Lecture Notes 2025

Linear Algebra

Shreas Arion
shreas.labib.arion@g.bracu.ac.bd

Preface

Hey, my name is Shreas. I am creating this to be intended as a stripped down version of Linear Algebra by Hoffman Kunze. I recommend you only read this if you have finished reading the main book, otherwise it will be difficult to follow.

If you find any issues please email me at **shreas.labib.arion@g.bracu.ac.bd**

Bibliography

[1] Kenneth Hoffman and Ray Kunze, *Linear Algebra*, 2nd Edition, Prentice-Hall, 1971.

Contents

Chapter 1 Linear Equations — **Page 5**

1.1	Fields	5
1.2	Systems of Linear Equations	6
1.3	Matrices and Elementary Row Operations	7
1.4	Row-Reduced Echelon Matrices	10
1.5	Matrix Multiplication	12
1.6	Invertible Matrices	15

Chapter 2 Vector Spaces — **Page 18**

2.1	Vector Spaces	18
2.2	Subspaces	20
2.3	Bases and Dimension	23
2.4	Coordinates	28
2.5	Summary of Row-Equivalence	30

Chapter 1

Linear Equations

1.1 Fields

We let F denote either the set of real numbers or the set of complex numbers.

1. Addition is commutative,
$$x + y = y + x$$
for all x and y in F.

2. Addition is associative,
$$x + (y + z) = (x + y) + z$$
for all x, y, and z in F.

3. There is a unique element 0 (zero) in F such that $x + 0 = x$, for every x in F.

4. To each x in F there corresponds a unique element $(-x)$ in F such that $x + (-x) = 0$.

5. Multiplication is commutative,
$$xy = yx$$
for all x and y in F.

6. Multiplication is associative,
$$x(yz) = (xy)z$$
for al x, y, and z in F.

7. There is a unique non-zero element 1 (one) in F such that $x1 = x$, for every x in F.

8. To each non-zero x in F there corresponds a unique element x^{-1} (or $\frac{1}{x}$) in F such that $xx^{-1} = 1$.

9. Multiplication distributes over addition; that is, $x(y + z) = xy + xz$, for all x, y, and z in F.

Suppose one has a set F of objects $x, y, z, \ldots$ and two operations on the elements of F as follows. The first operation, called addition, associates with each pair of elements x, y in F an element $(x + y)$ in F; the second operation, called multiplication, associates with each pair x, y an element xy in F; and these two operations satisfy conditions $(1) - (9)$ above. The set F, together with these two operations, is then called a **field**. The set $\mathbb{C}$ of complex numbers is a field, as is the set $\mathbb{R}$ of real numbers.

Example 1.1.1

The set of positive integers: $1, 2, 3, \ldots$, is not a subfield of $\mathbb{C}$, for a variety of reasons. For example, 0 is not a positive integer; for not positive integer n is $-n$ a positive integer; for no positive integer n except 1 is $1/n$ a positive integer.

Example 1.1.2

The set of integers: $\ldots, -2, -1, 0, 1, 2, \ldots$, is not a subfield of $\mathbb{C}$, because for an integer n, $1/n$ is not an integer unless n is 1 or -1. With the usual operations of addition and multiplication, the set of integers satisfies all of the conditions $(1) - (9)$ except condition (8).

1.2 Systems of Linear Equations

Suppose F is a field. We consider the problem of finding n scalars $x_1, \ldots, x_n$ which satisfy the conditions

$$\begin{array}{ccccccccc} A_{11}x_1 & + & A_{12}x_2 & + & \cdots & + & A_{1n}x_n & = & y_1 \\ A_{21}x_1 & + & A_{22}x_2 & + & \cdots & + & A_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ A_{m1}x_1 & + & A_{m2}x_2 & + & \cdots & + & A_{mn}x_n & = & y_m \end{array} \tag{1-1}$$

where $y_1, \ldots, y_m$ and A_{ij}, $1 \le i \le m, 1 \le j \le n$, are given elements of F. We call $(1-1)$ a **system of m linear equations in n unknowns**. Any n-tuple $(x_1, \ldots, x_n)$ of elements of F which satisfies each of the equations in $(1-1)$ is called a solution of the system. If $y_1 = \ldots = y_m = 0$, we say that the system is **homogeneous**, or that each of the equations is homogeneous. The most fundamental technique for finding the solutions of a system of linear equations is the technique of elimination.

$$\begin{array}{ccccccc} 2x_1 & - & 1x_2 & + & 1x_3 & = & 0 \\ 1x_1 & + & 3x_2 & + & 4x_3 & = & 0. \end{array} \tag{1.1}$$

If we add (-2) times the second equation to the first equation, we obtain

$$-7x_2 - 7x_3 = 0$$

or, $x_2 = -x_3$. If we add 3 times the first equation to the second equation, we obtain

$$7x_1 + 7x_3 = 0$$

or,$x_1 = -x_3$. So we conclude that if (x_1, x_2, x_3) is a solution then $x_1 = x_2 = -x_3$. The set of solutions consists of all triples $(-a, -a, a)$.

For the general system (1-1), suppose we select m scalars $c_1, \ldots, c_m$, multiply the jth equation by c_j and then add. We obtainn the equation

$$(c_1A_{11} + \ldots + c_mA_{m1})x_1 + \ldots + (c_1A_{1n} + \ldots + c_mA_{mn})x_n = c_1y_1 + \ldots + c_my_m.$$

Such an equation is called a **Linear Combination** of the equations in $(1-1)$. Any solution of the entire system of equations $(1-1)$ will also be a solution of this new equation. This is the fundamental idea of the elimination process. If we have another system of linear equations

$$\begin{array}{ccccccc} B_{11}x_1 & + & \cdots & + & B_{1n}x_n & = & z_1 \\ \vdots & & \vdots & & \vdots & & \vdots \\ B_{k1}x_1 & + & \cdots & + & B_{kn}x_n & = & z_k \end{array} \tag{1-2}$$

in which each of the k equations is a linear combination of the equations in $(1-1)$, then every solution of $(1-1)$ is a solution of this new system. Some solutions of $(1-2)$ may not be solutions of $(1-1)$. Two systems of linear equations are **equivalent** if each equation in each system is a linear combination of the equations in the other system. We can then formally state our observations as follows.

Theorem 1.2.1

Equivalent systems of linear equations have exactly the same solutions.

1.3 Matrices and Elementary Row Operations

We shall now abbreviate the system (1-1) by

$$AX = Y$$

$$\text{where} \quad A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

We call A the **matrix of coefficients** of the system. The rectangular array displayed above is not a matrix, bt is a representation of a matrix. An $m \times n$ **matrix over the field** F is a function A from the set of pairs of integers (i, j), $1 \le i \le m$, $1 \le j \le n$, into the field F. The **entries** of the matrix A are the scalars $A(i, j) = A_{ij}$. Above, X is, or defines an $n \times 1$ matrix and Y is an $m \times 1$ matrix.

We wish now to consider operations on the rows of the matrix A which correspond to forming inear combinations of the equations in the system $AX = Y$. We restrict our attention to three **elementary row operations** on an $m \times n$ matrix A over the field F:

1. multiplication of one row of A by a non-zero scalar c;
2. replacement of the rth row of A by row r plus c times row s, c any scalar and $r \neq s$;
3. interchange of two rows of A.

An elementary row operation iis a special type of function (rule) which associated with each $m \times n$ matrix A an $m \times n$ matrix $e(A)$. One can precisely describe e in the three cases as follows:

1. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e(A)_{rj} = cA_{rj}$.
2. $e(A)_{ij} = A_{ij}$ if $i \neq r$, $e(A)_{rj} = A_{rj} + cA_{sj}$.
3. $e(A)_{ij} = A_{ij}$ if i is different from both r and s, $e(A)_{rj} = A_{sj}$, $e(A)_{sj} = A_{rj}$.

In defining $e(A)$ the number of columns is not important, but the number of rows of A is crucial. An elementary row operation e is defined on the class of all $m \times n$ matrices over F, for some fixed m but any n. In other words, a particular e is defined on the class of all m-rowed matrices over F.

Theorem 1.3.1

To each elementary row operation e there corresponds an elementary row operation e_1, of the same type as e, such that $e_1(e(A)) = e(e_1(A)) = A$ for each A. In other words, the inverse operation of an elementary row operation exists and is an elementary row operation of the same type.

Proof: (1) Suppose e is the operation which multiplies the rth row of a matrix by the non-zero scalar c. Let e_1 be the operation which multiplies row r by c^{-1}. (2) Suppose e is the operation which replaces row r by row r plus c times row s, $r \neq s$. Let e_1 be the operation which replaces row r by r plus $(-c)$ times row s. (3) If e interchanges row r and s, let $e_1 = e$. In each of these three cases we clearly have $e_1(e(A)) = e(e_1(A)) = A$ for each A. ■

Definition 1.3.1: Row Equivalence

If A and B are $m \times n$ matrices over the field F, we say that B is **row-equivalent to** A if B can be obtained from A by a finite sequence of elementary row operations.

Each matrix is row-equivalent to itself; if B is row-equivalent to A, then A is row-equivalent to B; if B is row-equivalent to A and C is row-equivalent to B, then C is row-equivalent to A. In other words, row-equivalence is an equivalence relation.

Theorem 1.3.2

If A and B are row-equivalent $m \times n$ matrices, the homogeneous systems of linear equations $AX = 0$ and $BX = 0$ have exactly the same solutions.

Proof: Suppose we pass from A to B by a finite sequence of elementary row operations:

$$A = A_0 \longrightarrow A_1 \longrightarrow \ldots \longrightarrow A_k = B.$$

It is enough to prove that the systems $A_jX = 0$ and $A_{j+1}X = 0$ have the same solutions, i.e., that one elementary row operation does not disturb the set of solutions.

Suppose that B is obtained from A by a single elementary row operation. No matter which of the three types of operations is, (1), (2), or (3), each equation in the system $BX = 0$ will be a linear combination of the equations in the system $AX = 0$. Since the inverse of an elementary row operation is an elementary row operation, each equation in $AX = 0$ will also be a linear combination of the equations in $BX = 0$. Hence these two systems are equivalent, and by Theorem 1.2.1 they have the same solutions. ■

Example 1.3.1

Suppose F is the field of rational numbers, and

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

We shall perform a finite sequence of elementary row operations on A, indicating by numbers in parantheses the type of operation performed.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}$$

The row-equivalence of A with the final matrix in the above sequence tells us in particular that the solutions of

$$\begin{array}{rcrcrcrcl} 2x_1 & - & x_2 & + & 3x_3 & + & 2x_4 & = & 0 \\ x_1 & + & 4x_2 & & & - & x_4 & = & 0 \\ 2x_1 & + & 6x_2 & - & x_3 & + & 5x_4 & = & 0 \end{array} \tag{1.2}$$

and

$$\begin{array}{rrrcrcl} & & x_3 & - & \frac{11}{3}x_4 & = & 0 \\ x_1 & & & + & \frac{17}{3}x_4 & = & 0 \\ & x_2 & & - & \frac{5}{3}x_4 & = & 0 \end{array} \tag{1.3}$$

are exactly the same. In the second system it is apparent that if we assign any rational value to c to x_4 we obtain a solution $(-\frac{17}{3}c, \frac{5}{3}, \frac{11}{3}c, c)$, and also that every solution is of this form.

Definition 1.3.2: Row-Reduced Matrix

An $m \times n$ matrix R is called row-reduced if:

- the first non-zero entry in each non-zero row of R is equal to 1;
- each column of R which contains the leading non-zero entry of some row has all its other entries 0.

Example 1.3.2

One example of a row-reduced matrix is the $n \times n$ (square) **identity matrix** I. This is the $n \times n$ matrix defined by

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

This is the first of many occasions on which we shall use the **Kronecker delta**.

Theorem 1.3.3

Every $m \times n$ matrix over the field F is row-equivalent to a row-reduced matrix.

Proof: Let A be an $m \times n$ matrix over F. If every entry in the first row of A is 0, then condition (a) is satisfied in so far as row 1 is concerned. If row 1 has a non-zero entry, let k be the smallest positive integer j for which $A_{1j} \neq 0$. Multiply row 1 by $A_{1k}{}^{-1}$, and then condition (a) is satisfied with regard to row 1. Now for each $i \geq 2$, add $(-A_{ik})$ times row 1 to row i. Now the leading non-zero entry of row 1 occurs in column k, that entry is 1, and every other entry in column k is 0.

Now consider the matrix which has resulted from above. If every entry in row 2 is 0, we do nothing to row 2. If some entry in row 2 is different from 0, we multiply row 2 by a scalar so that the leading non-zero entry is 1. In the event that row 1 had a leading non-zero entry in column k, this leading non-zero entry of row 2 cannot occur in column k; say it occurs in column $k_r \neq k$. By adding suitable multiples of row 2 to the various rows, we can arrange that al entries in column k' are 0, except the 1 in row 2. The important thing to notice is this: In carrying out these last operations, we will not change the entries of row 1 in columns $1, \ldots, k$; nor will we change any entry of column k. Of course, if row 1 was identically 0, the operations with row 2 will not affect row 1.

Working with one row at a timein the above manner, it is clear that in a finite number of steps we will arrive at a row-reduced matrix. ■

1.4 Row-Reduced Echelon Matrices

Definition 1.4.1: Row-Reduced Echelon Matrix

An $m \times n$ matrix R is called a **row-reduced echelon matrix** if:

(a) R is row-reduced;

(b) every row of R which has all its entries 0 occurs below every row which has a non-zero entry.

(c) if rows $1, \ldots, r$ are the non-zero rows of R, and if the leading non-zero entry of row i occurs n column k_i, $i = 1, \ldots r$, then $k_1 < k_2 < \ldots < k_r$.

Theorem 1.4.1

Every $m \times n$ matrix A is row-equivalent to a row-reduced echelon matrix.

Proof: We know that A is row-equivalent to a row-reduced matrix. All that we need to observe is that by performinng a finite number of row interchanges on a row-reduced matrix we can bring it to row-reduced echelon form. ■

Let us now discuss briefly the system $RX = 0$, when R is a row-reduced echelon matrix. Let rows $1, \ldots, r$ be the non-zero rows of R, and supposet that the leading non-zero entry of row i occurus in column k_i. The system $RX = 0$ then consists of r non-trivia equations. Also the unknown x_{k_i} will occur (with non-zero coefficient) only in the ith equation. If we let $u_1, \ldots, u_{n-r}$ denote the $(n-r)$ unknowns which are different from $x_{k_1}, \ldots, x_{k_r}$, then the r non-trivial equations in $RX = 0$ are of the form

$$\begin{aligned} x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j &= 0 \\ \vdots \qquad\qquad & \vdots \\ x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j &= 0. \end{aligned} \tag{1-3}$$

All the solutions to the system of equations $RX = 0$ are obtained by assigning any values whatsoever to $u_1, \ldots, u_{n-r}$ and then computing the corresponding values of $x_{k_1}, \ldots, x_{k_r}$ from (1-3).

If the number r of non-zero rows in R is less than n, then the system $RX = 0$ has a non-trivial solution, that is, a solution $(x_1, \ldots, x_n)$ in which not every x_j is 0. For, since $r < n$, we can choose some x_j which is not among the r unknowns $x_{k_1}, \ldots, x_{k_r}$, and we can then construct a solution as above in which this x_j is 1. This observation leads us to one of the most fundamental facts concerning systems of homegeneous linear equations.

Theorem 1.4.2

If A is an $m \times n$ matrix and $m < n$, then the homogeneous system of linear equations $AX = 0$ has a non-trivial solution.

Proof: Let R be a row-reduced echelon matrix which is row-equivalent to A. Then the systems $AX = 0$ and $RX = 0$ have the same solutions. If r is the number of non-zero rows in R, then certainly $r \leq m$, and since $m < n$, we have $r < n$. It follows immediately from our remarks above that $AX = 0$ has a non-trivial solution. ■

Theorem 1.4.3

If A is an $n \times n$ (square) matrix, then A is row-equivalent to the $n \times n$ identity matrix if and only if the system of equations $AX = 0$ has only the trivial solution.

Proof: If A is row-equivalent to I, then $AX = 0$ and $IX = 0$ have the same solutions. Conversely, suppose $AX = 0$ has only the trivial solution $X = 0$. Let R be an $n \times n$ row-reduced echelon matrix which is row-equivalent to A, and let r be the number of non-zero rows of R. Then $RX = 0$ has no non-trivial solution. Thus $r \geq n$. But since R has n rows, certainly $r \leq n$, and we have $r = n$. Since this means that R actually has a leading non-zero entry of 1 in each of its n rows, and since these 1's occur each in a different one of the n columns, R must be the $n \times n$ identity matrix. ∎

The homogeneous system always has the trivial solution $x_1 = \ldots = x_n = 0$, an inhomogeneous system need have no solution at all.

We form the **augmented matrix** A' of the system $AX = Y$. This is the $m \times (n+1)$ matrix whose first $n+1$ columns are the columns of A and whose last column is Y. More precisely,

$$A'_{ij} = A_{ij}, \text{ if } j \leq n$$

$$A'_{i(n+1)} = y_i.$$

Suppose we form a sequence of elementary row operations on A, arriving at a row-reduced echelon matrix R. If we perform this same sequence of row operations on the augmented matrix A', we will arrive at a matrix R' whose first n columns are the columns of R and whose last column contains scalars $z_1, \ldots, z_m$. The scalars z_i are the entries of the $m \times 1$ matrix

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

which results from applying the sequence of row operations to the matrix Y. The systems $AX = Y$ and $RX = Z$ are equivalent and hence have the same solutions. If R has r non-zero rows, with the leading non-zero entry of row i occurring in column k_i, $i = 1, \ldots, r$, then the first r equations of $RX = Z$ effectively express $x_{k_1}, \ldots, x_{k_r}$ in terms of the $(n-r)$ remaining x_j annd the scalars $z_1, \ldots, z_r$. The last $(m-r)$ equations are

$$\begin{array}{ccc} 0 & = & z_{r+1} \\ \vdots & & \vdots \\ 0 & = & z_m \end{array}$$

and accordingly for the system to have a solution is $z_i = 0$ for $i > r$. If this condition is satisfied, all solutions to the system are found just as in the homogeneous case, by assigning arbitrary values to $(n-r$ of the x_j and then computing x_{k_i} from the ith equation.

1.5 Matrix Multiplication

Suppose B is an $n \times p$ matrix over a field F with rows $\beta_1, \ldots, \beta_n$ and that from B we construct a matrix C with rows $\gamma_1, \ldots, \gamma_m$ by forming certain linear combinations.

$$\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \ldots + A_{in}\beta_n.$$

The rows of C are determined by the mn scalars A_{ij} which are themselves the entries of an $m \times n$ matrix A. If (1-4) is expanded to

$$(C_{i1} \dots C_{ip}) = \sum_{r=1}^{n} (A_{ir}B_{r1} \dots A_{ir}B_{rp})$$

we see that the entries of C are given by

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Definition 1.5.1: Matrix Product

Let A be an $m \times n$ matrix over the field F and let B be an $n \times p$ matrix over F. The **product** AB is the $m \times p$ matrix C whose i, j entry is

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

The product of two matrices need not be defined; the product is defined if and only if the number of columns in the first matrix coincides with the number of rows in the second matrix. Even with the products AB and BA are both defined it need not be true that $AB = BA$; in other words, matrix multiplication is not commutative.

Theorem 1.5.1

If A, B, C are matrices over the field F such that the products BC and $A(BC)$ are defined, then so are the products AB, $(AB)C$ and

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

Proof: Suppose B is an $n \times p$ matrix. Since BC is defined, C is a matrix with p rows, and BC has n rows. Because $A(BC)$ is defined we may assume A is an $m \times n$ matrix. Thus the product AB exists and is an $m \times p$ matrix, from which it follows that the product $(AB)C$ exists. To show that $A(BC) = (AB)C$ means to show that

$$[A(BC)]_{ij} = [(AB)C]_{ij}$$

for each i, j. By definition

$$
\begin{aligned}
[A(BC)]_{ij} &= \sum_r A_{ir}(BC)_{rj} \\
&= \sum_r A_{ir} \sum_s B_{rs} C_{sj} \\
&= \sum_r \sum_s A_{ir} B_{rs} C_{sj} \\
&= \sum_s \sum_r A_{ir} B_{rs} C_{sj} \\
&= \sum_s \left(\sum_r A_{ir} B_{rs} \right) C_{sj} \\
&= \sum_s (AB)_{is} C_{sj} \\
&= [(AB)C]_{ij}
\end{aligned}
$$

∎

When A is an $n \times n$ matrix, the product $AA \cdots A$ (k times) is unambiguously defined, and we shall denote this product by A^k.

The relation $A(BC) = (AB)C$ implies that linear combinations of linear combinations of the rows of C are again linear combinations of the rows of C. If B is a given matrix and C is obtained from B by means of an elementary row operation, then each row of C is a linear combination of the rows of C, and hence there is a matrix A such that $AB = C$.

Definition 1.5.2: Elementary Matrix

An $m \times n$ matrix is said to be an elementary matrix if it can be obtained from the $m \times m$ identity matrix by means of a single elementary row operation.

Theorem 1.5.2

Let e be an elementary row operation and let E be the $m \times m$ elementary matrix $E = e(I)$. Then, for every $m \times n$ matrix A,

$$e(A) = EA.$$

Corollary 1.5.3

Let A and B be $m \times n$ matrices over the field F. Then B is row-equivalent to A if and only if $B = PA$, where P is a product of $m \times m$ elementary matrices.

Proof: Suppose $B = PA$ where $P = E_s \cdots E_2 E_1$ and the E_i are $m \times m$ elementary matrices. Then $E_1 A$ is row-equivalent to A, and $E_2(E_1 A)$ is row-equivalent to $E_1 A$. So $E_2 E_1 A$ is row-equivalent to A; and continuing this way we see that $(E_s \cdots E_1)A$ is row-equivalent to A.

Now suppose that B is row-equivalent to A. Let $E_1, E_2, \ldots, E_s$ be the elementary matrices corresponding to some sequence of elementary row operations which carries A into B. Then $B = (E_s \cdots E_1)A$.

∎

1.6 Invertible Matrices

Definition 1.6.1: Invertible Matrices

Let A be an $n \times n$ (square) matrix over the field F. An $n \times n$ matrix B such that $BA = I$ is called a **left inverse** of A; an $n \times n$ matrix B such that $AB = I$ is called a **right inverse** of A. If $AB = BA = I$, then B is called a **two-sided inverse** of A and A is said to be invertible.

Lemma 1.6.1

If A has a left inverse B and a right inverse C, then $B = C$.

Proof: Suppose $BA = I$ and $AC = I$. Then

$$B = BI = B(AC) = (BA)C = IC = C.$$

■

If A has a left and a right inverse, A is invertible and has a unique two-sided inverse, which we shall denote by A^{-1} and simply call **the inverse** of A.

Theorem 1.6.2

Let A and B be $n \times n$ matrices over F.

1. If A is invertible, so is A^{-1} and $(A^{-1})^{-1} = A$.
2. If both A and B are invertible, so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof: The first statement is evident from the symmetry of the definition. The second follows upon verification of the relations

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

■

Corollary 1.6.3

A product of invertible matrices is invertible.

Theorem 1.6.4

An elementary matrix is invertible.

Proof: Let E be an elementary matrix corresponding to the elementary row operation e. If e_1 is the inverse operation of e and $E_1 = e_1(I)$, then

$$EE_1 = e(E_1) = e(e_1(I)) = I$$

and

$$E_1 E = e_1(E) = e_1(e(I)) = I$$

so that E is invertible and $E_1 = E^{-1}$ ■

Theorem 1.6.5

If A is an $n \times n$ matrix, the following are equivalent.

1. A is invertible.
2. A is row-equivalent to the $n \times n$ identity matrix.
3. A is a product of elementary matrices.

Proof: Let R be a row-reduced echelon matrix which is row-equivalent to A. We know

$$R = E_k \cdots E_2 E_1 A$$

where $E_1, \ldots, E_k$ are elementary matrices. Each E_j is invertible, and so

$$A = E_1^{-1} \cdots E_k^{-1} R.$$

Since products of invertible matrices are invertible, we see that A is invertible if and only if R is invertible. Since R is a square row-reduced echelon matrix, R is invertible if and only if each row of R contains a non-zero entry, that is, if and only if $R = I$. We have now shown that A is invertible if and only if $R = I$, and if $R = I$ then $A = E_k^{-1} \cdots E_1^{-1}$. It should be now apparent that 1, 2 and 3 are equivalent statements about A. ■

Corollary 1.6.6

If A is an invertible $n \times n$ matrix and if a sequence of elementary row operations reduces A to the identity, then that same sequence of operations when applied to I yields A^{-1}.

Corollary 1.6.7

Let A and B be $m \times n$ matrices. Then B is row-equivalent to A if and only if $B = PA$ where P is an invertible $m \times m$ matrix.

Theorem 1.6.8

For an $n \times n$ matrix A, the following are equivalent.

1. A is invertible.
2. The homogenous system $AX = 0$ has only the trivial solution $X = 0$.
3. The system of equations $AX = Y$ has a solution X for each $n \times 1$ matrix Y.

Proof: Condition 2 is equivalent to the fact that A is row-equivalent to the identity matrix. Therefore, 1 and 2 are equivalent. If A is invertible, the solution of $AX = Y$ is $X = A^{-1}Y$. Conversely, suppose $AX = Y$ has a solution for each given Y. Let R be a row-reduced echelon matrix which is row-equivalent to A. We wish to show that $R = I$. That amounts to showing that the last row of R is not (identically) 0. Let

$$E = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

If the system $RX = E$ can be solved for X, the last row of R cannot be 0. We know that $R = PA$, where P is invertible. Thus $RX = E$ if and only if $AX = P^{-1}E$. According to 3, the latter system has a solution. ■

Corollary 1.6.9

A square matrix with either a left or right inverse is invertible.

Proof: Let A be an $n \times n$ matrix. Suppose A has a left inverse, i.e., a matrix B such that $BA = I$. Then $AX = 0$ has only the trivial solution, because $X = IX = B(AX)$. Therefore A is invertible. On the other hand, suppose A has a right inverse, i.e., a matrix C such that $AC = I$. Then C has a left inverse and is therefore invertible. It then follows that $A = C^{-1}$ and so A is invertible with inverse C. ■

Corollary 1.6.10

Let $A = A_1A_2 \cdots A_k$ where $A_1, \cdots, A_k$ are $n \times n$ (square) matrices. Then A is invertible if and only if each A_j is invertible.

Proof: We know the product of two invertible matrices is invertible. From this one sees easily that if each A_j is invertible then A is invertible.

Suppose now that A is invertible. We first prove that A_k is invertible. Suppose X is an $n \times 1$ matrix and $A_kX = 0$. Then $AX = (A_1 \cdots A_{k-1})A_kX = 0$. Since A is invertible we must have $X = 0$. The system of equations $A_kX = 0$ thus has no trivial solution, so A_k is invertible. But now $A_1 \cdots A_{k-1} = AA_k^{-1}$ is invertible. By the preceding argument, A_{k-1} is invertible. Continuing this way, we conclude that each A_j is invertible. ■

Suppose A is an $m \times n$ matrix and we wish to solve the system of equations $AX = Y$. If R is a row-reduced echelon matrix which is row-equivalent to A, then $R = PA$ where P is an $m \times m$ invertible matrix. The solutions of the system $AX = Y$ are exactly the same as the solutions of the system $RX = PY$ $(= Z)$.

Chapter 2

Vector Spaces

2.1 Vector Spaces

Definition 2.1.1: Vector Space

A **vector space** (or linear space) consists of the following:

1. a field F of scalars;
2. a set V of objects, called vectors;
3. a rule (or operation), called vector addition, which associates with each pair of vectors α, β in V a vector $\alpha + \beta$ in V, called the sum of α and β, in such a way that
   (a) addition is commutative, $\alpha + \beta = \beta + \alpha$;
   (b) addition is associative, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$;
   (c) there is a unique vector 0 in V, called the zero vector, such that $\alpha + 0 = \alpha$ for all α in V;
   (d) for each vector α in V there is a unique vector $-\alpha$ in V such that $\alpha + (-\alpha) = 0$;
4. a rule (or operation), called scalar multiplication, which associates with each scalar c in F and vector α in V a vector $c\alpha$ in V, called the product of c and α, in such a way that
   (a) $1\alpha = \alpha$ for every α in V;
   (b) $(c_1 c_2)\alpha = c_1(c_2\alpha)$;
   (c) $c(\alpha + \beta) = c\alpha + c\beta$;
   (d) $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$.

A vector space is a composite object consisting of a field, a set of vectors, and two operations with certain special properties. The same set of vectors may be part of a number of distinct vector spaces.

Example 2.1.1

The n-tuple space, F^n. Let F be any field, and let V be the set of all n-tuples $\alpha = (x_1, x_2, \cdots, x_n)$ of scalars x_i in F. If $\beta = (y_1, y_2, \cdots y_n)$ with y_i in F, the sum of α and β is defined by

$$\alpha + \beta = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n).$$

The product of a scalar c and vector α is defined by

$$c\alpha = (cx_1, cx_2, \cdots, cx_n).$$

Example 2.1.2

The space of $m \times n$ matrices, $F^{m\times n}$. Let F be any field and let m and n by positive integers. Let $F^{m\times n}$ be the set of all $m \times n$ matrices over the field F. The sum of two vectors A and B in $F^{m\times n}$ is defined by

$$(A + B)_{ij} = A_{ij} + B_{ij}.$$

The product of a scalar c and the matrix A is defined by

$$(cA)_{ij} = cA_{ij}.$$

Example 2.1.3

The space of functions from a set to a field. Let F be any field and let S be any non-empty set. Let V be the set of all functions from the set S into F. The sum of two vectors f and g in V is the vector $f + g$, i.e., the function from S into F, defined by

$$(f + g)(s) = f(s) + g(s).$$

The product of the scalar c and the function f is the function cf defined by

$$(cf)(s) = cf(s).$$

Example 2.1.4

The space of polynomial functions over a field F. Let F be a field and let V be the set of all functions f from F into F which have a rule of the form

$$f(x) = c_0 + c_1x + \cdots + c_nx^n$$

where $c_0, c_1, \cdots, c_n$ are fixed scalars in F. A function of this type is called a **polynomial function on F.** If f and g are polynomial functions and c is in F, then $f + g$ and cf are again polynomial functions.

If c is a scalar and 0 is the zero vector, then

$$c0 = c(0 + 0) = c0 + c0.$$

Adding $-(c0)$, we obtain

$$c0 = 0.$$

Similarly, for the scalar 0 and any vector α we find that

$$0\alpha = 0.$$

If c is a non-zero scalar and α is a vector such that $c\alpha = 0$, then, $c^{-1}(c\alpha) = 0$. But

$$c^{-1}(c\alpha) = (c^{-1}c)\alpha = 1\alpha = \alpha$$

hence, $\alpha = 0$. If c is a scalar and α a vector such that $c\alpha = 0$, then either c is the zero scalar or α is the zero vector.

If α is any vector in V, then

$$0 = 0\alpha = (1-1)\alpha = 1\alpha + (-1)\alpha = \alpha + (-1)\alpha$$

from which it follows that

$$(-1)\alpha = -\alpha.$$

Definition 2.1.2: Linear Combination

A vector β in V is said to be a linear combination of the vectors $\alpha_1, \cdots, \alpha_n$ in V provided there exists scalars $c_1, \cdots, c_n$ in F such that

$$\beta = c_a\alpha_1 + \cdots + c_n\alpha_n = \sum_{i=1}^{n} c_i\alpha_i.$$

2.2 Subspaces

Definition 2.2.1: Subspace

Let V be a vector space over the field F. A **subspace** of V is a subset W of V which is itself a vector space over F with the operations of vector addition and scalar multiplication on V.

The subset W of V is a subspace if for each α and β in W the vector $\alpha + \beta$ is again in W; the 0 vector is in W; for each α in W the vector $(-\alpha)$ is in W; for each α in W and each scalar c the vector $c\alpha$ is in W.

Theorem 2.2.1

A non-empty subset W of V is a subspace of V if and only if for each pair of vectors α, β in W and each scalar c in F the vector $c\alpha + \beta$ is again in W.

Proof: Suppose that W is a non-empty subset of V such that $c\alpha + \beta$ belongs to W for all vectors α, β in W and all scalars c in F. Since W is non-empty, there is a vector ρ in W, and hence $(-1)\rho + \rho = 0$ is in W. Then if α is any vector in W and c any scalar, the vector $c\alpha = c\alpha + 0$ is in W. In particular,

$(-1)\alpha = -\alpha$ is in W. Finally, if α and β are in W, then $\alpha + \beta = 1\alpha + \beta$ is in W. Thus W is a subspace of V.

Conversely, if W is a subspace of V, α and β are in W, and c is a scalar, certainly $c\alpha + \beta$ is in W. ■

If W is a non-empty subset of V such that $c\alpha + \beta$ is in V for all α, β in W and all c in F, then W is a vector space.

- If V is any vector space, V is a subspace of V; the subset consisting of the zero vector alone is a subspace of V, called the **zero sub-space** of V.
- In F^n, the set of n-tuples $(x_1, \cdots, x_n)$ with $x_1 = 0$ is a subspace; however, the set of n-tuples with $x_1 = 1 + x_2$ is not a subspace $(n \geq 2)$.
- The space of polynomial functions over the field F is a subspace of the space of all functions from F into F.
- An $n \times n$ matrix A over the field F is **symmetric** if $A_{ij} = A_{ji}$ for each i and j. The symmetric matrices form a subspace of the space of all $n \times n$ matrices over F.
- An $n \times n$ matrix A over the field $\mathbb{C}$ of complex numbers is **Hermitian** (or **self-adjoint**) if

  $$A_{jk} = \overline{A_{kj}}$$

  for each j, k, the bar denoting complex conjugation.

  A 2×2 matrix is Hermitian if and only if it has the form

  $$\begin{bmatrix} z & x + iy \\ x - iy & w \end{bmatrix}$$

where x, y, z and w are real numbers. The set of all Hermitean matrices is not a subspace of the space of all $n \times n$ matrices over $\mathbb{C}$.

Lemma 2.2.2

If A is an $m \times n$ matrix over F and B, C are $n \times p$ matrices over F then

$$A(dB + C) = d(AB) + AC$$

for each scalar d in F.

Proof:

$$\begin{aligned}
[A(dB + C)]_{ij} &= \sum_k A_{ik}(dB + C)_{kj} \\
&= \sum_k (dA_{ik}B_{kj} + A_{ik}C_{kj}) \\
&= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\
&= d(AB)_{ij} + (AC)_{ij} \\
&= [d(AB) + AC]_{ij}
\end{aligned}$$

■

Theorem 2.2.3

Let V be a vector space over the field F. The intersection of any collection of subspaces of V is a subspace of V.

Proof: Let $\{W_a\}$ be a collection of subspaces of V, and let $W = \bigcap_a W_a$ be their intersection. We know W is defined as the set of all elements belonging to every W_a. Since each W_a is a subspace, each contains the zero vector. Thus the zero vector is in the intersection W, and W is non-empty. Let α and β be vectors in W and let c be a scalar. By definition of W, both α and β belong to each W_a, and because each W_a is a subspace, the vector $(c\alpha + \beta)$ is in every W_a. Thus $(c\alpha + \beta)$ is again in W. Therefore, W is a subspace of V. ■

From the above theorem it follows that if S is any collection of vectors in V, then there is a smallest subspace of V which contains S, that is, a subspace which contains S and which is contained in every other subspace containing S.

Definition 2.2.2: The Subspace Spanned

Let S be a set of vectors in a vector space V. The subspace spanned by S is defined to be the intersection W of all subspaces of V which contains S. When S is a finite set of vectors, $S = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, we shall simply call W the **subspace spanned by the vectors** $\alpha_1, \alpha_2, \cdots, \alpha_n$**.**

Theorem 2.2.4

The subspace spanned by a non-empty subset S of a vector space V is the set of all linear combinations of vectors in S.

Proof: Let W be the subspace spanned by S. Then each linear combination

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

of vectors $\alpha_1, \alpha_2, \cdots, \alpha_m$ in S is clearly in W. Thus W contains the set L of all linear combinations of vectors in S. The set L, on the other hand, contains S and is non-empty. If α, β belong to L then α is a linear combination,

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m$$

of vectors α_i in S, and β is a linear combination,

$$\beta = y_1\beta_1 + y_2\beta_2 + \cdots + y_n\beta_n$$

of vectors β_j in S. For each scalar c,

$$c\alpha + \beta = \sum_{i=1}^{m}(cx_i)\alpha_i + \sum_{j=1}^{n} y_i\beta_j.$$

Hence $c\alpha + \beta$ belongs to L. Thus L is a subspace of V.

Now we have shown that L is a subspace of V which contains S, and also that any subspace which contains S contains L. It follows that L is the intersection of all subspaces containing S, i.e., that L is the subspace spanned by the set S. ■

Definition 2.2.3: Sum of the Subsets

If $S_1, S_2, \cdots, S_k$ are subsets of a vector space V, the set of all sums

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k$$

of vectors α_i in S_i is called the sum of the subsets $S_1, S_2, \ldots, S_k$ and is denoted by

$$S_1 + S_2 + \cdots + S_k$$

or by

$$\sum_{i=1}^{k} S_i.$$

Example 2.2.1

Let A be an $m \times n$ matrix over a field F. The **row vectors** of A are the vectors in F^n given by $\alpha_i = (A_{i1}, \ldots, A_{in}), i = 1, \ldots, m$. The subspace of F^n spanned by the row vectors of A is called the **row space** of A.

Example 2.2.2

Let V be the space of all polynomial functions over F. Let S be the subset of V consisting of the polynomial functions $f_0, f_1, f_2, \cdots$ defined by

$$f_n(x) = x^n, \qquad n = 0, 1, 2, \ldots$$

Then V is the subspace spanned by the set S.

2.3 Bases and Dimension

Definition 2.3.1: Linearly Dependent

Let V be a vector space over F. A subset S of V is said to be linearly dependent if there exist distinct vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ in S and scalars $c_1, c_2, \ldots, c_n$ in F, not all of which are 0, such that

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = 0.$$

A set which is not linearly dependent is called linearly independent.

The following are easy consequences of the definition.

1. Any set which contains a linearly dependent set is linearly dependent.
2. Any subset of a linearly independent set is linearly independent.
3. Any set which contains the 0 vector is linearly independent.
4. A set S of vectors is linearly independent if and only if each finite subset of S is linearly independent, i.e., if and only if for any distinct vectors $\alpha_1, \ldots, \alpha_n$ of S, $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$ implies each $c_i = 0$.

Definition 2.3.2: Basis

Let V be a vector space. A basis for V is a linearly independent set of vectors in V which spans the space V. The space V is finite-dimensional if it has a finite basis.

Example 2.3.1

Let P be an invertible $n \times n$ matrix with entries in the field F. Then $P_1, \ldots, P_n$, the columns of P, form a basis for the space of column matrices, $F^{n\times 1}$. We see that as follows. If X is a column matrix, then

$$PX = x_1P_1 + \ldots + x_nP_n.$$

Since $PX = 0$ has only the trivial solution $X = 0$, it follows that $\{P_1, \ldots, P_n\}$ is a linearly independent set. Why does it span $F^{n\times 1}$? Let Y be any column matrix. If $X = P^{-1}Y$, then $Y = PX$, that is,

$$Y = x_1P_1 + \ldots + x_nP_n.$$

So $\{P_1, \ldots, P_n\}$ is a basis for $F^{n\times 1}$.

Theorem 2.3.1

Let V be a vector space which is spanned by a finite set of vectors $\beta_1, \beta_2, \ldots, \beta_m$. Then any independent set of vectors in V is finite and contains no more than m elements.

Proof: To prove the theorem it suffices to show that every subset S of V which contains more than m vectors is linearly dependent. Let S be such a set. In S there are distinct vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ where $n > m$. Since $\beta_1, \ldots, \beta_m$ span V, there exist scalars A_{ij} in F such that

$$\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i.$$

For any n scalars $x_1, x_2, \ldots, x_n$ we have

$$\begin{aligned} x_1\alpha_1 + \ldots + x_n\alpha_n &= \sum_{j=1}^{n} x_j\alpha_j \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m}(A_{ij}x_j)\beta_i \\ &= \sum_{i=1}^{m}(\sum_{j=1}^{n} A_{ij}x_j)\beta_i. \end{aligned}$$

Since $n > m$, we know that there exist scalars $x_1, x_2, \ldots, x_n$ not all 0 such that

$$\sum_{j=1}^{n} A_{ij}x_j = 0, \qquad 1 \le i \le m.$$

Hence $x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n = 0$. This shows that S is a linearly dependent set. ■

Corollary 2.3.2

Let V be a finite-dimensional vector space and let $n = \dim V$. Then

1. any subset of V which contains more than n vectors is linearly dependent;
2. no subset of V which contains fewer than n vectors can span V.

Lemma 2.3.3

Let S be a linearly independent subset of a vector space V. Suppose β is a vector in V which is not in the subspace spanned by S. Then the set obtained by adjoining β to S is linearly independent.

Proof: Suppose $\alpha_1, \ldots, \alpha_m$ are distinct vectors in S and that

$$c_1\alpha_1 + \ldots + c_m\alpha_m + b\beta = 0.$$

Then $b = 0$; for otherwise,

$$\beta = \left(-\frac{c_1}{b}\right)\alpha_1 + \ldots + \left(-\frac{c_m}{b}\right)\alpha_m$$

and β is in the subspace spanned by S. Thus $c_1\alpha_1 + \ldots + c_m\alpha_m = 0$, and since S is a linearly independent set each $c_i = 0$. ■

Theorem 2.3.4

If W is a subspace of a finite-dimensional vector space V, every linearly independent subset of

> W is finite and is part of a (finite) basis for W.

Proof: Suppose S_0 is a linearly independent subset of W. If S is a linearly independent subset of W containing S_0, then S is also a linearly independent subset of V; since V is finite-dimensional, S contains no more than $\dim V$ elements.

We extend S_0 to a basis for W, as follows. If S_0 spans W, then S_0 is a basis for W and we are done. If S_0 does not span W, we use the preceding lemma to find a vector β_1 in W such that the set $S_1 = S_0 \cup \{\beta_1\}$ is independent. If S_1 spans W, fine. If not, apply the lemma to obtain a vector β_2 in W such that $S_2 = S_1 \cup \{\beta_2\}$ is independent. If we continue in this way, then (in not more than $\dim V$ steps) we reach a set

$$S_m = S_0 \cup \{\beta_1, \ldots, \beta_m\}$$

which is a basis for W. ■

> **Corollary 2.3.5**
>
> If W is a proper subspace of a finite-dimensional vector space V, then W is finite-dimensional and $\dim W < \dim V$.

Proof: We may suppose W contains a vector $\alpha \neq 0$. we know, there is a basis of W containing α which contains no more than $\dim V$ elements. Hence W is finite-dimensional, and $\dim W \leq \dim V$. Since W is a proper subspace, there is a vector β in V which is not in W. Adjoining β to any basis of W, we obtain a linearly independent subset of V. Thus $\dim W < \dim V$. ■

> **Corollary 2.3.6**
>
> In a finite-dimensional vector space V every non-empty linearly independent set of vectors is part of a basis.

> **Corollary 2.3.7**
>
> Let A be an $n \times n$ matrix over a field F, and suppose the row vectors of A form a linearly independent set of vectors in F^n. Then A is invertible.

Proof: Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the row vectors of A, and suppose W is the subspace of F^n spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Since $\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly independent, the dimension of W is n. Corollary 2.3.5 now shows that $W = F^n$. Hence there exist scalars B_{ij} in F such that

$$\epsilon_i = \sum_{j=1}^{n} B_{ij}\alpha_j, \qquad 1 \leq i \leq n$$

where $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ is the standard basis of F^n. Thus for the matrix B with entries B_{ij} we have

$$BA = I.$$

■

Theorem 2.3.8

If W_1 and W_2 are finite-dimensional subspaces of a vector space V, then $W_1 + W_2$ is finite-dimensional and

$$\dim W_1 + \dim W_2 = \dim (W_1 \cap W_2) + \dim (W_1 + W_2).$$

Proof: We know, $W_1 \cap W_2$ has a finite basis $\{\alpha_1, \ldots, \alpha_k\}$ which is part of a basis

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m\} \text{ for } W_1$$

and part of a basis

$$\{\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_n\} \text{ for } W_2.$$

The subspace $W_1 + W_2$ is spanned by the vectors

$$\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_n$$

and these vectors form an independent set. For suppose

$$\sum x_i \alpha_i + \sum y_j \beta_j + \sum z_r \gamma_r = 0.$$

Then

$$-\sum z_r \gamma_r = \sum x_i \alpha_i + \sum y_j \beta_j$$

which shows that $\sum z_r \gamma_r$ belongs to W_1. As $\sum z_r \gamma_r$ also belongs to W_2 it follows that

$$\sum z_r \gamma_r = \sum c_i \alpha_i$$

for certain scalars $c_1, \ldots, c_k$. Because the set

$$\{\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_n\}$$

is independent, each of the scalars $z_r = 0$. Thus

$$\sum x_i \alpha_i + \sum y_j \beta_j = 0$$

and since

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m\}$$

is also an independent set, each $x_i = 0$ and each $y_j = 0$. Thus,

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_n\}$$

is a basis for $W_1 + W_2$. Finally

$$\begin{aligned} \dim W_1 + \dim W_2 &= (k + m) + (k + n) \\ &= k + (m + k + n) \\ &= \dim (W_1 \cap W_2) + \dim (W_1 + W_2). \end{aligned}$$

■

2.4 Coordinates

Definition 2.4.1

If V is a finite-dimensional vector space, an **ordered basis** for V is a finite sequence of vectors which is linearly independent and spans V.

If the sequence $\alpha_1, \ldots, \alpha_n$ is an ordered basis for V, then the set $\{\alpha_1, \ldots, \alpha_n\}$ is a basis for V. Lets say that

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$$

is an ordered basis for V.

Now suppose V is a finite-dimensional vector space over the field F and that

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}$$

is an ordered basis for V. Given α in V, there is a unique n-tuple $(x_1, \ldots, x_n)$ of scalars such that

$$\alpha = \sum_{i=1}^{n} x_i \alpha_i.$$

The n-tuple is unique, because if we also have

$$\alpha = \sum_{i=1}^{n} z_i \alpha_i$$

then

$$\sum_{i=1}^{n} (x_i - z_i)\alpha_i = 0$$

and the linear independence of the α_i tells us that $x_i - z_i = 0$ for each i. We shall call x_i the ith **coordinate of α relative to the ordered basis**

$$\mathfrak{B} = \{\alpha_1, \ldots, \alpha_n\}.$$

If

$$\beta = \sum_{i=1}^{n} y_i \alpha_i$$

then

$$\alpha + \beta = \sum_{i=1}^{n} (x_i + y_i)\alpha_i$$

so that the ith coordinate of $(\alpha + \beta)$ in this ordered basis is $(x_i + y_i)$. Similarly, the ith coordinate of $(c\alpha)$ is cx_i. Every n-tuple $(x_1, \ldots, x_n)$ in F^n is the n-tuple of coordinates of some vector in V, namely the vector

$$\sum_{i=1}^{n} x_i \alpha_i.$$

It will be more convenient to use the **coordinate matrix of α relative to the ordered basis $\mathfrak{B}$:**

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

rather than the n-tuple $(x_1, \dots, x_n)$ of coordinates. To indicate the dependence of this coordinate matrix on the basis, we shall use the symbol

$$[\alpha]_{\mathfrak{B}}$$

for the coordinate matrix of the vector α relative to the ordered basis $\mathfrak{B}$.

Suppose that V is n-dimentional and that

$$\mathfrak{B} = \{\alpha_1, \dots, \alpha_n\} \quad \text{and} \quad \mathfrak{B}' = \{\alpha_1', \dots, \alpha_n'\}$$

are two ordered bases for V. There are unique scalars P_{ij} such that

$$\alpha_j' = \sum_{i=1}^{n} P_{ij}\alpha_i, \quad 1 \le j \le n.$$

Theorem 2.4.1

Let V be an n-dimensional vector space over the field F, and let $\mathfrak{B}$ and $\mathfrak{B}'$ be two ordered bases of V. Then there is a unique, necessarily invertible, $n \times n$ matrix P with entries in F such that

1. $[\alpha]_{\mathfrak{B}} = P[\alpha]_{\mathfrak{B}'}$
2. $[\alpha]_{\mathfrak{B}'} = P^{-1}[\alpha]_{\mathfrak{B}}$

for every vector α in V. The columns of P are given by

$$P_j = [\alpha_j']_{\mathfrak{B}}, \quad j = 1, \dots, n.$$

Theorem 2.4.2

Suppose P is an $n \times n$ invertible matrix over F. Let V be an n-dimensional vector space over F, and let $\mathfrak{B}$ be an ordered basis of V. Then there is a unique ordered basis $\mathfrak{B}'$ of V such that

1. $[\alpha]_{\mathfrak{B}} = P[\alpha]_{\mathfrak{B}'}$
2. $[\alpha]_{\mathfrak{B}'} = P^{-1}[\alpha]_{\mathfrak{B}}$

for every vector α in V.

Proof: Let $\mathfrak{B}$ consist of the vectors $\alpha_1, \dots, \alpha_n$. If $\mathfrak{B}' = \{\alpha_1', \dots, \alpha_n'\}$ is an ordered basis of V for which 1 is valid, it is clear that

$$\alpha_j' = \sum_{i=1}^{n} = P_{ij}\alpha_i.$$

Thus we only need to show that the vectors α'_j, defined by these equations, form a basis. Let $Q = P^{-1}$. Then

$$\begin{aligned}\sum_j Q_{jk}\alpha'_j &= \sum_j Q_{jk}\sum_i P_{ij}\alpha_i \\ &= \sum_j\sum_i P_{ij}Q_{jk}\alpha_i \\ &= \sum_i\left(\sum_j P_{ij}Q_{jk}\right)\alpha_i \\ &= \alpha_k\end{aligned}$$

Thus the subspace spanned by the set

$$\mathfrak{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

contains $\mathfrak{B}$ and hence equals V. Thus $\mathfrak{B}'$ is a basis, and from its definition and Theorem 2.4.1, it is clear that 1 is valid and hence also 2. ■

2.5 Summary of Row-Equivalence

We know that if A is an $m \times n$ matrix over the field F the row vectors of A are the vectors $\alpha_1, \ldots, \alpha_m$ in F^n defined by

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

and that the row space of A is the subspace of F^n spanned by these vectors. The **row rank** of A is the dimension of the row space of A.

Theorem 2.5.1

Let R be a non-zero row-reduced echelon matrix. Then the non-zero row vectors of R form a basis for the row space of R.

Proof: Let $\rho_1, \ldots, \rho_r$ be the non-zero row vectors of R:

$$\rho_i = (R_{i1}, \ldots, R_{in}).$$

These vectors span the row space of R; we need only prove they are linearly independent. Since R is a row-reduced echelon matrix, there are positive integers $k_1, \ldots, k_r$ such that, for $i \leq r$

1. $R(i, j) = 0$ if $j < k_i$
2. $R(i, k_j) = \delta_{ij}$
3. $k_1 < \cdots < k_r$.

Suppose $\beta = (b_1, \cdots, b_n)$ is a vector in the row space of R:

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r.$$

Then we claim that $c_j = b_{k_i}$. For,

$$\begin{aligned} b_{k_i} &= \sum_{i=1}^{r} c_i R(i, k_j) \\ &= \sum_{i=1}^{r} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

In particular, if $\beta = 0$, i.e., if $c_1\rho_1 + \cdots + c_r\rho_r = 0$, then c_j must be the k_jth coordinate of the zero vector so that $c_j = 0, j = 1, \dots, r$. Thus $\rho_1, \dots, \rho_r$ are linearly independent. ■

Theorem 2.5.2

Let m and n be positive integers and let F be a field. Suppose W is a subspace of F^n and $\dim W \leq m$. Then there is precisely one $m \times n$ row-reduced echelon matrix over F which has W as its row space.

Proof: There is atleast one $m \times n$ row-reduced echelon matrix with row space W. Since $\dim W \leq m$, we can select some m vectors $\alpha_1, \dots, \alpha_m$ in W which span W. Let A be the $m \times n$ matrix with row vectors $\alpha_1, \dots, \alpha_m$ and let R be a row-reduced echelon matrix which is row-equivalent to A. Then the row space of R is W.

Now let R be any row-reduced echelon matrix which has W as its row space. Let $\rho_1, \dots, \rho_r$ be the non-zero row vectors of R and suppose that the leading non-zero entry of ρ_i occurs in column $k_i, i = 1, \dots, r$. The vectors $\rho_1, \dots, \rho_r$ form a basis for W. We know if $\beta - (b_1, \dots, b_n)$ is in W, then

$$\beta = c_1\rho_1 + \cdots + c_r\rho_r,$$

and $c_i = b_{k_i}$; in other words, the unique expression for β as a linear combination of $\rho_1, \dots, \rho_r$ is

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i.$$

Thus any vector β is determined if one knows the coordinates $b_{k_i}, i = 1, \dots, r$.

Suppose β is in W and $\beta \neq 0$. We claim the first non-zero coordinate of β occurs in one of the columns k_s. Since

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i$$

and $\beta \neq 0$, we can write

$$\beta = \sum_{i=s}^{r} b_{k_i}\rho_i, \quad b_{k_s} \neq 0.$$

One has $R_{ij} = 0$ if $i > s$ and $j \leq k_s$. Thus

$$\beta = (0, \ldots, 0, \quad b_{k_s}, \ldots b_n), \quad b_{k_s} \neq 0$$

and the first non-zero coordinate of β occurs in column k_s. Note also that for each $k_s, s = 1, \ldots, r$, there exists a vector in W which has a non-zero k_sth coordinate, namely ρ_s.

It is now clear that R is uniquely determined by W. The description of R in terms of W is as follows. We consider all vectors $\beta = (b_1, \ldots, b_n)$ in W. If $\beta \neq 0$, then the first non-zero coordinate of β must occur in some column t:

$$\beta = (0, \ldots, 0, \quad , b_t, \ldots, b_n), \quad b_t \neq 0.$$

Let $k_1, \ldots, k_r$ be those positive integers t such that there is some $\beta \neq 0$ in W, the first non-zero coordinate of which occurs in column t. Arrange $k_1, \ldots, k_r$ in the order $k_1 < k_2 < \cdots < k_r$. For each of the positive integers k_s there will be one and only one vector ρ_s in W such that the k_sth coordinate of ρ_s is 1 and the k_ith coordinate of ρ_s is 0 for $i \neq s$. Then R is the $m \times n$ matrix which has the row vectors $\rho_1, \ldots, \rho_r, 0, \ldots, 0$. ■

Corollary 2.5.3

Each $m \times n$ matrix A is row-equivalent to one and only one row-reduced echelon matrix.

Proof: We know that A is row-equivalent to atleast one row-reduced echelon matrix R. If A is row-equivalent to another such matrix R', then R is row-equivalent to R'; hence R and R' have the same row space and must be identical. ■

Corollary 2.5.4

Let A and B be $m \times n$ matrices over the field F. Then A and B are row-equivalent if and only if they have the same row space.

Proof: We know that if A and B are row-equivalent, then they have the same row space. So suppose that A and B have the same row space. Now A is row-equivalent to a row-reduced echelon matrix R and B is row-equivalent to a row-reduced echelon matrix R'. Since A and B have the same row space, R and R' have the same row space. Thus $R = R'$ and A is row-equivalent to B. ■