|$. Hence, taking the limit as $x_1 \to x_2$, we get $\lim\limits_{x_1 \to x_2} \|x_1\mathbf{u} + \mathbf{v}\| = \|x_2\mathbf{u} + \mathbf{v}\|$.

Thus, we have proved the continuity of g and hence the prove of the required result. ■

7.6 Roots of a Polynomials

The main aim of this subsection is to prove the continuous dependence of the zeros of a polynomial on its coefficients and to recall Descartes's rule of signs.

Definition 7.6.1. [Jordan Curves] A **curve** in $\mathbb{C}$ is a continuous function $f : [a, b] \to \mathbb{C}$, where $[a, b] \subseteq \mathbb{R}$.

1. If the function f is one-one on $[a, b)$ and also on $(a, b]$, then it is called a **simple curve**.
2. If $f(b) = f(a)$, then it is called a **closed curve**.
3. A closed simple curve is called a **Jordan curve**.
4. The derivative (integral) of a curve $f = u + iv$ is defined component wise. If f' is continuous on $[a, b]$, we say f is a $\mathcal{C}^1$**-curve** (at end points we consider one sided derivatives and continuity).
5. A $\mathcal{C}^1$-curve on $[a, b]$ is called a **smooth curve**, if f' is never zero on (a, b).
6. A piecewise smooth curve is called a **contour**.
7. A positively oriented simple closed curve is called a **simple closed curve** such that while traveling on it the interior of the curve always stays to the left. (Camille Jordan has proved that such a curve always divides the plane into two connected regions, one of which is called the **bounded** region and the other is called the **unbounded** region. The one which is bounded is considered as the interior of the curve.)

We state the famous Rouche Theorem of complex analysis without proof.

Theorem 7.6.2. [Rouche's Theorem] *Let C be a positively oriented simple closed contour. Also, let f and g be two analytic functions on R_C, the union of the interior of C and the curve C itself. Assume also that $|f(x)| > |g(x)|$, for all $x \in C$. Then, f and $f + g$ have the same number of zeros in the interior of C.*

Corollary 7.6.3. **[Alen Alexanderian, The University of Texas at Austin, USA.]** *Let* $P(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ *have distinct roots* $\lambda_1, \ldots, \lambda_m$ *with multiplicities* $\alpha_1, \ldots, \alpha_m$*, respectively. Take any* $\epsilon > 0$ *for which the balls* $\overline{B_\epsilon(\lambda_i)}$ *are disjoint. Then, there exists a* $\delta > 0$ *such that the polynomial* $q(t) = t^n + a'_{n-1}t^{n-1} + \cdots + a'_0$ *has exactly* α_i *roots (counting with multiplicities) in* $B_\epsilon(\lambda_i)$*, whenever* $|a_j - a'_j| < \delta$*.*

Proof. For an $\epsilon > 0$ and $1 \le i \le m$, let $C_i = \{z \in \mathbb{C} : |z - \lambda_i| = \epsilon\}$. Now, for each $i, 1 \le i \le m$, take $\nu_i = \min\limits_{z \in C_i} |p(z)|$, $\rho_i = \max\limits_{z \in C_i}[1 + |z| + \cdots + |z|^{n-1}]$ and choose $\delta > 0$ such that $\rho_i \delta < \nu_i$. Then, for a fixed j and $z \in C_j$, we have

$$|q(z) - P(z)| = |(a'_{n-1} - a_{n-1})z^{n-1} + \cdots + (a'_0 - a_0)| \le \delta\rho_j < \nu_j \le |p(z)|.$$

Hence, by Rouche's theorem, $p(z)$ and $q(z)$ have the same number of zeros inside C_j, for each $j = 1, \ldots, m$. That is, the zeros of $q(t)$ are within the ϵ-neighborhood of the zeros of $P(t)$. ∎

As a direct application, we obtain the following corollary.

Corollary 7.6.4. *Eigenvalues of a matrix are continuous functions of its entries.*

Proof. Follows from Corollary 7.6.3. ∎

Remark 7.6.5.

1. **[Sign changes in a polynomial]** *Let* $P(x) = \sum_0^n a_i x^{n-i}$ *be a real polynomial, with* $a_0 \neq 0$*. Read the coefficient from the left* $a_0, a_1, \ldots$*. We say the* SIGN CHANGES OF a_i OCCUR AT $m_1 < m_2 < \cdots < m_k$ *to mean that* a_{m_1} *is the first after* a_0 *with sign opposite to* a_0*;* a_{m_2} *is the first after* a_{m_1} *with sign opposite to* a_{m_1}*; and so on.*

2. **[Descartes' Rule of Signs]** *Let* $P(x) = \sum_0^n a_i x^{n-i}$ *be a real polynomial. Then, the maximum number of positive roots of* $P(x) = 0$ *is the number of changes in sign of the coefficients and that the maximum number of negative roots is the number of sign changes in* $P(-x) = 0$*.*

   Proof. Assume that $a_0, a_1, \cdots, a_n$ *has* $k > 0$ *sign changes. Let* $b > 0$*. Then, the coefficients of* $(x - b)P(x)$ *are*

   $$a_0, a_1 - ba_0, a_2 - ba_1, \cdots, a_n - ba_{n-1}, -ba_n.$$

   This list has at least $k + 1$ *changes of signs. To see this, assume that* $a_0 > 0$ *and* $a_n \neq 0$*. Let the sign changes of* a_i *occur at* $m_1 < m_2 < \cdots < m_k$*. Then, setting*

   $$c_0 = a_0, c_1 = a_{m_1} - ba_{m_1 - 1}, c_2 = a_{m_2} - ba_{m_2 - 1}, \cdots, c_k = a_{m_k} - ba_{m_k - 1}, c_{k+1} = -ba_n,$$

   we see that $c_i > 0$ *when* i *is even and* $c_i < 0$*, when* i *is odd. That proves the claim.*

   Now, assume that $P(x) = 0$ *has* k *positive roots* $b_1, b_2, \cdots, b_k$*. Then,*

   $$P(x) = (x - b_1)(x - b_2) \cdots (x - b_k)Q(x),$$

   where $Q(x)$ *is a real polynomial. By the previous observation, the coefficients of* $(x - b_k)Q(x)$ *has at least one change of signs, coefficients of* $(x - b_{k-1})(x - b_k)Q(x)$ *has at least two, and so on. Thus coefficients of* $P(x)$ *has at least* k *change of signs. The rest of the proof is similar.* ∎

7.7 Variational characterizations of Hermitian Matrices

Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then, by Theorem 6.2.22, we know that all the eigenvalues of A are real. So, we write $\lambda_i(A)$ to mean the i-th smallest eigenvalue of A. That is, the i-th from the left in the list $\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_n(A)$.

Lemma 7.7.1. **[Rayleigh-Ritz Ratio]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Then,*

1. $\lambda_1(A)\mathbf{x}^*\mathbf{x} \leq \mathbf{x}^*A\mathbf{x} \leq \lambda_n(A)\mathbf{x}^*\mathbf{x}$, *for each* $\mathbf{x} \in \mathbb{C}^n$.
2. $\lambda_1(A) = \min\limits_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \min\limits_{\|\mathbf{x}\|=1} \mathbf{x}^*A\mathbf{x}$.
3. $\lambda_n(A) = \max\limits_{\mathbf{x}\neq 0} \frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \max\limits_{\|\mathbf{x}\|=1} \mathbf{x}^*A\mathbf{x}$.

Proof. Proof of Part 1: By spectral theorem (see Theorem 6.2.22, there exists a unitary matrix U such that $A = UDU^*$, where $D = \text{diag}(\lambda_1(A), \ldots, \lambda_n(A))$ is a real diagonal matrix. Thus, the set $\{U[:,1], \ldots, U[:,n]\}$ is a basis of $\mathbb{C}^n$. Hence, for each $\mathbf{x} \in \mathbb{C}^n$, there exists **Ans** $:_i$'s (scalar) such that $\mathbf{x} = \sum \alpha_i U[:,i]$. So, note that $\mathbf{x}^*\mathbf{x} = |\alpha_i|^2$ and

$$\lambda_1(A)\mathbf{x}^*\mathbf{x} = \lambda_1(A)\sum |\alpha_i|^2 \leq \sum |\alpha_i|^2\lambda_i(A) = \mathbf{x}^*A\mathbf{x} \leq \lambda_n \sum |\alpha_i|^2 = \lambda_n\mathbf{x}^*\mathbf{x}.$$

For Part 2 and Part 3, take $\mathbf{x} = U[:,1]$ and $\mathbf{x} = U(:,n)$, respectively. ∎

As an immediate corollary, we state the following result.

Corollary 7.7.2. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and $\alpha = \dfrac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}$. Then, A has an eigenvalue in the interval $(-\infty, \alpha]$ and has an eigenvalue in the interval $[\alpha, \infty)$.*

We now generalize the second and third parts of Theorem 7.7.2.

Proposition 7.7.3. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix with $A = UDU^*$, where U is a unitary matrix and D is a diagonal matrix consisting of the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then, for any positive integer $k, 1 \leq k \leq n$,*

$$\lambda_k = \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp U[:,1],\ldots,U[:,k-1]}} \mathbf{x}^*A\mathbf{x} = \max_{\substack{\|\mathbf{x}\|=1 \\ x\perp U[:,n],\ldots,U[:,k+1]}} \mathbf{x}^*A\mathbf{x}.$$

Proof. Let $\mathbf{x} \in \mathbb{C}^n$ such that $\mathbf{x}$ is orthogonal to $U[,1], \ldots, U[:,k-1]$. Then, we can write $\mathbf{x} = \sum\limits_{i=k}^{n} \alpha_i U[:,i]$, for some scalars α_i's. In that case,

$$\lambda_k\mathbf{x}^*\mathbf{x} = \lambda_k \sum_{i=k}^{n} |\alpha_i|^2 \leq \sum_{i=k}^{n} |\alpha_i|^2\lambda_i = \mathbf{x}^*A\mathbf{x}$$

and the equality occurs for $\mathbf{x} = U[:,k]$. Thus, the required result follows. ∎

Theorem 7.7.4. **[Courant-Fischer]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then,*

$$\lambda_k = \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} \mathbf{x}^*A\mathbf{x} = \min_{\mathbf{w}_n,\ldots,\mathbf{w}_{k+1}} \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp\mathbf{w}_n,\ldots,\mathbf{w}_{k+1}}} \mathbf{x}^*A\mathbf{x}.$$

Proof. Let $A = UDU^*$, where U is a unitary matrix and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Now, choose a set of $k-1$ linearly independent vectors from $\mathbb{C}^n$, say $S = \{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}$. Then, some of the eigenvectors $\{U[:,1], \ldots, U[:,k-1]\}$ may be an element of $S^\perp$. Thus, using Proposition 7.7.3, we see that

$$\lambda_k = \min_{\substack{\|\mathbf{x}\|=1,\\ \mathbf{x}\perp U[:,1],\ldots,U[:,k-1]}} \mathbf{x}^* A\mathbf{x} \geq \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\in S^\perp}} \mathbf{x}^* A\mathbf{x}.$$

Hence, $\lambda_k \geq \max\limits_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min\limits_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} \mathbf{x}^* A\mathbf{x}$, for each choice of $k-1$ linearly independent vectors. But, by Proposition 7.7.3, the equality holds for the linearly independent set $\{U[:,1], \ldots, U[:,k-1]\}$ which proves the first equality. A similar argument gives the second equality and hence the proof is omitted. ∎

Theorem 7.7.5. **[Weyl Interlacing Theorem]** *Let $A, B \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrices. Then, $\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B)$. In particular, if $B = P^*P$, for some matrix P, then $\lambda_k(A+B) \geq \lambda_k(A)$. In particular, for $\mathbf{z} \in \mathbb{C}^n$, $\lambda_k(A + \mathbf{z}\mathbf{z}^*) \leq \lambda_{k+1}(A)$.*

Proof. As A and B are Hermitian matrices, the matrix $A+B$ is also Hermitian. Hence, by Courant-Fischer theorem and Lemma 7.7.1.1,

$$\begin{aligned}
\lambda_k(A+B) &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} \mathbf{x}^*(A+B)\mathbf{x}\\
&\leq \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} [\mathbf{x}^* A\mathbf{x} + \lambda_n(B)] = \lambda_k(A) + \lambda_n(B)
\end{aligned}$$

and

$$\begin{aligned}
\lambda_k(A+B) &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} \mathbf{x}^*(A+B)\mathbf{x}\\
&\geq \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} [\mathbf{x}^* A\mathbf{x} + \lambda_1(B)] = \lambda_k(A) + \lambda_1(B).
\end{aligned}$$

If $B = P^*P$, then $\lambda_1(B) = \min\limits_{\|\mathbf{x}\|=1} \mathbf{x}^*(P^*P)\mathbf{x} = \min\limits_{\|\mathbf{x}\|=1} \|P\mathbf{x}\|^2 \geq 0$. Thus,

$$\lambda_k(A+B) \geq \lambda_k(A) + \lambda_1(B) \geq \lambda_k(A).$$

In particular, for $\mathbf{z} \in \mathbb{C}^n$, we have

$$\begin{aligned}
\lambda_k(A+\mathbf{z}\mathbf{z}^*) &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}}} [\mathbf{x}^* A\mathbf{x} + |\mathbf{x}^*\mathbf{z}|^2]\\
&\leq \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1},\mathbf{z}}} [\mathbf{x}^* A\mathbf{x} + |\mathbf{x}^*\mathbf{z}|^2]\\
&= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \min_{\substack{\|x\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1},\mathbf{z}}} \mathbf{x}^* A\mathbf{x}\\
&\leq \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1},\mathbf{w}_k} \min_{\substack{\|x\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{k-1},\mathbf{w}_k}} \mathbf{x}^* A\mathbf{x} = \lambda_{k+1}(A).
\end{aligned}$$

∎

Theorem 7.7.6. **[Cauchy Interlacing Theorem]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix. Define $\hat{A} = \begin{bmatrix} A & \mathbf{y} \\ \mathbf{y}^* & a \end{bmatrix}$, for some $a \in \mathbb{R}$ and $\mathbf{y} \in \mathbb{C}^n$. Then,*

$$\lambda_k(\hat{A}) \leq \lambda_k(A) \leq \lambda_{k+1}(\hat{A}).$$

Proof. Note that

$$\begin{aligned} \lambda_{k+1}(\hat{A}) &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathbb{C}^{n+1}} \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_k}} \mathbf{x}^*\hat{A}\mathbf{x} \leq \max_{\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathbb{C}^{n+1}} \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_k \\ \mathbf{x}_{n+1}=0}} \mathbf{x}^*\hat{A}\mathbf{x} \\ &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathbb{C}^{n}} \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_k}} \mathbf{x}^*A\mathbf{x} = \lambda_{k+1}(A) \end{aligned}$$

and

$$\begin{aligned} \lambda_{k+1}(\hat{A}) &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k} \in \mathbb{C}^{n+1}} \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_{n-k}}} \mathbf{x}^*\hat{A}\mathbf{x} \geq \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k} \in \mathbb{C}^{n+1}} \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_{n-k} \\ \mathbf{x}_{n+1}=0}} \mathbf{x}^*\hat{A}\mathbf{x} \\ &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k} \in \mathbb{C}^{n}} \max_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}\perp \mathbf{w}_1,\ldots,\mathbf{w}_{n-k}}} \mathbf{x}^*A\mathbf{x} = \lambda_k(A) \end{aligned}$$

■

As an immediate corollary, one has the following result.

Corollary 7.7.7. **[Inclusion principle]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and r be a positive integer with $1 \leq r \leq n$. If $B_{r\times r}$ is a principal submatrix of A then, $\lambda_k(A) \leq \lambda_k(B) \leq \lambda_{k+n-r}(A)$.*

Theorem 7.7.8. **[Poincare Separation Theorem]** *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \subseteq \mathbb{C}^n$ be an orthonormal set for some positive integer $r, 1 \leq r \leq n$. If further $B = [b_{ij}]$ is an $r \times r$ matrix with $b_{ij} = \mathbf{u}_i^* A \mathbf{u}_j$, $1 \leq i, j \leq r$ then, $\lambda_k(A) \leq \lambda_k(B) \leq \lambda_{k+n-r}(A)$.*

Proof. Let us extend the orthonormal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis, say $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathbb{C}^n$ and write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then, B is a $r \times r$ principal submatrix of U^*AU. Thus, by inclusion principle, $\lambda_k(U^*AU) \leq \lambda_k(B) \leq \lambda_{k+n-r}(U^*AU)$. But, we know that $\sigma(U^*AU) = \sigma(A)$ and hence the required result follows. ■

The proof of the next result is left for the reader.

Corollary 7.7.9. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and r be a positive integer with $1 \leq r \leq n$. Then,*

$$\lambda_1(A) + \cdots + \lambda_r(A) = \min_{U^*U=I_r} \mathrm{tr} U^*AU \quad \text{and} \quad \lambda_{n-r+1}(A) + \cdots + \lambda_n(A) = \max_{U^*U=I_r} \mathrm{tr} U^*AU.$$

Corollary 7.7.10. *Let $A \in \mathbb{M}_n(\mathbb{C})$ be a Hermitian matrix and W be a k-dimensional subspace of $\mathbb{C}^n$. Suppose, there exists a real number c such that $\mathbf{x}^*A\mathbf{x} \geq c\mathbf{x}^*\mathbf{x}$, for each $\mathbf{x} \in W$. Then, $\lambda_{n-k+1}(A) \geq c$. In particular, if $\mathbf{x}^*A\mathbf{x} > 0$, for each nonzero $x \in W$, then $\lambda_{n-k+1} > 0$. $\Big($Note that, a k-dimensional subspace need not contain an eigenvector of A. For example, the line $y = 2x$ does not contain an eigenvector of $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.$\Big)$*

Proof. Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n-k}\}$ be a basis of $W^\perp$. Then,

$$\lambda_{n-k+1}(A) = \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}} \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}}} \mathbf{x}^*A\mathbf{x} \geq \min_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{x}_1,\ldots,\mathbf{x}_{n-k}}} \mathbf{x}^*A\mathbf{x} \geq c.$$

Now assume that $\mathbf{x}^*A\mathbf{x} > 0$ holds for each nonzero $\mathbf{x} \in W$ and that $\lambda_{n-k+1} = 0$. Then, it follows that $\min\limits_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\perp\mathbf{x}_1,\ldots,\mathbf{x}_{n-k}}} x^*Ax = 0$. Now, define $f : \mathbb{C}^n \to \mathbb{C}$ by $f(\mathbf{x}) = \mathbf{x}^*A\mathbf{x}$.

Then, f is a continuous function and $\min\limits_{\substack{\|\mathbf{x}\|=1\\ \mathbf{x}\in W}} f(\mathbf{x}) = 0$. Thus, f must attain its bound on the unit sphere. That is, there exists $\mathbf{y} \in W$ with $\|\mathbf{y}\| = 1$ such that $\mathbf{y}^*A\mathbf{y} = 0$, a contradiction. Thus, the required result follows. ■