CHAPTER THREE

THEORY OF LINEAR EQUATIONS

3.1. Linear dependence and independence over a field

We begin our study of the linear dependence and independence of a set of vectors $x^1, x^2, \ldots, x^m$ by the following definition.

DEFINITION 1. If there exists a set of m vectors $x^1, x^2, \ldots, x^m$, and a set of m scalars $c_1, c_2, \ldots, c_m$ with a linear relation of the form

$$c_1x^1 + c_2x^2 + \cdots + c_mx^m = 0$$

then the vectors $x^1, x^2, \ldots, x^m$ are called *linearly independent* if the above relation implies that

$$c_1 = c_2 = \cdots = c_m = 0$$

But if one or more of the coefficients $c_1, c_2, \ldots, c_m$ can be non-zero, then the vectors are called *linearly dependent.*

If the coefficients are real the vectors are said to be dependent over the field of real numbers; likewise if the coefficients are complex the vectors are said to be dependent over the field of complex numbers.

Example

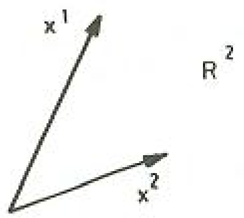

Fig. (1)

Fig. (2)

In figure 1 the vectors are linearly independent in the sense that there is no linear relation of the type

$$c_1x^1 + c_2x^2 = 0$$

In other words if such a relation did exist, then $c_1 = c_2 = 0$. On the other hand the vectors of figure 2 are linearly dependent for there exists a relation of the type

$$c_1x^1 + c_2x^2 = 0, \quad c_1, c_2 \neq 0$$

Example. Test the dependence of the following vectors

$$x^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad x^1 \text{ and } x^2 \in \mathbb{R}^2$$

We form the linear relation

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

From which we obtain $c_1 = c_2 = 0$; hence x^1 and x^2 are linearly independent.

Example. Test the dependence of the following vectors

$$x^1 = \begin{bmatrix} 1 \\ i \end{bmatrix}, \quad x^2 = \begin{bmatrix} i \\ -1 \end{bmatrix}, \quad i^2 = -1, \quad x^1 \text{ and } x^2 \in \mathbb{C}^2$$

We form the linear relation

$$c_1 \begin{bmatrix} 1 \\ i \end{bmatrix} + c_2 \begin{bmatrix} i \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

There are no non-zero real scalars c_1 and c_2 to make the vectors dependent; in other words

$$c_1 + ic_2 = 0$$
$$c_1 i - c_2 = 0$$

implies that

$$c_1 = c_2 = 0$$

Therefore the vectors are called independent over the field of real numbers. However there exist non-zero complex quantities c_1 and c_2 such that

$$c_1x^1 + c_2x^2 = 0$$

in other words such that x^1 and x^2 are linearly dependent. For example one can choose

$$c_1 = 1, \quad c_2 = i$$

The vectors x^1 and x^2 are hence said to be dependent over the field of complex numbers.

Now we are in a position to list some theorems on the linear dependence and independence of a set of vectors $x^1, x^2, \ldots, x^m$. All these theorems rely on the definition of linear dependence and independence cited before. A résumé of these theorems can be found in Kreko (1962), p. 117.

1. Any set of vectors containing the zero vector is linearly dependent.

Proof. If in the linear relation

$$c_1x^1 + c_2x^2 + \cdots + c_mx^m = 0$$

we have

$$c_1 = c_2 = \cdots = c_{m-1} = 0$$

but $c_m \neq 0$, we still obtain

$$0x^1 + 0x^2 + \cdots + 0x^{m-1} + c_mx^m = 0,$$

if $x^m = 0$. Hence $x^1, \ldots, x^m$ are linearly dependent.

2. Any non-empty subset of a set of linearly independent vectors is itself a linearly independent system.

Proof. Assume that the vectors $x^1, x^2, \ldots, x^m$ are linearly independent. If we cancel one of them, say x^1, the remaining system $x^2, \ldots, x^m$ is linearly independent. For if the equation

$$c_2x^2 + c_3x^3 + \cdots + c_mx^m = 0$$

has a non-trivial solution for $c_2, \ldots, c_m$, then so does

$$c_1x^1 + c_2x^2 + \cdots + c_mx^m = 0$$

if $c_1 = 0$. But this contradicts the assumption that $x^1, x^2, \ldots, x^m$ are linearly independent. Therefore, if we remove a vector from a system of linearly independent vectors, the remaining vectors are still linearly independent. If we remove another vector from the remaining ones, we still have a linearly independent system and so on, which completes the proof.

3. If the vectors $x^1, x^2, \ldots, x^m$ are linearly dependent, at least one of them can be written as a linear combination of the others.

Proof. As $x^1, x^2, \ldots, x^m$ are linearly dependent, there exists a set of coefficients $c_1, \ldots, c_m$, at least one of them non-zero, such that

$$c_1x^1 + c_2x^2 + \cdots + c_mx^m = 0$$

Let c_i be such a non-zero coefficient. Hence we obtain

$$x_i = -\frac{1}{c_i}(c_1x^1 + \cdots + c_mx^m)$$

4. If the vector y can be written as a linear combination of $x^1, \ldots, x^m$ the set of vectors $y, x^1, \ldots, x^m$ form a linearly dependent set.

Proof

$$y = c_1x^1 + \cdots + c_mx^m$$

Hence

$$-y + c_1x^1 + \cdots + c_mx^m = 0$$

The set $y, x^1, \ldots, x^m$ is linearly dependent, for at least the coefficient of y is non-zero.

Now we can study further a vector system, which is a finite set of vectors in the vector space V, and make the following definition.

DEFINITION 2. The rank of the set of vectors $x^1, x^2, \ldots, x^m$ is equal to the maximum number of linearly independent vectors in the set. This is equal to the total number of vectors in the set minus the number of linear relations existing among them.

The next theorems discuss properties of the rank of a vector system.

5. If r is the rank of a vector system, every vector in the system can be written as a linear combination of any r linearly independent vectors of the system, and this representation is unique.

Proof. Assume that r out of m vectors are linearly independent; call them $x^1, x^2, \ldots, x^r$. Any vector x^k from the remaining $(m - r)$ vectors is linearly dependent on $x^1, \ldots, x^r$ by definition of the rank; and we obtain

$$c_kx^k + c_1x^1 + \cdots + c_rx^r = 0$$

i.e.

$$\begin{aligned} x^k &= -\frac{1}{c_k}(c_1x^1 + \cdots + c_rx^r), \quad c_k \neq 0 \\ &= \alpha_1x^1 + \cdots + \alpha_rx^r \end{aligned}$$

Now for the second part of the theorem, assume that x^k takes a different representation in $x^1, \ldots, x^r$ such as the following

$$x^k = \tilde{\alpha}_1x^1 + \cdots + \tilde{\alpha}_rx^r$$

Subtracting the last two equations we obtain

$$0 = (\alpha_1 - \tilde{\alpha}_1)x^1 + \cdots + (\alpha_r - \tilde{\alpha}_r)x^r$$

But $x^1, \ldots, x^r$ are linearly independent; hence we obtain

$$\alpha_1 = \tilde{\alpha}_1, \ldots, \alpha_r = \tilde{\alpha}_r,$$

and the proof is complete.

6. If a vector that can be written as a linear combination of the other vectors in the system is removed from the given vector system, the rank of the system remains unchanged.

Proof. Let the vector system be $x^1, \ldots, x^r, x^{r+1}, \ldots, x^m$ of rank r, and whose first r vectors are linearly independent. Let us express x^r as a linear combination of the rest, i.e.

$$x^r = c_1 x^1 + \cdots + c_{r-1} x^{r-1} + c_{r+1} x^{r+1} + \cdots + c_m x^m$$

The validity of the above equation stems from the fact that all vectors are linearly dependent on $x^1, \ldots, x^r$. Now assume that the rank of the system without x^r is $r - 1$. According to the last theorem all the x^i (for $i = r + 1, \ldots, m$) can be written as a linear combination of the x^j (for $j = 1, \ldots, r - 1$), i.e.

$$x^i = \alpha_{i,1} x^1 + \cdots + \alpha_{i,r-1} x^{r-1}, \quad i = r + 1, \ldots, m$$

And consequently

$$x^r = (c_1 + c_{r+1}\alpha_{r+1,1} + \cdots + c_m \alpha_{m,1})\, x^1 + \cdots + \\ + (c_{r-1} + c_{r+1}\alpha_{r+1,r-1} + \cdots + c_m \alpha_{m,r-1})\, x^{r-1}$$

This shows that x^r is linearly dependent on $x^1, \ldots, x^{r-1}$, which contradicts the assumption that the first r vectors are linearly independent. Hence, even in this case the rank is preserved and the proof is complete.

7. If a vector system is changed by adding a vector which can be represented as a linear combination of vectors already in the system, the rank of the system remains unchanged.

Proof. Given a vector system $x^1, x^2, \ldots, x^m$, let y be a vector which can be represented as a linear combination of these vectors. Consider the vector system $x^1, x^2, \ldots, x^m, y$. According to the last theorem, removing y will not alter the rank, hence both systems have the same rank and the theorem is proved.

8. If the vectors of the system $x^1, x^2, \ldots, x^k$ can be represented as a linear combination of the vectors $y^1, y^2, \ldots, y^p$ in the system, the rank of the system $x^1, x^2, \ldots, x^k$ is at most equal to p.

Proof. Consider the system

$$x^1, x^2, \ldots, x^k, y^1, y^2, \ldots, y^p.$$

According to the last theorem, the rank of the above system is equal to that of the system $y^1, \ldots, y^p$. As the rank of this set is at most equal to p, the result follows immediately.

3.2. Dimension and basis

If in a vector space V there is a maximum number of linearly independent vectors, this number is called the *dimension* of the space. If m is such a number, then any vector system consisting of m linearly independent vectors in the space is called a *basis*. The vectors in a basis are called base vectors.

A fundamental property of a basis is that any vector in the vector space can be represented as a linear combination of the base vectors, and this representation is unique.

To prove this proposition, let $x^1, \ldots, x^m$ be a basis of the space of dimension m. Let y be any other vector in the same vector space. The vector system $x^1, x^2, \ldots, x^m$, y has a rank equal to m, since there are no more than m linearly independent vectors in the vector space, also since $x^1, x^2, \ldots, x^m$ is of rank m, then according to Theorem 5 in Section 3.1, y can be represented uniquely in terms of the basis $x^1, x^2, \ldots, x^m$, i.e.

$$y = \sum_{i=1}^{m} c_i x^i$$

Example. Let

$$x^1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

be a basis in $\mathbb{R}^2$. Any vector y, for example

$$y = \begin{bmatrix} 5 \\ 8 \end{bmatrix},$$

can be written as

$$\begin{bmatrix} 5 \\ 8 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

giving $c_1 = 1, c_2 = 2$.

3.3. Orthogonality and biorthogonality of vectors

A set of vectors $x^1, x^2, \ldots, x^m$ are called *orthogonal* to each other if

$$\langle x^j, x^i \rangle = 0, \quad \text{for all } i \neq j.$$

Example. The set of vectors

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

are orthogonal in $\mathbb{R}^3$.
The definition of orthogonality enables us to deduce the following theorem.

THEOREM 1. Non-zero orthogonal vectors are linearly independent.

Proof. Let $x^1, x^2, \ldots, x^m$ be such vectors. Form the linear relation

$$c_1 x^1 + c_2 x^2 + \cdots + c_m x^m = 0,$$

and proceed to prove that

$$c_1 = c_2 = \cdots = c_m = 0.$$

Taking the inner product with x^1, we obtain

$$c_1\langle x^1, x^1\rangle = 0$$

But

$$\langle x^1, x^1\rangle = \|x^1\|_2^2 > 0$$

as $x^1 \neq 0$; hence

$$c_1 = 0$$

Similarly we can show that

$$c_2 = c_3 = \cdots = c_m = 0$$

and the proof is complete.

The concept of orthogonality of vectors facilitates the expansion of any vector y in terms of orthogonal vectors. For let $x^1, x^2, \ldots, x^n$ be a set of non-zero orthogonal vectors in a space of dimension n, and y be any vector in the same space that we desire to expand in the form

$$y = \sum_{i=1}^{n} c_i x^i$$

Taking the inner product with x^j, we obtain

$$c_j = \frac{\langle x^j, y\rangle}{\langle x^j, x^j\rangle}$$

If in addition we have

$$\langle x^j, x^j\rangle = 1,$$

the vector system $x^1, \ldots, x^n$ is called an *orthonormal system* and the vectors $x^1, \ldots, x^n$ are called an *orthonormal basis*.

Example. Let

$$x^1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}, \quad x^3 = \begin{bmatrix} -2 \\ 1 \\ -5 \end{bmatrix}, \quad y = \begin{bmatrix} -3 \\ 0 \\ 4 \end{bmatrix}$$

Obtain c_1, c_2 and c_3 such that $y = c_1x^1 + c_2x^2 + c_3x^3$

Using the above technique we get:

$$c_1 = \frac{\langle y, x^1\rangle}{\langle x^1, x^1\rangle} = \frac{-3}{5}$$

$$c_2 = \frac{\langle y, x^2 \rangle}{\langle x^2, x^2 \rangle} = \frac{-10}{6}$$

$$c_3 = \frac{\langle y, x^3 \rangle}{\langle x^3, x^3 \rangle} = \frac{-14}{30}$$

Example. Transform the set of orthogonal vectors

$$x^1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}, \quad x^3 = \begin{bmatrix} -2 \\ 1 \\ -5 \end{bmatrix}$$

into an orthonormal basis.

Dividing by the norm of each we obtain the orthonormal basis as follows:

$$x_n^1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}, \quad x_n^2 = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix}, \quad x_n^3 = \begin{bmatrix} \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \\ \frac{-5}{\sqrt{30}} \end{bmatrix}$$

A very simple example of an orthonormal basis in the space $\mathbb{R}^n$ is the n-axis $e^1, \ldots, e^n$ where

$$e^1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e^2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, e^n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

If $x^1, \ldots, x^n$ are not orthogonal, but only independent, then to solve for the coefficients $c_1, c_2, \ldots, c_n$, the equations

$$y = \sum_{i=1}^{n} c_i x^i$$

is to solve n linear simultaneous equations in n unknowns. Until this is studied in later sections of this chapter, we use the concept of biorthogonality, which enables us to obtain $c_1, c_2, \ldots, c_n$ if the biorthogonal basis of $x^1, x^2, \ldots, x^n$ is known. We start by making the following definition.

DEFINITION. Two sets of vectors $x^1, x^2, \ldots, x^n$ and $y^1, y^2, \ldots, y^n$ are called *biorthogonal* if

$$\begin{aligned} \langle x^j, y^i \rangle &= 0, \quad \text{for } i \neq j \\ &\neq 0, \quad \text{for } i = j \end{aligned}$$

In the literature, we find some authors using the term *reciprocal* instead of *biorthogonal*. The concept of biorthogonality enables us to establish the following theorem.

THEOREM 2. If the two sets of vectors $x^1, x^2, \ldots, x^n$ and $y^1, y^2, \ldots, y^n$ are biorthogonal, the vectors of each set are linearly independent.

Proof. We prove here that $x^1, x^2, \ldots, x^n$ are linearly independent. To prove that $y^1, y^2, \ldots, y^n$ are linearly independent is similar. Form the relation

$$c_1 x^1 + c_2 x^2 + \cdots + c_n x^n = 0$$

and proceed to prove that

$$c_1 = c_2 = \cdots = c_n = 0$$

Taking the inner product with y^1, we obtain

$$c_1\langle y^1, x^1\rangle + c_2\langle y^1, x^2\rangle + \cdots + c_n\langle y^1, x^n\rangle = 0$$

All the terms from the second term up to the last are zero according to the biorthogonality condition. Hence

$$c_1\langle y^1, x^1\rangle = 0$$

However

$$\langle y^1, x^1\rangle \neq 0$$

according to the definition of biorthogonality. Hence

$$c_1 = 0$$

Similarly one can show that

$$c_2 = \cdots = c_n = 0$$

and the proof is complete.

Example

$$x^1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \quad y^1 = \begin{bmatrix} -4 \\ -3 \end{bmatrix}, \quad y^2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

The reader can check that the two sets x^1, x^2 and y^1, y^2 are biorthogonal.

The concept of biorthogonality facilitates the expansion of any vector v in terms of any set of linearly independent vectors $x^1, \ldots, x^n$ lying in a space of dimension n if their reciprocal vectors $y^1, \ldots, y^n$ are known. For let

$$v = \sum_{i=1}^{n} c_i x^i$$

To obtain c_j we take the inner product of the above equation with y^j giving

$$c_j = \frac{\langle y^j, v \rangle}{\langle y^j, x^j \rangle}$$

Example. Let

$$x^1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad x^3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$y^1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad y^2 = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}, \quad y^3 = \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix}$$

Then if

$$v = c_1 x^1 + c_2 x^2 + c_3 x^3$$

we have

$$c_1 = \frac{\langle y^1, v \rangle}{\langle y^1, x^1 \rangle} = \frac{-3}{4}$$

$$c_2 = \frac{\langle y^2, v \rangle}{\langle y^2, x^2 \rangle} = \frac{7}{4}$$

$$c_3 = \frac{\langle y^3, v \rangle}{\langle y^3, x^3 \rangle} = \frac{-5}{4}$$

3.4. The Grammian

Until now we have not given a method for testing the linear independence of vectors, except by relying upon the definition of independence itself. We have set out the linear relation

$$c_1 x^1 + c_2 x^2 + \cdots + c_m x^m = 0$$

and then proceeded to show that

$$c_1 = c_2 = \cdots = c_m = 0$$

Although the method seems simple, to show that the coefficients are all zero is a very tedious exercise. The method to be explained in this section gives a straightforward answer to whether the vectors are linearly independent or not.

We define the *Grammian matrix* or *Grammian* for a set of vectors $x^1, x^2, \ldots, x^m$ as follows:

$$\underset{m,m}{G} = \begin{bmatrix} \langle x^1, x^1 \rangle & \langle x^1, x^2 \rangle & \cdots & \langle x^1, x^m \rangle \\ \langle x^2, x^1 \rangle & \langle x^2, x^2 \rangle & \cdots & \\ \vdots & & & \\ \langle x^m, x^1 \rangle & & \cdots & \langle x^m, x^m \rangle \end{bmatrix}$$

Therefore if A is the matrix whose columns are $x^1, x^2, \ldots, x^m$, it appears directly that

$$G = A^*A$$

from which we conclude that the Grammian G is Hermitian and consequently its determinant is real. Moreover, the determinant of G can be shown to be greater than or equal to zero, and the reader should do this as an exercise.

Now we proceed to give the test of linear independence for a set of vectors $x^1, x^2, \ldots, x^m$ as a result of the following theorem.

THEOREM 1. If the determinant of the Grammian of a set of vectors $x^1, x^2, \ldots, x^m$ is greater than zero, the vectors are linearly independent.

Proof. Form the linear relation

$$c_1x^1 + c_2x^2 + \cdots + c_mx^m = 0,$$

and proceed to prove that

$$c_1 = c_2 = \cdots = c_m = 0.$$

Taking the inner product of the above equation first with x^1, then x^2, until x^m, we obtain

$$\begin{bmatrix} \langle x^1, x^1 \rangle & \langle x^1, x^2 \rangle & \cdots & \langle x^1, x^m \rangle \\ \langle x^2, x^1 \rangle & \langle x^2, x^2 \rangle & \cdots & \langle x^2, x^m \rangle \\ \vdots & & & \\ \langle x^m, x^1 \rangle & & \cdots & \langle x^m, x^m \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = 0$$

Hence if $|G| > 0$, then according to Exercise 2.1.14, we obtain

$$c_1 = c_2 = \cdots = c_m = 0$$

and the proof is complete.

THEOREM 2. A necessary and sufficient condition that $x^1, x^2, \ldots, x^m$ are linearly dependent is that their Grammian determinant is zero.

Proof. Necessity can be established from Exercise 2.1.14. To prove sufficiency we notice from Exercise 2.1.14 that if the Grammian determinant is zero we have two possibilities: either the coefficients are zero which is trivial, or the coefficients may not be zero, i.e. there exists a non-trivial solution for the coefficients, which completes the proof.

A special case arises when $x^1, \ldots, x^m$ lie in the space $\mathbb{R}^m$ or $\mathbb{C}^m$. In this case the matrix A, whose columns are the vectors $x^1, \ldots, x^m$, will be square, and we get the following results which are special cases of Theorems 1 and 2 above.

1. If $|A| \neq 0$, then $x^1, \ldots, x^m$ are linearly independent
2. A necessary and sufficient condition that $x^1, \ldots, x^m$ are linearly dependent is that $|A| = 0$

The proof of the above results is direct, and so it is left as an exercise for the reader.

One important result of the Grammian is that if $x^1, \ldots, x^m$ lie in the space $\mathbb{R}^n$ or $\mathbb{C}^n$ with $m > n$, then $x^1, \ldots, x^m$ are linearly dependent. In other words, one cannot find in the Euclidean space of dimension n more than n linearly independent vectors. See Exercise 2.1.18.

3.5. The Gram–Schmidt process

Sometimes it becomes useful to obtain a set of orthogonal vectors from a set of independent vectors. One such use is met when dealing with the eigenvalue problem of Hermitian matrices as will be seen in the next chapter. The method of generating n orthogonal vectors from n linearly independent vectors is called the Gram–Schmidt process and is as follows.

Let $x^1, x^2, \ldots, x^n$ be a set of linearly independent vectors and it is required to obtain from them a set of orthogonal vectors $y^1, y^2, \ldots, y^n$. We choose the last vectors as follows:

$$y^1 = x^1$$
$$y^2 = x^2 + \alpha x^1$$

and proceed to find α such that y^2 and y^1 are mutually orthogonal. Taking the inner product with y^1, we obtain

$$\langle y^1, y^2 \rangle = \langle x^1, y^2 \rangle = 0 = \langle x^1, x^2 \rangle + \alpha \langle x^1, x^1 \rangle$$

Hence

$$\alpha = -\frac{\langle x^1, x^2 \rangle}{\langle x^1, x^1 \rangle}$$

Choose y^3 in the following manner:

$$y^3 = x^3 + c_2 x^2 + c_1 x^1$$

and proceed to find c_1 and c_2 such that y^3, y^2 and y^1 are mutually orthogonal. Taking the inner product with y^1, we obtain

$$\langle y^1, y^3 \rangle = \langle x^1, y^3 \rangle = 0 = \langle x^1, x^3 \rangle + c_2 \langle x^1, x^2 \rangle + c_1 \langle x^1, x^1 \rangle$$

Now if y^3 is made orthogonal on x^2, then y^3 will consequently be orthogonal on

y^2. Hence taking the inner product with x^2 gives

$$\langle x^2, y^3 \rangle = 0 = \langle x^2, x^3 \rangle + c_2 \langle x^2, x^2 \rangle + c_1 \langle x^2, x^1 \rangle$$

The above two equations can be put in the following form:

$$\begin{bmatrix} \langle x^1, x^1 \rangle & \langle x^1, x^2 \rangle \\ \langle x^2, x^1 \rangle & \langle x^2, x^2 \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = - \begin{bmatrix} \langle x^1, x^3 \rangle \\ \langle x^2, x^3 \rangle \end{bmatrix}$$

The matrix on the left-hand side has a non-zero determinant, for it is the Grammian of the two linearly independent vectors x^1 and x^2. Therefore the coefficients c_1 and c_2 can be calculated uniquely.

The process can be prolonged similarly on y^4 if it is defined as

$$y^4 = x^4 + c_1 x^1 + c_2 x^2 + c_3 x^3$$

and so on until we take y^n to be

$$y^n = x^n + c_1 x^1 + c_2 x^2 + \cdots + c_{n-1} x^{n-1}$$

The coefficients $c_1, c_2, \ldots, c_{n-1}$ are obtained by solving the linear equations

$$\begin{bmatrix} \langle x^1, x^1 \rangle & \cdots & \langle x^1, x^{n-1} \rangle \\ \langle x^2, x^1 \rangle & \cdots & \\ \vdots & & \\ \langle x^{n-1}, x^1 \rangle & \cdots & \langle x^{n-1}, x^{n-1} \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix} = - \begin{bmatrix} \langle x^1, x^n \rangle \\ \\ \vdots \\ \langle x^{n-1}, x^n \rangle \end{bmatrix}$$

The matrix of the left-hand side is nonsingular for it is the Grammian of the set of linearly independent vectors $x^1, \ldots, x^{n-1}$.

Exercises 3.5

1. If $x^1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, $x^2 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, $y = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$, find c_1 and c_2 such that $y = c_1 x^1 + c_2 x^2$.
2. Determine the dimension of the vector space of the set of vectors

   $$x^T = [1 \quad 3 \quad 0], \quad y^T = [2 \quad -5 \quad 0], \quad z^T = [-1 \quad 7 \quad 0].$$

3. Explain why we cannot write $y = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ in terms of the two non-zero orthogonal vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}$.
4. Show that $\sum_{i=1}^{n} \langle x^i, y \rangle x^i = y$, if $\langle x^i, x^j \rangle = \delta_{ij}$, with $\delta_{ij} = 1 (i = j)$, $\delta_{ij} = 0 (i \neq j)$.
5. If $\langle x^i, x^j \rangle = 1$, $(i = j)$, and $\langle x^i, x^j \rangle = 0$, $(i \neq j)$, show that

   $$\langle y, y \rangle = \sum_{i=1}^{n} |c_i|^2, \quad \text{if } y = \sum_{i=1}^{n} c_i x^i$$

6. If y is orthogonal on $x^1, \ldots, x^m$, show that y is orthogonal on any vector which is a linear combination of $x^1, \ldots, x^m$.
7. If $\langle x^i, x^j \rangle = 1$, $(i = j)$, and $\langle x^i, x^j \rangle = 0$ $(i \neq j)$ show that the matrix $A_{n,n}$ whose columns are $x^1, \ldots, x^n$ is unitary.
8. Expand $y = \begin{bmatrix} i \\ 3 \end{bmatrix}$, $(i^2 = -1)$, in terms of $x^1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $x^2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

9. Determine the rank of the vectors $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 7 \end{bmatrix}$.

10. Determine the rank of the vectors $\begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}, \begin{bmatrix} 2+i \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4+3i \\ -5 \\ 2 \end{bmatrix}$ $(i^2 = -1)$.

11. Under what condition will the rank of the vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ r-2 \\ 2 \end{bmatrix} \begin{bmatrix} 0 \\ k-1 \\ k+2 \end{bmatrix}$ $\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$ be less than three?

12. Prove that A^*A is diagonal if and only if the columns of A are mutually orthogonal, and that $AA^* = I$ if and only if the rows of A are mutually orthogonal unit vectors (A need not be square).
13. Show that, if x, y and z are mutually orthogonal vectors,

$$\| x + y + z \|_2 = (\| x \|_2^2 + \| y \|_2^2 + \| z \|_2^2)^{1/2}.$$

14. If $x^1, \ldots, x^m$ are linearly independent, show that $\bar{x}^1, \ldots, \bar{x}^m$ are also linearly independent.
15. Let L be the linear space of finite trigonometric sums

$$x = c_1 \sin t + c_2 \sin 2t + \cdots + c_k \sin kt \quad (0 \leqslant t \leqslant 2\Pi)$$

where $c_1, c_2, \ldots$ are real coefficients. Show that the vectors $\sin t, \ldots, \sin kt$ constitute a basis for L. Hence show how to obtain $c_1, c_2, \ldots$.
16. If $\langle x^i, y^j \rangle = 1$, $(i = j)$, and $\langle x^i, y^j \rangle = 0$, $(i \neq j)$, show that

$\langle v, v \rangle = \sum_{i=1}^{n} \alpha_i \bar{c}_i = \sum_{i=1}^{n} \bar{\alpha}_i c_i$, if $v = \sum_{i=1}^{n} c_i x^i$ and $v = \sum_{i=1}^{n} \alpha_i y^i$, which gives a relation between c_i and α_i; hence show that $\sum_{i=1}^{n} \alpha_i \bar{c}_i$ and $\sum_{i=1}^{n} \bar{\alpha}_i c_i$ are real, for both are equal to $\langle v, v \rangle$, which is real. Finally show that $\sum_{i=1}^{n} \bar{\alpha}_i c_i = \langle \alpha, G_y \alpha \rangle$, where $\alpha^T = [\alpha_1 \alpha_2 \ldots \alpha_n]$ and G_y is the Grammian matrix for the vectors y^i.
17. If $x^1, x^2, \ldots, x^m$ are linearly independent as well as $y^1, y^2, \ldots, y^s$, and $\langle x^i, y^j \rangle = 0$ for all i, j, show that $x^1, \ldots, x^m, y^1, \ldots, y^s$ are linearly independent.
18. If $x^1, x^2, \ldots, x^n \in \mathbb{R}^n$ are linearly independent and are all orthogonal on a vector v, show that $v = 0$.

19. If $u^1, u^2, \ldots, u^m$ are a set of linearly independent vectors, show that the vectors $v^1, v^2, \ldots, v^m$ where $v^i = \sum_k \alpha_k^i u^k$, with α_k^i arbitrary and not zero, are also a set of linearly independent vectors. Show also that the set $v^1, v^2, \ldots, v^m, v^{m+1}$ are linearly dependent.

3.6. Rank of a matrix

Let $A_{m,n}$ be a matrix which contains m rows and n columns. Assume without loss of generality that $m \leqslant n$. A can be seen as consisting of m vectors in an n-dimensional Euclidean space, the rank of which will be denoted by ρ_r (row rank). Obviously

$$\rho_r \leqslant m$$

Also A can be seen as consisting of n vectors in an m-dimensional Euclidean space, the rank of which will be denoted by ρ_c (column rank). Obviously

$$\rho_c \leqslant m$$

The following theorem establishes the relation between the row rank and the column rank of a matrix.

THEOREM 1. In a matrix A, the number of linearly independent row vectors is equal to the number of linearly independent column vectors. In other words row rank ρ_r = column rank ρ_c = rank of A denoted by ρ.

Proof. Let the number of linearly independent column vectors in $A_{m,n}$ be ρ_c and the number of linearly independent row vectors be ρ_r. Select ρ_c linearly independent columns of A and put them in a matrix B_1 whose dimension is (m, ρ_c). As a result of Theorem 5 of Section 3.1 every column of A can be written as a linear combination of these independent vectors. If $x^1, x^2, \ldots, x^n$ are the columns of A, then

$$x^1 = B_1 v^1$$
$$x^2 = B_1 v^2$$
$$\vdots$$
$$x^n = B_1 v^n$$

where the components of v^i are the coordinates of x^i w.r.t. the basis consisting of the column vectors of B_1. Hence

$$A = [B_1 v^1 \; B_1 v^2 \ldots B_1 v^n] = B_1 [v^1 v^2 \ldots v^n] = B_1 B_2.$$

This means that the row vectors of A are linear combinations of the row vectors of B_2. It follows that the rank of the system of row vectors of A cannot exceed the number of row vectors in B_2. This means that

$$\rho_r \leqslant \rho_c.$$

If we consider the transpose of A we obtain

$$\rho_c \leqslant \rho_r,$$

from which we conclude that

$$\rho_r = \rho_c$$

Example

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

The matrix A has two columns in $\mathbb{R}^4$; hence $\rho_c \leqslant 2$. One can use the test of independence to show that $\rho_c = 2$. A also has four vectors in $\mathbb{R}^2$, hence $\rho_r \leqslant 2$. Similarly it can be shown that $\rho_r = 2$. Hence $\rho_r = \rho_c = 2$. In this example the rank is obtained easily because it is equal to the number of columns, and so calculating it is equivalent to making sure that the two columns are linearly independent, in other words by using the test of independence. However in some examples we find that the rank is less than both the number of rows and the number of columns; then we need methods for calculating the rank.

THEOREM 2. Let A be a matrix of order (m, n). Suppose that A has a sub-matrix S of order (r, r) with $|S| \neq 0$. And suppose that every sub-matrix T of order $(r + 1, r + 1)$ of which S is a sub-matrix has $|T| = 0$. Then $\rho(A) = r$.

Before proceeding with the proof, let us explain the theorem by an example. Let

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

The submatrix

$$S = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

has $|S| \neq 0$, but every (3, 3) sub-matrix T including S has $|T| = 0$, e.g.

$$\begin{vmatrix} 1 & 1 & 2 \\ 2 & 2 & 4 \\ 1 & 2 & 4 \end{vmatrix} = 0, \quad \begin{vmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \\ 1 & 2 & -1 \end{vmatrix} = 0, \quad \begin{vmatrix} 0 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 1 & 2 \end{vmatrix} = 0.$$

Hence from the theorem we conclude that $\rho(A) = 2$. Now we proceed with the proof.

Proof. Let S be composed of the linearly independent row vectors $x^1, \ldots, x^r$ in the following manner:

$$S = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^r \end{bmatrix}$$

Construct a sub-matrix S' as follows

$$S' = [S \mid v] = \begin{bmatrix} x'^1 \\ \vdots \\ x'^r \end{bmatrix}$$

where v is any column vector. Now S' has rank $= r$, as the reader can verify. And as T can be written without loss of generality in the following form

$$T = \begin{bmatrix} S' \\ \hdashline u \end{bmatrix};$$

then $|T| = 0$ implies that the row vectors of T are linearly dependent, i.e.

$$c_1 x'^1 + c_2 x'^2 + \cdots + c_r x'^r + c_{r+1} u = 0$$

where

$$c_{r+1} \neq 0;$$

otherwise all the other coefficients c_k will be zero since $x'^1, x'^2, \ldots, x'^r$ are linearly independent, hence

$$u = \alpha_1 x'^1 + \alpha_2 x'^2 + \cdots + \alpha_r x'^r$$

where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are scalars.

Applying the same procedure to all sub-matrices of order $(r + 1, r + 1)$ in which S is a submatrix; and use Theorem 7 in Section 3.1 to complete the proof.

One direct application of this theorem is when A is Hermitian of rank r; then at least one principal minor of order r is not zero. The proof of this corollary is left as an exercise for the reader.

3.7. Elementary row and column operations

There are sets of elementary row and column operations which can be executed on any matrix A to reduce it to an echelon form. These operations can be achieved by pre-multiplying A by a set of matrices (R_1, R_2, R_3) for rows or by post-multiplying A by a set of matrices (P_1, P_2, P_3) for columns. The elementary row operations are of three types:

R_1, responsible for interchanging any two rows

R_2, responsible for multiplying any row by a non-zero scalar
R_3, responsible for adding to any row any other row multiplied by a non-zero scalar.

R_1, R_2 and R_3 are all generated from the unit matrix I, by applying on the latter the same change which we require for the matrix A. For example if we want to interchange the first and the third rows of A, we multiply A by R_1, which is a unit matrix whose first and third rows are interchanged.

Example. Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ -2 & 1 & 0 & 4 \\ 5 & -2 & 3 & 1 \end{bmatrix}, \quad A' = \begin{bmatrix} 5 & -2 & 3 & 1 \\ -2 & 1 & 0 & 4 \\ 1 & 2 & 3 & 0 \end{bmatrix}$$

Hence

$$R_1 A = A' \text{ where } R_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The reader is asked to prove that elementary row operations do not change the rank of the matrix on which they operate; that R_1, R_2 and R_3 are nonsingular matrices; and that $|R_1| = -1$, $|R_3| = 1$. What is the value of $|R_2|$?

THEOREM 1. Any matrix can be reduced to an echelon form by a series of elementary row operations.

The proof is by construction and is best illustrated on an example. Let

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

Step 1: divide the first row by 5, giving

$$A_1 = \begin{bmatrix} 1 & 2/5 & 3/5 & -4/5 \\ 2 & 1 & 0 & 2 \\ -3 & -1 & -3 & 6 \end{bmatrix} \quad \text{with } R_2^1 = \begin{bmatrix} 1/5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Step 2: multiply the first row by 2 and subtract it from the second row, then multiply it by 3 and add it to the third row, giving

$$A_2 = \begin{bmatrix} 1 & 2/5 & 3/5 & -4/5 \\ 0 & 1/5 & -6/5 & 18/5 \\ 0 & 1/5 & -6/5 & 18/5 \end{bmatrix} \quad \text{with } R_3^1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

Step 3: multiply the second row by 5, giving

$$A_3 = \begin{bmatrix} 1 & 2/5 & 3/5 & -4/5 \\ 0 & 1 & -6 & 18 \\ 0 & 1/5 & -6/5 & 18/5 \end{bmatrix} \quad \text{with } R_2^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$