1 Matrices

If we examine the method of elimination described in Section 1.1, we can make the following observation: Only the numbers in front of the unknowns $x_1, x_2, \ldots, x_n$ and the numbers $b_1, b_2, \ldots, b_m$ on the right side are being changed as we perform the steps in the method of elimination. Thus we might think of looking for a way of writing a linear system without having to carry along the unknowns. Matrices enable us to do this - that is, to write linear systems in a compact form that makes it easier to automate the elimination method by using computer software in order to obtain a fast and efficient procedure for finding solutions. The use of matrices, however, is not merely that of a convenient notation. We now develop operations on matrices and will work with matrices according to the rules they obey; this will enable us to solve systems of linear equations and to handle other computational problems in a fast and efficient manner. Of course, as any good definition should do, the notion of a matrix not only provides a new way of looking at old problems, but also gives rise to a great many new questions, some of which we study in this lecture.

**Definition 1.** An $m \times n$ matrix $A$ is a rectangular array of $mn$ real or complex numbers arranged in $m$ horizontal rows and $n$ vertical columns:

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & \cdots & a_{mn}
\end{bmatrix}
\]

$\leftarrow$ $i$th row

$\leftarrow$ $j$th column
The \(i^{th}\) row of \(A\) is
\[\begin{bmatrix} a_{i1}, a_{i2}, \ldots, a_{im} \end{bmatrix} \quad (1 \leq i \leq m)\]

the \(j^{th}\) column of \(A\) is
\[\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad (1 \leq j \leq n)\]

We shall say that \(A\) is \(m\) by \(n\) (written as \(m \times n\)). If \(m = n\), we say that \(A\) is a square matrix of order \(n\) and that the numbers \(a_{11}, a_{22}, \ldots, a_{nn}\) form the main diagonal of \(A\). We refer to the number \(a_{ij}\), which is in the \(i^{th}\) row and \(j^{th}\) column of \(A\), as the \(i,j^{th}\) element of \(A\), or the \((i,j)\) entry of \(A\), and we often write \((11)\) as \(A = [a_{ij}]\).

Example 2. Let
\[
A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 + i & 4i \\ 2-3i & -3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 3 & -1 & 2 \end{bmatrix}, \quad E = [3], \quad F = \begin{bmatrix} -1 & 0 & 2 \end{bmatrix}
\]

Then \(A\) is a \(2 \times 3\) matrix with \(a_{12} = 2, a_{13} = 3, a_{22} = 0,\) and \(a_{23} = 1\); \(B\) is a \(2 \times 2\) matrix with \(b_{11} = 1 + i, b_{12} = 4i, b_{21} = 2 - 3i,\) and \(b_{22} = -3\); \(C\) is a \(3 \times 1\) matrix with \(c_{11} = 1, c_{21} = -1,\) and \(c_{31} = 2\); \(D\) is a \(3 \times 3\) matrix; \(E\) is a \(1 \times 1\) matrix; and \(F\) is a \(1 \times 3\) matrix. In \(D\), the elements \(d_{11} = 1, d_{22} = 0,\) and \(d_{33} = 2\) form the main diagonal.

An \(n \times 1\) matrix is also called an \(n\)-vector and is denoted by lowercase boldface letters. When \(n\) is understood, we refer to \(n\)-vectors merely as vectors. Vectors are discussed at future sections.

Example 3. \(u = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}\) is a 4-vector and \(v = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}\) is a 3-vector.

The \(n\)-vector all of whose entries are zero is denoted by \(\mathbf{0}\). Observe that if \(A\) is an \(n \times n\) matrix, then the rows of \(A\) are \(1 \times n\) matrices and the columns of \(A\) are \(n \times 1\) matrices. The set of all \(n\)-vectors with real entries is denoted by \(\mathbb{R}^n\). Similarly the set of all \(n\)-vectors with complex entries is denoted by \(\mathbb{C}^n\). As we have already pointed out, in the first six chapters of this book we work almost entirely with vectors in \(\mathbb{R}^n\).
Definition 4. Two \( m \times n \) matrices \( A = [a_{ij}] \) and \( B = [b_{ij}] \) are **equal** if they agree entry by entry, that is, if \( a_{ij} = b_{ij} \) for \( i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \).

Example 5. The matrices

\[
A = \begin{bmatrix}
1 & 2 & -1 \\
2 & -3 & 4 \\
0 & -4 & 5
\end{bmatrix}
\quad \text{and} \quad
B = \begin{bmatrix}
1 & 2 & \ w \\
2 & x & 4 \\
y & -4 & z
\end{bmatrix}
\]

are equal if and only if \( w = -1, x = -3, y = 0, \) and \( z = 5 \).

Definition 6. If \( A = [a_{ij}] \) is an \( n \times n \) matrix, then the **trace** of \( A \), \( \text{Tr}(A) \), is defined as the sum of all elements on the main diagonal of \( A \),

\[
\text{Tr}(A) = \sum_{i=1}^{n} a_{ii}.
\]

1.1 Matrix Operations

We next define a number of operations that will produce new matrices out of given matrices. When we are dealing with linear systems, for example, this will enable us to manipulate the matrices that arise and to avoid writing down systems over and over again. These operations and manipulations are also useful in other applications of matrices.

Definition 7 (Matrix Addition). If \( A = [a_{ij}] \) and \( B = [b_{ij}] \) are both \( m \times n \) matrices, then the sum \( A + B \) is an \( m \times n \) matrix \( C = [c_{ij}] \) defined by \( c_{ij} = a_{ij} + b_{ij}, i = 1, 2, \ldots, m; j = 1, 2, \ldots, n \). Thus, to obtain the sum of \( A \) and \( B \), we merely add corresponding entries.

Example 8. Let

\[
A = \begin{bmatrix}
1 & -2 & 3 \\
2 & -1 & 4
\end{bmatrix}
\quad \text{and} \quad
B = \begin{bmatrix}
0 & 2 & 1 \\
1 & 3 & -4
\end{bmatrix}.
\]

Then

\[
A + B = \begin{bmatrix}
1+0 & -2+2 & 3+1 \\
2+1 & -1+3 & 4+(-4)
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 4 \\
3 & 2 & 0
\end{bmatrix}.
\]

If \( x \) is an \( n \)-vector, then it is easy to show that \( x + 0 = x \), where 0 is the \( n \)-vector all of whose entries are zero.

It should be noted that the sum of the matrices \( A \) and \( B \) is defined only when \( A \) and \( B \) have the same number of rows and the same number of columns, that is, only when \( A \) and \( B \) are of the same size.

We now make the convention that when \( A + B \) is written, both \( A \) and \( B \) are of the same size.
Definition 9 (Scalar Multiplication). If \( A = [a_{ij}] \) is an \( m \times n \) matrix and \( r \) is a real number, then the scalar multiple of \( A \) by \( r \), \( rA \), is the \( m \times n \) matrix \( C = [c_{ij}] \) where \( c_{ij} = ra_{ij}, i = 1, 2, \ldots, m \) and \( j = 1, 2, \ldots, n \); that is, the matrix \( C \) is obtained by multiplying each entry of \( A \) by \( r \).

Example 10. We have

\[
-2 \begin{bmatrix} 4 & -2 & -3 \\ 7 & -3 & 2 \end{bmatrix} = \begin{bmatrix} (-2)(4) & (-2)(-2) & (-2)(-3) \\ (-2)(7) & (-2)(-3) & (-2)(2) \end{bmatrix} = \begin{bmatrix} -8 & 4 & 6 \\ -14 & 6 & -4 \end{bmatrix}
\]

Thus far, addition of matrices has been defined for only two matrices. Our work with matrices will call for adding more than two matrices. Theorem 1.1 in Section 1.4 shows that addition of matrices satisfies the associative property:

\[ A + (B + C) = (A + B) + C. \]

If \( A \) and \( B \) are \( m \times n \) matrices, we write \( A + (-1)B \) as \( A - B \) and call this the difference between \( A \) and \( B \).

Example 11. Let

\[ A = \begin{bmatrix} 2 & 3 & -5 \\ 4 & 2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 & 3 \\ 3 & 5 & -2 \end{bmatrix} \]

Then

\[ A - B = \begin{bmatrix} 2 - 2 & 3 + 1 & -5 - 3 \\ 4 - 3 & 2 - 5 & 1 + 2 \end{bmatrix} = \begin{bmatrix} 0 & 4 & -8 \\ 1 & -3 & 3 \end{bmatrix}. \]

We shall sometimes use the summation notation, and we now review this useful and compact notation.

By \( \sum_{i=1}^{n} a_i \) we mean \( a_1 + a_2 + \cdots + a_n \). The letter \( i \) is called the index of summation, it is a dummy variable that can be replaced by another letter. Hence we can write

\[ \sum_{i=1}^{n} u_i = \sum_{j=1}^{n} u_j = \sum_{k=1}^{n} u_k. \]

Thus

\[ \sum_{i=1}^{4} a_i = a_1 + a_2 + a_3 + a_4. \]

If \( A_1, A_2, \ldots, A_k \) are \( m \times n \) matrices and \( c_1, c_2, \ldots, c_k \) are real numbers, then an expression of the form

\[ c_1 A_1 + c_2 A_2 + \cdots + c_k A_k \]

is called a linear combination of \( A_1, A_2, \ldots, A_k \), and \( c_1, c_2, \ldots, c_k \) are called coefficients. The linear combination in Equation (12) can also be expressed in summation notation as

\[ \sum_{i=1}^{k} c_i A_i = c_1 A_1 + c_2 A_2 + \cdots + c_k A_k. \]
Example 12. The following are linear combinations of matrices:

\[
3 \begin{bmatrix}
0 & -3 & 5 \\
2 & 3 & 4 \\
1 & -2 & -3
\end{bmatrix} - \frac{1}{2} \begin{bmatrix}
5 & 2 & 3 \\
6 & 2 & 3 \\
-1 & -2 & 3
\end{bmatrix} + 2 \begin{bmatrix}
3 & -2 \\
-3 & 5 & 0 \\
-1 & -2
\end{bmatrix} \left(-0.5 \begin{bmatrix}
1 \\
-4 \\
-6
\end{bmatrix} + 0.4 \begin{bmatrix}
0.1 \\
-4 \\
0.2
\end{bmatrix}\right)
\]

Definition 13. If \(A = [a_{ij}]\) is an \(m \times n\) matrix, then the transpose of \(A\), \(A^T = [a_{ji}]\), is the \(n \times m\) matrix defined by \(a_{ji} = a_{ij}\). Thus the transpose of \(A\) is obtained from \(A\) by interchanging the rows and columns of \(A\).

Example 14. Let

\[
A = \begin{bmatrix}
4 & -2 & 3 \\
0 & 5 & -2
\end{bmatrix}, \quad B = \begin{bmatrix}
3 & 2 & -4 \\
0 & 4 & 3
\end{bmatrix}, \quad C = \begin{bmatrix}
5 & 4 \\
-3 & 2 \\
2 & -3
\end{bmatrix}, \quad D = \begin{bmatrix}
3 & -5 & 1
\end{bmatrix}, \quad E = \begin{bmatrix}
2 \\
-1 \\
3
\end{bmatrix}
\]

Then

\[
A^T = \begin{bmatrix}
4 & 0 \\
-2 & 5 \\
3 & -2
\end{bmatrix}, \quad B^T = \begin{bmatrix}
6 & 3 & 0 \\
2 & -1 & 4 \\
-4 & 2 & 3
\end{bmatrix}, \quad C^T = \begin{bmatrix}
5 & -3 & 2 \\
4 & 2 & -3
\end{bmatrix}, \quad D^T = \begin{bmatrix}
-5 \\
3 \\
1
\end{bmatrix}, \quad \text{and} \quad E^T = \begin{bmatrix}
2 \\
-1 \\
3
\end{bmatrix}
\]

1.2 Matrix Multiplication

In this section we introduce the operation of matrix multiplication. Unlike matrix addition, matrix multiplication has some properties that distinguish it from multiplication of real numbers.

Definition 15. The dot product, or inner product, of the \(n\)-vectors \(\mathbf{a} = [a_1, a_2, \ldots, a_n]\) and \(\mathbf{b} = [b_1, b_2, \ldots, b_n]\) in \(\mathbb{R}^n\) is

\[
\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n
\]
is defined as
\[ a \cdot b = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{i=1}^{n} a_ib_i. \]

The dot product is an important operation that will be used here and in later sections.

**Example 16.** The dot product of
\[
\begin{bmatrix}
1 \\
-2 \\
3 \\
4
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
2 \\
3 \\
-2 \\
1
\end{bmatrix}
\]
is
\[ u \cdot v = (1)(2) + (-2)(3) + (3)(-2) + (4)(1) = -6. \]

**Example 17.** Let \( a = \begin{bmatrix} x \\ 2 \\ 3 \end{bmatrix} \) and \( b = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} \). If \( a \cdot b = -4 \), find \( x \).

We have
\[ a \cdot b = 4x + 2 + 6 = -4 \]
\[ 4x + 8 = -4 \]
\[ x = -3. \]

**Definition 18 (Matrix Multiplication).** If \( A = [a_{ij}] \) is an \( m \times p \) matrix and \( B = [b_{ij}] \) is a \( p \times n \) matrix, then the product of \( A \) and \( B \), denoted \( AB \), is the \( m \times n \) matrix \( C = [c_{ij}] \), defined by
\[ c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} \]
\[ = \sum_{k=1}^{p} a_{ik}b_{kj} \quad (1 \leq i \leq m, 1 \leq j \leq n). \] (3)

Equation (13) says that the \( i \)th element in the product matrix is the dot product of the transpose of the \( i \)th row, \( \text{row}_i(A) \)-that is, \( (\text{row}_i(A))^T \)-and the \( j \)th column, \( \text{col}_j(B) \), of \( B \); this is shown in the below figure.
Observe that the product of $A$ and $B$ is defined only when the number of rows of $B$ is exactly the same as the number of columns of $A$, as indicated in the right figure.

**Example 19.** Let 

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -2 & 4 \\ 2 & 1 \end{bmatrix}.$$ 

Then 

$$AB = \begin{bmatrix} (1)(-2) + (2)(4) + (-1)(2) \\ (3)(-2) + (1)(4) + (4)(2) \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ 6 & 16 \end{bmatrix}.$$ 

**Example 20.** Let 

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

**Compute the** $(3,2)$ **entry of** $AB$. 

If \( AB = C \), then the \((3, 2)\) entry of \( AB \) is \( c_{32} \), which is \((\text{row}_3(A))^T \cdot \text{col}_2(B)\). We now have

\[
(\text{row}_3(A))^T \cdot \text{col}_2(B) = \begin{bmatrix} 0 & 4 \\ 1 & -1 \\ -2 & 2 \end{bmatrix} = -5.
\]

**Example 21.** Let

\[
A = \begin{bmatrix} 1 & x & 3 \\ 2 & -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 \\ 4 \\ y \end{bmatrix}
\]

If \( AB = \begin{bmatrix} 12 \\ 6 \end{bmatrix} \), find \( x \) and \( y \).

We have

\[
AB = \begin{bmatrix} 1 & x & 3 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ y \end{bmatrix} = \begin{bmatrix} 2 + 4x + 3y \\ 4 - 4 + y \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \end{bmatrix}
\]

Then

\[
2 + 4x + 3y = 12 \\
y = 6
\]

so \( x = -2 \) and \( y = 6 \).

The basic properties of matrix multiplication will be considered in the next section. However, multiplication of matrices requires much more care than their addition, since the algebraic properties of matrix multiplication differ from those satisfied by the real numbers. Part of the problem is due to the fact that \( AB \) is defined only when the number of columns of \( A \) is the same as the number of rows of \( B \). Thus, if \( A \) is an \( m \times p \) matrix and \( B \) is a \( p \times n \) matrix, then \( AB \) is an \( m \times n \) matrix. What about \( BA \)? Four different situations may occur:

1. \( BA \) may not be defined; this will take place if \( n \neq m \).

2. If \( BA \) is defined, which means that \( m = n \), then \( BA \) is \( p \times p \) while \( AB \) is \( m \times m \); thus if \( m \neq p \), \( AB \) and \( BA \) are of different sizes.

3. If \( AB \) and \( BA \) are both of the same size, they may be equal.

4. If \( AB \) and \( BA \) are both of the same size, they may be unequal.

**Example 22.** Let

\[
A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}
\]

Then

\[
AB = \begin{bmatrix} 2 & 3 \\ -2 & 2 \end{bmatrix} \quad \text{while} \quad BA = \begin{bmatrix} 1 & 7 \\ -1 & 3 \end{bmatrix}
\]

Thus \( AB \neq BA \).
Remark 23. If $\mathbf{u}$ and $\mathbf{v}$ are $n$-vectors ($n \times 1$ matrices), then it is easy to show by matrix multiplication (Exercise 41) that

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}.$$ 

1.3 Algebraic Properties of Matrix Operations

In this section we consider the algebraic properties of the matrix operations just defined. Many of these properties are similar to the familiar properties that hold for real numbers. However, there will be striking differences between the set of real numbers and the set of matrices in their algebraic behavior under certain operations— for example, under multiplication (as seen in Section 1.3). The proofs of most of the properties will be left as exercises.

Theorem 24 (Properties of Matrix Addition). Let $A, B$ and $C$ are $m \times n$ matrices.

a) $A + B = B + A$.

b) $A + (B + C) = (A + B) + C$.

c) There is a unique $m \times n$ matrix $O$ such that

$$A + O = A$$

for any $m \times n$ matrix $A$. The matrix $O$ is called the $m \times n$ zero matrix.

d) For each $m \times n$ matrix $A$, there is a unique $m \times n$ matrix $D$ such that

$$A + D = O$$

We shall write $D$ as $-A$, so (17) can be written as

$$A + (-A) = O.$$ 

The matrix $-A$ is called the negative of $A$. We also note that $-A$ is $(-1)A$. 

9
Proof.

a) Let

\[ A = [a_{ij}] \quad B = [b_{ij}] \]

\[ A + B = C = [c_{ij}] \text{, and } B + A = D = [d_{ij}] . \]

We must show that \( c_{ij} = d_{ij} \) for all \( i, j \). Now \( c_{ij} = a_{ij} + b_{ij} \) and \( d_{ij} = b_{ij} + a_{ij} \) for all \( i, j \). Since \( a_{ij} \) and \( b_{ij} \) are real numbers, we have \( a_{ij} + b_{ij} = b_{ij} + a_{ij} \) which implies that \( c_{ij} = d_{ij} \) for all \( i, j \).

c) Let \( U = [u_{ij}] \). Then \( A + U = A \) if and only if \( a_{ij} + u_{ij} = a_{ij} \), which holds if and only if \( u_{ij} = 0 \). Thus \( U \) is the \( m \times n \) matrix all of whose entries are zero: \( U \) is denoted by \( O \).

\[ \Box \]

Example 25. If

\[ \begin{bmatrix} 4 & -1 \\ 2 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4+0 & -1+0 \\ 2+0 & 3+0 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 2 & 3 \end{bmatrix} . \]

The 2 \( \times \) 3 zero matrix is

\[ O = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} . \]
**Theorem 26 (Properties of Matrix Multiplication).**

a) If $A, B,$ and $C$ are matrices of the appropriate sizes, then

$$A(BC) = (AB)C.$$  

b) If $A, B,$ and $C$ are matrices of the appropriate sizes, then

$$(A + B)C = AC + BC.$$  

c) If $A, B,$ and $C$ are matrices of the appropriate sizes, then

$$C(A + B) = CA + CB.$$  

\[ (6) \]

**Proof.** a) Suppose that $A$ is $m \times n$, $B$ is $n \times p$, and $C$ is $p \times q$. We shall prove the result for the special case $m = 2, n = 3, p = 4,$ and $q = 3$. The general proof is completely analogous.

Let $A = [a_{ij}], B = [b_{ij}], C = [c_{ij}], AB = D = [d_{ij}], BC = E = [e_{ij}], (AB)C = F = [f_{ij}],$ and $A(BC) = G = [g_{ij}]$. We must show that $f_{ij} = g_{ij}$ for all $i, j$. Now

$$f_{ij} = \sum_{k=1}^{4} d_{ik}c_{kj} = \sum_{k=1}^{4} \left( \sum_{r=1}^{3} a_{ir}b_{rk} \right) c_{kj}$$

and

$$g_{ij} = \sum_{r=1}^{3} a_{ir}c_{rj} = \sum_{r=1}^{3} a_{ir} \left( \sum_{k=1}^{4} b_{rk}c_{kj} \right)$$

Then, by the properties satisfied by the summation notation,

$$f_{ij} = \sum_{k=1}^{4} \left( a_{i1}b_{1k} + a_{i2}b_{2k} + a_{i3}b_{3k} \right) c_{kj}$$

$$= a_{i1} \sum_{k=1}^{4} b_{1k}c_{kj} + a_{i2} \sum_{k=1}^{4} b_{2k}c_{kj} + a_{i3} \sum_{k=1}^{4} b_{3k}c_{kj}.$$ 

$$= \sum_{r=1}^{3} a_{ir} \left( \sum_{k=1}^{4} b_{rk}c_{kj} \right) = g_{ij}$$ 

\[ \square \]
Example 27. Let

\[ A = \begin{bmatrix} 5 & 2 & 3 \\ 2 & -3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 2 & 2 \\ 3 & 0 & -1 \end{bmatrix} \]

and

\[ C = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -3 & 0 \\ 0 & 0 & 3 \\ 2 & 1 & 0 \end{bmatrix} \]

Then

\[ A(BC) = \begin{bmatrix} 5 & 2 & 3 \\ 2 & -3 & 4 \end{bmatrix} \begin{bmatrix} 8 & 3 & 7 \\ 9 & -4 & 6 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 43 & 16 & 56 \\ 12 & 30 & 8 \end{bmatrix} \]

and

\[ (AB)C = \begin{bmatrix} 19 & -1 & 6 \\ 16 & -8 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 3 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 43 & 16 & 56 \\ 12 & 30 & 8 \end{bmatrix} \]

Example 28. Let

\[ A = \begin{bmatrix} 2 & 2 & 3 \\ 3 & -1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 2 & 3 & -1 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 1 & 0 \\ 2 & 2 \\ 3 & -1 \end{bmatrix}. \]

Then

\[ (A+B)C = \begin{bmatrix} 2 & 2 & 4 \\ 5 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 18 & 0 \\ 12 & 3 \end{bmatrix} \]

and (verify)

\[ AC + BC = \begin{bmatrix} 15 & 1 \\ 7 & -4 \end{bmatrix} + \begin{bmatrix} 3 & -1 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} 18 & 0 \\ 12 & 3 \end{bmatrix}. \]

Theorem 29 (Properties of Scalar Multiplication). If \( r \) and \( s \) are real numbers and \( A \) and \( B \) are matrices of the appropriate sizes, then

\( a) r(sA) = (rs)A \)

\( b) (r + s)A = rA + sA \)

\( c) r(A + B) = rA + rB \)

\( d) A(rB) = r(AB) = (rA)B \)
Example 30. Let
\[ A = \begin{bmatrix} 4 & 2 & 3 \\ 2 & -3 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 3 & -2 & 1 \\ 2 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}. \]
Then
\[ 2(3A) = 2 \begin{bmatrix} 12 & 6 & 9 \\ 6 & -9 & 12 \end{bmatrix} = \begin{bmatrix} 24 & 12 & 18 \\ 12 & -18 & 24 \end{bmatrix} = 6A. \]
We also have
\[ A(2B) = \begin{bmatrix} 4 & 2 & 3 \\ 2 & -3 & 4 \\ 6 & -4 & 2 \\ 4 & 0 & -2 \\ 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 32 & -10 & 16 \\ 0 & 0 & 26 \end{bmatrix} = 2(AB). \]

Example 31. Scalar multiplication can be used to change the size of entries in a matrix to meet prescribed properties. Let
\[ A = \begin{bmatrix} 3 \\ 7 \\ 2 \\ 1 \end{bmatrix}. \]
Then for \( k = \frac{1}{7} \), the largest entry of \( kA \) is 1. Also if the entries of \( A \) represent the volume of products in gallons, for \( k = 4 \), \( kA \) gives the volume in quarts.

So far we have seen that multiplication and addition of matrices have much in common with multiplication and addition of real numbers. We now look at some properties of the transpose.

Theorem 32 (Properties of Transpose). If \( r \) is a scalar and \( A \) and \( B \) are matrices of the appropriate sizes, then

\[ a) \quad (A^T)^T = A. \]
\[ b) \quad (A + B)^T = A^T + B^T. \]
\[ c) \quad (AB)^T = B^T A^T. \]
\[ d) \quad (rA)^T = rA^T. \]

Proof. We leave the proofs of \( a) \), \( b) \), and \( d) \) as an exercise.
c) Let $A = [a_{ij}]$ and $B = [b_{ij}]$; let $AB = C = [c_{ij}]$. We must prove that $c_{ij}^T$ is the $(i, j)$ entry in $B^T A^T$. Now

$$c_{ij}^T = c_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} a^T_{kj} b^T_{ik}$$

$$= \sum_{k=1}^{n} b^T_{ik} a^T_{kj} = \text{the (i, j) entry in } B^T A^T.$$

\[ \square \]

**Example 33.** Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -2 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & -1 & 2 \\ 3 & 2 & -1 \end{bmatrix}.$$

Then

$$A^T = \begin{bmatrix} 1 & -2 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \quad \text{and} \quad B^T = \begin{bmatrix} 3 & 3 \\ -1 & 2 \\ 2 & -1 \end{bmatrix}.$$

Also,

$$A + B = \begin{bmatrix} 4 & 1 & 5 \\ 1 & 2 & 0 \end{bmatrix} \quad \text{and} \quad (A + B)^T = \begin{bmatrix} 4 & 1 \\ 1 & 2 \\ 5 & 0 \end{bmatrix}.$$

Now

$$A^T + B^T = \begin{bmatrix} 4 & 1 \\ 1 & 2 \\ 5 & 0 \end{bmatrix} = (A + B)^T.$$

**Example 34.** Let

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 12 & 5 \\ 7 & -3 \end{bmatrix} \quad \text{and} \quad (AB)^T = \begin{bmatrix} 12 & 7 \\ 5 & -3 \end{bmatrix}.$$

On the other hand,

$$A^T = \begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad B^T = \begin{bmatrix} 0 & 2 & 3 \\ 1 & 2 & -1 \end{bmatrix}.$$

Then

$$B^T A^T = \begin{bmatrix} 12 & 7 \\ 5 & -3 \end{bmatrix} = (AB)^T.$$
Remark 35. We also note two other peculiarities of matrix multiplication. If \( a \) and \( b \) are real numbers, then \( ab = 0 \) can hold only if \( a \) or \( b \) is zero. However, this is not true for matrices.

Example 36. If
\[
A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & -6 \\ -2 & 3 \end{bmatrix}
\]
then neither \( A \) nor \( B \) is the zero matrix, but \( AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \).

Remark 37. If \( a, b, \) and \( c \) are real numbers for which \( ab = ac \) and \( a \neq 0 \), it follows that \( b = c \). That is, we can cancel out the nonzero factor \( a \). However, the cancellation law does not hold for matrices, as the following example shows.

Example 38. If
\[
A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} -2 & 7 \\ 5 & -1 \end{bmatrix}
\]
then
\[
AB = AC = \begin{bmatrix} 8 & 5 \\ 16 & 10 \end{bmatrix}
\]
but \( B \neq C \).

Remark 39. We summarize some of the differences between matrix multiplication and the multiplication of real numbers as follows: For matrices \( A, B, \) and \( C \) of the appropriate sizes,

1. \( AB \) need not equal \( BA \).
2. \( AB \) may be the zero matrix with \( A \neq O \) and \( B \neq O \).
3. \( AB \) may equal \( AC \) with \( B \neq C \).

1.4 Special Types of Matrices and Partitioned Matrices

We have already introduced one special type of matrix \( O \), the matrix all of whose entries are zero. We now consider several other types of matrices whose structures are rather specialized and for which it will be convenient to have special names. An \( n \times n \) matrix \( A = [a_{ij}] \) is called a **diagonal matrix** if \( a_{ij} = 0 \) for \( i \neq j \). Thus, for a diagonal matrix, the terms off the main diagonal are all zero. Note that \( O \) is a diagonal matrix. A **scalar matrix** is a diagonal matrix whose diagonal elements are equal. The scalar matrix \( I_n = [d_{ij}] \), where \( d_{ii} = 1 \) and \( d_{ij} = 0 \) for \( i \neq j \), is called then \( n \times n \) **identity matrix**.
Example 40. Let

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
2 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 2 \\
\end{bmatrix}, \quad \text{and} \quad I_3 = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

Then \(A\), \(B\), and \(I_3\) are diagonal matrices; \(B\) and \(I_3\) are scalar matrices; and \(I_3\) is the \(3 \times 3\) identity matrix.

It is easy to show (Exercise 1) that if \(A\) is any \(m \times n\) matrix, then

\[
AI_n = A \quad \text{and} \quad I_mA = A
\]

Also, if \(A\) is a scalar matrix, then \(A = rI_n\) for some scalar \(r\).

Suppose that \(A\) is a square matrix. We now define the powers of a matrix, for \(p\) a positive integer, by

\[
A^p = A \cdot A \cdots A \quad \text{\(p\) factors}
\]

If \(A\) is \(n \times n\), we also define

\[
A^0 = I_n
\]

For non-negative integers \(p\) and \(q\), the familiar laws of exponents for the real numbers can also be proved for matrix multiplication of a square matrix \(A\) (Exercise 8):

\[
A^p A^q = A^{p+q} \quad \text{and} \quad (A^p)^q = A^{pq}
\]

It should also be noted that the rule

\[
(AB)^p = A^p B^p
\]

does not hold for square matrices unless \(AB = BA\) (Exercise 9).

An \(n \times n\) matrix \(A = [a_{ij}]\) is called upper triangular if \(a_{ij} = 0\) for \(i > j\). It is called lower triangular if \(a_{ij} = 0\) for \(i < j\). A diagonal matrix is both upper triangular and lower triangular.

Example 41. The matrix

\[
A = \begin{bmatrix}
1 & 3 & 3 \\
0 & 3 & 5 \\
0 & 0 & 2 \\
\end{bmatrix}
\]

is upper triangular; and

\[
B = \begin{bmatrix}
1 & 0 & 0 \\
2 & 3 & 0 \\
3 & 5 & 2 \\
\end{bmatrix}
\]

is lower triangular.
Definition 42. A matrix \( A \) with real entries is called **symmetric** if \( A^T = A \).

Definition 43. A matrix \( A \) with real entries is called **skew symmetric** if \( A^T = -A \).

Example 44. \( A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} \) is a symmetric matrix.

Example 45. \( B = \begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & -4 \\ -3 & 4 & 0 \end{bmatrix} \) is a skew symmetric matrix.

It follows from the preceding definitions that:

- If \( A \) is symmetric or skew symmetric, then \( A \) is a square matrix.
- If \( A \) is a symmetric matrix, then the entries of \( A \) are symmetric with respect to the main diagonal of \( A \). Also, \( A \) is symmetric if and only if \( a_{ij} = a_{ji} \).
- \( A \) is skew symmetric if and only if \( a_{ij} = -a_{ji} \). Moreover, if \( A \) is skew symmetric, then the entries on the main diagonal of \( A \) are all zero.
- An important property of symmetric and skew symmetric matrices is the following: If \( A \) is an \( n \times n \) matrix, then we can show that \( A = S + K \), where \( S \) is symmetric and \( K \) is skew symmetric. Moreover, this decomposition is unique (Exercise 29).

Definition 46. Let \( A \) be an \( n \times n \) matrix and \( p \) be a positive integer. If \( A^{p+1} = A \) then \( A \) is called a **periodic** matrix. If \( p \) is the least such integer, then the matrix is said to have **period** \( p \). If \( p = 1 \) then \( A^2 = A \) and \( A \) is called **idempotent**.

Definition 47. Let \( A \) be an \( n \times n \) matrix and \( q \) be a positive integer. If \( A^q = O \) then \( A \) is called a **nilpotent** matrix. The least such positive integer \( q \) is called the **index** (or, degree) of **nilpotency**.

Definition 48. Let \( A \) be an \( n \times n \) matrix. If \( A^2 = I_n \) then \( A \) is called an **involutory** matrix.

Definition 49. An \( n \times n \) real matrix \( A \) is called **orthogonal** if \( AA^T = A^T A = I_n \).

Example 50. The matrix \( A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \) is an orthogonal matrix.

**Partitioned Matrices**

If we start with an \( m \times n \) matrix \( A = [a_{ij}] \) and then cross out some, but not all of its rows or columns, we obtain a **submatrix** of \( A \).
Example 51. Let
\[
A = \begin{bmatrix}
1 & 2 & 3 & 4 \\
-2 & 4 & -3 & 5 \\
3 & 0 & 5 & -3
\end{bmatrix}.
\]
If we cross out the second row and third column, we get the submatrix
\[
\begin{bmatrix}
1 & 2 & 4 \\
3 & 0 & -3
\end{bmatrix}.
\]
A matrix can be partitioned into submatrices by drawing horizontal lines between rows and vertical lines between columns. Of course, the partitioning can be carried out in many different ways.

Example 52. The matrix
\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45}
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\]
can be partitioned as indicated previously. We could also write
\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45}
\end{bmatrix} = \begin{bmatrix}
\hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} \\
\hat{A}_{21} & \hat{A}_{22} & \hat{A}_{23}
\end{bmatrix}
\]
which gives another partitioning of \( A \). We thus speak of partitioned matrices.

Example 53. Let
\[
A = \begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 2 & 3 & -1 \\
2 & 0 & -4 & 0 \\
0 & 1 & 0 & 3
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\]
and let
\[
B = \begin{bmatrix}
2 & 0 & 0 & 1 & 1 & -1 \\
0 & 1 & 1 & -1 & 2 & 2 \\
1 & 3 & 0 & 0 & 1 & 0 \\
-3 & -1 & 2 & 1 & 0 & -1
\end{bmatrix} = \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix}
\]
Then
\[
AB = C = \begin{bmatrix}
3 & 3 & 0 & 1 & 2 & -1 \\
6 & 12 & 0 & -3 & 7 & 5 \\
0 & -12 & 0 & 2 & -2 & -2 \\
-9 & -2 & 7 & 2 & 2 & -1
\end{bmatrix} = \begin{bmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{bmatrix}
\]
where $C_{11}$ should be $A_{11}B_{11} + A_{12}B_{21}$. We verify that $C_{11}$ is this expression as

$$
A_{11}B_{11} + A_{12}B_{21} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ -3 & -1 & 2 \end{bmatrix}
= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 3 & 0 \\ 6 & 10 & -2 \end{bmatrix}
= \begin{bmatrix} 3 & 3 & 0 \\ 6 & 12 & 0 \end{bmatrix} = C_{11}
$$

- This method of multiplying partitioned matrices is also known as block multiplication.
- Partitioned matrices can be used to great advantage when matrices exceed the memory capacity of a computer.
- Thus, in multiplying two partitioned matrices, one can keep the matrices on disk and bring into memory only the submatrices required to form the submatrix products.
- The products, of course, can be downloaded as they are formed. The partitioning must be done in such a way that the products of corresponding submatrices are defined.

Partitioning of a matrix implies a subdivision of the information into blocks, or units. The reverse process is to consider individual matrices as blocks and adjoin them to form a partitioned matrix. The only requirement is that after the blocks have been joined, all rows have the same number of entries and all columns have the same number of entries.

**Definition 54.** Let $A$ be an $m \times n$ matrix with complex entries. The matrix $\overline{A}$, whose entries are complex conjugates of the entries of the matrix $A$ is called \textit{(complex) conjugate} of $A$.

**Example 55.** Let $A = \begin{bmatrix} 1 + i & 2 - i \\ 3 & 4i \end{bmatrix}$. Then the complex conjugate $\overline{A}$ of the matrix $A$ is

$$
\overline{A} = \begin{bmatrix} 1 - i & 2 + i \\ 3 & -4i \end{bmatrix}.
$$

**Properties of conjugate Matrices** If $A$ and $B$ are complex matrices of appropriate sizes and $k$ is a scalar then

a) $\overline{\left(\overline{A}\right)} = A$

b) $\overline{kA} = k\overline{A}$

c) $\overline{A + B} = \overline{A} + \overline{B}$

d) $\overline{AB} = \overline{A}\overline{B}$

e) $\overline{A}^T = \overline{A}^T$

19
Definition 56. An \( n \times n \) complex matrix \( A \) is called **Hermitian** if \( (A)^T = A \). This is equivalent to say that \( \overline{a_{ji}} = a_{ij} \) for all \( i \) and \( j \).

* Every real symmetric matrix is Hermitian, so we may consider Hermitian matrices as the analogs of real symmetric matrices.

* The entries on the main diagonal of a Hermitian matrix are all real numbers.

Example 57. The matrix

\[
A = \begin{bmatrix}
1 & i & 1+i \\
-i & -5 & 2+i \\
1-i & 2-i & 3
\end{bmatrix}
\]

is Hermitian since,

\[
(A)^T = \begin{bmatrix}
1 & -i & 1-i \\
i & -5 & 2-i \\
1+i & 2+i & 3
\end{bmatrix}^T = \begin{bmatrix}
1 & i & 1+i \\
-i & -5 & 2+i \\
1-i & 2-i & 3
\end{bmatrix} = A.
\]

Definition 58. An \( n \times n \) complex matrix \( A \) is called **skew-Hermitian** if \( (A)^T = -A \). This is equivalent to say that \( \overline{a_{ji}} = -a_{ij} \) for all \( i \) and \( j \).

* The entries on the main diagonal of a skew-Hermitian matrix are zero or a complex number with only imaginary part.

Example 59. Let

\[
A = \begin{bmatrix}
i & 1-i & 2 \\
-1-i & 3i & i \\
-2 & i & 0
\end{bmatrix}
\]

Then we have

\[
\overline{A} = \begin{bmatrix}
-i & 1+i & 2 \\
-1+i & -3i & -i \\
-2 & -i & 0
\end{bmatrix} \quad \text{and} \quad (A)^T = \begin{bmatrix}
-i & -1+i & -2 \\
1+i & -3i & -i \\
2 & -i & 0
\end{bmatrix} = -A.
\]

So, \( A \) is skew-Hermitian.

**Nonsingular Matrices**

We now come to a special type of square matrix and formulate the notion corresponding to the reciprocal of a nonzero real number.

Definition 60. An \( n \times n \) matrix \( A \) is called **nonsingular, or invertible (regular)**, if there exists an \( n \times n \) matrix \( B \) such that \( AB = BA = I_n \); such a \( B \) is called an inverse of \( A \). Otherwise, \( A \) is called **singular, or noninvertible**.

Remark 61. In Theorem 2.11, Section 2.3, we show that if \( AB = I_n \), then \( BA = I_n \). Thus, to verify that \( B \) is an inverse of \( A \), we need verify only that \( AB = I_n \).
Example 62. Let $A = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{bmatrix}$. Since $AB = BA = I_2$, we conclude that $B$ is an inverse of $A$.

**Theorem 63.** The inverse of a matrix, if it exists, is unique.

*Proof.* Let $B$ and $C$ be inverses of $A$. Then

$$AB = BA = I_n \quad \text{and} \quad AC = CA = I_n.$$ 

We then have $B = BI_n = B(AC) = (BA)C = I_nC = C$, which proves that the inverse of a matrix, if it exists, is unique. \qed

Because of this uniqueness, we write the inverse of a nonsingular matrix $A$ as $A^{-1}$. Thus

$$AA^{-1} = A^{-1}A = I_n.$$

**Example 64.** Let

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

If $A^{-1}$ exists, let

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$ 

Then we must have

$$AA^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so that

$$\begin{bmatrix} a + 2c & b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ 

Equating corresponding entries of these two matrices, we obtain the linear systems

$$a + 2c = 1 \quad \text{and} \quad 3a + 4c = 0 \quad \text{and} \quad b + 2d = 0 \quad \text{and} \quad 3b + 4d = 1.$$ 

The solutions are (verify) $a = -2, c = \frac{3}{2}, b = 1, \text{ and } d = -\frac{1}{2}$. Moreover, since the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix},$$

also satisfies the property that

$$\begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
we conclude that $A$ is nonsingular and that
$$A^{-1} = \begin{bmatrix}
-2 & 1 \\
3 & 1/2 \\
2 & -1/2
\end{bmatrix}.$$  

We next establish several properties of inverses of matrices.

**Theorem 65.** If $A$ and $B$ are both nonsingular $n \times n$ matrices, then $AB$ is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$.

*Proof.* We have $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = (AI_n)A^{-1} = AA^{-1} = I_n$. Similarly, $(B^{-1}A^{-1})(AB) = I_n$. Therefore $AB$ is nonsingular. since the inverse of a matrix is unique, we conclude that $(AB)^{-1} = B^{-1}A^{-1}$.  

**Corollary 66.** If $A_1, A_2, \ldots, A_r$ are $n \times n$ nonsingular matrices, then $A_1A_2 \cdots A_r$ is nonsingular and $(A_1A_2 \cdots A_r)^{-1} = A_r^{-1}A_{r-1}^{-1} \cdots A_1^{-1}$.

**Theorem 67.** If $A$ is a nonsingular matrix, then $A^{-1}$ is nonsingular and $(A^{-1})^{-1} = A$.

**Theorem 68.** If $A$ is a nonsingular matrix, then $A^T$ is nonsingular and $(A^{-1})^T = (A^T)^{-1}$.

*Proof.* We have $AA^{-1} = I_n$. Taking transposes of both sides, we get
$$(A^{-1})^TA^T = I_n^T = I_n,$$  

we find, similarly, that
$$(A^T)(A^{-1})^T = I_n.$$  

These equations imply that $(A^{-1})^T = (A^T)^{-1}$.  

**Example 69.** If
$$A = \begin{bmatrix} 1 & 2 \\
3 & 4 \end{bmatrix},$$  

then from Example 64
$$A^{-1} = \begin{bmatrix}
-2 & 1 \\
3 & 1/2 \\
2 & -1/2
\end{bmatrix} \text{ and } (A^{-1})^T = \begin{bmatrix}
-2 & 3 \\
1 & 1/2 \\
2 & -1/2
\end{bmatrix}.$$  

Also (verify)
$$A^T = \begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix} \text{ and } (A^T)^{-1} = \begin{bmatrix}
-2 & 3 \\
1 & 1/2 \\
2 & -1/2
\end{bmatrix}.$$  

**Remark 70.** Suppose that $A$ is nonsingular. Then $AB = AC$ implies that $B = C$ (Exercise 50), and $AB = O$ implies that $B = O$ (Exercise 51). It follows from Theorem 1.8 that if $A$ is a symmetric nonsingular matrix, then $A^{-1}$ is symmetric. (See Exercise 54.)
2 Echelon Form of a Matrix

Definition 71. An $m \times n$ matrix $A$ is said to be in reduced row echelon form if it satisfies the following properties:

a) All zero rows, if there are any, appear at the bottom of the matrix.

b) The first nonzero entry from the left of a nonzero row is a 1. This entry is called a leading one of its row.

c) For each nonzero row, the leading one appears to the right and below any leading ones in preceding rows.

d) If a column contains a leading one, then all other entries in that column are zero.

Remark 72. • A matrix in reduced row echelon form appears as a staircase ("echelon") pattern of leading ones descending from the upper left corner of the matrix.

• An $m \times n$ matrix satisfying properties a), b), and c) is said to be in row echelon form. In Definition 69, there may be no zero rows.

• A similar definition can be formulated in the obvious manner for reduced column echelon form and column echelon form.

Example 73. The following are matrices in reduced row echelon form, since they satisfy properties a), b), c), and d):

$$A = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}, \quad B = \begin{bmatrix}
1 & 0 & 0 & -2 & 4 \\
0 & 1 & 0 & 4 & 8 \\
0 & 0 & 1 & 7 & -2 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},$$

and

$$C = \begin{bmatrix}
1 & 2 & 0 & 0 & 1 \\
0 & 0 & 1 & 2 & 3 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}.$$
Example 74. The following are matrices in row echelon form:

\[ H = \begin{bmatrix} 1 & 5 & 0 & 2 & -2 & 4 \\ 0 & 1 & 0 & 3 & 4 & 8 \\ 0 & 0 & 0 & 1 & 7 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 0 & 1 & 3 & 5 & 7 & 9 \\ 0 & 0 & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \]

A useful property of matrices in reduced row echelon form (see Exercise 9) is that if \( A \) is an \( n \times n \) matrix in reduced row echelon form \( \neq I_n \), then \( A \) has a row consisting entirely of zeros.

We shall now show that every matrix can be put into row (column) echelon form, or into reduced row (column) echelon form, by means of certain row (column) operations.

Definition 75. An elementary row (column) operation on a matrix \( A \) is any one of the following operations:

a) **Type I:** Interchange any two rows (columns).

b) **Type II:** Multiply a row (column) by a nonzero number.

c) **Type III:** Add a multiple of one row (column) to another.

We now introduce the following notation for elementary row and elementary column operations on matrices:

- Interchange rows (columns) \( i \) and \( j \), Type I:

  \[ r_i \leftrightarrow r_j \quad (c_i \leftrightarrow c_j) \]

- Replace row (column) \( i \) by \( k \) times row (column) \( i \), Type II:

  \[ kr_i \rightarrow r_i \quad (kc_i \rightarrow c_i) \]

- Replace row (column) \( j \) by \( k \) times row (column) \( i \)+ row (column) \( j \), Type III:

  \[ kr_i + r_j \rightarrow r_j(kc_i + c_j \rightarrow c_j) \]
Using this notation, it is easy to keep track of the elementary row and column operations performed on a matrix. For example, we indicate that we have interchanged the $i$th and $j$th rows of $A$ as $A_{r_i \leftrightarrow r_j}$. We proceed similarly for column operations.

Observe that when a matrix is viewed as the augmented matrix of a linear system, the elementary row operations are equivalent, respectively, to interchanging two equations, multiplying an equation by a nonzero constant, and adding a multiple of one equation to another equation.

**Example 76.** Let

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ 3 & 6 & -9 \end{bmatrix}$$

*Interchanging rows 1 and 3 of $A$, we obtain*

$$B = A_{r_1 \leftrightarrow r_3} = \begin{bmatrix} 3 & 3 & 6 & -9 \\ 2 & 3 & 0 & -2 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

*Multiplying the third row of $A$ by $\frac{1}{3}$, we obtain*

$$C = A_{\frac{1}{3}r_3} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ 1 & 1 & 2 & -3 \end{bmatrix}.$$

*Adding $(-2)$ times row 2 of $A$ to row 3 of $A$, we obtain*

$$D = A_{-2r_2 + r_3} = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 2 & 3 & 0 & -2 \\ -1 & -3 & 6 & -5 \end{bmatrix}.$$

Observe that in obtaining $D$ from $A$, row 2 of $A$ did not change.

**Definition 77.** An $m \times n$ matrix $B$ is said to be **row (column) equivalent** to an $m \times n$ matrix $A$ if $B$ can be produced by applying a finite sequence of elementary row (column) operations to $A$.

**Example 78.** Let

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

*If we add 2 times row 3 of $A$ to its second row, we obtain*

$$B = A_{2r_3 + r_2} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 4 & -3 & 7 & 8 \\ 1 & -2 & 2 & 3 \end{bmatrix}.$$
so \( B \) is row equivalent to \( A \).

Interchanging rows 2 and 3 of \( B \), we obtain

\[
C = B_{r_2 \leftrightarrow r_3} = \begin{bmatrix}
1 & 2 & 4 & 3 \\
1 & -2 & 2 & 3 \\
4 & -3 & 7 & 8 \\
\end{bmatrix}
\]

so \( C \) is row equivalent to \( B \).

Multiplying row 1 of \( C \) by 2, we obtain

\[
D = C_{2r_1 \rightarrow r_1} = \begin{bmatrix}
2 & 4 & 8 & 6 \\
1 & -1 & 2 & 3 \\
4 & -3 & 7 & 8 \\
\end{bmatrix}
\]

so \( D \) is row equivalent to \( C \). It then follows that \( D \) is row equivalent to \( A \), since we obtained \( D \) by applying three successive elementary row operations to \( A \). Using the notation for elementary row operations, we have

\[
D = A_{2r_3 + r_2 \rightarrow r_2} \\
2r_1 \rightarrow r_1
\]

We adopt the convention that the row operations are applied in the order listed.

**Theorem 79.** Every nonzero \( m \times n \) matrix \( A = [a_{ij}] \) is row (column) equivalent to a matrix in row (column) echelon form.

**Definition 80.** The number of non-zero rows (columns) of a row (column) echelon form of the matrix \( A \) is called the **rank** of \( A \) and denoted by \( r_A \) or \( \text{rank}(A) \).

**Example 81.** Consider the following matrices:

\[
A = \begin{bmatrix}
1 & 4 & 3 \\
0 & 1 & 2 \\
0 & 0 & 1 \\
\end{bmatrix}, \quad B = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
\end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix}
1 & 2 & 3 \\
0 & 0 & 1 \\
0 & 0 & 0 \\
\end{bmatrix}.
\]

Then \( \text{rank}(A) = 3 \), \( \text{rank}(B) = 2 \) and \( \text{rank}(C) = 2 \).

In matrix \( A \), the first column with a nonzero entry is called the pivot column; the first nonzero entry in the pivot column is called the pivot.

**Example 82.** Let
Column 1 is the first (counting from left to right) column in $A$ with a nonzero entry, so column 1 is the pivot column of $A$. The first (counting from top to bottom) nonzero entry in the pivot column occurs in the third row, so the pivot is $a_{31} = 2$. We interchange the first and third rows of $A$, obtaining

$$B = A_{r1 \leftrightarrow r3} = \begin{bmatrix}
2 & 2 & -5 & 2 & 4 \\
0 & 0 & 2 & 3 & 4 \\
0 & 2 & 3 & -4 & 1 \\
2 & 0 & -6 & 9 & 7
\end{bmatrix}.$$  

Multiply the first row of $B$ by the reciprocal of the pivot, that is, by $\frac{1}{b_{11}} = \frac{1}{2}$, to obtain

$$C = B_{r_1 \rightarrow r_1} = \begin{bmatrix}
1 & 1 & -5 & 1 & 2 \\
0 & 0 & 2 & 3 & 4 \\
0 & 2 & 3 & -4 & 1 \\
2 & 0 & -6 & 9 & 7
\end{bmatrix}.$$  

Add $(-2)$ times the first row of $C$ to the fourth row of $C$ to produce a matrix $D$ in which the only nonzero entry in the pivot column is $d_{11} = 1$:

$$D = C_{-2r_1 + r_4 \rightarrow r_4} = \begin{bmatrix}
1 & 1 & -5 & 1 & 2 \\
0 & 0 & 2 & 3 & 4 \\
0 & 2 & 3 & -4 & 1 \\
0 & -2 & -1 & 7 & 3
\end{bmatrix}.$$  

Identify $A_1$ as the submatrix of $D$ obtained by deleting the first row of $D$: Do not erase the first row of $D$. Repeat the preceding steps with $A_1$ instead of $A$. 

27
Deleting the first row of $D_1$ yields the matrix $A_2$. We repeat the procedure with $A_2$ instead of $A$. No rows of $A_2$ have to be interchanged.
The matrix

\[ H = \begin{bmatrix}
1 & 1 & -\frac{5}{2} & 1 & 2 \\
0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \\
0 & 0 & -1 & \frac{3}{2} & 2 \\
0 & 0 & 0 & -1 & \frac{3}{2} \\
0 & 0 & 0 & 0 & 0
\end{bmatrix} \]

is in row echelon form and is row equivalent to \( A \).

**Theorem 83.** Every nonzero \( m \times n \) matrix \( A = [a_{ij}] \) is row (column) equivalent to a unique matrix in reduced row (column) echelon form.
Finding $A^{-1}$

**Remark 84.** Note that at this point we have shown that the following statements are equivalent for an $n \times n$ matrix $A$:

1. $A$ is nonsingular.
2. $A$ is row (column) equivalent to $I_n$ (The reduced row echelon form of $A$ is $I_n$.)

For a nonsingular matrix $A$, we transform the partitioned matrix $[A \mid I_n]$ to reduced row echelon form, obtaining $[I_n \mid A^{-1}]$.

**Example 85.** Let

\[
A = \begin{bmatrix}
1 & 1 & 1 \\
0 & 2 & 3 \\
5 & 5 & 1
\end{bmatrix}
\]

Assuming that $A$ is nonsingular, we form

\[
[A \mid I_3] = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
0 & 2 & 3 & 0 & 1 \\
5 & 5 & 1 & 0 & 1
\end{bmatrix}
\]

We now perform elementary row operations that transform $[A \mid I_3]$ to $[I_3 \mid A^{-1}]$: we consider $[A \mid I_3]$ as a $3 \times 6$ matrix, and whatever we do to a row of $A$ we also do to the corresponding row of $I_3$. In place of using elementary matrices directly, we arrange our computations, using elementary row operations as follows:
Hence

\[
A^{-1} = \begin{bmatrix}
\frac{13}{8} & -\frac{1}{2} & -\frac{1}{8} \\
-\frac{15}{8} & 1 & 3 \\
\frac{5}{4} & 0 & -\frac{1}{4}
\end{bmatrix}
\]

We can readily verify that \(AA^{-1} = A^{-1}A = I_3\).

**Theorem 86.** An \(n \times n\) matrix \(A\) is singular if and only if \(A\) is row equivalent to a matrix \(B\) that has a row of zeros. (That is, the reduced row echelon form of \(A\) has a row of zeros.)

**Remark 87.** This means that in order to find \(A^{-1}\), we do not have to determine, in advance, whether it exists. We merely start to calculate \(A^{-1}\); if at any point in the
computation we find a matrix $B$ that is row equivalent to $A$ and has a row of zeros, then $A^{-1}$ does not exist.

**Example 88.** Let

$$A = \begin{bmatrix} 1 & 2 & -3 \\ 1 & -2 & 1 \\ 5 & -2 & -3 \end{bmatrix}$$

To find $A^{-1}$, we proceed as follows:

At this point $A$ is row equivalent to

$$B = \begin{bmatrix} 1 & 2 & -3 \\ 0 & -4 & 4 \\ 0 & 0 & 0 \end{bmatrix},$$

the last matrix under $A$. Since $B$ has a row of zeros, we stop and conclude that $A$ is a singular matrix.

**Theorem 89.** If $A$ and $B$ are $n \times n$ matrices such that $AB = I_n$ then $BA = I_n$. Thus $B = A^{-1}$.

**Remark 90.** Theorem 96 implies that if we want to check whether a given matrix $B$ is $A^{-1}$, we need merely check whether $AB = I_n$ or $BA = I_n$. That is, we do not have to check both equalities.

## 3 Determinants

Determinants first arose in the solution of linear systems. First, we deal briefly with permutations, which are used in our definition of determinant.
Definition 91. Let $A = [a_{ij}]$ be an $n \times n$ matrix. The determinant function, denoted by $\text{det}$, is defined by

$$\det(A) = \sum (\pm) a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

where the summation is over all permutation $j_1, j_2, \cdots, j_n$ of the set $S = \{1, 2, \ldots, n\}$. The sign is taken as $+$ or $-$ according to whether the permutation $j_1, j_2, \cdots, j_n$ is even or odd.

In each term $(\pm) a_{1j_1} a_{2j_2} \cdots a_{nj_n}$ of $\det(A)$, the row subscripts are in natural order and the column subscripts are in the order $j_1, j_2, \cdots, j_n$. Thus each term in $\det(A)$, with its appropriate sign, is a product of $n$ entries of $A$, with exactly one entry from each row and exactly one entry from each column. Since we sum over all permutations of $S$, $\det(A)$ has $n!$ terms in the sum. Another notation for $\det(A)$ is $|A|$. We shall use both $\det(A)$ and $|A|$.

Example 92. If $A = [a_{11}]$ is a $1 \times 1$ matrix, then $\det(A) = a_{11}$.

Example 93. If

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then to obtain $\det(A)$, we write down the terms $a_{1-} a_{2-}$ and replace the dashes with all possible elements of $S_2$: The subscripts become 12 and 21. Now 12 is an even permutation and 21 is an odd permutation. Thus

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

Hence we see that $\det(A)$ can be obtained by forming the product of the entries on the line from left to right and subtracting from this number the product of the entries on the line from right to left.

Thus, if $A = \begin{bmatrix} 2 & -3 \\ 4 & 5 \end{bmatrix}$, then $|A| = (2)(5) - (-3)(4) = 22$.

Example 94. If

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

then to compute $\det(A)$, we write down the six terms $a_{1-} a_{2-} a_{3-}$, $a_{1-} a_{2-} a_{3-}$, $a_{1-} a_{2-} a_{3-}$, $a_{1-} a_{2-} a_{3-}$, $a_{1-} a_{2-} a_{3-}$, $a_{1-} a_{2-} a_{3-}$. All the elements of $S_3$ are used to replace the dashes, and if we prefix each term by $+$ or $-$ according to whether the permutation is even or odd, we find that (verify)

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$ (7)
We can also obtain $|A|$ as follows. Repeat the first and second columns of $A$, as shown right. Form the sum of the products of the entries on the lines from left to right, and subtract from this number the products of the entries on the lines from right to left (verify):

This method of calculating determinant is called the **rule of Sarrus**.

**Example 95.** Let

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

**Evaluate $|A|$.**

Substituting in (27), we find that


We could obtain the same result by using the easy method illustrated previously, as follows:


**Warning:** The methods used for computing det($A$) in above examples do not apply for $n \geq 4$.

### 3.1 Properties of Determinants

In this section we examine properties of determinants that simplify their computation.

**Theorem 96.** If $A$ is a matrix, then $\det(A) = \det(A^T)$

**Example 97.** Let $A$ be the matrix such that

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 1 \\ 3 & 3 & 2 \end{bmatrix}.$$
So, we have
\[
\]

**Theorem 98.** If matrix \(B\) results from matrix \(A\) by interchanging two different rows (columns) of \(A\), then \(\det(B) = -\det(A)\).

**Example 99.** We have \(|A| = \begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} = -\begin{vmatrix} 3 & 2 \\ 2 & -1 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ -1 & 2 \end{vmatrix} = |A^T| = 7.\)

**Theorem 100.** If two rows (columns) of \(A\) are equal, then \(\det(A) = 0\).

**Proof.** Suppose that rows \(r\) and \(s\) of \(A\) are equal. Interchange rows \(r\) and \(s\) of \(A\) to obtain a matrix \(B\). Then \(\det(B) = -\det(A)\). On the other hand, \(B = A\), so \(\det(B) = \det(A)\). Thus \(\det(A) = -\det(A)\), and so \(\det(A) = 0\).

**Example 101.** We have \( \begin{vmatrix} 1 & 2 & 3 \\ -1 & 0 & 7 \\ 1 & 2 & 3 \end{vmatrix} = 0.\) (Verify by the use of Definition 101.

**Theorem 102.** If a row (column) of \(A\) consists entirely of zeros, then \(\det(A) = 0\).

**Proof.** Let the \(i\)th row of \(A\) consist entirely of zeros. Since each term in Definition 101 for the determinant of \(A\) contains a factor from the \(i\)th row, each term in \(\det(A)\) is zero. Hence \(\det(A) = 0\).

**Theorem 103.** If \(B\) is obtained from \(A\) by multiplying a row (column) of \(A\) by a real number \(k\), then \(\det(B) = k\det(A)\).

**Example 104.** We have \( \begin{vmatrix} 2 & 6 \\ 1 & 12 \end{vmatrix} = 2 \begin{vmatrix} 1 & 3 \\ 1 & 12 \end{vmatrix} = (2)(3) \begin{vmatrix} 1 & 1 \\ 1 & 4 \end{vmatrix} = 6(4 - 1) = 18.\)

**Example 105.** We have \( \begin{vmatrix} 1 & 2 & 3 \\ 1 & 5 & 3 \\ 2 & 8 & 6 \end{vmatrix} = 2 \begin{vmatrix} 1 & 2 & 3 \\ 1 & 5 & 3 \\ 1 & 4 & 3 \end{vmatrix} = (2)(3) \begin{vmatrix} 1 & 5 & 1 \\ 1 & 1 & 1 \end{vmatrix} = (2)(3)(0) = 0.\)

*Here, we first factored out 2 from the third row and 3 from the third column, and then used Theorem 110, since the first and third columns are equal.*

**Theorem 106.** If \(B = [b_{ij}]\) is obtained from \(A = [a_{ij}]\) by adding to each element of the \(r\)th row (column) of \(A\), \(k\) times the corresponding element of the \(s\)th row (column), \(r \neq s\), of \(A\), then \(\det(B) = \det(A)\).
Example 107. We have

\[
\begin{vmatrix}
  1 & 2 & 3 \\
  2 & -1 & 3 \\
  1 & 0 & 1 \\
\end{vmatrix}
= 
\begin{vmatrix}
  5 & 0 & 9 \\
  2 & -1 & 3 \\
  1 & 0 & 1 \\
\end{vmatrix}
\]

obtained by adding twice the second row to the first row. By applying the definition of determinant to the first and second determinant, both are seen to have the value 4.

Theorem 108. If a matrix \( A = [a_{ij}] \) is upper (lower) triangular, then \( \det(A) = a_{11}a_{22}\cdots a_{nn} \) that is, the determinant of a triangular matrix is the product of the elements on the main diagonal.

Recall that in Section 2.1 we introduced compact notation for elementary row and elementary column operations on matrices. In this chapter, we use the same notation for rows and columns of a determinant:

- Interchange rows (columns) \( i \) and \( j \):
  \[ r_i \leftrightarrow r_j (c_i \leftrightarrow c_j) \]

- Replace row (column) \( i \) by \( k \) times row (column) \( i \):
  \[ kr_i \rightarrow r_i (kc_i \rightarrow c_i) \]

- Replace row (column) \( j \) by \( k \) times row (column) \( i \) + row (column) \( j \):
  \[ kr_i + r_j \rightarrow r_j (kc_i + c_j \rightarrow c_j) \]

Using this notation, it is easy to keep track of the elementary row and column operations performed on a matrix. For example, we indicate that we have interchanged the \( i \)th and \( j \)th rows of \( A \) as \( A \leftrightarrow \). We proceed similarly for column operations.

We can now interpret above theorems in terms of this notation as follows:

\[
\begin{align*}
\det(A_{r_i \leftrightarrow r_j}) &= -\det(A), \quad i \neq j \\
\det(A_{kr_i \rightarrow r_i}) &= k\det(A), \quad i \neq j \\
\det(A_{kr_i + r_j \rightarrow r_j}) &= \det(A), \quad i \neq j \\
\end{align*}
\]

It is convenient to rewrite these properties in terms of \( \det(A) \):

\[
\begin{align*}
\det(A) &= -\det(A_{r_i \leftrightarrow r_j}), \quad i \neq j \\
\det(A) &= \frac{1}{k}\det(A_{kr_i \rightarrow r_i}), \quad k \neq 0 \\
\det(A) &= \det(A_{kr_i + r_j \rightarrow r_j}), \quad i \neq j.
\end{align*}
\]
Theorems above are useful in evaluating determinants. What we do is transform $A$ by means of our elementary row or column operations to a triangular matrix. Of course, we must keep track of how the determinant of the resulting matrices changes as we perform the elementary row or column operations.

**Example 109.** Let $A = \begin{bmatrix} 4 & 3 & 2 \\ 3 & -2 & 5 \\ 2 & 4 & 6 \end{bmatrix}$. Compute $\det(A)$.

**Solution:** We have

$$\det(A) = 2 \det \left( A_{1 \to 3} \right)$$

Multiply row 3 by $\frac{1}{2}$.

$$= 2 \det \begin{bmatrix} 4 & 3 & 2 \\ 3 & -2 & 5 \\ 1 & 2 & 3 \end{bmatrix}$$

Interchange rows 1 and 3.

$$= (-1)2 \det \begin{bmatrix} 4 & 3 & 2 \\ 3 & -2 & 5 \\ 1 & 2 & 3 \end{bmatrix}_{r1 \leftrightarrow r3}$$

Zero out below the $(1,1)$ entry.

$$= -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 3 & -2 & 5 \\ 4 & 3 & 2 \end{bmatrix}$$

$$= -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 3 & -2 & 5 \\ 4 & 3 & 2 \end{bmatrix}_{-3r1 + r2 \to r2}$$

$$= -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & -8 & -4 \\ 0 & -5 & -10 \end{bmatrix}_{-4r1 + r3 \to r3}$$

$$= -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & -8 & -4 \\ 0 & -5 & -10 \end{bmatrix}_{-5r2 + r3 \to r3}$$

$$= -2 \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & -8 & -4 \\ 0 & 0 & \frac{30}{4} \end{bmatrix}.$$
Next we compute the determinant of the upper triangular matrix.

$$\det(A) = -2(1)(-8) \left( -\frac{30}{4} \right) = -120.$$  

The operations chosen are not the most efficient, but we do avoid fractions during the first few steps.

Remark 110. The method used to compute a determinant in Example above will be referred to as computation via reduction to triangular form.

Theorem 111. If $A$ is an $n \times n$ matrix, then $A$ is nonsingular if and only if $\det(A) \neq 0$.

Theorem 112. If $A$ and $B$ are $n \times n$ matrices, then $\det(AB) = \det(A) \det(B)$.

Example 113. Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$  

Then

$$|A| = -2 \quad \text{and} \quad |B| = 5.$$  

On the other hand, $AB = \begin{bmatrix} 4 & 3 \\ 10 & 5 \end{bmatrix}$, and $|AB| = -10 = |A||B|$.

Theorem 114. If $A$ is nonsingular, then $\det(A^{-1}) = \frac{1}{\det(A)}$.

3.2 Cofactor Expansion

Thus far we have evaluated determinants by using Definition 3.2 and the properties established in Section 3.2. We now develop a method for evaluating the determinant of an $n \times n$ matrix that reduces the problem to the evaluation of determinants of matrices of order $n-1$. We can then repeat the process for these $(n-1) \times (n-1)$ matrices until we get to $2 \times 2$ matrices.

Definition 115. Let $A = [a_{ij}]$ be an $n \times n$ matrix. Let $M_{ij}$ be the $(n - 1) \times (n - 1)$ submatrix of $A$ obtained by deleting the $i$th row and $j$th column of $A$. The determinant $\det(M_{ij})$ is called the minor of $a_{ij}$.

Definition 116. Let $A = [a_{ij}]$ be an $n \times n$ matrix. The cofactor $A_{ij}$ of $a_{ij}$ is defined as $A_{ij} = (-1)^{i+j} \det(M_{ij})$.

Example 117. Let

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 4 & 5 & 6 \\ 7 & 1 & 2 \end{bmatrix}.$$
Then \( \det(M_{12}) = \begin{vmatrix} 4 & 6 \\ 7 & 2 \end{vmatrix} = 8 - 42 = -34, \) \( \det(M_{23}) = \begin{vmatrix} 3 & -1 \\ 7 & 1 \end{vmatrix} = 3 + 7 = 10, \) and \\
\( \det(M_{31}) = \begin{vmatrix} -1 & 2 \\ 5 & 6 \end{vmatrix} = -6 - 10 = -16. \)

Also, 
\[ A_{12} = (-1)^{1+2} \det(M_{12}) = (-1)(-34) = 34, \]
\[ A_{23} = (-1)^{2+3} \det(M_{23}) = (-1)(10) = -10, \]
and 
\[ A_{31} = (-1)^{3+1} \det(M_{31}) = (1)(-16) = -16. \]

If we think of the sign \((-1)^{i+j}\) as being located in position \((i,j)\) of an \(n \times n\) matrix, then the signs form a checkerboard pattern that has a + in the \((1,1)\) position. The patterns for \(n=3\) and \(n=4\) are as follows:

\[
\begin{array}{ccc}
+ & - & + \\
- & + & - \\
+ & - & + \\
\end{array}
\quad
\begin{array}{ccc}
+ & - & + \\
- & + & - \\
- & + & - \\
+ & - & + \\
\end{array}
\quad
n=3
\quad
n=4
\]

**Theorem 118.** Let \(A = [a_{ij}]\) be an \(n \times n\) matrix. Then
\[
\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}
\]
[ expansion of \(\det(A)\) along the \(i\)th row ]

and
\[
\det(A) = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}
\]
[ expansion of \(\det(A)\) along the \(j\)th column ].

**Example 119.** To evaluate the determinant
\[
\begin{vmatrix} 1 & 2 & -3 & 4 \\ -4 & 2 & 1 & 3 \\ 3 & 0 & 0 & -3 \\ 2 & 0 & -2 & 3 \end{vmatrix}
\]
it is best to expand along either the second column or the third row because they each have two zeros. Obviously, the optimal course of action is to expand along the row or column that has the largest number of zeros, because in that case the cofactors \(A_{ij}\) of those \(a_{ij}\) which are zero do not have to be evaluated, since \(a_{ij}A_{ij} = (0)(A_{ij}) = 0. \) Thus, expanding along the third row, we have
Example 120. We have

\[
\begin{vmatrix}
1 & 2 & -3 & 4 \\
-4 & 2 & 1 & 3 \\
3 & 0 & 0 & -3 \\
2 & 0 & -2 & 3 \\
\end{vmatrix} = (-1)^{3+1}(3)
\begin{vmatrix}
2 & -3 & 4 \\
2 & 1 & 3 \\
0 & -2 & 3 \\
-4 & 1 & 3 \\
\end{vmatrix}
+ (-1)^{3+4}(0)
\begin{vmatrix}
1 & 2 & -3 \\
-4 & 2 & 3 \\
2 & 0 & 3 \\
-4 & 2 & 1 \\
\end{vmatrix}
+ (-1)^{3+4}(-3)
\begin{vmatrix}
1 & 2 & -3 \\
-4 & 2 & 3 \\
2 & 0 & 3 \\
-2 & 0 & 2 \\
\end{vmatrix}
= (+1)(3)(20) + 0 + 0 + (-1)(-3)(-4) = 48.
\]

3.3 Inverse of a Matrix

We saw in Section 3.3 that Theorem 3.10 provides formulas for expanding \( \det(A) \) along either a row or a column of \( A \). Thus

\[
\det(A) = a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn}
\]

is the expansion of \( \det(A) \) along the \( i \)th row. It is interesting to ask what

\[
a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn} = 0
\]

for \( i \neq k \), because as soon as we answer this question, we obtain another method for finding the inverse of a nonsingular matrix.

Theorem 121. If \( A = [a_{ij}] \) is an \( n \times n \) matrix, then

\[
a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn} = 0 \text{ for } i \neq k,
\]

\[
a_{1j}A_{1k} + a_{2j}A_{2k} + \cdots + a_{nj}A_{nk} = 0 \text{ for } j \neq k.
\]

Example 122. Let \( A = \begin{vmatrix}
1 & 2 & 3 \\
-2 & 3 & 1 \\
4 & 5 & -2
\end{vmatrix} \). Then

\[
A_{21} = (-1)^{2+1}
\begin{vmatrix}
2 & 3 \\
5 & -2
\end{vmatrix} = 19, \quad A_{22} = (-1)^{2+2}
\begin{vmatrix}
1 & 3 \\
4 & -2
\end{vmatrix} = -14, \quad \text{and} \quad A_{23} =
\]

\[
(-1)^{2+3}
\begin{vmatrix}
1 & 2 \\
4 & 5
\end{vmatrix} = 3.
\]

Now

\[
a_{31}A_{21} + a_{32}A_{22} + a_{33}A_{23} = (4)(19) + (5)(-14) + (-2)(3) = 0,
\]

and

\[
a_{11}A_{21} + a_{12}A_{22} + a_{13}A_{23} = (1)(19) + (2)(-14) + (3)(3) = 0.
\]
We may summarize our expansion results by writing

\[ a_{i1}A_{k1} + a_{i2}A_{k2} + \cdots + a_{in}A_{kn} = \det(A) \quad \text{if } i = k \]
\[ = 0 \quad \text{if } i \neq k \]

and

\[ a_{1j}A_{1k} + a_{2j}A_{2k} + \cdots + a_{nj}A_{nk} = \det(A) \quad \text{if } j = k \]
\[ = 0 \quad \text{if } j \neq k. \]

**Definition 123.** Let \( A = [a_{ij}] \) be an \( n \times n \) matrix. The \( n \times n \) matrix \( \text{adj } A \), called the **adjoint** of \( A \), is the matrix whose \((i,j)\) th entry is the cofactor \( A_{ji} \) of \( a_{ji} \). Thus

\[
\text{adj } A = 
\begin{bmatrix}
A_{11} & A_{21} & \cdots & A_{n1} \\
A_{12} & A_{22} & \cdots & A_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
A_{1n} & A_{2n} & \cdots & A_{nn}
\end{bmatrix}.
\]

**Example 124.** Let \( A = \begin{bmatrix} 3 & -2 & 1 \\ 5 & 6 & 2 \\ 1 & 0 & -3 \end{bmatrix} \). Compute \( \text{adj } A \)

**Solution:** We first compute the cofactors of \( A \). We have

\[
A_{11} = (-1)^{1+1} \begin{vmatrix} 6 & 2 \\ 0 & -3 \end{vmatrix} = -18, \quad A_{12} = (-1)^{1+2} \begin{vmatrix} 5 & 1 \\ 1 & 0 \end{vmatrix} = 17, \\
A_{13} = (-1)^{1+3} \begin{vmatrix} 5 & 6 \\ 1 & 0 \end{vmatrix} = -6, \quad A_{21} = (-1)^{2+1} \begin{vmatrix} -2 & 1 \\ 0 & -3 \end{vmatrix} = -6, \\
A_{22} = (-1)^{2+2} \begin{vmatrix} 3 & 1 \\ 1 & -3 \end{vmatrix} = -10, \quad A_{23} = (-1)^{2+3} \begin{vmatrix} 3 & -2 \\ 1 & 0 \end{vmatrix} = -2, \\
A_{31} = (-1)^{3+1} \begin{vmatrix} -2 & 1 \\ 6 & 2 \end{vmatrix} = -10, \\
A_{32} = (-1)^{3+2} \begin{vmatrix} 3 & 1 \\ 5 & 2 \end{vmatrix} = -1, \quad A_{33} = (-1)^{3+3} \begin{vmatrix} 3 & -2 \\ 5 & 6 \end{vmatrix} = 28.
\]

Then

\[
\text{adj } A = 
\begin{bmatrix}
-18 & -6 & -10 \\
17 & -10 & -1 \\
-6 & -2 & 28
\end{bmatrix}.
\]

**Theorem 125.** If \( A = [a_{ij}] \) is an \( n \times n \) matrix, then \( A(\text{adj } A) = (\text{adj } A)A = \det(A)I_n \).
Example 126. Consider the matrix of example above, then
\[
\begin{bmatrix}
3 & -2 & 1 \\
5 & 6 & 2 \\
1 & 0 & -3
\end{bmatrix}
\begin{bmatrix}
-18 & -6 & -10 \\
17 & -10 & -1 \\
-6 & -2 & 28
\end{bmatrix}
= \begin{bmatrix}
-94 & 0 & 0 \\
0 & -94 & 0 \\
0 & 0 & -94
\end{bmatrix}
= -94 \begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}
\]
and
\[
\begin{bmatrix}
-18 & -6 & -10 \\
17 & -10 & -1 \\
-6 & -2 & 28
\end{bmatrix}
\begin{bmatrix}
3 & -2 & 1 \\
5 & 6 & 2 \\
1 & 0 & -3
\end{bmatrix}
= -94 \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

Corollary 127. If $A$ is an $n \times n$ matrix and $\det(A) \neq 0$. Then

\[
A^{-1} = \frac{1}{\det(A)} (\text{adj } A) = \begin{bmatrix}
\frac{A_{11}}{\det(A)} & \frac{A_{21}}{\det(A)} & \cdots & \frac{A_{n1}}{\det(A)} \\
\frac{A_{12}}{\det(A)} & \frac{A_{22}}{\det(A)} & \cdots & \frac{A_{n2}}{\det(A)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{A_{1n}}{\det(A)} & \frac{A_{2n}}{\det(A)} & \cdots & \frac{A_{nn}}{\det(A)}
\end{bmatrix}
\]

Example 128. Again consider the matrix $A = \begin{bmatrix}
3 & -2 & 1 \\
5 & 6 & 2 \\
1 & 0 & -3
\end{bmatrix}$. Then $\det(A) = -94$
and

\[
A^{-1} = \frac{1}{\det(A)} (\text{adj } A) = \begin{bmatrix}
\frac{18}{94} & \frac{6}{94} & \frac{10}{94} \\
\frac{94}{17} & \frac{94}{10} & \frac{94}{1} \\
\frac{94}{6} & \frac{94}{2} & \frac{94}{28}
\end{bmatrix}
\]

Summary 129. Note that at this point we have shown that the following statements are equivalent for an $n \times n$ matrix $A$:

1. $A$ is nonsingular.
2. $A$ is row (column) equivalent to $I_n$.
3. $\det(A) \neq 0$.  

42
Alternative Definition for the Rank of a Matrix: The rank of a matrix can also be calculated using determinants. The rank of a matrix $A$ is $r$, where $r$ is the size of the largest non-zero $r \times r$ submatrix with non-zero determinant.

Example 130. Let $A = \begin{bmatrix} 2 & 1 & 3 & 2 & 0 \\ 3 & 2 & 5 & 1 & 0 \\ -1 & 1 & 0 & -7 & 0 \\ 3 & -2 & 1 & 17 & 0 \\ 0 & 1 & 1 & -4 & 0 \end{bmatrix}$. Find the rank of $A$.

* Column 5 can be discarded since it has all zero elements.

* Column 3 can be discarded since it is a linear combination of column 1 and column 2, that is, $c_3 = c_1 + c_2$.

Therefore we can consider the matrix $\begin{bmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ -1 & 1 & -7 \end{bmatrix}$.

\[ \begin{bmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ -1 & 1 & -7 \end{bmatrix} \]

i) Since the matrix is $5 \times 3$ we will look first to the determinants of the submatrices which have the dimension $3 \times 3$.

All $3 \times 3$ submatrices of the above matrix are and their determinants are:

\[
\begin{array}{c|c}
\begin{vmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ -1 & 1 & -7 \end{vmatrix} &= 28 - 1 + 6 + 4 + 21 - 2 = 0, & \begin{vmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ -1 & 1 & -7 \end{vmatrix} &= 68 - 3 - 12 - 12 - 51 + 4 = 0 \\
\begin{vmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ 0 & 1 & -4 \end{vmatrix} &= -16 + 6 + 12 - 2 = 0, & \begin{vmatrix} 3 & 2 & 1 \\ -1 & 1 & -7 \\ 3 & -2 & 17 \end{vmatrix} &= 34 - 21 + 4 - 17 - 28 = 0 \\
\begin{vmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ 0 & 1 & -4 \end{vmatrix} &= 28 - 1 + 6 + 4 + 21 - 2 = 0, & \begin{vmatrix} 3 & 2 & 1 \\ -1 & 1 & -7 \\ 3 & -2 & 17 \end{vmatrix} &= 16 + 6 + 12 - 34 = 0 \\
\begin{vmatrix} 2 & 1 & 2 \\ 3 & 2 & 1 \\ 0 & 1 & -4 \end{vmatrix} &= -16 + 6 + 12 - 2 = 0, & \begin{vmatrix} 3 & 2 & 1 \\ -1 & 1 & -7 \\ 3 & -2 & 17 \end{vmatrix} &= 51 - 42 + 2 - 3 - 42 + 34 = 0, & \begin{vmatrix} 3 & 2 & 1 \\ 3 & -2 & 17 \\ 0 & 1 & -4 \end{vmatrix} &= 24 + 3 + 24 - 51 = 0 \\
\begin{vmatrix} 3 & 2 & 1 \\ -1 & 1 & -7 \\ 3 & -2 & 17 \end{vmatrix} &= -8 - 21 + 12 + 17 = 0, & \begin{vmatrix} 3 & 2 & 1 \\ 3 & -2 & 17 \\ 0 & 1 & -4 \end{vmatrix} &= -12 - 1 - 8 + 21 = 0 \\
\end{array}
\]
Therefore, we have to look to the determinants of $2 \times 2$ submatrices. Since, \[
\begin{vmatrix}
2 & 1 \\
3 & 2
\end{vmatrix} = 2 \cdot 2 - 3 \cdot 1 = 1 \neq 0,
\] the rank of the matrix $A$ is 2.

The second midterm topics start here!!!

4 Linear Equation Systems

One of the most frequently recurring practical problems in many fields of study—such as mathematics, physics, biology, chemistry, economics, all phases of engineering, operations research, and the social sciences—is that of solving a system of linear equations. The equation

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$$

(8)

which expresses $b$ in terms of the unknowns $x_1, x_2, \ldots, x_n$ and the constants $a_1, a_2, \ldots, a_n$ is called a linear equation. In many applications we are given $b$ and must find numbers $x_1, x_2, \ldots, x_n$ satisfying (8). A solution to linear Equation (8) is a sequence of $n$ numbers $s_1, s_2, \ldots, s_n$, which has the property that (8) is satisfied when $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ are substituted in (8). Thus $x_1 = 2, x_2 = 3, x_3 = -4$ is a solution to the linear equation

$$6x_1 - 3x_2 + 4x_3 = -13$$

because

$$6(2) - 3(3) + 4(-4) = -13$$

More generally, a system of $m$ linear equations in $n$ unknowns, $x_1, x_2, \ldots, x_n$ or a linear system, is a set of $m$ linear equations each in $n$ unknowns. A linear system can conveniently be written as

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

(9)

Thus the $i$th equation is

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i.$$

In (9) the $a_{ij}$ are known constants. Given values of $b_1, b_2, \ldots, b_m$, we want to find values of $x_1, x_2, \ldots, x_n$ that will satisfy each equation in (9). A solution to linear system (9) is a sequence of $n$ numbers $s_1, s_2, \ldots, s_n$, which has the property that equation in (9) is satisfied when $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ are substituted.
If the linear system (9) has no solution, it is said to be \textit{inconsistent}; if it has a solution, it is called \textit{consistent}. If \( b_1 = b_2 = \cdots = b_m = 0 \), then (9) is called a \textit{homogeneous system}. Note that \( x_1 = x_2 = \cdots = x_n = 0 \) is always a solution to a homogeneous system; it is called the \textit{trivial solution}. A solution to a homogeneous system in which not all of \( x_1, x_2, \ldots, x_n \) are zero is called a \textit{nontrivial solution}.

Consider another system of \( r \) linear equations in \( n \) unknowns:

\[
\begin{align*}
c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n &= d_1 \\
c_{21}x_1 + c_{22}x_2 + \cdots + c_{2n}x_n &= d_2 \\
&\vdots \\
c_{r1}x_1 + c_{r2}x_2 + \cdots + c_{rn}x_n &= d_r
\end{align*}
\]

(10)

We say that (9) and (10) are \textit{equivalent} if they both have exactly the same solutions.

\textbf{Example 131.} The linear system

\[
\begin{align*}
x_1 - 3x_2 &= -7 \\
2x_1 + x_2 &= 7
\end{align*}
\]

(11)

has only the solution \( x_1 = 2 \) and \( x_2 = 3 \). The linear system

\[
\begin{align*}
8x_1 - 3x_2 &= 7 \\
3x_1 - 2x_2 &= 0 \\
10x_1 - 2x_2 &= 14
\end{align*}
\]

(12)

also has only the solution \( x_1 = 2 \) and \( x_2 = 3 \). Thus (11) and (12) are equivalent.

\section*{4.1 Solving Linear systems}

Consider the linear system of \( m \) equations in \( n \) unknowns,

\[
\begin{align*}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{align*}
\]

(13)

Now define the following matrices:

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

45
Then

\[
A \mathbf{x} = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{bmatrix} \quad (14)
\]

The entries in the product \( A \mathbf{x} \) at the end of (14) are merely the left sides of the equations in (13). Hence the linear system (13) can be written in matrix form as

\[ A \mathbf{x} = \mathbf{b}. \]

The matrix \( A \) is called the \textit{coefficient matrix} of the linear system (13), and the matrix

\[
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m
\end{bmatrix}
\]

obtained by adjoining column \( \mathbf{b} \) to \( A \), is called the \textit{augmented matrix} of the linear system (13). The augmented matrix of (13) is written as \([A : \mathbf{b}]\). Conversely, any matrix with more than one column can be thought of as the augmented matrix of a linear system. The coefficient and augmented matrices play key roles in our method for solving linear systems.

Recall from the previous section that if

\[ b_1 = b_2 = \cdots = b_m = 0 \]

in (14), the linear system is called a homogeneous system. A homogeneous system can be written as

\[ A \mathbf{x} = \mathbf{0} \]

where \( A \) is the coefficient matrix.

Let

\[
[A : \mathbf{b}] = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m
\end{bmatrix}
\]

be the augmented matrix of the linear system (13). Using elementary row operations, we can transform \([A : \mathbf{b}]\) to \([C : \mathbf{d}]\) which is in reduced row echelon form. Therefore,
1) If \( r_C \neq r \) where \( r_C \) is the rank of the matrix \( C \) (the number of nonzero rows in \( C \)) then the linear system (13) has no solution (inconsistent).

2) If \( r_C = r \) then the linear system (13) has a solution (consistent):
   
a) If \( r = n \) then the linear system (13) has a unique solution.
   
b) If \( r < n \) then the linear system (13) has infinitely many solutions with respect to \( n - r \) arbitrary variables.

This method of solving linear systems is called **Gauss-Jordan reduction** or **solving linear systems via equivalent matrices**.

**Example 132.** The linear system

\[
\begin{align*}
x + 2y + 3z &= 9 \\
2x - y + z &= 8 \\
3x - z &= 3
\end{align*}
\]

has the augmented matrix

\[
[A : b] = \begin{bmatrix}
1 & 2 & 3 : 9 \\
2 & -1 & 1 : 8 \\
3 & 0 & -1 : 3 \\
\end{bmatrix}
\]

Transforming this matrix to reduced row echelon form, we obtain (verify)

\[
[C : d] = \begin{bmatrix}
1 & 0 & 0 : 2 \\
0 & 1 & 0 : 1 \\
0 & 0 & 1 : 3 \\
\end{bmatrix}.
\]

Therefore, \( r_C = r \) = 3 = \( n \) and linear system has a unique solution which is \( x = 2, y = -1, z = 3 \).

**Example 133.** Solve the linear system

\[
\begin{align*}
x + y + 2z &= 1 \\
2y + 7z &= 4 \\
3x + 3y + 6z &= 3.
\end{align*}
\]

The augmented matrix of the linear system is

\[
[A : b] = \begin{bmatrix}
1 & 1 & 2 : 1 \\
0 & 2 & 7 : 4 \\
3 & 3 & 6 : 3 \\
\end{bmatrix}
\]
By using elementary row operations,

\[
\begin{bmatrix}
1 & 1 & 2 & 1 \\
0 & 2 & 7 & 4 \\
3 & 3 & 6 & 3 \\
\end{bmatrix}
\xrightarrow{r_2 \rightarrow r_2 - \frac{1}{2}r_1 \rightarrow r_2}
\begin{bmatrix}
1 & 1 & 2 & 1 \\
0 & 1 & \frac{7}{2} & 2 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\xrightarrow{-r_1 + r_3 \rightarrow r_3}
\begin{bmatrix}
1 & 0 & -\frac{3}{2} & -\frac{1}{2} \\
0 & 1 & \frac{7}{2} & 2 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}.
\]

Hence, \( r_C = r_{[C : d]} = 2 < 3 = n \). So, the linear system has infinitely many solutions with respect to \( 3 - 2 = 1 \) arbitrary variable. Using back substitution, we have

\[
x - \frac{3}{2}z = -1 \\
y + \frac{7}{2}z = 2 \\
z = \text{any real number}.
\]

So, the solution for the linear system is:

\[
x = -1 + \frac{3}{2}r \\
y = 2 - \frac{7}{2}r \\
z = r, \ \text{any real number}.
\]

Example 134. Solve the linear system

\[
\begin{align*}
3x + 2y &= 7 \\
17x + y &= 0 \\
6x + 4y &= 3.
\end{align*}
\]

The augmented matrix of the linear system is

\[
[A : b] = \begin{bmatrix}
3 & 2 & 7 \\
17 & 1 & 0 \\
6 & 4 & 3
\end{bmatrix}
\]

Transforming this matrix to row echelon form, we obtain (verify)

\[
[C : d] = \begin{bmatrix}
1 & 2/3 & 7/3 \\
0 & -31/3 & -119/3 \\
0 & 0 & -11
\end{bmatrix}.
\]

Since, \( r_C = 2 \neq r_{[C : d]} = 3 \) linear system has no solution (inconsistent). This can be verified by the last equation: \( 0x + 0y = -11 \) which can never be satisfied.
Homogeneous Systems

Now we study a homogeneous system $Ax = 0$ of $m$ linear equations in $n$ unknowns. Let $[C : 0]$ be a reduced row echelon form of the augmented matrix $[A : 0]$. Since we always have $r_C = r_{[C : 0]}$, a homogeneous system always has a solution:

1) If $r = n$, where $r = \text{rank}(C)$, the homogeneous system has only zero (trivial) solution.

2) If $r < n$ then the homogeneous system has a nontrivial solution; infinitely many solutions with respect to $n - r$ arbitrary variables.

Example 135. Solve the homogeneous system

$$
\begin{align*}
-3x + 2y - 3z &= 0 \\
2x + 5y + 2z &= 0 \\
4x + y - 3z &= 0.
\end{align*}
$$

The augmented matrix of the homogeneous system is

$$
[A : 0] = 
\begin{bmatrix}
-3 & 2 & -3 & 0 \\
2 & 5 & 2 & 0 \\
4 & 1 & -3 & 0
\end{bmatrix}
$$

By using elementary row operations,

$$
\begin{align*}
\begin{bmatrix}
-3 & 2 & -3 & 0 \\
2 & 5 & 2 & 0 \\
4 & 1 & -3 & 0
\end{bmatrix}
&\xrightarrow{r_1 + r_3 \rightarrow r_3}
\begin{bmatrix}
1 & 3 & -6 & 0 \\
2 & 5 & 2 & 0 \\
0 & 11 & -21 & 0
\end{bmatrix}
\xrightarrow{r_2 \rightarrow r_2}
\begin{bmatrix}
1 & 3 & -6 & 0 \\
0 & 1 & -14 & 0 \\
0 & 11 & -21 & 0
\end{bmatrix}
\xrightarrow{r_3 \rightarrow r_3}
\begin{bmatrix}
1 & 0 & 36 & 0 \\
0 & 1 & -14 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\xrightarrow{\frac{1}{175}r_3 \rightarrow r_3}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}.
\end{align*}
$$

Since $r = 3 = n$, the homogeneous system has only trivial (zero) solution, that is, $x = y = z = 0$.

Theorem 136. A homogeneous system of $m$ linear equations in $n$ unknowns always has a nontrivial solution if $m < n$, that is, if the number of unknowns exceeds the number of equations.
Cramer’s Rule

Let

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
    a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
    &\vdots \\
    a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{align*}
\]

be a linear system of \( n \) equations in \( n \) unknowns, and let \( A = [a_{ij}] \) be the coefficient matrix so that we can write the given system as \( Ax = b \), where

\[
b = \begin{bmatrix}
    b_1 \\
    b_2 \\
    \vdots \\
    b_n
\end{bmatrix}.
\]

If \( \det(A) \neq 0 \), then the system has the unique solution

\[
x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)}
\]

where \( A_i \) is the matrix obtained from \( A \) by replacing the \( i \)th column of \( A \) by \( b \).

**Example 137.** Consider the following linear system:

\[
\begin{align*}
    -2x_1 + 3x_2 - x_3 &= 1 \\
    x_1 + 2x_2 - x_3 &= 4 \\
    -2x_1 - x_2 + x_3 &= -3.
\end{align*}
\]

We have \( \det(A) = \begin{vmatrix}
    -2 & 3 & -1 \\
    1 & 2 & -1 \\
    -2 & -1 & 1
\end{vmatrix} = -2 \). Then
\[
x_1 = \frac{\begin{vmatrix} 1 & 3 & -1 \\ 4 & 2 & -1 \\ -3 & -1 & 1 \end{vmatrix}}{|A|} = \frac{-4}{-2} = 2,
\]

\[
x_2 = \frac{\begin{vmatrix} -2 & 1 & -1 \\ 1 & 4 & -1 \\ -2 & -3 & 1 \end{vmatrix}}{|A|} = \frac{-6}{-2} = 3,
\]

\[
x_3 = \frac{\begin{vmatrix} -2 & 3 & 1 \\ 1 & 2 & 4 \\ -2 & -1 & -3 \end{vmatrix}}{|A|} = \frac{-8}{-2} = 4.
\]

**Remark 138.** We note that Cramer’s rule is applicable only to the case in which we have \( n \) equations in \( n \) unknowns and the coefficient matrix \( A \) is nonsingular. If we have to solve a linear system of \( n \) equations in \( n \) unknowns whose coefficient matrix is singular, then we must use the Gaussian elimination or Gauss-Jordan reduction methods as discussed before. Cramer’s rule becomes computationally inefficient for \( n \geq 4 \), so it is better, in general, to use the Gaussian elimination or Gauss-Jordan reduction methods.

**Summary 139.** Note that at this point we have shown that the following statements are equivalent for an \( n \times n \) matrix \( A \):

1. \( A \) is nonsingular.
2. \( Ax = 0 \) has only the trivial solution.
3. \( A \) is row (column) equivalent to \( I_n \).
4. The linear system \( Ax = b \) has a unique solution for every \( n \times 1 \) matrix \( b \).
5. \( \det(A) \neq 0 \).
Solving Linear Systems via the Inverse of a Coefficient Matrix

Consider the linear system of \( n \) equations in \( n \) unknowns.

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
    a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
    & \vdots \\
    a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{align*}
\]

Let \( A = [a_{ij}] \) be the coefficient matrix of the linear system so that we can write the given system as \( Ax = b \), where

\[
b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.
\]

If \( \det(A) \neq 0 \) then \( A \) is nonsingular (invertible) and \( A^{-1} \) exists. Therefore, the equation \( Ax = b \) becomes:

\[
A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b.
\]

So, we can solve the linear system by using the inverse of the coefficient matrix.

Example 140. Solve the following linear system by using the inverse the coefficient matrix.

\[
\begin{align*}
    x + 3y + z &= 5 \\
    2x + y + z &= 4 \\
    -2x + 2y - z &= 3
\end{align*}
\]

We have \( A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 1 & 1 \\ -2 & 2 & -1 \end{bmatrix} \). Then \( A^{-1} = \frac{1}{\det(A)} \text{adj} A = \begin{bmatrix} -1 & 5/3 & 2/3 \\ 0 & 1/3 & 1/3 \\ 2 & -8/3 & -5/3 \end{bmatrix} \).

Therefore,

\[
x = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1}b = \begin{bmatrix} -1 & 5/3 & 2/3 \\ 0 & 1/3 & 1/3 \\ 2 & -8/3 & -5/3 \end{bmatrix} \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 11/3 \\ 7/3 \\ -17/3 \end{bmatrix}.
\]

So, \( x = \frac{11}{3}, y = \frac{7}{3} \) and \( z = -\frac{17}{3} \).
5 Vectors in the Plane and in 3-Space

Definition 141. A (nonzero) vector is a directed line segment drawn from a point \( P \) (called its initial point) to a point \( Q \) (called its terminal point) \( P \) and \( Q \) being distinct, points. The vector is denoted by \( \overrightarrow{PQ} \). Its magnitude is the length of the line segment, denoted by \( ||\overrightarrow{PQ}|| \) and its direction is the same as the directed line segment.

The zero vector is just a point, and it is denoted by \( \mathbf{0} \).

To indicate the direction of a vector, we draw an arrow from its initial point to its terminal point. We will often denote a vector by a single bold-faced letter (e.g. \( \mathbf{v} \)) and use the terms “magnitude” and “length” interchangeably. Note that our definition could apply to systems with any number of dimensions.

A few things need to be noted about the zero vector. Our motivation for what a vector is included the notions of magnitude and direction. What is the magnitude of the zero vector? We define it to be zero, i.e. \( ||\mathbf{0}|| = 0 \). This agrees with the definition of the zero vector as just a point, which has zero length. What about the direction of the zero vector? A single point really has no well-defined direction. Notice that we were careful to only define the direction of a nonzero vector, which is well-defined since the initial and terminal points are distinct. Not everyone agrees on the direction of the zero vector. Some contend that the zero vector has arbitrary direction (i.e. can take any direction), some say that it has indeterminate direction (i.e. the direction can not be determined), while others say that it has no direction.

Now that we know what a vector is, we need a way of determining when two vectors are equal. This leads us to the following definition.

Definition 142. Two nonzero vectors are equal if they have the same magnitude and the same direction. Any vector with zero magnitude is equal to the zero vector.
By this definition, vectors with the same magnitude and direction but with different initial points would be equal. For example, in Figure 1.1.5 the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ all have the same magnitude $\sqrt{5}$ (by the Pythagorean Theorem). And we see that $\mathbf{u}$ and $\mathbf{w}$ are parallel, since they lie on lines having the same slope $\frac{1}{2}$, and they point in the same direction. So $\mathbf{u} = \mathbf{w}$, even though they have different initial points. We also see that $\mathbf{v}$ is parallel to $\mathbf{u}$ but points in the opposite direction. So $\mathbf{u} \neq \mathbf{v}$.

Recall the distance formula for points in the Euclidean plane:

For points $P = (x_1, y_1), Q = (x_2, y_2)$ in $\mathbb{R}^2$, the distance $d$ between $P$ and $Q$ is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

By this formula, we have the following result: For a vector $\overrightarrow{PQ}$ in $\mathbb{R}^2$ with initial point $P = (x_1, y_1)$ and terminal point $Q = (x_2, y_2)$, the magnitude of $\overrightarrow{PQ}$ is:

$$\|\overrightarrow{PQ}\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Finding the magnitude of a vector $\mathbf{v} = (a, b)$ in $\mathbb{R}^2$ is a special case of above formula with $P = (0, 0)$ and $Q = (a, b)$: For a vector $\mathbf{v} = (a, b)$ in $\mathbb{R}^2$, the magnitude of $\mathbf{v}$ is:

$$\|\mathbf{v}\| = \sqrt{a^2 + b^2}$$

**Theorem 143.** The distance $d$ between points $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ in $\mathbb{R}^3$ is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

**Theorem 144.** For a vector $\mathbf{v} = (a, b, c)$ in $\mathbb{R}^3$, the magnitude of $\mathbf{v}$ is:

$$\|\mathbf{v}\| = \sqrt{a^2 + b^2 + c^2}$$

### 5.1 Vector Algebra

**Definition 145.** A scalar is a quantity that can be represented by a single number.

**Definition 146.** For a scalar $k$ and a nonzero vector $\mathbf{v}$, the scalar multiple of $\mathbf{v}$ by $k$, denoted by $k\mathbf{v}$, is the vector whose magnitude is $|k|\|\mathbf{v}\|$, points in the same direction as $\mathbf{v}$ if $k > 0$, points in the opposite direction as $\mathbf{v}$ if $k < 0$, and is the zero vector $0$ if $k = 0$. For the zero vector $\mathbf{0}$, we define $k\mathbf{0} = \mathbf{0}$ for any scalar $k$. 
Two vectors $v$ and $w$ are parallel (denoted by $v \parallel w$) if one is a scalar multiple of the other. You can think of scalar multiplication of a vector as stretching or shrinking the vector, and as flipping the vector in the opposite direction if the scalar is a negative number (see the below figure).

Recall that *translating* a nonzero vector means that the initial point of the vector is changed but the magnitude and direction are preserved. We are now ready to define the sum of two vectors.

**Definition 147.** The sum of vectors $v$ and $w$, denoted by $v + w$, is obtained by translating $w$ so that its initial point is at the terminal point of $v$; the initial point of $v + w$ is the initial point of $v$, and its terminal point is the new terminal point of $w$.

The term scalar was invented by 19th century Irish mathematician, physicist and astronomer William Rowan Hamilton, to convey the sense of something that could be represented by a point on a scale or graduated ruler. The word vector comes from Latin, where it means “carrier”.

An alternate definition of scalars and vectors, used in physics, is that under certain types of coordinate transformations (e.g. rotations), a quantity that is not affected is a scalar, while a quantity that is affected (in a certain way) is a vector. See MARION for details.

Intuitively, adding $w$ to $v$ means tacking on $w$ to the end of $v$ (see Figure below).

![Vectors, Translating, and Sum](image)

(a) Vectors $v$ and $w$  
(b) Translate $w$ to the end of $v$  
(c) The sum $v + w$

Notice that our definition is valid for the zero vector (which is just a point, and hence can be translated), and so we see that $v + 0 = v = 0 + v$ for any vector $v$. In particular, $0 + 0 = 0$. Also, it is easy to see that $v + (-v) = 0$, as we would expect. In general, since the scalar multiple $-v = -1v$ is a well-defined vector, we can define vector subtraction as follows: $v - w = v + (-w)$. 

55
The below figure shows the use of “geometric proofs” of various laws of vector algebra, that is, it uses laws from elementary geometry to prove statements about vectors. For example, (a) shows that $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ for any vectors $\mathbf{v}, \mathbf{w}$. And (c) shows how you can think of $\mathbf{v} - \mathbf{w}$ as the vector that is tacked on to the end of $\mathbf{w}$ to add up to $\mathbf{v}$.

Theorem 148. Let $\mathbf{v} = (v_1, v_2, v_3), \mathbf{w} = (w_1, w_2, w_3)$ be vectors in $\mathbb{R}^3$, let $k$ be a scalar. Then

(a) $k \mathbf{v} = (kv_1, kv_2, kv_3)$
(b) $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, v_3 + w_3)$

The following theorem summarizes the basic laws of vector algebra.

Theorem 149. For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, and scalars $k, l$, we have

(a) $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$  \quad \text{Commutative Law}
(b) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$  \quad \text{Associative Law}
(c) $\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}$  \quad \text{Additive Law}
(d) $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$  \quad \text{Additive Inverse}
(e) $k(l\mathbf{v}) = (kl)\mathbf{v}$  \quad \text{Associative Law}
(f) $k(\mathbf{v} + \mathbf{w}) = k\mathbf{v} + k\mathbf{w}$  \quad \text{Distributive Law}
(g) $(k + l)\mathbf{v} = k\mathbf{v} + l\mathbf{v}$  \quad \text{Distributive Law}

A unit vector is a vector with magnitude 1. Notice that for any nonzero vector $\mathbf{v}$, the vector $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is a unit vector which points in the same direction as $\mathbf{v}$, since $\frac{1}{\|\mathbf{v}\|} > 0$ and
\[ \frac{\|v\|}{\|v\|} = 1 \]. Dividing a nonzero vector \( v \) by \( \|v\| \) is often called normalizing \( v \).

**Analytic (Component) Form of Vectors**

There are specific unit vectors which we will often use, called the **basis vectors**:
- \( \mathbf{i} = (1, 0, 0) \), \( \mathbf{j} = (0, 1, 0) \), and \( \mathbf{k} = (0, 0, 1) \) in \( \mathbb{R}^3 \);
- \( \mathbf{i} = (1, 0) \) and \( \mathbf{j} = (0, 1) \) in \( \mathbb{R}^2 \).

These are useful for several reasons: they are mutually perpendicular, since they lie on distinct coordinate axes; they are all unit vectors: \( \|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1 \); every vector can be written as a unique scalar combination of the basis vectors:

- \( v = (a, b) = a\mathbf{i} + b\mathbf{j} \) in \( \mathbb{R}^2 \),
- \( v = (a, b, c) = a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \) in \( \mathbb{R}^3 \).

When a vector \( v = (a, b, c) \) is written as \( v = a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \), we say that \( v \) is in **component form** (analytic form), and that \( a, b, \) and \( c \) are the \( \mathbf{i}, \mathbf{j}, \) and \( \mathbf{k} \) components, respectively, of \( v \). We have:

\[
\begin{align*}
v &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}, \alpha \text{ is a scalar} \Rightarrow \alpha v &= \alpha v_1\mathbf{i} + \alpha v_2\mathbf{j} + \alpha v_3\mathbf{k} \\
v &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}, w = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k} \\
\Rightarrow v + w &= (v_1 + w_1)\mathbf{i} + (v_2 + w_2)\mathbf{j} + (v_3 + w_3)\mathbf{k} \\
v &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k} \Rightarrow \|v\| = \sqrt{v_1^2 + v_2^2 + v_3^2}
\end{align*}
\]

The distance \( d \) between two points \( P = (x_1, y_1, z_1) \) and \( Q = (x_2, y_2, z_2) \) in \( \mathbb{R}^3 \) is same as the length of the vector \( w - v \), where the vectors \( v \) and \( w \) are defined as \( v = (x_1, y_1, z_1) \) and \( w = (x_2, y_2, z_2) \). So since \( w - v = (x_2 - x_1, y_2 - y_1, z_2 - z_1) \), then

\[
d = \|w - v\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}
\]

**Dot Product**

**Definition 150.** Let \( v = (v_1, v_2, v_3) \) and \( w = (w_1, w_2, w_3) \) be vectors in \( \mathbb{R}^3 \). The **dot product** (inner product-scalar product) of \( v \) and \( w \), denoted by \( v \cdot w \), is given by:

\[
v \cdot w = v_1w_1 + v_2w_2 + v_3w_3
\]

Similarly, for vectors \( v = (v_1, v_2) \) and \( w = (w_1, w_2) \) in \( \mathbb{R}^2 \), the dot product is:

\[
v \cdot w = v_1w_1 + v_2w_2
\]
Notice that the dot product of two vectors is a scalar, not a vector. So the associative law that holds for multiplication of numbers and for addition of vectors does not hold for the dot product of vectors. Why? Because for vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, the dot product $\mathbf{u} \cdot \mathbf{v}$ is a scalar, and so $(\mathbf{u} \cdot \mathbf{v}) \cdot \mathbf{w}$ is not defined since the left side of that dot product (the part in parentheses) is a scalar and not a vector.

**Definition 151.** The angle between two nonzero vectors with the same initial point is the smallest angle between them.

We do not define the angle between the zero vector and any other vector. Any two nonzero vectors with the same initial point have two angles between them: $\theta$ and $360^\circ - \theta$. We will always choose the smallest nonnegative angle $\theta$ between them, so that $0^\circ \leq \theta \leq 180^\circ$.

**Theorem 152.** Let $\mathbf{v}$, $\mathbf{w}$ be nonzero vectors, and let $\theta$ be the angle between them. Then

$$\cos \theta = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}|| ||\mathbf{w}||}$$

For vectors $\mathbf{v} = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k}$ and $\mathbf{w} = w_1 \mathbf{i} + w_2 \mathbf{j} + w_3 \mathbf{k}$ in component form, the dot product is still $\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + v_3 w_3$, since $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors and perpendicular each other, that is,

$$||\mathbf{i}|| = ||\mathbf{j}|| = ||\mathbf{k}|| = 1, \quad i^2 = i \cdot i = 1, \quad j^2 = j \cdot j = 1, \quad k^2 = k \cdot k = 1$$

$$i \cdot j = 0, \quad i \cdot k = 0, \quad j \cdot k = 0$$

Two nonzero vectors are perpendicular if the angle between them is $90^\circ$. Since $\cos 90^\circ = 0$, we have the following important corollary:

**Corollary 153.** Two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ are perpendicular if and only if $\mathbf{v} \cdot \mathbf{w} = 0$.

We will write $\mathbf{v} \perp \mathbf{w}$ to indicate that $\mathbf{v}$ and $\mathbf{w}$ are perpendicular.

**Theorem 154.** For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, and scalar $k$, we have

\begin{align*}
(a) \quad & \mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v} & \text{Commutative Law} \\
(b) \quad & (k\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (k\mathbf{w}) = k(\mathbf{v} \cdot \mathbf{w}) & \text{Associative Law} \\
(c) \quad & (\mathbf{v}) \cdot 0 = 0 \cdot \mathbf{v} = 0 \\
(d) \quad & \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} & \text{Distributive Law} \\
(e) \quad & (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} & \text{Distributive Law} \\
(f) \quad & ||\mathbf{v} \cdot \mathbf{w}|| \leq ||\mathbf{v}|| ||\mathbf{w}|| & \text{Cauchy-Schwarz Inequality}
\end{align*}

Using the theorem above, we see that if $\mathbf{u} \cdot \mathbf{v} = 0$ and $\mathbf{u} \cdot \mathbf{w} = 0$, then $\mathbf{u} \cdot (k\mathbf{v} + l\mathbf{w}) = k(\mathbf{u} \cdot \mathbf{v}) + l(\mathbf{u} \cdot \mathbf{w}) = k(0) + l(0) = 0$ for all scalars $k, l$. Thus, we have the following fact:

**If $\mathbf{u} \perp \mathbf{v}$ and $\mathbf{u} \perp \mathbf{w}$, then $\mathbf{u} \perp (k\mathbf{v} + l\mathbf{w})$ for all scalars $k, l$.**
For vectors $v$ and $w$, the collection of all scalar combinations $k v + l w$ is called the span of $v$ and $w$. If nonzero vectors $v$ and $w$ are parallel, then their span is a line; if they are not parallel, then their span is a plane. So what we showed above is that a vector which is perpendicular to two other vectors is also perpendicular to their span.

**Theorem 155.** For any vectors $v, w$, we have

(a) $\|v\|^2 = v \cdot v$

(b) $\|v + w\| \leq \|v\| + \|w\|$  
**Triangle Inequality**

(c) $\|v - w\| \geq \|v\| - \|w\|

**Example 156.** Find the angle $\theta$ between the vectors $v = (2, 1, -1)$ and $w = (3, -4, 1)$.

**Solution:** Since $v \cdot w = (2)(3) + (1)(-4) + (-1)(1) = 1$, $\|v\| = \sqrt{6}$, and $\|w\| = \sqrt{26}$, then

$$\cos \theta = \frac{v \cdot w}{\|v\|\|w\|} = \frac{1}{\sqrt{6}\sqrt{26}} = \frac{1}{2\sqrt{39}} \approx 0.08 \Rightarrow \theta = 85.41^\circ$$

**Example 157.** Are the vectors $v = (-1, 5, -2)$ and $w = (3, 1, 1)$ perpendicular?

**Solution:** Yes, $v \perp w$ since $v \cdot w = (-1)(3) + (5)(1) + (-2)(1) = 0$

**Cross (Vector) Product**

In Section 1.3 we defined the dot product, which gave a way of multiplying two vectors. The resulting product, however, was a scalar, not a vector. In this section we will define a product of two vectors that does result in another vector. This product, called the cross product, is only defined for vectors in $\mathbb{R}^3$. The definition may appear strange and lacking motivation, but we will see the geometric basis for it shortly.

Suppose that $u = u_1 \mathbf{i} + u_2 \mathbf{j} + u_3 \mathbf{k}$ and $v = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k}$ and that we want to find a vector $w = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ orthogonal (perpendicular) to both $u$ and $v$. Thus we want $u \cdot w = 0$ and $v \cdot w = 0$, which leads to the linear system

$$u_1 x + u_2 y + u_3 z = 0$$
$$v_1 x + v_2 y + v_3 z = 0$$

It can be shown that

$$w = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}$$

is a solution to equation above (verify). Of course, we can also write $w$ as

$$w = (u_2 v_3 - u_3 v_2) \mathbf{i} + (u_3 v_1 - u_1 v_3) \mathbf{j} + (u_1 v_2 - u_2 v_1) \mathbf{k}$$

This vector is called the cross product of $u$ and $v$ and is denoted by $u \times v$. Note that the cross product, $u \times v$, is a vector, while the dot product, $u \cdot v$, is a scalar,
Definition 158. Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ be vectors in $\mathbb{R}^3$. The cross product of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} \times \mathbf{v}$, is the vector in $\mathbb{R}^3$ given by:

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

Example 159. Find $\mathbf{i} \times \mathbf{j}$.

Solution: Since $\mathbf{i} = (1, 0, 0)$ and $\mathbf{j} = (0, 1, 0)$, then

$$\mathbf{i} \times \mathbf{j} = ((0)(0) - (0)(1), (0)(0) - (1)(0), (1)(1) - (0)(0))$$
$$= (0, 0, 1)$$
$$= \mathbf{k}.$$

Similarly it can be shown that $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{i} = \mathbf{j}$.

Theorem 160. If the cross product $\mathbf{v} \times \mathbf{w}$ of two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ is also a nonzero vector, then it is perpendicular to both $\mathbf{v}$ and $\mathbf{w}$.

Proof. We will show that $(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{v} = 0$:

$$(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{v} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) \cdot (v_1, v_2, v_3)$$
$$= v_2w_3v_1 - v_3w_2v_1 + v_3w_1v_2 - v_1w_3v_2 + v_1w_2v_3 - v_2w_1v_3$$
$$= v_1v_2w_3 - v_1v_2w_3 + w_1v_2v_3 - w_1v_2v_3 + v_1w_2v_3 - v_1w_2v_3$$
$$= v_1v_2w_3 - v_1v_2w_3 + v_1w_2v_3 - v_1w_2v_3 + v_1w_2v_3 - v_1w_2v_3$$
$$= 0,$$ after rearranging the terms.

$\therefore \mathbf{v} \times \mathbf{w} \perp \mathbf{v}$ by Corollary 1.7.

We will now derive a formula for the magnitude of $\mathbf{v} \times \mathbf{w}$, for nonzero vectors $\mathbf{v}$, $\mathbf{w}$:

If $\theta$ is the angle between nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$, then

$$||\mathbf{v} \times \mathbf{w}|| = ||\mathbf{v}|| ||\mathbf{w}|| \sin \theta$$

Example 161. Let $\triangle PQR$ and $PQRS$ be a triangle and parallelogram, respectively, as shown in the below.

Think of the triangle as existing in $\mathbb{R}^3$, and identify the sides $QR$ and $QP$ with vectors $\mathbf{v}$ and $\mathbf{w}$, respectively, in $\mathbb{R}^3$. Let $\theta$ be the angle between $\mathbf{v}$ and $\mathbf{w}$. The area $A_{PQR}$ of
\( \Delta PQR \) is \( \frac{1}{2}bh \), where \( b \) is the base of the triangle and \( h \) is the height. So we see that

\[
b = \|v\| \quad \text{and} \quad h = \|w\| \sin \theta
\]

\[
A_{PQR} = \frac{1}{2} \|v\| \|w\| \sin \theta
\]

\[= \frac{1}{2} \|v \times w\|\]

So since the area \( A_{PQRS} \) of the parallelogram \( PQRS \) is twice the area of the triangle \( \Delta PQR \), then

\[
A_{PQRS} = \|v\| \|w\| \sin \theta
\]

**Theorem 162. Area of triangles and parallelograms**

(a) The area \( A \) of a triangle with adjacent sides \( v, w \) (as vectors in \( \mathbb{R}^3 \)) is:

\[
A = \frac{1}{2} \|v \times w\|
\]

(b) The area \( A \) of a parallelogram with adjacent sides \( v, w \) (as vectors in \( \mathbb{R}^3 \)) is:

\[
A = \|v \times w\|
\]

**Example 163.** Calculate the area of the triangle \( \Delta PQR \), where \( P = (2, 4, -7) \), \( Q = (3, 7, 18) \), and \( R = (-5, 12, 8) \).

**Solution:** Let \( v = \overrightarrow{PQ} \) and \( w = \overrightarrow{PR} \), as in Figure right. Then \( v = (3, 7, 18) - (2, 4, -7) = (1, 3, 25) \) and \( w = (-5, 12, 8) - (2, 4, -7) = (-7, 8, 15) \), so the area \( A \) of the triangle \( \Delta PQR \) is

\[
A = \frac{1}{2} \|v \times w\| = \frac{1}{2} \|(1, 3, 25) \times (-7, 8, 15)\|
\]

\[
= \frac{1}{2} \|((-3)(15) - (25)(8), (25)(-7) - (1)(15), (1)(8) - (3)(-7))\|
\]

\[
= \frac{1}{2} \|(-155, -190, 29)\|
\]

\[
= \frac{1}{2} \sqrt{(-155)^2 + (-190)^2 + 29^2} = \frac{1}{2} \sqrt{60966}
\]

\( A \approx 123.46 \)

**Example 164.** Calculate the area of the parallelogram \( PQRS \), where \( P = (1, 1) \), \( Q = (2, 3) \), \( R = (5, 4) \), and \( S = (4, 2) \).
Solution: Let \( \mathbf{v} = \overrightarrow{SP} \) and \( \mathbf{w} = \overrightarrow{SR} \), as in Figure right.
Then \( \mathbf{v} = (1, 1) - (4, 2) = (-3, -1) \) and \( \mathbf{w} = (5, 4) - (4, 2) = (1, 2) \). But these are vectors in \( \mathbb{R}^2 \), and the cross product is only defined for vectors in \( \mathbb{R}^3 \). However, \( \mathbb{R}^2 \) can be thought of as the subset of \( \mathbb{R}^3 \) such that the \( z \)-coordinate is always 0. So we can write \( \mathbf{v} = (-3, -1, 0) \) and \( \mathbf{w} = (1, 2, 0) \). Then the area \( A \) of \( PQRS \) is

\[
A = \| \mathbf{v} \times \mathbf{w} \| = \| (-3, -1, 0) \times (1, 2, 0) \| \\
= \| ((-1)(0) - (0)(2), (0)(1) - (-3)(0), (-3)(2) - (-1)(1)) \| \\
= \| (0, 0, -5) \| \\
A = 5
\]

**Theorem 165.** For any vectors \( \mathbf{u}, \mathbf{v}, \mathbf{w} \) in \( \mathbb{R}^3 \), and scalar \( k \), we have

(a) \( \mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v} \) \quad \text{Anticommutative Law}

(b) \( \mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w}) \) \quad \text{Distributive Law}

(c) \( (\mathbf{u} + \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{w}) \) \quad \text{Distributive Law}

(d) \( (k\mathbf{v}) \times \mathbf{w} = \mathbf{v} \times (k\mathbf{w}) = k(\mathbf{v} \times \mathbf{w}) \) \quad \text{Associative Law}

(e) \( \mathbf{v} \times \mathbf{0} = \mathbf{0} = \mathbf{0} \times \mathbf{v} \)

(f) \( \mathbf{v} \times \mathbf{v} = \mathbf{0} \)

(g) \( \mathbf{v} \times \mathbf{w} = \mathbf{0} \) \quad \text{if and only if} \( \mathbf{v} \parallel \mathbf{w} \)

**Example 166.** We have

\[
i \times i = j, \quad j \times j = k, \quad k \times k = 0 \\
i \times j = k, \quad j \times k = i, \quad k \times i = j
\]

Also,

\[
j \times i = -k, \quad k \times j = -i, \quad i \times k = -j
\]

**Volume of a parallelepiped: Scalar triple product**

Let the vectors \( \mathbf{u}, \mathbf{v}, \mathbf{w} \) in \( \mathbb{R}^3 \) represent adjacent sides of a parallelepiped \( P \) as in the figure. We will show that the volume of \( P \) is the scalar triple product \( \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \).
Recall that the volume \( \text{vol}(P) \) of a parallelepiped \( P \) is the area \( A \) of the base parallelogram times the height \( h \). We know that, the area \( A \) of the base parallelogram is \( \|v \times w\| \). And we can see that since \( v \times w \) is perpendicular to the base parallelogram determined by \( v \) and \( w \), then the height \( h \) is \( \|u\| \cos \theta \), where \( \theta \) is the angle between \( u \) and \( v \times w \).

We also know that \( \cos \theta = \frac{u \cdot (v \times w)}{\|u\| \|v \times w\|} \). Hence,

\[
\text{vol}(P) = Ah = \|v \times w\| \frac{u \cdot (v \times w)}{\|u\| \|v \times w\|} = u \cdot (v \times w)
\]

For any vectors \( u, v, w \) in \( \mathbb{R}^3 \),

\[
u \cdot (v \times w) = w \cdot (u \times v) = v \cdot (w \times u)
\]

Since \( v \times w = -w \times v \) for any vectors \( v, w \) in \( \mathbb{R}^3 \), then picking the wrong order for the three adjacent sides in the scalar triple product will give you the negative of the volume of the parallelepiped. So taking the absolute value of the scalar triple product for any order of the three adjacent sides will always give the volume.

Another type of triple product is the **vector triple product** \( u \times (v \times w) \).

**Theorem 167.** For any vectors \( u, v, w \) in \( \mathbb{R}^3 \),

\[
u \times (v \times w) = (u \cdot w)v - (u \cdot v)w
\]

**Example 168.** Find \( u \times (v \times w) \) for \( u = (1, 2, 4), v = (2, 2, 0), w = (1, 3, 0) \)

**Solution:** Since \( u \cdot v = 6 \) and \( u \cdot w = 7 \), then

\[
u \times (v \times w) = (u \cdot w)v - (u \cdot v)w
\]

\[
= 7(2, 2, 0) - 6(1, 3, 0)
\]

\[
= (14, 14, 0) - (6, 18, 0)
\]

\[
= (8, -4, 0)
\]

For vectors \( v = v_1 \mathbf{i} + v_2 \mathbf{j} + v_3 \mathbf{k} \) and \( w = w_1 \mathbf{i} + w_2 \mathbf{j} + w_3 \mathbf{k} \) in component form, the cross product is written as: \( v \times w = (v_2w_3 - v_3w_2) \mathbf{i} + (v_3w_1 - v_1w_3) \mathbf{j} + (v_1w_2 - v_2w_1) \mathbf{k} \). It is often to use the component form for the cross product, because it can be represented as a determinant.
Example 169. Let \( \mathbf{v} = 4\mathbf{i} - \mathbf{j} + 3\mathbf{k} \) and \( \mathbf{w} = \mathbf{i} + 2\mathbf{k} \). Then

\[
\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \begin{vmatrix} v_2 & v_3 & \mathbf{i} \\ v_1 & v_3 & \mathbf{j} \\ w_1 & w_3 & \mathbf{k} \end{vmatrix} - \begin{vmatrix} v_1 & v_3 & \mathbf{i} \\ v_2 & v_3 & \mathbf{j} \\ w_1 & w_2 & \mathbf{k} \end{vmatrix} = (v_2w_3 - v_3w_2)\mathbf{i} + (v_3w_1 - v_1w_3)\mathbf{j} + (v_1w_2 - v_2w_1)\mathbf{k}
\]

\[
= (4\cdot1 - 3\cdot2)\mathbf{i} + (3\cdot1 - 2\cdot1)\mathbf{j} + (1\cdot2 - 1\cdot1)\mathbf{k} = 2\mathbf{i} - 5\mathbf{j} + \mathbf{k}
\]

Theorem 170. For any vectors \( \mathbf{u} = (u_1, u_2, u_3), \mathbf{v} = (v_1, v_2, v_3), \mathbf{w} = (w_1, w_2, w_3) \) in \( \mathbb{R}^3 \):

\[
\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}
\]

Example 171. Find the volume of the parallelepiped with adjacent sides \( \mathbf{u} = (2, 1, 3), \mathbf{v} = (-1, 3, 2), \mathbf{w} = (1, 1, -2) \). By the previous theorem, the volume \( \text{vol}(P) \) of the parallelepiped \( P \) is the absolute value of the scalar triple product of the three adjacent sides (in any order).

\[
\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} 2 & 1 & 3 \\ -1 & 3 & 2 \\ 1 & 1 & -2 \end{vmatrix} = 2\begin{vmatrix} 3 & 2 \\ 1 & -2 \end{vmatrix} - 1\begin{vmatrix} -1 & 2 \\ 1 & -2 \end{vmatrix} + 3\begin{vmatrix} -1 & 3 \\ 1 & 1 \end{vmatrix} = 2(-8) - 1(0) + 3(-4) = -28,
\]

so

\[
\text{vol}(P) = |-28| = 28.
\]

6 Vector spaces

A useful procedure in mathematics and other disciplines involves classification schemes. That is, we form classes of objects based on properties they have in common. This allows us to treat all members of the class as a single unit. Thus, instead of dealing with each distinct member of the class, we can develop notions that apply to each and every member of the class based on the properties that they have in common. In many ways this helps us work with more abstract and comprehensive ideas.

Linear algebra has such a classification scheme that has become very important. This is the notion of a vector space. A vector space consists of a set of objects and two operations on these objects that satisfy a certain set of rules. If we have a vector
space, we will automatically be able to attribute to it certain properties that hold for all vector spaces. Thus, upon meeting some new vector space, we will not have to verify everything from scratch.

The name "vector space" conjures up the image of directed line segments from the plane, or 3-space, as discussed in the previous section. This is, of course, where the name of the classification scheme is derived from. We will see that matrices and \( n \)-vectors will give us examples of vector spaces, but other collections of objects, like polynomials, functions, and solutions to various types of equations, will also be vector spaces. For particular vector spaces, the members of the set of objects and the operations on these objects can vary, but the rules governing the properties satisfied by the operations involved will always be the same.

**Definition 172.** A *real vector space* is a set \( V \) of elements on which we have two operations \( \oplus \) and \( \odot \) defined with the following properties:

a) If \( u \) and \( v \) are any elements in \( V \), then \( u \oplus v \) is in \( V \). (We say that \( V \) is closed under the operation \( \oplus \).)

(1) \( u \oplus v = v \oplus u \) for all \( u, v \) in \( V \).

(2) \( u \oplus (v \oplus w) = (u \oplus v) \oplus w \) for all \( u, v, w \) in \( V \).

(3) There exists an element \( 0 \) in \( V \) such that \( u \oplus 0 = 0 \oplus u = u \) for any \( u \) in \( V \).

(4) For each \( u \) in \( V \) there exists an element \( -u \) in \( V \) such that \( u \oplus -u = -u \oplus u = 0 \).

b) If \( u \) is any element in \( V \) and \( c \) is any real number, then \( c \odot u \) is in \( V \) (i.e., \( V \) is closed under the operation \( \odot \)).

(5) \( c \odot (u \oplus v) = (c \odot u) \oplus (c \odot v) \) for any \( u, v \) in \( V \) and any real number \( c \)

(6) \( (c + d) \odot u = (c \odot u) \oplus (d \odot u) \) for any \( u \) in \( V \) and any real numbers \( c \) and \( d \).

(7) \( c \odot (d \odot u) = (cd) \odot u \) for any \( u \) in \( V \) and any real numbers \( c \) and \( d \).

(8) \( 1 \odot u = u \) for any \( u \) in \( V \).

* The elements of \( V \) are called vectors; the elements of the set of real numbers \( \mathbb{R} \) are called scalars.

* The operation \( \oplus \) is called vector addition; the operation \( \odot \) is called scalar multiplication.

* The vector \( 0 \) in property (3) is called a zero vector. The vector \( -u \) in property (4) is called a negative of \( u \). It can be shown that \( 0 \) and \( -u \) are unique.

In order to specify a vector space, we must be given a set \( V \) and two operations \( \oplus \) and \( \odot \) satisfying all the properties of the definition. We shall often refer to a real vector
space merely as a \textit{vector space}. Thus a "vector" is now an element of a vector space and no longer needs to be interpreted as a directed line segment. In our examples we shall see, however, how this name came about in a natural manner. We now consider some examples of vector spaces, leaving it to the reader to verify that all the properties of definition above.

\textbf{Example 173.} Consider $\mathbb{R}^n$, the set of all $n \times 1$ matrices

$$
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_n
\end{bmatrix}
$$

with real entries. Let the operation $\oplus$ be matrix addition and let the operation $\odot$ be multiplication of a matrix by a real number (scalar multiplication). By the use of the properties of matrices established in previous section, it is not difficult to show that $\mathbb{R}^n$ is a vector space by verifying that the properties of Definition 172. Thus the matrix

$$
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_n
\end{bmatrix}
$$

as an element of $\mathbb{R}^n$, is now called an $n$-vector or merely a vector.

\textbf{Example 174.} The set of all $m \times n$ matrices with matrix addition as $\oplus$ and multiplication of a matrix by a real number as $\odot$ is a vector space (verify). We denote this vector space by $M_{m \times n}$.

\textbf{Example 175.} The set of all real numbers with $\oplus$ as the usual addition of real numbers and $\odot$ as the usual multiplication of real numbers is a vector space (verify). In this case the real numbers play the dual roles of both vectors and scalars. This vector space is essentially the case with $n = 1$ of Example 173.

\textbf{Example 176.} Let $R_n$ be the set of all $1 \times n$ matrices $[a_1 \ a_2 \ \ldots \ a_n]$, where we define $\oplus$ by

$$
\begin{bmatrix}
a_1 & a_2 & \cdots & a_n
\end{bmatrix} \oplus \begin{bmatrix}
b_1 & b_2 & \cdots & b_n
\end{bmatrix} = \begin{bmatrix}
a_1 + b_1 & a_2 + b_2 & \cdots & a_n + b_n
\end{bmatrix}
$$

and we define $\odot$ by

$$
c \odot \begin{bmatrix}
a_1 & a_2 & \cdots & a_n
\end{bmatrix} = \begin{bmatrix}
ca_1 & ca_2 & \cdots & ca_n
\end{bmatrix}.
$$

Then $R_n$ is a vector space (verify). This is just a special case of Example 175.

\textbf{Example 177.} Let $V$ be the set of all $2 \times 2$ matrices with trace equal to zero; that is,

$$A = \begin{bmatrix}
a & b \\
c & d
\end{bmatrix} \text{ is in } V \text{ provided } \text{Tr}(A) = a + d = 0$$
for the definition and properties of the trace of a matrix. The operation \( \oplus \) is standard matrix addition and the operation \( \odot \) is standard scalar multiplication of matrices; then \( V \) is a vector space. We verify properties (a), (3), (4), (b), and (7) of Definition 172. The remaining properties are left for the student to verify.

Let
\[
A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} r & s \\ t & p \end{bmatrix}
\]
be any elements of \( V \). Then \( \text{Tr}(A) = a + d = 0 \) and \( \text{Tr}(B) = r + p = 0 \). For property (a), we have
\[
A \oplus B = \begin{bmatrix} a + r & b + s \\ c + t & d + p \end{bmatrix}
\]
and
\[
\text{Tr}(A \oplus B) = (a + r) + (d + p) = (a + d) + (r + p) = 0 + 0 = 0,
\]
so \( A \oplus B \) is in \( V \); that is \( V \) is closed under the operation \( \oplus \). To verify property (3), observe that the matrix has trace equal to zero, so it is in \( V \). Then it follows from the definition of \( \oplus \) that property (3) is valid in \( V \), so \( \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \) is the zero vector, which we denote as \( \mathbf{0} \). To verify property (4), let \( A \), as given previously, be an element of \( V \) and let
\[
C = \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}.
\]
We first show that \( C \) is in \( V \):
\[
\text{Tr}(C) = (-a) + (-d) = -(a + d) = 0.
\]
Then we have
\[
A \oplus C = \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}
\]
so \( C = -A \). For property (b), let \( k \) be any real number. We have
\[
k \odot A = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}
\]
and \( \text{Tr}(k \odot A) = ka + kd = k(a + d) = 0 \), so \( k \odot A \) is in \( V \); that is, \( V \) is closed under the operation \( \odot \). For property (7), let \( k \) and \( m \) be any real numbers. Then
\[
k \odot (m \odot A) = k \odot \begin{bmatrix} ma & mb \\ mc & md \end{bmatrix} = \begin{bmatrix} kma & kmb \\ kmc & kmd \end{bmatrix}
\]
and
\[
(km) \odot A = \begin{bmatrix} kma & kmb \\ kmc & kmd \end{bmatrix}
\]
It follows that \( k \odot (m \odot A) = (km) \odot A \).
Example 178. Another source of examples are sets of polynomials; therefore, we recall some well-known facts about such functions. A \textbf{polynomial} (in \( t \)) is a function that is expressible as

\[ p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0 \]

where \( a_0, a_1, \ldots, a_n \) are real numbers and \( n \) is a nonnegative integer. If \( a_n \neq 0 \), where \( p(t) \) is said to have \textbf{degree} \( n \). Thus the degree of a polynomial is the highest power of a term having a nonzero coefficient; \( p(t) = 2t + 1 \) has degree 1, and the constant polynomial \( p(t) = 3 \) has degree 0. The \textbf{zero polynomial}, denoted by \( 0 \), has no degree. We now let \( P_n \) be the set of all polynomials of degree \( \leq n \) together with the zero polynomial. If \( p(t) \) and \( q(t) \) are in \( P_n \), we can write

\[ p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0 \quad \text{and} \quad q(t) = b_n t^n + b_{n-1} t^{n-1} + \cdots + b_1 t + b_0. \]

We define \( p(t) \odot q(t) \) as

\[ p(t) \odot q(t) = (a_n + b_n) t^n + (a_{n-1} + b_{n-1}) t^{n-1} + \cdots + (a_1 + b_1) t + (a_0 + b_0). \]

If \( c \) is a scalar, we also define \( c \odot p(t) \) as

\[ c \odot p(t) = (ca_n) t^n + (ca_{n-1}) t^{n-1} + \cdots + (ca_1) t + (ca_0). \]

We now show that \( P_n \) is a vector space. Let \( p(t) \) and \( q(t) \), as before, be elements of \( P_n \); that is, they are polynomials of degree \( \leq n \) or the zero polynomial. Then the previous definitions of the operations \( \odot \) and \( \odot \) show that \( p(t) \odot q(t) \) and \( c \odot p(t) \), for any scalar \( c \), are polynomials of degree \( \leq n \) or the zero polynomial. That is, \( p(t) \odot q(t) \) and \( c \odot p(t) \) are in \( P_n \) so that (a) and (b) in Definition 172 hold. To verify property (1), we observe that

\[ q(t) \odot p(t) = (b_n + a_n) t^n + (b_{n-1} + a_{n-1}) t^{n-1} + \cdots + (b_1 + a_1) t + (a_0 + b_0), \]

and since \( a_i + b_i = b_i + a_i \) holds for the real numbers, we conclude that \( p(t) \odot q(t) = q(t) \odot p(t) \). Similarly, we verify property (2). The zero polynomial is the element \( 0 \) needed in property (3). If \( p(t) \) is as given previously, then its negative, \( -p(t) \), is

\[ -a_n t^n - a_{n-1} t^{n-1} - \cdots - a_1 t - a_0. \]

We shall now verify property (6) and will leave the verification of the remaining properties to the reader. Thus

\[ (c + d) \odot p(t) = (c + d) a_n t^n + (c + d) a_{n-1} t^{n-1} + \cdots + (c + d) a_1 t + (c + d) a_0 \]

\[ = ca_n t^n + da_n t^n + ca_{n-1} t^{n-1} + da_{n-1} t^{n-1} + \cdots + ca_1 t + da_1 t + ca_0 + da_0 \]

\[ = c \left( a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0 \right) + d \left( a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0 \right) \]

\[ = (c \odot p(t)) \odot (d \odot p(t)). \]
Example 179. Let $V$ be the set of all real-valued continuous functions defined on $R^1$. If $f$ and $g$ are in $V$, we define $f \oplus g$ by $(f \oplus g)(t) = f(t) + g(t)$. If $f$ in $V$ and $c$ is a scalar, we define $c \odot f$ by $(c \odot f)(t) = cf(t)$. Then $V$ is a vector space, which is denoted by $C(-\infty, \infty)$.

Example 180. Let $V$ be the set of all real numbers with the operations $u \oplus v = u - v$ (is ordinary subtraction) and $c \odot u = cu$ (ordinary multiplication). Is $V$ a vector space? If it is not, which properties in Definition 172 fail to hold?

**Solution:** If $u$ and $v$ are in $V$, and $c$ is a scalar, then $u \oplus v$ and $c \odot u$ are in $V$, so that (a) and (b) in Definition 172 hold. However, property (1) fails to hold, since

$$u \oplus v = u - v \quad \text{and} \quad v \oplus u = v - u$$

and these are not the same, in general. (Find $u$ and $v$ such that $u - v \neq v - u$). Also, we shall let the reader verify that properties (2), (3), and (4) fail to hold. Properties (5), (7), and (8) hold, but property (6) does not hold, because

$$(c + d) \odot u = (c + d)u = cu + du,$$

whereas

$$c \odot u \oplus d \odot u = cu \oplus du = cu - du$$

and these are not equal, in general. Thus $V$ is not a vector space.

Example 181. Let $V$ be the set of all integers; define $\oplus$ as ordinary addition and $\odot$ as ordinary multiplication. Here $V$ is not a vector, because if $u$ is any nonzero vector in $V$ and $c = \sqrt{3}$, then $c \odot u$ is not in $V$. Thus (b) fails to hold.

Remark 182. To verify that a given set $V$ with two operations $\oplus$ and $\odot$ is a real vector space, we must show that it satisfies all the properties of Definition 172.

- If both (a) and (b) hold, it is recommended that (3), the existence of a zero element, be verified next. Naturally, if (3) fails to hold, we do not have a vector space and do not have to check the remaining properties.

6.1 Subspaces

In this section we begin to analyze the structure of a vector space. First, it is convenient to have a name for a subset of a given vector space that is itself a vector space with respect to the same operations as those in $V$. Thus we have a definition.

**Definition 183.** Let $V$ be a vector space and $W$ a nonempty subset of $V$. If $W$ is a vector space with respect to the operations in $V$, then $W$ is called a **subspace** of $V$. 

69
Theorem 184. Let $V$ be a vector space with operations $\oplus$ and $\odot$ and let $W$ be a nonempty subset of $V$. Then $W$ is a subspace of $V$ if and only if the following conditions hold:

(a) If $u$ and $v$ are any vectors in $W$, then $u \oplus v$ is in $W$.
(b) If $c$ is any real number and $u$ is any vector in $W$, then $c \odot u$ is in $W$.

Example 185. Every vector space has at least two subspaces, itself and the subspace $\{0\}$ consisting only of the zero vector. (Recall that $0 \oplus 0 = 0$ and $c \odot 0 = 0$ in any vector space.) Thus $\{0\}$ is closed for both operations and hence is a subspace of $V$. The subspace $\{0\}$ is called the zero subspace of $V$.

Example 186. Let $P_2$ be the set consisting of all polynomials of degree $\leq 2$ and the zero polynomial; $P_2$ is a subset of $P$, the vector space of all polynomials. To verify that $P_2$ is a subspace of $P$, show it is closed for $\oplus$ and $\odot$. In general, the set $P_n$ consisting of all polynomials of degree $\leq n$ and the zero polynomial is a subspace of $P$. Also, $P_n$ is a subspace of $P_{n+1}$.

Example 187. Let $V$ be the set of all polynomials of degree exactly $= 2$; $V$ is a subset of $P$, the vector space of all polynomials; but $V$ is not a subspace of $P$, because the sum of the polynomials $2t^2 + 3t + 1$ and $-2t^2 + t + 2$ is not in $V$, since it is a polynomial of degree $1$.

Example 188. Let $W$ be the set of all vectors in $\mathbb{R}^3$ of the form $\begin{bmatrix} a \\ b \\ a + b \end{bmatrix}$ where $a$ and $b$ are any real numbers. To verify Theorem 184 a) and (b), we let

\[
\begin{align*}
  u &= \begin{bmatrix} a_1 \\ b_1 \\ a_1 + b_1 \end{bmatrix} \\
  v &= \begin{bmatrix} a_2 \\ b_2 \\ a_2 + b_2 \end{bmatrix}
\end{align*}
\]

be two vectors in $W$. Then

\[
\begin{align*}
  u \oplus v &= \begin{bmatrix} a_1 + a_2 \\ b_1 + b_2 \\ (a_1 + b_1) + (a_2 + b_2) \end{bmatrix} \\
  &= \begin{bmatrix} a_1 + a_2 \\ b_1 + b_2 \\ (a_1 + a_2) + (b_1 + b_2) \end{bmatrix}
\end{align*}
\]

is in $W$, for $W$ consists of all those vectors whose third entry is the sum of the first two entries. Similarly,

\[
\begin{align*}
  c \odot \begin{bmatrix} a_1 \\ b_1 \\ a_1 + b_1 \end{bmatrix} &= \begin{bmatrix} ca_1 \\ cb_1 \\ c(a_1 + b_1) \end{bmatrix} \\
  &= \begin{bmatrix} ca_1 \\ cb_1 \\ ca_1 + cb_1 \end{bmatrix}
\end{align*}
\]

is in $W$. Hence $W$ is a subspace of $\mathbb{R}^3$. 

70
Note 189. Henceforth, we shall usually denote $u \oplus v$ and $c \odot u$ in a vector space $V$ as $u + v$ and $cu$, respectively.

We can also show that a nonempty subset $W$ of a vector space $V$ is a subspace of $V$ if and only if $cu + dv$ is in $W$ for any vectors $u$ and $v$ in $W$ and any scalars $c$ and $d$.

Definition 190. Let $v_1, v_2, \ldots, v_k$ be vectors in a vector space $V$. A vector $v$ in $V$ is called a linear combination of $v_1, v_2, \ldots, v_k$ if

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = \sum_{j=1}^{k} a_j v_j$$

for some real numbers $a_1, a_2, \ldots, a_k$.

Example 191. In $\mathbb{R}^3$ let

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \text{and} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$ 

The vector

$$v = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$$

is a linear combination of $v_1, v_2,$ and $v_3$ if we can find real numbers $a_1, a_2,$ and $a_3$ so that

$$a_1 v_1 + a_2 v_2 + a_3 v_3 = v.$$ 

Substituting for $v, v_1, v_2,$ and $v_3,$ we have

$$a_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + a_3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}.$$ 

Equating corresponding entries leads to the linear system (verity)

$$a_1 + a_2 + a_3 = 2,$$

$$2a_1 + a_3 = 1,$$

$$a_1 + 2a_2 = 5.$$ 

Solving this linear system by the methods of previous Chapters gives (verify) $a_1 = 1, a_2 = 2,$ and $a_3 = -1$, which means that $v$ is a linear combination of $v_1, v_2,$ and $v_3$. Thus

$$v = v_1 + 2v_2 - v_3.$$
6.2 Span

Linear combinations play an important role in describing vector spaces. The set of all possible linear combinations of a pair of vectors in a vector space $V$ gives us a subspace. We have the following definition to help with such constructions:

**Definition 192.** If $S = \{v_1, v_2, \ldots, v_k\}$ is a set of vectors in a vector space $V$, then the set of all vectors in $V$ that are linear combinations of the vectors in $S$ is denoted by \[ \text{span } S = \langle S \rangle = \{a_1v_1 + a_2v_2 + \ldots + a_kv_k \mid a_1, a_2, \ldots, a_k \in \mathbb{R}\}. \]

**Example 193.** Consider the set $S$ of $2 \times 3$ matrices given by

\[
S = \begin{Bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\end{Bmatrix}.
\]

Then $\text{span } S$ is the set in $M_{23}$ consisting of all vectors of the form

\[
a\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ 0 & c & d \end{bmatrix}, \text{ where } a, b, c, d \in \mathbb{R}.
\]

That is, $\text{span } S$ is the subset of $M_{23}$ consisting of all matrices of the form

\[
\begin{bmatrix} a & b & 0 \\ 0 & c & d \end{bmatrix}
\]

where $a, b, c,$ and $d$ are real numbers.

**Theorem 194.** Let $S = \{v_1, v_2, \ldots, v_k\}$ be a set of vectors in a vector space $V$. Then $\text{span } S$ is a subspace of $V$.

**Proof.** Let

\[
u = \sum_{j=1}^{k} a_jv_j \quad \text{and} \quad w = \sum_{j=1}^{k} b_jv_j
\]

for some real numbers $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_k$. We have

\[
u + w = \sum_{j=1}^{k} a_jv_j + \sum_{j=1}^{k} b_jv_j = \sum_{j=1}^{k} (a_j + b_j)v_j.
\]

Moreover, for any real number $c$,

\[
cu = c\left(\sum_{j=1}^{k} a_jv_j\right) = \sum_{j=1}^{k} (ca_j)v_j.
\]

Since the sum $\nu + w$ and the scalar multiple $cu$ are linear combinations of the vectors in $S$, then $\text{span } S$ is a subspace of $V$. \qed
Definition 195. Let $S$ be a set of vectors in a vector space $V$. If every vector in $V$ is a linear combination of the vectors in $S$, then the set $S$ is said to span $V$, or $V$ is spanned (generated) by the set $S$; that is, $\text{span } S = V$.

Remark 196. If $\text{span } S = V$, $S$ is called a spanning set of $V$. A vector space can have many spanning sets.

Example 197. In $\mathbb{R}^3$, let

$$v_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}.$$ 

Determine whether the vector

$$v = \begin{bmatrix} 1 \\ 5 \\ -7 \end{bmatrix}$$

belongs to $\text{span}\{v_1, v_2\}$.

Solution: If we can find scalars $a_1, a_2$ such that $v = a_1 v_1 + a_2 v_2$, then $v$ belongs to $\text{span}\{v_1, v_2\}$. Substituting for $v_1, v_2$ and $v$, we have

$$a_1 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -7 \end{bmatrix}.$$ 

This expression corresponds to the linear system whose augmented matrix is (verify)

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

The reduced row echelon form of this system is (verify)

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

which indicates that the linear system is consistent, $a_1 = 2$, and $a_2 = 3$. Hence $v$ belongs to $\text{span}\{v_1, v_2\}$.

Example 198. In $P_2$, let

$$v_1 = 2t^2 + t + 2, \quad v_2 = t^2 - 2t, \quad v_3 = 5t^2 - 5t + 2, \quad v_4 = -t^2 - 3t - 2.$$ 

Determine whether the vector

$$v = t^2 + t + 2$$

belongs to $\text{span}\{v_1, v_2, v_3, v_4\}$. 

73
belongs to span \{v_1, v_2, v_3, v_4\}.

**Solution:** If we can find scalars \(a_1, a_2, a_3, \text{ and } a_4\) so that

\[
a_1v_1 + a_2v_2 + a_3v_3 + a_4v_4 = v
\]

then \(v\) belongs to span \{v_1, v_2, v_3, v_4\}. Substituting for \(v_1, v_2, v_3, \text{ and } v_4\), we have

\[
a_1 (2t^2 + t + 2) + a_2 (t^2 - 2t) + a_3 (5t^2 - 5t + 2) + a_4 (-t^2 - 3t - 2) = t^2 + t + 2,
\]
or

\[
(2a_1 + a_2 + 5a_3 - a_4) t^2 + (a_1 - 2a_2 - 5a_3 - 3a_4) t + (2a_1 + 2a_3 - 2a_4) = t^2 + t + 2.
\]

Now two polynomials agree for all values of \(t\) only if the coefficients of respective powers of \(t\) agree. Thus we get the linear system

\[
\begin{align*}
2a_1 + a_2 + 5a_3 - a_4 &= 1 \\
a_1 - 2a_2 - 5a_3 - 3a_4 &= 1 \\
2a_1 + 2a_3 - 2a_4 &= 2.
\end{align*}
\]

To determine whether this system of linear equations is consistent, we form the augmented matrix and transform it to reduced row echelon form, obtaining (verify)

\[
\begin{bmatrix}
1 & 0 & 1 & -1 & 0 \\
0 & 1 & 3 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix},
\]

which indicates that the system is inconsistent; that is, it has no solution. Hence \(v\) does not belong to span\{v_1, v_2, v_3, v_4\}.

**Remark 199.** In general, to determine whether a specific vector \(v\) belongs to span \(S\), we investigate the consistency of an appropriate linear system.

**Example 200.** Let \(V\) be the vector space \(\mathbb{R}^3\). Let

\[
v_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \text{and} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.
\]

To find out whether \(v_1, v_2, v_3\) span \(V\), we pick any vector \(v = \begin{bmatrix} a \\ b \\ c \end{bmatrix}\) in \(V\) (\(a, b\) and \(c\) are arbitrary real numbers) and determine whether there are constants \(a_1, a_2, \text{ and } a_3\) such that

\[
a_1v_1 + a_2v_2 + a_3v_3 = v.
\]
This leads to the linear system (verify)

\[ a_1 + a_2 + a_3 = a \]
\[ 2a_1 + a_3 = b \]
\[ a_1 + 2a_2 = c. \]

A solution is (verify)

\[ a_1 = \frac{-2a + 2b + c}{3}, \quad a_2 = \frac{a - b + c}{3}, \quad a_3 = \frac{4a - b - 2c}{3}. \]

Thus \( v_1, v_2, v_3 \) span \( V \). This is equivalent to saying that \( \text{span} \{v_1, v_2, v_3\} = \mathbb{R}^3 \).

**Example 201.** Let \( V \) be the vector space \( P_2 \). Let \( v_1 = t^2 + 2t + 1 \) and \( v_2 = t^2 + 2 \). Does \( \{v_1, v_2\} \) span \( V \)?

**Solution:** Let \( v = at^2 + bt + c \) be any vector in \( V \), where \( a, b, \) and \( c \) are any real numbers. We must find out whether there are constants \( a_1 \) and \( a_2 \) such that

\[ a_1v_1 + a_2v_2 = v, \]

or

\[ a_1\left(t^2 + 2t + 1\right) + a_2\left(t^2 + 2\right) = at^2 + bt + c. \]

Thus

\[(a_1 + a_2)t^2 + (2a_1)t + (a_1 + 2a_2) = at^2 + bt + c.\]

Equating the coefficients of respective powers of \( t \), we get the linear system

\[ a_1 + a_2 = a \]
\[ 2a_1 = b \]
\[ a_1 + 2a_2 = c. \]

Transforming the augmented matrix of this linear system, we obtain (verify)

\[
\begin{bmatrix}
1 & 0 & 2a - c \\
0 & 1 & c - a \\
0 & 0 & b - 4a + 2c
\end{bmatrix}
.
\]

If \( b - 4a + 2c \neq 0 \), then the system is inconsistent and there is no solution. Hence \( \{v_1, v_2\} \) does not span \( V \).

### 6.3 Linear Independence

**Definition 202.** The vectors \( v_1, v_2, \ldots, v_k \) in a vector space \( V \) are said to be **linearly dependent** if there exist constants \( a_1, a_2, \ldots, a_k \), not all zero, such that

\[
\sum_{j=1}^{k} a_jv_j = a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0. \quad (15)
\]
Otherwise, \( v_1, v_2, \ldots, v_k \) are called **linearly independent**. That is,

\[
v_1, v_k, \ldots, v_k \text{ are linearly independent if, whenever } a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0, a_1 = a_2 = \cdots = a_k = 0.
\]

If \( S = \{v_1, v_2, \ldots, v_k\} \), then we also say that the set \( S \) is linearly dependent or linearly independent if the vectors have the corresponding property.

It should be emphasized that for any vectors \( v_1, v_2, \ldots, v_k \), Equation (15) always holds if we choose all the scalars \( a_1, a_2, \cdots, a_k \) equal to zero. The important point in this definition is whether it is possible to satisfy (15) with at least one or the scalars different from zero.

**Example 203.** Determine whether the vectors

\[
v_1 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}
\]

are linearly independent.

**Solution:** Forming Equation (15),

\[
a_1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + a_3 \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},
\]

we obtain the homogeneous system (verify)

\[
3a_1 + a_2 - a_3 = 0 \\
2a_1 + 2a_2 + 2a_3 = 0 \\
a_1 - a_3 = 0
\]

The corresponding augmented matrix is

\[
\begin{bmatrix} 3 & 1 & -1 & | & 0 \\ 2 & 2 & 2 & | & 0 \\ 1 & 0 & -1 & | & 0 \end{bmatrix},
\]

whose reduced row echelon form is (verify)

\[
\begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.
\]
Thus there is a nontrivial solution
\[
\begin{bmatrix}
k \\
-2k \\
k
\end{bmatrix}, \quad k \neq 0 \text{ (verify)},
\]
so the vectors are linearly dependent.

**Example 204.** Are the vectors \( v_1 = \begin{bmatrix} 1 & 0 & 1 & 2 \end{bmatrix}, v_2 = \begin{bmatrix} 0 & 1 & 1 & 2 \end{bmatrix}, \) and \( v_3 = \begin{bmatrix} 1 & 1 & 1 & 3 \end{bmatrix} \) in \( \mathbb{R}^4 \) linearly dependent or linearly independent?

**Solution:** We form Equation (15),
\[
a_1v_1 + a_2v_2 + a_3v_3 = 0
\]
and solve for \( a_1, a_2, \) and \( a_3. \) The resulting homogeneous system is (verify) and solve for \( a_1, a_2, \) and \( a_3. \) The resulting homogeneous system is (verify)
\[
\begin{align*}
a_1 + a_3 &= 0 \\
a_2 + a_3 &= 0 \\
a_1 + a_2 + a_3 &= 0 \\
2a_1 + 2a_2 + 3a_3 &= 0.
\end{align*}
\]
The corresponding augmented matrix is (verify)
\[
\begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 \\
2 & 2 & 3 & 0
\end{bmatrix}
\]
and its reduced row echelon form is (verify)
\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]
Thus the only solution is the trivial solution \( a_1 = a_2 = a_3 = 0, \) so the vectors are linearly independent.

**Example 205.** Are the vectors
\[
v_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 & -3 \\ -2 & 1 \end{bmatrix}
\]
in \( M_{22} \) linearly independent?
**Solution:** We form Equation (15),

\[
\begin{bmatrix}
2 & 1 \\
0 & 1
\end{bmatrix} +
\begin{bmatrix}
1 & 2 \\
1 & 0
\end{bmatrix} +
\begin{bmatrix}
0 & -3 \\
-2 & 1
\end{bmatrix} =
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix},
\]

and solve for \(a_1, a_2, \) and \(a_3\). Performing the scalar multiplications and adding the resulting matrices gives

\[
\begin{bmatrix}
2a_1 + a_2 & a_1 + 2a_2 - 3a_3 \\
a_2 - 2a_3 & a_1 + a_3
\end{bmatrix} =
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}.
\]

Using the definition for equal matrices, we have the linear system

\[
\begin{align*}
2a_1 + a_2 &= 0 \\
a_1 + 2a_2 - 3a_3 &= 0 \\
a_2 - 2a_3 &= 0 \\
a_1 + a_3 &= 0
\end{align*}
\]

The corresponding augmented matrix is

\[
\begin{bmatrix}
2 & 1 & 0 & 0 \\
1 & 2 & -3 & 0 \\
0 & 1 & -2 & 0 \\
1 & 0 & 1 & 0
\end{bmatrix},
\]

and its reduced row echelon form is \(\text{(verify)}\)

\[
\begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & -2 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

Thus there is a nontrivial solution

\[
\begin{bmatrix}
-k \\
2k \\
k
\end{bmatrix}, \quad k \neq 0 \text{ (verify)}
\]

so the vectors are linearly dependent.

**Example 206.** Are the vectors \(v_1 = t^2 + t + 2, v_2 = 2t^2 + t, \) and \(v_3 = 3t^2 + 2t + 2\) in \(P_2\) linearly dependent or linearly independent?

**Solution:** Forming Equation (15), we have \(\text{(verify)}\)

\[
\begin{align*}
a_1 + 2a_2 + 3a_3 &= 0 \\
a_1 + a_2 + 2a_3 &= 0 \\
2a_1 + 2a_3 &= 0,
\end{align*}
\]

which has infinitely many solutions \(\text{(verify)}\). A particular solution is \(a_1 = 1, a_2 = 1, a_3 = -1, \) so, \(v_1 + v_2 - v_3 = 0\). Hence the given vectors are linearly dependent.
Theorem 207. Let $S = \{v_1, v_2, \ldots, v_n\}$ be a set of $n$ vectors in $\mathbb{R}^n (R_n)$. Let $A$ be the matrix whose columns (rows) are the elements of $S$. Then $S$ is linearly independent if and only if $\det(A) \neq 0$.

Example 208. Is $S = \{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 3 & 0 & -1 \end{bmatrix}\}$ a linearly independent set of vectors in $\mathbb{R}^3$?

Solution: We form the matrix $A$ whose rows are the vectors in $S$:

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

Since $\det(A) = 2$ (verify), we conclude that $S$ is linearly independent.

Theorem 209. Let $S_1$ and $S_2$ be finite subsets of a vector space and let $S_1$ be a subset of $S_2$. Then the following statements are true:

a) If $S_1$ is linearly dependent, so is $S_2$.

b) If $S_2$ is linearly independent, so is $S_1$.

Summary 210. At this point, we have established the following results:

- The set $S = \{0\}$ consisting only of 0 is linearly dependent, since, for example, $50 = 0$, and $5 \neq 0$.

- From this it follows that if $S$ is any set of vectors that contains 0, then $S$ must be linearly dependent.

- A set of vectors consisting of a single nonzero vector is linearly independent (verify).

- If $v_1, v_2, \ldots, v_k$ are vectors in a vector space $V$ and any two of them are equal, then $v_1, v_2, \ldots, v_k$ are linearly dependent (verify).

Theorem 211. The nonzero vectors $v_1, v_2, \ldots, v_n$ in a vector space $V$ are linearly dependent if and only if one of the vectors $v_j (j \geq 2)$ is a linear combination of the preceding vectors $v_1, v_2, \ldots, v_{j-1}$.

Example 212. Let $V = R_3$ and also $v_1 = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}, v_3 = \begin{bmatrix} -3 & 2 & -1 \end{bmatrix}$ and $v_4 = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$. We find (verify) that

$$v_1 + v_2 + 0v_3 - v_4 = 0$$

so $v_1, v_2, v_3,$ and $v_4$ are linearly dependent. We then have

$$v_4 = v_1 + v_2 + 0v_3.$$
Remark 213. 1) We observe that Theorem 211 does not say that every vector \( \mathbf{v} \) is a linear combination of the preceding vectors. Thus, in Example 212, we also have \( \mathbf{v}_1 + 2\mathbf{v}_2 + \mathbf{v}_3 + 0\mathbf{v}_4 = 0 \). We cannot solve, in this equation, for \( \mathbf{v}_4 \) as a linear combination of \( \mathbf{v}_1, \mathbf{v}_2, \) and \( \mathbf{v}_3 \), since its coefficient is zero.

2) We can also prove that if \( S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \) is a set of vectors in a vector space \( V \), then \( S \) is linearly dependent if and only if one of the vectors in \( S \) is a linear combination of all the other vectors in \( S \). For instance, in Example 212
\[
\mathbf{v}_1 = -\mathbf{v}_2 - 0\mathbf{v}_3 + \mathbf{v}_4; \quad \mathbf{v}_2 = -\frac{1}{2}\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_3 - 0\mathbf{v}_4.
\]

3) Observe that if \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \) are linearly independent vectors in a vector space, then they must be distinct and nonzero.

Suppose that \( S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \) spans a vector space \( V \), and \( \mathbf{v}_j \) is a linear combination of the preceding vectors in \( S \). Then the set
\[
S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n\}
\]
consisting of \( S \) with \( \mathbf{v}_j \) deleted, also spans \( V \). To show this result, observe that if \( \mathbf{v} \) is any vector in \( V \), then, since \( S \) spans \( V \), we can find scalars \( a_1, a_2, \ldots, a_n \) such that
\[
\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_j\mathbf{v}_{j-1} + a_j\mathbf{v}_j + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_n\mathbf{v}_n.
\]
Now if
\[
\mathbf{v}_j = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_{j-1}\mathbf{v}_{j-1},
\]
then
\[
\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_{j-1}\mathbf{v}_{j-1} + a_j(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_{j-1}\mathbf{v}_{j-1})
+ a_{j+1}\mathbf{v}_{j+1} + \cdots + a_n\mathbf{v}_n
= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{j-1}\mathbf{v}_{j-1} + c_j\mathbf{v}_j + c_{j+1}\mathbf{v}_{j+1} + \cdots + c_n\mathbf{v}_n
\]
which means that span \( S_1 = V \).

Example 214. Consider the set of vectors \( S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \) in \( \mathbb{R}^4 \), where
\[
\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}
\]
and let \( W = \text{span } S \). Since
\[
\mathbf{v}_4 = \mathbf{v}_1 + \mathbf{v}_2,
\]
we conclude that \( W = \text{span } S_1 \), where \( S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \).
7 Basis and Dimension

In this section we continue our study of the structure of a vector space $V$ by determining a set of vectors in $V$ that completely describes $V$.

7.1 Basis

Definition 215. The vectors $v_1, v_2, \ldots, v_k$ in a vector space $V$ are said to form a basis for $V$ if

a) $v_1, v_2, \ldots, v_k$ span $V$ and

b) $v_1, v_2, \ldots, v_k$ are linearly independent.

Remark 216. If $v_1, v_2, \ldots, v_k$ forms a basis for a vector space $V$, then they must be distinct and nonzero.

Example 217. Let $V = \mathbb{R}^3$. The vectors 
\[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}
\]
form a basis for $\mathbb{R}^3$, called the natural basis or standard basis, for $\mathbb{R}^3$. We can readily see how to generalize this to obtain the natural basis for $\mathbb{R}^n$. Similarly, 
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
is the natural basis for $\mathbb{R}^3$.

The natural basis for $\mathbb{R}^n$ is denoted by $\{e_1, e_2, \ldots, e_n\}$, where

\[
\begin{bmatrix}
0 \\
\vdots \\
0 \\
1 \\
0 \\
\vdots \\
0
\end{bmatrix} \quad \text{← $i$th row;}
\]

that is, $e_i$ is an $n \times 1$ matrix with a 1 in the $i$th row and zeros elsewhere. The natural basis for $\mathbb{R}^3$ is also often denoted by

\[
i = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad j = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad k = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.
\]

Example 218. Show that the set $S = \{t^2 + 1, t - 1, 2t + 2\}$ is a basis for the vector space $P_2$.

Solution: To do this, we must show that $S$ spans $V$ and is linearly independent. To show that it spans $V$, we take any vector in $V$, that is, a polynomial $at^2 + bt + c$, and
find constants $a_1, a_2$ and $a_3$ such that

$$at^2 + bt + c = a_1 (t^2 + 1) + a_2 (t - 1) + a_3 (2t + 2)$$

$$= a_1 t^2 + (a_2 + 2a_3) t + (a_1 - a_2 + 2a_3).$$

Since two polynomials agree for all values of $t$ only if the coefficients of respective powers of $t$ agree, we get the linear system

$$a_1 = a$$

$$a_2 + 2a_3 = b$$

$$a_1 - a_2 + 2a_3 = c.$$ 

Solving, we have

$$a_1 = a, \quad a_2 = \frac{a + b - c}{2}, \quad a_3 = \frac{c + b - a}{2}.$$ 

Hence $S$ spans $V$.

To illustrate this result, suppose that we are given the vector $2t^2 + 6t + 13$. Here, $a = 2, b = 6, \text{ and } c = 13$. Substituting in the foregoing expressions for $a, b, \text{ and } c$, we find that

$$a_1 = 2, \quad a_2 = -\frac{5}{2}, \quad a_3 = \frac{17}{4}.$$ 

Hence

$$2t^2 + 6t + 13 = 2(t^2 + 1) - \frac{5}{2}(t - 1) + \frac{17}{4}(2t + 2).$$ 

To show that $S$ is linearly independent, we form

$$a_1 (t^2 + 1) + a_2 (t - 1) + a_3 (2t + 2) = 0.$$ 

Then

$$a_1 t^2 + (a_2 + 2a_3) t + (a_1 - a_2 + 2a_3) = 0.$$ 

Again, this can hold for all values of $t$ only if

$$a_1 = 0$$

$$a_2 + 2a_3 = 0$$

$$a_1 - a_2 + 2a_3 = 0.$$ 

The only solution to this homogeneous system is $a_1 = a_2 = a_3 = 0$, which implies that $S$ is linearly independent. Thus $S$ is a basis for $P_2$. 

82
Example: Show that the set \( S = \{ v_1, v_2, v_3, v_4 \} \) where
\[ v_1 = [1, 0, 1, 0], \quad v_2 = [0, 1, 0, 1], \quad v_3 = [0, 2, 2, 1] \]
and \( v_4 = [1, 0, 0, 1] \) is a basis for \( R_4 \).

Solution: To show that \( S \) is linearly independent, we form the equation
\[ a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4 = 0 \]
and solve for \( a_1, a_2, a_3 \) and \( a_4 \). Substituting for \( v_1, v_2, v_3 \) and \( v_4 \) we obtain
the linear system
\[
\begin{align*}
0a_1 + a_2 &= 0 \\
0a_1 + 2a_2 &= 0 \\
0a_1 + 0a_2 + a_3 &= 0 \\
0a_1 + 0a_2 + a_3 &= 0
\end{align*}
\]
which has its only solution \( a_1 = a_2 = a_3 = a_4 = 0 \), showing that \( S \) is linearly independent.

Observe that the columns of the coefficient matrix of the preceding linear system are \( v_1^T, v_2^T, v_3^T \) and \( v_4^T \). To show that \( S \) spans \( R_4 \), we let
\[ v = [a, b, c, d] \]
be any vector in \( R_4 \). We now seek constants \( a_1, a_2, a_3 \) and \( a_4 \) such that
\[ a_1 v_1 + a_2 v_2 + a_3 v_3 + a_4 v_4 = v. \]

Substituting for \( v_1, v_2, v_3, v_4 \) and \( v \) we find a solution for \( a_1, a_2, a_3 \) and \( a_4 \) to
the resulting linear system. Hence \( S \) spans \( R_4 \) and is a basis for \( R_4 \).

Example: The set \( W \) of all \( 2 \times 2 \) matrices with trace equal to zero is a subspace of \( M_{22} \). Show that the set \( S = \{ v_1, v_2, v_3 \} \), where
\[ v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \]
and \( v_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \), is a basis for \( W \).

Solution: We must show \( \text{span} S = W \) and \( S \) is linearly independent. To show
that \( \text{span} S = W \), we take any vector \( v \) in \( W \), that is \( v = \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \)
and find constants \( a_1, a_2 \) and \( a_3 \) such that
\[ a_1 v_1 + a_2 v_2 + a_3 v_3 = v. \]
Substituting for \( v_1, v_2 \) and \( v_3 \),
\[ \begin{bmatrix} 0 & a_1 \\ a_2 & -a_3 \end{bmatrix} = \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \]
and hence
\[ \begin{bmatrix} a_2 & a_1 \\ -a_3 & a_1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}. \]

Equating corresponding entries gives \( a_3 = 0, a_1 = b \) and \( a_2 = c \), so \( \text{span} S = W \).

If we replace the vector \( v \) by the zero matrix, it follows in a similar
fashion that \( a_1 = a_2 = a_3 = 0 \), so \( S \) is linearly independent set. Hence \( S \) is a
basis for \( W \).
Example: Find a basis for the subspace \( V \) of \( P_2 \), consisting of all vectors of the form \( at^2 + bt + c \), where \( c = a - b \).

Solution: Every vector \( V \) is of the form \( at^2 + bt + c \), which can be written as \( a(t^2 + 1) + b(t + 1) \), so the vectors \( t^2 + 1 \) and \( t + 1 \) span \( V \). Moreover, these vectors are linearly independent because neither one is a multiple of the other.

A vector space \( V \) is called finite dimensional if there is a finite subset of \( V \) that is a basis for \( V \).

Theorem: If \( S = \{v_1, v_2, \ldots, v_n\} \) is a basis for a vector space \( V \), then every vector in \( V \) can be written in one and only one way as a linear combination of the vectors in \( S \).

Theorem: If \( S = \{v_1, v_2, \ldots, v_n\} \) is a basis for a vector space \( V \) and \( T = \{w_1, w_2, \ldots, w_r\} \) is a linearly independent set of vectors in \( V \), then \( r \leq n \).

Corollary: If \( S = \{v_1, v_2, \ldots, v_n\} \) and \( T = \{w_1, w_2, \ldots, w_m\} \) are bases for a vector space \( V \), then \( n = m \).

Dimension: The dimension of a non-zero vector space \( V \) is the number of vectors in a basis for \( V \). We often write \( \dim V \) for the dimension of \( V \). The dimension of the trivial vector space \( \{0\} \) is zero.

Example: The set \( S = \{t^2, t, 1\} \) is a basis for \( P_2 \), so \( \dim P_2 = 3 \).

* If vector space \( V \) has dimension \( n \), then any subset \( m \geq n \) vectors must be linearly dependent.

* If vector space \( V \) has dimension \( n \), then any subset of \( m < n \) vectors cannot span \( V \).

Theorem: Let \( V \) be an \( n \)-dimensional vector space.

a) If \( S = \{v_1, v_2, \ldots, v_n\} \) is a linearly independent set of vectors in \( V \), then \( S \) is a basis for \( V \).

b) If \( S = \{v_1, v_2, \ldots, v_n\} \) spans \( V \), then \( S \) is a basis for \( V \).
Example: Show that the set \( T = \{ v_1, v_2, v_3 \} \), where \( v_1 = [1 \ 1 \ 1] \), \( v_2 = [2 \ 3 \ 1] \), and \( v_3 = [1 \ -1 \ 0] \), is a basis for \( \mathbb{R}^3 \).

Solution: \( \mathbb{R}^3 \) is a 3-dimensional vector space. Since \( T \) has 3 vectors, we only need to show that \( T \) is linearly independent. Forming the equation

\[
\begin{align*}
0_1 v_1 + 0_2 v_2 + 0_3 v_3 &= 0, \\
-0_1 + 0_3 &= 0, \\
0_1 + 20_2 - 0_3 &= 0, \\
0_1 + 30_2 &= 0
\end{align*}
\]

Solving this homogeneous system, we obtain \( 0_1 = 0_2 = 0_3 = 0 \). Therefore the set \( T \) is linearly independent and so, \( \Rightarrow \) a basis for \( \mathbb{R}^3 \).

Coordinates and Transition Matrices:

Definition: A set of vectors \( S = \{ v_1, v_2, \ldots, v_n \} \) in a vector space \( V \) is called an ordered basis for \( V \) provided \( S \) is a basis for \( V \) and if we reorder the vectors in \( S \), this new ordering of the vectors in \( S \) is considered a different basis for \( V \). Therefore, if \( S = \{ v_1, v_2, \ldots, v_n \} \) is a ordered basis for \( V \), \( S_1 = \{ v_2, v_1, \ldots, v_n \} \) is a different ordered basis for \( V \).

* If \( S = \{ v_1, v_2, \ldots, v_n \} \) is an ordered basis for the \( n \)-dimensional vector space \( V \) then every vector \( v \) in \( V \) can be uniquely expressed in the form

\[
\begin{align*}
v &= a_1 v_1 + a_2 v_2 + \cdots + a_n v_n,
\end{align*}
\]

where \( a_1, a_2, \ldots, a_n \) are real numbers. We shall refer to

\[
[v]_S = \begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_n
\end{bmatrix}
\]

as the coordinate vector of \( v \) with respect to the ordered basis \( S \). The entries of \( [v]_S \) are called the coordinates of \( v \) with respect to \( S \).

Example: Consider the vector space \( \mathbb{R}^3 \) and let \( S = \{ v_1, v_2, v_3 \} \) be an ordered basis for \( \mathbb{R}^3 \), where \( v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \), \( v_2 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \) and \( v_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \). If \( v = \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix} \), compute \( [v]_S \).

To find \( [v]_S \), we need to find the constants \( a_1, a_2, \text{ and } a_3 \) such that

\[
0_1 v_1 + 0_2 v_2 + 0_3 v_3 = v,
\]

which leads to the linear system whose augmented matrix is

\[
\begin{bmatrix}
1 & 2 & -1 & | & 1 \\
0 & 1 & 1 & | & 2 \\
-1 & 0 & 2 & | & -5
\end{bmatrix}
\]

or equivalently \( [v_1 \ v_2 \ v_3 \ v] \).
Transforming the matrix to reduced row echelon form, we obtain the solution $a_1=3, a_2=-1$ and $a_3=-2$. So, $[V]_S = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}$.

**Transition Matrices:**

Let $S = \{v_1, v_2, \ldots, v_n\}$ and $T = \{w_1, w_2, \ldots, w_n\}$ be two ordered bases for the $n$-dimensional vector space $V$. If $v$ is any vector in $V$, then

$$ v = c_1 w_1 + c_2 w_2 + \cdots + c_n w_n \quad \text{and} \quad [V]_T = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}. $$

Then $[V]_S = [c_1 v_1 + c_2 v_2 + \cdots + c_n v_n]_S$

$$ = [c_1 v_1]_S + [c_2 v_2]_S + \cdots + [c_n v_n]_S \quad (\ast) $$

$$ = c_1 [v_1]_S + c_2 [v_2]_S + \cdots + c_n [v_n]_S. $$

Let the coordinate vector of $v$ with respect to $S$ be denoted by $[v]_S = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$. The $n \times n$ matrix whose $j$th column is $[v_j]_S$ is called the transition matrix from the $T$-basis to the $S$-basis and is denoted by $P_{S \leftarrow T}$ ($[M]_T^S$).

The equation in $(\ast)$ can be written in matrix form as $[V]_S = P_{S \leftarrow T} [V]_T$.

**Example:** Let $V$ be $\mathbb{R}^3$ and let $S = \{v_1, v_2, v_3\}$ and $T = \{w_1, w_2, w_3\}$ be ordered bases for $\mathbb{R}^3$, where $v_1 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $w_1 = \begin{bmatrix} 6 \\ 3 \\ 3 \end{bmatrix}$, $w_2 = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$, $w_3 = \begin{bmatrix} 5 \\ 5 \\ 2 \end{bmatrix}$.

a) Compute the transition matrix $P_{S \leftarrow T}$ ($[M]_T^S$) from the $T$-basis to the $S$-basis.

b) Verify Equation $(\ast)$ for $v = \begin{bmatrix} 4 \\ 2 \\ -5 \end{bmatrix}$.

**Solution:** a) To compute $P_{S \leftarrow T}$, we need to find $a_1, a_2, a_3$ such that

$$ a_1 v_1 + a_2 v_2 + a_3 v_3 = w_1, $$

which leads to a linear system of three equations in three unknowns, whose augmented matrix is $[v_1, v_2, v_3 \mid w_1]$, that is $\begin{bmatrix} 2 & 1 & 1 & 6 \\ 1 & 2 & 1 & 3 \\ 1 & 1 & 1 & 3 \end{bmatrix}$.

Similarly, we need to find $b_1, b_2, b_3, c_1, c_2, c_3$ such that

$$ b_1 v_1 + b_2 v_2 + b_3 v_3 = w_2 \quad \text{and} \quad c_1 v_1 + c_2 v_2 + c_3 v_3 = w_3. $$

These vector equations lead to two linear systems, each of three equations in three unknowns.
or specifically \[
\begin{bmatrix}
2 & 1 & 1 & 4 \\
0 & 2 & 1 & -1 \\
1 & 0 & 1 & 3
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
2 & 1 & 1 & 5 \\
0 & 2 & 1 & 5 \\
1 & 0 & 1 & 3
\end{bmatrix}
\]

Since the coefficient matrix of all three linear systems is \([V_1, V_2, V_3]\), we can transform the three augmented matrices to reduced row echelon form simultaneously by transforming the partitioned matrix \([V_1, V_2, V_3; W_1, W_2, W_3]\) to reduced row echelon form. Thus we transform:

\[
\begin{bmatrix}
2 & 1 & 1 & 6 & 1 & 4 & 5 \\
0 & 2 & 1 & 3 & 1 & -1 & 5 \\
1 & 0 & 1 & 3 & 3 & 1 & 2
\end{bmatrix}
\quad \text{to reduced row echelon form} \quad
\begin{bmatrix}
1 & 0 & 0 & 1 & 2 & 1 & 1 \\
0 & 1 & 0 & 1 & -1 & 2 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

which implies that the transition matrix from \(T\)-basis to the \(S\)-basis is

\[
P_{S \leftarrow T} = \begin{bmatrix}
2 & 2 & 1 \\
1 & -1 & 2 \\
1 & 1 & 1
\end{bmatrix}
\]

b) If \(v = \begin{bmatrix} 4 \\ -9 \\ 5 \end{bmatrix}\), then expressing \(v\) in terms of the \(T\)-basis, we have

\[
v = -\frac{4}{3} \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} + \frac{5}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.
\]

So, \([v]_T = \begin{bmatrix} 1 \\ -2 \end{bmatrix}\). Then \([v]_S = P_{S \leftarrow T} [v]_T = \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = -\frac{5}{2} \begin{bmatrix} 1 \\ -2 \end{bmatrix}\].

If we compute \([v]_S\) directly we find that

\[
v = -\frac{4}{3} \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} - 5 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \text{ so, } [v]_S = -\frac{2}{3} \begin{bmatrix} 1 \\ -2 \end{bmatrix}.
\]

**Theorem:** If \(S\) and \(T\) are ordered bases for the vector space \(\mathbb{R}^3\), then

\[
P_{S \leftarrow T} = M_S^{-1} M_T,
\]

where \(M_S\) is the \(n \times n\) matrix whose \(j\)-th column is \(v_j\) and \(M_T\) is the \(n \times n\) matrix whose \(j\)-th column is \(w_j\). This formula implies that \(P_{S \leftarrow T}\) is nonsingular and 

\[
(P_{S \leftarrow T})^{-1} = P_{T \leftarrow S}, \quad (M_S^{-1})^{-1} = M_S.
\]
Example: Compute the transition matrix $P_{T \leftarrow S} (\text{IMS}^T)$ from the $S$-basis to $T$-basis and show that $P_{T \leftarrow S} = (P_{S \leftarrow T})^{-1}$, where $V = \mathbb{R}^3$, $S = \{v_1, v_2, v_3\}$, $T = \{w_1, w_2, w_3\}$ and

\[
v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad w_1 = \begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix}, \quad w_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}, \quad w_3 = \begin{bmatrix} 5 \\ 5 \\ 2 \end{bmatrix}.
\]

$P_{T \leftarrow S}$ is the matrix whose columns are the solution vectors to the linear system:

\[
\begin{align*}
& a_1 w_1 + a_2 w_2 + a_3 w_3 = v_1 \\
& b_1 w_1 + b_2 w_2 + b_3 w_3 = v_2 \\
& c_1 w_1 + c_2 w_2 + c_3 w_3 = v_3
\end{align*}
\]

We can solve these linear systems simultaneously by transforming the partitioned matrix

\[
\begin{bmatrix}
| & | & \\
| w_1 & w_2 & w_3 \; | & v_1 & v_2 & v_3 \\
| & | & \\
\end{bmatrix}
\]

to reduced row echelon form. That is,

\[
\begin{bmatrix}
6 & 4 & 5 & 2 & 1 & 1 \\
3 & -1 & 5 & 0 & 2 & 1 \\
3 & 3 & 2 & 1 & 1 & 1
\end{bmatrix}
\] 

N

\[
\begin{bmatrix}
-1 & 2 & 5 & 1 & 1 & 1 \\
0 & 0 & 1 & -1 & 0 & 2 \\
0 & 1 & -1 & 0 & -1 & 2
\end{bmatrix}
\] 

N

\[
\begin{bmatrix}
1 & 0 & 3/2 & 1/2 & -1/2 \\
0 & 1 & -1 & 0 & 3/2 \\
0 & 0 & 1 & -1 & 0
\end{bmatrix}
\] 

Therefore, $P_{T \leftarrow S} = \begin{bmatrix} 3/2 & 1/2 & -1/2 \\
-1/2 & -1/2 & 3/2 \\
-1 & 0 & 2
\end{bmatrix}$.

We have found that $P_{S \leftarrow T} = \begin{bmatrix} 2 & 2 & 1 \\
1 & -1 & 2 \\
1 & 1 & 1
\end{bmatrix}$, hence,

\[
P_{T \leftarrow S} P_{S \leftarrow T} = \begin{bmatrix} 3/2 & 1/2 & -1/2 \\
-1/2 & -1/2 & 3/2 \\
-1 & 0 & 2
\end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\
1 & -1 & 2 \\
1 & 1 & 1
\end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{bmatrix}
\]

and then $P_{T \leftarrow S} = (P_{S \leftarrow T})^{-1}$. 


Example: Let \( V \) be \( P_1 \), and let \( S = \{ v_1, v_2 \} \) and \( T = \{ w_1, w_2 \} \) be ordered bases for \( P_1 \), where \( v_1 = t, v_2 = t+3, w_1 = t+1, w_2 = t+1 \).

a) Compute the transition matrix \( P_{S \rightarrow T} \left( [M]^S_T \right) \) from the \( T \)-basis to the \( S \)-basis.

b) Verify Equation (*) for \( v = 5t+1 \).

c) Compute the transition matrix \( P_{T \rightarrow S} \) from the \( S \)-basis to the \( T \)-basis and show that \( P_{T \rightarrow S} = (P_{S \rightarrow T})^{-1} \).

d) To compute \( P_{S \rightarrow T} \) we need to solve the vector equations

\[
\begin{align*}
0v_1 + a_2v_2 &= w_1 \\
b_1v_1 + b_2v_2 &= w_2
\end{align*}
\]
simultaneously by transforming the resulting partitioned matrix

\[
\begin{bmatrix}
v_1 & v_2 \\
w_1 & w_2
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
0 & -3 & 1
\end{bmatrix}
\]
to reduced row echelon form. The result is (Verify)

\[
\begin{bmatrix}
1 & 0 & 2/3 & 4/3 \\
0 & 1/3 & -1/3 & 4/3
\end{bmatrix}, \quad \text{so,} \quad P_{S \rightarrow T} = \begin{bmatrix}
2/3 & 4/3 \\
1/3 & -1/3
\end{bmatrix}
\]

b) If \( v = 5t+1 \), then expressing \( v \) in terms of the \( T \)-basis, we have

\( v = 5t+1 = 2(t-1) + 3(t+1) \), so \( [V]^T = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \). Then

\[
[V]^T = [V]^S = P_{S \rightarrow T} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/3 & 4/3 \\ 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 16/3 \\ -1/3 \end{bmatrix}
\]

Computing \([V]^T\) directly, we find that \( v = 5t+1 = 16 + \frac{1}{3}(t-3) \),

so, \( [V]^S = \begin{bmatrix} 16/3 \\ -1/3 \end{bmatrix} \). Hence, \([V]^T = P_{S \rightarrow T} [V]^T\).

c) \( P_{T \rightarrow S} = \begin{bmatrix} \frac{1}{2} & 2 \\ \frac{1}{2} & 1 \end{bmatrix} \)
Let $A$ be an $n \times n$ matrix. A scalar $\lambda$ is said to be an eigenvalue or a characteristic value of $A$ if there exists a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda \mathbf{x}$.

The vector $\mathbf{x}$ said to be eigenvector or characteristic vector of $A$ associated with the eigenvalue $\lambda$. (eigen is German means proper).

Since $A\mathbf{x} = \lambda \mathbf{x}$, we have $A\mathbf{x} - \lambda \mathbf{x} = 0 \Rightarrow (A - \lambda I)\mathbf{x} = 0$.

If $A - \lambda I$ is a singular matrix then the above equation has infinitely many solutions.

If $A - \lambda I$ is nonsingular, then the only solution is the trivial solution.

**Example:** Let $A = \begin{bmatrix} -1 & 1 \\ 2 & 4 \end{bmatrix}$. We wish to find the eigenvalues of $A$ and their associated eigenvectors. Thus we wish to find all numbers $\lambda$ and all nonzero vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ such that $A\mathbf{x} = \lambda \mathbf{x}$. So, $A\mathbf{x} = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ yields,

\[
\begin{align*}
-1x_1 + x_2 &= \lambda x_1 \\
2x_1 - 4x_2 &= \lambda x_2
\end{align*}
\]

This homogeneous system of two equations in two unknowns has a nontrivial solution if and only if the determinant of the coefficient matrix is zero. Thus,

\[
\begin{vmatrix}
-1 - \lambda & 1 \\
2 & 4 - \lambda
\end{vmatrix} = 0 \Rightarrow (4-\lambda)(-\lambda - 2) + 2 = 0 \Rightarrow \lambda^2 - 6\lambda + 6 = 0 \Rightarrow (\lambda - 3)(\lambda - 2) = 0
\]

So, $\lambda_1 = 2$ and $\lambda_2 = 3$ are the eigenvalues of $A$. This means that the equation (2) will have a nontrivial solution only when $\lambda_1 = 2$ or $\lambda_2 = 3$. To find all eigenvectors of $A$ associated with $\lambda_1 = 2$, we substitute $\lambda = 2$ in equation (1):

\[
\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow
\begin{align*}
x_1 + x_2 &= 2x_1 \\
-2x_1 + 4x_2 &= 2x_2
\end{align*}
\]

All solutions for this homogeneous system are: $x_1 = x_2 \\
x_2 = 0$ or a real number $r$.

Hence all eigenvectors associated with the eigenvalue $\lambda_1 = 2$ are given by $\begin{bmatrix} r \\ r \end{bmatrix}$, $r$ being any real number. Similarly, substituting $\lambda = 3$ in Equation (2), we obtain $x_1 + x_2 = 0$, and all solutions to this homogeneous system are given by $x_1 = \frac{1}{2} x_2 \\
x_2 = 0$ or a number $s$. Therefore, eigenvectors associated with $\lambda_2 = 3$ are given by $\begin{bmatrix} s \\ 2 \end{bmatrix}$, $s$ being any number.
Definition: Let \( A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \) be an \( n \times n \) matrix. Then the determinant of the matrix \( \det(\lambda I_n - A) = \begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix} \) is called the characteristic polynomial of \( A \). The equation \( p(\lambda) = \det(\lambda I_n - A) = 0 \) is called the characteristic equation of \( A \).

Each term in the expansion of the determinant of an \( n \times n \) matrix is a product of \( n \) entries of the matrix, containing exactly one entry from each row and exactly one entry from each column. Thus if we expand \( \det(\lambda I_n - A) \) we obtain a polynomial of degree \( n \). 

\[ \det(\lambda I_n - A) = \lambda^n - a_1 \lambda^{n-1} + \cdots + (-1)^n a_n \]

If \( a = 0 \) then \( \det(A) = 0 \), and thus the constant term of the characteristic polynomial of \( A \) is \( a_n = (-1)^n \det(A) \). And \( a_1 = -\text{Tr}(A) \).
EXAMPLE 11
Let \( A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{bmatrix} \). The characteristic polynomial of \( A \) is

\[ p(\lambda) = \det(\lambda I_3 - A) = \begin{vmatrix} \lambda - 1 & -2 & 1 \\ -1 & \lambda & -1 \\ -4 & 4 & \lambda - 5 \end{vmatrix} = \lambda^3 - 6\lambda^2 + 11\lambda - 6 \]

(verify).

Theorem 7.1
Let \( A \) be an \( n \times n \) matrix. The eigenvalues of \( A \) are the roots of the characteristic polynomial of \( A \).

Proof
Let \( x \) in \( \mathbb{R}^n \) be an eigenvector of \( A \) associated with the eigenvalue \( \lambda \). Then

\[ Ax = \lambda x \quad \text{or} \quad Ax = (\lambda I_n)x \quad \text{or} \quad (\lambda I_n - A)x = 0. \]

This is a homogeneous system of \( n \) equations in \( n \) unknowns; a nontrivial solution exists if and only if \( \det(\lambda I_n - A) = 0 \). Hence \( \lambda \) is a root of the characteristic polynomial of \( A \).

Conversely, if \( \lambda \) is a root of the characteristic polynomial of \( A \), then \( \det(\lambda I_n - A) = 0 \), so the homogeneous system \( (\lambda I_n - A)x = 0 \) has a nontrivial solution. Hence \( \lambda \) is an eigenvalue of \( A \).

Thus, to find the eigenvalues of a given matrix \( A \), we must find the roots of its characteristic polynomial \( p(\lambda) \). There are many methods for finding approximations to the roots of a polynomial, some of them more effective than others. Two results that are sometimes useful in this connection are as follows: (1) The product of all the roots of the polynomial

\[ p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n \]

is \((-1)^n a_n\); and (2) If \( a_1, a_2, \ldots, a_n \) are integers, then \( p(\lambda) \) cannot have a rational root that is not already an integer. Thus we need to try only the integer factors of \( a_n \) as possible rational roots of \( p(\lambda) \). Of course, \( p(\lambda) \) might well have irrational roots or complex roots.

The corresponding eigenvectors are obtained by substituting for \( \lambda \) in the matrix equation

\[ (\lambda I_n - A)x = 0 \quad (6) \]

and solving the resulting homogeneous system.
Example 12

Compute the eigenvalues and associated eigenvectors of the matrix $A$ defined in Example 11.

Solution

In Example 11 we found the characteristic polynomial of $A$ to be

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

The possible integer roots of $p(\lambda)$ are $\pm1, \pm2, \pm3, \text{and} \pm6$. By substituting these values in $p(\lambda)$, we find that $p(1) = 0$, so $\lambda = 1$ is a root of $p(\lambda)$. Hence $(\lambda - 1)$ is a factor of $p(\lambda)$. Dividing $p(\lambda)$ by $(\lambda - 1)$, we obtain

$$p(\lambda) = (\lambda - 1)(\lambda^2 - 5\lambda + 6)$$

(verify).

Factoring $\lambda^2 - 5\lambda + 6$, we have

$$p(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 3).$$

The eigenvalues of $A$ are then $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$. To find an eigenvector $x_1$ associated with $\lambda_1 = 1$, we substitute $\lambda = 1$ in (6) to get

$$\begin{bmatrix} 1 & -1 & -2 & 1 \\ -1 & 1 & -1 \\ -4 & 4 & 1 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} 0 & -2 & 1 \\ -1 & 1 & -1 \\ -4 & 4 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$ 

The vector

$$\begin{bmatrix} -\frac{r}{2} \\ \frac{r}{2} \end{bmatrix}$$

is a solution for any number $r$. Thus

$$x_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

is an eigenvector of $A$ associated with $\lambda_1 = 1$ ($r$ was taken as 2).

To find an eigenvector $x_2$ associated with $\lambda_2 = 2$, we substitute $\lambda = 2$ in (6), obtaining

$$\begin{bmatrix} 2 & -1 & -2 & 1 \\ -1 & 2 & -1 \\ -4 & 4 & 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ -4 & 4 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$
7.1 Eigenvalues and Eigenvectors

The vector \( \begin{bmatrix} -\frac{5}{2} \\ \frac{5}{4} \\ r \end{bmatrix} \) is a solution for any number \( r \). Thus \( \mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} \) is an eigenvector of \( A \) associated with \( \lambda_2 = 2 \) (\( r \) was taken as 4).

To find an eigenvector \( \mathbf{x}_3 \) associated with \( \lambda_3 = 3 \), we substitute \( \lambda = 3 \) in (6), obtaining

\[
\begin{bmatrix}
3 & -1 & -2 & 1 \\
-1 & 3 & -1 & 0 \\
-4 & 4 & 3 & -5 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix} = \begin{bmatrix} 0 \\
0 \\
0 \\
\end{bmatrix}
\]

or

\[
\begin{bmatrix}
2 & -2 & -1 \\
-1 & 3 & -1 & 0 \\
-4 & 4 & -2 & 0 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix} = \begin{bmatrix} 0 \\
0 \\
0 \\
\end{bmatrix}.
\]

The vector \( \begin{bmatrix} -\frac{r}{4} \\ \frac{r}{4} \\ r \end{bmatrix} \) is a solution for any number \( r \). Thus

\[
\mathbf{x}_3 = \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix}
\]

is an eigenvector of \( A \) associated with \( \lambda_3 = 3 \) (\( r \) was taken as 4).

**EXAMPLE 13**

Compute the eigenvalues and associated eigenvectors of

\[
A = \begin{bmatrix}
0 & 0 & 3 \\
1 & 0 & -1 \\
0 & 1 & 3 \\
\end{bmatrix}
\]

**Solution**

The characteristic polynomial of \( A \) is

\[
p(\lambda) = \det(\lambda I_3 - A) = \begin{vmatrix}
\lambda - 0 & 0 & -3 \\
-1 & \lambda - 0 & 1 \\
0 & -1 & \lambda - 3 \\
\end{vmatrix} = \lambda^3 - 3\lambda^2 + \lambda - 3
\]

(verify). We find that \( \lambda = 3 \) is a root of \( p(\lambda) \). Dividing \( p(\lambda) \) by \( (\lambda - 3) \), we get \( p(\lambda) = (\lambda - 3)(\lambda^2 + 1) \). The eigenvalues of \( A \) are then

\[
\lambda_1 = 3, \quad \lambda_2 = i, \quad \lambda_3 = -i.
\]

To compute an eigenvector \( \mathbf{x}_1 \) associated with \( \lambda_1 = 3 \), we substitute \( \lambda = 3 \) in (6), obtaining

\[
\begin{bmatrix}
3 & 0 & -3 \\
-1 & 3 & 0 \\
0 & -1 & 3 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix} = \begin{bmatrix} 0 \\
0 \\
0 \\
\end{bmatrix}.
\]
EXAMPLE 14

We find that the vector
\[
\begin{bmatrix}
r \\
0 \\
r
\end{bmatrix}
\]
is a solution for any number \( r \) (verify). Letting \( r = 1 \), we conclude that
\[
\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}
\]
is an eigenvector of \( A \) associated with \( \lambda_1 = 3 \). To obtain an eigenvector \( \mathbf{x}_2 \) associated with \( \lambda_2 = i \), we substitute \( \lambda = i \) in (6), which yields
\[
\begin{bmatrix}
i - 0 & 0 & -3 \\ -1 & i - 0 & 1 \\ 0 & -1 & i - 3
\end{bmatrix}
\begin{bmatrix}
x_1 \\ x_2 \\ x_3
\end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\]

We find that the vector
\[
\begin{bmatrix}
(\cdot - 3i) r \\ (\cdot - 3 + i) r \\ r
\end{bmatrix}
\]
is a solution for any number \( r \) (verify). Letting \( r = 1 \), we conclude that
\[
\mathbf{x}_2 = \begin{bmatrix} -3i \\ -3 + i \\ 1 \end{bmatrix}
\]
is an eigenvector of \( A \) associated with \( \lambda_2 = i \). Similarly, we find that
\[
\mathbf{x}_3 = \begin{bmatrix} 3i \\ -3 - i \\ 1 \end{bmatrix}
\]
is an eigenvector of \( A \) associated with \( \lambda_3 = -i \).
The procedure for finding the eigenvalues and associated eigenvectors of a matrix is as follows:

**Step 1.** Determine the roots of the characteristic polynomial

\[ p(\lambda) = \det(\lambda I_n - A). \]

These are the eigenvalues of \( A \).

**Step 2.** For each eigenvalue \( \lambda \), find all the nontrivial solutions to the homogeneous system \((\lambda I_n - A)x = 0\). These are the eigenvectors of \( A \) associated with the eigenvalue \( \lambda \).

Eigenvalues and eigenvectors satisfy many important and interesting properties. For example, if \( A \) is an upper (lower) triangular matrix, then the eigenvalues of \( A \) are the elements on the main diagonal of \( A \) (Exercise 11). Other properties are developed in the exercises for this section.

It must be pointed out that the method for finding the eigenvalues of a linear transformation or matrix by obtaining the real roots of the characteristic polynomial is not practical for \( n > 4 \), since it involves evaluating a determinant. Efficient numerical methods for finding eigenvalues and associated eigenvectors are studied in numerical analysis courses.

**Warning** When finding the eigenvalues and associated eigenvectors of a matrix \( A \), do not make the common mistake of first transforming \( A \) to reduced row echelon form \( B \) and then finding the eigenvalues and eigenvectors of \( B \). To see quickly how this approach fails, consider the matrix \( A \) defined in Example 10. Its eigenvalues are \( \lambda_1 = 2 \) and \( \lambda_2 = 3 \). Since \( A \) is a nonsingular matrix, when we transform it to reduced row echelon form \( B \), we have \( B = I_2 \). The eigenvalues of \( I_2 \) are \( \lambda_1 = 1 \) and \( \lambda_2 = 1 \).
The Cayley–Hamilton theorem states that a matrix satisfies its characteristic equation; that is, if \( A \) is an \( n \times n \) matrix with characteristic polynomial
\[
p(\lambda) = \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1} \lambda + a_n,
\]
then
\[
A^n + a_1 A^{n-1} + \cdots + a_{n-1} A + a_n I_n = O.
\]

The Cayley–Hamilton Theorem provides a method for calculating powers of a matrix. For example, if \( A \) is a 2 \times 2 matrix with characteristic equation
\[
c_0 + c_1 \lambda + \lambda^2 = 0
\]
then \( c_0 I + c_1 A + A^2 = 0 \), so
\[
A^2 = -c_1 A - c_0 I
\]
Multiplying through by \( A \) yields \( A^3 = -c_1 A^2 - c_0 A \), which expresses \( A^3 \) in terms of \( A^2 \) and \( A \), and multiplying through by \( A^2 \) yields \( A^4 = -c_1 A^3 - c_0 A^2 \), which expresses \( A^4 \) in terms of \( A^3 \) and \( A^2 \). Continuing in this way, we can calculate successive powers of \( A \) by expressing them in terms of lower powers.
Solution  The characteristic polynomial of $A$ is

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

so the characteristic equation is

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

and the distinct eigenvalues of $A$ are $\lambda = 1$ and $\lambda = 2$. We leave it for you to show that bases for the eigenspaces are

$$\lambda = 1: \quad p_1 = \begin{bmatrix}
\frac{1}{8} \\
-\frac{1}{8} \\
1
\end{bmatrix} \quad \lambda = 2: \quad p_2 = \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}$$

Since $A$ is a $3 \times 3$ matrix and there are only two basis vectors in total, $A$ is not diagonalizable.

Alternative Solution  If you are concerned only in determining whether a matrix is diagonalizable and not with actually finding a diagonalizing matrix $P$, then it is not necessary to compute bases for the eigenspaces—it suffices to find the dimensions of the eigenspaces. For this example, the eigenspace corresponding to $\lambda = 1$ is the solution space of the system

$$
\begin{bmatrix}
0 & 0 & 0 \\
-1 & -1 & 0 \\
3 & -5 & -1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} =
\begin{bmatrix} 0 \\
0 \\
0
\end{bmatrix}
$$

Since the coefficient matrix has rank 2 (verify), the nullity of this matrix is 1 by Theorem 4.8.2, and hence the eigenspace corresponding to $\lambda = 1$ is one-dimensional.

The eigenspace corresponding to $\lambda = 2$ is the solution space of the system

$$
\begin{bmatrix}
1 & 0 & 0 \\
-1 & 0 & 0 \\
3 & -5 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} =
\begin{bmatrix} 0 \\
0 \\
0
\end{bmatrix}
$$

This coefficient matrix also has rank 2 and nullity 1 (verify), so the eigenspace corresponding to $\lambda = 2$ is also one-dimensional. Since the eigenspaces produce a total of two basis vectors, and since three are needed, the matrix $A$ is not diagonalizable.

T2. (The Cayley–Hamilton Theorem) states that every square matrix satisfies its characteristic equation; that is, if $A$ is an $n \times n$ matrix whose characteristic equation is

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

then $A^n + c_1A^{n-1} + \cdots + c_n = 0$.
9. The Cayley–Hamilton Theorem provides a method for calculating powers of a matrix. For example, if \( A \) is a \( 2 \times 2 \) matrix with characteristic equation
\[
c_0 + c_1 \lambda + \lambda^2 = 0
\]
then \( c_0 I + c_1 A + A^2 = 0 \), so
\[
A^2 = -c_1 A - c_0 I
\]
Multiplying through by \( A \) yields \( A^3 = -c_1 A^2 - c_0 A \), which expresses \( A^3 \) in terms of \( A^2 \) and \( A \), and multiplying through by \( A^2 \) yields \( A^4 = -c_1 A^3 - c_0 A^2 \), which expresses \( A^4 \) in terms of \( A^3 \) and \( A^2 \). Continuing in this way, we can calculate successive powers of \( A \) by expressing them in terms of lower powers. Use this procedure to calculate \( A^2 \), \( A^3 \), \( A^4 \), and \( A^5 \) for
\[
A = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}
\]
* Cayley–Hamilton Theorem also provides a method calculating the inverse of a nonsingular matrix. Let \( A \) is an \( n \times n \) matrix whose characteristic polynomial is \( p(\lambda) = \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 \). If \( A \) is nonsingular, then \( \det(A) \neq 0 \). So, \( a_n = (-1)^n \det(A) \neq 0 \)
Since \( A \) satisfies its characteristic equation,
\[
A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I = 0
\]
Therefore,
\[
a_n I = -A(a_{n-1} A^{n-2} + \cdots + a_1 I)
\]
(Multiplying both sides by \( \frac{1}{a_n} \))
\[
A^{-1} = \frac{1}{a_n} \left( A^{n-1} + a_1 A^{n-2} + \cdots + a_{n-1} I \right)
\]
Example: Let \( A = \begin{bmatrix} 1 & 0 & 5 \\ -2 & -4 & -3 \\ 3 & 6 & 0 \end{bmatrix} \) be a nonsingular matrix. Find the inverse \( A' \) of \( A \) and \( A^5 \) by using Cayley–Hamilton Theorem.
First, we have to find the characteristic polynomial.
\[
\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & 0 & -5 \\ 2 & \lambda + 4 & 3 \\ -3 & -6 & \lambda \end{vmatrix} = 0
\Rightarrow (\lambda - 1)(\lambda + 4) + 60 - 15(\lambda + 4) + 18(\lambda - 1) = 0
\Rightarrow \lambda^3 + 3\lambda^2 - 18 = 0
\Rightarrow \lambda = \frac{1}{2} \left( -3 \pm \sqrt{9 + 36} \right)
\Rightarrow \lambda = -3, 2, 3
\Rightarrow A^{-1} = \frac{1}{a_n} \left( A^{n-1} + a_1 A^{n-2} + \cdots + a_{n-1} I \right)
\]
Therefore, from Cayley-Hamilton Theorem:

\[ A^3 + 3A^2 - A - 18I_3 = 0 \Rightarrow 18A^{-1} = A^4 + 3A - I_3 \]

\[ A^{-1} = \frac{1}{18} (A^2 + 3A - I_3) \Rightarrow A^{-1} = \frac{1}{18} \begin{bmatrix} 1 & 0 & 5 \\ -2 & -4 & -3 \\ 3 & 6 & 0 \end{bmatrix} + \begin{bmatrix} -3 & 0 & 15 \\ -6 & -12 & -9 \\ 9 & 18 & 0 \end{bmatrix} - I_3 \]

\[ = \frac{1}{18} \begin{bmatrix} 16 & 30 & 5 \\ -3 & -2 & 2 \\ -9 & -24 & -3 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 15 \\ -6 & -13 & -9 \\ 9 & 18 & -1 \end{bmatrix} = \begin{bmatrix} 1/2 & 5/3 & 10/9 \\ -1/2 & -5/6 & -7/18 \\ 0 & -1/3 & 2/9 \end{bmatrix} \]

\[ A^3 = -3A^2 + A + 18I_3 \text{ (Multiplying both sides by } A) \]

\[ A^4 = 3A^3 + A^2 + 18A = -3(-3A^2 + A + 18I_3) + A^2 + 18A \]

\[ = -9A^2 - 3A - 54I_3 + A^2 + 18A \]

\[ A^4 = 10A^3 + 15A^2 - 54A = 10(-3A^2 + A + 18I_3) + 15A^2 - 54A \]

\[ A^5 = 10A^4 + 15A^3 - 54A = 10(10A^3 + 15A^2 - 54A) + 15A^3 - 54A \]

\[ = -30A^2 + 10A + 180I_3 + 15A^2 - 54A \]

Therefore, \[ A^5 = -15A^2 + 25A + 180I_3 \]

\[ A^5 = \begin{bmatrix} -240 & -450 & -75 \\ 45 & 30 & -30 \\ -15 & 360 & 45 \end{bmatrix} + \begin{bmatrix} -44 & 0 & -220 \\ 88 & 176 & 132 \\ -132 & -264 & 0 \end{bmatrix} + \begin{bmatrix} 180 & 0 & 0 \\ 0 & 180 & 0 \\ 0 & 0 & 180 \end{bmatrix} \]

\[ = \begin{bmatrix} -104 & -450 & -295 \\ 133 & 386 & 102 \\ 3 & 96 & 215 \end{bmatrix} \]

THE END...