Lecture Notes on Linear Algebra

Arbind K Lal Sukant Pati

July 10, 2016

Contents

Chapter 1

Introduction to Matrices

1.1 Definition of a Matrix

Definition 1.1.1.1. Matrix A rectangular array of numbers is called a **matrix**.

The horizontal arrays of a matrix are called its **rows** and the vertical arrays are called its **columns**. 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. To be concise, one writes $A_{m\times n} = [a_{ij}]$ or, in short, $A = [a_{ij}]$. We will also use $A[i,:]$ to denote the i-th row of A, $A[:,j]$ to denote the j-th column of A. We shall mostly be concerned with matrices having complex numbers as entries.

For example, if $A = \begin{bmatrix} 1 & 3+\mathbf{i} & 7 \\ 4 & 5 & 6-5\mathbf{i} \end{bmatrix}$ then $A[1,:] = [1 \quad 3+\mathbf{i} \quad 7]$, $A[:,3] = \begin{bmatrix} 7 \\ 6-5\mathbf{i} \end{bmatrix}$ and $a_{22} = 5$. In general, in row vector commas are inserted to differentiate between entries. Thus, $A[1,:] = [1,\ 3+\mathbf{i},\ 7]$. 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. Thus, $A[1,:]$ is a row vector and $A[:,3]$ is a column vector.

Definition 1.1.1.2. Equality of two Matrices Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ having the same order $m \times n$ are **equal** if $a_{ij} = b_{ij}$, for each $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

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

Example 1.1.1.3. The linear system of equations $2x+3y = 5$ and $3x+2y = 5$ can be identified with the matrix $A = \left[\begin{array}{cc|c} 2 & 3 & 5 \\ 3 & 2 & 5 \end{array}\right]$. Note that x and y are unknowns with the understanding that x is associated with $A[:,1]$ and y is associated with $A[:,2]$.

1.1.A Special Matrices

Definition 1.1.1.4. 1. A matrix in which each entry is zero is called a zero-matrix, denoted $\mathbf{0}$. For example,

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

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

3. Let $A = [a_{ij}]$ be an $n \times n$ square matrix.

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

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

   (c) If $A = \text{diag}[a_{11}, \ldots, a_{nn}]$ and $a_{ii} = d$ for all $i = 1, \ldots, n$ then the diagonal matrix A is called a **scalar matrix**.

   (d) 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 $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

   (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.

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.

1.2 Operations on Matrices

Definition 1.1.2.1. Transpose of a Matrix Let $A = [a_{ij}]$ be an $m \times n$ matrix with real entries. Then the **transpose** of A, denoted $A^T = [b_{ij}]$, is an $n \times m$ matrix with $b_{ij} = a_{ji}$, for all i, j.

Definition 1.1.2.2. Conjugate Transpose of a Matrix Let $A = [a_{ij}]$ be an $m \times n$ matrix with complex entries. Then the **conjugate transpose** of A, denoted A^*, is an $n \times m$ matrix with $(A^*)_{ij} = \overline{a_{ji}}$, for all i, j, where for $a \in \mathbb{C}$, $\overline{a}$ denotes the complex-conjugate of a.

Thus, if $\mathbf{x}$ is a column vector then $\mathbf{x}^T$ and $\mathbf{x}^*$ are row vectors and vice-versa. For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ then $A^* = A^T = \begin{bmatrix} 1 & 0 \\ 4 & 1 \\ 5 & 2 \end{bmatrix}$, whereas if $A = \begin{bmatrix} 1 & 4+\mathbf{i} \\ 0 & 1-\mathbf{i} \end{bmatrix}$ then $A^* = \begin{bmatrix} 1 & 0 \\ 4-\mathbf{i} & 1+\mathbf{i} \end{bmatrix}$ and note that $A^* \neq A^T$.

Theorem 1.1.2.3. *For any matrix A, $(A^*)^* = A$. Thus, $(A^T)^T = A$.*

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 and hence the result follows. ∎

Remark 1.1.2.4. *Note that transpose is studied whenever the entries of the matrix are real. Since, we are allowing the matrix entries to be complex numbers, we will state and prove the results for complex-conjugate. The readers should separately very that similar results hold for transpose whenever the matrix has real entries.*

Definition 1.1.2.5. Addition of Matrices Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $m \times n$ matrices. Then, the **sum** of A and B, denoted $A + B$, is defined to be the matrix $C = [c_{ij}]$ with $c_{ij} = a_{ij} + b_{ij}$.

Definition 1.1.2.6. Multiplying a Scalar to a Matrix Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then the **product** of $k \in \mathbb{C}$ with A, denoted kA, is defined as $kA = [ka_{ij}]$.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ then $5A = \begin{bmatrix} 5 & 20 & 25 \\ 0 & 5 & 10 \end{bmatrix}$ and $(2+\mathbf{i})A = \begin{bmatrix} 2+\mathbf{i} & 8+4\mathbf{i} & 10+5\mathbf{i} \\ 0 & 2+\mathbf{i} & 4+2\mathbf{i} \end{bmatrix}$.

Theorem 1.1.2.7. *Let A, B and C be matrices of order $m \times n$, 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. Part 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 reader is required to prove the other parts as all the results follow from the properties of complex numbers. ∎

Definition 1.1.2.8. Additive Inverse Let A be an $m \times n$ matrix.

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

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

EXERCISE **1.1.2.9.** *1. Find a 3×3 non-zero matrix A with real entries satisfying*

   (a) $A^T = A$.

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

2. Find a 3×3 non-zero matrix A with complex entries satisfying

   (a) $A^* = A$.

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

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

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

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

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

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

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

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

1.2.A Multiplication of Matrices

Definition 1.1.2.10. Matrix Multiplication / Product Let $A = [a_{ij}]$ be an $m \times n$ matrix and $B = [b_{ij}]$ be an $n \times r$ matrix. The **product** of A and B, denoted AB, is a matrix $C = [c_{ij}]$ of order $m \times r$ with

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

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

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

$$AB = \begin{bmatrix} a\alpha + bx + cu & a\beta + by + cv & a\gamma + bz + cw & a\delta + bt + cs \\ d\alpha + ex + fu & d\beta + ey + fv & d\gamma + ez + fw & d\delta + et + fs \end{bmatrix}. \qquad (1.1.2.1)$$

Thus, note that the rows of the matrix AB can be written directly as

$$\begin{aligned}(AB)[1,:] &= a\,[\alpha,\beta,\gamma,\delta] + b\,[x,y,z,t] + c\,[u,v,w,s] = aB[1,:] + bB[2,:] + cB[3,:]\\ (AB)[2,:] &= dB[1,:] + eB[2,:] + fB[3,:] \end{aligned} \tag{1.1.2.2}$$

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

$$(AB)[:,1] = \begin{bmatrix} a\alpha + bx + cu \\ d\alpha + ex + fu \end{bmatrix} = \alpha\ A[:,1] + x\ A[:,2] + u\ A[:,3], \tag{1.1.2.3}$$

$(AB)[:,2] = \beta\ A[:,1] + y\ A[:,2] + v\ A[:,3], \cdots, (AB)[:,4] = \delta\ A[:,1] + t\ A[:,2] + s\ A[:,3].$

Remark 1.1.2.11. *Observe the following:*

1. *In this example, while AB is defined, the product BA is not defined. However, for square matrices A and B of the same order, both the product AB and BA are defined.*

2. *The product AB corresponds to operating on the* rows of the matrix B *(see Equation (1.1.2.2)). This is* **row method** *for calculating the matrix product.*

3. *The product AB also corresponds to operating on the* columns of the matrix A *(see Equation (1.1.2.3)). This is* **column method** *for calculating the matrix product.*

4. *Let A and B be two matrices such that the product AB is defined. Then verify that*

$$AB = \begin{bmatrix} A[1,:]B \\ A[2,:]B \\ \vdots \\ A[n,:]B \end{bmatrix} = [A\ B[:,1],\ A\ B[:,2],\ \ldots,\ A\ B[:,p]]. \tag{1.1.2.4}$$

Example 1.1.2.12. Let $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$. Use the row/column method of matrix multiplication to

1. find the second row of the matrix AB.
   Solution: By Remark 1.1.2.11.4, $(AB)[2,:] = A[2,:]B$ and hence

$$(AB)[2,:] = 1 \cdot [1,0,-1] + 0 \cdot [0,0,1] + 1 \cdot [0,-1,1] = [1,-1,0].$$

2. find the third column of the matrix AB.
   Solution: Again, by Remark 1.1.2.11.4, $(AB)[:,3] = A\ B[:,3]$ and hence

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

Definition 1.1.2.13. [Commutativity of Matrix Product] Two square matrices A and B are said to **commute** if $AB = BA$.

Remark 1.1.2.14. *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 $AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = BA$.*

Theorem 1.1.2.15. *Suppose that the matrices A, B and C are so chosen that the matrix multiplications are defined.*

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

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

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

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

5. *Now let A be a square matrix of order n and $D = diag[d_1, d_2, \ldots, d_n]$. Then*
   - *$(DA)[i, :] = d_i\ A[i, :]$, for $1 \leq i \leq n$, and*
   - *$(AD)[:, j] = d_jA[:, j]$, for $1 \leq j \leq n$.*

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

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

Therefore,

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

Part 5. As D is a diagonal matrix, using Remark 1.1.2.11.4, we have

$$(DA)[i, :] = D[i, :]\ A = \sum_{j\neq i} 0 \cdot A[j, :] + d_iA[i, :].$$

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

Exercise **1.1.2.16.**

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

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

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

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

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

6. Let A be a $1 \times n$ matrix and B be an $n \times 1$ matrix. Then verify that AB is a 1×1 matrix, whereas BA has order $n \times n$.

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

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

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

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

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

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

8. 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 $(B^*A^*)[2,:]$.

9. 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?

10. 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 $C^3 = 3C^2$?

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

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

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

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

12. Let a, b and c be indeterminate. Then, can we find A with complex entries satisfying $A \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a+b \\ b-c \\ 3a-5b+c \end{bmatrix}$? What if $A \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \cdot b \\ a \end{bmatrix}$? Give reasons for your answer.

1.2.B Inverse of a Matrix

Definition 1.1.2.17. [Inverse of a Matrix] Let A be a square matrix of order n.

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

Lemma 1.1.2.18. *Let A be an $n \times n$ matrix. Suppose that there exist $n \times n$ matrices B and C such that $AB = I_n$ and $CA = I_n$ then $B = C$.*

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

Remark 1.1.2.19. 1. *Lemma 1.1.2.18 implies that whenever A is invertible, the inverse is unique.*

2. *Therefore the inverse of A is denoted by A^{-1}. That is, $AA^{-1} = A^{-1}A = I$.*

Example 1.1.2.20. 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) The matrices $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 4 & 0 \end{bmatrix}$ and $\begin{bmatrix} 4 & 2 \\ 6 & 3 \end{bmatrix}$ do not have inverses.

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

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 1.1.2.21. *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. Proof of Part 1. Let $B = A^{-1}$ be the inverse of A. Then $AB = BA = I$. Thus, by definition, B is invertible and $B^{-1} = A$. Or equivalently, $(A^{-1})^{-1} = A$.

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

Proof of Part 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 2.3.A.

EXERCISE **1.1.2.22.** 1. *Let A be an invertible matrix. Then $(A^{-1})^r = A^{-r}$ for all integer r.*

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

3. *Let $A_1, \ldots, A_r$ be invertible matrices. Then the matrix $B = A_1A_2\cdots A_r$ is also invertible.*

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

5. *Let A be an $n \times n$ invertible matrix. Then prove that*
   (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 $n \geq 3$.*

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

7. *Determine A that satisfies* $(I - A)^{-1} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -2 \\ 1 & -2 & 1 \end{bmatrix}$. *[See Example 1.1.2.20.2].*

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

9. *Let $A = [a_{ij}]$ be an invertible matrix. If $B = [p^{i-j}a_{ij}]$ for some $p \in \mathbb{C}$, $p \neq 0$ then determine B^{-1}.*

1.3 Some More Special Matrices

Definition 1.1.3.1. 1. Let A be a square matrix with real entries. Then, A is called

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

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

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

2. Let A be a square matrix with complex entries. Then, A is called

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

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

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

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

3. A matrix A is said to be **idempotent** if $A^2 = A$. For example, $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ is idempotent.

4. A matrix that is symmetric and idempotent is called a **projection** matrix. For example, let $\mathbf{u} \in \mathbb{R}^n$ be a column vector with $\mathbf{u}^T\mathbf{u} = 1$ then $A = \mathbf{u}\mathbf{u}^T$ is an idempotent matrix. Moreover, A is symmetric and hence is a projection matrix. In particular, let $\mathbf{u} = \frac{1}{\sqrt{5}}(1,2)^T$ and $A = \mathbf{u}\mathbf{u}^T$. Then $\mathbf{u}^T\mathbf{u} = 1$ and for any vector $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$ note that

$$A\mathbf{x} = (\mathbf{u}\mathbf{u}^*)\mathbf{x} = \mathbf{u}(\mathbf{u}^*\mathbf{x}) = \frac{x_1 + 2x_2}{\sqrt{5}}\mathbf{u} = \left[\frac{x_1 + 2x_2}{5}, \frac{2x_1 + 4x_2}{5}\right]^T.$$

Thus, $A\mathbf{x}$ is the feet of the perpendicular from the point $\mathbf{x}$ on the vector $[1\ \ 2]^T$.

5. A square matrix 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 $A = [a_{ij}]$ is an $n \times n$ matrix with a_{ij} equal to 1 if $i - j = 1$ and 0, otherwise then $A^n = \mathbf{0}$ and $A^\ell \neq \mathbf{0}$ for $1 \le \ell \le n-1$.

EXERCISE **1.1.3.2.** *1. 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 $A = S_1 + S_2$.*

2. Let A and B be two lower triangular matrices. Then prove that AB is a lower triangular matrix. A similar statement holds for upper triangular matrices.

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

4. *Show that the diagonal entries of a skew-Hermitian matrix are zero or purely imaginary.*

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

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

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

8. *Are the matrices* $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$ *and* $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \sin\theta & -\cos\theta \end{bmatrix}$ *orthogonal, for* $\theta \in [0,\ 2\pi]$?

9. *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 $i \ne j$. Then prove that the 3×3 matrix*

   (a) *$U = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$ satisfies $U^*U = I$. Thus, $UU^* = I$.*

   (b) *$A = \mathbf{u}_i\mathbf{u}_i^*$, for $1 \le i \le 3$, satisfies $A^2 = A$. Is A symmetric?*

   (c) *$A = \mathbf{u}_i\mathbf{u}_i^* + \mathbf{u}_j\mathbf{u}_j^*$, for $i \ne j$, satisfies $A^2 = A$. Is A symmetric?*

10. *Verify that the matrices in Exercises 9.9b and 9.9c are projection matrices.*

11. *Let A and B be two $n \times n$ orthogonal matrices. Then prove that AB is also an orthogonal matrix.*

1.3.A sub-matrix of a Matrix

Definition 1.1.3.3. A matrix obtained by deleting some of the rows and/or columns of a matrix is said to be a **sub-matrix** of the given matrix.

For example, if $A = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 2 \end{bmatrix}$ then $[1], [2], \begin{bmatrix} 1 \\ 0 \end{bmatrix}, [1\ 5], \begin{bmatrix} 1 & 5 \\ 0 & 2 \end{bmatrix}$, A are a few sub-matrices of A. But the matrices $\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 4 \\ 0 & 2 \end{bmatrix}$ are not sub-matrices of A (the reader is advised to give reasons).

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

Theorem 1.1.3.4. *Let $A = [a_{ij}] = [P \; Q]$ and $B = [b_{ij}] = \begin{bmatrix} H \\ K \end{bmatrix}$ be defined as above. Then*

$$AB = PH + QK.$$

Proof. The matrix products PH and QK are valid as the order of the matrices P, H, Q and K are respectively, $n \times r$, $r \times p$, $n \times (m-r)$ and $(m-r) \times p$. Also, the matrices PH and QK are of the same order and hence their sum is justified. 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 1.1.3.5. *Theorem 1.1.3.4 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 nice forms. This partition may be quite useful during different matrix operations.*
3. *If we want to prove results using induction then after proving the initial step, one assume the result for all $r \times r$ sub-matrices and then try to prove it for $(r+1) \times (r+1)$ sub-matrices.*

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

Suppose $A = \begin{array}{c c} & \begin{array}{cc} m_1 & m_2 \end{array} \\ \begin{array}{c} n_1 \\ n_2 \end{array} & \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \end{array}$ and $B = \begin{array}{c c} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \end{array} & \begin{bmatrix} E & F \\ G & H \end{bmatrix} \end{array}$. Then the matrices P, Q, R, S and E, F, G, H, are called the blocks of the matrices A and B, respectively. Note that even if $A + B$ is defined, the orders of P and E need not be the same. But, if the block sums are defined then $A + B = \begin{bmatrix} P+E & Q+F \\ R+G & S+H \end{bmatrix}$. Similarly, if the product AB is defined, the product PE may not be defined. Again, if the block products are defined, one can verify that $AB = \begin{bmatrix} PE+QG & PF+QH \\ RE+SG & RF+SH \end{bmatrix}$. That is, once a partition of A is fixed, the partition of B has to be properly chosen for purposes of block addition or multiplication.

EXERCISE **1.1.3.6.**

1. *Complete the proofs of Theorems 1.1.2.7 and 1.1.2.15.*
2. *Let $A = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 2 & 2 & 6 \\ 2 & 1 & 2 & 5 \\ 3 & 3 & 4 & 10 \end{bmatrix}$. Compute*

(a) $(AC)[1,:]$,

(b) $(B(AC))[1,:]$, $(B(AC))[2,:]$ and $(B(AC))[3,:]$.

(c) Note that $(B(AC))[:,1] + (B(AC))[:,2] + (B(AC))[:,3] - (B(AC))[:,4] = \mathbf{0}$.

(d) Let $\mathbf{x}^T = [1, 1, 1, -1]$. Use previous result to prove $C\mathbf{x} = \mathbf{0}$.

3. 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 $B = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$. Then

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

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

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

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

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

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

4. Fix a unit vector $\mathbf{a} \in \mathbb{R}^n$ and define $f : \mathbb{R}^n \to \mathbb{R}^n$ by $f(\mathbf{y}) = 2(\mathbf{a}^T\mathbf{y})\mathbf{a} - \mathbf{y}$. Does this function give a reflection about the line that contains the points $\mathbf{0}$ and $\mathbf{a}$.

5. Consider the two coordinate transformations

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

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

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

(c) Is $C = AB$? Give reasons for your answer.

6. For $A_{n\times n} = [a_{ij}]$, the **trace** of A, denoted $\text{Tr}(A)$, is defined by $\text{Tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$.

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

(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 $\text{Tr}(AB)$.

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

i. $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$.

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

(d) Prove that one cannot find matrices A and B such that $AB - BA = cI$ for any $c \neq 0$.

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

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

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

8. *Let A be an $n \times n$ matrix such that $AB = BA$ for all $n \times n$ matrices B. Then prove that A is a scalar matrix. That is, $A = \alpha I$ for some $\alpha \in \mathbb{C}$.*

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

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

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

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

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

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

12. *Let* $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & c \end{bmatrix}$, *where A_{11} is an $n \times n$ invertible matrix and $c \in \mathbb{C}$.*

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

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

    is the inverse of A.

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

13. *Let $\mathbf{x}$ be an $n \times 1$ vector with real entries and satisfying $\mathbf{x}^T\mathbf{x} = 1$.*

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

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

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

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

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

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

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

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

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

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

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

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

17. *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.*

18. *Let A be an $m \times n$ matrix. Then a matrix G of order $n \times m$ is called a* **generalized 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.*

1.4 Summary

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

1. The zero matrix of size $m \times n$, denoted $\mathbf{0}_{m\times n}$ or $\mathbf{0}$.

2. The identity matrix of size $n \times n$, denoted I_n or I.

3. Triangular matrices.

4. Hermitian/Symmetric matrices.

5. Skew-Hermitian/skew-symmetric matrices.

6. Unitary/Orthogonal matrices.

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 to a matrix A is same as operating on the rows of A.

2. right to a matrix A is same as operating on the columns of A.

Chapter 2

System of Linear Equations over $\mathbb{R}$

2.1 Introduction

Let us look at some examples of linear systems.

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

$$a_1x + b_1y = c_1,\ a_2x + b_2y = c_2$$

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

   (a) UNIQUE SOLUTION
   $x + 2y = 1$ and $x + 3y = 1$. The unique solution is $[x, y]^T = [1, 0]^T$.
   Observe that in this case, $a_1b_2 - a_2b_1 \neq 0$.

   (b) INFINITE NUMBER OF SOLUTIONS
   $x + 2y = 1$ and $2x + 4y = 2$. As both equations represent the same line, the solution set is $[x, y]^T = [1 - 2y, y]^T = [1, 0]^T + y[-2, 1]^T$ with y arbitrary. Observe that

   i. $a_1b_2 - a_2b_1 = 0,\ a_1c_2 - a_2c_1 = 0$ and $b_1c_2 - b_2c_1 = 0$.

   ii. the vector $[1, 0]^T$ corresponds to the solution $x = 1, y = 0$ of the given system.

   iii. the vector $[-2, 1]^T$ corresponds to the solution $x = -2, y = 1$ of the system $x + 2y = 0, 2x + 4y = 0$.

(c) NO SOLUTION

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

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

(a) UNIQUE SOLUTION

Consider the system $x+y+z = 3$, $x+4y+2z = 7$ and $4x+10y-z = 13$. The unique solution to this system is $[x, y, z]^T = [1, 1, 1]^T$, *i.e.*, THE THREE PLANES INTERSECT AT A POINT.

(b) INFINITE NUMBER OF SOLUTIONS

Consider the system $x + y + z = 3$, $x + 2y + 2z = 5$ and $3x + 4y + 4z = 11$. The solution set is $[x, y, z]^T = [1, 2 - z, z]^T = [1, 2, 0]^T + z[0, -1, 1]^T$, with z arbitrary. Observe the following:

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

ii. The vector $[1, 2, 0]^T$ corresponds to the solution $x = 1, y = 2$ and $z = 0$ of the linear system $x + y + z = 3$, $x + 2y + 2z = 5$ and $3x + 4y + 4z = 11$. Also, the vector $[0, -1, 1]^T$ corresponds to the solution $x = 0, y = -1$ and $z = 1$ of the linear system $x + y + z = 0$, $x + 2y + 2z = 0$ and $3x + 4y + 4z = 0$.

(c) NO SOLUTION

The system $x+y+z = 3$, $2x+2y+2z = 5$ and $3x+3y+3z = 3$ has no solution. In this case, we have three parallel planes. The readers are advised to supply the proof.

Definition 2.2.1.1. [Linear System] A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \tag{2.2.1.1}$$

where for $1 \leq i \leq n$ and $1 \leq j \leq m$; $a_{ij}, b_i \in \mathbb{R}$. Linear System (2.2.1.1) is called **homogeneous** if $b_1 = 0 = b_2 = \cdots = b_m$ and **non-homogeneous** , otherwise.

Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$. Then (2.2.1.1) 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 (2.2.1.1).

Remark 2.2.1.2. *Consider the augmented matrix $[A\ \mathbf{b}]$ of the linear system $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix, $\mathbf{b}$ and $\mathbf{x}$ are column vectors of appropriate size. If $\mathbf{x}^T = [x_1, \ldots, x_n]$ then it is important to note that*

1. *the unknown x_1 corresponds to the column $([A\ \mathbf{b}])[:, 1]$.*
2. *in general, for $j = 1, 2, \ldots, n$, the unknown x_j corresponds to the column $([A\ \mathbf{b}])[:, j]$.*
3. *the vector $\mathbf{b} = ([A\ \mathbf{b}])[:, n+1]$.*
4. *for $i = 1, 2, \ldots, m$, the i^{th} equation corresponds to the row $([A\ \mathbf{b}])[i, :]$.*

Definition 2.2.1.3. [Consistent, Inconsistent] A 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 the system $x + y = 2, 2x + 2y = 1$ is inconsistent.

Definition 2.2.1.4. Consider the linear system $A\mathbf{x} = \mathbf{b}$. Then the corresponding linear system $A\mathbf{x} = \mathbf{0}$ is called the **associated homogeneous system**. As mentioned in the previous paragraph, the associated homogeneous system is always consistent.

Definition 2.2.1.5. [Solution of a Linear System] 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 $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 2 \\ 4 & 10 & -1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 7 \\ 13 \end{bmatrix}$ equals $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

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

Theorem 2.2.1.6. *Consider the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.*

1. *Then $\mathbf{x} = \mathbf{0}$, the zero vector, is always a 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 $c \in \mathbb{R}$.*
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 all $a_i \in \mathbb{R}, 1 \leq i \leq k$.*

Remark 2.2.1.7. *Consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then*

1. *the vector $\mathbf{0}$ is called the **trivial** solution.*
2. *a non-zero solution is called a **non-trivial** solution. For example, for the system $A\mathbf{x} = \mathbf{0}$, where $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, the vector $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a non-trivial solution.*

3. *Thus, by Theorem 2.2.1.6, the existence of a non-trivial solution of* $A\mathbf{x} = \mathbf{0}$ *is equivalent to having an infinite number of solutions for the system* $A\mathbf{x} = \mathbf{0}$.

4. *Let* $\mathbf{u}, \mathbf{v}$ *be two distinct solutions of the non-homogeneous system* $A\mathbf{x} = \mathbf{b}$. *Then* $\mathbf{x}_h = \mathbf{u} - \mathbf{v}$ *is a solution of the homogeneous system* $A\mathbf{x} = \mathbf{0}$. *That is, any two solutions of* $A\mathbf{x} = \mathbf{b}$ *differ by a solution of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$. *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 **2.2.1.8.** 1. *Consider a system of* 2 *equations in* 3 *unknowns. If this system is consistent then how many solutions does it have?*

2. *Define a linear system of* 3 *equations in* 2 *unknowns such that the system is inconsistent.*

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

4. *Let* $A\mathbf{x} = \mathbf{b}$ *be a system of* m *equations and* n *unknowns. Then*

   (a) *determine the possible solution set if* $m \geq 3$ *and* $n = 2$.

   (b) *determine the possible solution set if* $m \geq 4$ *and* $n = 3$.

   (c) *can this system have exactly two distinct solutions?*

   (d) *can have only a finitely many (greater than* 1*) solutions?*

2.1.A Elementary Row Operations

Example 2.2.1.9. Solve the linear system $y + z = 2,\ 2x + 3z = 5,\ x + y + z = 3$.

Solution: Let $B_0 = [A \quad \mathbf{b}]$, 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} equation (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 $\frac{-2}{3}$ (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 2.2.1.9, observe that for each operation on the system of linear equations there is a corresponding operation on the row of the augmented matrix. We use this idea to define elementary row operations and the equivalence of two linear systems.

Definition 2.2.1.10. [Elementary Row Operations] Let A be an $m \times n$ matrix. Then the **elementary row operations** are

1. E_{ij}: Interchange of $A[i,:]$ and $A[j,:]$.
2. $E_k(c)$ for $c \neq 0$: Multiply $A[k,:]$ by c.
3. $E_{ij}(c)$ for $c \neq 0$: Replace $A[i,:]$ by $A[i,:] + cA[j,:]$.

Definition 2.2.1.11. [Equivalent Linear Systems] Consider the linear systems $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ with corresponding augmented matrices, $[A \ \ \mathbf{b}]$ and $[C \ \ \mathbf{d}]$, respectively. Then the two linear systems are said to be **equivalent** if $[C \ \ \mathbf{d}]$ can be obtained from $[A \ \ \mathbf{b}]$ by application of a finite number of elementary row operations.

Definition 2.2.1.12. [Equivalent Matrices] Two matrices are said to be **equivalent** if one can be obtained from the other by a finite number of elementary row operations.

Thus, note that the linear systems at each step in Example 2.2.1.9 are equivalent to each other. We now prove that the solution set of two equivalent linear systems are same.

Lemma 2.2.1.13. *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

$$(a_{j1} + ca_{k1})\alpha_1 + \cdots + (a_{jn} + ca_{kn})\alpha_n = b_j + cb_k. \tag{2.2.1.2}$$

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

$$(a_{j1} + ca_{k1})x_1 + \cdots + (a_{jn} + ca_{kn})x_n = b_j + cb_k. \tag{2.2.1.3}$$

Therefore, using Equation (2.2.1.2), we see that $\mathbf{y}^T = [\alpha_1, \ldots, \alpha_n]$ is also a solution for Equation (2.2.1.3). 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 2.2.1.13 as an induction step to prove the main result of this subsection which is stated next.

Theorem 2.2.1.14. *Let $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ be two equivalent linear systems. Then they have the same solution set.*

2.2 System of Linear Equations

In the previous section, we saw that if one linear system can be obtained from another by a repeated application of elementary row operations then the two linear systems have the same solution set. Sometimes it helps to imagine an elementary row operation as a product on the left by elementary matrix. In this section, we will try to understand this relationship and use them to first obtain results for the system of linear equations and then to the theory of square matrices.

2.2.A Elementary Matrices and the Row-Reduced Echelon Form (RREF)

Definition 2.2.2.1. A square matrix E of order n is called an **elementary matrix** if it is obtained by applying exactly one elementary row operation to the identity matrix I_n.

Remark 2.2.2.2. *The elementary matrices are of three types and they correspond to elementary row operations.*

1. *E_{ij}: Matrix obtained by applying elementary row operation E_{ij} to I_n.*
2. *$E_k(c)$ for $c \neq 0$: Matrix obtained by applying elementary row operation $E_k(C)$ to I_n.*

3. *$E_{ij}(c)$ for $c \neq 0$: Matrix obtained by applying elementary row operation $E_{ij}(c)$ to I_n.*

Example 2.2.2.3. 1. In particular, for $n = 3$ and a real number $c \neq 0$, one has

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

2. Let $A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 4 & 5 & 6 \\ 2 & 0 & 3 & 4 \end{bmatrix}$. Then verify that

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

3. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 4 & 4 \end{bmatrix}$. Then $E_{21}(-1)E_{32}(-2)A = E_{21}(-1) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

4. Let $A = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix}$. Then verify that $E_{31}(-2)E_{13}E_{31}(-1)A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 3 & 3 \end{bmatrix}$.

EXERCISE **2.2.2.4.** *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 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. *Determine elementary matrices $F_1, \ldots, F_k$ such that $E_k \cdots E_1 \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix} = I_3$.*

Remark 2.2.2.5. *Observe that*

1. *$(E_{ij})^{-1} = E_{ij}$ as $E_{ij}E_{ij} = I = E_{ij}E_{ij}$.*

2. *Let $c \neq 0$. Then $(E_k(c))^{-1} = E_k(1/c)$ as $E_k(c)E_k(1/c) = I = E_k(1/c)E_k(c)$.*

3. *Let $c \neq 0$. Then $(E_{ij}(c))^{-1} = E_{ij}(-c)$ as $E_{ij}(c)E_{ij}(-c) = I = E_{ij}(-c)E_{ij}(c)$.*

Thus, each elementary matrix is invertible and the inverse is also an elementary matrix.

Based on the above observation and the fact that product of invertible matrices is invertible, the readers are advised to prove the next result.

Lemma 2.2.2.6. *Prove that applying elementary row operations is equivalent to multiplying on the left by the corresponding elementary matrix and vice-versa.*

Proposition 2.2.2.7. *Let A and B be two equivalent matrices. Then prove that $B = E_1 \cdots E_k A$, for some elementary matrices $E_1, \ldots, E_k$.*

Proof. By definition of equivalence, the matrix B can be obtained from A by a finite number of elementary row operations. But by Lemma 2.2.2.6, each elementary row operation on A corresponds to multiplying on the left of A by an elementary matrix. Thus, the required result follows. ■

We now give a direct prove of Theorem 2.2.1.14. To do so, we state the theorem once again.

Theorem 2.2.2.8. *Let $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ be two equivalent linear systems. Then they have the same solution set.*

Proof. Let $E_1, \ldots, E_k$ be the elementary matrices such that $E_1 \cdots E_k[A \;\; \mathbf{b}] = [C \;\; \mathbf{d}]$. Then, by definition of matrix product and Remark 2.2.2.5

$$E_1 \cdots E_k A = C, E_1 \cdots E_k \mathbf{b} = \mathbf{d} \text{ and } A = (E_1 \cdots E_k)^{-1} C, \mathbf{b} = (E_1 \cdots E_k)^{-1} \mathbf{d}. \tag{2.2.2.1}$$

Now assume that $A\mathbf{y} = \mathbf{b}$ holds. Then, by Equation (2.2.2.1)

$$C\mathbf{y} = E_1 \cdots E_k A\mathbf{y} = E_1 \cdots E_k \mathbf{b} = \mathbf{d}. \tag{2.2.2.2}$$

On the other hand if $C\mathbf{z} = \mathbf{d}$ holds then using Equation (2.2.2.1), we have

$$A\mathbf{z} = (E_1 \cdots E_k)^{-1} C\mathbf{z} = (E_1 \cdots E_k)^{-1} \mathbf{d} = \mathbf{b}. \tag{2.2.2.3}$$

Therefore, using Equations (2.2.2.2) and (2.2.2.3) the required result follows. ■

As an immediate corollary, the readers are advised to prove the following result.

Corollary 2.2.2.9. *Let A and B be two equivalent matrices. Then the systems $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solution set.*

Example 2.2.2.10. Are the matrices $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix}$ 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 2.2.2.11. [Pivot/Leading Term] Let A be a non-zero matrix. Then, a **pivot/leading term** is the first (from left) nonzero element of a non-zero row in A and the column containing the pivot term is called the **pivot 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, the columns 1 and 2 are pivot columns.

Definition 2.2.2.12. [Echelon Form] A matrix A is in **echelon form (EF)** (ladder like) if

1. pivot of the $(i+1)$-th row comes to the right of the i-th.
2. entries below the pivot in a pivotal column are 0.
3. the zero rows are at the bottom.

Example 2.2.2.13. 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 2.2.2.14. [Row-Reduced Echelon Form] A matrix C is said to be in the **row-reduced echelon form (RREF)** if

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

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

Example 2.2.2.15. 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).

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

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

Theorem 2.2.2.16. *Let A and B be two matrices in RREF. If they are equivalent then $A = B$.*

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

Corollary 2.2.2.17. *The RREF of a matrix is unique.*

Proof. Suppose there exists a matrix A with two different RREFs, say B and C. As the RREFs are obtained by 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 $C = F_1 \cdots F_\ell A$. Thus,

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

As inverse of elementary matrices are elementary matrices, we see that the matrices B and C are equivalent. As B and C are in RREF, using Theorem 2.2.2.16, we see that $B = C$. ∎

Theorem 2.2.2.18. *Let A be an $m \times n$ matrix. If B consists of the first s columns of A then RREF(B) equals the first s columns of RREF(A).*

Proof. Let us write $F = \text{RREF}(A)$. By definition of RREF, there exist elementary matrices $E_1, \ldots, E_k$ such that $E_1 \cdots E_k A = F$. Then, by matrix multiplication

$$E_1 \cdots E_k A = [E_1 \cdots E_k A[:,1], \ldots, E_1 \cdots E_k A[:,n]] = [F[:,1], \ldots, F[:,n]].$$

Thus, $E_1 \cdots E_k B = [E_1 \cdots E_k A[:,1], \ldots, E_1 \cdots E_k A[:,s]] = [F[:,1], \ldots, F[:,s]]$. Since the matrix F is in RREF, by definition, it's first s columns are also in RREF. Hence, by Corollary 2.2.2.17 we see that $\text{RREF}(B) = [F[:,1], \ldots, F[:,s]]$. Thus, the required result follows. ∎

Let A an $m \times n$ matrix. Then by Corollary 2.2.2.17, it's RREF is unique. We use it to define our next object.

2.2.B Rank of a Matrix

Definition 2.2.2.19. [Row-Rank of a Matrix] Let A be an $m \times n$ matrix and let the number of pivots (number of non-zero rows) in it's RREF. Then the **row rank** of A, denoted row-rank(A), equals r. For example, row-rank$(I_n) = n$ and row-rank$(\mathbf{0}) = 0$.

Remark 2.2.2.20.

1. *Even though, row-rank is defined using the RREF of a matrix, we just need to compute the echelon form as the number of non-zero rows/pivots do not change when we proceed to compute the RREF from the echelon form.*

2. *Let A be an $m \times n$ matrix. Then by the definition of RREF, the number of pivots, say r, satisfies $r \leq \min\{m, n\}$. Thus, row-rank$(A) \leq \min\{m, n\}$.*

Example 2.2.2.21. Determine the row-rank of the following matrices.

1. $\text{diag}(d_1, \ldots, d_n)$.
   Solution: Let $S = \{i \mid 1 \leq i \leq n, d_i \neq 0\}$. Then, verify that row-rank equals the number of elements in S.

2. $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix}$.
   Solution: The echelon form of A is obtained as follows:

$$\begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix} \xrightarrow{E_{31}(-1), E_{21}(-2)} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \xrightarrow{E_{32}(-1)} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & -2 & 0 \end{bmatrix}.$$

   As the echelon form of A has 3 non-zero rows row-rank$(A) = 3$.

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

Solution: row-rank$(A) = 2$ as the echelon form of A (given below) has two non-zero rows:

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix} \overset{E_{31}(-1), E_{21}(-2)}{\to} \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \end{bmatrix} \overset{E_{32}(-1)}{\to} \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Remark 2.2.2.22. *Let $A\mathbf{x} = \mathbf{b}$ be a linear system with m equations in n unknowns. Let RREF$([A\ \mathbf{b}]) = [C\ \mathbf{d}]$, row-rank$(A) = r$ and row-rank$([A\ \mathbf{b}]) = r_a$.*

1. *Then, using Theorem 2.2.2.18 conclude that $r \leq r_a$.*
2. *If $r < r_a$ then again using Theorem 2.2.2.18, note that $r_a = r + 1$ and $([C\ \mathbf{d}])[:, n+1]$ has a pivot at the $(r+1)$-th place. Hence, by definition of RREF, $([C\ \mathbf{d}])[r+1, :] = [\mathbf{0}^T,\ 1]$.*
3. *If $r = r_a$ then $([C\ \mathbf{d}])[:, n+1]$ has no pivot. Thus, $[\mathbf{0}^T,\ 1]$ is not a row of $[C\ \mathbf{d}]$.*

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

Definition 2.2.2.23. The column transformations obtained by right multiplication of elementary matrices are called COLUMN OPERATIONS. For example, if $A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 0 & 3 & 2 \\ 3 & 4 & 5 & 3 \end{bmatrix}$ then

$AE_{23} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 2 & 3 & 0 & 2 \\ 3 & 5 & 4 & 3 \end{bmatrix}$ and $AE_{14}(-1) = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 3 & 0 \\ 3 & 4 & 5 & 0 \end{bmatrix}$.

Remark 2.2.2.24 (Rank of a Matrix)**.**

1. *The idea of row-rank was based on RREF and RREF was the result of systematically applying a finite number of elementary row operations. So, starting with a matrix A, we can systematically apply a finite number of elementary column operations (see Definition 2.2.2.23) to get a matrix which in some sense looks similar to RREF, call it say B, and then use the non-zero columns in that to define the* **column-rank**. *Note that B will have the following form:*
   - *(a) The zero columns appear after the non-zero columns.*
   - *(b) The first non-zero entry of a non-zero column is 1, called the pivot.*
   - *(c) The pivots in non-zero column move down in successive columns.*
2. *It will be proved later that row-rank(A) = column-rank(A). Thus, we just talk of the "***rank***", denoted* RANK(A).

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

Theorem 2.2.2.25. *Let A be a matrix of rank r. Then there exist a finite number of elementary matrices $E_1, \ldots, E_s$ and $F_1, \ldots, F_\ell$ such that*

$$E_1 \cdots E_s\, A\, F_1 \cdots F_\ell = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Proof. Let $C = \text{RREF}(A)$. As $\text{Rank}(A) = r$, by definition of RREF, there exist elementary matrices $E_1, \ldots, E_s$ such that $C = E_1 \cdots E_s A$. Note that C has r pivots and they appear in columns, say $i_1 < i_2 < \cdots < i_r$.

Now, let D be the matrix obtained from C by successively multiplying the elementary matrices E_{ji_j}, for $1 \le j \le r$, on the right of C. Then observe that $D = \begin{bmatrix} I_r & B \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where B is a matrix of an appropriate size.

As the $(1,1)$ block of D is an identity matrix, the block $(1,2)$ can be made the zero matrix by elementary column operations to D. Thus, the required result follows. ∎

Exercise **2.2.2.26.**

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 be a matrix of rank r. Then prove that there exist invertible matrices B_i, C_i such that $B_1 A = \begin{bmatrix} R_1 & R_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $AC_1 = \begin{bmatrix} S_1 & \mathbf{0} \\ S_3 & \mathbf{0} \end{bmatrix}$, $B_2AC_2 = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $B_3AC_3 = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where the $(1,1)$ block of each matrix is of size $r \times r$. Also, prove that A_1 is an invertible matrix.*

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

4. *Prove that if the product AB is defined and Rank(A) = Rank(AB) then $A = ABX$ for some matrix X. Similarly, if BA is defined and Rank(A) = Rank(BA) then $A = YBA$ for some matrix Y. [Hint: Choose invertible matrices P, Q satisfying $PAQ = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $P(AB) = (PAQ)(Q^{-1}B) = \begin{bmatrix} A_2 & A_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Now find an invertible matrix R such that $P(AB)R = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$. Now, define $X = R \begin{bmatrix} C^{-1}A_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Q^{-1}$ to get the required result.]*

5. *Prove that if AB is defined then Rank$(AB) \le$ Rank(A) and Rank$(AB) \le$ Rank(B).*

6. *Let P and Q be invertible matrices such that the matrix product PAQ is defined. Then prove that Rank(PAQ) = Rank(A).*

7. *Prove that if $A + B$ is defined then Rank$(A + B) \le$ Rank(A) + Rank(B).*

2.2.C Gauss-Jordan Elimination and System of Linear Equations

Let A be an $m \times n$ matrix. We now present an algorithm, commonly known as the Gauss-Jordan Elimination (GJE) method, 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 non-zero entry. Suppose this non-zero entry is a_{ij}. Box it. This is a pivot.
5. **Step 3:** Replace the row containing the pivot with the top row of the region. Also, make the pivot entry 1. 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 2.2.2.27. 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 $= \begin{bmatrix} 2 & 3 & 7 \\ 2 & 3 & 7 \\ 0 & 0 & 1 \end{bmatrix} \neq \mathbf{0}$. Let $C = \begin{bmatrix} 2 & 3 & 7 \\ 2 & 3 & 7 \\ 0 & 0 & 1 \end{bmatrix}$.
4. Then $E_1(\frac{1}{2})C = \begin{bmatrix} \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 2 & 3 & 7 \\ 0 & 0 & 1 \end{bmatrix}$ and $E_{21}(-2)E_1(\frac{1}{2})C = \begin{bmatrix} \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Or equivalently,

   $E_{32}(-2)E_2(\frac{1}{2})B = \begin{bmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.
5. Thus, $E_{34}E_{32}(-2)E_2(\frac{1}{2})E_{31}(-1)E_{12}A = \begin{bmatrix} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{1} & \frac{3}{2} & \frac{7}{2} \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Hence,

   $$E_{12}(-\frac{3}{2})E_{13}(-1)E_{23}(-\frac{7}{2})E_{34}E_{32}(-2)E_2(\frac{1}{2})E_{31}(-1)E_{12}A = \begin{bmatrix} \boxed{1} & 0 & -\frac{1}{2} & 0 \\ 0 & \boxed{1} & \frac{3}{2} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$
6. As the matrix A has been multiplied with elementary matrices on the left the RREF matrix $\begin{bmatrix} \boxed{1} & 0 & -\frac{1}{2} & 0 \\ 0 & \boxed{1} & \frac{3}{2} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is equivalent to A.

Remark 2.2.2.28 (Gauss Elimination (GE)). *GE is similar to the GJE except that*

1. *the pivots need not be made* 1 *and*
2. *the entries above the pivots need not be made* 0.

For example, if $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$ *then GE gives* $E_{32}(-3)E_{21}(-1)E_{31}(-1)A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$.

Thus, Gauss-Jordan Elimination may be viewed as an extension of the Gauss Elimination.

Example 2.2.2.29. Consider the system $A\mathbf{x} = \mathbf{b}$ with A a matrix of order 3×3 and $A[:,1] \neq \mathbf{0}$. If $[C \;\; \mathbf{d}] = \text{RREF}([A \;\; \mathbf{b}])$ then the possible choices for $[C \;\; \mathbf{d}]$ are given below:

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 and with UNIQUE SOLUTION $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$.

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.

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.

EXERCISE **2.2.2.30.**

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

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

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

3. *Find the solutions of the linear system of equations using Gauss-Jordan method.*

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

Now, using Proposition 2.2.2.7, Theorem 2.2.2.8 and the definition of RREF of a matrix, we obtain the following remark.

Remark 2.2.2.31. *Let RREF*$([A \;\; \mathbf{b}]) = [C \;\; \mathbf{d}]$. *Then*

1. *there exist elementary matrices, say $E_1, \ldots, E_k$, such that $E_1 \cdots E_k[A \;\; \mathbf{b}] = [C \;\; \mathbf{d}]$. Thus, the GJE (or the GE) is equivalent to multiplying by a finite number of elementary matrices on the left of $[A \;\; \mathbf{b}]$.*

2. *by Theorem 2.2.2.18 RREF$(A) = C$.*

Definition 2.2.2.32. [Basic, Free Variables] Consider the linear system $A\mathbf{x} = \mathbf{b}$. If RREF$([A \;\; \mathbf{b}]) = [C \;\; \mathbf{d}]$ then the unknowns

1. corresponding to the pivotal columns of C are called the **basic** variables.

2. that are not basic are called **free** variables.

Example 2.2.2.33. 1. Let RREF$([A \;\; \mathbf{b}]) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Then, $A\mathbf{x} = \mathbf{b}$ has

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

as it's solution set. Note that x and y are basic variables and z is the free variable.

2. Let RREF$([A \;\; \mathbf{b}]) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Then, the system $A\mathbf{x} = \mathbf{b}$ has no solution as RREF$([A \;\; \mathbf{b}])[3, :] = [0, 0, 0, 1]$ which corresponds to the equation $0 \cdot x + 0 \cdot y + 0 \cdot z = 1$.

3. Suppose the system $A\mathbf{x} = \mathbf{b}$ is consistent and RREF(A) has r non-zero rows. Then the system has r basic variables and $n - r$ free variables.

We now prove the main result in the theory of linear system (recall Remark 2.2.2.22).

Theorem 2.2.2.34. *Let A be an $m \times n$ matrix and let RREF$([A \;\; \mathbf{b}]) = [C \;\; \mathbf{d}]$, Rank$(A) = r$ and Rank$([A \, \mathbf{b}]) = r_a$. Then $A\mathbf{x} = \mathbf{b}$*

1. *is inconsistent if $r < r_a$*

2. *is consistent if $r = r_a$. Furthermore, $A\mathbf{x} = \mathbf{b}$ has*

   (a) A UNIQUE SOLUTION *if $r = n$.*

   (b) INFINITE NUMBER OF SOLUTIONS *if $r < n$. In this case, the solution set equals*

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

where $\mathbf{x}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r} \in \mathbb{R}^n$ with $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{u}_i = \mathbf{0}$, for $1 \le i \le n - r$.

Proof. PART 1: As $r < r_a$, by Remark 2.2.2.22 $([C \;\; \mathbf{d}])[r + 1, :] = [\mathbf{0}^T, \; 1]$. Then, 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 2.2.1.14, $A\mathbf{x} = \mathbf{b}$ is inconsistent.

PART 2: As $r = r_a$, by Remark 2.2.2.22, $[C \;\; \mathbf{d}]$ doesn't have a row of the form $[\mathbf{0}^T, \; 1]$ and there are r pivots in C. Suppose the pivots appear in columns $i_1, \ldots, i_r$ with $1 \le i_1 < \cdots < i_r \le n$. Thus, the unknowns 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 ℓ, $1 \le \ell \le r$, corresponds to the equation

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

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

$$\begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_r} \\ x_{t_1} \\ x_{t_2} \\ \vdots \\ x_{t_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 - \sum\limits_{k=1}^{n-r} c_{1t_k} x_{t_k} \\ \vdots \\ d_r - \sum\limits_{k=1}^{n-r} c_{rt_k} x_{t_k} \\ x_{t_1} \\ x_{t_2} \\ \vdots \\ x_{t_{n-r}} \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_{t_1} \begin{bmatrix} c_{1t_1} \\ \vdots \\ c_{rt_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_{t_2} \begin{bmatrix} c_{1t_2} \\ \vdots \\ c_{rt_2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_{t_{n-r}} \begin{bmatrix} c_{1t_{n-r}} \\ \vdots \\ c_{rt_{n-r}} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (2.2.2.4)$$

Thus, by Theorem 2.2.1.14 the system $A\mathbf{x} = \mathbf{b}$ is consistent. In case of Part 2a, $r = n$ and hence there are no free variables. Thus, the unique solution equals $x_i = d_i$, for $1 \le i \le n$.

In case of 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 $1 \le i \le n-r$, $A\mathbf{u}_i = \mathbf{0}$ and by Equation (2.2.2.4) 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. ∎

Let A be an $m \times n$ matrix. Then by Remark 2.2.2.20, $\text{Rank}(A) \le m$ and hence using Theorem 2.2.2.34 the next result follows.

Corollary 2.2.2.35. *Let A be a matrix of order $m \times n$ and consider $A\mathbf{x} = \mathbf{0}$. If*

1. *$Rank(A) = r < \min\{m, n\}$ then $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions.*
2. *$m < n$, then $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions.*

Thus, in either case, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has at least one non-trivial solution.

Remark 2.2.2.36. *Let A be an $m \times n$ matrix. Then Theorem 2.2.2.34 implies that*

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

2. *the vectors associated to the free variables in Equation (2.2.2.4) are solutions to the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$.

We end this subsection with some applications.

Example 2.2.2.37. 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 unknowns. Since this curve passes through the given points, we get a homogeneous system in 3 equations and 4 unknowns, 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, taking $d = 13$, the equation of the required circle is $3(x^2 + y^2) - 16y + 13 = 0$.

2. Determine the equation of the plane that contains the points $(1, 1, 1), (1, 3, 2)$ and $(2, -1, 2)$.

Solution: The general equation of a plane in space is given by $ax + by + cz + d = 0$, where a, b, c and d are unknowns. Since this plane passes through the 3 given points, we get a homogeneous system in 3 equations and 4 unknowns. 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, taking $d = 3$, the equation of the required plane is $-4x - y + 2z + 3 = 0$.

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

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

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

Solution of Part 3a: Solving for $A\mathbf{x} = 2\mathbf{x}$ is equivalent to solving $(A - 2I)\mathbf{x} = \mathbf{0}$ whose augmented matrix 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 non-zero solution.

Part 3b: As above, $A\mathbf{y} = 4\mathbf{y}$ is equivalent to solving $(A - 4I)\mathbf{y} = \mathbf{0}$ whose augmented matrix equals $\begin{bmatrix} -2 & 3 & 4 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & -3 & 0 & 0 \end{bmatrix}$. Now, verify that $\mathbf{y}^T = [2, 0, 1]$ is a non-zero solution.

EXERCISE **2.2.2.38.** *1. In the first part of this chapter* 3 *figures (see Figure* **??***) were given to illustrate different cases in Euclidean plane (*2 *equations in* 2 *unknowns). It is well known that in the case of Euclidean space (*3 *equations in* 3 *unknowns) there*

(a) is a way to place the 3 *planes so that the system has a unique solution.*

(b) are 4 *distinct ways to place the* 3 *planes so that the system has no solution.*

(c) are 3 distinct ways to place the 3 planes so that the system has an infinite number of solutions.

Determine the position of planes by drawing diagram to explain the above cases. Do these diagrams and the RREF matrices that appear in Example 2.2.2.29 have any relationship? Justify your answer.

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

3. *Solve the following linear systems.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

8. *Prove that* 5 *distinct points are needed to specify a general conic in Euclidean plane.*

9. *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 unknowns* c *and* d *is consistent.*

10. *Consider the linear system* $A\mathbf{x} = \mathbf{b}$ *in* m *equations and* 3 *unknowns. 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) $(1, 1, 1)^T$ *is the only solution.*
    (c) $\{(1, 1, 1)^T + c(1, 2, 1)^T | c \in \mathbb{R}\}$ *as the solution set.*
    (d) $\{c(1, 2, 1)^T | c \in \mathbb{R}\}$ *as the solution set.*
    (e) $\{(1, 1, 1)^T + c(1, 2, 1)^T + d(2, 2, -1)^T | c, d \in \mathbb{R}\}$ *as the solution set.*
    (f) $\{c(1, 2, 1)^T + d(2, 2, -1)^T | c, d \in \mathbb{R}\}$ *as the solution set.*

2.3 Square Matrices and System of Linear Equations

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

1. *A is invertible.*
2. *The homogeneous system* $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution.*
3. *Rank*$(A) = n$.
4. *The RREF of* A *is* I_n.
5. *A is a product of elementary matrices.*

Proof. $1 \Longrightarrow 2$ As A is invertible, A^{-1} exists and $A^{-1}A = I_n$. So, if $\mathbf{x}_0$ is any solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then

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

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

$2 \Longrightarrow 3$ Let if possible Rank$(A) = r < n$. Then, by Corollary 2.2.2.35, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has infinitely many solution. A contradiction. Thus, A has full rank.

$3 \Longrightarrow 4$ Suppose Rank$(A) = n$ and let $B = \text{RREF}(A)$. As $B = [b_{ij}]$ is a square matrix of order n, each column of B contains a pivot. Since the pivots move to the right as we go down the row, the i-th pivot must be at position b_{ii}, for $1 \leq i \leq n$. Thus, $B = I_n$ and hence, the RREF of A is I_n.

$4 \Longrightarrow 5$ Suppose RREF$(A) = I_n$. Then using Proposition 2.2.2.7, there exist elementary matrices $E_1, \ldots, E_k$ such that $E_1 \cdots E_k A = I_n$. Or equivalently,

$$A = (E_1 \cdots E_k)^{-1} = E_k^{-1} \cdots E_1^{-1} \tag{2.2.3.1}$$

which gives the desired result as by Remark 2.2.2.5 we know that the inverse of an elementary matrix is also an elementary matrix.

$5 \Longrightarrow 1$ Suppose $A = E_1 \cdots E_k$ for some elementary matrices $E_1, \ldots, E_k$. As the elementary matrices are invertible (see Remark 2.2.2.5) and the product of invertible matrices is also invertible, we get the required result. ■

As an immediate consequence of Theorem 2.2.3.1, we have the following important result which implies that one needs to compute either the left or the right inverse to prove invertibility.

Corollary 2.2.3.2. *Let A be a square matrix of order n. Suppose there exists a matrix*

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

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

Proof. Part 1: Let $CA = I_n$ for some matrix C and let $\mathbf{x}_0$ be a solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then $A\mathbf{x}_0 = \mathbf{0}$ and

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

Thus, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Hence, using Theorem 2.2.3.1, the matrix A is invertible.

Part 2: Using the first part, B is invertible. Hence, $B^{-1} = A$ or equivalently $A^{-1} = B$ and thus A is invertible as well. ■

Another important consequence of Theorem 2.2.3.1 is stated next which uses Equation (2.2.3.1) to get the required result. This result is used to compute the inverse of a matrix using the Gauss-Jordan Elimination.

Corollary 2.2.3.3. *Let A be an invertible matrix of order n. Suppose there exist elementary matrices $E_1, \ldots, E_k$ such that $E_1 \cdots E_k A = I_n$. Then $A^{-1} = E_1 \cdots E_k$.*

Remark 2.2.3.4. *Let A be an $n \times n$ matrix. Apply GJE to $[A \;\; I_n]$ and let RREF$([A \;\; I_n]) = [B \;\; C]$. If $B = I_n$, then $A^{-1} = C$ or else A is not invertible.*

Example 2.2.3.5. 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 **2.2.3.6.** 1. *Find the inverse of the following matrices using GJE.*

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

2. *Let A and B be two matrices having positive entries and of order 1×2 and 2×1, respectively. Which of BA or AB is invertible? Give reasons.*

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

   (a) $I - BA$ is invertible if and only if $I - AB$ is invertible [Use Theorem 2.2.3.1.2].

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

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

   (d) if A, B and $A + B$ are invertible then $(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B$.

4. *Let A be a square matrix. Then prove that*

   (a) A is invertible if and only if $A^T A$ is invertible.

   (b) A is invertible if and only if AA^T is invertible.

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

Theorem 2.2.3.7. *The following statements are equivalent for an $n \times n$ matrix A.*

1. *A is invertible.*

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

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

Proof. $1 \Longrightarrow 2$ 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, by Corollary 2.2.3.2, the matrix A is invertible. ■

We now give an immediate application of Theorem 2.2.3.7 and Theorem 2.2.3.1 without proof.

Theorem 2.2.3.8. *The following two statements cannot hold together for an $n \times n$ matrix A.*

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

EXERCISE **2.2.3.9.**

1. *Let A and B be square matrices of order n with $B = PA$, for an invertible matrix P. Then prove that A is invertible if and only if B is invertible.*
2. *Let A and B be two $m \times n$ matrices. Then prove that A and B are equivalent if and only if $B = PA$, where P is product of elementary matrices. When is this P unique?*
3. *Let $\mathbf{b}^T = [1, 2, -1, -2]$. Suppose A is a 4×4 matrix such that the linear system $A\mathbf{x} = \mathbf{b}$ has no solution. Mark each of the statements given below as* TRUE *or* FALSE*?*
   (a) *The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
   (b) *The matrix A is invertible.*
   (c) *Let $\mathbf{c}^T = [-1, -2, 1, 2]$. Then the system $A\mathbf{x} = \mathbf{c}$ has no solution.*
   (d) *Let $B = RREF(A)$. 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.*

2.3.A Determinant

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

Example 2.2.3.10. For $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 4 & 7 \end{bmatrix}$, $A(1 \mid 2) = \begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix}$ and $A(1, 2 \mid 1, 3) = [4]$.

With the notations as above, we are ready to give an inductive definition of the determinant of a square matrix. The advanced students can find the actual definition of the determinant in Appendix 7.7.1.22, where it is proved that the definition given below corresponds to the expansion of determinant along the first row.

Definition 2.2.3.11. [Determinant] Let A be a square matrix of order n. Then the determinant of A, denoted $\det(A)$ (or $|A|$) is defined by

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

Example 2.2.3.12. 1. Let $A = [-2]$. Then $\det(A) = |A| = -2$.

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

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

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

EXERCISE **2.2.3.13.** *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 2.2.3.14. [Singular, Non-Singular] 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 7.2.

Theorem 2.2.3.15. *Let A be an $n \times n$ matrix. If*

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

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

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

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

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

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

Example 2.2.3.16. 1. Since $\begin{vmatrix} 2 & 2 & 6 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{vmatrix} \xrightarrow{E_1(\frac{1}{2})} \begin{vmatrix} 1 & 1 & 3 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{vmatrix} \xrightarrow{E_{21}(-1), E_{31}(-1)} \begin{vmatrix} 1 & 1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{vmatrix}$, using Theorem 2.2.3.15, we see that, for $A = \begin{bmatrix} 2 & 2 & 6 \\ 1 & 3 & 2 \\ 1 & 1 & 2 \end{bmatrix}$, $\det(A) = 2 \cdot (1 \cdot 2 \cdot (-1)) = -4$, where the first 2 appears from the elementary matrix $E_1(\frac{1}{2})$.

2. For $A = \begin{bmatrix} 2 & 2 & 6 & 8 \\ 1 & 1 & 2 & 4 \\ 1 & 3 & 2 & 6 \\ 3 & 3 & 5 & 8 \end{bmatrix}$ verify that $|\, A \,| \xrightarrow{E_1(\frac{1}{2})} \begin{vmatrix} 1 & 1 & 3 & 4 \\ 1 & 1 & 2 & 4 \\ 1 & 3 & 2 & 6 \\ 3 & 3 & 5 & 8 \end{vmatrix}$. Now, a successive application of $E_{21}(-1), E_{31}(-1)$ and $E_{41}(-3)$ gives $\begin{vmatrix} 1 & 1 & 3 & 4 \\ 0 & 0 & -1 & 0 \\ 0 & 2 & -1 & 2 \\ 0 & 0 & -4 & -4 \end{vmatrix}$ and then applying E_{32} and $E_{43}(-4)$, we get $\begin{vmatrix} 1 & 1 & 3 & 4 \\ 0 & 2 & -1 & 2 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -4 \end{vmatrix}$. Thus, by Theorem 2.2.3.15 $\det(A) = 2 \cdot (-1) \cdot (8) = -16$ as 2 gets contributed due to $E_1(\frac{1}{2})$ and -1 due to E_{32}.

EXERCISE **2.2.3.17.** *Use Theorem 2.2.3.15 to arrive at the answer.*

1. *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. Prove that* $\det(B) = \alpha \det(A)$ *and* $\det(C) = \det(A)$.

2. *Prove that* 3 *divides* $\begin{vmatrix} 1 & 2 & 1 \\ 3 & 3 & 5 \\ 2 & 1 & 3 \end{vmatrix}$ *and* $\begin{vmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{vmatrix} = 0$.

By Theorem 2.2.3.15.6 $\det(I_n) = 1$. The next result about the determinant of the elementary matrices is an immediate consequence of Theorem 2.2.3.15 and hence the proof is omitted.

Corollary 2.2.3.18. *Fix a positive integer n. Then*

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

2. *For* $c \neq 0$, $\det(E_k(c)) = c$.

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

Remark 2.2.3.19. *Theorem 2.2.3.15.1 implies that the determinant can be calculated by expanding along any row. Hence, the readers are advised to verify that*

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

Example 2.2.3.20. Let us use Remark 2.2.3.19 to compute the determinant.

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

2. $\begin{vmatrix} 2 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{vmatrix} = (-1)^{2+2} \cdot \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 2 - 1 = 1.$

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

Remark 2.2.3.21. 1. *Let* $\mathbf{u}^T = [u_1, u_2], \mathbf{v}^T = [v_1, v_2] \in \mathbb{R}^2$. *Now, consider the parallelogram on vertices* $P = [0,0]^T, Q = \mathbf{u}, R = \mathbf{u}+\mathbf{v}$ *and* $S = \mathbf{v}$ *(see Figure 3). Then Area* $(PQRS) = \mid u_1v_2 - u_2v_1 \mid$, *the absolute value of* $\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$.

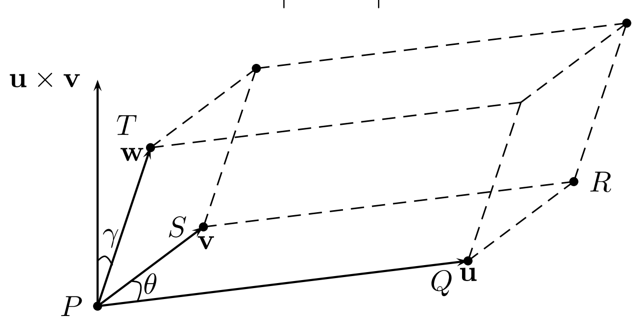


Figure 3: Parallelepiped with vertices P, Q, R *and* S *as base*

Recall that the dot product of $\mathbf{u}^T, \mathbf{v}^T$, *denoted* $\mathbf{u} \bullet \mathbf{v} = u_1v_1 + u_2v_2$, *the length of* $\mathbf{u}$, *denoted* $\ell(\mathbf{u}) = \sqrt{u_1^2 + u_2^2}$ *and* $\cos(\theta) = \frac{\mathbf{u} \bullet \mathbf{v}}{\ell(\mathbf{u})\ell(\mathbf{v})}$, *where* θ *is the angle between* $\mathbf{u}$ *and* $\mathbf{v}$. *Therefore*

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

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

2. *Consider Figure 3 again. Let* $\mathbf{u} = [u_1, u_2, u_3]^T, \mathbf{v} = [v_1, v_2, v_3]^T, \mathbf{w} = [w_1, w_2, w_3]^T \in \mathbb{R}^3$. *Then* $\mathbf{u} \bullet \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$ *and the cross product of* $\mathbf{u}$ *and* $\mathbf{v}$*, denoted*

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

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

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

3. *In general, for an* $n \times n$ *matrix* A, $\mid \det(A) \mid$ *satisfies all the properties associated with the volume of the* n*-dimensional parallelepiped. The actual proof is beyond the scope of this book. But, one can observe the following:*
   (a) *The volume of the* n*-dimensional unit cube is* $1 = \det(I_n)$.
   (b) *If one vector is multiplied by* $c \neq 0$ *then the volume either reduces or expands by* c*. The same holds for determinant.*
   (c) *Recall that if the height and base of a parallelogram is not changed then the area remains the same. This corresponds to replacing the perpendicular vector with the perpendicular vector plus* $c \neq 0$ *times the base vector for different choices of* c.

2.3.B Adjugate (classically Adjoint) of a Matrix

Definition 2.2.3.22. [Minor, Cofactor] Let A be an $n \times n$ matrix. Then the

1. $(i, j)^{\text{th}}$ **minor** of A, denoted $A_{ij} = \det(A(i \mid j))$, for $1 \leq i, j \leq n$.
2. $(i, j)^{\text{th}}$ **cofactor** of A, denoted $C_{ij} = (-1)^{i+j} A_{ij}$.
3. the **Adjugate** (classically Adjoint) of A, denoted $Adj(A) = [b_{ij}]$ with $b_{ij} = C_{ji}$, for $1 \leq i, j \leq n$. For example, for $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$, $A_{11} = \det(A(1 \mid 1)) = \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} = 4, A_{12} = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3, A_{13} = \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 1, \ldots, A_{32} = \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} = -5$ and $A_{33} = \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = -1$. So,
$$Adj(A) = \begin{bmatrix} (-1)^{1+1}A_{11} & (-1)^{2+1}A_{21} & (-1)^{3+1}A_{31} \\ (-1)^{1+2}A_{12} & (-1)^{2+2}A_{22} & (-1)^{3+2}A_{32} \\ (-1)^{1+3}A_{13} & (-1)^{2+3}A_{23} & (-1)^{3+3}A_{33} \end{bmatrix} = \begin{bmatrix} 4 & 2 & -7 \\ -3 & -1 & 5 \\ 1 & 0 & -1 \end{bmatrix}.$$

We now prove a very important result that relates adjugate matrix with the inverse.

Theorem 2.2.3.23. *Let* A *be an* $n \times n$ *matrix. Then*

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

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

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

$$\text{whenever } \det(A) \neq 0 \;\; \text{one has } \; A^{-1} = \frac{1}{\det(A)} Adj(A). \tag{2.2.3.2}$$

Proof. **Part 1:** It follows directly from Remark 2.2.3.19 and the definition of the cofactor.

Part 2: Fix positive integers i, ℓ with $1 \le i \neq \ell \le n$ 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 2.2.3.15.5, $\det(B) = 0$. As $A(\ell \mid j) = B(\ell \mid j)$, for $1 \le j \le n$, using Remark 2.2.3.19

$$\begin{aligned} 0 = \det(B) &= \sum_{j=1}^{n} (-1)^{\ell+j} b_{\ell j} \det\big(B(\ell \mid j)\big) = \sum_{j=1}^{n} (-1)^{\ell+j} a_{ij} \det\big(B(\ell \mid j)\big) \\ &= \sum_{j=1}^{n} (-1)^{\ell+j} a_{ij} \det\big(A(\ell \mid j)\big) = \sum_{j=1}^{n} a_{ij} C_{\ell j}. \end{aligned} \tag{2.2.3.3}$$

This completes the proof of Part 2.

Part 3:, Using Equation (2.2.3.3) and Remark 2.2.3.19, 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 Corollary 2.2.3.2, $A^{-1} = \frac{1}{\det(A)} Adj(A)$. ∎

Example 2.2.3.24. 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 2.2.3.23.3, $A^{-1} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ -1/2 & -1/2 & 1/2 \\ 1/2 & 3/2 & -1/2 \end{bmatrix}$.

Let A be a non-singular matrix. Then, by Theorem 2.2.3.23.3, $A^{-1} = \frac{1}{\det(A)} Adj(A)$. 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 2.2.3.25. *Let A be a non-singular matrix. Then*

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

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

Theorem 2.2.3.26. *A square matrix A is non-singular if and only if A is invertible.*

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

Now, let us assume that A is invertible. Then, using Theorem 2.2.3.1, $A = E_1 \cdots E_k$ a product of elementary matrices. Also, by Corollary 2.2.3.18, $\det(E_i) \neq 0$ for $1 \leq i \leq k$. Thus, a repeated application of Parts $1, 2$ and 3 of Theorem 2.2.3.15 gives $\det(A) \neq 0$. ∎

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

Theorem 2.2.3.27. *Let A and B be square matrices of order n. Then*

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

Proof. **Step 1.** Let A be non-singular. Then, by Theorem 2.2.3.23.3, A is invertible and by Theorem 2.2.3.1, $A = E_1 \cdots E_k$, a product of elementary matrices. Then a repeated application of Parts $1, 2$ and 3 of Theorem 2.2.3.15 gives the desired result as

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

Step 2. Let A be singular. Then, by Theorem 2.2.3.26 A is not invertible. So, by Theorem 2.2.3.1 and Exercise 2.2.2.26.2 there exists an invertible matrix P such that $PA = C$, where $C = \begin{bmatrix} C_1 \\ \mathbf{0} \end{bmatrix}$. 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. ∎

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

Proof. If A is a non-singular, Corollary 2.2.3.25 gives $\det(A) = \det(A^T)$.

If A is singular then, by Theorem 2.2.3.26, A is not invertible. So, A^T is also not invertible and hence by Theorem 2.2.3.26, $\det(A^T) = 0 = \det(A)$. ∎

Example 2.2.3.29. Let A be an orthogonal matrix then, by definition, $AA^T = I$. Thus, by Theorems 2.2.3.27 and 2.2.3.28

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

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 $\theta \in [0,\ 2\pi]$ 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 across the line $y = m\mathbf{x}$. Determine m? (see Exercise 3.3b).

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 3.3a).

EXERCISE **2.2.3.30.** *1. Let A be an $n \times n$ upper triangular matrix with non-zero entries on the diagonal. Then prove that A^{-1} is also an upper triangular matrix.*

2. Let A be an $n \times n$ matrix. Then $\det(A) = 0$ *if either*

(a) Prove that $\begin{vmatrix} a & b & c \\ e & f & g \\ h & j & \ell \end{vmatrix} = \begin{vmatrix} a & e & \alpha a + \beta e + h \\ b & f & \alpha b + \beta f + j \\ c & g & \alpha c + \beta g + \ell \end{vmatrix}$ *for some* $\alpha, \beta \in \mathbb{C}$.

(b) Prove that 17 *divides* $\begin{vmatrix} 3 & 1 & 1 \\ 4 & 8 & 1 \\ 0 & 7 & 9 \end{vmatrix}$.

(c) $A[i,:]^T = \mathbf{0}$ *or* $A[:,i] = \mathbf{0}$, *for some* $i, 1 \le i \le n$.

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

2.3.C Cramer's Rule

Let A be a square matrix. Then combining Theorem 2.2.3.7 and Theorem 2.2.3.26, one has the following result.

Corollary 2.2.3.31. *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 2.2.3.32 (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*

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

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 $P[A \mid \mathbf{b}] = [I \mid P\mathbf{b}]$. Let $\mathbf{d} = A\mathbf{b}$. 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 2.2.3.33. 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

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

2.4 Miscellaneous Exercises

EXERCISE **2.2.4.1.** *1. 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 $B = \begin{bmatrix} A_{11}^{-1} + (A_{11}^{-1}A_{12})P^{-1}(A_{21}A_{11}^{-1}) & -(A_{11}^{-1}A_{12})P^{-1} \\ -P^{-1}(A_{21}A_{11}^{-1}) & P^{-1} \end{bmatrix}$.

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

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

4. Let A be a 2×2 matrix with $\mathrm{TR}(A) = 0$ and $\det(A) = 0$. Then A is a nilpotent matrix.

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

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

(a) A is not invertible.

(b) $\mathrm{RANK}(A) \neq n$.

(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.*

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

8. *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.*

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

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

11. *Let $p \in \mathbb{C}, p \neq 0$. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $n \times n$ matrices with $b_{ij} = p^{i-j} a_{ij}$, for $1 \le i, j \le n$. Then compute* $\det(B)$ *in terms of* $\det(A)$.

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

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

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

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

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

2.5 Summary

In this chapter, we started with a system of m linear equations in n unknowns 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 $[A \mid \mathbf{b}]$ and thus giving as the RREF which in turn gave us rank of a matrix. If $\text{Rank}(A) = r$ and $\text{Rank}([A \mid \mathbf{b}]) = r_a$ and

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

We have also see that the following conditions are equivalent for an $n \times n$ matrix A.

1. A is invertible.
2. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
3. The row reduced echelon form of A is I.
4. A is a product of elementary matrices.
5. The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$.
6. The system $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.
7. $\text{rank}(A) = n$.
8. $\det(A) \neq 0$.

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, if $A\mathbf{x} = \mathbf{0}$ has
   (a) a unique solution then the columns of A are linear independent.
   (b) else, 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 $\mathbb{C}$. The last section will consist of results in infinite dimensional vector spaces that are similar but different as compared with he finite dimensional case. We have given lots of examples of vector spaces that are infinite dimensional or are vector spaces over fields that are different from $\mathbb{R}$ and $\mathbb{C}$. See appendix to have some ideas about fields that are different from $\mathbb{R}$ and $\mathbb{C}$.

3.1 Vector Spaces: Definition and Examples

In this chapter, $\mathbb{F}$ denotes either $\mathbb{R}$, the set of real numbers or $\mathbb{C}$, the set of complex numbers.

Let A be an $m \times n$ complex matrix and let $\mathbb{V}$ denote the solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$. Then, by Theorem 2.2.1.6, $\mathbb{V}$ 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 $\alpha \in \mathbb{C}$. In particular, for $\alpha = -1$, $-\mathbf{x} \in \mathbb{V}$.

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

4. if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{V}$ then, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.

That is, the solution set of a homogeneous linear system satisfies some nice properties. The Euclidean plane, $\mathbb{R}^2$ and the Euclidean space, $\mathbb{R}^3$, also satisfy the above properties. In this chapter, our aim is to understand sets that satisfy such properties. We start with the following definition.

Definition 3.3.1.1. [Vector Space] 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 axioms:

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 a unique element $-\mathbf{u} \in \mathbb{V}$ that satisfies $\mathbf{u} \oplus (-\mathbf{u}) = \mathbf{0}$ (called **the additive inverse**).

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) $\alpha \cdot (\beta \odot \mathbf{u}) = (\alpha\beta) \odot \mathbf{u}$ 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 $\mathbb{F}$).

Definition 3.3.1.2. 1. The number $0 \in \mathbb{F}$, whereas $\mathbf{0} \in \mathbb{V}$.

2. The elements of $\mathbb{F}$ are called **scalars**.
3. The elements of $\mathbb{V}$ are called **vectors**.
4. If $\mathbb{F} = \mathbb{R}$ then $\mathbb{V}$ is called a **real vector space**.
5. If $\mathbb{F} = \mathbb{C}$ then $\mathbb{V}$ is called a **complex vector space**.
6. In general, a vector space over $\mathbb{R}$ or $\mathbb{C}$ is called a **linear space**.

Some interesting consequences of Definition 3.3.1.1 is stated next. Intuitively, these results seem obvious but for better understanding of the axioms it is desirable to go through the proof.

Theorem 3.3.1.3. *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 Axiom 3.3.1.1.1d, 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 Axiom 3.3.1.1.3, 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 Axiom 3.3.1.1.3b,

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

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 Axiom 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 3.3.1.4. The readers are advised to justify the statements given below.

1. Let A be an $m \times n$ matrix with complex entries with $\text{Rank}(A) = r \leq n$. Then, using Theorem 2.2.2.34, $\mathbb{V} = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$ is a vector space.

2. Consider $\mathbb{R}$ with the usual addition and multiplication. That is, $\oplus \equiv +$ and $\odot \equiv \cdot$. Then, $\mathbb{R}$ forms a real vector space.

3. 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.

4. Let $\mathbb{R}^n = \{(a_1, \ldots, a_n)^T \mid a_i \in \mathbb{R}, 1 \leq i \leq n\}$. 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 $\sqrt{-1}$.

5. 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 $\alpha \in \mathbb{R}$. Then $\alpha \odot \mathbf{z}_1 = (\alpha x_1) + i(\alpha y_1)$ and we call $\mathbb{C}$ the real vector space.

   (b) let $\alpha + i\beta \in \mathbb{C}$. Then $(\alpha + i\beta) \odot (x_1 + iy_1) = (\alpha x_1 - \beta y_1) + i(\alpha y_1 + \beta x_1)$ and we call $\mathbb{C}$ the complex vector space.

6. Let $\mathbb{C}^n = \{(z_1, \ldots, z_n)^T \mid z_i \in \mathbb{C}, 1 \leq i \leq n\}$. 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 complex vector space) as well as over $\mathbb{R}$ (called real vector space). In general, we assume $\mathbb{C}^n$ to be a complex vector space.

Remark 3.3.1.5. *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 $i \notin \mathbb{R}$.*

7. 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 $\mathbb{M}_n(\mathbb{C})$.
8. Let S be a non-empty set and let $\mathbb{R}^S = \{f \mid f$ is a function from S 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,
   (a) for $S = \mathbb{N}$, observe that $\mathbb{R}^{\mathbb{N}}$, consisting of all real sequences, forms a real vector space.
   (b) Let $\mathbb{V}$ be the set of all bounded real sequences. Then $\mathbb{V}$ is a real vector space.
   (c) Let $\mathbb{V}$ be the set of all real sequences that converge to 0. Then $\mathbb{V}$ is a real vector space.
   (d) Let S be the set of all real sequences that converge to 1. Then check that S is not a vector space. Determine the conditions that fail.
9. Fix $a, b \in \mathbb{R}$ with $a < b$ and let $\mathcal{C}([a,b], \mathbb{R}) = \{f : [a,b] \to \mathbb{R} \mid f$ 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.
10. Let $\mathcal{C}(\mathbb{R}, \mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f$ is continuous$\}$. Then $\mathcal{C}(\mathbb{R}, \mathbb{R})$ with $(f + \alpha g)(x) = f(x) + \alpha g(x)$, for all $x \in \mathbb{R}$, is a real vector space.
11. Fix $a < b \in \mathbb{R}$ and let $\mathcal{C}^2((a,b), \mathbb{R}) = \{f : (a,b) \to \mathbb{R} \mid f''$exists and f'' 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.
12. Fix $a < b \in \mathbb{R}$ and let $\mathcal{C}^\infty((a,b), \mathbb{R}) = \{f : (a,b) \to \mathbb{R} \mid f$ 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.
13. 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.
14. Let $\mathbb{R}[x] = \{p(x) \mid p(x)$ is a real polynomial in $x\}$. Then, with the usual addition of polynomials and $\alpha(a_0 + a_1x + \cdots + a_nx^n) = (\alpha a_0) + \cdots + (\alpha a_n)x^n$, for $\alpha \in \mathbb{R}$, gives $\mathbb{R}[x]$ a real vector space structure.
15. Fix $n \in \mathbb{N}$ and let $\mathbb{R}[x;n] = \{p(x) \in \mathbb{R}[x] \mid p(x)$ has degree $\leq n\}$.Then, with the usual addition of polynomials and $\alpha(a_0 + a_1x + \cdots + a_nx^n) = (\alpha a_0) + \cdots + (\alpha a_n)x^n$, for $\alpha \in \mathbb{R}$, gives $\mathbb{R}[x;n]$ a real vector space structure.
16. Let $\mathbb{C}[x] = \{p(x) \mid p(x)$ is a complex polynomial in $x\}$. Then, with the usual addition of polynomials and $\alpha(a_0 + a_1x + \cdots + a_nx^n) = (\alpha a_0) + \cdots + (\alpha a_n)x^n$, for $\alpha \in \mathbb{C}$, gives $\mathbb{C}[x]$ a real vector space structure.
17. Let $\mathbb{V} = \{\mathbf{0}\}$. Then $\mathbb{V}$ is a real as well as a complex vector space.
18. 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

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

19. 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.

20. Let $\mathbb{V} = \{A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C}) \mid a_{11} = 0\}$. Then $\mathbb{V}$ is a complex vector space.

21. Let $\mathbb{V} = \{A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C}) \mid A = A^*\}$. Then $\mathbb{V}$ is a real vector space but not a complex vector space.

22. 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.

23. Let $\mathbb{Q}$ be the set of scalars. Then $\mathbb{R}$ is a vector space over $\mathbb{Q}$. As $e, \pi - \sqrt{2} \notin \mathbb{Q}$, these real numbers are vectors but not scalars in this space.

24. Similarly, $\mathbb{C}$ is a vector space over $\mathbb{Q}$. Since $e - \pi, i + \sqrt{2}, i \notin \mathbb{Q}$, these complex numbers are vectors but not scalars in this space.

25. Let $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ with addition and multiplication, respectively, given by

+	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, $\mathbb{V} = \{(a, b) \mid a, b \in \mathbb{Z}_5\}$ is a vector space having 25 vectors.

Note that all our vector spaces, except the last three, 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 **3.3.1.6.** *1. Verify the axioms for vector spaces that appear in Example 3.3.1.4.*

2. *Does the set $\mathbb{V}$ given below form a real/complex or both real and complex vector space? Give reasons for your answer.*
   (a) *For* $\mathbf{x} = (x_1, x_2)^T, \mathbf{y} = (y_1, y_2)^T \in \mathbb{R}^2$, *define* $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2)^T$ *and* $\alpha\mathbf{x} = (\alpha x_1, 0)^T$ *for all* $\alpha \in \mathbb{R}$.
   (b) *Let* $\mathbb{V} = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a, b, c, d \in \mathbb{C}, a + c = 0 \right\}$.
   (c) *Let* $\mathbb{V} = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a = \overline{b},\ a, b, c, d \in \mathbb{C} \right\}$.
   (d) *Let* $\mathbb{V} = \{(x, y, z)^T \mid x + y + z = 1\}$.
   (e) *Let* $\mathbb{V} = \{(x, y)^T \in \mathbb{R}^2 \mid x \cdot y = 0\}$.
   (f) *Let* $\mathbb{V} = \{(x, y)^T \in \mathbb{R}^2 \mid x = y^2\}$.
   (g) *Let* $\mathbb{V} = \{\alpha(1, 1, 1)^T + \beta(1, 1, -1)^T \mid \alpha, \beta \in \mathbb{R}\}$.
   (h) *Let* $\mathbb{V} = \mathbb{R}$ *with* $x \oplus y = x - y$ *and* $\alpha \odot x = -\alpha x$, *for all* $x, y \in \mathbb{V}$ *and* $\alpha \in \mathbb{R}$.
   (i) *Let* $\mathbb{V} = \mathbb{R}^2$. *Define* $(x_1, y_1)^T \oplus (x_2, y_2)^T = (x_1 + x_2, 0)^T$ *and* $\alpha \odot (x_1, y_1)^T = (\alpha x_1, 0)^T$, *for* $\alpha, x_1, x_2, y_1, y_2 \in \mathbb{R}$.

3.1.A Subspaces

Definition 3.3.1.7. [Vector Subspace] 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 3.3.1.8. 1. If $\mathbb{V}$ is a vector space then $\mathbb{V}$ and $\{\mathbf{0}\}$ are subspaces, called **trivial subspaces**.

2. The real vector space $\mathbb{R}$ has no non-trivial subspace.

3. $\mathbb{W} = \{\mathbf{x} \in \mathbb{R}^3 \mid [1, 2, -1]\mathbf{x} = 0\}$ is a plane in $\mathbb{R}^3$ containing $\mathbf{0}$ (so a subspace).

4. $\mathbb{W} = \left\{\mathbf{x} \in \mathbb{R}^3 \mid \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix}\mathbf{x} = \mathbf{0}\right\}$ is a line in $\mathbb{R}^3$ containing $\mathbf{0}$ (so a subspace).

5. The vector space $\mathbb{R}[x; n]$ is a subspace of $\mathbb{R}[x]$.

6. Prove that $\mathcal{C}^2(a, b)$ is a subspace of $\mathcal{C}(a, b)$.

7. Prove that $\mathbb{W} = \{(x, 0)^T \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

8. Is the set of sequences converging to 0 a subspace of the set of all bounded sequences?

9. Let $\mathbb{V}$ be the vector space of Example 3.3.1.4.19. 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) $\mathbb{W} = \{(x, 3)^T \mid x \in \mathbb{R}\}$ is a subspace of $\mathbb{V}$.

10. The vector space $\mathbb{R}^+$ defined in Example 3.3.1.4.18 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 3.3.1.9. *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 $\mathbf{u}, \mathbf{v} \in \mathbb{W}$.*

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 3.3.1.3.3, $-\mathbf{u} = (-1)\mathbf{u} \in \mathbb{W}$ as well.

4. The commutative and associative laws of vector addition hold as they hold in $\mathbb{V}$.

5. The axioms related with scalar multiplication and the distributive laws also hold as they hold in $\mathbb{V}$.

Thus, one obtains the required result. ■

EXERCISE **3.3.1.10.** *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 $(0,0) \in \mathbb{R}^2$.

3. 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) $\mathbb{W} = \{f \in C([-1,1]) \mid f(-1/2) = 0, f(1/2) = 0\}$.
   (c) $\mathbb{W} = \{f \in C([-1,1]) \mid f'(\frac{1}{4})$ *exists* $\}$.

4. Are all the sets given below subspaces of $\mathbb{R}[x]$?

   (a) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid \deg(f(x)) = 3\}$.
   (b) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid \deg(f(x)) = 0\}$.
   (c) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid f(1) = 0\}$.
   (d) $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid f(0) = 0, f(1/2) = 0\}$.

5. Which of the following are subspaces of $\mathbb{R}^n(\mathbb{R})$?

   (a) $\{(x_1, x_2, \ldots, x_n)^T \mid x_1 \geq 0\}$.
   (b) $\{(x_1, x_2, \ldots, x_n)^T \mid x_1$ *is rational*$\}$.
   (c) $\{(x_1, x_2, \ldots, x_n)^T \mid \; | x_1 | \leq 1\}$.

6. *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) $\{(z_1, z_2, \ldots, z_n)^T \mid \mid z_1 \mid = \mid z_2 \mid\}$.

7. *Fix $n \in \mathbb{N}$. Then, is 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) $\mathbb{W} = \{A \mid A \text{ is a diagonal matrix}\}$?

   (e) $\mathbb{W} = \{A \mid trace(A) = 0\}$?

   (f) $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A^T = 2A\}$?

   (g) $\mathbb{W} = \{A = [a_{ij}] \mid a_{11} + a_{21} + a_{34} = 0\}$?

8. *Fix $n \in \mathbb{N}$. Then, is $\mathbb{W} = \{A = [a_{ij}] \mid a_{11} + \overline{a_{22}} = 0\}$ a subspace of the complex vector space $M_n(\mathbb{C})$? What if $M_n(\mathbb{C})$ is a real vector space?*

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) $G = \{A \in M_n(\mathbb{R}) \mid \det(A) \neq 0\}$.

   (c) $G = \{A \in M_n(\mathbb{R}) \mid \det(A) = 1\}$.

3.1.B Linear Span

Definition 3.3.1.11. [Linear Combination] Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{V}$. Then, a vector $\mathbf{u} \in \mathbb{V}$ is said to be a **linear combination** of $\mathbf{u}_1, \ldots, \mathbf{u}_n$ if we can find scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n$.

Example 3.3.1.12. 1. $(3, 4, 3) = 2(1, 1, 1) + (1, 2, 1)$ but we cannot find $a, b \in \mathbb{R}$ such that $(3, 4, 5) = a(1, 1, 1) + b(1, 2, 1)$.

2. 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

$$a(1, 0, 0) + b(2, 1, 0) + c(3, 3, 1) = (4, 5, 5) \qquad (3.3.1.1)$$

   in the unknowns $a, b, c \in \mathbb{R}$ has a solution. Clearly, Equation (3.3.1.1) has solution $a = 9, b = -10$ and $c = 5$.

3. Is $(4,5,5)$ a linear combination of the vectors $(1,2,3), (-1,1,4)$ and $(3,3,2)$?
   Solution: The vector $(4,5,5)$ is a linear combination if the linear system
   $$a(1,2,3) + b(-1,1,4) + c(3,3,2) = (4,5,5) \tag{3.3.1.2}$$
   in the unknowns $a, b, c \in \mathbb{R}$ has a solution. The RREF of the corresponding augmented matrix equals $\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, implying infinite number of solutions. For example, $(4,5,5) = 3(1,2,3) - (-1,1,4)$.

4. Is $(4,5,5)$ a linear combination of the vectors $(1,2,1), (1,0,-1)$ and $(1,1,0)$?
   Solution: The vector $(4,5,5)$ is a linear combination if the linear system
   $$a(1,2,1) + b(1,0,-1) + c(1,1,0) = (4,5,5) \tag{3.3.1.3}$$
   in the unknowns $a, b, c \in \mathbb{R}$ has a solution. The RREF of the corresponding augmented matrix equals $\begin{bmatrix} 1 & 0 & 1/2 & 5/2 \\ 0 & 1 & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$. So, Equation (3.3.1.3) has no solution. Thus, $(4,5,5)$ is not a linear combination of the given vectors.

EXERCISE **3.3.1.13.** *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 (x,y,z) is a linear combination of
   (a) $(1,2,3), (-1,1,4)$ and $(3,3,2)$.
   (b) $(1,2,1), (1,0,-1)$ and $(1,1,0)$.
   (c) $(1,1,1), (1,1,0)$ and $(1,-1,0)$.

Definition 3.3.1.14. [Linear Span] Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{V}$. Then, the **linear span** of S, denoted $LS(S)$, equals
$$LS(S) \;=\; \{\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n \mid \alpha_i \in \mathbb{F}, 1 \le i \le n\}.$$
If S is an empty set, we define $LS(S) = \{\mathbf{0}\}$.

Example 3.3.1.15. 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 is the geometrical representation of $LS(S)$?
   Solution: $LS(S) = \{a(1,1,1)^T + b(2,1,3)^T \mid a, b \in \mathbb{R}\} = \{(a+2b, a+b, a+3b)^T \mid a, b \in \mathbb{R}\}$. For geometrical representation, we need to find conditions on x, y and z such that

$(a+2b, a+b, a+3b) = (x,y,z)$. Or equivalently, $a+2b = x, a+b = y, a+3b = z$ has a solution for all $a, b \in \mathbb{R}$. Check that 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, we need $2x-y-z=0$. Hence, $LS(S)$ is a plane given by

$$LS(S) = \{a(1,1,1)^T + b(2,1,3)^T \mid a, b \in \mathbb{R}\} = \{(x,y,z)^T \in \mathbb{R}^3 \mid 2x-y-z=0\}.$$

3. $S = \{(1,2,1)^T, (1,0,-1)^T, (1,1,0)^T\}$. What is the geometrical representation of $LS(S)$?
Solution: As above, we need to find condition(s) on x, y, z such that the linear system

$$a(1,2,1) + b(1,0,-1) + c(1,1,0) = (x,y,z) \tag{3.3.1.4}$$

in the unknowns a, b, c is always consistent. An application of GJE to Equation (3.3.1.4) gives $\begin{bmatrix} 1 & 0 & 1 & \frac{x+y}{3} \\ 0 & 1 & \frac{1}{2} & \frac{2x-y}{3} \\ 0 & 0 & 0 & x-y+z \end{bmatrix}$. Thus,

$$LS(S) = \{(x,y,z)^T \in \mathbb{R}^3 \mid x-y+z=0\}.$$

4. $S = \{(1,2,3)^T, (-1,1,4)^T, (3,3,2)^T\}$.
Solution: As above, need to find condition(s) on x, y, z such that the linear system

$$a(1,2,3) + b(-1,1,4) + c(3,3,2) = (x,y,z)$$

in the unknowns a, b, c is always consistent. An application of GJE method gives $5x-7y+3z=0$ as the required condition. Thus,

$$LS(S) = \{(x,y,z)^T \in \mathbb{R}^3 \mid \ 5x-7y+3z=0\}.$$

5. $S = \{(1,2,3,4)^T, (-1,1,4,5)^T, (3,3,2,3)^T\} \subseteq \mathbb{R}^4$.
Solution: The readers are advised to show that

$$LS(S) = \{(x,y,z,w)^T \in \mathbb{R}^4 \mid \ 2x-3y+w=0, \ 5x-7y+3z=0\}.$$

EXERCISE **3.3.1.16.** *For each S, determine the geometric representation of $LS(S)$.*

1. $S = \{-1\} \subseteq \mathbb{R}$.
2. $S = \{\pi\} \subseteq \mathbb{R}$.
3. $S = \{(1,0,1)^T, (0,1,0)^T, (3,0,3)^T\} \subseteq \mathbb{R}^3$.
4. $S = \{(1,2,1)^T, (2,0,1)^T, (1,1,1)^T\} \subseteq \mathbb{R}^3$.
5. $S = \{(1,0,1,1)^T, (0,1,0,1)^T, (3,0,3,1)^T\} \subseteq \mathbb{R}^4$.

Definition 3.3.1.17. [Finite Dimensional Vector Space] 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 3.3.1.18.

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, $\mathbb{R}[x;n]$ is finite dimensional as $\mathbb{R}[x;n] = LS(\{1, x, x^2, \ldots, x^n\})$.
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.
6. The vector space $\mathbb{C}$ over $\mathbb{Q}$ is infinite dimensional.

Lemma 3.3.1.19 (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 3.3.1.9, $LS(S)$ is a vector subspace. ∎

Remark 3.3.1.20. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. If W is a subspace of $\mathbb{V}$ and $S \subseteq W$ then $LS(S)$ is a subspace of W as well.*

Theorem 3.3.1.21. *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 Remark 3.3.1.20, $LS(S) \subseteq \mathbb{W}$ and hence the result follows. ∎

Definition 3.3.1.22. 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 $\mathbb{V}$. Then, we define their **sum**, denoted $P + Q$, as $P + Q = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in P, \mathbf{v} \in Q\}$. For example, $P + Q = \mathbf{R}^2$, 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 $(x,y) = \dfrac{2y-x}{3}(1,2) + \dfrac{2x-y}{3}(2,1)$.

We leave the proof of the next result for readers.

Lemma 3.3.1.23. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let P and Q be two subspaces of $\mathbb{V}$. Then $P+Q$ is the smallest subspace of $\mathbb{V}$ containing both P and Q.*

EXERCISE **3.3.1.24.**

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. *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 $d = 0$.*

4. *Let $\mathbb{W} = \{f(x) \in \mathbb{R}[x] \mid \deg(f(x)) = 5\}$. Prove that $\mathbb{W}$ is not a subspace $\mathbb{R}[x]$.*

5. *Determine all subspaces of the vector space in Example 3.3.1.4.19.*

6. *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 $\mathbb{U}+\mathbb{W}$?*

7. *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 $\mathbb{W}\cup\mathbb{W}$ a subspace of $\mathbb{V}$?*

   (d) *Prove that $LS(\mathbb{W}\cup\mathbb{U}) = \mathbb{W}+\mathbb{U}$.*

8. *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 $LS(S) = LS(S \setminus \{\mathbf{x}_i\})$.*

9. *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 $\mathbf{u} \in \mathbb{R}^3$, determine $\mathbf{u}_W \in \mathbb{W}$ and $\mathbf{u}_U \in \mathbb{U}$ such that $\mathbf{u} = \mathbf{u}_W + \mathbf{u}_U$. Is it necessary that $\mathbf{u}_W$ and $\mathbf{u}_U$ are unique?*

10. *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 $\mathbf{u} \in \mathbb{R}^3$ such that when we write $\mathbf{u} = \mathbf{u}_W + \mathbf{u}_U$, with $\mathbf{u}_W \in \mathbb{W}$ and $\mathbf{u}_U \in \mathbb{U}$, the vectors $\mathbf{u}_W$ and $\mathbf{u}_U$ are not unique.*

3.1.C Fundamental Subspaces Associated with a Matrix

Definition 3.3.1.25. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. Then, we use the functions $P : \mathbb{C}^n \to \mathbb{C}^m$ and $Q : \mathbb{C}^m \to \mathbb{C}^n$ defined by $P(\mathbf{x}) = A\mathbf{x}$ and $Q(\mathbf{y}) = A^*\mathbf{y}$, respectively, for $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$, to get the four fundamental subspaces associated with A, namely,

1. $\text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n\} = \text{Rng}(P)$ is a subspace of $\mathbb{C}^m$, called the **Column / Range space**. Observe that $\text{Col}(A)$ is the linear span of columns of A.
2. $\text{Null}(A) = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{C}^n\} = \text{Null}(P)$ is a subspace of $\mathbb{C}^n$, called the **Null space**.
3. $\text{Col}(A^*) = \{A^*\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n\} = \text{Rng}(Q)$ is the linear span of rows of $\overline{A} = [\overline{a_{ij}}]$. If $A \in \mathbb{M}_{m,n}(\mathbb{R})$ then $\text{Col}(A^*)$ reduces to $\text{Row}(A) = \{\mathbf{x}^T A \mid \mathbf{x} \in \mathbb{R}^m\}$, the **row space** of A.
4. $\text{Null}(A^*) = \{\mathbf{x} \mid A^*\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{C}^n\} = \text{Null}(Q)$. If $A \in \mathbb{M}_{m,n}(\mathbb{R})$ then $\text{Null}(A^*)$ reduces to $\text{Null}(A^T) = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{x}^T A = \mathbf{0}^T\}$.

Example 3.3.1.26. 1. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$. Then

(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) $\text{Row}(A) = \{\mathbf{x} = (x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1 + x_2 - 2x_3 = 0\}$.

(c) $\text{Null}(A) = LS((1, 1, -2)^T)$.

(d) $\text{Null}(A^T) = LS((1, -1, 1)^T)$.

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

(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 $A = \begin{bmatrix} 1 & i & 2i \\ i & -2 & -3 \\ 1 & 1 & 1+i \end{bmatrix}$. Then

(a) $\text{Col}(A) = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid (2+i)x_1 - (1-i)x_2 - x_3 = 0\}$.

(b) $\text{Col}(A^*) = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid ix_1 - x_2 + x_3 = 0\}$.

(c) $\text{Null}(A) = LS((i, 1, -1)^T)$.

(d) $\text{Null}(A^*) = LS((-2+i, 1+i, 1)^T)$.

Remark 3.3.1.27. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$. Then, in Example 3.3.1.26, observe that the direction ratios of* $\text{Col}(A)$ *matches vector(s) in* $\text{Null}(A^T)$. *Similarly, the direction ratios of* $\text{Row}(A)$ *matches with vectors in* $\text{Null}(A)$. *What are the relationships in case $A \in \mathbb{M}_{m,n}(\mathbb{C})$? We will come back to these spaces again and again.*

Let $\mathbb{V}$ be a vector space over either $\mathbb{R}$ or $\mathbb{C}$. Then, we have learnt that

1. for any $S \subseteq \mathbb{V}$, $LS(S)$ is again a vector space. Moreover, $LS(S)$ is the smallest subspace containing S.

2. unless $S = \emptyset$, $LS(S)$ has infinite number of vectors.

   Therefore, the following questions arise:

   (a) Are there conditions under which $LS(S_1) = LS(S_2)$ for $S_1 \neq S_2$?

   (b) Is it always possible to find S so that $LS(S) = \mathbb{V}$?

   (c) Suppose we have found $S \subseteq \mathbb{V}$ such that $LS(S) = \mathbb{V}$. Can we find the minimum number of vectors in S?

   We try to answer these questions in the subsequent sections.

3.2 Linear Independence

Definition 3.3.2.1. [Linear Independence and Dependence] Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a non-empty subset of a vector space $\mathbb{V}$ over $\mathbb{F}$. Then the set S is said to be **linearly independent** if the system of linear equations

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}, \tag{3.3.2.1}$$

in the unknowns α_i's, $1 \le i \le m$, has only the trivial solution. If the system (3.3.2.1) has a non-trivial solution then the set S is said to be **linearly dependent**.

If the set 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, else S is **linearly dependent**.

Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subseteq \mathbb{V}$ with $S \neq \emptyset$. Then, one needs to solve the linear system of equation

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0} \tag{3.3.2.2}$$

in the unknowns $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. If $\alpha_1 = \cdots = \alpha_m = 0$ is THE ONLY SOLUTION of (3.3.2.2) then S is a linearly independent subset of $\mathbb{V}$. Otherwise, S is a linearly dependent subset of $\mathbb{V}$. Since one is solving a linear system over $\mathbb{F}$, linear independence and dependence depend on $\mathbb{F}$, the set of scalars.

Example 3.3.2.2. Is the set S a linear independent set? Give reasons.

1. Let $S = \{(1,2,1)^T, (2,1,4)^T, (3,3,5)^T\}$.
   Solution: Consider the system $a(1,2,1) + b(2,1,4) + c(3,3,5) = (0,0,0)$ in the unknowns a, b and c. As rank of coefficient matrix is $2 < 3$, the number of unknowns, the system has a non-trivial solution. Thus, S is a linearly dependent subset of $\mathbb{R}^3$.

2. Let $S = \{(1,1,1)^T, (1,1,0)^T, (1,0,1)^T\}$.
   Solution: Consider the system $a(1,1,1) + b(1,1,0) + c(1,0,1) = (0,0,0)$ in the unknowns a, b and c. As rank of coefficient matrix is $3 =$ the number of unknowns, the system has only the trivial solution. Hence, S is a linearly independent subset of $\mathbb{R}^3$.

3. 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 the complex vector space $\mathbb{C}$.

4. 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 unknowns $a, b \in \mathbb{R}$. Since $a, b \in \mathbb{R}$, equating real and imaginary parts, we get $a = b = 0$. So, S is a linear independent subset the real vector space $\mathbb{C}$.

5. Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$. If $\text{Rank}(A) < \min\{m, n\}$ then the rows of A are linearly dependent.
Solution: Let $B = \text{RREF}(A)$. Then, there exists an invertible matrix $P = [p_{ij}]$ such that $B = PA$. Since $\text{Rank}(A) < \min\{m, n\}$, $B[m, :] = \mathbf{0}^T$. Thus, $\mathbf{0}^T = B[m, :] = \sum\limits_{i=1}^{n} p_{mi} A[i, :]$. As P is invertible, at least one $p_{mi} \neq 0$. Thus, the required result follows.

3.2.A Basic Results related to Linear Independence

The reader is expected to supply the proof of parts that are not given.

Proposition 3.3.2.3. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then,*

1. $\mathbf{0}$*, the zero-vector, cannot belong to a linearly independent set.*

2. *every non-empty subset of a linearly independent set in $\mathbb{V}$ is also linearly independent.*

3. *a set containing a linearly dependent set of $\mathbb{V}$ is also linearly dependent.*

Proof. Let $S = \{\mathbf{0} = \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$. Then, $1 \cdot \mathbf{u}_1 + 0 \cdot \mathbf{u}_2 + \cdots + 0 \cdot \mathbf{u}_n = \mathbf{0}$. Hence, the system $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}$ has a non-trivial solution $[\alpha_1, \alpha_2, \ldots, \alpha_n] = [1, 0 \ldots, 0]$. Thus, the set S is linearly dependent. ■

Theorem 3.3.2.4. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subseteq \mathbb{V}$ with $S \neq \emptyset$. If $T \subseteq LS(S)$ such that $m = \mid T \mid > k$ then T is a linearly dependent set.*

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 $\mathbf{w}_i = a_{i1}\mathbf{u}_1 + \cdots + a_{ik}\mathbf{u}_k$, for $1 \leq i \leq m$. 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 2.2.2.35, the system $A^T\mathbf{x} = \mathbf{0}$ has a non-trivial solution, say $\mathbf{x}^T = [\alpha_1, \ldots, \alpha_m] \neq \mathbf{0}^T$. That is, $\sum\limits_{i=1}^{m} \alpha_i A^T[:, i] = \mathbf{0}$. Or equivalently, $\sum\limits_{i=1}^{m} \alpha_i A[i, :] = \mathbf{0}^T$. Thus,

$$\sum_{i=1}^{m} \alpha_i \mathbf{w}_i = \sum_{i=1}^{m} \alpha_i \left(A[i, :] \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} \right) = \left(\sum_{i=1}^{m} \alpha_i A[i, :] \right) \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{0}^T \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{0}.$$

Thus, the set T is linearly dependent. ■

Corollary 3.3.2.5. *Fix $n \in \mathbb{N}$. Then, any set $S \subseteq \mathbb{R}^n$ with $|\, S \,| \geq n+1$ is linearly dependent.*

Proof. Observe that $\mathbb{R}^n = LS(\{\mathbf{e}_1, \ldots, \mathbf{e}_n\})$, where $\mathbf{e}_i = I_n[:, i]$, the i-th column of I_n. Hence, the required result follows using Theorem 3.3.2.4. ∎

Theorem 3.3.2.6. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let S be a linearly independent subset of $\mathbb{V}$. Suppose $\mathbf{v} \in \mathbb{V}$. Then $S \cup \{\mathbf{v}\}$ 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 $\mathbf{v}_i \in S$ such that the system $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p + \alpha_{p+1} \mathbf{v} = \mathbf{0}$ has a non-trivial solution, say $\alpha_i = c_i$, for $1 \leq i \leq p+1$. As the solution is non-trivial one of the c_i's is non-zero. We claim that $c_{p+1} \neq 0$.

For, if $c_{p+1} = 0$ then the system $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p = \mathbf{0}$ in the unknowns $\alpha_1, \ldots, \alpha_p$ has a non-trivial solution $[c_1, \ldots, c_p]$. This contradicts Proposition 3.3.2.3.2 as $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subset of the linearly independent set S. Thus, $c_{p+1} \neq 0$ and we get

$$\mathbf{v} = -\frac{1}{c_{p+1}}(c_1 \mathbf{v}_1 + \cdots + c_p \mathbf{v}_p) \in L(\mathbf{v}_1, \ldots, \mathbf{v}_p)$$

as $-\frac{c_i}{c_{p+1}} \in \mathbb{F}$, for $1 \leq i \leq p$. Hence, $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

Now, assume that $\mathbf{v} \in LS(S)$. Then, there exists $c_i \in \mathbb{F}$, not all zero and $\mathbf{v}_i \in S$ such that $\mathbf{v} = \sum\limits_{i=1}^{p} c_i \mathbf{v}_i$. Thus, the system $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p + \alpha_{p+1} \mathbf{v} = \mathbf{0}$ in the unknowns 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 3.3.2.6 without proof.

Corollary 3.3.2.7. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{V}$ with $\mathbf{u}_1 \neq \mathbf{0}$. If S is*

1. *linearly dependent then there exists k, $2 \leq k \leq n$ with $LS(\mathbf{u}_1, \ldots, \mathbf{u}_k) = LS(\mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$.*
2. *linearly independent then $\mathbf{v} \in \mathbb{V} \setminus LS(S)$ if and only if $S \cup \{\mathbf{v}\} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}\}$ is also a linearly independent subset of $\mathbb{V}$.*
3. *linearly independent then $LS(S) = \mathbb{V}$ if and only if each proper superset of S is linearly dependent.*

3.2.B Application to Matrices

We leave the proof of the next result for readers.

Theorem 3.3.2.8. *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.*

A generalization of Theorem 3.3.2.8 is stated next.

Theorem 3.3.2.9. *Let $A \in \mathbb{M}_{m,n}(\mathbb{C})$ with $B = RREF(A)$. Then, the rows of A corresponding to the pivotal rows of B are linearly independent. Also, the columns of A corresponding to the pivotal columns of B are linearly independent.*

Proof. 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. Let A_1 be the submatrix of A which gives B_1. As the RREF of a matrix is unique (see Corollary 2.2.2.17) 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 2.2.3.1). Thus, it contradicts the linear independence of the pivotal rows of B.

Let $B[:, i_1], \ldots, B[:, i_r]$ be the pivotal columns of B which are linearly independent due to pivotal 1's. Let $B = PA$ then the corresponding columns of A satisfy

$$[A[:, i_1], \ldots, A[:, i_r]] = P^{-1}[B[:, i_1], \ldots, B[:, i_r]].$$

Hence, the required result follows as P is an invertible matrix. ■

We give an example for better understanding.

Example 3.3.2.10. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$ with $\mathrm{RREF}(A) = B = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Then $B[:, 3] = -B[:, 1] + 2B[:, 2]$. Thus, $A[:, 3] = -A[:, 1] + 2A[:, 2]$. 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. Also, note that during the application of GJE to get RREF, we have interchanged the 3-rd and 4-th rows. Hence, the rows $A[1, :], A[2, :]$ and $A[4, :]$ are linearly independent.

3.2.C 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 3.3.2.11. *Let S be a linearly independent set in a vector space $\mathbb{V}$ over $\mathbb{F}$. Then each $\mathbf{v} \in LS(S)$ is a unique linear combination vectors from S.*

Proof. On the contrary, suppose there exists $\mathbf{v} \in LS(S)$ such that $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_p \mathbf{v}_p$ and $\mathbf{v} = \beta_1 \mathbf{v}_1 + \cdots + \beta_p \mathbf{v}_p$, for $\alpha_i, \beta_i \in \mathbb{F}$ and $\mathbf{v}_i \in S$, for $1 \leq i \leq p$. Equating the two expressions for $\mathbf{v}$ gives

$$(\alpha_1 - \beta_1)\mathbf{v}_1 + \cdots + (\alpha_p - \beta_p)\mathbf{v}_p = \mathbf{0}. \tag{3.3.2.3}$$

As $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subseteq S$ is a linearly independent subset in $LS(S)$, the system $c_1 \mathbf{v}_1 + \cdots + c_p \mathbf{v}_p = \mathbf{0}$ in the unknowns $c_1, \ldots, c_p$ has only the trivial solution. Thus, each of the scalars $\alpha_i - \beta_i$, appearing in Equation (3.3.2.3), must be equal to 0. That is, $\alpha_i - \beta_i = 0$, for $1 \leq i \leq p$. Thus, for $1 \leq i \leq p$, $\alpha_i = \beta_i$ and the result follows. ■

EXERCISE **3.3.2.12.**

1. *Consider the Euclidean plane $\mathbb{R}^2$. Let $\mathbf{u}_1 = (1,0)^T$. Determine condition on $\mathbf{u}_2$ such that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a linearly independent subset of $\mathbb{R}^2$.*

2. *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 $(x,y,z,u)^T \in LS(S)$.*

3. *Show that $S = \{(1,2,3)^T, (-2,1,1)^T, (8,6,10)^T\} \subseteq \mathbb{R}^3$ is linearly dependent.*

4. *Prove that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{C}^n$ is linearly independent if and only if $\{A\mathbf{u}_1, \ldots, A\mathbf{u}_n\}$ is linearly independent for every invertible matrix A.*

5. *Let $\mathbb{V}$ be a complex vector space and let $A \in \mathbb{M}_n(\mathbb{C})$ be invertible. Then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{V}$ is a linearly independent if and only if $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \subseteq \mathbb{V}$ is linearly independent, where $\mathbf{w}_i = \sum\limits_{j=1}^{n} a_{ij}\mathbf{u}_j$, for $1 \le i \le n$.*

6. *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.*

7. *Is $\{(1,0)^T, (i,0)^T\}$ a linearly independent subset of the real vector space $\mathbb{C}^2$?*

8. *Suppose $\mathbb{V}$ is a collection of vectors such that $\mathbb{V}$ is a real as well as a complex vector space. Then prove that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, i\mathbf{u}_1, \ldots, i\mathbf{u}_k\}$ is a linearly independent subset of $\mathbb{V}$ over $\mathbb{R}$ if and only if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a linear independent subset of $\mathbb{V}$ over $\mathbb{C}$.*

9. *Let $\mathbb{V}$ be a vector space and M be a subspace of $\mathbb{V}$. For $\mathbf{u}, \mathbf{v} \in \mathbb{V} \setminus M$, define $K = LS(M, \mathbf{u})$ and $H = LS(M, \mathbf{v})$. Then prove that $\mathbf{v} \in K$ if and only if $\mathbf{u} \in H$.*

10. *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.*

11. *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 $AP = PD$.*

12. *Prove that the rows/columns of*

    (a) *$A \in M_n(\mathbb{C})$ are linearly independent if and only if $\det(A) \neq 0$.*

    (b) *$A \in M_n(\mathbb{C})$ span $\mathbb{C}^n$ if and only if A is an invertible matrix.*

(c) a skew-symmetric matrix A of odd order are linearly dependent.

13. *Let P and Q be subspaces of $\mathbb{R}^n$ such that $P + Q = \mathbb{R}^n$ and $P \cap Q = \{\mathbf{0}\}$. Prove that each $\mathbf{u} \in \mathbb{R}^n$ is uniquely expressible as $\mathbf{u} = \mathbf{u}_P + \mathbf{u}_Q$, where $\mathbf{u}_P \in P$ and $\mathbf{u}_Q \in Q$.*

3.3 Basis of a Vector Space

Definition 3.3.3.1. 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 S of T has property P.

Example 3.3.3.2. 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 3.3.3.3. 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 S of $\mathbb{V}$ linearly independent.

Example 3.3.3.4.
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 any collection of 4 or more vectors from $\mathbb{R}^3$ is linearly dependent (see Corollary 3.3.2.5).
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 3.3.2.9, 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.

Theorem 3.3.3.5. *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 3.3.2.7.2, 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 $|S| = k$. Then, using Theorem 3.3.2.4, we see that if $T \subseteq \mathbb{V}$ is linearly independent then $|T| \leq k$. Hence, a maximal linearly independent subset of $\mathbb{V}$ can have at most k vectors. Thus, we arrive at the following important result.

Theorem 3.3.3.6. *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 $|S| = |T|$.*

Proof. By Theorem 3.3.3.5, 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. ∎

Definition 3.3.3.7. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\mathbb{V} \neq \{\mathbf{0}\}$. Suppose $\mathbb{V}$ has a finite maximal linearly independent set S. Then $\mid S \mid$ is called the **dimension** of $\mathbb{V}$, denoted $\dim(\mathbb{V})$. By convention, $\dim(\{\mathbf{0}\}) = 0$.

Example 3.3.3.8.

1. As $\{\pi\}$ is a maximal linearly independent subset of $\mathbb{R}$, $\dim(\mathbb{R}) = 1$.
2. As $\{(1,0,1)^T, (0,1,1)^T, (1,1,0)^T\} \subseteq \mathbb{R}^3$ is maximal linearly independent, $\dim(\mathbb{R}^3) = 3$.
3. As $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a maximal linearly independent set in $\mathbb{R}^n$, $\dim(\mathbb{R}^n) = n$.
4. As $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a maximal linearly independent subset of the complex vector space $\mathbb{C}^n$, $\dim(\mathbb{C}^n) = n$.
5. Using Exercise 3.3.2.12.8, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n, i\mathbf{e}_1, \ldots, i\mathbf{e}_n\}$ is a maximal linearly independent subset of the real vector space $\mathbb{C}^n$. Thus, as a real vector space $\dim(\mathbb{C}^n) = 2n$.
6. 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 3.3.3.4.3, we see that $\dim(LS(S)) = \text{Rank}(A)$.

Definition 3.3.3.9. [Basis of a Vector Space] Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\mathbb{V} \neq \{\mathbf{0}\}$. Then a maximal linearly independent subset of $\mathbb{V}$ is called a **basis** of $\mathbb{V}$. The vectors in a basis are called **basis** vectors. Note that a basis of $\{\mathbf{0}\}$ is either not defined or is the empty set.

Definition 3.3.3.10. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\mathbb{V} \neq \{\mathbf{0}\}$. Then a set $S \subseteq \mathbb{V}$ is called **minimal spanning** if $LS(S) = \mathbb{V}$ and no proper subset of S spans $\mathbb{V}$.

Example 3.3.3.11. [Standard Basis] Fix $n \in \mathbb{N}$ and let $\mathbf{e}_i = I_n[:, i]$, the i-th column of the identity matrix. Then $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$ or $\mathbb{C}^n$. In particular,

1. $\mathcal{B} = \{e_1\} = \{1\}$ is a standard basis of $\mathbb{R}$ over $\mathbb{R}$.
2. $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ with $\mathbf{e}_1^T = (1,0)^T$ and $\mathbf{e}_2^T = (0,1)^T$ is the standard basis of $\mathbb{R}^2$.
3. $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \{(1,0,0)^T, (0,1,0)^T, (0,0,1)^T\}$ is the standard basis of $\mathbb{R}^3$.
4. $\{1\}$ is a basis of $\mathbb{C}$ over $\mathbb{C}$.
5. $\{\mathbf{e}_1, \ldots, \mathbf{e}_n, i\mathbf{e}_1, \ldots, i\mathbf{e}_n\}$ is a basis of $\mathbb{C}^n$ over $\mathbb{R}$. So, $\{1, i\}$ is a basis of $\mathbb{C}$ over $\mathbb{R}$.

Example 3.3.3.12.

1. Note that $\{-2\}$ is a basis and a minimal spanning set of $\mathbb{R}$.
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 $S = \{1, 2, \ldots, n\}$ and consider the vector space $\mathbb{R}^S$ (see Example 3.3.1.4.8). Then, for $1 \le i \le n$, define $\mathbf{e}_i(j) = \begin{cases} 1, & \text{if } j = i \\ 0, & \text{if } j \neq i. \end{cases}$ Prove that $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a linearly independent set. Is it a basis of $\mathbb{R}^S$?

7. Let $S = \mathbb{R}^n$ and consider the vector space $\mathbb{R}^S$ (see Example 3.3.1.4.8). 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 set. Is it a basis of $\mathbb{R}^S$?

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 Theorem 3.3.2.9, the columns of A corresponding to the pivotal columns in RREF(A) form a basis as well as a minimal spanning subset of $LS(S)$.

9. Let $S = \{a_1, \ldots, a_n\}$. Then recall that $\mathbb{R}^S$ is a real vector space (see Example 8). For $1 \le i \le n$, define $f_i(a_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$. Then, verify that $\{f_1, \ldots, f_n\}$ is a basis of $\mathbb{R}^S$. What can you say if S is a countable set?

10. Recall the vector space $\mathcal{C}[a, b]$, where $a < b \in \mathbb{R}$. For each $\alpha \in [a, b] \cap \mathbb{Q}$, define $f_\alpha(x) = x - \alpha$, for all $x \in [a, b]$. Is the set $\{f_\alpha \mid \alpha \in [a, b]\}$ linearly independent? What if $\alpha \in [a, b]$? Can we write any function in $\mathcal{C}[a, b]$ as a finite linear combination? Give reasons for your answer.

3.3.A Main Results associated with Bases

Theorem 3.3.3.13. *Let $\mathbb{V} \neq \{\mathbf{0}\}$ be a 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 it spans $\mathbb{V}$.*
3. *$\mathcal{B}$ is a minimal spanning set of $\mathbb{V}$.*

Proof. $1 \Rightarrow 2$ By definition, every basis is a maximal linearly independent subset of $\mathbb{V}$. Thus, using Corollary 3.3.2.7.2, 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 $\mathbf{x} \in S$, $\mathbf{x} \notin LS\left(S - \{\mathbf{x}\}\right)$. Hence $LS\left(S - \{\mathbf{x}\}\right) \neq \mathbb{V}$.

$3 \Rightarrow 1$ If $\mathcal{B}$ is linearly dependent then using Corollary 3.3.2.7.1 $\mathcal{B}$ 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 $\mathcal{B}$ spans $\mathbb{V}$, for any $\mathbf{x} \in \mathbb{V} \setminus \mathcal{B}$, 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 3.3.2.11, we get the following result.

Remark 3.3.3.14. *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 in a finite dimensional vector space".

Theorem 3.3.3.15. *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 3.3.2.7.2, 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 3.3.3.13, this T is indeed a required basis. ■

3.3.B 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 3.3.2.7.2, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

Step i: Either $\mathbb{V} = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$ or $LS(\mathbf{v}_1, \ldots, \mathbf{v}_i) \neq \mathbb{V}$. 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 3.3.2.7.2, 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 **3.3.3.16.**

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 $1 \leq i \leq n$?*
2. *Find a basis of $\mathbb{R}^3$ containing the vector $(1, 1, -2)^T$.*
3. *Find a basis of $\mathbb{R}^3$ containing the vector $(1, 1, -2)^T$ and $(1, 2, -1)^T$.*
4. *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 $(1, -2, 7, -11)^T$?*
5. *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?*
6. *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 $\mathbb{C}$.*
7. *Find a basis of $\mathbb{C}^3$ over $\mathbb{R}$ containing the basis $\mathcal{B}$ given in Example 3.3.3.16.6.*
8. *Determine a basis and dimension of $W = \{(x, y, z, w)^T \in \mathbb{R}^4 \mid x + y - z + w = 0\}$.*
9. *Find a basis of $\mathbb{V} = \{(x, y, z, u) \in \mathbb{R}^4 \mid x - y - z = 0, x + z - u = 0\}$.*
10. *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 $\mathbb{V} = \{\mathbf{x} \in \mathbb{R}^5 \mid A\mathbf{x} = \mathbf{0}\}$.*

11. *Prove that* $\mathcal{B} = \{1, x, \ldots, x^n, \ldots\}$ *is a basis of* $\mathbb{R}[x]$*.* $\mathcal{B}$ *is called the standard basis of* $\mathbb{R}[x]$*.*

12. *Let* $\mathbf{u}^T = (1, 1, -2), \mathbf{v}^T = (-1, 2, 3)$ *and* $\mathbf{w}^T = (1, 10, 1)$*. Find a basis of* $L(\mathbf{u}, \mathbf{v}, \mathbf{w})$*. Determine a geometrical representation of* $LS(\mathbf{u}, \mathbf{v}, \mathbf{w})$*.*

13. *Let* $\mathbb{V}$ *be a vector space of dimension* n*. Then any set*

    (a) *consisting of* n *linearly independent vectors forms a basis of* $\mathbb{V}$*.*

    (b) S *in* $\mathbb{V}$ *having* n *vectors with* $LS(S) = \mathbb{V}$ *forms a basis of* $\mathbb{V}$*.*

14. *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.*

15. *Let* $\mathbb{W}_1$ *and* $\mathbb{W}_2$ *be two subspaces of a vector space* $\mathbb{V}$ *such that* $\mathbb{W}_1 \subseteq \mathbb{W}_2$*. Show that* $\mathbb{W}_1 = \mathbb{W}_2$ *if and only if* $\dim(\mathbb{W}_1) = \dim(\mathbb{W}_2)$*.*

16. *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* $1 \leq t\ell$*. 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*
$$\int_{-\pi}^{\pi} \sin(mx) \left(\alpha_1 \sin(k_1 x) + \cdots + \alpha_\ell \sin(k_\ell x)\right)\ dx$$
replace m *with* k_i*'s to show that* $\alpha_i = 0$*, for all* $i, 1 \leq i \leq \ell$ *to get the required contradiction.]*

17. *Let* $\mathbb{V}$ *be a vector space over* $\mathbb{F}$ *with* $\dim(\mathbb{V}) = n$*. If* $\mathbb{W}_1$ *is a* k*-dimensional subspace of* $\mathbb{V}$ *then prove that there exists a subspace* $\mathbb{W}_2$ *of* $\mathbb{V}$ *such that* $\mathbb{W}_1 \cap \mathbb{W}_2 = \{\mathbf{0}\}$*,* $\mathbb{W}_1 + \mathbb{W}_2 = \mathbb{V}$ *and* $\dim(\mathbb{W}_2) = n - k$*. Also, prove that for each* $\mathbf{v} \in \mathbb{V}$ *there exist unique vectors* $\mathbf{w}_1 \in \mathbb{W}_1$ *and* $\mathbf{w}_2 \in \mathbb{W}_2$ *such that* $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$*. The subspace* $\mathbb{W}_2$ *is called the complementary subspace of* $\mathbb{W}_1$ *in* $\mathbb{V}$*.*

18. *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 Application to the subspaces of $\mathbb{C}^n$

In this subsection, we will study results that are intrinsic to the understanding of linear algebra from the point of view of matrices, especially the fundamental subspaces (see Definition 3.3.1.25) associated with matrices. We start with an example.

Example 3.3.4.1. 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) $\text{Col}(A^*) = \text{Row}(A) = \{(x,y,z,u)^T \in \mathbb{C}^4 \mid 3x-2y=z, 5x-3y+u=0\}$.

(b) $\text{Col}(A) = \text{Row}(A^*) = \{(x,y,z)^T \in \mathbb{C}^3 \mid 4x-3y-z=0\}$.

(c) $\text{Null}(A) = \{(x,y,z,u)^T \in \mathbb{C}^4 \mid x+3z-5u=0, y-2z+3u=0\}$.

(d) $\text{Null}(A^*) = \{(x,y,z)^T \in \mathbb{C}^3 \mid x+4z=0, y-3z=0\}$.

2. Let $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 & 3 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$. Find a basis and dimension of $\text{Null}(A)$.

Solution: Writinf 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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} x_7 - x_2 - x_4 - x_5 \\ x_2 \\ 2x_7 - 2x_4 - 3x_5 \\ x_4 \\ x_5 \\ -x_7 \\ x_7 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}. \quad (3.3.4.1)$$

Now, let $\mathbf{u}_1^T = \left[-1,1,0,0,0,0,0\right]$, $\mathbf{u}_2^T = \left[-1,0,-2,1,0,0,0\right]$, $\mathbf{u}_3^T = \left[-1,0,-3,0,1,0,0\right]$ and $\mathbf{u}_4^T = \left[1,0,2,0,0,-1,1\right]$. 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 (3.3.4.1) $\text{Null}(A) = LS(S)$.

(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 unknowns 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 non-zero 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 non-zero entry at the 5-th place ($\mathbf{u}_3$ corresponds to x_5) and hence $c_3 = 0$.

iii. Similar arguments hold for the unknowns c_2 and c_1.

Exercise **3.3.4.2.** *Let* $A = \begin{bmatrix} 1 & 2 & 1 & 3 & 2 \\ 0 & 2 & 2 & 2 & 4 \\ 2 & -2 & 4 & 0 & 8 \\ 4 & 2 & 5 & 6 & 10 \end{bmatrix}$ *and* $B = \begin{bmatrix} 2 & 4 & 0 & 6 \\ -1 & 0 & -2 & 5 \\ -3 & -5 & 1 & -4 \\ -1 & -1 & 1 & 2 \end{bmatrix}$.

1. *Find RREF(A) and RREF(B).*

2. *Find invertible matrices P_1 and P_2 such that $P_1 A = RREF(A)$ and $P_2 B = RREF(B)$.*

3. *Find bases for* $\text{Col}(A)$, $\text{Col}(A^*)$, $\text{Col}(B)$ *and* $\text{Col}(B^*)$.

4. *Find bases of* $\text{Null}(A)$, $\text{Null}(A^*)$, $\text{Null}(B)$ *and* $\text{Null}(B^*)$.

5. *Find the dimensions of all the vector subspaces so obtained.*

6. *Does there exist relationship between the bases of different spaces?*

Lemma 3.3.4.3. *Let* $A \in M_{m\times n}(\mathbb{C})$ *and let* E *be an elementary matrix. If*

1. $B = EA$ *then* $\text{Col}(A^*) = \text{Col}(B^*)$ *and* $\text{Col}(A^T) = \text{Col}(B^T)$. *Hence* $\dim(\text{Col}(A^*)) = \dim(\text{Col}(B^*))$ *and* $\dim(\text{Col}(A^T)) = \dim(\text{Col}(B^T))$.

2. $B = AE$ *then* $\text{Col}(A) = \text{Col}(B)$ *and* $\text{Col}(\overline{A}) = \text{Col}(\overline{B})$. *Hence* $\dim(\text{Col}(A)) = \dim(\text{Col}(B))$ *and* $\dim(\text{Col}(\overline{A})) = \dim(\text{Col}(\overline{B}))$.

Proof. Note that $B = EA$ if and only if $\overline{B} = \overline{EA}$. As E is invertible, A are B are equivalent and hence they have the same RREF. Also, $\overline{E}$ is invertible as well and hence $\overline{A}$ are $\overline{B}$ have the same RREF. Now, use Theorem 3.3.2.9 to get the required result.

For the second part, note that $B^* = E^*A^*$ and E^* is invertible. Hence, using the first part $\text{Col}((A^*)^*) = \text{Col}((B^*)^*)$, or equivalently, $\text{Col}(A) = \text{Col}(B)$. ∎

Let $A \in M_{m\times n}(\mathbb{C})$ and let $B = \text{RREF}(A)$. Then as an immediate application of Lemma 3.3.4.3, we get $\dim(\text{Col}(A^*)) = \text{Row rank}(A)$. Hence, $\dim(\text{Col}(A)) = \text{Column rank}(A)$ as well. We now prove that $\text{Row rank}(A) = \text{Column rank}(A)$.

Theorem 3.3.4.4. *Let* $A \in M_{m\times n}(\mathbb{C})$. *Then* $\text{Row rank}(A) = \text{Column rank}(A)$.

Proof. Let $\text{Row rank}(A) = r = \dim(\text{Col}(A^T))$. Then there exist $i_1, \ldots, i_r$ such that $\{A[i_1,:], \ldots, A[i_r,:]\}$ form a basis of $\text{Col}(A^T)$. Then, $B = \begin{bmatrix} A[i_1,:] \\ \vdots \\ A[i_r,:] \end{bmatrix}$ is an $r \times n$ matrix and it's rows are a basis of $\text{Col}(A^T)$. Therefore, there exist $\alpha_{ij} \in \mathbb{C}$, $1 \le i \le m$, $1 \le j \le r$ such that $A[t,:] = [\alpha_{t1}, \ldots, \alpha_{tr}]B$, for $1 \le t \le m$. So, using matrix multiplication (see Remark 1.1.2.11.4)

$$A = \begin{bmatrix} A[1,:] \\ \vdots \\ A[m,:] \end{bmatrix} = \begin{bmatrix} [\alpha_{11}, \ldots, \alpha_{1r}]B \\ \vdots \\ [\alpha_{m1}, \ldots, \alpha_{mr}]B \end{bmatrix} = CB,$$

where $C = [\alpha_{ij}]$ is an $m \times r$ matrix. Thus, matrix multiplication implies that each column of A is a linear combination of r columns of C. Hence, $\text{Column rank}(A) = \dim(\text{Col}(A)) \le r = \text{Row rank}$. A similar argument gives $\text{Row rank}(A) \le \text{Column rank}(A)$. Hence, we have the required result. ∎

Remark 3.3.4.5. *The proof also shows that for every* $A \in M_{m\times n}(\mathbb{C})$ *of rank* r *there exists matrices* $B_{r\times n}$ *and* $C_{m\times r}$, *each of rank* r, *such that* $A = CB$.

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 3.3.1.24.7d) $\mathbb{W}_1 + \mathbb{W}_2 = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in \mathbb{W}_1, \mathbf{v} \in \mathbb{W}_2\} = LS(\mathbb{W}_1 \cup \mathbb{W}_2)$ 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 7.7.4.1).

Theorem 3.3.4.6. *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*

$$\dim(\mathbb{W}_1) + \dim(\mathbb{W}_2) = \dim(\mathbb{W}_1 + \mathbb{W}_2) + \dim(\mathbb{W}_1 \cap \mathbb{W}_2). \tag{3.3.4.2}$$

For better understanding, we give an example for finite subsets of $\mathbb{R}^n$. The example uses Theorem 3.3.2.9 to obtain bases of $LS(S)$, for different choices S. The readers are advised to see Example 3.3.2.9 before proceeding further.

Example 3.3.4.7. Let $\mathbb{V}$ and $\mathbb{W}$ be two spaces with $\mathbb{V} = \{(v, w, x, y, z)^T \in \mathbb{R}^5 \mid v+x+z = 3y\}$ 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: $(v, w, x, y, z)^T \in \mathbb{V} \cap \mathbb{W}$ 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(3.3.4.3)) and apply row operations other than E_{ij}. Then after a few row operations, we get

$$\left[\begin{array}{ccccc} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ \hline -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccccc} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{3.3.4.3}$$

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 **3.3.4.8.**

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 $\mathbb{R}^4$. 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 4-dimensional subspaces of a vector space $\mathbb{V}$ of dimension 7. Then prove that* $\dim(\mathbb{W}_1 \cap \mathbb{W}_2) \geq 1$.

4. *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* $\dim(\mathbb{W}_1 \cap \mathbb{W}_2) \geq 1$.

5. *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 3.3.4.9 (Rank-Nullity Theorem)**.** *Let* $A \in M_{m\times n}(\mathbb{C})$. *Then*

$$\dim(\textsc{Col}(A)) + \dim(\textsc{Null}(A)) = n. \tag{3.3.4.4}$$

Proof. Let $\dim(\textsc{Null}(A)) = r \leq n$ and let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis of $\textsc{Null}(A)$. Since $\mathcal{B}$ is a linearly independent set in $\mathbb{R}^n$, extend it to get $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ as a basis of $\mathbb{R}^n$. Then,

$$\begin{aligned} \textsc{Col}(A) &= LS(\mathcal{B}) = LS(A\mathbf{u}_1, \ldots, A\mathbf{u}_n) \\ &= LS(\mathbf{0}, \ldots, \mathbf{0}, A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n) = LS(A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n). \end{aligned}$$

So, $\mathcal{C} = \{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ spans $\textsc{Col}(A)$. We further need to show that $\mathcal{C}$ is linearly independent. So, consider the linear system

$$\alpha_1 A\mathbf{u}_{r+1} + \cdots + \alpha_{n-r} A\mathbf{u}_n = \mathbf{0} \Leftrightarrow A(\alpha_1 \mathbf{u}_{r+1} + \cdots + \alpha_{n-r}\mathbf{u}_n) = \mathbf{0} \tag{3.3.4.5}$$

in the unknowns $\alpha_1, \ldots, \alpha_{n-r}$. Thus, $\alpha_1\mathbf{u}_{r+1} + \cdots + \alpha_{n-r}\mathbf{u}_n \in \textsc{Null}(A) = LS(\mathcal{B})$. Therefore, there exist scalars $\beta_i,\ 1 \leq i \leq r$, 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,

$$\beta_1\mathbf{u}_1 + \cdots + \beta_r\mathbf{u}_r - \alpha_1\mathbf{u}_{r+1} - \cdots - \alpha_{n-r}\mathbf{u}_n = \mathbf{0}. \tag{3.3.4.6}$$

As $\mathcal{B}$ is a linearly independent set, the only solution of Equation (3.3.4.6) 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 (3.3.4.5) is the trivial solution. Hence, $\{A\mathbf{u}_{r+1}, \ldots, A\mathbf{u}_n\}$ is a basis of $\textsc{Col}(A)$. Thus, the required result follows. ∎

Theorem 3.3.4.9 is part of what is known as the fundamental theorem of linear algebra (see Theorem 5.5.1.23). The following are some of the consequences of the rank-nullity theorem. The proof is left as an exercise for the reader.

Exercise **3.3.4.10.** 1. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{C})$. *If*

(a) $n > m$ *then the system* $A\mathbf{x} = \mathbf{0}$ *has infinitely many solutions,*

(b) $n < m$ *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) $\dim(\text{COL}(A)) = k$.

(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* $1 \le i \le k$, *is consistent.*

(g) $\dim(\text{NULL}(A)) = n - k$.

3.5 Ordered Bases

Let $\mathbb{V}$ be a vector subspace of $\mathbb{C}^n$ for some $n \in \mathbb{N}$ with $\dim(\mathbb{V}) = k$. Then, a basis of $\mathbb{V}$ may not look like a standard basis. Our problem may force us to look for some other basis. In such a case, it is always 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. It also may happen that k is very-very small as compared to n and hence it is always better to work with vectors of small size.

Definition 3.3.5.1. [Ordered Basis, Basis Matrix] Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with a basis $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Then an **ordered basis** for $\mathbb{V}$ is a basis $\mathcal{B}$ together with a one-to-one correspondence between $\mathcal{B}$ and $\{1, 2, \ldots, n\}$. Since there is an order among the elements of $\mathcal{B}$, we write $\mathcal{B} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. The matrix $B = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is called the **basis matrix**.

Thus, $\mathcal{B} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ is different from $\mathcal{C} = [\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n, \mathbf{u}_1]$ and both of them are different from $\mathcal{D} = [\mathbf{u}_n, \mathbf{u}_{n-1}, \ldots, \mathbf{u}_2, \mathbf{u}_1]$ even though they have the same set of vectors as elements. We now define the notion of coordinates of a vector with respect to an ordered basis.

Definition 3.3.5.2. [Coordinates of a Vector] Let $B = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ be the basis matrix corresponding to an ordered basis $\mathcal{B}$ of $\mathbb{V}$. Since $\mathcal{B}$ is a basis of $\mathbb{V}$, for each $\mathbf{v} \in \mathbb{V}$, there exist $\beta_i, 1 \le i \le n$, such that $\mathbf{v} = \sum\limits_{i=1}^{n} \beta_i \mathbf{v}_i = B \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$. The column vector $\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$ is called **the coordinates of v with respect to** $\mathcal{B}$, denoted $[\mathbf{v}]_{\mathcal{B}}$. Thus, using notation $\mathbf{v} = B[\mathbf{v}]_{\mathcal{B}}$.

Example 3.3.5.3. 1. 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, $\begin{bmatrix} \pi \\ e \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \pi \\ e \end{bmatrix}_{\mathcal{B}}$.
Thus, $\begin{bmatrix} \pi \\ e \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} \pi \\ e \end{bmatrix}$.

2. Consider the vector space $\mathbb{R}[x; 2]$ with basis $\{1 - x, 1 + x, x^2\}$. Then an ordered basis can either be $\mathcal{B} = [1 - x, 1 + x, x^2]$ or $\mathcal{C} = [1 + x, 1 - x, x^2]$ or .... Note that there are 3! different ordered bases. Also, for $a_0 + a_1x + a_2x^2 \in \mathbb{R}[x; 2]$, one has

$$a_0 + a_1x + a_2x^2 = [1 - x, 1 + x, x^2] \begin{bmatrix} \frac{a_0 - a_1}{2} \\ \frac{a_0 + a_1}{2} \\ a_2 \end{bmatrix}.$$

Thus, $[a_0 + a_1x + a_2x^2]_{\mathcal{B}} = \begin{bmatrix} \frac{a_0-a_1}{2} \\ \frac{a_0+a_1}{2} \\ a_2 \end{bmatrix}$, whereas $[a_0 + a_1x + a_2x^2]_{\mathcal{C}} = \begin{bmatrix} \frac{a_0+a_1}{2} \\ \frac{a_0-a_1}{2} \\ a_2 \end{bmatrix}$.

3. Let $\mathbb{V} = \{(x, y, z)^T \mid x + y = z\}$. If $\mathcal{B} = [(-1, 1, 0)^T, (1, 0, 1)^T]$ is an ordered basis of $\mathbb{V}$ then $\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ x+y \end{bmatrix}$ and hence $\begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$.

4. Let $\mathbb{V} = \{(v, w, x, y, z)^T \in \mathbb{R}^5 \mid w - x = z, v = y, v + x + z = 3y\}$. So, if $\mathcal{B} = [(1, 2, 0, 1, 2)^T, (0, 0, 1, 0, -1)^T]$ is an ordered basis then $[(3, 6, 0, 3, 1)]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

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 $\mathcal{B}$ and $\mathcal{C}$, respectively. So, using the notation in Definition 3.3.5.2, we have

$$A = [\mathbf{v}_1, \ldots, \mathbf{v}_n] = [B[\mathbf{v}_1]_{\mathcal{C}}, \ldots, B[\mathbf{v}_n]_{\mathcal{C}}] = B\left[[\mathbf{v}_1]_{\mathcal{C}}, \ldots, [\mathbf{v}_n]_{\mathcal{C}}\right].$$

So, the matrix $[A]_{\mathcal{C}} = [[\mathbf{v}_1]_{\mathcal{C}}, \ldots, [\mathbf{v}_n]_{\mathcal{C}}]$, denoted $[\mathcal{B}]_{\mathcal{C}}$, is called the matrix of $\mathcal{B}$ **with respect to the ordered basis** $\mathcal{C}$ or the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{C}$. We now summarize the above discussion.

Theorem 3.3.5.4. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\dim(\mathbb{V}) = n$. Let $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ be two ordered bases of $\mathbb{V}$.*

1. *Then, the matrix $[\mathcal{B}]_{\mathcal{C}}$ is invertible and $[\mathbf{w}]_{\mathcal{C}} = [\mathcal{B}]_{\mathcal{C}}[\mathbf{w}]_{\mathcal{B}}$, for all $\mathbf{w} \in \mathbb{V}$.*

2. *Similarly, verify that $[\mathcal{C}]_{\mathcal{B}}$ is invertible and $[\mathbf{w}]_{\mathcal{B}} = [\mathcal{C}]_{\mathcal{B}}[\mathbf{w}]_{\mathcal{C}}$, for all $\mathbf{w} \in \mathbb{V}$.*

3. *Furthermore, $([\mathcal{B}]_{\mathcal{C}})^{-1} = [\mathcal{C}]_{\mathcal{B}}$.*

Proof. **Part 1:** We prove all the parts together. Let $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n], B = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ and $C = [\mathcal{B}]_{\mathcal{C}}$ and $D = [\mathcal{C}]_{\mathcal{B}}$. Then, by previous paragraph $A = BC$. Similarly,

$$B = [\mathbf{u}_1, \ldots, \mathbf{u}_n] = [A[\mathbf{u}_1]_{\mathcal{B}}, \ldots, A[\mathbf{u}_n]_{\mathcal{B}}] = A\left[[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_n]_{\mathcal{B}}\right] = AD.$$

But, by Exercise 3.3.3.16.14, A and B are invertible and thus $C = B^{-1}A$ and $D = A^{-1}B$ are invertible as well. Clearly, $C^{-1} = \left(B^{-1}A\right)^{-1} = A^{-1}B = D$ which proves the third part. For the first two parts, note that for any $\mathbf{w} \in \mathbb{V}$, $\mathbf{w} = A[\mathbf{w}]_{\mathcal{B}}$, $\mathbf{w} = B[\mathbf{w}]_{\mathcal{C}}$. Hence,

$$B[\mathbf{w}]_{\mathcal{C}} = \mathbf{w} = A[\mathbf{w}]_{\mathcal{B}} = BC[\mathbf{w}]_{\mathcal{B}} = B[\mathcal{B}]_{\mathcal{C}}[\mathbf{w}]_{\mathcal{B}}$$

and thus $[\mathbf{w}]_{\mathcal{C}} = [\mathcal{B}]_{\mathcal{C}}[\mathbf{w}]_{\mathcal{B}}$. Similarly, $[\mathbf{w}]_{\mathcal{B}} = [\mathcal{C}]_{\mathcal{B}}[\mathbf{w}]_{\mathcal{C}}$ and the required result follows. ∎

Example 3.3.5.5.

1. Note that if $\mathcal{C} = [\mathbf{e}_1, \ldots, \mathbf{e}_n]$ then

$$[A]_{\mathcal{C}} = [[\mathbf{v}_1]_{\mathcal{C}}, \ldots, [\mathbf{v}_n]_{\mathcal{C}}] = [\mathbf{v}_1, \ldots, \mathbf{v}_n] = A.$$

2. Suppose $\mathcal{B} = \left[(1, 0, 0)^T, (1, 1, 0)^T, (1, 1, 1)^T\right]$ and $\mathcal{C} = \left[(1, 1, 1)^T, (1, -1, 1)^T, (1, 1, 0)^T\right]$ are two bases of $\mathbb{R}^3$. Then, verify the statements in the previous result.

(a) Then $\begin{bmatrix} x\\ y\\ z\end{bmatrix} = \begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}\begin{bmatrix} x\\ y\\ z\end{bmatrix}_{\mathcal{B}}$. Thus, $\begin{bmatrix} x\\ y\\ z\end{bmatrix}_{\mathcal{B}} = \left(\begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}\right)^{-1}\begin{bmatrix} x\\ y\\ z\end{bmatrix} = \begin{bmatrix} x-y\\ y-z\\ z\end{bmatrix}$.

(b) Similarly, $\begin{bmatrix} x\\ y\\ z\end{bmatrix}_{\mathcal{C}} = \begin{bmatrix}1&1&1\\1&-1&1\\1&1&0\end{bmatrix}^{-1}\begin{bmatrix} x\\ y\\ z\end{bmatrix} = \dfrac{1}{2}\begin{bmatrix}-1&1&2\\1&-1&0\\2&0&-2\end{bmatrix}\begin{bmatrix} x\\ y\\ z\end{bmatrix} = \dfrac{1}{2}\begin{bmatrix}-x+y+2z\\ x-y+z\\ 2x-2z\end{bmatrix}$.

(c) Verify that $[\mathcal{B}]_{\mathcal{C}} = \begin{bmatrix}-1/2&0&1\\1/2&0&0\\1&1&0\end{bmatrix}$ and $[\mathcal{C}]_{\mathcal{B}} = \begin{bmatrix}0&2&0\\0&-2&1\\1&1&0\end{bmatrix}$.

Remark 3.3.5.6. *Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with $\mathcal{B} = [\mathbf{v}_1,\ldots,\mathbf{v}_n]$ as an ordered basis. Then, by Theorem 3.3.5.4, $\mathbf{v}_{\mathcal{B}}$ is an element of $\mathbb{F}^n$, for each $\mathbf{v} \in \mathbb{V}$. Therefore, if*

1. *$\mathbb{F} = \mathbb{R}$ then the elements of $\mathbb{V}$ look like the elements of $\mathbb{R}^n$.*
2. *$\mathbb{F} = \mathbb{C}$ then the elements of $\mathbb{V}$ look like the elements of $\mathbb{C}^n$.*

EXERCISE **3.3.5.7.** *Let $\mathcal{B} = \left[(1,2,0)^T,(1,3,2)^T,(0,1,3)^T\right]$ and $\mathcal{C} = \left[(1,2,1)^T,(0,1,2)^T,(1,4,6)^T\right]$ be two ordered bases of $\mathbb{R}^3$. Find the change of basis matrix*

1. *P from $\mathcal{B}$ to $\mathcal{C}$.*
2. *Q from $\mathcal{C}$ to $\mathcal{B}$.*
3. *from the standard basis of $\mathbb{R}^3$ to $\mathcal{B}$. What do you notice?*

Is it true that $PQ = I = QP$? Give reasons for your answer.

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 axioms in Definition 3.3.1.1.

If all axioms in Definition 3.3.1.1 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 unknowns $c_1,\ldots,c_n$. 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, where as 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. $\text{Col}(A) = \mathbb{R}^n$.
12. Columns of A form a basis of $\mathbb{R}^n$.
13. $\text{Null}(A) = \{\mathbf{0}\}$.

Chapter 4

Linear Transformations

4.1 Definitions and Basic Properties

In the previous chapter, it was shown that if $\mathbb{V}$ is a real vector space with $\dim(\mathbb{V}) = n$ then with respect to an ordered basis, the elements of $\mathbb{V}$ were column vectors of size n. So, in some sense the vector in $\mathbb{V}$ look like elements of $\mathbb{R}^n$. 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 4.4.1.1. [Linear Transformation, Linear Operator] Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. A function (map) $T : \mathbb{V} \to \mathbb{W}$ is called a **linear transformation** if for all $\alpha \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ the function T satisfies

$$T(\alpha \cdot \mathbf{u}) = \alpha \odot T(\mathbf{u}) \ \text{ and } \ T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) \oplus T(\mathbf{v}),$$

where $+, \cdot$ are binary operations in $\mathbb{V}$ 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 T is called a **linear operator** and the corresponding set of linear operators is denoted by $\mathcal{L}(\mathbb{V})$.

Definition 4.4.1.2. [Equality of Linear Transformation] Let $S, T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, S and T are said to be **equal** if $T(\mathbf{x}) = S(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{V}$.

We now give examples of linear transformations.

Example 4.4.1.3.

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 two 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 $T(x) = x$, for all $x \in \mathbb{R}$, is an element of $\mathcal{L}(\mathbb{R})$ as $T(ax) = ax = aT(x)$ and $T(x+y) = x+y = T(x)+T(y)$.

4. The map $T(x) = (x, 3x)^T$, for all $x \in \mathbb{R}$, is an element of $\mathcal{L}(\mathbb{R}, \mathbb{R}^2)$ as $T(\lambda x) = (\lambda x, 3\lambda x)^T = \lambda(x, 3x)^T = \lambda T(x)$ and $T(x+y) = (x+y, 3(x+y)^T = (x, 3x)^T + (y, 3y)^T = T(x) + T(y)$.

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 $S \circ T \in \mathcal{L}(\mathbb{V}, \mathbb{Z})$, where $(S \circ T)(\mathbf{v}) = S\big(T(\mathbf{v})\big)$, for all $\mathbf{v} \in \mathbb{V}$, is called the **composition** of maps. The readers should verify that $S \circ T$, in short ST, is an element of $\mathcal{L}(\mathbb{V}, \mathbb{Z})$.

6. Fix $\mathbf{a} \in \mathbb{R}^n$ and define $T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{a} \rangle$, for all $\mathbf{x} \in \mathbb{R}^n$. Then $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$. For example, if
   (a) $\mathbf{a} = (1, \ldots, 1)^T$ then $T(\mathbf{x}) = \sum\limits_{i=1}^{n} \mathbf{x}_i$, for all $\mathbf{x} \in \mathbb{R}^n$.
   (b) $\mathbf{a} = \mathbf{e}_i$, for a fixed i, $1 \le i \le n$, then $T_i(\mathbf{x}) = x_i$, for all $\mathbf{x} \in \mathbb{R}^n$.

7. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T\left((x, y)^T\right) = (x+y, 2x-y, x+3y)^T$. Then $T \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^3)$ with $T(\mathbf{e}_1) = (1, 2, 1)^T$ and $T(\mathbf{e}_2) = (1, -1, 3)^T$.

8. Let $A \in M_{m \times n}(\mathbb{C})$. Define $T_A(\mathbf{x}) = A\mathbf{x}$, for every $\mathbf{x} \in \mathbb{C}^n$. Then, $T_A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$. Thus, for each $A \in \mathbb{M}_{m,n}(\mathbb{C})$, there exists a map $T_A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$.

9. Define $T : \mathbb{R}^{n+1} \to \mathbb{R}[x; n]$ by $T\left((a_1, \ldots, a_{n+1})^T\right) = a_1 + a_2 x + \cdots + a_{n+1} x^n$, for $(a_1, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. Then T is a linear transformation.

10. Fix $A \in M_n(\mathbb{C})$. Then $T_A : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ and $S_A : M_n(\mathbb{C}) \to \mathbb{C}$ are both linear transformations, where $T_A(B) = BA^*$ and $S_A(B) = \mathrm{T{\scriptsize R}}(BA^*)$, for every $B \in M_n(\mathbb{C})$.

11. The map $T : \mathbb{R}[x; n] \to \mathbb{R}[x; n]$ defined by $T(f(x)) = \frac{d}{dx} f(x) = f'(x)$, for all $f(x) \in \mathbb{R}[x; n]$ is 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 $TS = \mathrm{Id}$? What about ST?

13. Recall the vector space $\mathbb{R}^{\mathbb{N}}$ in Example 3.3.1.4.8. 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 3.3.1.4.10). Then, the map $g = T(f)$, for each $f \in \mathcal{C}(\mathbb{R}, \mathbb{R})$, defined by $g(x) = \int\limits_0^x f(t)dt$ is an element of $\mathcal{L}(\mathcal{C}(\mathbb{R}, \mathbb{R}))$. For example, $(T(\sin))(x) = \int\limits_0^x \sin(t)dt = 1 - \cos(x)$. So, $(T(\sin))(x) = 1 - \cos(x)$, for all $x \in \mathbb{R}$.

We now prove that any linear transformation sends the zero vector to a zero vector.

Proposition 4.4.1.4. *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 have $T(\mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}} + \mathbf{0}_{\mathbb{V}}) = T(\mathbf{0}_{\mathbb{V}}) + T(\mathbf{0}_{\mathbb{V}})$. Thus, $T(\mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{W}}$ as $T(\mathbf{0}_{\mathbb{V}}) \in \mathbb{W}$. ■

From now on $\mathbf{0}$ will be used as the zero vector of the domain and codomain. The next result states that a linear transformation is known if we know its image on a basis of the domain space.

Example 4.4.1.5. Does there exist a linear transformation

1. $T : \mathbb{V} \to \mathbb{W}$ 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 4.4.1.4).
2. $T : \mathbb{R} \to \mathbb{R}$ 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. $T : \mathbb{R} \to \mathbb{R}$ 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. $T : \mathbb{R} \to \mathbb{R}$ such that $T(x) = \sin(x)$, for all $x \in \mathbb{R}$?
   Solution: No, as $T(ax) \neq aT(x)$.
5. $T : \mathbb{R} \to \mathbb{R}$ 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. $T : \mathbb{R} \to \mathbb{R}$ such that $T(5) = \pi$ and $T(e) = \pi$?
   Solution: No, as $5T(1) = T(5) = \pi$ implies that $T(1) = \dfrac{\pi}{5}$. So, $T(e) = eT(1) = \dfrac{e\pi}{5}$.
7. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((x,y)^T) = (x+y, 2)^T$?
   Solution: No, as $T(\mathbf{0}) \neq \mathbf{0}$.
8. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((x,y)^T) = (x+y, xy)^T$?
   Solution: No, as $T((2,2)^T) = (4,4)^T \neq 2(2,1)^T = 2T((1,1)^T)$.

Theorem 4.4.1.6. *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then T is known if the image of T on basis vectors of $\mathbb{V}$ are known.*

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of $\mathbb{V}$ and $\mathbf{v} \in \mathbb{V}$. Then, there exist $c_1, \ldots, c_n \in \mathbb{F}$ such that $\mathbf{v} = \sum\limits_{i=1}^{n} \mathbf{u}_i c_i = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. Hence, by definition

$$T(\mathbf{v}) = T(c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n) = c_1T(\mathbf{u}_1) + \cdots + c_nT(\mathbf{u}_n) = [T(\mathbf{u}_1), \ldots, T(\mathbf{u}_n)] \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Thus, the required result follows. ■

As a direct application, we have the following result.

Corollary 4.4.1.7 (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 Theorem 4.4.1.6, T is known if we know the image of T on $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, the standard basis of $\mathbb{R}^n$. As T is given, for $1 \le i \le n$, $T(\mathbf{e}_i) = a_i$, for some $a_i \in \mathbb{R}$. So, let us take $\mathbf{a} = (a_1, \ldots, a_n)^T$. Then, for $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$,

$$T(\mathbf{x}) = T\left(\sum_{i=1}^{n} x_i \mathbf{e}_i\right) = \sum_{i=1}^{n} x_i T(\mathbf{e}_i) = \sum_{i=1}^{n} x_i a_i = \mathbf{a}^T\mathbf{x}.$$

Thus, the required result follows. ∎

Example 4.4.1.8. Does there exist a linear transformation

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 $\mathbb{R}^2$, the matrix $\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$ is invertible. Also, $T\left(\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}\right) = T\left(a\begin{bmatrix}1\\ 1\end{bmatrix} + b\begin{bmatrix}1\\ -1\end{bmatrix}\right) = a\begin{bmatrix}1\\ 2\end{bmatrix} + b\begin{bmatrix}5\\ 10\end{bmatrix} = \begin{bmatrix}1 & 5\\ 2 & 10\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}$. So,

$$\begin{aligned} T\left(\begin{bmatrix}x\\ y\end{bmatrix}\right) &= T\left(\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\left(\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix}\right)\right) = \begin{bmatrix}1 & 5\\ 2 & 10\end{bmatrix}\left(\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix}\right) \\ &= \begin{bmatrix}1 & 5\\ 2 & 10\end{bmatrix}\begin{bmatrix}\frac{x+y}{2}\\ \frac{x-y}{2}\end{bmatrix} = \begin{bmatrix}3x-2y\\ 6x-4y\end{bmatrix}. \end{aligned}$$

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, let $\{(1,1)^T, \mathbf{u}\}$ be a basis of $\mathbb{R}^2$ and define $T(\mathbf{u}) = \mathbf{v} = (v_1, v_2)^T$, for some $\mathbf{v} \in \mathbb{R}^2$. For example, if $\mathbf{u} = (1,0)^T$ then

$$T\left(\begin{bmatrix}x\\ y\end{bmatrix}\right) = T\left(\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\left(\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix}\right)\right) = \begin{bmatrix}1 & v_1\\ 2 & v_2\end{bmatrix}\left(\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix}\right) = y\begin{bmatrix}1\\ 2\end{bmatrix} + (x-y)\mathbf{v}.$$

3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ 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 $T(\mathbf{e}_2) = \mathbf{0}$ or $T(\mathbf{e}_1) = (1,\pi)^T$ and $T(\mathbf{e}_2) = a(1,\pi)^T)$, for some $a \in \mathbb{R}$.

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Rng}(T) = \{T(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^2\} = \mathbb{R}^2$?
   Solution: Yes. Define $T(\mathbf{e}_1) = (1,\pi)^T$ and $T(\mathbf{e}_2) = (\pi, e)^T$. Or, let $\{\mathbf{u}, \mathbf{v}\}$ be a basis of $\mathbb{R}^2$ and define $T(\mathbf{e}_1) = \mathbf{u}$ and $T(\mathbf{e}_2) = \mathbf{v}$.

5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ 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 $T(\mathbf{e}_2) = \mathbf{0}$.

6. $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\textsc{Null}(T) = \{\mathbf{x} \in \mathbb{R}^2 \mid T(\mathbf{x}) = \mathbf{0}\} = LS\{(1,\pi)^T\}$?
   Solution: Yes. Let a basis of $\mathbb{R}^2 = \{(1,\pi)^T, (1,0)^T\}$ and define $T((1,\pi)^T) = \mathbf{0}$ and $T((1,0)^T) = \mathbf{u} \neq \mathbf{0}$.

EXERCISE **4.4.1.9.** *1. 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 $\mathbf{a} = \mathbf{0}$.*

2. *Are the maps $T : \mathbb{V} \to \mathbb{W}$ given below, linear transformations?*

   (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?*

   (a) $T(A) = A^T$ (b) $T(A) = I + A$ (c) $T(A) = A^2$

   (d) $T(A) = BAB^{-1}$, *where B is a fixed 2×2 matrix.*

4. *Prove that a map $T : \mathbb{R} \to \mathbb{R}$ is a linear transformation if and only if there exists a unique $c \in \mathbb{R}$ such that $T(\mathbf{x}) = c\mathbf{x}$, for every $\mathbf{x} \in \mathbb{R}$.*

5. *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 $\mathbf{x} \in \mathbb{C}^n$. Prove that for any positive integer k, $T_A^k(\mathbf{x}) = A^k\mathbf{x}$.*

6. *Use matrices to give examples of linear operators $T, S : \mathbb{R}^3 \to \mathbb{R}^3$ that satisfy:*

   (a) $T \neq \mathbf{0}$, $T^2 \neq \mathbf{0}$, $T^3 = \mathbf{0}$.

   (b) $T \neq \mathbf{0}$, $S \neq \mathbf{0}$, $S \circ T \neq \mathbf{0}$, $T \circ S = \mathbf{0}$.

   (c) $S^2 = T^2$, $S \neq T$.

   (d) $T^2 = I$, $T \neq I$.

7. *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.*

8. *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.*

9. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with $T(\mathbf{x}_0) = \mathbf{y}_0$ for some $\mathbf{x}_0 \in \mathbb{R}^n$ and $\mathbf{y}_0 \in \mathbb{R}^m$. Define $T^{-1}(\mathbf{y}_0) = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{y}_0\}$. Then prove that for every $\mathbf{x} \in T^{-1}(\mathbf{y}_0)$ there exists $\mathbf{z} \in T^{-1}(\mathbf{0})$ such that $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$. Also, prove that $T^{-1}(\mathbf{y}_0)$ is a subspace of $\mathbb{R}^n$ if and only if $\mathbf{0} \in T^{-1}(\mathbf{y}_0)$.*

10. *Define a map $T : \mathbb{C} \to \mathbb{C}$ by $T(z) = \overline{z}$, the complex conjugate of z. Is T a linear transformation over the real vector space $\mathbb{C}$?*

11. *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 $T((-1, 1, 2)^T) = (1, 0)^T$?*

12. *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 $T((1,1,2)^T) = (2,3)^T$?

13. 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 $T(\mathbf{x}) = (9,3,k)^T$.

14. 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 $T(\mathbf{x}) = (1,1,-1)^T$.

15. 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?

16. 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?

17. 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 $c, d \in \mathbb{R}$?

18. 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) $A \in \left\{ \frac{1}{2}\begin{bmatrix} \sqrt{3} & 1 \\ 1 & -\sqrt{3} \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \begin{bmatrix} \cos\left(\frac{2\pi}{3}\right) & \sin\left(\frac{2\pi}{3}\right) \\ \sin\left(\frac{2\pi}{3}\right) & -\cos\left(\frac{2\pi}{3}\right) \end{bmatrix} \right\}$.

19. 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 $(x,y) \in \mathbb{R}^2$.

20. 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 $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$ form a basis of $\mathbb{C}^3$.

4.2 Rank-Nullity Theorem

The readers are advised to see Theorem 3.3.4.9 on Page 81 for clarity and similarity with the results in this section. We start with the following result.

Theorem 4.4.2.1. *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If $S \subseteq \mathbb{V}$ is linearly dependent then $T(S) = \{T(\mathbf{v}) \mid \mathbf{v} \in \mathbb{V}\}$ is linearly dependent.*

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 unknown x_i's, has a non-trivial solution, say $x_i = a_i \in \mathbb{F}, 1 \le i \le k$. Thus, $\sum\limits_{i=1}^{k} a_i \mathbf{v}_i = \mathbf{0}$. Now, consider the system $\sum\limits_{i=1}^{k} y_i T(\mathbf{v}_i) = \mathbf{0}$, in the unknown y_i's. Then,

$$\sum_{i=1}^{k} a_i T(\mathbf{v}_i) = \sum_{i=1}^{k} T(a_i \mathbf{v}_i) = T\left(\sum_{i=1}^{k} a_i \mathbf{v}_i\right) = T(\mathbf{0}) = \mathbf{0}.$$

Thus, a_i's give a non-trivial solution of $\sum\limits_{i=1}^{k} y_i T(\mathbf{v}_i) = \mathbf{0}$ and hence the required result follows. ∎

As an immediate corollary, we get the following result.

Remark 4.4.2.2. *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector space over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Suppose $S \subseteq \mathbb{V}$ such that $T(S)$ is linearly independent then S is linearly independent.*

We now give some important definitions.

Definition 4.4.2.3. [Range Space and Null Space] Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, we define

1. $\mathrm{RNG}(T) = \{T(\mathbf{x}) \mid \mathbf{x} \in \mathbb{V}\}$ and call it the **range space** of T and
2. $\mathrm{NULL}(T) = \{\mathbf{x} \in \mathbb{V} \mid T(\mathbf{x}) = \mathbf{0}\}$ and call it the **null space** of T.

Example 4.4.2.4. Determine $\mathrm{RNG}(T)$ and $\mathrm{NULL}(T)$ of $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^4)$, where we define $T((x,y,z)^T) = (x-y+z, y-z, x, 2x-5y+5z)^T$.

Solution: Consider the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of $\mathbb{R}^3$. Then

$$\begin{aligned}
\mathrm{RNG}(T) &= LS(T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)) = LS\big((1,0,1,2)^T, (-1,1,0,-5)^T, (1,-1,0,5)^T\big) \\
&= LS\big((1,0,1,2)^T, (1,-1,0,5)^T\big) = \{\lambda(1,0,1,2)^T + \beta(1,-1,0,5)^T \mid \lambda, \beta \in \mathbb{R}\} \\
&= \{(\lambda+\beta, -\beta, \lambda, 2\lambda+5\beta) \ : \lambda, \beta \in \mathbb{R}\} \\
&= \{(x,y,z,w)^T \in \mathbb{R}^4 \mid x+y-z=0, 5y-2z+w=0\}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{NULL}(T) &= \{(x,y,z)^T \in \mathbb{R}^3 \ : T((x,y,z)^T) = \mathbf{0}\} \\
&= \{(x,y,z)^T \in \mathbb{R}^3 \ : (x-y+z, y-z, x, 2x-5y+5z)^T = \mathbf{0}\} \\
&= \{(x,y,z)^T \in \mathbb{R}^3 \ : x-y+z=0, y-z=0, x=0, 2x-5y+5z=0\} \\
&= \{(x,y,z)^T \in \mathbb{R}^3 \ : y-z=0, x=0\} \\
&= \{(0,y,y)^T \in \mathbb{R}^3 \ : y \in \mathbb{R}\} = LS((0,1,1)^T)
\end{aligned}$$

EXERCISE **4.4.2.5.** *1. Let $\mathbb{V}$ and $\mathbb{W}$ be two 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) $\text{NULL}(T)$ *is a subspace of* $\mathbb{V}$.

Furthermore, if $\mathbb{V}$ *is finite dimensional then*

(a) $\dim(\text{NULL}(T)) \le \dim(\mathbb{V})$.

(b) $\dim(\text{RNG}(T))$ *is finite and whenever* $\dim(\mathbb{W})$ *is finite* $\dim(\text{RNG}(T)) \le \dim(\mathbb{W})$.

2. *Describe* $\text{NULL}(D)$ *and* $\text{RNG}(D)$*, where* $D \in \mathcal{L}(\mathbb{R}[x;n])$ *and is defined by* $D(f(x)) = f'(x)$*. Note that* $\text{RNG}(D) \subseteq \mathbb{R}[x;n-1]$.

3. *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* $T^3 = T$*?*

4. *Find* $T \in \mathcal{L}(\mathbb{R}^3)$ *for which* $\text{RNG}(T) = LS\big((1,2,0)^T, (0,1,1)^T, (1,3,1)^T\big)$.

5. *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be two vector spaces over* $\mathbb{F}$*. If* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *is a* $\mathbb{V}$ *and* $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbb{W}$ *then prove that there exists a unique* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ *such that* $T(\mathbf{v}_i) = \mathbf{w}_i$*, for* $i = 1, \ldots, n$.

Definition 4.4.2.6. [Rank and Nullity] Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$. If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $\dim(\mathbb{V})$ is finite then we define $\text{RANK}(T) = \dim(\text{RNG}(T))$ and $\text{NULLITY}(T) = \dim(\text{NULL}(T))$.

We now prove the rank-nullity Theorem. The proof of this result is similar to the proof of Theorem 3.3.4.9. We give it again for the sake of completeness.

Theorem 4.4.2.7 (Rank-Nullity Theorem)**.** *Let* $\mathbb{V}$ *and* $\mathbb{W}$ *be two vector spaces over* $\mathbb{F}$*. If* $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ *and* $\dim(\mathbb{V})$ *is finite then*

$$\text{RANK}(T) + \text{NULLITY}(T) = \dim(\text{RNG}(T)) + \dim(\text{NULL}(T)) = \dim(\mathbb{V}).$$

Proof. By Exercise 4.4.2.5.1.1a, $\dim(\text{NULL}(T)) \le \dim(\mathbb{V})$. Let $\mathcal{B}$ be a basis of $\text{NULL}(T)$. We extend it to form a basis $\mathcal{C}$ of $\mathbb{V}$. So, by definition $\text{RNG}(T) = LS(\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C}\}) = LS(\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C} \setminus \mathcal{B}\})$. We claim that $\{T(\mathbf{v}) | \mathbf{v} \in \mathcal{C} \setminus \mathcal{B}\}$ is linearly independent subset of $\mathbb{W}$.

Let if possible the claim be false. Then, 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 $\sum\limits_{i=1}^{k} a_i T(\mathbf{v}_i) = \mathbf{0}$. Thus, we see that

$$T\left(\sum_{i=1}^{k} a_i \mathbf{v}_i\right) = \sum_{i=1}^{k} a_i T(\mathbf{v}_i) = \mathbf{0}.$$

That is, $\sum\limits_{i=1}^{k} a_i \mathbf{v}_i \in \text{NULL}(T)$. Hence, there exists $b_1, \ldots, b_\ell \in \mathbb{F}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_\ell \in \mathcal{B}$ such that $\sum\limits_{i=1}^{k} a_i \mathbf{v}_i = \sum\limits_{j=1}^{k} b_j \mathbf{u}_j$. Or equivalently, the system $\sum\limits_{i=1}^{k} x_i \mathbf{v}_i + \sum\limits_{j=1}^{k} y_j \mathbf{u}_j = \mathbf{0}$, 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 in $\mathbb{V}$. A contradiction to $S \subseteq \mathcal{C}$. That is,

$$\dim(\textsc{Rng}(T)) + \dim(\textsc{Null}(T)) = |\mathcal{C} \setminus \mathcal{B}| + |\mathcal{B}| = |\mathcal{C}| = \dim(\mathbb{V}).$$

Thus, we have proved the required result. ■

As an immediate corollary, we have the following result.

Corollary 4.4.2.8. *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. $\min\{\textsc{Rank}(S), \textsc{Rank}(T)\} \geq \textsc{Rank}(ST) \geq n - \textsc{Rank}(S) - \textsc{Rank}(T)$.

Proof. The prove of Part 2 is omitted as it directly follows from Part 1 and Theorem 4.4.2.7. Part 2: We first prove the second inequality. Suppose $\mathbf{v} \in \textsc{Null}(T)$. Then $(ST)(\mathbf{v}) = S(T(\mathbf{v}) = S(\mathbf{0}) = \mathbf{0}$ gives $\textsc{Null}(T) \subseteq \textsc{Null}(ST)$. Thus, $\textsc{Nullity}(T) \leq \textsc{Nullity}(ST)$.

By Theorem 4.4.2.7, $\textsc{Nullity}(S) \leq \textsc{Nullity}(ST)$ is equivalent to $\textsc{Rng}(ST) \subseteq \textsc{Rng}(S)$. And 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 $\textsc{Null}(T)$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq \textsc{Null}(ST)$. So, let us extend it to get a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ of $\textsc{Null}(ST)$.

Claim: $\{T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_\ell)\}$ is a linearly independent subset of $\textsc{Null}(S)$.

Clearly, $\{T(\mathbf{u}_1), \ldots, T(\mathbf{u}_\ell)\} \subseteq \textsc{Null}(S)$. Now, consider the system $c_1T(\mathbf{u}_1) + \cdots + c_\ell T(\mathbf{u}_\ell) = \mathbf{0}$ in the unknowns $c_1, \ldots, c_\ell$. As $T \in \mathcal{L}(\mathbb{V})$, we get $T\left(\sum\limits_{i=1}^{\ell} c_i\mathbf{u}_i\right) = \mathbf{0}$. Thus, $\sum\limits_{i=1}^{\ell} c_i\mathbf{u}_i \in \textsc{Null}(T)$. Hence, $\sum\limits_{i=1}^{\ell} c_i\mathbf{u}_i$ is a unique linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Therefore,

$$c_1\mathbf{u}_1 + \cdots + c_\ell\mathbf{u}_\ell = \alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k \tag{4.4.2.1}$$

for some scalars $\alpha_1, \ldots, \alpha_k$. But by assumption, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ is a basis of $\textsc{Null}(ST)$ and hence linearly independent. Therefore, the only solution of Equation (4.4.2.1) is given by $c_i = 0$ for $1 \leq i \leq \ell$ and $\alpha_j = 0$ for $1 \leq j \leq k$. Thus, we have proved the claim. Hence, $\textsc{Nullity}(S) \geq \ell$ and $\textsc{Nullity}(ST) = k + \ell \leq \textsc{Nullity}(T) + \textsc{Nullity}(S)$. ■

Exercise **4.4.2.9.**

1. *Let $A \in M_n(\mathbb{R})$ with $A^2 = A$. Define $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ by $T(\mathbf{v}) = A\mathbf{v}$, for all $\mathbf{v} \in \mathbb{R}^n$. Then prove that*
   (a) *$T^2 = T$. Equivalently, $T(Id - T) = \mathbf{0}$.*
   (b) $\textsc{Null}(T) \cap \textsc{Rng}(T) = \{\mathbf{0}\}$.
   (c) $\mathbb{R}^n = \textsc{Rng}(T) + \textsc{Null}(T)$. *[Hint: $\mathbf{x} = T(\mathbf{x}) + (Id - T)(\mathbf{x})$]*
2. *Define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x - y + z \\ x + 2z \end{bmatrix}$. Find a basis and the dimension of* $\textsc{Rng}(T)$ *and* $\textsc{Null}(T)$.
3. *Let $z_i \in \mathbb{C}$, for $1 \leq i \leq k$. Define $T \in \mathcal{L}(\mathbb{C}[x; n], \mathbb{C}^k)$ by $T\big(P(z)\big) = \big(P(z_1), \ldots, P(z_k)\big)$. If z_i's are distinct then for each $k \geq 1$, determine* $\textsc{Rank}(T)$.

4.2.A Algebra of Linear Transformation

We start with the following definition.

Definition 4.4.2.10. [Sum and Scalar Multiplication of Linear Transformations] 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 4.4.2.11. *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 for $S, T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$, if we define $(S+\alpha T)(\mathbf{v}) = S(\mathbf{v}) + \alpha T(\mathbf{v})$ (point-wise addition and scalar multiplication) then $\mathcal{L}(\mathbb{V}, \mathbb{W})$ 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

$$\mathbf{f}_{ij}(\mathbf{v}_k) = \begin{cases} \mathbf{w}_j, & \text{if } k = i \\ \mathbf{0}, & k \neq i. \end{cases}$$

For other vectors of $\mathbb{V}$, we extend the definition by linearity. That is, if $\mathbf{v} = \sum\limits_{s=1}^{n} \alpha_s \mathbf{v}_s$ then

$$\mathbf{f}_{ij}(\mathbf{v}) = \mathbf{f}_{ij}\left(\sum_{s=1}^{n} \alpha_s \mathbf{v}_s\right) = \sum_{s=1}^{n} \alpha_s \mathbf{f}_{ij}(\mathbf{v}_s) = \alpha_i \mathbf{f}_{ij}(\mathbf{v}_i) = \alpha_i \mathbf{w}_j. \tag{4.4.2.2}$$

Thus, $\mathbf{f}_{ij} \in \mathcal{L}(\mathbb{V}, \mathbb{W})$.

Claim: $\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}$ is a basis of $\mathcal{L}(\mathbb{V}, \mathbb{W})$.

So, let us 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 k from 1 to 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, $f(\mathbf{v}_s) \in \mathbb{W}$ and hence there exists β_{st}'s such that $f(\mathbf{v}_s) = \sum\limits_{t=1}^{m} \beta_{st} \mathbf{w}_t$, for $1 \le s \le n$. So, if $\mathbf{v} = \sum\limits_{s=1}^{n} \alpha_s \mathbf{v}_s \in \mathbb{V}$ then, using Equation (4.4.2.2), we get

$$\begin{aligned} f(\mathbf{v}) &= f\left(\sum_{s=1}^{n} \alpha_s \mathbf{v}_s\right) = \sum_{s=1}^{n} \alpha_s f(\mathbf{v}_s) = \sum_{s=1}^{n} \alpha_s \left(\sum_{t=1}^{m} \beta_{st} \mathbf{w}_t\right) = \sum_{s=1}^{n} \sum_{t=1}^{m} \beta_{st} (\alpha_s \mathbf{w}_t) \\ &= \sum_{s=1}^{n} \sum_{t=1}^{m} \beta_{st} \mathbf{f}_{st}(\mathbf{v}_s) = \left(\sum_{s=1}^{n} \sum_{t=1}^{m} \beta_{st} \mathbf{f}_{st}\right)(\mathbf{v}). \end{aligned}$$

Since the above is true for every $\mathbf{v} \in \mathbb{V}$, $LS(\{\mathbf{f}_{ij} | 1 \le i \le n, 1 \le j \le m\}) = \mathcal{L}(\mathbb{V}, \mathbb{W})$ and thus the required result follows. ■

Definition 4.4.2.12. 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 4.4.2.13. *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}. The reader should prove that f is invertible if and only if f is both one-one and onto.*

Theorem 4.4.2.14. *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Also assume that T is one-one and onto. Then*

1. *for each $\mathbf{w} \in \mathbb{W}$, the set $|T^{-1}(\mathbf{w})| = 1$, where $T^{-1}(\mathbf{w}) = \{\mathbf{v} \in \mathbb{V} | T(\mathbf{v}) = \mathbf{w}\}$.*
2. *the map $T^{-1} \in \mathcal{L}(\mathbb{W}, \mathbb{V})$, where one defines $T^{-1}(\mathbf{w}) = \mathbf{v}$ whenever $T(\mathbf{v}) = \mathbf{w}$.*

Proof. PART 1. As T is onto, for each $\mathbf{w} \in \mathbb{W}$ there exists $\mathbf{v} \in \mathbb{V}$ such that $T(\mathbf{v}) = \mathbf{w}$. So, $T^{-1}(\mathbf{w}) \neq \emptyset$. Now, let us assume that there exist vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$ such that $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. Then T is one-one implies $\mathbf{v}_1 = \mathbf{v}_2$. Hence, $|T^{-1}(\mathbf{w})| = 1$. This completes the proof of Part 1.

PART 2. 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 Part 1, 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, by definition of T^{-1}, 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 proof of Part 2 is complete. ■

Definition 4.4.2.15. [Inverse Linear Transformation] Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If T is one-one and onto then $T^{-1} \in \mathcal{L}(\mathbb{W}, \mathbb{V})$, where $T^{-1}(\mathbf{w}) = \mathbf{v}$ whenever $T(\mathbf{v}) = \mathbf{w}$. The map T^{-1} is called the **inverse** of the linear transformation T.

Example 4.4.2.16.

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T((x, y)^T) = (x + y, x - y)^T$. Then $T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}$ as $(T \circ T^{-1})\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = T\left(T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)\right) = T\left(\begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$. Thus, the map T^{-1} is indeed the inverse of T.
2. Define $T \in \mathcal{L}(\mathbb{R}^{n+1}, \mathbb{R}[x; n])$ by $T(a_1, \ldots, a_{n+1}) = \sum\limits_{i=1}^{n+1} a_i x^{i-1}$, for $(a_1, \ldots, a_{n+1}) \in \mathbb{R}^{n+1}$. Then, one defines $T^{-1}\left(\sum\limits_{i=1}^{n+1} a_i x^{i-1}\right) = (a_1, \ldots, a_{n+1})$, for all $\sum\limits_{i=1}^{n+1} a_i x^{i-1} \in \mathbb{R}[x; n]$. Verify that $T^{-1} \in \mathcal{L}(\mathbb{R}[x; n], \mathbb{R}^{n+1})$.

Definition 4.4.2.17. Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, T is said to be **singular** if there exists $\mathbf{v} \in \mathbb{V}$ such that $\mathbf{v} \neq \mathbf{0}$ but $T(\mathbf{v}) = \mathbf{0}$. If such a $\mathbf{v} \in \mathbb{V}$ does not exist then T is called **non-singular**.

Example 4.4.2.18. 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 4.4.2.19. *Let $\mathbb{V}$ and $\mathbb{W}$ be two 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. Thus, the required result follows. ∎

Definition 4.4.2.20. Let $\mathbb{V}$ and $\mathbb{W}$ be two 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 3.3.5.6.

Theorem 4.4.2.21. *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. ∎

We now summarize the different definitions related with a linear operator on a finite dimensional vector space. The prove basically uses the rank-nullity theorem.

Theorem 4.4.2.22. *Let $\mathbb{V}$ be a 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. $\text{NULL}(T) = \{\mathbf{0}\}$.
3. $\text{RANK}(T) = n$.
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.*

Proof. $1 \Rightarrow 2$ Let $\mathbf{x} \in \text{NULL}(T)$. Then $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$. So, T is one-one implies $\mathbf{x} = \mathbf{0}$. Thus $\text{NULL}(T) = \{\mathbf{0}\}$.

$2 \Rightarrow 3$ As $\text{NULL}(T) = \{\mathbf{0}\}$, $\text{NULLITY}(T) = 0$ and hence by Theorem 4.4.2.7 $\text{RANK}(T) = n$.

$3 \Rightarrow 4$ As $\text{RANK}(T) = n$, $\text{RNG}(T) \subseteq \mathbb{V}$ and $\dim(\mathbb{V}) = n$, we get $\text{RNG}(T) = \mathbb{V}$. Thus T is onto.

$4 \Rightarrow 1$ As T is onto, $\dim(\text{RNG}(T)) = n$. So, by Theorem 4.4.2.7 $\text{NULL}(T) = \{\mathbf{0}\}$. Now, let $\mathbf{x}, \mathbf{y} \in \mathbb{V}$ such that $T(\mathbf{x}) = T(\mathbf{y})$. Or equivalently, $\mathbf{x} - \mathbf{y} \in \text{NULL}(T) = \{\mathbf{0}\}$. Thus $\mathbf{x} = \mathbf{y}$ and T is one-one.

The equivalence of 1 and 2 gives the equivalence with 5. Also, using Theorem 4.4.2.19, one has the equivalence of 1, 6 and 7. Further note that the equivalence of 1 and 2 with Theorem 4.4.2.14 implies that T is invertible. For the other way implication, note that by definition T is invertible implies that T is one-one and onto. Thus, all the statements are equivalent. ■

EXERCISE **4.4.2.23.** *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. If $\dim(\mathbb{V})$ is finite then prove that*

1. *T cannot be onto if $\dim(\mathbb{V}) < \dim(\mathbb{W})$.*
2. *T cannot be one-one if $\dim(\mathbb{V}) > \dim(\mathbb{W})$.*

4.3 Matrix of a linear transformation

In Example 4.4.1.3.8, 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 an $m \times n$ matrix. Before proceeding further, the readers should recall the results on ordered basis (see Section 3.5).

So, let $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. Then, recall that if $A = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $B = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ then $\mathbf{v} = A[\mathbf{v}]_{\mathcal{B}}$ and $\mathbf{w} = B[\mathbf{w}]_{\mathcal{C}}$, for all $\mathbf{v} \in \mathbb{V}$ and $\mathbf{w} \in \mathbb{W}$. So, if $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then, for any $\mathbf{v} \in \mathbb{V}$,

$$\begin{aligned} B[\mathbf{T}(\mathbf{v})]_{\mathcal{C}} &= T(\mathbf{v}) = T(A[\mathbf{v}]_{\mathcal{B}}) = T(A)[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)][\mathbf{v}]_{\mathcal{B}} \\ &= [B[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, B[T(\mathbf{v}_n)]_{\mathcal{C}}]\,[\mathbf{v}]_{\mathcal{B}} = B\,[[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{C}}]\,[\mathbf{v}]_{\mathcal{B}}. \end{aligned}$$

As B is invertible, we get $[\mathbf{T}(\mathbf{v})]_{\mathcal{C}} = [[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{C}}]\,[\mathbf{v}]_{\mathcal{B}}$. Note that the matrix $[[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{C}}]$, denoted $T[\mathcal{B}, \mathcal{C}]$, is an $m \times n$ matrix and is unique as the i-th column equals $[T(\mathbf{v}_i)]_{\mathcal{C}}$, for $1 \le i \le n$. So, we immediately have the following definition and result.

Definition 4.4.3.1. [Matrix of a Linear Transformation] Let $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then the matrix $T[\mathcal{B}, \mathcal{C}]$ is called the **coordinate matrix** of T or the **matrix of the linear transformation** T with respect to the basis $\mathcal{B}$ and $\mathcal{C}$, respectively. When there is no mention of bases, we take the standard bases and denote the matrix by $[T]$.

Theorem 4.4.3.2. *Let $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ be ordered bases of $\mathbb{V}$ and $\mathbb{W}$, respectively. If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then there exists a matrix $A \in M_{m \times n}(\mathbb{F})$ with*

$$A = T[\mathcal{B}, \mathcal{C}] = [[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{C}}] \text{ and } [T(\mathbf{x})]_{\mathcal{C}} = A \; [\mathbf{x}]_{\mathcal{B}}, \text{ for all } \mathbf{x} \in \mathbb{V}.$$

We now give a few examples to understand the above discussion and Theorem 4.4.3.2.

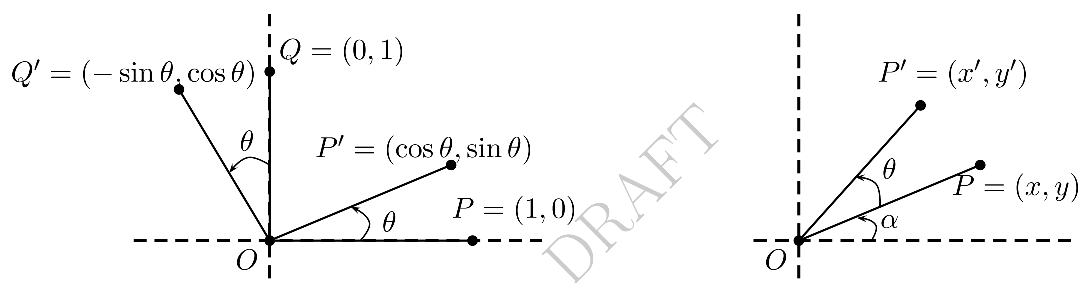


Figure 4.1: Counter-clockwise Rotation by an angle θ

Example 4.4.3.3. 1. Let $T \in \mathcal{L}(\mathbb{R}^2)$ represent a counterclockwise rotation by an angle $\theta, 0 \le \theta < 2\pi$. 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

$$[T] = \Big[T(\mathbf{e}_1), T(\mathbf{e}_2)\Big] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{4.4.3.1}$$

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 $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Then
   $$[T] = [[T(\mathbf{e}_1)]_{\mathcal{C}} \quad [T(\mathbf{e}_2)]_{\mathcal{C}}] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{C}} \quad \begin{bmatrix} 1 \\ -1 \end{bmatrix}_{\mathcal{C}} \right] = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}.$$
   (c) In the above, let the ordered basis of the domain space be $\mathcal{B} = \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$. Then
   $$T[\mathcal{B}, \mathcal{C}] = \left[\left[T\begin{bmatrix} -1 \\ 1 \end{bmatrix} \right]_{\mathcal{C}} \quad \left[T\begin{bmatrix} 3 \\ 1 \end{bmatrix} \right]_{\mathcal{C}} \right] = \left[\begin{bmatrix} 0 \\ -2 \end{bmatrix}_{\mathcal{C}} \quad \begin{bmatrix} 4 \\ 2 \end{bmatrix}_{\mathcal{C}} \right] = \begin{bmatrix} 2 & 2 \\ -2 & 2 \end{bmatrix}.$$

3. Let $\mathcal{B} = [\mathbf{e}_1, \mathbf{e}_2]$ and $\mathcal{C} = [\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_2]$ be two ordered bases of $\mathbb{R}^2$. Then Compute $T[\mathcal{B},\mathcal{B}]$ and $T[\mathcal{C},\mathcal{C}]$, 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 $B^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. So,

$$T[\mathcal{B},\mathcal{B}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{B}}, \left[T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}}\right] = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \text{ and}$$

$$T[\mathcal{C},\mathcal{C}] = \left[\left[T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{C}}, \left[T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\right]_{\mathcal{C}}\right] = \left[\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{C}}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{C}}\right] = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{3}{2} \end{bmatrix}$$

as $\begin{bmatrix} 2 \\ -1 \end{bmatrix}_{\mathcal{C}} = B^{-1}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 3 \end{bmatrix}_{\mathcal{C}} = B^{-1}\begin{bmatrix} 0 \\ 3 \end{bmatrix}$.

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 $[T]$.
Solution: 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 coordinates with respect to the ordered basis $\mathcal{B} = \left[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\right]$ and $\mathcal{C} = \left[(1,0,0), (1,1,0), (1,1,1)\right]$.
Solution: By definition, verify that

$$T[\mathcal{B},\mathcal{C}] = [[T(\mathbf{e}_1)]_{\mathcal{C}}, [T(\mathbf{e}_2)]_{\mathcal{C}}, [T(\mathbf{e}_3)]_{\mathcal{C}}] = \left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{C}}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{C}}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{C}}\right] = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$T[\mathcal{C},\mathcal{B}] = \left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}_{\mathcal{B}}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{B}}\right] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, verify that $T[\mathcal{C},\mathcal{B}]^{-1} = T[\mathcal{B},\mathcal{C}]$ and $T[\mathcal{B},\mathcal{B}] = T[\mathcal{C},\mathcal{C}] = I_3$ as the given map is indeed the identity map.

6. Fix $A \in \mathbb{M}_n(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n)$ by $T(\mathbf{x}) = A\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. If $\mathcal{B}$ is the standard basis of $\mathbb{C}^n$ then

$$[T][:, i] = [T(\mathbf{e}_i)]_{\mathcal{B}} = [A(\mathbf{e}_i)]_{\mathcal{B}} = [A[:, i]]_{\mathcal{B}} = A[:, i].$$

7. Fix $A \in \mathbb{M}_{m,n}(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ by $T(\mathbf{x}) = A\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. Let $\mathcal{B}$ and $\mathcal{C}$ be the standard ordered bases of $\mathbb{C}^n$ and $\mathbb{C}^m$, respectively. Then $T[\mathcal{B},\mathcal{C}] = A$ as

$$(T[\mathcal{B},\mathcal{C}])[:, i] = [T(\mathbf{e}_i)]_{\mathcal{C}} = [A(\mathbf{e}_i)]_{\mathcal{C}} = [A[:, i]]_{\mathcal{C}} = A[:, i].$$

8. Fix $A \in \mathbb{M}_n(\mathbb{C})$ and define $T \in \mathcal{L}(\mathbb{C}^n)$ by $T(\mathbf{x}) = A\mathbf{x}$, for all $\mathbf{x} \in \mathbb{C}^n$. Let $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ be two ordered basses of $\mathbb{C}^n$ with respective matrices B and C. Then

$$\begin{aligned} T[\mathcal{B}, \mathcal{C}] &= [[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_1)]_{\mathcal{C}}] = \left[C^{-1}T(\mathbf{v}_1), \ldots, C^{-1}T(\mathbf{v}_1)\right] \\ &= \left[C^{-1}A\mathbf{v}_1, \ldots, C^{-1}A\mathbf{v}_1\right] = C^{-1}A[\mathbf{v}_1, \ldots, \mathbf{v}_n] = C^{-1}AB. \end{aligned}$$

In particular, if

(a) $\mathcal{B} = \mathcal{C}$ then $T[\mathcal{B}, \mathcal{B}] = B^{-1}AB$.

(b) $A = I_n$ so that $T = \mathrm{Id}$ then $\mathrm{Id}[\mathcal{B}, \mathcal{C}] = C^{-1}B$, an invertible matrix. Similarly, $\mathrm{Id}[\mathcal{C}, \mathcal{B}] = B^{-1}C$. So, $\mathrm{Id}[\mathcal{C}, \mathcal{B}] \cdot \mathrm{Id}[\mathcal{B}, \mathcal{C}] = (B^{-1}C)(C^{-1}B) = I_n$.

(c) $A = I_n$ so that $T = \mathrm{Id}$ and $\mathcal{B} = \mathcal{C}$ then $\mathrm{Id}[\mathcal{B}, \mathcal{B}] = I_n$.

EXERCISE **4.4.3.4.** *1. Let $T \in \mathcal{L}(\mathbb{R}^2)$ represent the reflection about the line $y = mx$. Find its matrix with respect to the standard ordered basis of $\mathbb{R}^2$.*

2. Let $T \in \mathcal{L}(\mathbb{R}^3)$ represent the reflection about the X-axis. Find its matrix with respect to the standard ordered basis of $\mathbb{R}^3$.

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 $\mathrm{RNG}(D) \subseteq \mathbb{R}[x; n-1]$.

4.3.A Dual Space*

Definition 4.4.3.5. 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 4.4.3.6. The following linear transformations are linear functionals.

1. $T(A) = \mathrm{trace}(A)$ for $T \in \mathcal{L}(\mathbb{M}_n(\mathbb{R}), \mathbb{R})$.
2. $T(f) = \int\limits_a^b f(t)dt$ for $T \in \mathcal{L}(\mathcal{C}([a, b], \mathbb{R}), \mathbb{R})$.
3. $T(f) = \int\limits_a^b t^2 f(t)dt$ for $T \in \mathcal{L}(\mathcal{C}([a, b], \mathbb{R}), \mathbb{R})$.

EXERCISE **4.4.3.7.** *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 $\mathbf{v} = \mathbf{0}$.*

Definition 4.4.3.8. 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 4.4.2.21.

Corollary 4.4.3.9. *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 $\mathbf{f}_i(\mathbf{v}_k) = \begin{cases} 1, & \text{if } k = i \\ \mathbf{0}, & k \neq i. \end{cases}$.*

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 $L_{\mathbf{v}} \in \mathbb{V}^{**}$. We use it to give the required canonical isomorphism.

Theorem 4.4.3.10. *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 the map T satisfies the following:

1. For each $\mathbf{f} \in \mathbb{V}^*$, note that

   $$L_{\alpha\mathbf{v}+\mathbf{u}}(\mathbf{f}) = \mathbf{f}(\alpha\mathbf{v} + \mathbf{u}) = \alpha\mathbf{f}(\mathbf{v}) + \mathbf{f}(\mathbf{u}) = \alpha L_{\mathbf{v}}(\mathbf{f}) + L_{\mathbf{u}}(\mathbf{f}) = (\alpha L_{\mathbf{v}} + L_{\mathbf{u}})(\mathbf{f}).$$

   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.
2. We now show that T is one-one. So, suppose assume that $T(\mathbf{v}) = T(\mathbf{u})$, for some $\mathbf{u}, \mathbf{v} \in \mathbb{V}$. Then, $L_{\mathbf{v}} = L_{\mathbf{u}}$. That is, $L_{\mathbf{v}}(\mathbf{f}) = L_{\mathbf{u}}(\mathbf{f})$, for all $\mathbf{f} \in \mathbb{V}^*$. Or equivalently, $\mathbf{f}(\mathbf{v} - \mathbf{u}) = 0$, for all $\mathbf{f} \in \mathbb{V}^*$. Hence, by Exercise 4.4.3.7 $\mathbf{v} - \mathbf{u} = \mathbf{0}$. So, $\mathbf{v} = \mathbf{u}$. Therefore T is one-one.

Thus, T gives an inclusion map from $\mathbb{V}$ to $\mathbb{V}^{**}$. Further, applying Corollary 4.4.3.9.2 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 4.4.3.10.

Corollary 4.4.3.11. *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 $\alpha \in \mathbb{V}$ such that $T(\mathbf{f}) = \mathbf{f}(\alpha)$, for all $\mathbf{f} \in \mathbb{V}^*$.*
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 4.4.3.9.2) 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 4.4.3.9.2. ∎

Let $\mathbb{V}$ be a finite dimensional vector space. Then Corollary 4.4.3.11 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 $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then, we want to define a map $\widehat{T} : \mathbb{W}^* \to \mathbb{V}^*$. As $g \in \mathbb{W}^*$, $\widehat{T}(\mathbf{g}) \in \mathbb{V}^*$, a linear functional. So, we need to be evaluate at an element of $\mathbb{V}$. Thus, we define $\widehat{T}(\mathbf{g})(\mathbf{v}) = g(T(\mathbf{v}))$, for all $\mathbf{v} \in \mathbb{V}$. 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 4.4.3.12. *Let $\mathbb{V}$ and $\mathbb{W}$ be two vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{B} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\mathcal{C} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$, respectively. Also, let $\mathcal{B}^* = [\mathbf{f}_1, \ldots, \mathbf{f}_n]$ and $\mathcal{C}^* = [\mathbf{g}_1, \ldots, \mathbf{g}_m]$ be the corresponding ordered bases of the dual spaces $\mathbb{V}^*$ and $\mathbb{W}^*$, respectively. Then, $\widehat{T}[\mathcal{C}^*, \mathcal{B}^*] = (T[\mathcal{B}, \mathcal{C}])^T$, the transpose of the coordinate matrix T.*

Proof. Note that we need to compute $\widehat{T}[\mathcal{C}^*, \mathcal{B}^*] = \left[\left[\widehat{T}(\mathbf{g}_1)\right]_{\mathcal{B}^*}, \ldots, \left[\widehat{T}(\mathbf{g}_m)\right]_{\mathcal{B}^*}\right]$ and prove that it equals the transpose of the matrix $T[\mathcal{B}, \mathcal{C}]$. So, let $T[\mathcal{B}, \mathcal{C}] = [[T(\mathbf{v}_1)]_{\mathcal{C}}, \ldots, [T(\mathbf{v}_n)]_{\mathcal{C}}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. Thus, to prove the required result, we need to show that

$$\left[\widehat{T}(\mathbf{g}_j)\right]_{\mathcal{B}^*} = [\mathbf{f}_1, \ldots, \mathbf{f}_n] \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{bmatrix} = \sum_{k=1}^{n} a_{jk}\mathbf{f}_k, \text{ for } 1 \leq j \leq m. \tag{4.4.3.2}$$

Now, recall that the functionals $\mathbf{f}_i$'s and $\mathbf{g}_j$'s satisfy $\left(\sum\limits_{k=1}^{n} \alpha_k \mathbf{f}_k\right)(\mathbf{v}_t) = \sum\limits_{k=1}^{n} \alpha_k (\mathbf{f}_k(\mathbf{v}_t)) = \alpha_t$, for $1 \leq t \leq n$ 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 $C = [\mathbf{w}_1, \ldots, \mathbf{w}_m]$ and evaluate $\widehat{T}(\mathbf{g}_j)$ at $\mathbf{v}_t$'s, the elements of $\mathcal{B}$.

$$\begin{aligned} \left(\widehat{T}(\mathbf{g}_j)\right)(\mathbf{v}_t) &= \mathbf{g}_j(T(\mathbf{v}_t)) = \mathbf{g}_j(C[T(\mathbf{v}_t)]_{\mathcal{C}}) = [\mathbf{g}_j(\mathbf{w}_1), \ldots, \mathbf{g}_j(\mathbf{w}_m)][T(\mathbf{v}_t)]_{\mathcal{C}} \\ &= \mathbf{e}_j^T \begin{bmatrix} a_{1t} \\ a_{2t} \\ \vdots \\ a_{mt} \end{bmatrix} = a_{jt} = \left(\sum_{k=1}^{n} a_{jk}\mathbf{f}_k\right)(\mathbf{v}_t). \end{aligned}$$

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 (4.4.3.2). ∎

Remark 4.4.3.13. *The proof of Theorem 4.4.3.12 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 $T[\mathcal{B},\mathcal{C}]$ and $\widehat{T}[\mathcal{C}^*,\mathcal{B}^*]$ are transpose of each other, where the ordered bases $\mathcal{B}^*$ of $\mathbb{V}^*$ and $\mathcal{C}^*$ of $\mathbb{W}^*$ correspond, respectively, to the ordered bases $\mathcal{B}$ of $\mathbb{V}$ and $\mathcal{C}$ 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.*

4.4 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.

$$\begin{tikzcd}
(\mathbb{V},\mathcal{B},n) \arrow[r, "{T[\mathcal{B},\mathcal{C}]_{m\times n}}"] \arrow[rr, bend right, "{(ST)[\mathcal{B},\mathcal{D}]_{p\times n} = S[\mathcal{C},\mathcal{D}]\cdot T[\mathcal{B},\mathcal{C}]}"'] & (\mathbb{W},\mathcal{B}_2,m) \arrow[r, "{S[\mathcal{C},\mathcal{D}]_{p\times m}}"] & (Z,\mathcal{D},p)
\end{tikzcd}$$

Figure 4.2: Composition of Linear Transformations

Theorem 4.4.4.1 (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 Z, respectively. Then using Theorem 4.4.3.2, 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 4.4.4.1 we have the following result.

Theorem 4.4.4.2 (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 $\mathcal{B}$, $(T[\mathcal{B},\mathcal{B}])^{-1} = T^{-1}[\mathcal{B},\mathcal{B}]$. That is, the coordinate matrix is invertible.*

Proof. As T is invertible, $TT^{-1} = \mathrm{Id}$. Thus, Example 4.4.3.3.8c and Theorem 4.4.4.1 imply

$$I_n = \mathrm{Id}[\mathcal{B},\mathcal{B}] = (TT^{-1})[\mathcal{B},\mathcal{B}] = T[\mathcal{B},\mathcal{B}]\, T^{-1}[\mathcal{B},\mathcal{B}].$$

Hence, by definition of inverse, $T^{-1}[\mathcal{B},\mathcal{B}] = (T[\mathcal{B},\mathcal{B}])^{-1}$ and the required result follows. ∎

Exercise **4.4.4.3.** *Find the matrix of the linear transformations given below.*

1. *Define $T \in \mathcal{L}(\mathbb{R}^3)$ by $T(\mathbf{x}_1) = \mathbf{x}_2$, $T(\mathbf{x}_2) = \mathbf{x}_3$ and $T(\mathbf{x}_3) = \mathbf{x}_1$. Find $T[\mathcal{B},\mathcal{B}]$, where $\mathcal{B} = \left[, \mathbf{x}_2, \mathbf{x}_3\right]$ is an ordered basis of $\mathbb{R}^3$. Is T invertible?*

2. *Let $\mathcal{B} = \left[1, x, x^2, x^3\right]$ 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 $T^{-1}[\mathcal{B},\mathcal{B}]$.*

Let $\mathbb{V}$ be a finite dimensional vector space. Then, the next result answer the question "what happens to the matrix $T[\mathcal{B},\mathcal{B}]$ if the ordered basis $\mathcal{B}$ changes to $\mathcal{C}$?"

$$\begin{tikzcd}
(\mathbb{V},\mathcal{B}) \arrow[r, "{T[\mathcal{B},\mathcal{B}]}"] \arrow[d, "{\mathrm{Id}[\mathcal{B},\mathcal{C}]}"'] \arrow[dr, bend left, "\mathrm{Id}\circ T"] \arrow[dr, bend right, "T\circ \mathrm{Id}"'] & (\mathbb{V},\mathcal{B}_1) \arrow[d, "{\mathrm{Id}[\mathcal{B},\mathcal{C}]}"] \\
(\mathbb{V},\mathcal{C}) \arrow[r, "{T[\mathcal{C},\mathcal{C}]}"'] & (\mathbb{V},\mathcal{C})
\end{tikzcd}$$

Figure 4.3: Commutative Diagram for Similarity of Matrices

Theorem 4.4.4.4. *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 $T \in \mathcal{L}(\mathbb{V})$*

$$T[\mathcal{C},\mathcal{C}] = (Id[\mathcal{C},\mathcal{B}])^{-1} \cdot T[\mathcal{B},\mathcal{B}] \cdot Id[\mathcal{C},\mathcal{B}]. \tag{4.4.4.1}$$

Proof. The proof uses Theorem 4.4.4.1 to represent $T[\mathcal{B},\mathcal{C}]$ as $(\mathrm{Id}\circ T)[\mathcal{B},\mathcal{C}]$ and $(T \circ \mathrm{Id})[\mathcal{B},\mathcal{C}]$ (see Figure 4.3 for clarity). Now, by Theorem 4.4.4.1, $T[\mathcal{B},\mathcal{C}] = (\mathrm{Id}\circ T)[\mathcal{B},\mathcal{C}] = \mathrm{Id}[\mathcal{B},\mathcal{C}]\cdot T[\mathcal{B},\mathcal{B}]$ and $T[\mathcal{B},\mathcal{C}] = (T\circ \mathrm{Id})[\mathcal{B},\mathcal{C}] = T[\mathcal{C},\mathcal{C}]\cdot \mathrm{Id}[\mathcal{B},\mathcal{C}]$. Hence, $T[\mathcal{C},\mathcal{C}]\cdot \mathrm{Id}[\mathcal{B},\mathcal{C}] = \mathrm{Id}[\mathcal{B},\mathcal{C}]\cdot T[\mathcal{B},\mathcal{B}]$ and hence $T[\mathcal{C},\mathcal{C}] = (\mathrm{Id}[\mathcal{C},\mathcal{B}])^{-1}\cdot T[\mathcal{B},\mathcal{B}]\cdot \mathrm{Id}[\mathcal{C},\mathcal{B}]$ and the 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 4.4.4.4 tells us that all these matrices are related by an invertible matrix. Thus we are led to the following remark and the definition.

Remark 4.4.4.5. *As $T[\mathcal{C},\mathcal{C}] = Id[\mathcal{B},\mathcal{C}] \cdot T[\mathcal{B},\mathcal{B}] \cdot Id[\mathcal{C},\mathcal{B}]$, the matrix $Id[\mathcal{B},\mathcal{C}]$ is called the $\mathcal{B}:\mathcal{C}$ change of basis matrix (also, see Theorem 3.3.5.4).*

Definition 4.4.4.6. [Similar Matrices] Let $B, C \in \mathbb{M}_n(\mathbb{C})$. Then, B and C are said to be **similar** if there exists a non-singular matrix P such that $P^{-1}BP = C \Leftrightarrow BP = PC$.

Example 4.4.4.7. 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, for $\mathrm{Id}(a+bx+cx^2) = a+bx+cx^2$, verify that verify that $\mathrm{Id}[\mathcal{B},\mathcal{C}]^{-1} = \mathrm{Id}[\mathcal{C},\mathcal{B}]$, where

$$\begin{aligned}
\mathrm{Id}[\mathcal{C},\mathcal{B}] &= [[1]_{\mathcal{B}}, [1+x]_{\mathcal{B}}, [1+x+x^2]_{\mathcal{B}}] = \begin{bmatrix} -1 & 1 & -2 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and} \\
\mathrm{Id}[\mathcal{B},\mathcal{C}] &= [[1+x]_{\mathcal{C}}, [1+2x+x^2]_{\mathcal{C}}, [2+x]_{\mathcal{C}}] = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.
\end{aligned}$$

EXERCISE **4.4.4.8.**

1. *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 $\mathcal{C} = \left[(1,1,1), (1,-1,1), (1,1,2)\right]$ 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 $P^{-1}T[\mathcal{B},\mathcal{B}]\ P = T[\mathcal{C},\mathcal{C}]$.*

2. *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 $T^n = \mathbf{0}$.*
   - (a) *Prove that there exists $\mathbf{u} \in \mathbb{V}$ with $\{\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\}$, a basis of $\mathbb{V}$.*
   - (b) *If $\mathcal{B} = \left[\mathbf{u}, T(\mathbf{u}), \ldots, T^{n-1}(\mathbf{u})\right]$ then* $T[\mathcal{B},\mathcal{B}] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$.
   - (c) *Let A be an $n \times n$ matrix satisfying $A^{n-1} \neq \mathbf{0}$ but $A^n = \mathbf{0}$. Then prove that A is similar to the matrix given in Part 1b.*

3. *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.*

4.5 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.

Definition 5.5.1.1. [Inner Product] 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 5.5.1.2. *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 5.5.1.3. [Inner Product Space] 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 5.5.1.4. 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 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T A \mathbf{x}$. 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) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1$.

   (b) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1^2 + y_1^2 + x_2^2 + y_2^2$.

   (c) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1^3 + x_2 y_2^3$.

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 $\langle A, B \rangle = tr(B^T A)$. Then

$$\begin{aligned} \langle A + B, C \rangle &= \text{TR}\big(C^T(A + B)\big) = \text{TR}(C^T A) + \text{TR}(C^T B) = \langle A, C \rangle + \langle B, C \rangle \text{ and} \\ \langle A, B \rangle &= \text{TR}(B^T A) = \text{TR}(\ (B^T A)^T\ ) = \text{TR}(A^T B) = \langle B, A \rangle. \end{aligned}$$

   If $A = [a_{ij}]$ then $\langle A, A \rangle = \text{TR}r(A^T A) = \sum\limits_{i=1}^{n} (A^T A)_{ii} = \sum\limits_{i,j=1}^{n} a_{ij} a_{ij} = \sum\limits_{i,j=1}^{n} a_{ij}^2$ and therefore, $\langle A, A \rangle > 0$ for all non-zero matrix A.

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.A 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 5.5.1.5. [Length/Norm of a Vector] 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 5.5.1.6. 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 **5.5.1.7.** *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 $A \in \mathbb{M}_n(\mathbb{C})$ *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. 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 unknowns* a, b, c*.]*

7. 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* $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

A very useful and a fundamental inequality, commonly called the Cauchy-Schwartz inequality, concerning the inner product is proved next.

Theorem 5.5.1.8 (Cauchy-Bunyakovskii-Schwartz inequality). *Let $\mathbb{V}$ be an inner product space over $\mathbb{F}$. Then, for any $\mathbf{u}, \mathbf{v} \in \mathbb{V}$*

$$|\ \langle \mathbf{u}, \mathbf{v} \rangle\ |\ \le \|\mathbf{u}\|\ \|\mathbf{v}\|. \tag{5.5.1.1}$$

Moreover, equality holds in Inequality (5.5.1.1) if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. Furthermore, if $\mathbf{u} \neq \mathbf{0}$ then $\mathbf{v} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$.

Proof. If $\mathbf{u} = \mathbf{0}$ then Inequality (5.5.1.1) holds. Hence, let $\mathbf{u} \neq \mathbf{0}$. Then, by Definition 5.5.1.1.3, $\langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v} \rangle \ge 0$ for all $\lambda \in \mathbb{F}$ and $\mathbf{v} \in \mathbb{V}$. In particular, for $\lambda = -\dfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}$,

$$\begin{aligned} 0 &\le \langle \lambda\mathbf{u} + \mathbf{v}, \lambda\mathbf{u} + \mathbf{v} \rangle = \lambda\overline{\lambda}\|\mathbf{u}\|^2 + \lambda\langle \mathbf{u}, \mathbf{v} \rangle + \overline{\lambda}\langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{v}\|^2 \\ &= \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \frac{\overline{\langle \mathbf{v}, \mathbf{u} \rangle}}{\|\mathbf{u}\|^2} \|\mathbf{u}\|^2 - \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \langle \mathbf{u}, \mathbf{v} \rangle - \frac{\overline{\langle \mathbf{v}, \mathbf{u} \rangle}}{\|\mathbf{u}\|^2} \langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{v}\|^2 = \|\mathbf{v}\|^2 - \frac{|\ \langle \mathbf{v}, \mathbf{u} \rangle\ |^2}{\|\mathbf{u}\|^2}. \end{aligned}$$

Or, in other words $|\ \langle \mathbf{v}, \mathbf{u} \rangle\ |^2 \le \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$ and the proof of the inequality is over.

Now, note that equality holds in Inequality (5.5.1.1) 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 5.5.1.9. *Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then* $\left(\sum\limits_{i=1}^{n} \mathbf{x}_i \mathbf{y}_i \right)^2 \le \left(\sum\limits_{i=1}^{n} \mathbf{x}_i^2 \right) \left(\sum\limits_{i=1}^{n} \mathbf{y}_i^2 \right)$.

5.1.B 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 $-1 \le \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|} \le 1$. We use this together with the properties of the cosine function to define the angle between two vectors in an inner product space.

Definition 5.5.1.10. [Angle between two vectors] Let $\mathbb{V}$ be a real vector space. If $\theta \in [0,\ \pi]$ is the angle between $\mathbf{u}, \mathbf{v} \in \mathbb{V} \setminus \{\mathbf{0}\}$ then we define

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|}.$$

Example 5.5.1.11.

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 $(a+b)\left(\frac{1}{a}+\frac{1}{b}\right) \geq 4$.

6. For $1 \leq i \leq n$, let $a_i \in \mathbb{R}$ with $a_i > 0$. Then, use Corollary 5.5.1.9 to show that $\left(\sum\limits_{i=1}^{n} a_i\right)\left(\sum\limits_{i=1}^{n} \frac{1}{a_i}\right) \geq n^2$.

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 $\alpha = 1$?

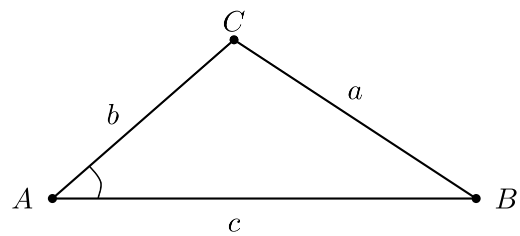


Figure 2: Triangle with vertices A, B and C

We will now prove that if A, B and C are the vertices of a triangle (see Figure 5.1.B) 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 5.5.1.12. *Let A, B and C be the vertices of a triangle (see Figure 5.1.B) 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 5.5.1.13. [Orthogonality] 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 5.5.1.14. 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 non-zero 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 unknowns 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 **5.5.1.15.** *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 5.5.1.4.8). If* $\mathbb{W} = \{A \in \mathbb{M}_n(\mathbb{R}) \mid A^T = A\}$ *then* $\mathbb{W}^\perp$*?*

5.1.C Normed Linear Space

To proceed further, recall that a vector space over $\mathbb{R}$ or $\mathbb{C}$ was a linear space.

Definition 5.5.1.16. 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 5.5.1.17. *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 5.5.1.18. 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 **5.5.1.19.** *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 $\|\mathbf{x}+i\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|i\mathbf{y}\|^2$.

The next result is stated without proof as the proof is beyond the scope of this book.

Theorem 5.5.1.20. *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 5.5.1.21. 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

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \|\,(\mathbf{x}-\mathbf{z})-(\mathbf{y}-\mathbf{z})\,\| \le \|\,(\mathbf{x}-\mathbf{z})\,\|+\|\,(\mathbf{y}-\mathbf{z})\,\| = d(\mathbf{x},\mathbf{z})+d(\mathbf{z},\mathbf{y}).$$

5.1.D Application to Fundamental Spaces

We end this section by proving the fundamental theorem of linear algebra. So, the readers are advised to recall the four fundamental subspaces and also to go through Theorem 3.3.4.9 (the rank-nullity theorem for matrices). We start with the following result.

Lemma 5.5.1.22. *Let* $A \in \mathbb{M}_{m,n}(\mathbb{R})$. *Then* $\text{NULL}(A) = \text{NULL}(A^TA)$.

Proof. Let $\mathbf{x} \in \text{NULL}(A)$. Then $A\mathbf{x} = \mathbf{0}$. So, $(A^TA)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}$. Thus, $\mathbf{x} \in \text{NULL}(A^TA)$. That is, $\text{NULL}(A) \subseteq \text{NULL}(A^TA)$.

Suppose that $\mathbf{x} \in \text{NULL}(A^TA)$. Then $(A^TA)\mathbf{x} = \mathbf{0}$ and $0 = \mathbf{x}^T\mathbf{0} = \mathbf{x}^T(A^TA)\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2$. Thus, $A\mathbf{x} = \mathbf{0}$ and the required result follows. ■

Theorem 5.5.1.23 (Fundamental Theorem of Linear Algebra)**.** *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *Then*

1. $\dim(\text{NULL}(A)) + \dim(\text{COL}(A)) = n$.
2. $\text{NULL}(A) = \big(\text{COL}(A^*)\big)^{\perp}$ *and* $\text{NULL}(A^*) = \big(\text{COL}(A)\big)^{\perp}$.
3. $\dim(\text{COL}(A)) = \dim(\text{COL}(A^*))$.

Proof. PART 1: Proved in Theorem 3.3.4.9.

PART 2: We first prove that $\text{NULL}(A) \subseteq \text{COL}(A^*)^{\perp}$. Let $\mathbf{x} \in \text{NULL}(A)$. Then $A\mathbf{x} = \mathbf{0}$ and

$$0 = \langle \mathbf{0}, \mathbf{u}\rangle = \langle A\mathbf{x}, \mathbf{u}\rangle = \mathbf{u}^*A\mathbf{x} = (A^*\mathbf{u})^*\mathbf{x} = \langle \mathbf{x}, A^*\mathbf{u}\rangle, \text{ for all } \mathbf{u} \in \mathbb{C}^n.$$

But $\text{COL}(A^*) = \{A^*\mathbf{u} \mid \mathbf{u} \in \mathbb{C}^n\}$. Thus, $\mathbf{x} \in \text{COL}(A^*)^{\perp}$ and $\text{NULL}(A) \subseteq \text{COL}(A^*)^{\perp}$.

We now prove that $\text{COL}(A^*)^{\perp} \subseteq \text{NULL}(A)$. Let $\mathbf{x} \in \text{COL}(A^*)^{\perp}$. Then, for every $\mathbf{y} \in \mathbb{C}^n$,

$$0 = \langle \mathbf{x}, A^*\mathbf{y}\rangle = (A^*\mathbf{y})^*\mathbf{x} = \mathbf{y}^*(A^*)^*\mathbf{x} = \mathbf{y}^*A\mathbf{x} = \langle A\mathbf{x}, \mathbf{y}\rangle.$$

In particular, for $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^n$, we get $\|A\mathbf{x}\|^2 = 0$. Hence $A\mathbf{x} = \mathbf{0}$. That is, $\mathbf{x} \in \text{NULL}(A)$. 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. ■

We now give some implications of the above theorem.

Corollary 5.5.1.24. *Let* $A \in \mathbb{M}_n(\mathbb{R})$. *Then the function* $T : \text{COL}(A^T) \to \text{COL}(A)$ *defined by* $T(\mathbf{x}) = A\mathbf{x}$ *is one-one and onto.*

Proof. In view of Theorem 5.5.1.23.3, we just need to show that the map is one-one. So, let us assume that there exist $\mathbf{x}, \mathbf{y} \in \textsc{Col}(A^T)$ such that $T(\mathbf{x}) = T(\mathbf{y})$. Or equivalently, $A\mathbf{x} = A\mathbf{y}$. Thus, $\mathbf{x} - \mathbf{y} \in \textsc{Null}(A) = (\textsc{Col}(A^T))^{\perp}$ (by Theorem 5.5.1.23.2). Therefore, $\mathbf{x} - \mathbf{y} \in (\textsc{Col}(A^T))^{\perp} \cap \textsc{Col}(A^T) = \{\mathbf{0}\}$ (by Example 2). Thus, $\mathbf{x} = \mathbf{y}$ and hence the map is one-one. ∎

Remark 5.5.1.25. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$.*

1. *Then the spaces $\textsc{Col}(A)$ and $\textsc{Null}(A^T)$ are not only orthogonal but are orthogonal complement of each other.*
2. *Further if Rank$(A) = r$ then, using Corollary 5.5.1.24, we observe the following:*
   (a) *If $i_1, \ldots, i_r$ are the pivot rows of A then $\{A(A[i_1,:]^T), \ldots, A(A[i_r,:]^T)\}$ form a basis of $\textsc{Col}(A)$.*
   (b) *Similarly, if $j_1, \ldots, j_r$ are the pivot columns of A then $\{A^T(A[:,j_1]), \ldots, A^T(A[:,j_r])\}$ form a basis of $\textsc{Col}(A^T)$.*
   (c) *So, if we choose the rows and columns corresponding to the pivot entries then the corresponding $r \times r$ submatrix of A is invertible.*

The readers should look at Example 3.3.1.26 and Remark 3.3.1.27. We give one more example.

Example 5.5.1.26. 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 $\textsc{Col}(A)$.
2. $\{(1,1,-1)^T\}$ is a basis of $\textsc{Null}(A^T)$.
3. $\textsc{Null}(A^T) = (\textsc{Col}(A))^{\perp}$.

Exercise **5.5.1.27.**

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."

5.1.E Properties of Orthonormal Vectors

At the end of the previous section, we saw that $\text{Col}(A)$ is orthogonal to $\text{Null}(A^T)$. So, in this section, we try to understand the orthogonal vectors.

Definition 5.5.1.28. [Orthonormal Set] 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 **orthonormal set** if

1. $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$, for all $1 \leq i \neq j \leq n$. That is, $\mathbf{v}_i$ and $\mathbf{v}_j$ are **mutually orthogonal**, for $1 \leq i \neq j \leq n$.
2. $\|\mathbf{v}_i\| = 1$, for $1 \leq i \leq n$

If S is also a basis of $\mathbb{V}$ then S is called an **orthonormal basis** of $\mathbb{V}$.

Example 5.5.1.29. 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 $1 \leq i \leq n$.
   (b) $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$, for $1 \leq i \neq j \leq n$.
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 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 $S = \{\mathbf{1}\} \cup \{\mathbf{e}_m \mid m \geq 1\} \cup \{\mathbf{f}_n \mid n \geq 1\}$, where $\mathbf{1}(x) = 1$, $\mathbf{e}_m(x) = \cos(mx)$ and $\mathbf{f}_n(x) = \sin(nx)$, for all $m, n \geq 1$ 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, $\left\{\frac{\mathbf{1}}{\sqrt{2\pi}}\right\} \cup \left\{\frac{1}{\sqrt{\pi}}\mathbf{e}_m \mid m \geq 1\right\} \cup \left\{\frac{1}{\sqrt{\pi}}\mathbf{f}_n \mid n \geq 1\right\}$ is an orthonormal set in $\mathcal{C}[-\pi,\ \pi]$.

Theorem 5.5.1.30. *Let $\mathbb{V}$ be an* IPS *with $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ as a set of mutually orthogonal vectors.*

1. *Then the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is linearly independent.*
2. *Let $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i \in \mathbb{V}$. Then $\|\mathbf{v}\|^2 = \|\sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 = \sum\limits_{i=1}^{n} \mid \alpha_i \mid^2 \|\mathbf{u}_i\|^2$;*
3. *Let $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{u}_i \in \mathbb{V}$. So, for $1 \leq i \leq n$, if $\|\mathbf{u}_i\| = 1$ then $\alpha_i = \langle \mathbf{v}, \mathbf{u}_i \rangle$. That is, $\mathbf{v} = \sum\limits_{i=1}^{n} \langle \mathbf{v}, \mathbf{u}_i \rangle \mathbf{u}_i$ and $\|\mathbf{v}\|^2 = \sum\limits_{i=1}^{n} \mid \langle \mathbf{v}, \mathbf{u}_i \rangle \mid^2$.*
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 unknowns $c_1, \ldots, c_n$. 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

$$0 = \langle \mathbf{0}, \mathbf{u}_i\rangle = \langle c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n, \mathbf{u}_i\rangle = \sum_{j=1}^{n} c_j \langle \mathbf{u}_j, \mathbf{u}_i\rangle = c_i\langle \mathbf{u}_i, \mathbf{u}_i\rangle.$$

As $\mathbf{u}_i \neq \mathbf{0}$, $\langle \mathbf{u}_i, \mathbf{u}_i\rangle \neq 0$ and therefore $c_i = 0$, for $1 \le i \le n$. Thus, the above linear system has only the trivial solution. Hence, this completes the proof of Part 1.

Part 2: A similar argument gives

$$\begin{aligned} \|\sum_{i=1}^{n} \alpha_i \mathbf{u}_i\|^2 &= \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{u}_i, \sum_{i=1}^{n} \alpha_i \mathbf{u}_i \right\rangle = \sum_{i=1}^{n} \alpha_i \left\langle \mathbf{u}_i, \sum_{j=1}^{n} \alpha_j \mathbf{u}_j \right\rangle \\ &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} \overline{\alpha_j}\, \langle \mathbf{u}_i, \mathbf{u}_j\rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i}\, \langle \mathbf{u}_i, \mathbf{u}_i\rangle = \sum_{i=1}^{n} \mid \alpha_i \mid^2 \|\mathbf{u}_i\|^2. \end{aligned}$$

Part 3: If $\|\mathbf{u}_i\| = 1$, for $1 \le i \le n$ then $\langle \mathbf{v}, \mathbf{u}_i\rangle = \left\langle \sum\limits_{j=1}^{n} \alpha_j \mathbf{u}_j, \mathbf{u}_i \right\rangle = \sum\limits_{j=1}^{n} \alpha_j \langle \mathbf{u}_j, \mathbf{u}_i\rangle = \alpha_j$.

Part 4: Follows directly using Part 3 as $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis of $\mathbb{V}$. ∎

Remark 5.5.1.31. *Using Theorem 5.5.1.30, we see that if* $\mathcal{B} = \big[\mathbf{v}_1, \ldots, \mathbf{v}_n\big]$ *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 **5.5.1.32.**

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* $\mathbb{R}^3$, $[(2,3,1)^T]_{\mathcal{B}} = \begin{bmatrix}2\sqrt{3}\\ \frac{-1}{\sqrt{2}}\\ \frac{3}{\sqrt{6}}\end{bmatrix}$.
4. *Let* $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subseteq \mathbb{R}^n$ *and define* $A = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. *Then prove that* A *is an orthogonal matrix if and only if* S *is an orthonormal basis of* $\mathbb{R}^n$.
5. *Let* A *be an* $n \times n$ *orthogonal matrix. Then prove that*
   (a) *the rows/columns of* A *form an orthonormal basis of* $\mathbb{R}^n$.
   (b) *for any two vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ ORTHOGONAL MATRICES PRESERVE ANGLE.
   (c) *for any vector* $\mathbf{x} \in \mathbb{R}^n$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$ ORTHOGONAL MATRICES PRESERVE LENGTH.
6. *Let* A *be an* $n \times n$ *unitary matrix. Then prove that*

(a) *the rows/columns of A form an orthonormal basis of the complex vector space $\mathbb{C}^n$.*

(b) *for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$* UNITARY MATRICES PRESERVE ANGLE.

(c) *for any vector $\mathbf{x} \in \mathbb{C}^n$, $\|A\mathbf{x}\| = \|\mathbf{x}\|$* UNITARY MATRICES PRESERVE LENGTH.

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

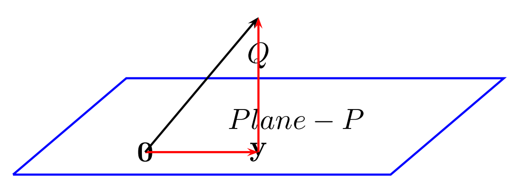
8. *If $A = [a_{ij}]$ and $B = [b_{ij}]$ are unitarily equivalent then prove that $\sum\limits_{ij} |a_{ij}|^2 = \sum\limits_{ij} |b_{ij}|^2$.*

9. *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.*

5.2 Gram-Schmidt Orthogonalization Process

In view of the importance of Theorem 5.5.1.30, 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 Orthogonalization process**. We first consider a few examples.

Example 5.5.2.1. 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, $\mathbf{y} = \vec{Q} - \overrightarrow{\mathbf{y}Q}$. So, need to find a way to compute $\overrightarrow{\mathbf{y}Q}$, a line on the plane passing through $\mathbf{0}$ and $\mathbf{y}$.

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 $\mathbf{v}$.

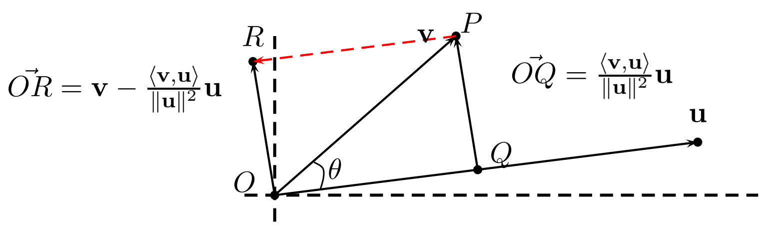


Figure 3: Decomposition of vector $\mathbf{v}$

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 5.5.1.10, $\|\vec{OQ}\| = \|\mathbf{v}\| \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{v}\|\,\|\mathbf{u}\|}\right| = \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{u}\|}\right|$, 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, $\mathbf{y} = \vec{OQ} = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$ and $\mathbf{z} = \mathbf{v} - \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|}$. 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,

$$\mathrm{Proj}_{\mathbf{u}}(\mathbf{v}) = \left\langle \mathbf{v}, \frac{\mathbf{u}}{\|\mathbf{u}\|}\right\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|} \ \text{ and } \ \|\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})\| = \|\vec{OQ}\| = \left|\frac{\langle \mathbf{v},\mathbf{u}\rangle}{\|\mathbf{u}\|}\right|. \tag{5.5.2.1}$$

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 5.5.2.2.

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 (5.5.2.1), we have $\mathrm{Proj}_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{v},\mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{3}(1,1,-1,0)^T$ and $\mathbf{w} = (1,1,1,1)^T - \mathrm{Proj}_{\mathbf{v}}(\mathbf{u}) = \frac{1}{3}(2,2,4,3)^T$ 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) $\mathbf{v}_1 = \mathrm{Proj}_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{v},\mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^T$ 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 $\mathrm{Proj}_{\mathbf{u}}(\mathbf{w})$ is parallel to $\mathbf{u}$ and $\mathrm{Proj}_{\mathbf{v}_2}(\mathbf{w})$ is parallel to $\mathbf{v}_2$. Hence, we have

   (a) $\mathbf{w}_1 = \mathrm{Proj}_{\mathbf{u}}(\mathbf{w}) = \langle \mathbf{w},\mathbf{u}\rangle \frac{\mathbf{u}}{\|\mathbf{u}\|^2} = \frac{1}{4}\mathbf{u} = \frac{1}{4}(1,1,1,1)^T$ is parallel to $\mathbf{u}$,

   (b) $\mathbf{w}_2 = \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{w}) = \langle \mathbf{w},\mathbf{v}_2\rangle \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|^2} = \frac{7}{44}(3,3,-5,-1)^T$ is parallel to $\mathbf{v}_2$ and

   (c) $\mathbf{w}_3 = \mathbf{w} - \mathbf{w}_1 - \mathbf{w}_2 = \frac{3}{11}(1,1,2,-4)^T$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}_2$.

That is, from the given vector subtract all the orthogonal projections/components. If the new vector is non-zero then this vector is orthogonal to the previous ones. This idea is generalized to give the Gram-Schmidt Orthogonalization process.

Theorem 5.5.2.3 (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 $1 \leq i \leq n$.*

Proof. Note that for orthonormality, we need $\|\mathbf{w}_i\| = 1$, for $1 \leq i \leq n$ and $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$, for $1 \leq i \neq j \leq n$. Also, by Corollary 3.3.2.7.2, $\mathbf{v}_i \notin LS(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1})$, for $2 \leq i \leq n$, 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 $\mathbf{u}_2 \neq \mathbf{0}$ 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 $\mathbf{u}_k = \mathbf{v}_k - \sum\limits_{i=1}^{k-1} \langle \mathbf{v}_k, \mathbf{w}_i \rangle \mathbf{w}_i = \mathbf{v}_k - \sum\limits_{i=1}^{k-1} \text{Proj}_{\mathbf{w}_i}(\mathbf{v}_k) \neq \mathbf{0}$ 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 $2 \leq i \leq k-1$. 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 5.5.2.4.

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 $S = \{\mathbf{v}_1 = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}^T, \mathbf{v}_2 = \begin{bmatrix} \frac{3}{2} & 2 & 0 \end{bmatrix}^T, \mathbf{v}_3 = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} & 0 \end{bmatrix}^T, \mathbf{v}_4 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T\}$. 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 5.5.2.5. *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 5.5.2.6. *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 **5.5.2.7.**

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* $\mathbf{v}_0 \in S^{\perp}$. *That is,* $\mathbb{V} = LS(\mathbf{u}, S^{\perp})$.

5.3 Orthogonal Operator and Rigid Motion

We now give the definition and a few properties of an orthogonal operator.

Definition 5.5.3.1. [Orthogonal Operator] 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 5.5.3.2. 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 $T(\mathbf{x}) = 2\langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a} - \mathbf{x}$, for all $\mathbf{x} \in \mathbb{V}$.
   Solution: Note that $\mathrm{Proj}_{\mathbf{a}}(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}$. So, $\big\langle \langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a},\ \mathbf{x} - \langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}\big\rangle = 0$. Also, by Pythagoras theorem $\|\mathbf{x} - \langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}\|^2 = \|\mathbf{x}\|^2 - (\langle \mathbf{a}, \mathbf{x}\rangle)^2$. Thus,

$$\|T(\mathbf{x})\|^2 = \|(\langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}) + (\langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a} - \mathbf{x})\|^2 = \|\langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}\|^2 + \|\mathbf{x} - \langle \mathbf{a}, \mathbf{x}\rangle \mathbf{a}\|^2 = \|\mathbf{x}\|^2.$$

2. Let $n = 2, \mathbb{V} = \mathbb{R}^2$ and $0 \leq \theta < 2\pi$. 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 5.5.3.3. *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 5.5.3.4. *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 5.5.3.5. [Isometry, Rigid Motion] 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 5.5.3.6. 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 1), 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 5.5.3.3, we see that every orthogonal operator is an isometry.

We now prove that every rigid motion that fixes origin is an orthogonal operator.

Theorem 5.5.3.7. *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 5.5.3.3. 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. ■

5.4 Orthogonal Projections and Applications

Till now, our main interest was to understand the linear system $A\mathbf{x} = \mathbf{b}$ from different points of view. But, in most practical situations the system has no solution. So, we try to find $\mathbf{x}$ such that the vector $\mathbf{err} = \mathbf{b} - A\mathbf{x}$ has minimum norm. The next result gives the existence of an orthogonal subspace of a finite dimensional inner product space.

Theorem 5.5.4.1 (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 only closest point in $\mathbb{W}$ from $\mathbf{b}$. Furthermore, $\mathbf{b} - \mathbf{y} \in \mathbb{W}^\perp$.*

Proof. Clearly $\mathbf{y} = \sum\limits_{i=1}^{k} \langle \mathbf{b}, \mathbf{f}_i\rangle \mathbf{f}_i \in \mathbb{W}$. As the closet point is the feet of the perpendicular, we need to show that $\mathbf{b} - \mathbf{y} \in \mathbb{W}^\perp$. To do so, we verify that $\langle \mathbf{b} - \mathbf{y}, \mathbf{f}_i\rangle = 0$, for $1 \le i \le k$.

$$\langle \mathbf{b} - \mathbf{y}, \mathbf{f}_i\rangle = \langle \mathbf{b}, \mathbf{f}_i\rangle - \left\langle \sum_{j=1}^{k} \langle \mathbf{b}, \mathbf{f}_j\rangle \mathbf{f_j}, \mathbf{f}_i \right\rangle = \langle \mathbf{b}, \mathbf{f}_i\rangle - \sum_{j=1}^{k} \langle \mathbf{b}, \mathbf{f}_j\rangle \langle \mathbf{f_j}, \mathbf{f}_i\rangle = \langle \mathbf{b}, \mathbf{f}_i\rangle - \langle \mathbf{b}, \mathbf{f}_i\rangle = 0.$$

Also, note that for each $\mathbf{w} \in W$, $\mathbf{y} - \mathbf{w} \in \mathbb{W}$ and hence

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \mathbf{y} + \mathbf{y} - \mathbf{w}\|^2 = \|\mathbf{b} - \mathbf{y}\|^2 + \|\mathbf{y} - \mathbf{w}\|^2 \ge \|\mathbf{b} - \mathbf{y}\|^2.$$

Thus, $\mathbf{y}$ is the closet point in $\mathbb{W}$ from $\mathbf{b}$. Now, use Pythagoras theorem to conclude that $\mathbf{y}$ is unique. Thus, the required result follows. ■

We now give a definition and then an implication of Theorem 5.5.4.1.

Definition 5.5.4.2. [Orthogonal Projection] Let $\mathbb{W}$ be a finite dimensional subspace of an IPS $\mathbb{V}$. Then, by Theorem 5.5.4.1, 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

$$P_{\mathbb{W}} : \mathbb{V} \to \mathbb{V} \text{ by } P_{\mathbb{W}}(\mathbf{v}) = \mathbf{w}.$$

The vector $\mathbf{w}$ is called the **projection** of $\mathbf{v}$ on $\mathbb{W}$.

Remark 5.5.4.3. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$. Then, to find the orthogonal projection $\mathbf{y}$ of a vector $\mathbf{b}$ on $\text{Col}(A)$, 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 $\mathbf{y} = \sum\limits_{i=1}^{k} \langle \mathbf{b}, \mathbf{f}_i \rangle \mathbf{f}_i$.*
2. *By Remark 5.5.1.25.1, the spaces $\text{Col}(A)$ and $\text{Null}(A^T)$ are completely orthogonal. Hence, every $\mathbf{b} \in \mathbb{R}^m$ equals $\mathbf{b} = \mathbf{u} + \mathbf{v}$ for unique $\mathbf{u} \in \text{Col}(A)$ and $\mathbf{v} \in \text{Null}(A^T)$. Thus, using Definition 5.5.4.2 and Theorem 5.5.4.1, $\mathbf{y} = \mathbf{u}$.*

Before proceeding to projections, we give an application of Theorem 5.5.4.1 to a linear system.

Corollary 5.5.4.4. *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 $A^T A\mathbf{x} = A^T\mathbf{b}$. Conversely, every solution of $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least square solution of $A\mathbf{x} = \mathbf{b}$.*

Proof. Let $\mathbb{W} = \text{Col}(A)$. Then, by Remark 5.5.4.3, $\mathbf{b} = \mathbf{y} + \mathbf{v}$, where $\mathbf{y} \in \mathbb{W}$, $\mathbf{v} \in \text{Null}(A^T)$ and $\min\{\|\mathbf{b} - \mathbf{w}\| \mid \mathbf{w} \in \mathbb{W}\} = \|\mathbf{b} - \mathbf{y}\|$.

As $\mathbf{y} \in \mathbb{W}$ there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $A\mathbf{x}_0 = \mathbf{y}$. That is, $\mathbf{x}_0$ is the least square solution of $A\mathbf{x} = \mathbf{b}$. Hence,

$$(A^T A)\mathbf{x}_0 = A^T (A\mathbf{x}_0) = A^T\mathbf{y} = A^T (\mathbf{b} - \mathbf{v}) = A^T\mathbf{b} - \mathbf{0} = A^T\mathbf{b}.$$

Conversely, we need to show that $\min\{\|\mathbf{b} - A\mathbf{x}\| \mid \mathbf{x} \in \mathbb{R}^n\} = \|\mathbf{b} - A\mathbf{x}_1\|$, where $\mathbf{x}_1 \in \mathbb{R}^n$ is a solution of $A^T A\mathbf{x} = A^T\mathbf{b}$. Thus, $A^T(A\mathbf{x}_1 - \mathbf{b}) = \mathbf{0}$. Hence, for any $\mathbf{x} \in \mathbb{R}^n$, $\langle \mathbf{b} - A\mathbf{x}_1, A(\mathbf{x} - \mathbf{x}_1) \rangle = (\mathbf{x} - \mathbf{x}_1)^T A^T (\mathbf{b} - A\mathbf{x}_1) = (\mathbf{x} - \mathbf{x}_1)^T \mathbf{0} = 0$. Thus,

$$\|\mathbf{b} - A\mathbf{x}\|^2 = \|\mathbf{b} - A\mathbf{x}_1 + A\mathbf{x}_1 - A\mathbf{x}\|^2 = \|\mathbf{b} - A\mathbf{x}_1\|^2 + \|A\mathbf{x}_1 - A\mathbf{x}\|^2 \geq \|\mathbf{b} - \mathbf{y}\|^2.$$

Hence, the required result follows. ∎

The above corollary gives the following result.

Corollary 5.5.4.5. *Let $A \in \mathbb{M}_{m,n}(\mathbb{R})$ and $\mathbf{b} \in \mathbb{R}^m$. If*

1. *$A^T A$ is invertible then the least square solution of $A\mathbf{x} = \mathbf{b}$ equals $\mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$.*
2. *$A^T A$ is not invertible then the least square solution of $A\mathbf{x} = \mathbf{b}$ equals $\mathbf{x} = (A^T A)^{-} A^T\mathbf{b}$, where $(A^T A)^{-}$ is the pseudo-inverse of $A^T A$.*

Proof. Part 1 directly follows from Corollary 5.5.4.5. For Part 1, let $\mathbb{W} = \text{Col}(A)$. Then, by Remark 5.5.4.3, $\mathbf{b} = \mathbf{y} + \mathbf{v}$, where $\mathbf{y} \in \mathbb{W}$ and $\mathbf{v} \in \text{Null}(A^T)$. So, $A^T\mathbf{b} = A^T(\mathbf{y} + \mathbf{v}) = A^T\mathbf{y}$. Since $\mathbf{y} \in \mathbb{W}$, there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $A\mathbf{x}_0 = \mathbf{y}$. Thus, $A^T\mathbf{b} = A^T A\mathbf{x}_0$. Now, using the definition of pseudo-inverse (see Exercise 1.1.3.6.18), we see that

$$(A^A A)\left((A^T A)^{-} A^T\mathbf{b}\right) = (A^T A)(A^T A)^{-}(A^T A)\mathbf{x}_0 = (A^T A)\mathbf{x}_0 = A^T\mathbf{b}.$$

Thus, we see that $(A^T A)^{-} A^T\mathbf{b}$ is a solution of the system $A^T A\mathbf{x} = A^T\mathbf{b}$. Hence, by Corollary 5.5.4.4, the required result follows. ∎

We now give a few examples to understand projections.

Example 5.5.4.6. 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 $\text{Null}([1,-1,1,-1,1])$.
   Solution: Here $A=[1,-1,1,-1,1]$. So, a basis of $\text{Col}(A^T)$ equals $\{(1,-1,1,-1,1)^T\}$ and that of $\text{Null}(A)$ equals $\{(1,1,0,0,0)^T,(1,0,-1,0,0)^T,(1,0,0,1,0)^T,(1,0,0,0,-1)^T\}$. Then, the solution of the linear system $B\mathbf{x}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$, where $B=\begin{bmatrix}1&1&1&1&1\\1&0&0&0&-1\\0&-1&0&0&1\\0&0&1&0&-1\\0&0&0&-1&1\end{bmatrix}$ equals $\mathbf{x}=\frac{1}{5}\begin{bmatrix}6\\-4\\6\\-4\\1\end{bmatrix}$. Thus, the projection is $\frac{1}{5}\left(6(1,1,0,0,0)^T-4(1,0,-1,0,0)^T+6(1,0,0,1,0)^T-4(1,0,0,0,-1)^T\right)=\frac{2}{5}(2,3,2,3,2)^T$.
2. Determine the projection of $(1,1,1)^T$ on $\text{Null}([1,1,-1])$.
   Solution: Here $A=[1,1,-1]$. So, a basis of $\text{Null}(A)$ equals $\{(1,-1,0)^T,(1,0,1)^T\}$ and that of $\text{Col}(A^T)$ 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 $\text{Col}\left([1,2,1]^T\right)$.
   Solution: Here, $A^T=[1,2,1]$, a basis of $\text{Col}(A)$ equals $\{(1,2,1)^T\}$ and that of $\text{Null}(A^T)$ 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 5.5.4.3, we prove the following result.

Theorem 5.5.4.7. *[Matrix of Orthogonal Projection] Let $\mathbb{W}$ be a subspace of an* IPS *$\mathbb{V}$ with $\dim(\mathbb{W})<\infty$. If $\{\mathbf{f}_1,\ldots,\mathbf{f}_k\}$ is an orthonormal basis of $\mathbb{W}$ then $P_{\mathbb{W}}=\sum\limits_{i=1}^{k}\mathbf{f}_i\mathbf{f}_i^T$.*

Proof. Let $\mathbf{v}\in\mathbb{V}$. Then $P_{\mathbb{W}}\mathbf{v}=\left(\sum\limits_{i=1}^{k}\mathbf{f}_i\mathbf{f}_i^T\right)\mathbf{v}=\sum\limits_{i=1}^{k}\mathbf{f}_i\left(\mathbf{f}_i^T\mathbf{v}\right)=\sum\limits_{i=1}^{k}\langle\mathbf{v},\mathbf{f}_i\rangle\mathbf{f}_i$. As $P_{\mathbb{W}}\mathbf{v}$ indeed gives the only closet point (see Theorem 5.5.4.1), the required result follows. ∎

Example 5.5.4.8. 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 5.5.4.6.

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. $\mathbb{W} = \{(x,y,z)^T \in \mathbb{R}^3 \mid x+y-z=0\} = \textsc{Null}\,([1,1,-1])$.

   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. $\mathbb{W} = LS(\ (1,2,1)\ ) = \textsc{Col}\left([1,2,1]^T\right) \subseteq \mathbb{R}^3$.

   Solution: Using Example 5.5.2.4.3 and Equation (5.5.2.1)

$$\mathbb{W}^\perp = LS(\{(-2,1,0),(-1,0,1)\}) = LS(\{(-2,1,0),(1,2,-5)\}).$$

   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 5.5.4.9. *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 **5.5.4.10.**

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) *$\textsc{Null}(P) \cap \textsc{Rng}(P) = \textsc{Null}(A) \cap \textsc{Col}(A) = \{\mathbf{0}\}$.*

(c) *$\mathbb{R}^2 = \textsc{Null}(P) + \textsc{Rng}(P)$. But, $(\textsc{Rng}(P))^{\perp} = (\textsc{Col}(A))^{\perp} \neq \textsc{Null}(A)$.*

(d) *Since $(\textsc{Col}(A))^{\perp} \neq \textsc{Null}(A)$, the map P is not an orthogonal projector. In this case, P is called a projection of $\mathbb{R}^2$ onto $\textsc{Rng}(P)$ along $\textsc{Null}(P)$.*

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 $\mathbb{W}^{\perp}$.*

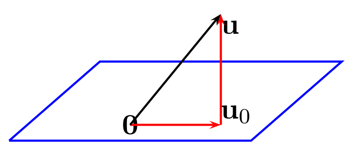
(b) *Q sends every vector $\mathbf{w} \in \mathbb{W}$ to $-\mathbf{w}$.*

(c) *$Q \circ Q = I_n$.*

The function Q is called the **reflection operator** *with respect to $\mathbb{W}^{\perp}$.*

Theorem 5.5.4.11 (Bessel's Inequality)**.** *Let $\mathbb{V}$ be an IPS with $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ as an orthogonal set. Then $\sum\limits_{k=1}^{n} \dfrac{|\langle \mathbf{u}, \mathbf{v}_k \rangle|^2}{\|\mathbf{v}_k\|^2} \leq \|\mathbf{u}\|^2$, for each $\mathbf{u} \in \mathbb{V}$. Equality holds if and only if $\mathbf{u} = \sum\limits_{k=1}^{n} \dfrac{\langle \mathbf{u}, \mathbf{v}_k \rangle}{\|\mathbf{v}_k\|^2} \mathbf{v}_k$.*

Proof. For $1 \leq k \leq n$, define $\mathbf{w}_k = \dfrac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$ and $\mathbf{u}_0 = \sum\limits_{k=1}^{n} \langle \mathbf{u}, \mathbf{w}_k \rangle \mathbf{w}_k$. Then, by Theorem 5.5.4.1 $\mathbf{u}_0$ is the nearest the vector to $\mathbf{u}$ in $LS(\mathbf{v}_1, \cdots, \mathbf{v}_n)$. Also, $\langle \mathbf{u} - \mathbf{u}_0, \mathbf{u}_0 \rangle = 0$. So, $\|\mathbf{u}\|^2 = \|\mathbf{u} - \mathbf{u}_0 + \mathbf{u}_0\|^2 = \|\mathbf{u} - \mathbf{u}_0\|^2 + \|\mathbf{u}_0\|^2 \geq \|\mathbf{u}_0\|^2$. Thus, we have obtained the required inequality. The equality is attained if and only if $\mathbf{u} - \mathbf{u}_0 = \mathbf{0}$ or equivalently, $\mathbf{u} = \mathbf{u}_0$.



■

We now give a generalization of the pythagoras theorem. The proof is left as an exercise for the reader.

Theorem 5.5.4.12 (Parseval's formula)**.** *Let $\mathbb{V}$ be an IPS with $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ as an orthonormal basis of $\mathbb{V}$. Then, for each $\mathbf{x}, \mathbf{y} \in \mathbb{V}$, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum\limits_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle \overline{\langle \mathbf{y}, \mathbf{v}_i \rangle}$. Furthermore, if $\mathbf{x} = \mathbf{y}$ then $\|\mathbf{x}\|^2 = \sum\limits_{i=1}^{n} |\langle \mathbf{x}, \mathbf{v}_i \rangle|^2$, giving us a generalization of the* **Pythagoras Theorem***.*

Exercise **5.5.4.13.** *Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$. Then there exists a unique B such that $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, B\mathbf{y} \rangle$, for all $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m$. In fact $B = A^T$.*

5.4.A Orthogonal Projections as Self-Adjoint Operators*

Theorem 5.5.4.9 implies that the matrix of the projection operator is symmetric. We use this idea to proceed further.

Definition 5.5.4.14. [Self-Adjoint Operator] 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 5.5.4.15. 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 5.5.4.15.1) 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 5.5.4.16. *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 5.5.4.17. *Theorem 5.5.4.16 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, $\|\mathbf{v} - \mathbf{w}\| \ge \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|$ 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

$$d(\mathbf{v}, \mathbb{W}) = \inf \{\|\mathbf{v} - \mathbf{w}\| \ : \mathbf{w} \in W\} = \|\mathbf{v} - P_{\mathbb{W}}(\mathbf{v})\|.$$

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 5.5.4.16. 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 $1 \le i \le k$,
2. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for each $\mathbf{v}_i \in \mathbb{W}_i, \mathbf{v}_j \in \mathbb{W}_j, 1 \le i \ne j \le k$ and

3. $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_k$.

Theorem 5.5.4.18. *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, $1 \le i \ne j \le k$, there exist orthogonal projectors $P_{\mathbb{W}_i} : \mathbb{V} \to \mathbb{V}$ of $\mathbb{V}$ onto $\mathbb{W}_i$ satisfying the following:*

1. $\text{NULL}(P_{\mathbb{W}_i}) = W_i^{\perp} = \mathbb{W}_1 \oplus \mathbb{W}_2 \oplus \cdots \oplus W_{i-1} \oplus \mathbb{W}_{i+1} \oplus \cdots \oplus W_k$.

2. $\text{RNG}(P_{\mathbb{W}_i}) = \mathbb{W}_i$.

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. $I_{\mathbb{V}} = P_{\mathbb{W}_1} \oplus P_{\mathbb{W}_2} \oplus \cdots \oplus P_{\mathbb{W}_k}$.

5.5 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 5.5.5.1 (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 $\det(A) \ne 0$ 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

$$LS(A[:,1], \ldots, A[:,i]) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i), \quad \text{for } 1 \le i \le n.$$

Since $A[:,i] \in LS(\mathbf{v}_1, \ldots, \mathbf{v}_i)$, for $1 \le i \le n$, there exist $\alpha_{ji} \in \mathbb{R}, 1 \le j \le i$, 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 5.5.1.32.4),

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 $1 \le i \le n$, 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 2.2.3.30.1, 1.1.3.2.2, the matrix $R_2R_1^{-1}$ is an upper triangular matrix.
2. Exercises 1.1.3.2.11, $Q_2^{-1}Q_1$ is an orthogonal matrix.

So, the matrix $R_2R_1^{-1}$ is an orthogonal upper triangular matrix and hence, by Exercise 1.1.3.6.17, $R_2R_1^{-1} = I_n$. 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 5.5.2.6, 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

$$LS(A[:,1], \ldots, A[:,j]) = LS(\mathbf{v}_1, \ldots, \mathbf{v}_i), \quad \text{for } 1 \le i \le j \le k.$$

Hence, proceeding on the lines of the above theorem, we have the following result.

Theorem 5.5.5.2 (Generalized QR Decomposition)**.** *Let A be an $n \times k$ matrix of rank r. Then $A = QR$, where*

1. *$Q = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ 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 5.5.5.3. 1. Let $A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. Find an orthogonal matrix Q and an upper triangular matrix R such that $A = QR$.

Solution: From Example 5.5.2.4, 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) $\text{Rank}(A) = 3$,

   (b) $A = QR$ with $Q^TQ = I_3$, and

   (c) R is a 3×4 upper triangular matrix with $\text{Rank}(R) = 3$.

Remark 5.5.5.4. *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 Rank$(A) = n$.*

   (a) *Then A^TA is invertible (see Exercise 5.5.1.27.4).*

   (b) *By Theorem 5.5.5.2, $A = QR$ 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 4.4a, 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 5.5.4.7, 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.*

5.6 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 $\text{Col}(A)$ or $\text{Null}(A)$ 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. $\dim(\text{Null}(A)) + \dim(\text{Col}(A)) = n$.
2. $\text{Null}(A) = \big(\text{Col}(A^*)\big)^\perp$ and $\text{Null}(A^*) = \big(\text{Col}(A)\big)^\perp$.
3. $\dim(\text{Col}(A)) = \dim(\text{Col}(A^*))$.

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. $\mathbf{b} \in \text{Col}(A)$ then we can use the Gauss-Jordan method to get a solution.
2. $\mathbf{b} \notin \text{Col}(A)$ 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 $\text{Col}(A)$. 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 $\text{Col}(A)$.

Chapter 6

Eigenvalues, Eigenvectors and Diagonalization

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 6.6.1.1. 1. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$. Then,

(a) A magnifies the nonzero vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ three times as $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Verify that $B \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and hence B magnifies 5 times.

(b) A behaves by changing the direction of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ as $\begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, whereas B fixes it.

(c) $\mathbf{x}^T A \mathbf{x} = 3\dfrac{(x+y)^2}{2} - \dfrac{(x-y)^2}{2}$ and $\mathbf{x}^T B \mathbf{x} = 5\dfrac{(x+y)^2}{2} + \dfrac{(x-y)^2}{2}$. So, maximum and minimum displacement occurs along lines $x + y = 0$ and $x - y = 0$, where $x + y = (x, y) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $x - y = (x, y) \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

(d) the curve $\mathbf{x}^T A \mathbf{x} = 1$ represents a hyperbola, where as the curve $\mathbf{x}^T B \mathbf{x} = 1$ represents an ellipse (see Figure 6.1 drawn using the package "MATHEMATICA").

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 $\begin{bmatrix} 1 + \sqrt{3} & -2 \\ -1 & \sqrt{3} - 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$

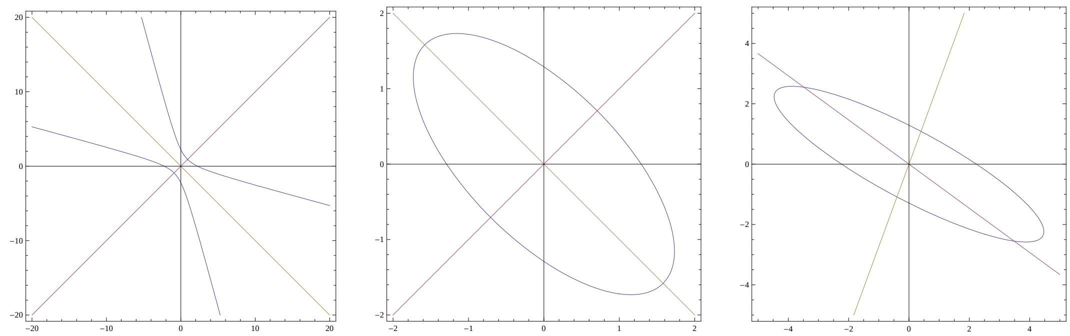


Figure 6.1: A Hyperbola and two Ellipses (first one has orthogonal axes)

.

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

$$L(\mathbf{x},\lambda) = \mathbf{x}^T A\mathbf{x} - \lambda(\mathbf{x}^T\mathbf{x}-1) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j - \lambda\left(\sum_{i=1}^{n} x_i^2 \quad -1\right).$$

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 6.6.1.1.

1. $\det(A) = -3 = 3 \times -1$, $\det(B) = 5 = 5 \times 1$ and $\det(C) = 1 = (2+\sqrt{3}) \times (2-\sqrt{3})$.

2. $\text{Tr}(A) = 2 = 3 - 1$, $\text{Tr}(B) = 6 = 5 + 1$ and $\det(C) = 4 = (2 + \sqrt{3}) + (2 - \sqrt{3})$.
3. Both the sets $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ 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) $AS = [A\mathbf{v}_1, A\mathbf{v}_2] = [3\mathbf{v}_1, -\mathbf{v}_2] = S \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \Leftrightarrow S^{-1}AS = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} = \text{diag}(3, -1)$.
   (b) $BS = [B\mathbf{v}_1, B\mathbf{v}_2] = [5\mathbf{v}_1, \mathbf{v}_2] = S \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \Leftrightarrow S^{-1}AS = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} = \text{diag}(5, 1)$.
   (c) 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. That is, if $U = [\mathbf{u}_1, \mathbf{u}_2]$ then $I = UU^* = \mathbf{u}_1\mathbf{u}_1^* + \mathbf{u}_2\mathbf{u}_2^*$ and
      i. $A = 3\mathbf{u}_1\mathbf{u}_1^* - \mathbf{u}_2\mathbf{u}_2^*$.
      ii. $B = 5\mathbf{u}_1\mathbf{u}_1^* + \mathbf{u}_2\mathbf{u}_2^*$.
5. 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
$$S^{-1}CS = \begin{bmatrix} 2 + \sqrt{3} & 0 \\ 0 & 2 - \sqrt{3} \end{bmatrix} = \text{diag}(2 + \sqrt{3}, 2 - \sqrt{3}).$$

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

1. the equation
$$A\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow (A - \lambda I_n)\mathbf{x} = \mathbf{0} \tag{6.6.1.1}$$
   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 a non-zero vector $\mathbf{x}$ satisfying $A\mathbf{x} = \alpha\mathbf{x}$.
3. a non-zero vector $\mathbf{x}$ satisfying Equation (6.6.1.1) 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 **eigen-space** or **characteristic vector space** of A corresponding to α.

Theorem 6.6.1.3. *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 2.2.2.34 this holds if and only if $\det(A - \alpha I) = 0$. ∎

Definition 6.6.1.4. 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.

We thus observe the following.

Remark 6.6.1.5. *1. Let $A \in \mathbb{M}_n(\mathbb{C})$. If $\alpha \in \mathbb{C}$ is a root of $p_A(\lambda) = 0$ then α is an eigenvalue. As* $\text{NULL}(A - \alpha I)$ *is a subspace of $\mathbb{C}^n$, the following statements hold.*

(a) $(\alpha, \mathbf{x})$ is an eigen-pair of A if and only if $(\alpha, c\mathbf{x})$ is an eigen-pair of A, for $c \in \mathbb{C} \setminus \{0\}$.

(b) If $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are linearly independent eigenvectors of A for the eigenvalue α then $\sum\limits_{i=1}^{r} c_i \mathbf{x}_i$, with at least one $c_i \neq 0$, is also an eigenvector of A for α.

Hence, if S is a collection of eigenvectors, S needs to be LINEARLY INDEPENDENT.

2. $A - \alpha I$ is singular. Therefore, if $\text{RANK}(A - \alpha I) = r$ *then $r < n$. Hence, by Theorem 2.2.2.34, the system $(A - \alpha I)\mathbf{x} = \mathbf{0}$ has $n - r$ linearly independent solutions.*

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 6.6.1.6. *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 6.6.1.7. 1. Let $A = \text{diag}(d_1, \ldots, d_n)$ with $d_i \in \mathbb{C}, 1 \leq i \leq n$. 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 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then $p(\lambda) = (1 - \lambda)^2$. Hence, 1 is a repeated eigenvalue. 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 6.6.1.5.1, $\mathbf{e}_1$ is an eigenvector. Therefore, 1 IS A REPEATED EIGENVALUE WHEREAS THERE IS ONLY ONE EIGENVECTOR.

3. 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 $\mathbf{x}^t, \mathbf{y}^t$ from $\mathbb{C}^2$ 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.

4. Let $A = \begin{bmatrix} 1 & 3 \\ -2 & 5 \end{bmatrix}$. Then, $\left(3 + \sqrt{2}i, \begin{bmatrix} \frac{3}{2+\sqrt{2}i} \\ 1 \end{bmatrix}\right)$ and $\left(3 - \sqrt{2}i, \begin{bmatrix} \frac{3}{2-\sqrt{2}i} \\ 1 \end{bmatrix}\right)$ are the eigen-pairs of A.

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. Verify that $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ has the eigenvalue 0 repeated 3 times with $\mathbf{e}_1$ as the only eigenvector as $A\mathbf{x} = \mathbf{0}$ with $\mathbf{x} = (x_1, x_2, x_3)^T$ implies $x_2 = 0 = x_3$.

7. Verify that $A = \left[\begin{array}{ccc|cc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$ has the eigenvalue 0 repeated 5 times with $\mathbf{e}_1$ and $\mathbf{e}_4$ as the only eigenvectors as $A\mathbf{x} = \mathbf{0}$ with $\mathbf{x} = (x_1, x_2, x_3)^T$ implies $x_2 = 0 = x_3 = x_5$. Note that the diagonal blocks of A are nilpotent matrices.

EXERCISE **6.6.1.8.** *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. What is the corresponding eigenvector?

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.

5. Let A be an idempotent matrix. Then prove that its eigenvalues are either 0 or 1 or both.

6. Let A be a nilpotent matrix. Then prove that its eigenvalues are all 0.

Theorem 6.6.1.9. *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* $\mathrm{TR}(A) = \sum\limits_{i=1}^{n} a_{ii} = \sum\limits_{i=1}^{n} \lambda_i$.

Proof. Since $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, by definition,

$$\det(A - xI_n) = (-1)^n \prod_{i=1}^{n} (x - \lambda_i) \tag{6.6.1.2}$$

is an identity in x as polynomials. Therefore, by substituting $x = 0$ in Equation (6.6.1.2), we get $\det(A) = (-1)^n(-1)^n \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i$. Also,

$$\begin{aligned} \det(A - xI_n) &= \begin{bmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{bmatrix} && (6.6.1.3) \\ &= a_0 - xa_1 + \cdots + (-1)^{n-1}x^{n-1}a_{n-1} + (-1)^n x^n && (6.6.1.4) \end{aligned}$$

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, $a_{n-1} = \sum\limits_{i=1}^{n} a_{ii} = \mathrm{Tr}(A)$, the trace of A. Also, from Equation (6.6.1.2) and (6.6.1.4), 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

$$\mathrm{Tr}(A) = a_{n-1} = (-1)\left\{(-1)\sum_{i=1}^{n} \lambda_i\right\} = \sum_{i=1}^{n} \lambda_i.$$

Hence, we get the required result. ■

Exercise 6.6.1.10. 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.*

6.1.A Spectrum of an eigenvalue

Definition 6.6.1.11. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then,

1. the collection of eigenvalues of A, counting with multiplicities, is called the **spectrum** of A, denoted $\sigma(A)$.
2. the multiplicity of $\alpha \in \sigma(A)$ is called the **algebraic multiplicity** of A, denoted $\mathrm{Alg.Mul}_\alpha(A)$.
3. for $\alpha \in \sigma(A)$, $\dim(\mathrm{Null}(A - \alpha I))$ is called the **geometric multiplicity** of A, $\mathrm{Geo.Mul}_\alpha(A)$.

We now state the following observations.

Remark 6.6.1.12. *Let $A \in \mathbb{M}_n(\mathbb{C})$.*

1. *Then, for each $\alpha \in \sigma(A)$, using Theorem 2.2.2.34* $\dim(\textsc{Null}(A - \alpha I)) \geq 1$. *So, we have at least one eigenvector.*

2. *If the algebraic multiplicity of $\alpha \in \sigma(A)$ is $r \geq 2$ then the Example 6.6.1.7.7 implies that we need not have r linearly independent eigenvectors.*

Theorem 6.6.1.13. *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 $\alpha \in \sigma(A)$,* $\textsc{Alg.Mul}_\alpha(A) = \textsc{Alg.Mul}_\alpha(B)$ *and* $\textsc{Geo.Mul}_\alpha(A) = \textsc{Geo.Mul}_\alpha(B)$.

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

$$\begin{aligned} \det(A - xI) &= \det(SBS^{-1} - xI) = \det\left(S(B - xI)S^{-1}\right) \\ &= \det(S)\det(B - xI)\det(A^{-1}) = \det(B - xI). \end{aligned} \tag{6.6.1.5}$$

Note that Equation (6.6.1.5) also implies that $\textsc{Alg.Mul}_\alpha(A) = \textsc{Alg.Mul}_\alpha(B)$. We will now show that $\textsc{Geo.Mul}_\alpha(A) = \textsc{Geo.Mul}_\alpha(B)$.

So, let $Q_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of $\textsc{Null}(A - \alpha I)$. Then, $B = SAS^{-1}$ implies that $Q_2 = \{S\mathbf{v}_1, \ldots, S\mathbf{v}_k\} \subseteq \textsc{Null}(B - \alpha I)$. Since Q_1 is linearly independent and S is invertible, we get Q_2 is linearly independent. So, $\textsc{Geo.Mul}_\alpha(A) \leq \textsc{Geo.Mul}_\alpha(B)$. Now, we can start with eigenvectors of B and use similar arguments to get $\textsc{Geo.Mul}_\alpha(B) \leq \textsc{Geo.Mul}_\alpha(A)$ and hence the required result follows. ∎

Remark 6.6.1.14. *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*

1. *the same set of eigenvalues.*

2. $\textsc{Alg.Mul}_\alpha(A) = \textsc{Alg.Mul}_\alpha(B)$, *for each $\alpha \in \sigma(A)$.*

3. $\textsc{Geo.Mul}_\alpha(A) = \textsc{Geo.Mul}_\alpha(B)$, *for each $\alpha \in \sigma(A)$.*

We will now give a relation between the geometric multiplicity and the algebraic multiplicity.

Theorem 6.6.1.15. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, for $\alpha \in \sigma(A)$,* $\textsc{Geo.Mul}_\alpha(A) \leq \textsc{Alg.Mul}_\alpha(A)$.

Proof. Let $\textsc{Geo.Mul}_\alpha(A) = k$. Suppose $Q_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthonormal basis of $\textsc{Null}(A - \alpha I)$. 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

$$\begin{aligned} P^*AP &= P^*\left[A\mathbf{v}_1, \ldots, A\mathbf{v}_k, A\mathbf{v}_{k+1}, \ldots, A\mathbf{v}_n\right] \\ &= \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_k^* \\ \mathbf{v}_{k+1}^* \\ \vdots \\ \mathbf{v}_n^* \end{bmatrix} [\alpha\mathbf{v}_1, \ldots, \alpha\mathbf{v}_k, *, \ldots, *] = \left[\begin{array}{ccc|ccc} \alpha & \cdots & 0 & * & \cdots & * \\ 0 & \ddots & 0 & * & \cdots & * \\ 0 & \cdots & \alpha & * & \cdots & * \\ \hline 0 & \cdots & 0 & * & \cdots & * \\ \vdots & & & & & \\ 0 & \cdots & 0 & * & \cdots & * \end{array}\right]. \end{aligned}$$

Now, if we denote the lower diagonal submatrix as D then

$$p_A(x) = \det(A - xI) = \det(P^*AP - xI) = (\alpha - x)^k \det(D - xI).$$

So, $\text{ALG.MUL}_\alpha(A) = \text{ALG.MUL}_\alpha(P^*AP) \geq k = \text{GEO.MUL}_\alpha(A)$. ∎

EXERCISE **6.6.1.16.** *1. 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 $\alpha \neq 0$ then
 i. $\alpha \in \sigma(BA)$.
 ii. $\text{ALG.MUL}_\alpha(AB) = \text{ALG.MUL}_\alpha(BA)$.
 iii. $\text{GEO.MUL}_\alpha(AB) = \text{GEO.MUL}_\alpha(BA)$.

(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 $n = m$.

2. 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 6.6.1.16.1a].

(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.

3. Let $A, B \in \mathbb{M}_2(\mathbb{R})$ such that $\det(A) = \det(B)$ and $\text{TR}(A) = \text{TR}(B)$.

(a) Do A and B have the same set of eigenvalues?

(b) Give examples to show that the matrices A and B need not be similar.

4. Let $A, B \in \mathbb{M}_n(\mathbb{R})$. Also, let $(\lambda_1, \mathbf{u})$ 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 $A + B$.

5. 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 $T[\mathcal{B}, \mathcal{B}]$ 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 6.6.2.1. [Matrix Digitalization] 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 6.6.2.2.
1. 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 6.6.1.13, $\{d_1, d_2\} = \sigma(D) = \sigma(A) = \{0, 0\}$. Hence, $D = \mathbf{0}$ and therefore, $A = SDS^{-1} = \mathbf{0}$, a contradiction.

2. 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 6.6.1.13, $\{d_1, d_2, d_3\} = \sigma(D) = \sigma(A) = \{2, 2, 2\}$. Hence, $D = 2I_3$ and therefore, $A = SDS^{-1} = 2I_3$, a contradiction.

3. 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 6.6.2.3. *Let $A \in \mathbb{M}_n(\mathbb{R})$.*
1. *Let S be an invertible matrix such that $S^{-1}AS = diag(d_1, \ldots, d_n)$. 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 Theorem 3.3.2.8, $S = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ 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. ∎

Theorem 6.6.2.4. *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,

$$\alpha_{\ell+1}\mathbf{v}_{\ell+1} = \alpha_{\ell+1}\beta_1\mathbf{v}_1 + \cdots + \alpha_{\ell+1}\beta_\ell\mathbf{v}_\ell. \tag{6.6.2.1}$$

and

$$\alpha_{\ell+1}\mathbf{v}_{\ell+1} \quad = \quad A\mathbf{v}_{\ell+1} = A\left(\beta_1\mathbf{v}_1 + \cdots + \beta_\ell\mathbf{v}_\ell\right) = \alpha_1\beta_1\mathbf{v}_1 + \cdots + \alpha_\ell\beta_\ell\mathbf{v}_\ell. \tag{6.6.2.2}$$

Now, subtracting Equation (6.6.2.2) from Equation (6.6.2.1), 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 6.6.2.3 and Theorem 6.6.2.4 is stated next without proof.

Corollary 6.6.2.5. *Let $A \in \mathbb{M}_n(\mathbb{C})$ have n distinct eigenvalues. Then A is diagonalizable.*

The converse of Theorem 6.6.2.4 is not true as I_n has n linearly independent eigenvectors corresponding to the eigenvalue 1, repeated n times.

Corollary 6.6.2.6. *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

$$p_j(A)\mathbf{u} = \begin{cases} \mathbf{0}, & \text{if } \mathbf{u} \in \textsc{Null}(A - \alpha_i I_n), \text{ for some } i \neq j \\ \prod\limits_{i \neq j}(\alpha_j - \alpha_i)\mathbf{u} & \text{if } \mathbf{u} \in \textsc{Null}(A - \alpha_j I_n) \end{cases} \tag{6.6.2.3}$$

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 unknowns c_{ij}'s. Now, applying the matrix $p_j(A)$ and using Equation (6.6.2.3), 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 6.6.2.7. *Let $A \in \mathbb{M}_n(\mathbb{C})$ with distinct eigenvalues $\alpha_1, \ldots, \alpha_k$. Then A is diagonalizable if and only if $\textsc{Geo.Mul}_{\alpha_i}(A) = \textsc{Alg.Mul}_{\alpha_i}(A)$, for each $1 \leq i \leq k$.*

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 6.6.2.6 A has $\sum\limits_{i=1}^{k} n_i$ linearly independent eigenvectors. Also, by Theorem 6.6.1.15, $n_i \le m_i$, for $1 \le i \le m_i$.

Now, let A be diagonalizable. Then, by Theorem 6.6.2.3, 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 6.6.2.3, A is diagonalizable. ■

Example 6.6.2.8. 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 6.6.2.3, A is not diagonalizable.

EXERCISE **6.6.2.9.** *1. Is the matrix* $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ *diagonalizable?*

2. *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.*

3. *Let J_n be an $n \times n$ matrix with all entries 1. Then, prove that $\text{GEO.MUL}_1(J_n) = \text{ALG.MUL}_1(J_n) = 1$ and $\text{GEO.MUL}_0(J_n) = \text{ALG.MUL}_0(J_n) = n - 1$.*

4. *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?*

5. *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.

6. *Let $A \in \mathbb{M}_n(\mathbb{R})$ with $A \ne \mathbf{0}$ but $A^2 = \mathbf{0}$. Prove that A cannot be diagonalized.*

7. *Are the following matrices diagonalizable?*
$$i) \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 2 & 3 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad ii) \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix}, \quad iii) \begin{bmatrix} 1 & -3 & 3 \\ 0 & -5 & 6 \\ 0 & -3 & 4 \end{bmatrix} \text{ and } iv) \begin{bmatrix} 2 & i \\ i & 0 \end{bmatrix}.$$

6.2.A Schur's Unitary Triangularization

We now prove one of the most important results in diagonalization, called the Schur's Lemma or Schur's unitary triangularization.

Lemma 6.6.2.10 (Schur's unitary triangularization (SUT)). *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then there exists a unitary matrix U such that A is an upper triangular matrix. Further, if $A \in \mathbb{M}_n(\mathbb{R})$ and $\sigma(A)$ have real entries then U is 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

$$X^*AX = X^*[A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n] = \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \\ \vdots \\ \mathbf{x}_n^* \end{bmatrix} [\lambda_1\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n] = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & B \end{bmatrix}, \qquad (6.6.2.4)$$

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 7.7, 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 6.6.2.11. *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,*

$$\{\alpha_1, \ldots, \alpha_n\} = \sigma(A) = \sigma(U^*AU) = \{t_{11}, \ldots, t_{nn}\}. \qquad (6.6.2.5)$$

Furthermore, we can get the α_i's in the diagonal of T in any prescribed order.

Definition 6.6.2.12. [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 $A = U^*BU$.

EXERCISE **6.6.2.13.** *Use the exercises given below to conclude that the upper triangular matrix obtained in the "Schur's Lemma" need not be unique.*

1. *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.*

2. *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.*

3. *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*

   (a) A_1 *and* D *are unitarily equivalent.*

   (b) A_2 *and* B *are unitarily equivalent.*

   (c) *Do the above results contradict Exercise 5.5.1.32.6.6c? Give reasons for your answer.*

4. *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.*

5. *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?*

6. *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?*

7. *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* $A = U^* B U$?

Remark 6.6.2.14. *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 6.6.2.13.7). In numerical calculations, unitary transformations are preferred as compared to similarity transformations due to the following main reasons:*

1. *Exercise 5.5.1.32.6.6c? 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.*

We use Lemma 6.6.2.10 to give another proof of Theorem 6.6.1.9.

Corollary 6.6.2.15. *Let* $A \in \mathbb{M}_n(\mathbb{C})$. *If* $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\}$ *then* $\det(A) = \prod_{i=1}^{n} \alpha_i$ *and* $\mathrm{Tr}(A) = \sum_{i=1}^{n} \alpha_i$.

Proof. By Schur's Lemma there exists a unitary matrix U such that $U^*AU = T = [t_{ij}]$, a triangular matrix. By Remark 6.6.2.11, $\sigma(A) = \sigma(T)$. Hence, $\det(A) = \det(T) = \prod_{i=1}^{n} t_{ii} = \prod_{i=1}^{n} \alpha_i$ and $\text{Tr}(A) = \text{Tr}(A(UU^*)) = \text{Tr}(U^*(AU)) = \text{Tr}(T) = \sum_{i=1}^{n} t_{ii} = \sum_{i=1}^{n} \alpha_i$. ∎

6.2.B 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 6.6.2.16 (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 $U^*AU = diag(\alpha_1, \ldots, \alpha_n)$.*

Proof. By Schur's Lemma there exists a unitary matrix U such that $U^*AU = T = [t_{ij}]$, a triangular matrix. Since A is upper triangular, we see that

$$T^*T = (U^*AU)^*(U^*AU) = U^*A^*AU = U^*AA^*U = (U^*AU)(U^*AU)^* = TT^*.$$

Thus, we see that T is an upper triangular matrix with $T^*T = TT^*$. Thus, by Exercise 1.1.3.6.17, T is a diagonal matrix and this completes the proof. ∎

Exercise **6.6.2.17.** *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 6.6.2.16 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 6.6.2.18. *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 6.6.2.3).*
   (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 $W_i = \text{Null}(A - \alpha_i I_n)$, 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 6.6.2.19. *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 such that $U^*AU = D$, where $D =$ diag$(\alpha_1, \ldots, \alpha_n)$.*

Proof. The second part is immediate from Theorem 6.6.2.16 and hence the proof is omitted. 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 $\lambda = \overline{\lambda}$. That is, $\lambda \in \mathbb{R}$. ■

As an immediate corollary of Theorem 6.6.2.19 and the second part of Lemma 6.6.2.10, we give the following result without proof.

Corollary 6.6.2.20. *Let $A \in \mathbb{M}_n(\mathbb{R})$ be symmetric. Then $A = U\,\text{diag}(\alpha_1, \ldots, \alpha_n)U^*$, 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 **6.6.2.21.**

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. *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 $\|A^*\mathbf{x} - \overline{\lambda}\mathbf{x}\|^2 = \|A\mathbf{x} - \lambda\mathbf{x}\|^2$.]*
4. *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.*
5. *Let A be a 2×2 orthogonal matrix. Then prove the following:*
   (a) *if $\det(A) = 1$ then $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, for some $\theta, 0 \leq \theta < 2\pi$. 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 $\theta, 0 \leq \theta < 2\pi$. That is, A reflects every point in $\mathbb{R}^2$ about a line passing through origin. Determine this line. Or equivalently, there exists a non-singular matrix P such that $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.*
6. *Let A be a 3×3 orthogonal matrix. Then prove the following:*

(a) *if* $\det(A) = 1$ *then* A *is a rotation about a fixed axis, in the sense that* A *has an eigen-pair* $(1, \mathbf{x})$ *such that the restriction of* A *to the plane* $\mathbf{x}^{\perp}$ *is a two dimensional rotation in* $\mathbf{x}^{\perp}$.

(b) *if* $\det A = -1$ *then* A *corresponds to a reflection through a plane* P, *followed by a rotation about the line through origin that is orthogonal to* P.

7. *Let* A *be a normal matrix. Then prove that* $\text{Rank}(A)$ *equals the number of non-zero eigenvalues of* A.

6.2.C Cayley Hamilton Theorem

Let $A \in \mathbb{M}_n(\mathbb{C})$. Then, in Theorem 6.6.1.9, we saw that

$$p_A(x) = \det(A - xI) = (-1)^n \left(x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + (-1)^{n-1}a_1x + (-1)^n a_0\right) \tag{6.6.2.6}$$

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, $x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + (-1)^{n-1}a_1x + (-1)^n a_0 = 0$ is satisfied by n complex numbers. It turns out that the expression

$$A^n - a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \cdots + (-1)^{n-1}a_1A + (-1)^n a_0 I = \mathbf{0}$$

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 6.6.2.22. *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,* $A_1A_2\cdots A_n = \mathbf{0}$.

Proof. We use induction to prove that the first k columns of $A_1A_2\cdots A_k$ 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_1A_2 is $\mathbf{0}$. Let $B = A_1A_2$. 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 $j \ge 2$. 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 $1 \le i \le n$, 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 $1 \le j \le n-1$ and $(A_n)_{ni} = 0$, for $i = n-1, n$. Thus, by induction hypothesis the required result follows. ∎

We now prove the Cayley Hamilton Theorem using Schur's unitary triangularization.

Theorem 6.6.2.23 (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*

$$A^n - a_{n-1}A^{n-1} + a_{n-2}A^{n-2} + \cdots + (-1)^{n-1}a_1A + (-1)^n a_0 I = \mathbf{0}$$

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 $1 \le i \le n$. Now, observe that if $A_i = T - \alpha_i I$ then the A_i's satisfy the conditions of Lemma 6.6.2.22. 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 6.6.2.24. *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 $A^{-1} = \frac{1}{5}(A + 2I_2) = \frac{1}{5}\begin{bmatrix} 3 & 2 \\ 1 & -1 \end{bmatrix}$.

2. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. *Then* $p_A(x) = x^2$. *So, even though* $A \neq \mathbf{0}$, $A^2 = \mathbf{0}$.

3. 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* $A^2 = \mathbf{0}$.

4. 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* $1 \le i \le n-1$. *This is quite helpful in numerical computation as computing powers takes much more time that division.*

ii. Note that $LS\{I, A, A^2, \ldots\}$ *is a subspace of* $\mathbb{M}_n(\mathbb{C})$. *Also,* $\dim(\mathbb{M}_n(\mathbb{C})) = n^2$. *But, the above argument implies that* $\dim\left(LS\{I, A, A^2, \ldots\}\right) \le n$.

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 **6.6.2.25.** *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 **6.6.2.26.** *Miscellaneous Exercises:*

1. *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.*

2. *Let B and C be $n \times n$ matrices and $A = \begin{bmatrix} B & C \\ -C & B \end{bmatrix}$. 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) *$\det(A) = |\det(B + iC)|^2$.*

We end this chapter with an application to the study of conic sections in analytic geometry.

6.3 Quadratic Forms

Definition 6.6.3.1. Let $A \in \mathbb{M}_n(\mathbb{C})$. Then A is said to be

1. **positive semi-definite**(psd) if $\mathbf{x}^* A \mathbf{x} \geq 0$, for all $\mathbf{x} \in \mathbb{C}^n$.
2. **positive definite**(pd) if $\mathbf{x}^* A \mathbf{x} > 0$, for all $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.
3. **negative semi-definite**(psd) if $\mathbf{x}^* A \mathbf{x} \leq 0$, for all $\mathbf{x} \in \mathbb{C}^n$.
4. **negative definite**(pd) if $\mathbf{x}^* A \mathbf{x} < 0$, for all $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.

Remark 6.6.3.2. *Let $A = [a_{ij}] \in \mathbb{M}_n(\mathbb{C})$ be positive semi-definite (positive definite/negative semi-definite or negative definite) matrix. Then, A is Hermitian.*

Solution: *By definition, $a_{ii} = \mathbf{e}_i^* A \mathbf{e}_i \in \mathbb{R}$. 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, $Im(a_{ij}) = -Im(a_{ji})$. 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 $Re(a_{ij}) = Re(a_{ji})$.*

Example 6.6.3.3. 1. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Then, A is positive definite.

2. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then, A is positive semi-definite but not positive definite.

3. Let $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$. Then, A is negative definite.

4. Let $A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$. Then, A is negative semi-definite.

5. Let $A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$. Then, A is neither positive semi-definite nor positive definite nor negative semi-definite and nor negative definite.

Theorem 6.6.3.4. *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 Remark 6.6.3.2 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 $D = \text{diag}(\alpha_1, \ldots, \alpha_n)$. As $\alpha_i \geq 0$, for $1 \leq i \leq n$, define $D^{\frac{1}{2}} = \text{diag}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n})$. 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 give the next result and hence the proof is omitted.

Theorem 6.6.3.5. *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 $B \in \mathbb{M}_n(\mathbb{C})$.*

Definition 6.6.3.6. 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 6.6.3.7. [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 the **Hermitian form**. In case $A \in \mathbb{M}_n(\mathbb{R})$, $H(\mathbf{x})$ is called the **quadratic** form.

Remark 6.6.3.8. *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 Hermitian form $H(\mathbf{x})$ is a real number. Hence, for $\alpha \in \mathbb{R}$, the equation $H(\mathbf{x}) = \alpha$, represents a conic in $\mathbb{C}^n$.*

Example 6.6.3.9. 1. Let $\mathbf{v}_i \in \mathbb{C}^n$, for $1 \le i \le n$. 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, $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T A\mathbf{x}$, 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.A 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. Also, in Example 6.6.1.1, we expressed $\mathbf{x}^T A\mathbf{x} = 3\dfrac{(x+y)^2}{2} - \dfrac{(x-y)^2}{2}$ and $\mathbf{x}^T B\mathbf{x} = 5\dfrac{(x+y)^2}{2} + \dfrac{(x-y)^2}{2}$. But, we can also express them as $\mathbf{x}^T A\mathbf{x} = 2(x+y)^2 - (x^2+y^2)$ and $\mathbf{x}^T B\mathbf{x} = 2(x+y)^2 + (x^2+y^2)$. Note that the first expression clearly gives the direction of maximum and minimum displacements or the axes of the curves that they represent whereas such deductions cannot be made from the other expression. So, in this subsection, we proceed to clarify these ideas.

Let $A \in \mathbb{M}_n(\mathbb{C})$ be Hermitian. Then, by Theorem 6.6.2.19, $\sigma(A) = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{R}$ and there exists a unitary matrix U such that $U^*AU = D = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$. 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
$$H(\mathbf{x}) = \mathbf{z}^* U^* AU\mathbf{z} = \mathbf{z}^* D\mathbf{z} = \sum_{i=1}^{n} \lambda_i |\mathbf{z}_i|^2 = \sum_{i=1}^{p} \left|\sqrt{|\lambda_i|}\ \mathbf{z}_i\right|^2 - \sum_{i=p+1}^{r} \left|\sqrt{|\lambda_i|}\ \mathbf{z}_i\right|^2. \tag{6.6.3.7}$$

Thus, the possible values of $H(\mathbf{x})$ depend only on the eigenvalues of A. Since U is an invertible matrix, the components $\mathbf{z}_i$'s of $\mathbf{z} = U^*\mathbf{x}$ 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 (6.6.3.7), p corresponds to the number of positive eigenvalues and $r-p$ to the number of negative eigenvalues. So, 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 6.6.3.10 (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 (6.6.3.7) 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

$$\begin{aligned} H(\mathbf{x}) &= |\mathbf{y}_1|^2 + \cdots + |\mathbf{y}_p|^2 - |\mathbf{y}_{p+1}|^2 - \cdots - |\mathbf{y}_r|^2 \\ &= |\mathbf{z}_1|^2 + \cdots + |\mathbf{z}_q|^2 - |\mathbf{z}_{q+1}|^2 - \cdots - |\mathbf{y}_s|^2, \end{aligned} \tag{6.6.3.8}$$

where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^* = M\mathbf{x}$ and $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^* = N\mathbf{x}$ for some invertible matrices M and N. Hence, $\mathbf{z} = B\mathbf{y}$, for $B = NM^{-1}$, an invertible matrix. Let us write $Y_1 = (\mathbf{y}_1, \ldots, \mathbf{y}_p)^*$, $Z_1 = (\mathbf{z}_1, \ldots, \mathbf{z}_q)^*$ and $B = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$, where B_1 is a $q \times p$ matrix. As $p > q$, the homogeneous linear system $B_1 Y_1 = \mathbf{0}$ has a non-zero solution, say $\widetilde{Y_1} = (\tilde{y_1}, \ldots, \tilde{y_p})^*$ and let $\tilde{\mathbf{y}} = \begin{bmatrix} \widetilde{Y_1} \\ \mathbf{0}^* \end{bmatrix}$. Then $Z_1 = \mathbf{0}$ and thus, using Equation (6.6.3.8), we have

$$H(\tilde{\mathbf{y}}) = |\tilde{y_1}|^2 + |\tilde{y_2}|^2 + \cdots + |\tilde{y_p}|^2 = -(|z_{q+1}|^2 + \cdots + |z_s|^2).$$

Now, this can hold only if $\tilde{y_1} = \cdots = \tilde{y_p} = 0$, a contradiction. Hence $p = q$. Similarly, the case $r > s$ can be resolved. Thus, the proof of the lemma is over. ☐

Remark 6.6.3.11. *Since A is Hermitian,* $\text{RANK}(A)$ *equals the number of non-zero 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 $H(\mathbf{x})$.*

Definition 6.6.3.12. [Associate 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 $f(x, y) = 0$.

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.

Proposition 6.6.3.13. *Consider the general quadratic $f(x,y)$, for $a,b,c,g,f,h \in \mathbb{R}$. Then $f(x,y)=0$ represents*

1. *an ellipse or a circle if $ab-h^2>0$,*
2. *a parabola or a pair of parallel lines if $ab-h^2=0$,*
3. *a hyperbola or a pair of intersecting lines if $ab-h^2<0$.*

Proof. As A is symmetric, by Corollary 6.6.2.20, $A = U\text{diag}(\alpha_1,\alpha_2)U^T$, 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. Let $[u,\ v]=\mathbf{x}^TU$. As $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, u and v represent orthogonal lines passing through origin in the (x,y)-plane. In most cases, these lines form the principal axes of the conic.

We also have $\mathbf{x}^TA\mathbf{x}=\alpha_1u^2+\alpha_2v^2$ and hence $f(x,y)=0$ reduces to

$$\lambda_1u^2+\lambda_2v^2+2g_1u+2f_1v+c=0. \tag{6.6.3.9}$$

for some $g_1,f_1\in\mathbb{R}$. Now, we consider different cases depending of the values of α_1,α_2:

1. If $\alpha_1=0=\alpha_2$ then $A=\mathbf{0}$ and Equation (6.6.3.9) gives the straight line $2gx+2fy+c=0$.
2. if $\alpha_1=0$ and $\alpha_2\neq 0$ then $ab-h^2=\det(A)=\alpha_1\alpha_2=0$. So, after dividing by α_2, Equation (6.6.3.9) reduces to $(v+d_1)^2=d_2u+d_3$, for some $d_1,d_2,d_3\in\mathbb{R}$. Hence, let us look at the possible subcases:
   (a) Let $d_2=d_3=0$. Then $v+d_1=0$ is a pair of coincident lines.
   (b) Let $d_2=0,\ d_3\neq 0$.
      i. If $d_3>0$, then we get a pair of parallel lines given by $v=-d_1\pm\sqrt{\frac{d_3}{\alpha_2}}$.
      ii. If $d_3<0$, the solution set of the corresponding conic is an empty set.
   (c) If $d_2\neq 0$. Then the given equation is of the form $Y^2=4aX$ for some translates $X=x+\alpha$ and $Y=y+\beta$ and thus represents a parabola.

   Let $H(\mathbf{x})=x^2+4y^2+4xy$ be the associated quadratic form for a class of curves. Then, $A=\begin{bmatrix}1&2\\2&4\end{bmatrix}$, $\alpha_1=0,\alpha_2=5$ and $v=x+2y$. Now, let $d_1=-3$ and vary d_2 and d_3 to get different curves (see Figure 6.2 drawn using the package "MATHEMATICA").
3. $\alpha_1>0$ and $\alpha_2<0$. Then $ab-h^2=\det(A)=\lambda_1\lambda_2<0$. If $\alpha_2=-\beta_2$, for $\beta_2>0$, then Equation (6.6.3.9) reduces to

$$\alpha_1(u+d_1)^2-\beta_2(v+d_2)^2=d_3,\quad\text{for some}\quad d_1,d_2,d_3\in\mathbb{R} \tag{6.6.3.10}$$

   whose understanding requires the following subcases:
   (a) If $d_3=0$ then Equation (6.6.3.10) equals

$$\left(\sqrt{\alpha_1}(u+d_1)+\sqrt{\beta_2}(v+d_2)\right)\cdot\left(\sqrt{\alpha_1}(u+d_1)-\sqrt{\beta_2}(v+d_2)\right)=0$$

   or equivalently, a pair of intersecting straight lines in the (u,v)-plane.

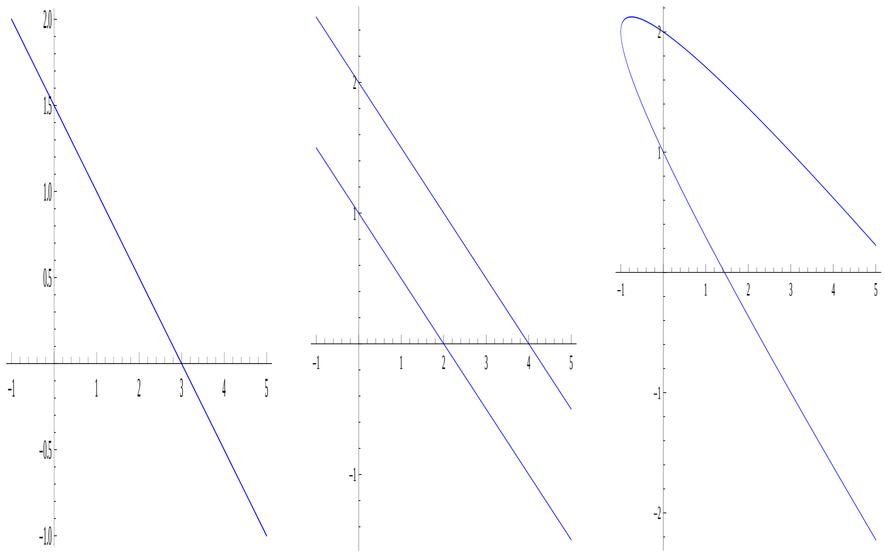


Figure 6.2: Curves for $d_2 = 0 = d_3$, $d_2 = 0, d_3 = 1$ and $d_2 = 1, d_3 = 1$

(b) Without loss of generality, let $d_3 > 0$. Then Equation (6.6.3.10) equals

$$\frac{\lambda_1(u+d_1)^2}{d_3} - \frac{\alpha_2(v+d_2)^2}{d_3} = 1$$

or equivalently, a hyperbola with orthogonal principal axes $u+d_1 = 0$ and $v+d_2 = 0$. Let $H(\mathbf{x}) = 10x^2 - 5y^2 + 20xy$ be the associated quadratic form for a class of curves. Then, $A = \begin{bmatrix} 10 & 10 \\ 10 & -5 \end{bmatrix}$, $\alpha_1 = 15, \alpha_2 = -10$ and $\sqrt{5}u = 2x+y, \sqrt{5}v = x - 2y$. Now, let $d_1 = \sqrt{5}, d_2 = -\sqrt{5}$ to get $3(2x+y+1)^2 - 2(x-2y-1)^2 = d_3$. Now vary d_3 to get different curves (see Figure 6.3 drawn using the package "MATHEMATICA").

4. $\alpha_1, \alpha_2 > 0$. Then, $ab - h^2 = \det(A) = \alpha_1\alpha_2 > 0$ and Equation (6.6.3.9) reduces to

$$\lambda_1(u+d_1)^2 + \lambda_2(v+d_2)^2 = d_3, \quad \text{for some} \quad d_1, d_2, d_3 \in \mathbb{R}. \tag{6.6.3.11}$$

We consider the following three subcases to understand this.

(a) If $d_3 = 0$ then we get a pair of orthogonal lines $u + d_1 = 0$ and $v + d_2 = 0$.

(b) If $d_3 < 0$ then the solution set of Equation (6.6.3.11) is an empty set.

(c) If $d_3 > 0$ then Equation (6.6.3.11) reduces to $\dfrac{\alpha_1(u+d_1)^2}{d_3} + \dfrac{\alpha_2(v+d_2)^2}{d_3} = 1$, an ellipse or circle with $u + d_1 = 0$ and $v + d_2 = 0$ as the orthogonal principal axes.

Let $H(\mathbf{x}) = 6x^2 + 9y^2 + 4xy$ be the associated quadratic form for a class of curves. Then, $A = \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix}$, $\alpha_1 = 10, \alpha_2 = 5$ and $\sqrt{5}u = x+2y, \sqrt{5}v = 2x-y$. Now, let $d_1 = \sqrt{5}, d_2 = -\sqrt{5}$ to get $2(x+2y+1)^2+(2x-y-1)^2 = d_3$. Now vary d_3 to get different curves (see Figure 6.4 drawn using the package "MATHEMATICA").

Thus, we have considered all the possible cases and the required result follows. ■

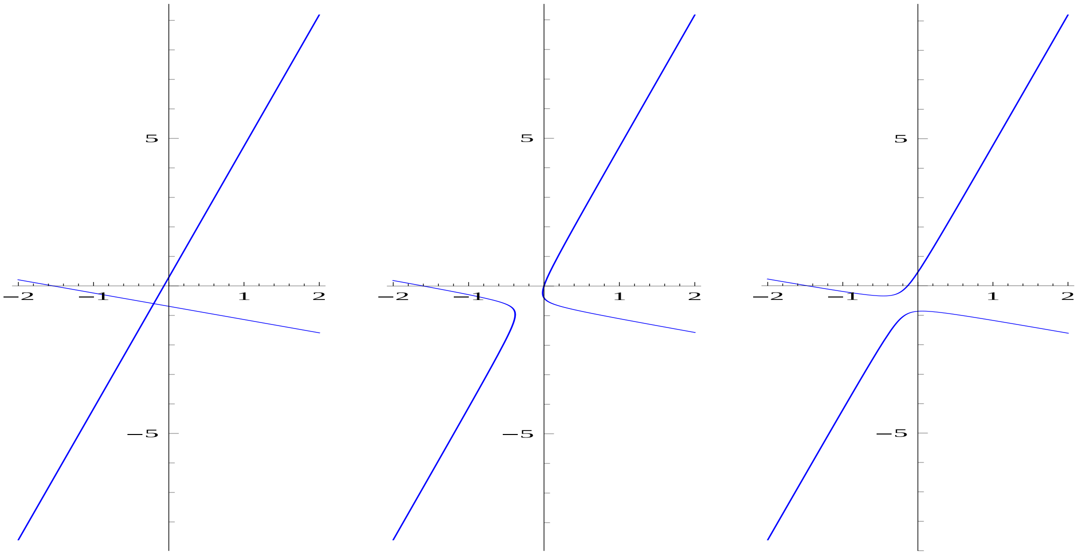


Figure 6.3: Curves for $d_3 = 0$, $d_3 = 1$ and $d_3 = -1$

.

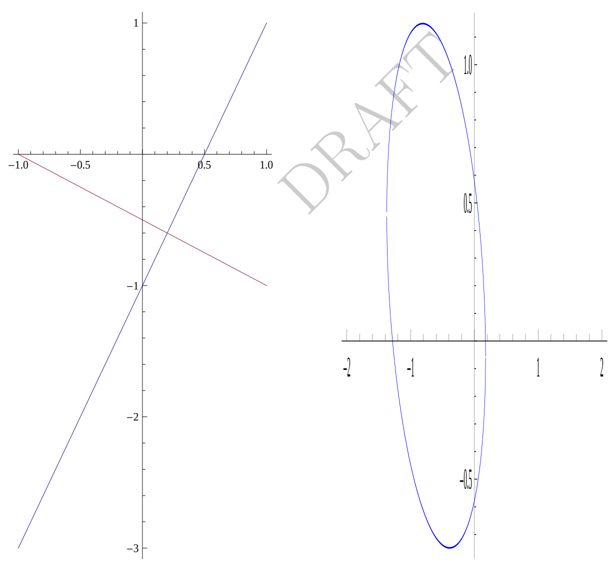


Figure 6.4: Curves for $d_3 = 0$ and $d_3 = 5$

.

Remark 6.6.3.14. *Observe that the condition* $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$ *implies that the principal axes of the conic are functions of the eigenvectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$.

EXERCISE **6.6.3.15.** *Sketch the graph of the following surfaces:*

1. $x^2 + 2xy + y^2 - 6x - 10y = 3.$

2. $2x^2 + 6xy + 3y^2 - 12x - 6y = 5.$

3. $4x^2 - 4xy + 2y^2 + 12x - 8y = 10.$

4. $2x^2 - 6xy + 5y^2 - 10x + 4y = 7$.

As a last application, we consider a quadratic in 3 variables, namely x, y and z. To do so, let $A = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} l \\ m \\ n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ with

$$\begin{aligned} f(x,y,z) &= \mathbf{x}^T A\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + q && (6.6.3.12) \\ &= ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + 2lx + 2my + 2nz + q && (6.6.3.13) \end{aligned}$$

Then, we now observe the following:

1. As A is symmetric, $P^TAP = \mathrm{diag}(\alpha_1, \alpha_2, \alpha_3)$, 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 $f(x, y, z)$ reduces to
$$g(y_1, y_2, y_3) = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \alpha_3 y_3^2 + 2l_1y_1 + 2l_2y_2 + 2l_3y_3 + q. \qquad (6.6.3.14)$$
3. Depending on the values of α_i's, rewrite $g(y_1, y_2, y_3)$ to determine to determine the center and the planes of symmetry of $f(x, y, z) = 0$.

Example 6.6.3.16. 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: Part 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, the orthogonal matrices $P = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix}$ and $P^TAP = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Hence, $f(x, y, z) = 0$ reduces to

$$4(y_1 + \frac{5}{4\sqrt{3}})^2 + (y_2 + \frac{1}{\sqrt{2}})^2 + (y_3 - \frac{1}{\sqrt{6}})^2 = \frac{9}{12}.$$

So, the standard form of the quadric is $4z_1^2 + z_2^2 + z_3^2 = \frac{9}{12}$, where $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = P \begin{bmatrix} \frac{-5}{4\sqrt{3}} \\ \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} \frac{-3}{4} \\ \frac{1}{4} \\ \frac{-3}{4} \end{bmatrix}$ is the center and $x + y + z = 0, x - y = 0$ and $x + y - 2z = 0$ as the principal axes.

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.

The different curves are given in Figure 6.5. These curves have been drawn using the package "MATHEMATICA".

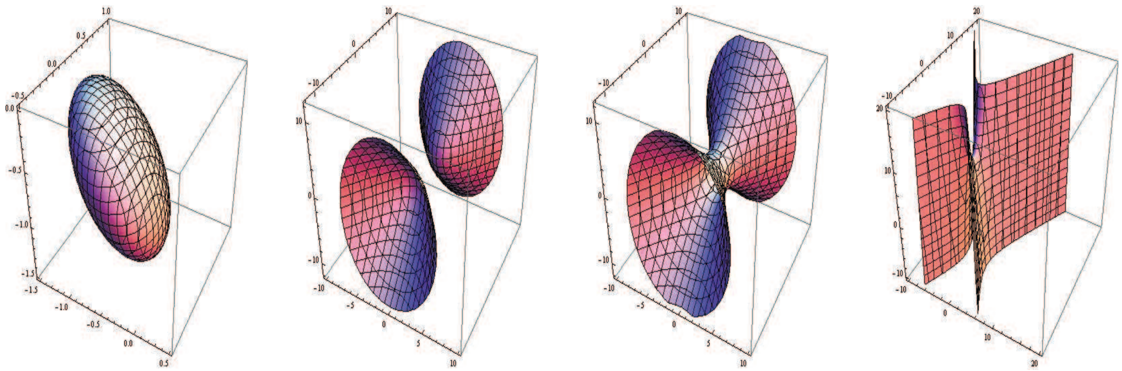


Figure 6.5: Ellipsoid, Hyperboloid of two sheets and one sheet, Hyperbolic Paraboloid

.

Chapter 7

Appendix

7.1 Permutation/Symmetric Groups

Definition 7.7.1.1. 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 7.7.1.2. 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 7.7.1.3. *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 7.7.1.4. 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 **7.7.1.5.** *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 7.7.1.6. *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 7.7.1.5, note that $f_i^{-1} = f_i$, for $i = 1, 2, 3, 6$ and $f_4^{-1} = f_5$.*

Remark 7.7.1.7. *Let $\mathcal{S}_n = \{f : [n] \to [n] \ : \ \sigma$ 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 7.7.1.3, $f \circ g \in \mathcal{S}_n$, 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 7.7.1.8. *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 7.7.1.9. Let $f \in \mathcal{S}_n$. Then, the number of inversions of f, denoted $n(f)$, equals

$$\begin{aligned} n(f) &= |\{(i,j) : \ i < j, \ f(i) > f(j) \ \}| \\ &= |\{j : i+1 \le j \le n, f(j) < f(i)\}| \text{ using two row notation.} \end{aligned} \tag{7.7.1.1}$$

Example 7.7.1.10. 1. For $f = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$, $n(f) = \mid \{(1,2),(1,3),(2,3)\} \mid = 3$.

2. In Exercise 7.7.1.5, $n(f_5) = 2 + 0 = 2$.

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 7.7.1.11. [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 $j \neq i_1, \ldots, i_r$. 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 7.7.1.12. *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 7.7.1.13. A permutation $f \in \mathcal{S}_n$ is called a **transposition** if there exist $m, r \in [n]$ such that $f = (m, r)$.

Remark 7.7.1.14. *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 7.7.1.15. *Let $f \in \mathcal{S}_n$. Then f can be written as product of transpositions.*

Proof. Note that using use Remark 7.7.1.14, 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. □

Remark 7.7.1.16. *Note that when one writes a permutation as product of disjoint cycles, cycles of length 1 are suppressed so as to match Definition 7.7.1.11. For example, the algorithm in the proof of Theorem 7.7.1.15 implies*

1. *Using Remark 7.7.1.14.3, 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 7.7.1.17. *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, $t \geq 2$. If $t = 2$, we are done. So, let us assume that the result holds for all expressions in which the number of transpositions $t \leq k$. 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 applied mathematical induction to get the required result. $\square$

Theorem 7.7.1.18. *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 7.7.1.17, $k+\ell$ is even. Thus, either k and ℓ are both even or both odd. $\square$

Definition 7.7.1.19. [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 7.7.1.20. 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 $\text{sgn} : \mathcal{S}_n \to \{1, -1\}$, 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 7.7.1.21. Consider the set $\mathcal{S}_n$. Then

1. by Lemma 7.7.1.17, 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 7.7.1.20, $\text{sgn}(f \circ g) = \text{sgn}(f) \cdot \text{sgn}(g)$ for any two permutations $f, g \in \mathcal{S}_n$.

We are now ready to define determinant of a square matrix A.

Definition 7.7.1.22. Let $A = [a_{ij}]$ be an $n \times n$ matrix with complex entries. Then the **determinant** of A, denoted $\det(A)$, is defined as

$$\det(A) = \sum_{g \in \mathcal{S}_n} \text{sgn}(g) a_{1g(1)} a_{2g(2)} \ldots a_{ng(n)} = \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(g) \prod_{i=1}^{n} a_{ig(i)}. \tag{7.7.1.2}$$

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 (7.7.1.2),

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma) \prod_{i=1}^{3} a_{i\sigma(i)} \\
&= \text{sgn}(f_1) \prod_{i=1}^{3} a_{if_1(i)} + \text{sgn}(f_2) \prod_{i=1}^{3} a_{if_2(i)} + \text{sgn}(f_3) \prod_{i=1}^{3} a_{if_3(i)} + \\
&\qquad \text{sgn}(f_4) \prod_{i=1}^{3} a_{if_4(i)} + \text{sgn}(f_5) \prod_{i=1}^{3} a_{if_5(i)} + \text{sgn}(f_6) \prod_{i=1}^{3} a_{if_6(i)} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.
\end{aligned}$$

7.2 Properties of Determinant

Theorem 7.7.2.1 (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 7.7.1.8, $\mathcal{S}_n = \{\sigma \circ \tau : \sigma \in \mathcal{S}_n\}$ 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,

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} b_{k\sigma(k)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} b_{k\sigma(k)} \right) b_{i\sigma(i)} \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) (a_{i\sigma(i)} + c a_{j\sigma(j)}) \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) a_{i\sigma(i)} + c \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \left(\prod_{k \neq i} a_{k\sigma(k)} \right) a_{j\sigma(j)}) \\
&= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} a_{k\sigma(k)} + c \cdot 0 = \det(A).
\end{aligned}$$

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 2.2.3.1, $A = E_1 \cdots E_k$, 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 2.2.3.1, $\operatorname{Rank}(A) < n$. So, by Exercise 2.2.2.26.2, 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 2.2.3.1, $A = E_1 \cdots E_k$, 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). 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 $1 \leq i, j \leq n$. By Lemma 7.7.1.8, 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,

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma^{-1}) \prod_{i=1}^{n} b_{\sigma^{-1}(i),i} = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma^{-1}) \prod_{i=1}^{n} a_{i\sigma(i)} \\
&= \det(A).
\end{aligned}$$

$\square$

Remark 7.7.2.2. 1. *As* $\det(A) = \det(A^T)$, *we observe that in Theorem 7.7.2.1, the condition on "row" can be replaced by the condition on "column".*

2. *Let* $A = [a_{ij}]$ *be a matrix satisfying* $a_{11} = 1$ *and* $a_{1j} = 0$, *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, prove that* $\det(A) = \det(B)$.

   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,*

$$\begin{aligned}\det(A) &= \sum_{\sigma \in \mathcal{S}_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} = \sum_{\sigma \in \mathcal{S}_n, \sigma(1)=1} sgn(\sigma) \prod_{i=2}^{n} a_{i\sigma(i)} \sum_{\sigma \in \mathcal{S}_{n-1}} sgn(\sigma) \prod_{i=1}^{n-1} b_{i\sigma(i)} \\ &= \det(B).\end{aligned}$$

We now relate this definition of determinant with the one given in Definition 2.2.3.11.

Theorem 7.7.2.3. *Let* A *be an* $n \times n$ *matrix. Then* $\det(A) = \sum\limits_{j=1}^{n} (-1)^{1+j} a_{1j} \det\big(A(1 \mid j)\big)$, *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 7.7.2.1.5,

$$\det(A) = \sum_{j=1}^{n} \det(B_j). \tag{7.7.2.3}$$

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 7.7.2.1.3, we get $\det(B_j) = (-1)^{j-1} \det(C_j)$. So, using Remark 7.7.2.2.2 and Theorem 7.7.2.1.2 and Equation (7.7.2.3), 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 Chapter 2 is valid. $\square$

7.3 Uniqueness of RREF

Definition 7.7.3.1. 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 $p_{ij} = \delta_{i,f(i)}$, where $\delta_{i,j} = 1$ if $i = j$ and 0, otherwise is the famous Kronecker delta function. 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 7.7.3.2. *Recall that in Remark 7.7.1.16.1, 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 non-zero entry. This non-zero 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 2.2.2.16.

Theorem 7.7.3.3. *Let A and B be two matrices in RREF. If they are 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 $A, B \neq \mathbf{0}$. On the contrary, assume that $A \neq B$. Then, there exists the smallest k such that $A[:, k] \neq B[:, k]$ but $A[:, i] = B[:, i]$, for $1 \le i \le k - 1$. For $1 \le j \le n$, we define matrices $A_j = [A[:, 1], \ldots, A[:, j]]$ and $B_j = [B[:, 1], \ldots, B[:, j]]$. Then, by the choice of k, $A_{k-1} = B_{k-1}$. Also, let there be r pivots in A_{k-1}. To get our result, we consider the following three cases.

Case 1: Neither $A[:,k]$ nor $B[:,k]$ are pivotal. As $A_{k-1} = B_{k-1}$ and the number of pivots is r, we can find a permutation matrix P such that $A_kP = \left[\begin{array}{c|c} I_r & X \\ \mathbf{0} & \mathbf{0} \end{array}\right]$ and $B_kP = \left[\begin{array}{c|c} I_r & Y \\ \mathbf{0} & \mathbf{0} \end{array}\right]$. But, A_k and B_k are equivalent and thus there exists an invertible matrix C such that $A_k = CB_k$. That is,

$$\left[\begin{array}{c|c} I_r & X \\ \mathbf{0} & \mathbf{0} \end{array}\right] = A_kP = CB_kP = \left[\begin{array}{c|c} C_1 & C_2 \\ C_3 & C_4 \end{array}\right]\left[\begin{array}{c|c} I_r & Y \\ \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} C_1 & C_1Y \\ C_3 & C_3Y \end{array}\right].$$

Hence, $C_1 = I_r$ and therefore $X = Y$, contradicting $A[:,k] \neq B[:,k]$.

Case 2: $A[:,k]$ is pivotal but $B[:,k]$ in non-pivotal. Let $\{i_1, \ldots, i_\ell\}$ be the pivotal columns of A_k. Define $F = [A[:,i_1], \ldots, A[:,i_\ell]]$ and $G = B[:,i_1], \ldots, B[:,i_\ell]]$. So, using matrix multiplication, F and G are equivalent, they are in RREF and are of the same size. But,

$$F = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ and } G = \begin{bmatrix} 1 & 0 & \cdots & 0 & b_{1k} \\ 0 & 1 & \cdots & 0 & b_{2k} \\ & \ddots & \ddots & & \\ 0 & \cdots & 0 & 1 & b_{rk} \\ 0 & \cdots & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Clearly F and G are not equivalent.

Case 3: Both $A[:,k]$ and $B[:,k]$ are pivotal. As $A_{k-1} = B_{k-1}$ and they have r pivots, the next pivot will appear in the $(r+1)$-th row. But, both $A[:,k]$ and $B[:,k]$ are pivotal and hence they will have 1 in the $(r+1)$-th entry and 0, everywhere else. Thus, $A[:,k] = B[:,k]$, a contradiction.

Therefore, combining all the three cases, we get the required result. ∎

7.4 Dimension of $\mathbb{W}_1 + \mathbb{W}_2$

Theorem 7.7.4.1. *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*

$$\dim(\mathbb{W}_1) + \dim(\mathbb{W}_2) = \dim(\mathbb{W}_1 + \mathbb{W}_2) + \dim(\mathbb{W}_1 \cap \mathbb{W}_2). \tag{7.7.4.4}$$

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 $\mathbb{W}_2$, we 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

$$\alpha_1\mathbf{u}_1 + \cdots + \alpha_r\mathbf{u}_r + \beta_1\mathbf{w}_1 + \cdots + \beta_s\mathbf{w}_s + \gamma_1\mathbf{v}_1 + \cdots + \gamma_t\mathbf{v}_t = \mathbf{0} \tag{7.7.4.5}$$

in the unknowns 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 (7.7.4.5), 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 3.3.3.16.1, $\alpha_i - \delta_i = 0$, for $1 \le i \le r$ and $\beta_j = 0$, for $1 \le j \le t$. Thus, the system (7.7.4.5) 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 (7.7.4.5) has no non-zero 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 7.7.5.1. *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*

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \qquad \text{(PARALLELOGRAM LAW)}. \tag{7.7.5.6}$$

Proof. Suppose that $\|\cdot\|$ is indeed induced by an inner product. Then by Exercise 5.5.1.7.3 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, $f(\mathbf{x}, \mathbf{x}) \ge 0$. 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,

$$2f(\mathbf{x},\mathbf{y}) = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2, \text{ for } \mathbf{x},\mathbf{y} \in \mathbb{V}. \tag{7.7.5.7}$$

Thus, for $\mathbf{x},\mathbf{y},\mathbf{z} \in \mathbb{V}$, we have

$$\begin{aligned} 4\left(f(\mathbf{x},\mathbf{y}) + f(\mathbf{z},\mathbf{y})\right) &= \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 + \|\mathbf{z}+\mathbf{y}\|^2 - \|\mathbf{z}-\mathbf{y}\|^2 \\ &= 2\left(\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{z}+\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{z}\|^2 - 2\|\mathbf{y}\|^2\right) \\ &= \|\mathbf{x}+\mathbf{z}+2\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{z}\|^2 - \left(\|\mathbf{x}+\mathbf{z}\|^2 + \|\mathbf{x}-\mathbf{z}\|^2\right) - 4\|\mathbf{y}\|^2 \\ &= \|\mathbf{x}+\mathbf{z}+2\mathbf{y}\|^2 - \|\mathbf{x}+\mathbf{z}\|^2 - \|2\mathbf{y}\|^2 \\ &= 2f(\mathbf{x}+\mathbf{z},2\mathbf{y}) \text{ using Equation (7.7.5.7).} \end{aligned} \tag{7.7.5.8}$$

Now, substituting $\mathbf{z} = \mathbf{0}$ in Equation (7.7.5.8) and using Equation (7.7.5.7), 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,

$$f(\mathbf{x}+\mathbf{z},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) + f(\mathbf{z},\mathbf{y}), \text{ for all } \mathbf{x},\mathbf{y} \in \mathbb{V}. \tag{7.7.5.9}$$

STEP 4: Using Equation (7.7.5.9), $f(\mathbf{x},\mathbf{y}) = f(\mathbf{y},\mathbf{x})$ 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 (7.7.5.9) 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. ∎

Index