CHAPTER 2

Linear Algebra and Matrix Theory

1. Solving a System of Linear Equations with Numerical Coefficients by the Method of Elimination.

In this section we shall solve some systems of linear equations by the method of elimination (also called the Gauss procedure). The student has undoubtedly already learned this method in high school. Here, however, we will systematize the procedure. The purpose of this systematization is twofold. (1) When the problems are to be programmed for the digital computer, the procedure must be laid out systematically and (2) (the more important purpose) by systematizing we can construct a general theory of linear equations; indeed, the development of the theory in the latter part of this chapter depends essentially on the elimination procedures described in this section. Therefore, the student should go through this section carefully and follow through the various "tableaus," even though the systematic solution may not be the quickest in an individual case.

As to the number of solutions of a given system of linear equations, there are just three possibilities:

(1) There is precisely one solution (called the unique solution).

(2) There is no solution (the equations are inconsistent).

(3) There is more than one solution. (In this case, as we shall see, there are infinitely many solutions. The solutions are expressed in terms of "free parameters" that can be assigned arbitrarily.)

We shall now work through examples illustrating each of these situations.

Example 1.1. Solve the system

$$3x_1 + 6x_2 + 3x_3 = 21, \tag{1.1}$$

$$4x_1 + 2x_2 - 5x_3 = -20, \tag{1.2}$$

$$-2x_1 + x_2 + 2x_3 = 12. \tag{1.3}$$

Our first step is to multiply each side of equation (1.1) by 1/3. The purpose of this is to get a coefficient of 1 for x_1; this serves to standarize the subsequent steps. Thus the system becomes

$$x_1 + 2x_2 + x_3 = 7, \tag{1.1a}$$

$$4x_1 + 2x_2 - 5x_3 = -20, \tag{1.2a}$$

$$-2x_1 + x_2 + 2x_3 = 12. \tag{1.3a}$$

Now we add (-4) times each side of equation (1.1a) to the corresponding side of equation (1.2a) and (2) times each side of equation (1.1a) to the corresponding side of equation (1.3a). The effect of this is to eliminate x_1 from the last two equations:

$$x_1 + 2x_2 + x_3 = 7, \tag{1.1b}$$

$$- 6x_2 - 9x_3 = -48, \tag{1.2b}$$

$$5x_2 + 4x_3 = 26. \tag{1.3b}$$

Now we regard the last two equations as two equations in two unknowns and apply the analogous procedure; namely, we multiply each side of equation (1.2b) by (-1/6) to obtain the system

$$x_1 + 2x_2 + x_3 = 7, \tag{1.1c}$$

$$x_2 + \frac{3}{2}x_3 = 8, \tag{1.2c}$$

$$5x_2 + 4x_3 = 26. \tag{1.3c}$$

Next we add (-5) times each side of equation (1.2c) to the corresponding side of equation (1.3c), thus eliminating x_2 from the last equation:

$$\begin{aligned}
&(1.1d) \qquad & x_1 + 2x_2 + x_3 &= 7, \\
&(1.2d) \qquad & x_2 + \tfrac{3}{2}x_3 &= 8, \\
&(1.3d) \qquad & -\tfrac{7}{2}x_3 &= -14.
\end{aligned}$$

We multiply each side of equation (1.3d) by $(-2/7)$ to put the system in the form

$$\begin{aligned}
&(1.1e) \qquad & x_1 + 2x_2 + x_3 &= 7, \\
&(1.2e) \qquad & x_2 + \tfrac{3}{2}x_3 &= 8, \\
&(1.3e) \qquad & x_3 &= 4.
\end{aligned}$$

The system is now in what is called "echelon form." Since (1.3e) is solved for x_3, we could use this value in (1.2e) to determine x_2 and, with x_2 and x_3 determined, solve (1.1e) for x_1. We shall follow a different procedure. The rule we have been following up to this point may be described as follows. We start with the entry in the "Northwest Corner" and wipe out the variables in the vertical column below it. Then we proceed to the "Northwest Corner" of the smaller system, and so on. Now we shall use the analogous procedure working from entry in the "Southeast Corner" and wiping out variables in the column vertically above it. Specifically we add $(-\frac{3}{2})$ of each side of equation (1.3e) to the corresponding side of equation (1.2e) and we add (-1) times each side of equation (1.3e) to the corresponding side of (1.1e) to eliminate the x_3 from the first two equations:

$$\begin{array}{ll} (1.1f) & x_1 + 2x_2 \qquad = 3, \\ (1.2f) & \qquad\quad x_2 \qquad = 2, \\ (1.3f) & \qquad\qquad\quad x_3 = 4. \end{array}$$

Now we add (-2) times each side of equation (1.2f) to equation (1.1f) to eliminate x_2 from the first equation:

$$\begin{array}{ll} (1.1g) & x_1 \qquad\qquad = -1, \\ (1.2g) & \qquad x_2 \qquad = 2, \\ (1.3g) & \qquad\qquad x_3 = 4. \end{array}$$

We have proved that <u>if there is a solution</u> to the system of equations (1.1), (1.2), (1.3), it is given by equations (1.1g), (1.2g), (1.3g). One can easily verify that these values indeed provide a solution by substituting them into the original system (1.1), (1.2), (1.3), or by observing that each step we took could be reversed, so that we could pass from (1.1g), (1.2g), (1.3g) to (1.1), (1.2), (1.3) in the reverse series of steps. Thus we see that in this example there is a unique solution.

The above computations can be put in more concise form by observing that the symbols for the variables themselves, x_1, x_2, x_3 played no part in the computation. They serve only to mark a column position and need not be carried along if we keep the individual columns separated. Thus the equations (1.1), (1.2), (1.3) -- (1.1g), (1.2g), (1.3g) and the instructions for obtaining them given in the text above, can be put into the compact form of Table 1.1a. (Tables 1.1b and 1.1c will be discussed later.)

Eqn. No. in Text	x_1	x_2	x_3	Constant Term				
(1.1)	3	6	3	21	12	1	0	0
(1.2)	4	2	-5	-20	25	0	1	0
(1.3)	-2	1	2	12	-12	0	0	1
Multiply row 1 by (1/3)								
(1.1a)	1	2	1	7	4	1/3	0	0
(1.2a)	4	2	-5	-20	25	0	1	0
(1.3a)	-2	1	2	12	-12	0	0	1
Add (-4) times row 1 to row 2 Add (2) times row 1 to row 3								
(1.1b)	1	2	1	7	4	1/3	0	0
(1.2b)	0	-6	-9	-48	9	-4/3	1	0
(1.3b)	0	5	4	26	-4	2/3	0	1
Multiply row 2 by (-1/6)								
(1.1c)	1	2	1	7	4	1/3	0	0
(1.2c)	0	1	3/2	8	-3/2	2/9	-1/6	0
(1.3c)	0	5	4	26	-4	2/3	0	1
Add (-5) times row 2 to row 3								
(1.1d)	1	2	1	7	4	1/3	0	0
(1.2d)	0	1	3/2	8	-3/2	2/9	-1/6	0
(1.3d)	0	0	-7/2	-14	7/2	-4/9	5/6	1
Multiply row 3 by (-2/7)								
(1.1e)	1	2	1	7	4	1/3	0	0
(1.2e)	0	1	3/2	8	-3/2	2/9	-1/6	0
(1.3e)	0	0	1	4	-1	8/63	-5/21	-2/7
Add (-3/2) times row 3 to row 2 Add (-1) times row 3 to row 1								
(1.1f)	1	2	0	3	5	13/63	5/21	2/7
(1.2f)	0	1	0	2	0	2/63	4/21	3/7
(1.3f)	0	0	1	4	-1	8/63	-5/21	-2/7
Add (-2) times row 2 to row 1								
(1.1g)	1	0	0	-1	5	1/7	-1/7	-4/7
(1.2g)	0	1	0	2	0	2/63	4/21	3/7
(1.3g)	0	0	1	4	-1	8/63	-5/21	-2/7
Table 1.1a					Table 1.1b	Table 1.1c		

We shall now show a useful aspect of the Table 1.1a. Suppose we wish to solve the system:

$$(1.4) \qquad 3x_1 + 6x_2 + 3x_3 = 12,$$

$$(1.5) \qquad 4x_1 + 2x_2 - 5x_3 = 25,$$

$$(1.6) \qquad -2x_1 + x_2 + 2x_3 = -12.$$

Note that the system of equations (1.4), (1.5), (1.6) differs from that of equations (1.1), (1.2), (1.3) only in the right hand sides. To solve (1.4), (1.5), (1.6) we do not have to go through the whole business again, since the reduction of the left hand sides to the final form will consist of exactly the same sequence of operations. Therefore, in Table 1.1b we enter in the first three lines the new right hand sides and perform the same sequence of operations on these numbers. Thus we obtain the solution to the system (1.4), (1.5), (1.6)

$$x_1 = 5, \; x_2 = 0, \; x_3 = -1,$$

as is seen in the last three entries of Table 1.1b.

In Table 1.1c we solve three more such problems, namely those for which the right hand sides of equations (1.1), (1.2), (1.3) are replaced by

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

From the last three lines of Table 1.1c we see that the solution to the system

$$(1.7) \qquad 3x_1 + 6x_2 + 3x_3 = 1,$$

$$(1.8) \qquad 4x_1 + 2x_2 - 5x_3 = 0,$$

$$(1.9) \qquad -2x_1 + x_2 + 2x_3 = 0,$$

is $x_1 = 1/7$, $x_2 = 2/63$, $x_3 = 8/63$. The solution to the system

(1.10) $$3x_1 + 6x_2 + 3x_3 = 0,$$

(1.11) $$4x_1 + 2x_2 - 5x_3 = 1,$$

(1.12) $$-2x_1 + x_2 + 2x_3 = 0,$$

is $x_1 = -1/7$, $x_2 = 4/21$, $x_3 = -5/21$. Finally, the solution to the system

(1.13) $$3x_1 + 6x_2 + 3x_3 = 0,$$

(1.14) $$4x_1 + 2x_2 - 5x_3 = 0,$$

(1.15) $$-2x_1 + x_2 + 2x_3 = 1,$$

is $x_1 = -4/7$, $x_2 = 3/7$, $x_3 = -2/7$.

We shall see in Section 11 how the solution for any given right hand sides can be obtained easily from the solutions of these last three systems which appear as the last three rows of Table 1.1c.

Example 1.1a. To illustrate a slight difficulty that may arise with the above procedure and to give the procedure for circumventing it, we consider the following problem.

Solve the system

(1.16) $$x_1 + 2x_2 + x_3 = 1,$$

(1.17) $$2x_1 + 4x_2 + x_3 = -1,$$

(1.18) $$3x_1 - x_2 + 4x_3 = 2.$$

Since the coefficient of x_1 in the equation (1.16) is already unity, we proceed to the next step, namely adding (-2) times each side of equation (1.16) to the corresponding side of equation (1.17); and (-3) times each side of equation (1.16) to the corresponding side of equation (1.18). Thus we obtain the system:

$$(1.16a)\qquad x_1 + 2x_2 + x_3 = 1,$$

$$(1.17a)\qquad 0x_2 - x_3 = -3,$$

$$(1.18a)\qquad -7x_2 + x_3 = -1.$$

In the normal procedure we would next multiply each side of equation (1.17a) by the reciprocal of the coefficient of x_2 in that equation. Since this coefficient is zero in this case, we cannot do this. We note instead that the coefficient of x_2 in a later equation (i.e. equation (1.18a)) is not zero and so we simply interchange equations (1.17a) and (1.18a) and proceed. [If the coefficient of x_2 were zero in <u>all</u> <u>subsequent</u> equations then we would leave the x_2 column alone and proceed to the x_3 column; see Example 1.3a below.] The complete solution shown in Table 1.2 is $x_1 = -22/7$, $x_2 = 4/7$, $x_3 = 3$.

<u>Example 1.2</u>. Solve the system:

$$(1.19)\qquad -x_1 + 2x_2 + 3x_3 = 1,$$

$$(1.20)\qquad 2x_1 + 5x_2 - 3x_3 = 2,$$

$$(1.21)\qquad 11x_1 + 14x_2 - 21x_3 = -1.$$

The same procedure is employed. The work is shown in Table 1.3a. The last line of this table shows that <u>if there is a solution</u> it must satisfy the equation

$$0x_1 + 0x_2 + 0x_3 = -6.$$

Cleary this equation has no solution, so we conclude that no solution to the system of equations (1.19), (1.20), (1.21) exists. That is, the equations are inconsistent. Inconsistency always shows up

x_1	x_2	x_3	
1	2	1	1
2	4	1	-1
3	-1	4	2
Add (-2) times row 1 to row 2 Add (-3) times row 1 to row 3			
1	2	1	1
0	0	-1	-3
0	-7	1	-1
Interchange rows 2 and 3			
1	2	1	1
0	-7	1	-1
0	0	-1	-3
Multiply row 2 by (-1/7)			
1	2	1	1
0	1	-1/7	1/7
0	0	-1	-3
Multiply row 3 by (-1)			
1	2	1	1
0	1	-1/7	1/7
0	0	1	3
Add (1/7) times row 3 to row 2 Add (-1) times row 3 to row 1			
1	2	0	-2
0	1	0	4/7
0	0	1	3
Add (-2) times row 2 to row 1			
1	0	0	-22/7
0	1	0	4/7
0	0	1	3

Table 1.2

in essentially this way; whenever a system of equations is inconsistent, we always arrive at an equation in which the coefficients are all zero, but the constant term is not. Conversely, as soon as we obtain an equation in which all the coefficients are zero but the constant term is not, we can without further work draw the conclusion that the system is inconsistent.

Example 1.3. Solve the system:

$$-x_1 + 2x_2 + 3x_3 = 1, \tag{1.22}$$

$$2x_1 + 5x_2 - 3x_3 = 2, \tag{1.23}$$

$$11x_1 + 14x_2 - 21x_3 = 5. \tag{1.24}$$

Since the left hand sides of (1.22), (1.23), (1.24) are the same as the left hand sides of (1.19), (1.20), (1.21) we do not need to go through the whole reduction again but merely change the right hand sides to agree with (1.22), (1.23), (1.24). This is shown in Table 1.3b. At the end of the Table 1.3b our equations are in the form:

$$x_1 - 2x_2 - 3x_3 = -1, \tag{1.22a}$$

$$x_2 + \frac{1}{3}x_3 = \frac{4}{9}, \tag{1.23a}$$

$$0x_3 = 0. \tag{1.24a}$$

Our next step in the standard procedure would be to divide by the coefficient of x_3 in the last equation. We cannot do this, since it is zero. However equation (1.24a) will be satisfied no matter what value is assigned to x_3. Therefore let x_3 be assigned any arbitrary value, say t. We cannot now wipe out the column above x_3 using the "Southeast Corner" rule. We can, however, back up to

x_1	x_2	x_3		
-1	2	3	1	1
2	5	-3	2	2
11	14	-21	-1	5
Multiply row 1 by (-1)				
1	-2	-3	-1	-1
2	5	-3	2	2
11	14	-21	-1	5
Add (-2) times row 1 to row 2 Add (-11) times row 1 to row 3				
1	-2	-3	-1	-1
0	9	3	4	4
0	36	12	10	16
Multiply row 2 by (1/9)				
1	-2	-3	-1	-1
0	1	1/3	4/9	4/9
0	36	12	10	16
Add (-36) times row 2 to row 3				
1	-2	-3	-1	-1
0	1	1/3	4/9	4/9
0	0	0	-6	0
Table 1.3a				Table 1.3b

x_2 and do it there. Specifically we add (2) times each side of equation (1.23a) to equation (1.22a), putting the system in the standard form

$$x_1 \qquad - \frac{7}{3}x_3 = -\frac{1}{9}, \tag{1.22b}$$

$$x_2 + \frac{1}{3}x_3 = \frac{4}{9}, \tag{1.23b}$$

$$0x_3 = 0. \tag{1.24b}$$

From this we can write the solution

$$x_1 = -\frac{1}{9} + \frac{7}{3}t, \tag{1.25}$$

$$x_2 = \frac{4}{9} - \frac{1}{3}t, \tag{1.26}$$

$$x_3 = t. \tag{1.27}$$

Example 1.3a. The solution of Example 1.3 contained one arbitrary parameter t. To illustrate a case in which two arbitrary parameters are involved, consider the following problem.

Solve the system:

$$x_1 + 2x_2 - x_3 + 4x_4 = 1, \tag{1.28}$$

$$3x_1 + 6x_2 + x_3 + 12x_4 = 3, \tag{1.29}$$

$$9x_1 + 18x_2 + x_3 + 36x_4 = 9. \tag{1.30}$$

The computations are shown in Table 1.4.

From the last three rows of Table 1.4 we see that (1.28), (1.29), (1.30) have been reduced to the form:

$$x_1 + 2x_2 \qquad + 4x_4 = 1, \tag{1.31}$$

$$x_3 \qquad = 0. \tag{1.32}$$

Consequently we can take x_2 and x_4 to be arbitrary, say equal to s and t respectively. Thus the solution is

x_1	x_2	x_3	x_4	
1	2	-1	4	1
3	6	1	12	3
9	18	1	36	9
Add (-3) times row 1 to row 2 Add (-9) times row 1 to row 3				
1	2	-1	4	1
0	0	4	0	0
0	0	10	0	0
Multiply row 2 by (1/4)				
1	2	-1	4	1
0	0	1	0	0
0	0	10	0	0
Add (-1/10) times row 2 to row 3				
1	2	-1	4	1
0	0	1	0	0
0	0	0	0	0
Add (1) times row 2 to row 1				
1	2	0	4	1
0	0	1	0	0
0	0	0	0	0

Table 1.4

(1.33) $$x_1 = 1 - 2s - 4t,$$

(1.34) $$x_2 = s,$$

(1.35) $$x_3 = 0,$$

(1.36) $$x_4 = t,$$

where s and t are arbitrary numbers.

Summary: The method of elimination demonstrated above is applicable regardless of whether the number of equations is greater than, less than, or equal to the number of unknown variables. Three basic operations are employed:

(i) Multiplying each side of an equation by a non-zero constant.

(ii) Adding a multiple of each side of one equation to the corresponding side of another equation.

(iii) Interchanging two equations.

Let us call these the Elementary Operations on Equations. By means of these operations we have seen that it is possible to reduce any system of linear equations to a form, called the echelon form, having the following properties:

(a) Reading from left to right along any particular equation the first non-zero coefficient which we encounter, if any, will be a 1.

(b) In any given equation in which not all the coefficients are zero the number of initial zeros (that is zeros before the first 1 is encountered) is at least one more than in the previous equation.

(c) The equations (if any) in which all the coefficients are zero follow the other equations.

A further simplification is possible to what we shall call the reduced echelon form, having the additional property:

(d) Any variable whose coefficient is the first non-zero one in some one of the equations (that coefficient therefore being 1, by property (a)) has coefficient zero in all other equations.

Figure 1.1 indicates a typical situation, where the symbols A_i and B_i denote quantities which might have any values in the echelon form. In the reduced echelon form the B_i would all be zero.

x_1	x_2	x_3	x_4	x_5	x_6		Right hand side
1	B_1	A_1	A_2	B_2	A_3	=	A_4
0	1	A_5	A_6	B_3	A_7	=	A_8
0	0	0	0	1	A_9	=	A_{10}
0	0	0	0	0	0	=	A_{11}
0	0	0	0	0	0	=	A_{12}

Figure 1.1

In the reduced echelon form the situation in regard to solvability is clear. There will be no solution unless the quantities A_{11} and A_{12} are both zero. If $A_{11} = A_{12} = 0$ then x_6 may be taken arbitrarily, $x_5 = A_{10} - A_9x_6$; x_3 and x_4 may be taken arbitrarily, and $x_2 = A_8 - A_7x_6 - A_6x_4 - A_5x_3$ and $x_1 = A_4 - A_3x_6 - A_2x_4 - A_1x_3$, i.e. in the rows with leading 1's we can solve for x_1, x_2, x_5 in terms of x_3, x_4, x_6 which are arbitrary. To put this another way

$$
\begin{aligned}
x_1 &= A_4 - A_3u - A_2v - A_1w, \\
x_2 &= A_8 - A_7u - A_6v - A_5w, \\
x_3 &= w, \\
x_4 &= v, \\
x_5 &= A_{10} - A_9u, \\
x_6 &= u,
\end{aligned}
$$

where u, v, and w are arbitrary.

Notice that if $A_{11} = A_{12} = 0$, so that a solution exists, then the number of arbitrary parameters P in the solution (3 in this case) is equal to V the number of variables (6 in this case) minus N the number of initial 1's (3 in this case). $P = V - N$. This result is true in general, as will be shown in Section 11.

A linear equation whose constant term is zero is said to be <u>homogeneous</u>. For the important special case of a system of homogeneous equations we have the following useful theorems.

<u>Theorem 1.1</u>. A system of homogeneous linear equations is never inconsistent, having always the solution $x_1 = x_2 = \dots = 0$.

The solution $x_1 = x_2 = \dots = 0$ is called the <u>trivial</u> solution.

<u>Theorem 1.2</u>. If $m < n$ a system of m homogeneous linear equations in n unknowns always has a non-trivial solution.

<u>Proof</u>. Since there are more columns than rows, in the echelon form there must be at least one column that does not contain any of the 1's which are the leading non-zero elements of the rows. The variable corresponding to that column can be chosen arbitrarily and hence not equal to zero.

We can summarize the results of this section as follows.

<u>Theorem 1.3</u>. A system of m linear equations in n variables might have

(a) One solution;

(b) No solution;

(c) An infinite number of solutions expressed in terms of one or more arbitrary parameters.

If the equations are homogeneous case (b) cannot arise; if $m < n$ case (a) cannot arise.

At the beginning of this section we remarked that in individual cases the systematic Gauss method of elimination is not necessarily the quickest method of solution. Once we get beyond the practice problems associated with this section the student is free to use any legitimate methods at his disposal to solve any systems of linear equations that he may encounter, unless, of course, the method of solution is specified.

Problems

(The reader is asked, in the following four problems, to solve a system of equations. However, he should not assume that the request to solve implies the existence of a solution. This warning applies, in fact, not only to these problems, and not only to this course, but to all his mathematical work.)

1.1 Solve the system:

$$\begin{aligned} x_1 - x_2 + x_3 &= 1, \\ 2x_1 + x_2 - x_3 &= -2, \\ x_1 - 2x_2 + 3x_3 &= 0. \end{aligned}$$

Solve it also when the right hand side is replaced

by $\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$; by $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

1.2 Solve the system:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ x_1 - x_2 + 2x_3 &= 4, \\ x_1 + 5x_2 - x_3 &= 1. \end{aligned}$$

Solve it also when the right hand side is replaced

$$\text{by} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad ; \quad \text{by} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

1.3 Solve the system:

$$\begin{aligned} 3x_1 + 4x_2 - x_3 &= 5, \\ -x_1 + 2x_2 + 3x_3 &= -7, \\ 2x_1 + 3x_2 - 2x_3 &= 1, \\ 5x_1 + 20x_2 + 9x_3 &= -13. \end{aligned}$$

1.4 Solve the following system:

$$\begin{aligned} 2x_1 - x_2 - x_3 + x_4 &= 1, \\ x_1 + 2x_2 - x_3 - x_4 &= 3, \\ 3x_1 - 7x_2 + x_3 + 2x_4 &= -1. \end{aligned}$$

Solve it also when the right hand side is replaced

$$\text{by} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \quad ; \quad \text{by} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

1.5 Show that elementary operation (iii) can be acheived by using operations (i) and (ii).

1.6

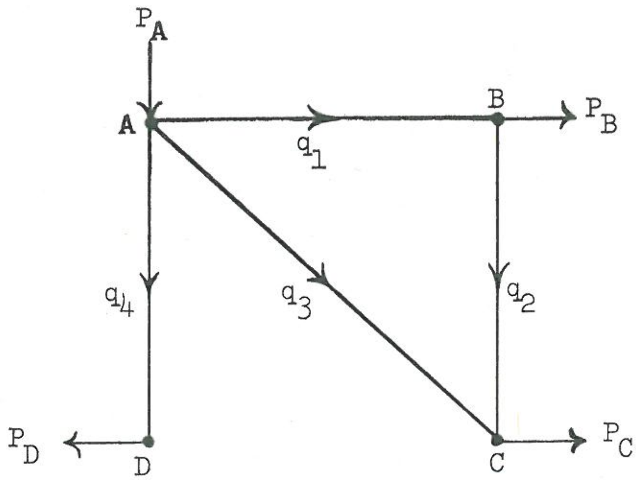


Figure 1.2

Figure 1.2 represents a water supply network. Water at the rate of P_A gpm is pumped in at point A. Water is taken out at the given rates P_B, P_C, P_D at points B, C, D. The flows in the lines are unknowns represented by q_1, q_2, q_3, q_4 with the assumed directions as indicated by the arrows. A negative value for any of these variables represents flow in the direction opposite to that indicated by the arrows. The equations expressing the conservation of mass at the four points A, B, C, D are, respectively,

$$\begin{aligned} q_1 + q_3 + q_4 &= P_A, \\ q_1 - q_2 &= P_B, \\ q_2 + q_3 &= P_C, \\ q_4 &= P_D. \end{aligned}$$

(a) Show that a solution cannot exist unless $P_A = P_B + P_C + P_D$. Interpret this physically in terms of conservation of mass.

(b) If $P_A = P_B + P_C + P_D$, show that the general solution is given by $q_1 = P_B + P_C - t$, $q_2 = P_C - t$, $q_3 = t$, $q_4 = P_A - P_C - P_B$

for any value of t. Interpret this physically for $t = 0$.

(c) Show that the arbitrary part of the solution, which may be added in any strength,

$$q_1 = -1,\ q_2 = -1,\ q_3 = 1,\ q_4 = 0,$$

represents a conservative circulation through the loop ABC and is a solution of the homogeneous system.

(d) Carry through a similar analysis for the system shown in Figure 1.3, which has an inlet at A and outlets at E and F.

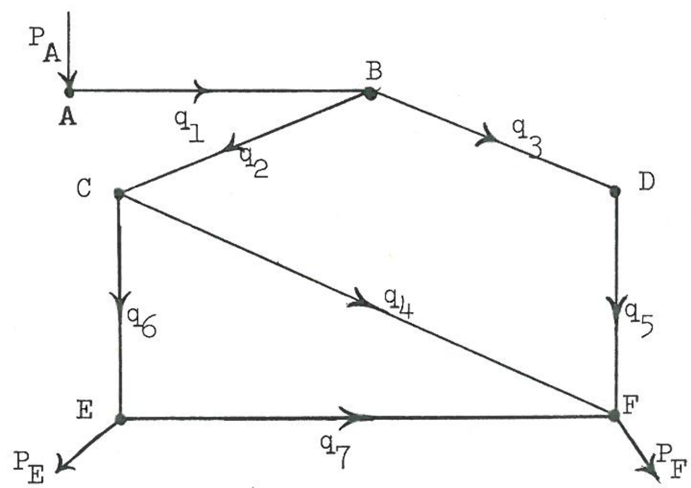


Figure 1.3

1.7 Consider the differential equation

$$A(x)y'' + B(x)y' + C(x)y = F(x) \tag{1.37}$$

subject to the boundary conditions

$$y(a) = \alpha\ , \quad y(b) = \beta\ .$$

To find an approximate solution by numerical methods we divide the interval (a,b) into N equal parts, each of length $h = \frac{b-a}{N}$, as indicated in Figure 1.4.

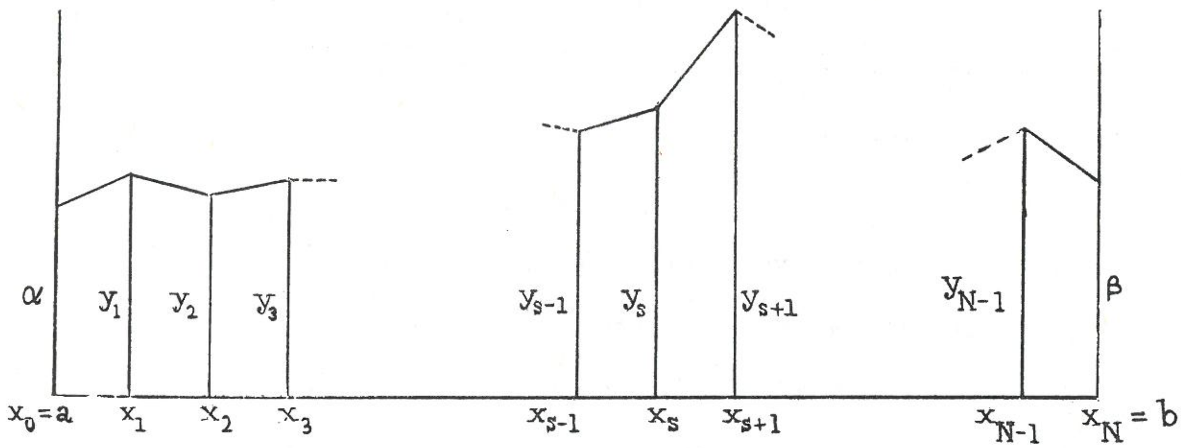


Figure 1.4

Let y_s denote the value of the approximate solution at the point x_s. By methods similar to those used in getting equation (2.4) of Chapter 1, one can derive the approximations

$$(1.38)\qquad \begin{aligned} y'(x_s) &= \frac{1}{2h}(y_{s+1} - y_{s-1}), \\ y''(x_s) &= \frac{1}{h^2}(y_{s+1} - 2y_s + y_{s-1}); \end{aligned}$$

each of these approximations having a truncation error of the order of h^2.

(a) Writing equation (1.37) for each value of x from x_1 to x_{N-1}, and replacing $y'(x_s)$ and $y''(x_s)$ by their approximate values from (1.38), show that we obtain the system of linear equations

$$\left[A(x_s)-\frac{h}{2}B(x_s)\right] y_{s-1} + \left[h^2C(x_s)-2A(x_s)\right] y_s + \left[A(x_s)+\frac{h}{2}B(x_s)\right] y_{s+1} = h^2F(x_s) ,$$
$$\text{for } s = 1, 2, \ldots , N-1$$

where $y_0 = \alpha$, $y_N = \beta$.

For example if $A(x) = B(x) = C(x) = F(x) \equiv 1$, $N = 5$, $a = 0$, $b = 1$ we get

$$0.9y_{s-1} - 1.96y_s + 1.1y_{s+1} = .04 \quad \text{for } s = 1, 2, 3, 4$$

or written out

$$\begin{aligned}
-1.96y_1 + 1.1\,y_2 \qquad\qquad\qquad\qquad &= .04 - 0.9\alpha, \\
0.9\,y_1 - 1.96y_2 + 1.1\,y_3 \qquad\qquad &= .04, \\
0.9\,y_2 - 1.96y_3 + 1.1\,y_4 &= .04, \\
0.9\,y_3 - 1.96y_4 &= .04 - 1.1\beta .
\end{aligned}$$

Note that it is particularly easy to reduce these equations to echelon form in the process of solving them.

(b) Solve this system for $\alpha = \beta = 0$. (Slide rule accuracy is good enough.) Ans. $y_1 = -.0969$, $y_2 = -.1363$, $y_3 = -.1273$, $y_4 = -.0788$.

(c) Set up and solve the analogous system if $A(x) \equiv 1$, $B(x) = x$, $C(x) = x + \frac{1}{2}$, $F(x) = x^2$, $N = 5$, $a = 0$, $b = 2$, $\alpha = 1$, $\beta = 4$. Ans. $y_1 = 4.016$, $y_2 = 6.074$, $y_3 = 6.563$, $y_4 = 5.609$.

(d) How does the discussion at the end of Section 4 of Chapter 1, in particular the three possibilities mentioned there, relate to this computational method?

1.8 The moments at three consecutive supports of a uniformly loaded beam are related by the "Three Moments Equation" *

$$\mathbf{A}M_1 + 2(\mathbf{A} + B)M_2 + BM_3 = \frac{w}{4}(\mathbf{A}^3 + B^3),$$

where w is the load per unit length along the beam and $\mathbf{A}$ and B are the distances between supports as indicated in Figure 1.5.

---

* E.P. Popov, Mechanics of Materials, Prentice-Hall Inc., 1952, p. 329.

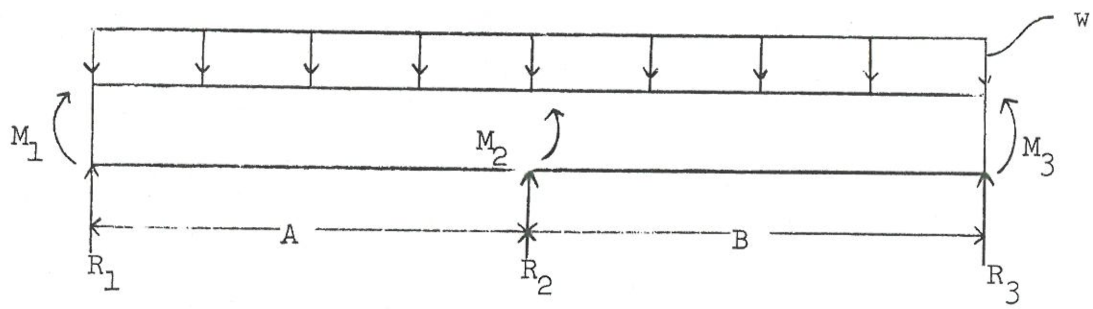


Figure 1.5

Find the moments at the supports of the structure shown in Figure 1.6. Ans. M_2 = 977 lb.-ft., M_3 = 330 lb.-ft., M_4 = 190 lb.-ft., M_5 = 1362 lb.-ft.

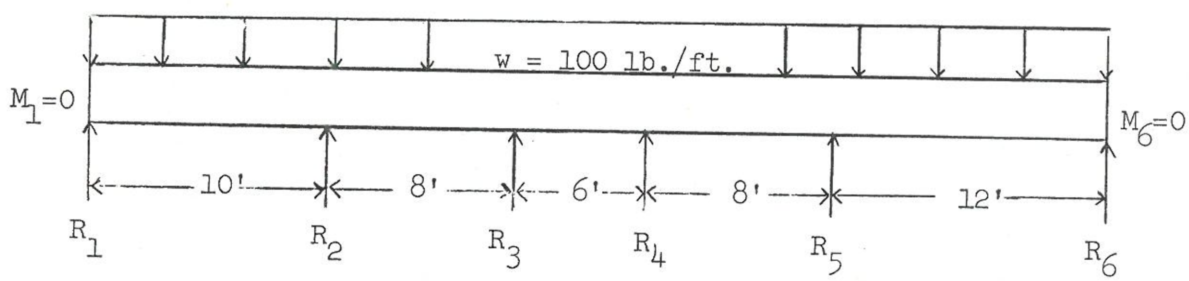


Figure 1.6

1.9 Find the currents in the various resistors of Figure 1.7 by setting up all of the Kirchoff node and loop equations and reducing the system of equations to echelon form. Take E = 10 volts, $R_1 = R_2$ = 100Ω, $R_3 = R_4 = R_5$ = 200Ω. Ans. $i_1 = i_2$ = .04 amps, $i_3 = i_5$ = .03 amps, i_4 = .01 amps.

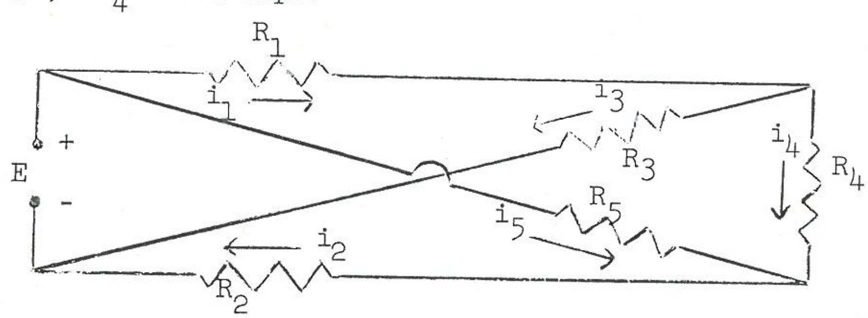


Figure 1.7

1.10

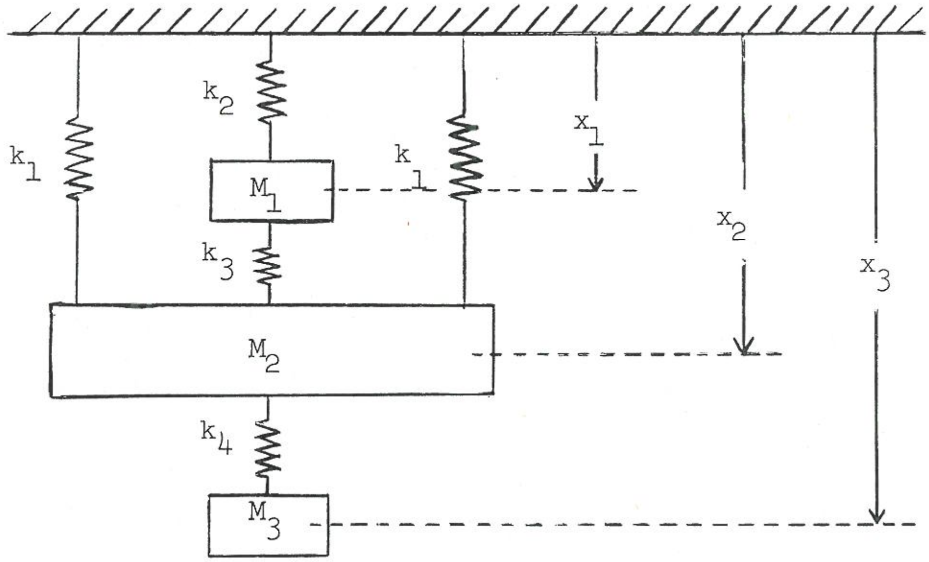


Figure 1.8

The system shown in Figure 1.8 is hanging at rest. Let the natural length of spring k_i be L_i $(i = 1,2,3,4)$, assumed measured between the same reference marks as the x's, i.e. when $x_1 = L_2$ spring k_2 exerts no force, etc. Assume that the values of the k's, L's and M's are given and write the force balance equation for each mass to find three equations for the three unknowns (x_1, x_2, x_3). Solve for the case $k_1 = 10$ lb./in., $k_2 = 20$ lb./in., $k_3 = 5$ lb./in., $k_4 = 10$ lb./in., $M_1 = 5$ lb., $M_2 = 20$ lb., $M_3 = 5$ lb., $L_1 = 2$ in., $L_2 = 1$ in., $L_3 = 1$ in., $L_4 = 2$ in. Ans. $x_1 = 17/12$ in., $x_2 = 37/12$ in., $x_3 = 67/12$ in.

1.11 Equations are said to be in the "triple diagonal form" if they are of the form

$$
\begin{aligned}
b_1x_1 + c_1x_2 \qquad\qquad\qquad &= d_1, \\
a_2x_1 + b_2x_2 + c_2x_3 \qquad &= d_2, \\
a_3x_2 + b_3x_3 + c_3x_4 &= d_3, \\
\text{- - - - - - -}
\end{aligned}
$$

$$a_{n-1}x_{n-2} + b_{n-1}x_{n-1} + c_{n-1}x_n = d_{n-1},$$
$$a_n \; x_{n-1} + b_n \; x_n = d_n.$$

Such systems of equations arise frequently (see Problems 1.7 and 1.8 for examples). If, as is usually the case, the b's are sufficiently large compared with the a's and the c's there is a unique solution. The Gauss elimination method is particularly easy for triple diagonal systems. Program it, assuming initial input of N and A(I), B(I), C(I), D(I) for I = (1,1,N). (Since A(1) and C(N) are not used the values assigned to them are immaterial.)

Use your program to solve Problems 1.7b and 1.8

The programming of a more general Gauss process will be delayed until additional notation and technique have been introduced.

VECTOR SPACES

2. Physical Vectors

In this section we utilize some of the features of the vectors that the student has already met in mathematics, physics, and mechanics in order to provide motivation for the abstract vector spaces which will be introduced in Section 3.

The student will recall from his course in Mechanics that if a system of forces is applied at a point (as in Figure 2.1a), the net effect of the system of forces is the same as that of the resultant force, computed by the "head-to-tail" rule (as in Figure 2.1b). This is also called the rule for addition of vectors. The resultant in

Figure 2.1b is denoted by

(2.1) $$\vec{R} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4$$

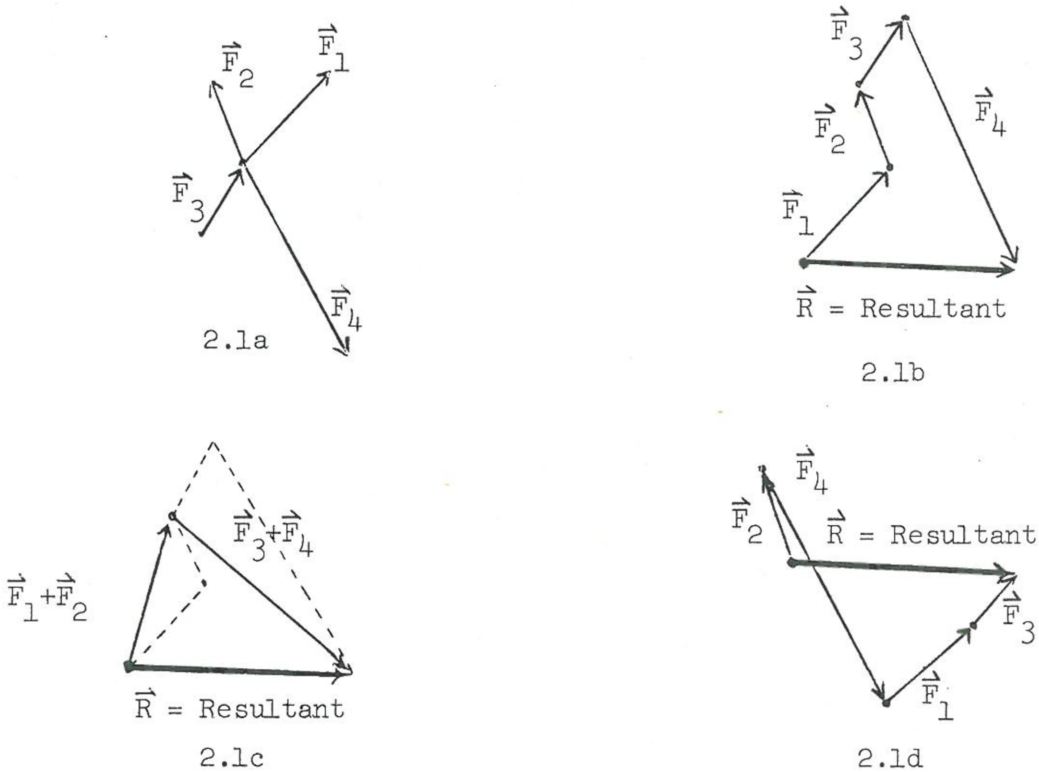


Figure 2.1

Two properties of this rule are worth noting. First, parentheses may be inserted on the right hand side of (2.1) without changing the resultant. For example, by $(\vec{F}_1 + \vec{F}_2) + (\vec{F}_3 + \vec{F}_4)$ we mean that we first apply the "head-to-tail" rule to $\vec{F}_1$ and $\vec{F}_2$, then to $\vec{F}_3$ and $\vec{F}_4$ and then apply the rule again to $(\vec{F}_1 + \vec{F}_2)$ and $(\vec{F}_3 + \vec{F}_4)$, as indicated in Figure 2.1c. Similarly $\vec{F}_1 + (\vec{F}_2 + (\vec{F}_3 + \vec{F}_4))$, $((\vec{F}_1 + \vec{F}_2) + \vec{F}_3) + \vec{F}_4$, $\vec{F}_1 + (\vec{F}_2 + \vec{F}_3) + \vec{F}_4$, and so on, all give the same resultant. These facts are implied by the Associative Law for Addition of Vectors:

For any three vectors $\vec{V}_1$, $\vec{V}_2$, $\vec{V}_3$

$$\vec{V}_1 + (\vec{V}_2 + \vec{V}_3) = (\vec{V}_1 + \vec{V}_2) + \vec{V}_3 \tag{2.2}$$

In view of (2.2) we write the sum simply as $\vec{V}_1 + \vec{V}_2 + \vec{V}_3$, with the freedom to insert parentheses as we please. The proof of (2.2) is indicated in Figure 2.2. From the associative law for three vectors the associative law for any number of vectors is easy to verify.

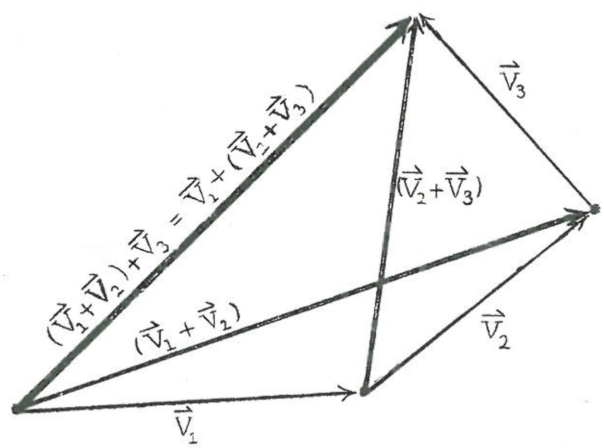


Figure 2.2

Associative Law of Addition of Vectors

Secondly, the order in which terms on the right hand side of (2.1) appear does not affect the resultant. For example $\vec{F}_2 + \vec{F}_4 + \vec{F}_1 + \vec{F}_3$ yields the same resultant, as indicated in Figure 2.1d. Similarly $\vec{F}_2 + \vec{F}_1 + \vec{F}_3 + \vec{F}_4$, $\vec{F}_4 + \vec{F}_1 + \vec{F}_2 + \vec{F}_3$, etc. all give the same resultant. These results are implied by the <u>Commutative Law for Addition of Vectors</u>:

For any two vectors $\vec{V}_1$, $\vec{V}_2$

$$\vec{V}_1 + \vec{V}_2 = \vec{V}_2 + \vec{V}_1 . \tag{2.3}$$

The proof of (2.3) is indicated in Figure 2.3. From the commutative law for two vectors the extension to any number of vectors is easy to verify, since any ordering of the terms can be achieved by successive transpositions of adjacent terms.

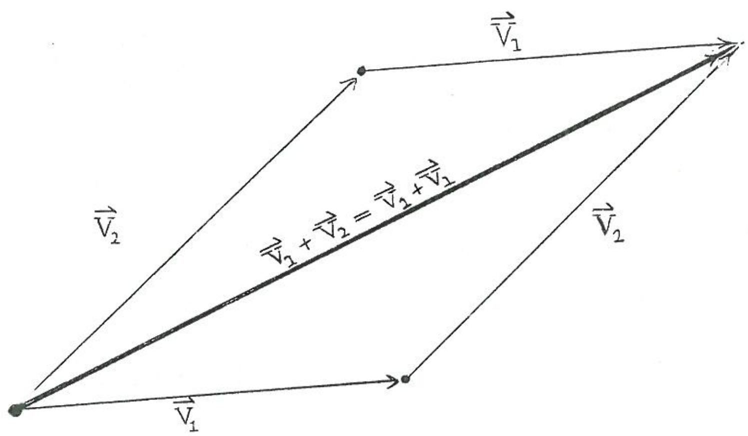


Figure 2.3

Commutative Law of Addition of Vectors

Another concept which the student has already met in Mechanics is that of multiplication of a vector by a scalar; i.e. by a real number. If a is a positive number and $\vec{F}$ is a vector then $a\vec{F}$ denotes a vector whose direction is the same as that of $\vec{F}$ and whose magnitude is a times the magnitude of $\vec{F}$. If a is a negative number then $a\vec{F}$ denotes a vector in the opposite direction to that of $\vec{F}$ whose magnitude is $|a|$ times the magnitude of $\vec{F}$. For the case $a = 0$ it is convenient to define $0\vec{F}$ as the zero vector $\vec{0}$ which has length zero and no direction associated with it. The following rules are then true.

(2.4) Addition of Zero Vector: $\vec{0} + \vec{F} = \vec{F}$.

(2.5) Distributive Laws: $(a + b)\vec{F} = a\vec{F} + b\vec{F}$,

(2.6) $a(\vec{F}_1 + \vec{F}_2) = a\vec{F}_1 + a\vec{F}_2$.

(2.7) Mixed Associative Law: $a(b\vec{F}) = (ab)\vec{F}$.

(2.8) Multiplication by Zero: $0\vec{F} = \vec{0}$.

(2.9) Multiplication by Unit: $1\vec{F} = \vec{F}$.

The student will find it easy to verify these rules (equation (2.6) follows from similar triangles.)

Many other physical quantities obey these same laws; for example, velocity of a particle, angular velocity of a rigid body, and electric and magnetic fields. It is, in fact, the wide application of the vector properties that makes them of importance in applied mathematics.

It should be pointed out that there are other properties of forces, as they occur in Mechanics, which are <u>not</u> reflected in the above discussion. For example, the line along which a force acts is needed to compute the moment of the force; the point at which a force is applied to a body is needed to determine the stresses and deformations of that body. The previous discussion considers only those aspects of a force which can be described in terms of its magnitude and direction. These constitute the vector aspects of the force and form a background for the developments of this chapter.

3. Vector Spaces

In manipulating physical vectors it is often convenient to express them in terms of certain suitably chosen "basis" vectors. If for example we are concerned with vectors lying in a plane, we can (cf. Thomas, Section 12-3) express any vector $\vec{v}$ in the form

$$\vec{v} = a\vec{i} + b\vec{j}, \tag{3.1}$$

where $\vec{i}$ and $\vec{j}$ are unit vectors along the axes of a suitably chosen Cartesian coordinate system (Figure 3.1). The vector $\vec{v}$ can then be represented by the ordered pair of scalars (a,b). Conversely, any ordered pair of scalars (a,b) represents a vector, given by (3.1).

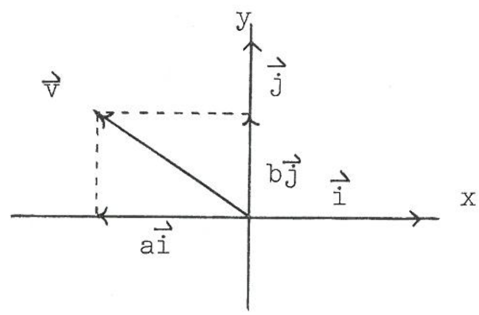


Figure 3.1

It is easy to show (cf. Thomas, loc. cit.) that if (a,b) represents $\vec{v}$ then (ca,cb) represents $c\vec{v}$, and if (a',b') represents $\vec{w}$ then (a+a',b+b') represents $\vec{v} + \vec{w}$. We are therefore led to make the following definitions:

$$\begin{aligned} (a,b) + (a',b') &= (a+a',b+b'), \\ c(a,b) &= (ca,cb). \end{aligned} \tag{3.2}$$

That is, instead of saying that (a,b) "represents" a vector we say that (a,b) "is" a vector, and that equations (3.2) define the two fundamental operations on these vectors just as certain geometric processes define the fundamental operations on the force vectors.

For these vectors it is a routine matter to prove the rules expressed in (2.2) to (2.9). First we define the zero vector, $\vec{0}$, to be the ordered pair (0,0). Then we easily verify the following:

$$(3.3)\qquad (a,b) + [(c,d) + (e,f)] = (a,b) + (c+e,d+f)$$
$$= (a+c+e,b+d+f)$$
$$= (a+c,b+d) + (e,f)$$
$$= [(a,b) + (c,d)] + (e,f);$$

$$(3.4)\qquad (a,b) + (c,d) = (a+c,b+d) = (c+a,d+b) = (c,d) + (a,b);$$

$$(3.5)\qquad (0,0) + (a,b) = (0+a,0+b) = (a,b);$$

$$(3.6)\qquad (a+b)(e,f) = ((a+b)e,(a+b)f) = (ae+be,af+bf)$$
$$= (ae,af)+(be,bf) = a(e,f)+b(e,f);$$

$$(3.7)\qquad a[(c,d)+(e,f)] = a(c+e,d+f) = (a(c+e),a(d+f))$$
$$= (ac+ae,ad+af) = (ac,ad)+(ae,af)$$
$$= a(c,d)+a(e,f);$$

$$(3.8)\qquad a(b(e,f)) = a(be,bf) = (abe,abf) = (ab)(e,f);$$

$$(3.9)\qquad 0(a,b) = (0a,0b) = (0,0);$$

$$(3.10)\qquad 1(a,b) = (1\cdot a,1\cdot b) = (a,b).$$

Thus ordered pairs have all the basic algebraic properties of physical vectors in the plane.

In the same manner, if we consider physical vectors in space, we arrive at the notion of a vector consisting of an ordered triple of scalars (a,b,c). In place of (3.1) we define

$$(3.11)\qquad \begin{aligned}(a,b,c) + (a',b',c') &= (a+a',b+b',c+c'),\\ p(a,b,c) &= (pa,pb,pc),\end{aligned}$$

and the zero vector is $\vec{0} = (0,0,0)$. The analogs of (3.3) to (3.10) can be proved in just the same fashion.

Now the physical picture stops at this point but the algebraic process does not. Instead of considering ordered pairs or ordered triples we can just as well consider ordered quadruples, or more generally, ordered n-tuples $(a_1, a_2, \ldots, a_n)$. We define, as before

$$
(3.12)\qquad
\begin{aligned}
(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) &= (a_1+b_1, a_2+b_2, \ldots, a_n+b_n),\\
c(a_1, a_2, \ldots, a_n) &= (ca_1, ca_2, \ldots, ca_n),\\
\vec{0} &= (0, 0, \ldots, 0),
\end{aligned}
$$

and once again the eight basic properties are proveable as before.

As an illustration of the use of this more general type of vector, consider first the case of a stockroom storing four items: hatchets, ratchets, latchets, and matchets! If on a given day there are on hand 896 hatchets, 218 ratchets, 898 latchets and 351 matchets, we can represent this information very concisely as the 4-tuple (896, 218, 898, 351) where we agree permanently that the first entry refers to hatchets, the second to ratchets, the third to latchets and the fourth to matchets. If now a shipment arrives containing 200 hatchets and 100 latchets we can represent the shipment by the 4-tuple (200, 0, 100, 0) and the new inventory by (896, 218, 898, 351) + (200, 0, 100, 0) = (1096, 218, 998, 351). If on the next day we hand out 100 hatchets, 20 ratchets, 100 latchets and have 10 matchets returned to stock, the increment can be represented by the vector (-100, -20, -100, 10) and the new gross inventory by (1096, 218, 998, 351) + (-100, -20, -100, 10) = (996, 198, 898, 361). If the general manager now says he wants the inventory to be raised by 10% then we must put in orders represented by the vector $\frac{1}{10}(996, 198, 898, 361) =$ (99.6, 19.8, 89.8, 36.1) i.e. we must order 100 hatchets, 20 ratchets, 90 latchets and 37 matchets.

As another illustration, consider the flow in the network shown in Figure 1.2 of Problem 1.6. The 4-tuple (q_1, q_2, q_3, q_4) represents a certain flow in the system as a result of given requirements (P_A, P_B, P_C, P_D) at the stations A, B, C, and D. If we take requirement vector (P'_A, P'_B, P'_C, P'_D) we get another flow vector (q'_1, q'_2, q'_3, q'_4). If we 'superpose' the requirement vectors $(P_A+P'_A, P_B+P'_B, P_C+P'_C, P_D+P'_D)$ the resulting flow is given by $(q_1+q'_1, q_2+q'_2, q_3+q'_3, q_4+q'_4)$. If the requirement vector (P_A, P_B, P_C, P_D) is scaled down to $\frac{1}{5}(P_A, P_B, P_C, P_D)$ the resulting flow is scaled down to $\frac{1}{5}(q_1, q_2, q_3, q_4)$. The answer to Problem 1.6 shows that $(q_1, q_2, q_3, q_4) = (P_B+P_C,\ P_C,\ 0,\ P_A-P_C-P_B) + t(-1, -1, 1, 0)$. The last vector represents a conservative circulation in the system. This situation will be discussed further in Section 11.

We have considered several different kinds of vectors - forces, velocities, and n-tuples for various values of n - and have seen that they all have certain algebraic properties, embodied in equations (2.2) to (2.9). We now take the important step of mathematical abstraction. Any collection of objects having the properties expressed by equations (2.2) - (2.9) will be called a vector space,* and the objects themselves will be called vectors. In detail:

Definition 3.1. Let V be any collection of objects $\{\vec{u}, \vec{v}, \vec{w}, \dots\}$ which is closed under operations of addition (i.e. if $\vec{u}$ and $\vec{v}$ are in V then $(\vec{u} + \vec{v})$ is in V) and under multiplication by sca-

---

*More properly a vector space over the real numbers, because the scalars here are real numbers.

lars (i.e. if $\vec{u}$ is in V and a is a real number then $(a\vec{u})$ is in V.) Suppose also that there is an element $\vec{0}$ in V (the zero vector) and that the following eight identities are true for any elements $\vec{u}, \vec{v}, \vec{w}$ in V and any real numbers a, b:

(1) Associative Law of Vector Addition $\quad \vec{u} + (\vec{v}+\vec{w}) = (\vec{u}+\vec{v}) + \vec{w}$,

(2) Commutative Law of Vector Addition $\quad \vec{u}+\vec{v} = \vec{v}+\vec{u}$,

(3) Zero Addition Law $\quad \vec{0}+\vec{u} = \vec{u}$,

(4) Scalar Distributive Law $\quad a(\vec{u}+\vec{v}) = a\vec{u} + a\vec{v}$,

(5) Vector Distributive Law $\quad (a+b)\vec{u} = a\vec{u} + b\vec{u}$,

(6) Mixed Associative Law $\quad a(b\vec{u}) = (ab)\vec{u}$,

(7) Zero Multiplier Law $\quad 0\vec{u} = \vec{0}$,

(8) Unit Multiplier Law $\quad 1\vec{u} = \vec{u}$.

Then V is called a <u>vector space</u> and its elements $(\vec{u}, \vec{v}, \vec{w} \ldots)$ are called <u>vectors</u>.

The sum $\vec{u} + (-1)\vec{v}$ is usually denoted by $\vec{u}-\vec{v}$ and the operation is called <u>subtraction</u>, since $\vec{u}-\vec{v}$ is the solution of the equation $\vec{v}+\vec{x} = \vec{u}$. Thus far, we have used a special typographical device to single out those symbols that stand for vectors, namely an arrow above the symbol. (The more usual device in type-set material is a bold-face symbol.) While the use of some such device is sometimes helpful in avoiding confusion between vectors and scalars it is by no means necessary, any more than it is necessary to use a special kind of f in expressing a function $f(x)$. We shall feel free to omit the arrow when there is no danger of confusion, and the reader is advised to get used to both notations. Note that if arrows are omitted the symbol 0 may stand either for the scalar zero or the vector $\vec{0}$; the context should always make the meaning clear.

Example 3.1. The vector space V_n.

We have seen that n-tuples of scalars $(a_1, a_2, \ldots, a_n)$ can be considered as vectors, n being any integer greater than 1. The vector space of n-tuples is designated by V_n. Thus, if we say that v (or $\vec{v}$) is a "member of V_n," or an "element of V_n," or that "v is in V_n," or "v belongs to V_n," we mean that v is an n-tuple. The scalars that make up this n-tuple are called the components of v. Thus $v = (2,3,0,1,-7.5)$ is an element of V_5, and the fifth component of v is -7.5. Two elements of V_n are equal if and only if their corresponding components are equal. An important consequence of this is that a vector equation is equivalent to n scalar equations. For example, to solve the equation $\vec{a} + 3\vec{x} = \vec{b}$, where $\vec{a} = (1,2,3)$ and $\vec{b} = (-2,5,6)$ we let $\vec{x} = (x_1, x_2, x_3)$. Then the equation becomes

$$(1,2,3) + 3(x_1, x_2, x_3) = (-2,5,6), \tag{3.13}$$

or, combining the vectors on the left hand side,

$$(1+3x_1, 2+3x_2, 3+3x_3) = (-2,5,6). \tag{3.14}$$

This vector equation means exactly the same as the three scalar equations

$$\begin{aligned} 1 + 3x_1 &= -2, \\ 2 + 3x_2 &= 5, \\ 3 + 3x_3 &= 6, \end{aligned} \tag{3.15}$$

from which we get $x_1 = -1$, $x_2 = 1$, $x_3 = 1$, or $\vec{x} = (-1,1,1)$ as the solution. The passage from (3.13) to (3.15) is generally referred to as "writing out the equation in terms of its components" and is usually done in one step, omitting the middle step (3.14); it is a very common way of handling problems in V_n.

One might ask, "Is there a V_1?" There is - it is just the set of scalars. The reader can readily verify that they satisfy Definition 3.1 and so are entitled to be called "vectors." Those readers interested in abstract mathematics might like to know that there is even a V_0!

<u>Example 3.2</u>. Spaces of polynomials in one variable.

Consider polynomials in a variable t, that is, expressions of the form

$$a_0 + a_1 t + \ldots + a_m t^m,$$

where the coefficients $a_0, a_1, \ldots, a_m$ are real numbers. If $p(t)$ and $q(t)$ are two such polynomials (not necessarily with the same values for m) so are $p(t) + q(t)$ and $cp(t)$ for any scalar c. If we define the zero polynomial as the one whose coefficients are all zero it is easy to check that these polynomials satisfy Definition 3.1 and so constitute a vector space.

Instead of considering <u>all</u> polynomials we often find it useful to restrict our attention to those polynomials whose degrees do not exceed some specified value n. These also form a vector space, which we designate by P^n; for example, P^3 consists of all polynomials of the form $a_0+a_1t+a_2t^2+a_3t^3$, where some, or all, of the coefficients a_0, a_1, a_2, a_3 may be zero. In line with this designation we use P^{∞} to designate the vector space with no restriction on the degree. As to the proof that P^n is a vector space, it can be done by referring back to Definition 3.1 but a much quicker proof will be given in Section 5.

One can also consider vector spaces of polynomials in two or more variables but we shall have no use for them in what follows.

Example 3.3. Function spaces.

To see that vector spaces can be made out of functions let us first review some of the facts concerning real-valued functions of a real variable t. Such a function, f, is a rule whereby, given a value of the independent variable t, the value $f(t)$ of the function, sometimes called the dependent variable, is determined. However, the values to be assigned to t may be restricted; for example, in considering the function $f(t) = \sqrt{t}$ we restrict t to non-negative values. (We have seen in the previous chapter how such restrictions can arise and later chapters will bring other examples.) In general let us assume that t is restricted to some interval I defined by inequalities such as $a < t < b$. Then if f and g are any functions defined on I and if c is any real number we define the functions

$$h = f + g \quad \text{and} \quad k = cf$$

by the equations

$$h(t) = f(t) + g(t), \qquad k(t) = cf(t),$$

for all values of t on I. (See Figure 3.2.) The zero function is defined to be that function which is identically zero on I. Once again it is easy to check Definition 3.1. to show that with these definitions of addition and multiplication by scalars the set of all functions in I is a vector space.

For applications, and even for pure theory, the space of all functions in I is too inclusive to be of much value.

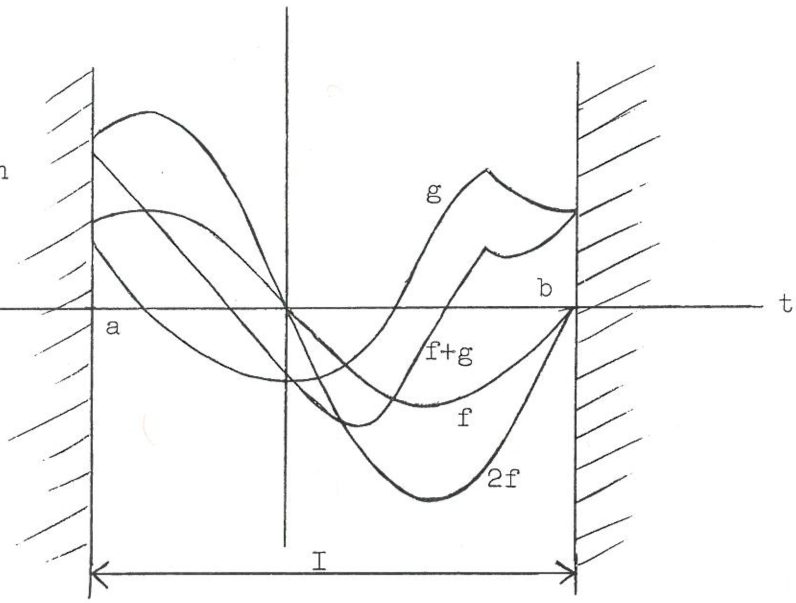


Figure 3.2

We find it best to restrict our functions by requiring them to have various useful properties, and among the most generally useful properties are continuity and differentiability. The space of all <u>continuous</u> functions on I is designated by C^0 (or $C^0(a,b)$ if we wish to indicate the interval of definition), the space of all functions having a first derivative which is continuous by $C^1, \ldots,$ the space of all functions having a continuous n-th derivative by C^n, and, finally, the space of all functions having an infinite number of derivatives by C^∞. (Functions of this type will be found in the next chapter to be very important.) That all these are vector spaces will be shown in Section 5. In Figure 3.2, f and g are elements of C^0, but g is not an element of C^1. (Why?) f is part of the graph of the function $f(t) = \sin t$ and so belongs to C^∞ - and, of course, also to C^n for any value of n.

Problems

3.1 In V_5, if $u = (1,0,-1,3,-2)$ and $v = (0,0,-2,-4,1)$ compute $u+v$, $u-v$, $5u+2v$. If also $w = (1,1,-1,3,2)$ compute $-w$, $2u+v-3w$.

3.2 Solve for the vector u in V_5:

$$2u+3(0,1,0,1,1) = -5(0,1,5,5,2).$$

3.3 Solve for the vectors u, v in V_4:

$$3u+4v = (0,1,0,1),$$

$$-2u+7v = (-1,2,-7,0).$$

3.4 An $m \times n$ (read "m by n") matrix is a rectangular array of numbers having m horizontal rows and n vertical columns. For example

$$\begin{pmatrix} 2 & 0 & -4 \\ 6 & .3 & 2 \end{pmatrix}$$

is a 2×3 matrix. Can you define addition of $m \times n$ matrices and multiplication by scalars so as to obtain a vector space? What is the zero matrix? Tell how you would prove that you actually have a vector space (do not go into all the details of the proof).

3.5 A polynomial $p(t)$ can be regarded as a function, since giving the variable t a value determines a value of the polynomial. There is no restriction on t, so $p(t)$ is defined on any interval (including the interval $-\infty < t < \infty$, i.e. the entire line). Is $p(t)$ a member of C^0, that is, is $p(t)$ continuous? Is $p(t)$ a member of C^1, of C^2, of C^∞?

3.6 Do the polynomials whose degrees are exactly n form a vector space? If so, what is the zero element $\vec{0}$ in this space? If not, why not?

3.7 How would you construct a vector space that might reasonably be designated by V_∞? [There are at least two ways of doing so.]

4. Further Properties of Physical Vectors

The object of this section is to provide an intuitive background, in terms of the physical vectors discussed in Section 2, for the more abstract development in Section 5. The concepts discussed here will be generalized in Section 5 and developed in more precise terms there.

The vectors that we are usually concerned with in Mechanics are physical vectors in three-dimensional space. Sometimes, due to the special nature of a particular problem, we need consider only vectors lying in a particular plane. The set of all vectors lying in this plane themselves constitute a vector space of two dimensions. This is called a vector subspace of the three-dimensional space. Similarly we might find it convenient to take components along a given line. The set of all vectors lying along this line constitutes a one-dimensional vector space. It also is a subspace of the three-dimensional physical space.

In Figure 4.1 we show two vectors $\vec{v}_1$ and $\vec{v}_2$ in three-dimensional space. A vector $\vec{v}_3 = a_1\vec{v}_1 + a_2\vec{v}_2$ where a_1 and a_2 are scalars, is called a linear superposition, or a <u>linear combination</u> of $\vec{v}_1$ and $\vec{v}_2$. The totality of all linear combinations of $\vec{v}_1$ and $\vec{v}_2$ sweeps out the plane P passing through the origin and containing $\vec{v}_1$ and $\vec{v}_2$. It is called the vector subspace <u>generated</u> Spanned by the set $\{\vec{v}_1, \vec{v}_2\}$. Since any vector in the plane P can be

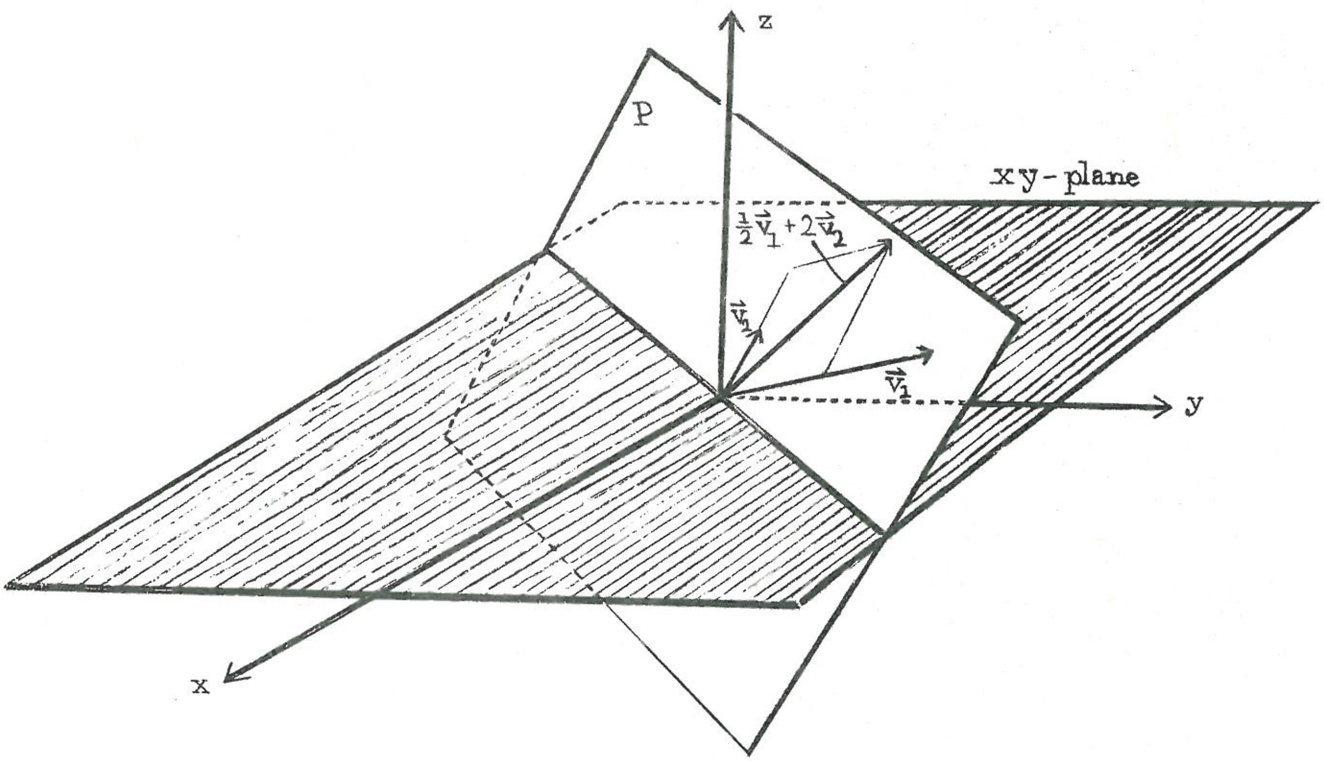


Figure 4.1

represented as a linear combination of $\vec{v}_1$ and $\vec{v}_2$, we say that the set $\{\vec{v}_1, \vec{v}_2\}$ spans plane P. If $\vec{w}$ is any vector in P then the set $\{\vec{v}_1, \vec{v}_2, \vec{w}\}$ is said to be <u>linearly dependent</u>, because we can express one of the vectors, $\vec{w}$, as a linear combination of the other two. A set of vectors in which none can be expressed as linear combinations of the others, such as the set $\{\vec{v}_1, \vec{v}_2\}$ of

Figure 4.1, is said to be linearly independent.

In Figure 4.2 we show three non-coplanar vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Since, e.g. $\vec{v}_3$ does not lie in the plane Q through the origin and containing $\vec{v}_1$ and $\vec{v}_2$, it is clear that $\vec{v}_3$ cannot be expressed as a linear combination of $\vec{v}_1$ and $\vec{v}_2$. Similarly none

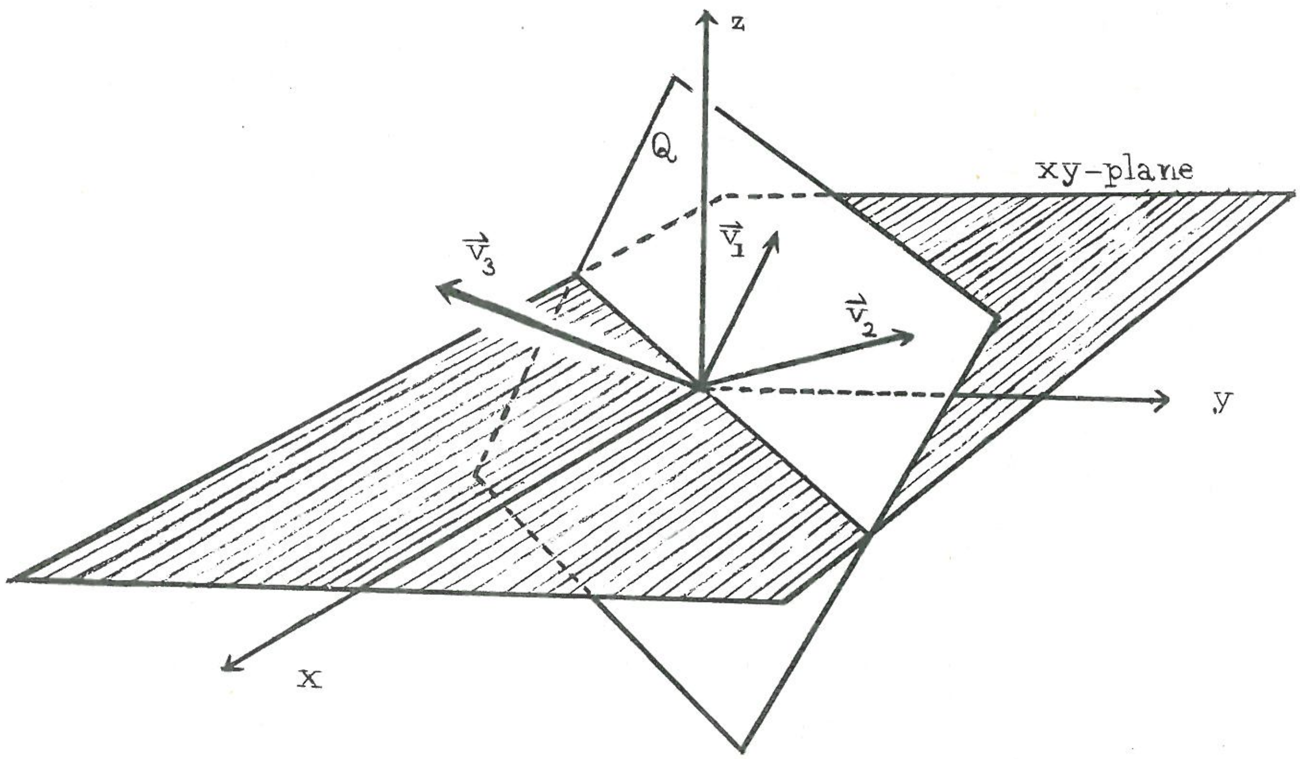


Figure 4.2

of the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ can be expressed as a linear combination of the others and consequently the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is independent. Furthermore, any vector in the three-dimensional space can be expressed as a linear combination of $\vec{v}_1, \vec{v}_2$ and $\vec{v}_3$. Thus the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent and spans all of the three-dimensional space. It is called a basis for the three-dimensional space. Similarly, in a two-dimensional plane, any two non-collinear vectors are independent and span the plane. They are said to be a basis for the two-dimensional plane.

The following facts will be evident to students with a good geometric intuition; the others may take comfort from the fact that they follow readily from the algebraic results of the next section.

(1) In three-dimensional space (physical space),

(a) Any set of four or more vectors is linearly dependent;

(b) Any three independent vectors span the space and hence form a basis;

(c) Any basis consists of a set of three non-coplanar vectors and conversely.

(2) In two-dimensional space (the plane),

(a) Any set of three or more vectors is linearly dependent;

(b) Any two independent vectors span the space and hence form a basis;

(c) Any basis consists of a set of two non-collinear vectors and conversely.

5. <u>Vector Subspaces</u>. <u>Linear Combination</u>.

We now return to the general vector spaces of Section 3. In Sections 5, 6, and 7 we shall give precise definitions of many of the terms used in a loose fashion in Section 4, and investigate various relations between them.

Let V be a vector space. A set S of vectors of V is called a <u>vector subspace</u> of V if S itself is a vector space under the laws of addition and multiplication by scalars as defined in V.

Example 5.1. Let V be V_5 and let S be the set of all 5-tuples of the form $(x_1,x_2,x_3,x_4,0)$. It is easy to show that S is closed under addition and under multiplication by scalars. Also, S contains the zero vector $(0,0,0,0,0)$. And finally, S satisfies the eight identities of Definition 3.1. Hence S is a vector space and so it is a vector subspace of V_5.

Since a vector subspace is, by definition, a vector space, any definition or theorem referring to vector spaces in general can also be applied to any vector subspace.

Two extreme cases of vector subspaces must be noted. In any vector space V the set consisting of the single vector $\vec{0}$ is a vector subspace of V; and also, V is a vector subspace of itself. A vector subspace of V which is neither $\vec{0}$ nor V itself is called a proper subspace.

The determination of whether or not a set S is a vector subspace of V is made easier by the following theorem. Often, in fact, the best way to prove that a set S (provided with rules of addition and scalar multiplication) forms a vector space, is to exhibit S as a subset of a known vector space V and use the test given by Theorem 5.1.

Theorem 5.1. A non-empty set S of vectors of V is a vector subspace of V if S is closed under addition and under multiplication by scalars.

Proof. First we note that S must contain the zero vector $\vec{0}$. For if $\vec{v}$ is any vector of S, then since S is assumed closed under multiplication by scalars, $0\vec{v} = \vec{0}$ is a vector of S. We have thus only to prove that the eight identities hold in S. But since they hold for all vectors in V they will assuredly hold for those particular vectors which lie in S. Hence S is a vector subspace.

Example 5.2. We can now supply the proof, for Examples 3.2 and 3.3, that P^n and C^n are vector spaces. Since every element (polynomial) of P^n is certainly an element of P^∞, by virtue of Theorem 5.1 we need only show that P^n is closed under addition and multiplication by scalars. This is evident, for if $p(t)$ and $q(t)$ have degrees at most n, the same is true for $p(t) + q(t)$ and $ap(t)$. Hence P^n is a vector subspace of P^∞, and so, by definition, it is a vector space. Similarly, if the functions $f(t)$ and $g(t)$ belong to C^n, that is, if they have continuous n-th derivatives, then so do $f(t) + g(t)$ and $af(t)$. C^n is therefore a vector subspace of the vector space of all functions, and so it is a vector space. In the same manner we can see that C^∞ is a vector space.

Example 5.3. Let $a_1, a_2, \ldots, a_n$ be any fixed set of scalars. Then the set of all vectors $\vec{x} = (x_1, x_2, \ldots, x_n)$ of V_n satisfying the condition $\sum_{i=1}^{n} a_i x_i = 0$ is a vector subspace of V_n. For it is easy to check that if $\vec{x}$ satisfies the condition so does $a\vec{x}$, and if $\vec{x}$ and $\vec{y}$ do, so does $\vec{x} + \vec{y}$.

Example 5.4. If u and v are two vectors in V, the set of all vectors w of the form $w = au + bv$, where a and b are arbitrary scalars, is a vector subspace of V. Here again it is easy to check the closure of the set of vectors $\{w\}$ under addition and multiplication by scalars. It is evident that the example can be generalized to vectors of the form $w = \sum_{i=1}^{m} a_i v_i$, if $\{v_1, \ldots, v_m\}$ is any fixed set of vectors.

This last example leads us to an important definition. Consider a set of vectors $\{v_1, v_2, \ldots, v_m\}$ in V. By a <u>linear combination</u> of these vectors we mean a vector of the form

$$\sum_{i=1}^{m} a_i v_i = a_1 v_1 + a_2 v_2 + \cdots + a_m v_m, \tag{5.1}$$

where the a_i's are scalars. A linear combination is said to be <u>non-trivial</u> if at least one of the a_i's is different from zero.

Let V be a vector space and let $\{v_1, v_2, \ldots, v_m\}$ be a set of vectors in V. The set of all linear combinations

$$\sum_{i=1}^{m} a_i v_i$$

of the set $\{v_1, v_2, \ldots, v_m\}$, obtained by taking all possible choices of the coefficients a_i, is called the vector subspace <u>spanned</u> by (or <u>generated</u> by) the set $\{v_1, v_2, \ldots, v_n\}$. In particular, V is itself spanned by $\{v_1, v_2, \cdots, v_m\}$ if every vector in V is a linear combination of $\{v_1, v_2, \cdots, v_m\}$.

The following examples show that questions involving linear combinations of vectors often reduce to the kind of problems in Section 1.

<u>Example 5.5</u>. Is (1,2,5) a linear combination of (-1,2,11), (2,5,14), (3,-3,-21)? Is (1,2,-1)?

To answer the first question we try to find a_1, a_2, a_3 such that

$$a_1(-1,2,11) + a_2(2,5,14) + a_3(3,-3,-21) = (1,2,5). \tag{5.2}$$

Writing this out in terms of components gives the system of linear equations:

$$(5.3)\qquad \begin{aligned} -a_1 + 2a_2 + 3a_3 &= 1, \\ 2a_1 + 5a_2 - 3a_3 &= 2, \\ 11a_1 + 14a_2 - 21a_3 &= 5. \end{aligned}$$

In Example 1.3 this system was seen to have solutions; for example, $a_1 = -1/9$, $a_2 = 4/9$, $a_3 = 0$. Hence $(1,2,5)$ is a linear combination of the three given vectors. On the other hand, Example 1.2 shows that $(1,2,-1)$ is not a linear combination of these three, since if we replace $(1,2,5)$ in (5.2) by $(1,2,-1)$ the equations corresponding to (5.3) have no solution.

Note that a given vector may be expressible in more than one way as a linear combination of vectors. In the above example, for instance, equations (5.3) do not have a unique solution, and we could equally well have taken the solution $a_1 = 3$, $a_2 = 0$, $a_3 = 4/3$. We shall see later that the case when this <u>cannot</u> happen is of special importance.

<u>Example 5.6.</u> Does the set of vectors $\{(1,2,-3),(1,1,4),(3,1,26)\}$ span V_3?

We must determine whether, given <u>any</u> vector (c_1,c_2,c_3) in V_3, there exist numbers a_1,a_2,a_3 such that

$$a_1(1,2,-3) + a_2(1,1,4) + a_3(3,1,26) = (c_1,c_2,c_3).$$

Writing this equation in terms of components gives us

$$(5.4)\qquad \begin{aligned} a_1 + a_2 + 3a_3 &= c_1, \\ 2a_1 + a_2 + a_3 &= c_2, \\ -3a_1 + 4a_2 + 26a_3 &= c_3. \end{aligned}$$

These reduce to the echelon form

$$(5.5)\qquad \begin{matrix} 1 & 1 & 3 & c_1 \\ 0 & 1 & 5 & 2c_1-c_2 \\ 0 & 0 & 0 & -11c_1+7c_2+c_3; \end{matrix}$$

consequently there is no solution unless $-11c_1+7c_2+c_3=0$. For instance, $(1,0,0)$ is not in the subspace spanned by the given set. Hence the set $\{(1,2,-3), (1,1,4), (3,1,26)\}$ does not span V_3.

(We really did not need to carry along the column of constants in going from (5.4) to (5.5). It is evident that <u>some</u> combination of c_1, c_2, c_3 will appear in the last row of (5.5). Regardless of what this is, the appearance of the row of zeros in the echelon form of the coefficients is enough to tell us that the given set of vectors does not span the whole space.)

The following theorem about linear combinations is often useful.

<u>Theorem 5.2</u>. If w is a linear combination of $\{v_1, \ldots, v_n\}$, and if each v_i is a linear combination of $\{u_1, \ldots, u_m\}$ then w is a linear combination of $\{u_1, \ldots, u_m\}$.

<u>Proof</u>. If

$$w = a_1v_1 + a_2v_2 + \ldots + a_nv_n$$

and

$$v_i = b_{i1}u_1 + b_{i2}u_2 + \ldots + b_{im}u_m$$

for $i = 1, \ldots, n$, then

$$\begin{aligned} w = {} & a_1(b_{11}u_1 + b_{12}u_2 + \dots + b_{1m}u_m) \\ & + a_2(b_{21}u_1 + b_{22}u_2 + \dots + b_{2m}u_m) \\ & + \dots \\ & + a_n(b_{n1}u_1 + b_{n2}u_2 + \dots + b_{nm}u_m) \\ & = (a_1b_{11} + a_2b_{21} + \dots + a_nb_{n1})u_1 \\ & + (a_1b_{12} + a_2b_{22} + \dots + a_nb_{n2})u_2 \\ & + \dots \\ & + (a_1b_{1m} + a_2b_{2m} + \dots + a_nb_{nm})u_m. \end{aligned}$$

Since this is of the form $c_1u_1 + \dots + c_mu_m$, w is a combination of $u_1, \dots, u_m$.

Corollary 5.1. If $\{v_1, \dots, v_n\}$ span W and if each v_i is a linear combination of vectors $\{u_1, \dots, u_m\}$ in W, then $\{u_1, \dots, u_m\}$ span W.

Proof. If w is any vector in W, then w is a combination of $\{v_1, \dots, v_n\}$, and hence, by the Theorem, a combination of $\{u_1, \dots, u_m\}$. That is, $\{u_1, \dots, u_m\}$ span W.

Problems

5.1 In each of the following cases, does the set of vectors form a vector subspace of the indicated V_n?

(a) The set of all vectors of V_4 of the form (a,b,a,b)?

(b) The set of all vectors of V_3 of the form $(a,b,1)$?

(c) The set of all vectors of V_4 of the form $(a,-a,b,a-b)$?

(d) The set of all vectors of V_2 of the form (a,a^2)?

(e) The set of all vectors of V_3 whose components are positive?

Ans. (a) Yes, (c) Yes, (e) No.

5.2 In each of the following cases does the set of all polynomials $p(t)$ in P^n satisfying the given condition form a vector subspace of P^n?

(a) $p(t_o) = 0$, for some fixed t_o.

(b) $p(t_o) \geqq 0$, for some fixed t_o.

(c) $p(t)$ is divisible by $t^2 + t + 1$.

(d) $p(t)$ is an <u>even</u> polynomial; i.e. $p(-t) = p(t)$.

(e) The equation $p(t) = 0$ has no real roots.

(f) The equation $p(t) = 0$ has at least one real root.
Ans. (a) Yes, (c) Yes, (e) No.

5.3 Is $(3,4,1)$ a linear combination of $(1,1,1)$, $(1,-1,5)$, $(1,2,-1)$? Is $(1,0,0)$? Is $(0,0,0)$? Ans. (a) Yes.

5.4 A machinist's supply house has three mixtures of washers, as indicated by the following schedule of percentages by weight:

	1/4"	5/16"	3/8"	1/2"
Mixture I	10	20	30	40
Mixture II	25	25	25	25
Mixture III	50	30	20	0

(a) A customer orders 100 lbs. of a 32,26,24,18 mixture. Can this be made up out of the mixtures on hand? State this as a problem in linear combinations and solve the problem.

(b) Do the same for a 20,25,25,30 mixture.
Ans. (a) Yes, (b) No.

5.5 Do the vectors $\{(1,0,1), (2,1,-1), (5,2,1)\}$ span V_3? Ans. Yes.

5.6 Do the vectors $\{(0,1,2,-1), (-1,2,1,0), (3,1,1,0), (7,8,7,-1)\}$ span V_4?

5.7 Show that the two sets of vectors $\{(0,1,1,0), (1,0,0,1)\}$ and $\{(2,3,3,2), (1,2,2,1)\}$ span the same subspace of V_4.

6. Linear Dependence.

A finite set of vectors is said to be linearly dependent if we can express one of them as a linear combination of the others. If the set is not linearly dependent it is said to be linearly independent.

[Note. When there is no danger of ambiguity it is customary to omit the adverb "linearly" in speaking of combinations, dependence or independence of vectors.]

Example 6.1. From Example 5.5 it follows that the set $\{(1,2,5), (-1,2,11), (2,5,14), (3,-3,-21)\}$ is dependent. On the other hand we cannot conclude from this example that $\{(1,2,-1), (-1,2,11), (2,5,14), (3,-3,-21)\}$ is independent, even though $(1,2,-1)$ is not a combination of the other three, for one of the last three may be a combination of the others. (Example 6.2 will show that this is indeed the case.)

The following theorem gives another way of deciding whether or not a set of vectors is dependent. In many situations this alternative criterion is more useful than the definition.

Theorem 6.1. A finite set of vectors is linearly dependent if and only if there is a non-trivial linear combination of them which is equal to zero.

Proof. If $\{v_1, v_2, \ldots, v_n\}$ is dependent then at least one of the vectors, say v_j, is a linear combination of the rest;

$$v_j = \sum_{i \neq j} a_i v_i \ .$$

Writing this in the form

$$v_j - \sum_{i \neq j} a_i v_i = 0 \ ,$$

we have a non-trivial linear combination (since the coefficient of v_j is certainly not zero) equal to zero.

Conversely if we have a linear combination equal to zero,

$$\sum_{i=1}^{n} a_i v_i = 0 \ ,$$

with at least one non-zero coefficient, say $a_j \neq 0$, we can solve for v_j in terms of the others,

$$v_j = \sum_{i \neq j} (-a_i/a_j) v_i \ ,$$

and so the set is dependent.

Notice that we could assume here that j is the largest suffix such that $a_j \neq 0$. Then v_j is a linear combination of its predecessors $v_1, v_2, \ldots, v_{j-1}$. We thus have

Corollary 6.1. The vectors $v_1, v_2, \ldots, v_n$ are linearly independent if and only if no v_j is a linear combination of its predecessors $v_1, v_2, \ldots, v_{j-1}$.

Example 6.2. Are the vectors $\{(-1,2,11), (2,5,14), (3,-3,-21)\}$ dependent?

By the criterion of Theorem 6.1 they are dependent if there exist scalars x_1, x_2, x_3, not all zero, such that

$$x_1(-1,2,11) + x_2(2,5,14) + x_3(3,-3,-21) = 0.$$

This gives the linear system,

$$(6.1)\qquad \begin{aligned} -x_1 + 2x_2 + 3x_3 &= 0, \\ 2x_1 + 5x_2 - 3x_3 &= 0, \\ 11x_1 + 14x_2 - 21x_3 &= 0. \end{aligned}$$

From the material in Example 1.3 it is easily seen that these homogeneous equations have a non-trivial solution. Hence the given vectors are dependent.

It is evident from this example that dependence of vectors in V_n is closely related to the solution of homogeneous linear equations. In particular a very useful fact follows at once from Theorem 1.2.

Theorem 6.2. In V_n, any set containing more than n vectors is dependent.

Proof. Suppose the set has m vectors, with $m > n$. On writing out the equations corresponding to (6.1) we get a system of n homogeneous equations in m variables. By Theorem 1.2 (with the roles of m and n interchanged) there is a non-trivial solution, and so the vectors are dependent.

Example 6.3. Referring back to Example 6.1 we see now that the second set of four vectors is in fact dependent.

The above theorem refers only to the special vector space V_n of n-tuples. The next theorem does not suffer from this restriction but applies to any vector space V. Its proof, while more complicated than that of Theorem 6.2, uses the same basic idea.

Theorem 6.3. If $m > n$ and if each of the m vectors $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$ is a combination of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then the set $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is dependent.

Proof. We are given that scalars a_{ij} exist such that

$$(6.2)\qquad \begin{aligned} \vec{w}_1 &= a_{11}\vec{v}_1 + a_{12}\vec{v}_2 + \dots + a_{1n}\vec{v}_n, \\ \vec{w}_2 &= a_{21}\vec{v}_1 + a_{22}\vec{v}_2 + \dots + a_{2n}\vec{v}_n, \\ &\cdots\cdots\cdots \\ \vec{w}_m &= a_{m1}\vec{v}_1 + a_{m2}\vec{v}_2 + \dots + a_{mn}\vec{v}_n, \end{aligned}$$

and we wish to show that there are scalars $c_1, c_2, \dots, c_m$, not all zero, such that

$$(6.3)\qquad c_1\vec{w}_1 + c_2\vec{w}_2 + \dots + c_m\vec{w}_m = \vec{0}.$$

If we substitute from (6.2) into (6.3) we get

$$(6.4)\qquad \begin{aligned} & c_1(a_{11}\vec{v}_1 + a_{12}\vec{v}_2 + \dots + a_{1n}\vec{v}_n) \\ + \; & c_2(a_{21}\vec{v}_1 + a_{22}\vec{v}_2 + \dots + a_{2n}\vec{v}_n) \\ + \; & \cdots\cdots\cdots \\ + \; & c_m(a_{m1}\vec{v}_1 + a_{m2}\vec{v}_2 + \dots + a_{mn}\vec{v}_n) = \vec{0}. \end{aligned}$$

Rearranging terms by adding first along columns of (6.4) instead of across rows enables us to write this in the form

$$(6.5)\qquad \begin{aligned} & (c_1a_{11} + c_2a_{21} + \dots + c_ma_{m1})\vec{v}_1 \\ + \; & (c_1a_{12} + c_2a_{22} + \dots + c_ma_{m2})\vec{v}_2 \\ + \; & \cdots\cdots\cdots \\ + \; & (c_1a_{1n} + c_2a_{2n} + \dots + c_ma_{mn})\vec{v}_n = \vec{0}. \end{aligned}$$

(6.5) will be true if we can find c's that are a non-trivial solution of

$$\begin{aligned}
c_1a_{11} + c_2a_{21} + \dots + c_ma_{m1} &= 0,\\
c_1a_{12} + c_2a_{22} + \dots + c_ma_{m2} &= 0,\\
&\dots\dots\\
c_1a_{1n} + c_2a_{2n} + \dots + c_ma_{mn} &= 0.
\end{aligned} \tag{6.6}$$

But Theorem 1.2 applies to this system and tells us that there is a non-trivial solution for the c's. Hence we can satisfy (6.3) and so the set $\left\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\right\}$ is dependent.

The details of this proof provide a good illustration of the techniques involved in using the summation notation to abbreviate the writing of linear equations. Equation (6.2) can be condensed to

$$\vec{w}_i = \sum_{j=1}^{n} a_{ij}\vec{v}_j, \qquad i = 1, \dots, m, \tag{6.2a}$$

and (6.3) to

$$\sum_{i=1}^{m} c_i\vec{w}_i = \vec{0}; \tag{6.3a}$$

combining these two gives

$$\sum_{i=1}^{m} c_i \sum_{j=1}^{n} a_{ij}\vec{v}_j = \vec{0}. \tag{6.4a}$$

The passage from this to (6.5a) actually involves two extra steps. First we carry out the multiplication implied in $c_i \sum_{j=1}^{n} a_{ij}\vec{v}_j$ to get $\sum_{j=1}^{n} c_ia_{ij}\vec{v}_j$, so that (6.4a) becomes

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_ia_{ij}\vec{v}_j = \vec{0}.$$

Then we interchange the order of the two summations, to get

$$\sum_{j=1}^{n} \sum_{i=1}^{m} c_i a_{ij} \vec{v}_j = \vec{0};$$

this corresponds to the passage from summing first by rows to summing first by columns. Then the inner summation $\sum_{i=1}^{m} c_i a_{ij} \vec{v}_j$ can be factored to give $(\sum_{i=1}^{m} c_i a_{ij})\vec{v}_j$, and we get

$$\sum_{j=1}^{n} (\sum_{i=1}^{m} c_i a_{ij}) \vec{v}_j = \vec{0}. \tag{6.5a}$$

Equation (6.5a) is clearly satisfied for any scalars $c_1, c_2, \cdots c_m$ satisfying

$$\sum_{i=1}^{m} c_i a_{ij} = 0, \qquad j = 1, 2, \ldots, n. \tag{6.6a}$$

Such manipulation of summation signs is very useful in deriving properties of vectors and matrices.

We close this section with some useful facts about linear dependence and independence of vectors.

<u>Theorem 6.4.</u>

(a) If a set of vectors is dependent then any larger set containing these vectors is dependent.

(b) If the set $\{v_1, \ldots, v_n\}$ is independent and $\{v_1, \ldots, v_n, v_{n+1}\}$ is dependent then v_{n+1} is a combination of $\{v_1, \ldots, v_n\}$.

(c) If $\{v_1, \ldots, v_n\}$ is independent and spans a space W then any element w of W is expressible in only one way as a linear combination of $\{v_1, \ldots, v_n\}$.

Proof.

(a) Let $S = \{v_1, \dots, v_n\}$ be a dependent set of vectors and $S' = \{v_1, \dots, v_n, v_{n+1}, \dots, v_m\}$ a larger set. By Theorem 6.1, applied to the set S, there exist scalars, $a_1, \dots, a_n$, not all zero, such that

$$a_1v_1 + \dots + a_nv_n = 0.$$

But this equation is the same as

$$a_1v_1 + \dots + a_nv_n + 0v_{n+1} + \dots + 0v_m = 0,$$

which says that the set S' is dependent.

(b) If $\{v_1, \dots, v_n, v_{n+1}\}$ is dependent we have

$$a_1v_1 + \dots + a_nv_n + a_{n+1}v_{n+1} = 0, \tag{6.7}$$

with not all $a_1, \dots, a_{n+1}$ equal to zero. Now if a_{n+1} were zero, (6.7) would become

$$a_1v_1 + \dots + a_nv_n = 0,$$

which is impossible since $\{v_1, \dots, v_n\}$ are independent. Hence we must have $a_{n+1} \neq 0$, and so (6.7) can be solved for v_{n+1},

$$v_{n+1} = -\frac{a_1}{a_{n+1}}v_1 - \dots - \frac{a_n}{a_{n+1}}v_n .$$

(c) Suppose we have both

$$w = a_1v_1 + \dots + a_nv_n$$

and

$$w = b_1v_1 + \dots + b_nv_n.$$

Subtracting these two equation gives

$$(a_1-b_1)v_1 + \dots + (a_n-b_n)v_n = 0.$$

Since the v's are independent their coefficients must all be zero, so

$$a_1=b_1, \dots, a_n=b_n.$$

That is, the two expressions for w in terms of the v's are really the same.

Problems

6.1 Determine whether the following sets of vectors are independent.

(a) $\{(1,0,1,-1), (2,0,-1,1), (1,-2,1,0), (1,0,1,0)\}$.

(b) $\{(1,0,3), (2,1,-2), (1,3,4), (0,1,5)\}$.

(c) $\{(1,0,1,2,-1), (2,3,1,0,2), (4,9,1,-4,8)\}$.

(d) $\{(1,0,0,0), (4,1,0,0), (7,2.5,1,0), (\pi,\sqrt{3},4,1)\}$.

6.2 Show that any set of vectors containing the zero vector is dependent.

6.3 (a) Show that in V_6 the vectors

$$(2,1,-1,0,4,3),$$
$$(0,0,3,1,0,4),$$
$$(0,0,0,-1,2,3),$$
$$(0,0,0,0,0,1)$$

are independent.

(b) Generalize the above situation to show that in V_n, m non-zero vectors in this type of "echelon form" are independent.

6.4 (a) If $\{u,v,w\}$ are independent show that $\{u+v, u+w, v+w\}$ are independent.

(b) If $\{u,v,w\}$ are independent are $\{u-v, u-w, v-w\}$ independent?

(c) If $\{u,v,w\}$ are dependent, what, if anything, can be said about $\{u+v, u+w, v+w\}$?

6.5 Give a different proof of Theorem 6.4(b) by making use of Corollary 6.1.

6.6 Prove that if the set $\{v_1, \ldots, v_n\}$ is dependent and if each of $w_1, \ldots, w_n$ is a combination of $\{v_1, \ldots, v_n\}$, then $\{w_1, \ldots, w_n\}$ is dependent.

7. Bases. Dimension of a Vector Space.

A finite set of vectors $\{v_1, v_2, \ldots, v_n\}$ that is linearly independent and spans a vector space V is called a basis of V.*

Example 7.1. $\{(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)\}$ is an obvious basis of V_4, and might be thought of as the most natural basis to choose. But many other bases are possible; for example $\{(1,1,1,1), (0,1,1,0), (1,0,1,0), (0,1,1,1)\}$, (prove this).

The most important theorem about bases is the following.

Theorem 7.1. Any two bases of the same vector space must contain the same number of vectors.

Proof. Let B_1 be a basis of V with n vectors and B_2 a basis with m vectors. Since B_1 spans V, every vector of B_2 is a combination of vectors of B_1. If m were greater than n then, by Theorem 6.3, B_2 would be dependent. This contradicts the fact that B_2

---

*Strictly speaking, this is a finite basis. One sometimes considers infinite bases; for example, $\{1, x, x^2, \ldots\}$ is a basis for the vector space P^∞ of all polynomials.

is a basis. Hence m is not greater than n. By interchanging the roles of B_1 and B_2 we can show similarly that n is not greater than m. Hence $m = n$.

Thus if a vector space V has a basis at all, each of its bases has the same number of vectors. This number, which is thus uniquely associated with V, is called the dimension of V. A vector space which has a (finite) basis, and therefore a dimension, is said to be finite-dimensional; a space that has no (finite) basis is infinite-dimensional. For the statement of certain theorems it is convenient to regard a vector space consisting of the single vector $\vec{0}$ to be a finite-dimensional space of dimension 0.

Example 7.2.

(a) The dimension of V_n is n, (cf. Example 7.1 for $n = 4$).

(b) Physical experiment convinces us that any force acting at the origin is expressible as a linear combination of three unit forces $\vec{i}, \vec{j}, \vec{k}$ acting along the x-, y- and z-axes, and also that none of $\vec{i}, \vec{j}, \vec{k}$ is a combination of the other two. Hence $\{\vec{i}, \vec{j}, \vec{k}\}$ is a basis for forces in space and the dimension of the vector space of forces at a point is 3. Any basis for forces in space must therefore consist of precisely three vectors.

(c) In P^{∞} the set $S = \{1, x, x^2, \ldots, x^n\}$ is independent no matter how large n is. (Why?) If P^{∞} had dimension m then by Theorem 6.3 we would have to have $n \leq m$. Hence P^{∞} is infinite-dimensional. On the other hand the set S obviously spans P^n, and so S is a basis for P^n; P^n is therefore of dimension $n+1$.

This example suggests the question, "How do we determine the dimension, if any, of a given vector space?" There is no general answer to this question, for vector spaces can arise in a bewildering variety of ways and with all sorts of mathematical entities as their elements. However, the following theorem gives an answer in one important case, that in which the vector space is specified as the span of a finite set of vectors.

Theorem 7.2. A spanning set of a vector space W contains a basis of W.

Proof. Let $\{v_1, v_2, \ldots, v_m\}$ span W. Taking the v's one at a time in the order indicated by their subscripts, discard each one that is a combination of previous undiscarded ones. In the set of vectors that are left (call this set S) no vector is a combination of its predecessors (otherwise it would have been discarded) and so, by Corollary 6.1, S is independent. Furthermore, each of the original v's is a combination of the vectors in S, and so, by Corollary 5.1, S spans W. Thus S, being an independent spanning set, is a basis of W.

Corollary 7.1. A space spanned by n vectors has dimension at most n.

Example 7.3. Find the dimension of the subspace W of V_5 spanned by the set $\{(1,2,3,4,5), (2,4,6,1,-4), (-3,-6,-9,-2,5), (0,1,1,1,0), (1,3,4,3,1)\}$, and select a basis from this spanning set.

The elimination technique of Section 1 can be nicely adapted to the method of Theorem 7.2. We write down the matrix whose rows are the given vectors,

$$
(7.1) \qquad
\begin{array}{rrrrrcl}
1 & 2 & 3 & 4 & 5 & = & v_1 \\
2 & 4 & 6 & 1 & -4 & = & v_2 \\
-3 & -6 & -9 & -2 & 5 & = & v_3 \\
0 & 1 & 1 & 1 & 0 & = & v_4 \\
1 & 3 & 4 & 3 & 1 & = & v_5
\end{array}
$$

and proceed to reduce it to a "semi-echelon" form. In this reduction we allow ourselves to multiply any row (vector) by a non-zero scalar, or to add any scalar multiple of one row to another row, but we do not allow rows to be interchanged. We shall carry out the reduction row by row.

The first two rows of (7.1) reduce to

$$
(7.2) \qquad
\begin{array}{rrrrrcl}
1 & 2 & 3 & 4 & 5 & = & v_1 \\
0 & 0 & 0 & 1 & 2 & = & v_2'.
\end{array}
$$

Here v_2' is a combination of $\{v_1, v_2\}$; explicitly, $v_2' = -\frac{1}{7}(v_2 - 2v_1)$. The set $\{v_1, v_2'\}$ is independent (see Problem 6.3). Since v_1 and v_2' are combinations of $\{v_1, v_2\}$ it follows (see Problem 6.6) that $\{v_1, v_2\}$ are independent.

Now we change the third row by adding appropriate multiples of the rows (7.2) so as to get zeros in the first and fourth positions. We find that the third row reduces to

$$
\begin{array}{rrrrrcl}
0 & 0 & 0 & 0 & 0 & = & v_3'.
\end{array}
$$

Since v_3' is a combination of $\{v_1, v_2, v_3\}$ this shows that $\{v_1, v_2, v_3\}$ is a dependent set. We have just seen that $\{v_1, v_2\}$ is independent, and so an application of Theorem 6.4(b) tells us that v_3 is a combination of $\{v_1, v_2\}$. We therefore discard v_3.

The fourth row of (7.1) now reduces to

$$0 \quad 1 \quad 1 \quad 0 \quad -2 \;=\; v_4{}',$$

where $v_4{}'$ is a combination of $\{v_1, v_2{}', v_4\}$, and hence (Theorem 5.2) of $\{v_1, v_2, v_4\}$. An argument similar to that in the second paragraph above then proves that $\{v_1, v_2, v_4\}$ are independent.

Finally, the fifth row of (7.1) reduces to

$$0 \quad 0 \quad 0 \quad 0 \quad 0 \;=\; v_5{}'.$$

As above we see from this that v_5 is a combination of $\{v_1, v_2, v_4\}$, and so we discard v_5.

We are thus left with the basis $S = \{v_1, v_2, v_4\}$ for W, and so the dimension of W is 3.

Some useful relations between bases, dimension, independence, etc. are stated in the following theorem.

Theorem 7.3. Let V be a vector space of dimension n.

(a) A set of fewer than n vectors cannot span V.

(b) A set of more than n vectors in V is dependent.

(c) A set of n independent vectors in V spans V, and the set is a basis of V.

(d) A set of n vectors spanning V is independent, and the set is a basis of V.

(e) Any vector subspace of V has dimension at most n; any proper subspace has dimension less than n.

Proof. (a) This follows at once from Corollary 7.1.

(b) Since any vector in V is a combination of the vectors in any basis $\{u_1, \dots u_n\}$, this follows at once from Theorem 6.3.

(c) Let $\{v_1, \dots, v_n\}$ be independent and let v be any vector in V. By (b), $\{v_1, \dots, v_n, v\}$ are dependent, and it follows from Theorem 6.4(b) that v is a combination of $\{v_1, \dots, v_n\}$. Since v was <u>any</u> vector in V this means that $\{v_1, \dots, v_n\}$ span V. Being independent, $\{v_1, \dots, v_n\}$ is therefore a basis.

(d) Let $\{v_1, \dots, v_n\}$ span V. If $\{v_1, \dots, v_n\}$ are not independent then one of them, which we may call v_n, is a combination of the rest. Thus each v_i, $i=1,\dots,n$, is a combination of $\{v_1, \dots, v_{n-1}\}$, and so, by Corollary 5.1, $\{v_1, \dots, v_{n-1}\}$ span V. This contradicts (a), and so $\{v_1, \dots, v_n\}$ must be independent. Since $\{v_1, \dots, v_n\}$ also spans V, it is a basis.

(e) Let W be a vector subspace of V. Let v_1 be any vector in W. If v_1 does not span W there must be a v_2 in W which is not a combination of v_1. Then by Theorem 6.4(b), $\{v_1, v_2\}$ are independent. If $\{v_1, v_2\}$ does not span W there must be a v_3 in W which is not a combination of $\{v_1, v_2\}$, and by the same argument $\{v_1, v_2, v_3\}$ are independent. Continuing this process we get an independent set $\{v_1, \dots, v_m\}$ in W. But by (b) we must have $m \leqq n$, and so the process must stop. The only way it can stop is for $\{v_1, \dots, v_m\}$ to span W. Then $\{v_1, \dots, v_m\}$ is a basis of W, and W is of dimension $m \leqq n$. If $m = n$, $\{v_1, \dots, v_m\}$ is a basis of V, by (c); and in this case $W = V$ and so W is not a proper subspace of V.

7A. The Steinitz Replacement Theorem.

Theorem 7.1 is critical for the definition of dimension of a vector subspace. Our proof of this key theorem goes back through Theorem 6.3 to Theorem 1.2, about solutions of homogeneous linear equations. Quite apart from the fact that we never gave a complete mathematical proof of Theorem 1.2, it is interesting and sometimes useful to be able to develop the whole theory of vector spaces without recourse to the solution of linear equations ; so that, in fact, the whole theory of linear equations is a consequence of the theory of vector spaces rather than a foundation.

As a basis for this alternate development we use the so-called Steinitz Replacement Theorem, or Exchange Theorem. This theorem uses nothing beyond Theorem 6.1 and so could replace Theorems 6.2 and 6.3.

Theorem 7A.1. Let $\{v_1, \dots, v_n\}$ span V and let $\{u_1, \dots, u_m\}$ be independent vectors in V. Then $n \geqq m$, and we can discard m of the v's, suitably chosen, and replace them by the u's so that the resulting set still spans V.

Proof. Since $\{v_1, \dots, v_n\}$ span V and u_1 is in V, u_1 is a combination of the v's,

$$u_1 = a_1v_1 + \dots + a_nv_n.$$

Now $u_1 \neq 0$, since otherwise $\{u_1, \dots, u_m\}$ would be dependent, and so at least one of the a's is not zero. By renumbering the v's if necessary we can assume $a_1 \neq 0$. Then

$$v_1 = \frac{1}{a_1}(u_1 - a_2v_2 - \dots - a_nv_n).$$

Hence $\{v_1, v_2, \dots, v_n\}$ are combinations of $\{u_1, v_2, \dots, v_n\}$ and so, by Corollary 5.1, $\{u_1, v_2, \dots, v_n\}$ span V. We have thus made the first replacement.

Proceeding similarly,

$$u_2 = b_1u_1 + b_2v_2 + \dots + b_nv_n,$$

or

$$u_2 - b_1u_1 = b_2v_2 + \dots + b_nv_n.$$

As before, we cannot have $u_2 - b_1u_1 = 0$ since this would mean that $\{u_1, \dots, u_n\}$ are dependent. Hence one of $b_2, \dots, b_n$ is different from zero, and by another renumbering if necessary we can assume that $b_2 \neq 0$. Then

$$v_2 = \frac{1}{b_2}(-b_1u_1 + u_2 - b_3v_3 - \dots - b_nv_n)$$

and by the same argument as before $\{u_1, u_2, v_3, \dots, v_n\}$ span V.

We continue in this way, replacing a v by a u. Suppose that we had $n < m$. Then we would exhaust the v's before the u's, $\{u_1, \dots, u_n\}$ would span V, and u_{n+1} would be a combination of $\{u_1, \dots, u_n\}$, contrary to the assumption that $\{u_1, \ldots, u_m\}$ are independent. Hence we must have $n \geqq m$. The exchange process will then terminate when we have exhausted all the u's, and we will have achieved the result stated in the theorem.

Our two previous theorems, 6.3 and 7.1, are corollaries of the Exchange Theorem.

Corollary 7A.1. (Theorem 7.1) Any two bases of the same vector space must contain the same number of vectors.

Proof. If the sets $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_m\}$ of Theorem 7A.1 are bases then $n \geqq m$. But the roles of the u's and the v's can be interchanged in this case, and so we also have $m \geqq n$. Hence $m = n$.

Corollary 7A.2. (Theorem 6.3) If $m > n$ and if each of the m vectors $\{u_1, \ldots, u_m\}$ is a combination of $\{v_1, \ldots, v_n\}$ then the set $\{u_1, \ldots, u_m\}$ is dependent.

Proof. Call V the vector space spanned by $\{v_1, \ldots, v_n\}$. By Theorem 7A.1, if $\{u_1, \ldots, u_m\}$ were independent we would have $n \geqq m$. Since we are given that $n < m$, it follows that the set $\{u_1, \ldots, u_m\}$ is dependent.

Problems

7.1 Find the dimension of the vector subspace of V spanned by each of the following sets and select a basis from the given spanning set.

(a) $\{(1,2,3,4),\ (4,3,2,1),\ (1,1,1,1),\ (0,1,1,0)\}$; $V = V_4$.

(b) $\{(1,1,1,1,1),\ (1,0,1,0,1),\ (0,1,1,1,0),\ (2,1,0,1,2),\ (2,1,0,-1,-2),$
$(1,2,3,4,5)\}$; $V = V_5$.

(c) $\{\sin^2 t,\ \cos^2 t,\ \cos 2t,\ 5\}$; $V = C^\circ$.

7.2 Prove: Any independent set $\{v_1, \ldots, v_m\}$ in a space V of dimension n $(> m)$ can be enlarged to form a basis of V. [Let $\{u_1, \ldots, u_n\}$ be a basis of V, and apply the method of Theorem 7.2 to the set $\{v_1, \ldots, v_m, u_1, \ldots, u_n\}$.]

7.3 Use the above result to enlarge each of the following sets to be a basis of the appropriate V.

(a) $\{(1,1,1,1),\ (1,-1,1,-1)\}$; $V = V_4$.

(b) $\{(1,2,0)\}$; $V = V_3$.

(c) $\{t^2 + t^4,\ t^3 + t^4,\ t^4\}$; $V = P^4$.

7.4 Let U be the subspace of V_4 spanned by $\{(1,0,1,0),\ (0,1,1,0),\ (1,-1,0,0),\ (0,0,1,1)\}$. Find a basis for U containing the vector $(1,1,2,0)$.

7.5 Prove Theorem 7.2 by working backwards on the v's, starting with v_n and discarding any v_i if it is a combination of $\{v_1, \ldots, v_{i-1}\}$.

7.6 What is the dimension of the vector space of $m \times n$ matrices? (cf. Problem 3.4.)

7.7 In each of the following cases find a basis and the dimension of the subspace formed by vectors satisfying the given condition.

(a) Problem 5.1 (a).

(b) Problem 5.1 (c).

(c) Problem 5.2 (a), $n = 2$.

(d) Problem 5.2 (c), $n = 4$.

(e) Problem 5.2 (d), $n = 7$.

(f) Vectors $\vec{x} = (x_1, x_2, x_3)$ such that $x_1 + x_2 + x_3 = 0$.

(g) Vectors $\vec{x} = (x_1, x_2, x_3, x_4)$ such that $x_1 + x_2 + x_3 + x_4 = 0$ and $x_1 - x_2 + x_3 - x_4 = 0$.

7.8 Show that all functions of the form $A \sin (t+\theta)$, where A and θ are arbitrary constants, form a 2-dimensional subspace of C^{∞}.

8. Isomorphism. Change of Basis.

The reader may have noticed that we have been placing considerable emphasis on the special vector spaces of n-tuples, V_n, especially in examples and problems. In the present section we shall justify this emphasis by showing that V_n is a prototype of any n-dimensional vector space, and that the techniques for handling problems in V_n can be applied to any finite-dimensional vector spaces. We first introduce a concept which, in more general forms, is basic to many branches of mathematics.

Definition 8.1. Let U and V be two vector spaces between whose elements there is a one-to-one correspondence, $u \longleftrightarrow v$. This correspondence is said to be an isomorphism between U and V if it is preserved under the two basic vector operations, addition and multiplication by scalars; that is, if it is true that whenever $u \longleftrightarrow v$ and $u' \longleftrightarrow v'$ then necessarily $u+u' \longleftrightarrow v+v'$ and $cu \longleftrightarrow cv$. The vector spaces U and V are said to be <u>isomorphic</u>.

The proof of the following theorem is left to the reader.

<u>Theorem 8.1</u>. If U and V are isomorphic and V and W are isomorphic then U and W are isomorphic.

Since linear combination, dependence, subspaces, bases, dimension, etc. are defined in terms of addition and multiplication by scalars it follows that isomorphic vector spaces behave alike as regards these properties. In particular they have the same dimension. The converse follows from the next theorem.

<u>Theorem 8.2</u>. Any n-dimensional vector space is isomorphic to V_n.

<u>Proof</u>. Let V be an n-dimensional vector space and let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis of V. Any vector $\vec{v}$ in V can be expressed as a linear combination of the basis, in a unique way

$$\vec{v} = a_1\vec{v}_1 + \dots + a_n\vec{v}_n. \tag{8.1}$$

Moreover, since $\{\vec{v}_1, \dots, \vec{v}_n\}$ are independent it follows from Theorem 6.4(c) that the n-tuple $(a_1, \dots, a_n)$ is uniquely determined by $\vec{v}$. Conversely, if an n-tuple $(a_1, \dots, a_n)$ is given, equation (8.1) determines a unique vector $\vec{v}$ of V. Hence the relation

$$\vec{v} \longleftrightarrow (a_1, \dots, a_n)$$

determined by (8.1) is a one-to-one correspondence between V and V_n. To show that this correspondence is an isomorphism consider another vector $\vec{v}'$ in V and its corresponding vector $(a_1', \ldots, a_n')$ in V_n determined by

$$\vec{v}' = a_1'\vec{v}_1 + \ldots + a_n'\vec{v}_n. \tag{8.2}$$

From (8.1) and (8.2) we get

$$\vec{v}+\vec{v}' = (a_1+a_1')\vec{v}_1 + \ldots + (a_n+a_n')\vec{v}_n,$$

and, if c is any scalar,

$$c\vec{v} = (ca_1)\vec{v}_1 + \ldots + (ca_n)\vec{v}_n.$$

These equations say that

$$\vec{v}+\vec{v}' \longleftrightarrow (a_1, \ldots, a_n) + (a_1', \ldots, a_n'),$$

$$c\vec{v} \longleftrightarrow c(a_1, \ldots, a_n),$$

and so the correspondence is an isomorphism.

Corollary 8.1. Any two n-dimensional vector spaces are isomorphic.

Proof. Use Theorem 8.1.

Example 8.1. We have seen that $\{1, t, t^2, \ldots, t^n\}$ is a natural basis for P^n. The normal way of writing a polynomial

$$p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n$$

suggests at once the obvious isomorphism

$$p(t) \longleftrightarrow (a_0, a_1, \ldots, a_n)$$

between P^n and V_{n+1}.

Example 8.2. Let V consist of physical vectors lying in the plane, and let $\vec{i}$ and $\vec{j}$ be perpendicular unit vectors in this plane. Then any vector $\vec{v}$ of V is expressible in the form

$$\vec{v} = a\vec{i} + b\vec{j}.$$

Since $\vec{i}$ and $\vec{j}$ are independent the set $\{\vec{i}, \vec{j}\}$ is a basis and the correspondence

$$\vec{v} \longleftrightarrow (a, b)$$

is an isomorphism between V and V_2.

This process is just what we did at the beginning of Section 3, and the reader should review that earlier discussion in the light of our present, more precise, point of view.

From the physical example we carry over to the general case the term "component," calling the coefficients $a_1, \dots, a_n$ of equation (8.1) the components of $\vec{v}$ with respect to the basis $\{\vec{v}_1, \dots, \vec{v}_n\}$. The isomorphism of V and V_n can thus be described as the correspondence between a vector and the ordered set of its components.

It is highly important to note that the components of a vector depend on the chosen basis. If we pick a new basis we get a new isomorphism, and the components of a vector can be expected to change. A natural question to ask is then "How does a change of basis affect the components?"

To answer this question consider two bases of V, $\{\vec{v}_1, \dots, \vec{v}_n\}$ and $\{\vec{v}'_1, \dots, \vec{v}'_n\}$. If $\vec{u}$ is any vector of V we have

(8.3) $$\vec{u} = \sum_{i=1}^{n} a_i \vec{v}_i$$

and also

(8.4) $$\vec{u} = \sum_{i=1}^{n} a_i' \vec{v}_i'.$$

What we want to know is how the a_i' are related to the a_i. To discover this we must obviously know how the $\vec{v}_i'$ are related to the $\vec{v}_i$'s, and one way of specifying this is to express each $\vec{v}_i$ in terms of the basis $\{\vec{v}_1', \ldots, \vec{v}_n'\}$, thus

(8.5) $$\vec{v}_i = \sum_{j=1}^{n} c_{ji} \vec{v}_j', \qquad i = 1, \ldots, n.$$

We can then substitute from (8.5) into (8.3) to get

(8.6) $$\begin{aligned} \vec{u} &= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} c_{ji} \vec{v}_j' \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ji} a_i \vec{v}_j' \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} c_{ji} a_i \vec{v}_j' \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} c_{ji} a_i\right) \vec{v}_j'. \end{aligned}$$

We can equally well write (8.4) as

(8.7) $$\vec{u} = \sum_{j=1}^{n} a_j' \vec{v}_j'.$$

(8.6) and (8.7) give two expansions of $\vec{u}$ as linear combinations of the basis $\{\vec{v}_1', \ldots, \vec{v}_n'\}$. As we have seen above, such an expansion is unique, and so we must have

$$a'_j = \sum_{i=1}^{n} c_{ji} a_i \,, \qquad j = 1, \ldots, n. \tag{8.8}$$

This, then, is our desired relation between the components with respect to the two bases.

Example 8.3. Rotation of axes in a plane. (cf. Thomas, p. 492)

Let $\vec{v}'_1$ and $\vec{v}'_2$ be perpendicular unit vectors forming a basis for the space V of vectors from a point O and lying in a plane. Let $\{\vec{v}_1, \vec{v}_2\}$ be a similar basis for V making an angle θ with the basis $\{\vec{v}'_1, \vec{v}'_2\}$. (See Figure 8.1.) Considering the components of the unit vectors $\vec{v}_1$ and $\vec{v}_2$ with respect to the basis $\{\vec{v}'_1, \vec{v}'_2\}$, we have for (8.5),

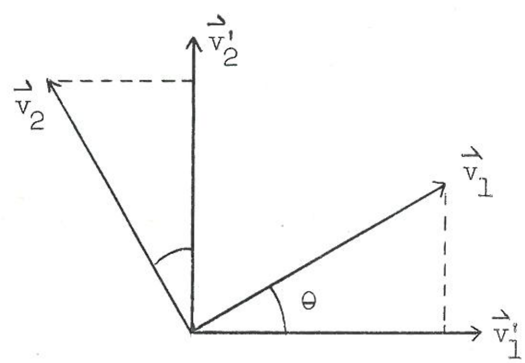


Figure 8.1

$$\vec{v}_1 = \vec{v}'_1 \cos\theta + \vec{v}'_2 \sin\theta \,,$$

$$\vec{v}_2 = -\vec{v}'_1 \sin\theta + \vec{v}'_2 \cos\theta \,.$$

If $\vec{u} = a_1\vec{v}_1 + a_2\vec{v}_2 = a'_1\vec{v}'_1 + a'_2\vec{v}'_2$, then replacing $\vec{v}_1$ and $\vec{v}_2$ the values above gives

$$\vec{u} = (a_1\cos\theta - a_2\sin\theta)\vec{v}'_1 + (a_1\sin\theta + a_2\cos\theta)\vec{v}'_2$$

$$= a'_1\vec{v}'_1 + a'_2\vec{v}'_2,$$

and so

$$a_1' = a_1 \cos \theta - a_2 \sin \theta ,$$

(8.9)

$$a_2' = a_1 \sin \theta + a_2 \cos \theta .$$

A more extensive discussion of the relation (8.8) is deferred until we have the appropriate machinery of matrix algebra at our disposal.

Problems

8.1 (a) Show that $B_1 = \{ (t+1)^2, (t+2)^2, (t+3)^2 \}$ is a basis of P^2.

(b) Express each vector of the basis $B_2 = \{ 1, t, t^2 \}$ of P^2 in terms of the basis B_1.

Ans. $1 = \frac{1}{2} (t+1)^2 - (t+2)^2 + \frac{1}{2} (t+3)^2,$

$t = - \frac{5}{4} (t+1)^2 + 2(t+2)^2 - \frac{3}{4} (t+3)^2,$

$t^2 = 3(t+1)^2 - 3(t+2)^2 + (t+3)^2.$

(c) Express each of the following members of P^2 in terms of the basis B_1.

(i) $1 + 2t + 3t^2,$

(ii) $(t+4)^2,$

(iii) $a_0 + a_1 t + a_2 t^2.$

8.2 In deriving equations (8.9) we needed to use only the definitions of the sine and cosine functions. In particular we did not need the addition formulas for the sine and cosine. Show that these addition formulas can be <u>derived</u> from equations (8.9) by introducing the angle ϕ between $\vec{v}_1$ and $\vec{u}$.

LINEAR TRANSFORMATIONS AND MATRICES

9. Examples of Linear Transformations.

Example 9.1. Gizmo-Dynamics, Inc. produces four types of gizmos: Mark I, Mark II, Mark III, Mark IV. All the gizmos consist of nuts, screws and bolts; in particular Mark I consists of 37 nuts, 41 screws, and 53 bolts. We represent the constituents of Mark I by the vector (37, 41, 53). Similarly the constituents of Mark II, III, IV are represented respectively by the vectors (129, 312, 252), (6329, 7175, 8251) and (2,3,2). (Some design improvements were incorporated in later models.) The corporation plans to produce x_1 Mark I's, x_2 Mark II's, x_3 Mark III's and x_4 Mark IV's. Under this production plan how many nuts, bolts and screws should be ordered? Let y_1 represent the number of nuts to be ordered, y_2 the number of screws, and y_3 the number of bolts. Then from the given data we find

$$\begin{aligned} 37x_1 + 129x_2 + 6329x_3 + 2x_4 &= y_1, \\ 41x_1 + 312x_2 + 7175x_3 + 3x_4 &= y_2, \\ 53x_1 + 252x_2 + 8251x_3 + 2x_4 &= y_3. \end{aligned} \tag{9.1}$$

Note that here, in contrast to Section 1, we are regarding the x_i's as given and using (9.1) to determine the y_i's. Thus we may say that <u>(9.1) defines the V_3 vector y as a function of the V_4 vector x</u>. We shall pursue this point further, but before doing so, we shall find it convenient to use a more tractable notation. To this end we make some definitions.

<u>Definition 9.1</u>. An $m \times n$ <u>matrix</u> is a rectangular array of numbers having m rows and n columns. The number appearing in the i^{th} row and j^{th} column of a matrix A is denoted by a_{ij}. The matrix whose entry in the i^{th} row and j^{th} column is a_{ij} is denoted by (a_{ij}).

The coefficients appearing on the left side of (9.1) may be represented in the form of a 3×4 matrix, which we shall call A;

$$\mathbf{A} = \begin{pmatrix} 37 & 129 & 6329 & 2 \\ 41 & 312 & 7175 & 3 \\ 53 & 252 & 8251 & 2 \end{pmatrix}.$$

Let a_{ij} be the number in the i^{th} row and j^{th} column of A; e.g., $a_{23} = 7175$, $a_{11} = 37$. Then (9.1) can be written in the form

$$\sum_{j=1}^{4} a_{ij}x_j = y_i, \qquad i = 1,2,3. \tag{9.2}$$

More generally, any $m \times n$ matrix A defines a function from V_n to V_m (i.e. a rule whereby given an n-component vector $x = (x_1, \ldots, x_n)$ an m-component vector $y = (y_1, \ldots, y_m)$ is determined) by the equation

(9.3) $$\sum_{j=1}^{n} a_{ij}x_j = y_i \,, \qquad i = 1,2,\ldots,m.$$

To emphasize the functional role of the matrix A, and to simplify the notation still further, we can write equation (9.3) in the form

(9.4) $$A(x) = y.$$

Note that $A(x)$ is defined only when the number of components in the vector x is equal to the number of columns of A, and that the number of components of $A(x)$ is equal to the number of rows of A.

The function $A(x)$ is multiplicative. By this we mean that if a is any scalar then

(9.5) $$A(ax) = aA(x).$$

This is easily verified from (9.3). The physical interpretation is simple. If we double the "production plan" we double the number of nuts, screws and bolts required. If we triple the production plan we triple the number of nuts, screws and bolts required, and so on.

Secondly the function is additive. This means that

(9.6) $$A(x + x') = A(x) + A(x').$$

This is also easily verified from (9.3). For the physical interpretation let x be the production plan based on satisfying the national market and x' a production plan to supply a newly opened foreign market. Then the total number of nuts, screws and bolts required is the sum of that required for the national market plus that required for the foreign market, i.e. $A(x) + A(x')$.

The multiplicative and additive properties expressed in (9.5) and (9.6) are usually combined to give the statement that the function $\mathbf{A}(x)$ is <u>linear</u>:

$$A(ax + bx') = aA(x) + bA(x'). \tag{9.7}$$

We shall return to Gizmo-Dynamics, Inc. in Section 10.

<u>Example 9.2</u>. Consider the flow of water in the network shown in Figure 9.1. This time we regard the flows $(q_1, q_2, q_3, q_4, q_5, q_6)$ as given. These flows determine

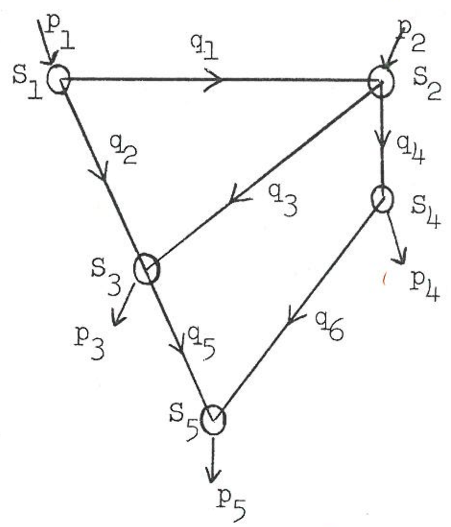


Figure 9.1

the flow rates $(p_1, p_2, p_3, p_4, p_5)$ at the stations S_1, S_2, S_3, S_4, S_5 by the equations:

$$\begin{aligned}
q_1 + q_2 \qquad\qquad\qquad\qquad &= p_1, \\
-q_1 \qquad + q_3 + q_4 \qquad\qquad &= p_2, \\
- q_2 - q_3 \qquad + q_5 \qquad &= -p_3, \\
- q_4 \qquad + q_6 &= -p_4, \\
- q_5 - q_6 &= -p_5.
\end{aligned} \tag{9.8}$$

Let q and p be the vectors

$$q = (q_1, q_2, q_3, q_4, q_5, q_6),$$

$$p = (p_1, p_2, -p_3, -p_4, -p_5),$$

and let A be the 5×6 matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

Then (9.8) can be written in the form

$$A(q) = p.$$

Here the multiplicative property $A(aq) = aA(q)$ means, for example, that if all flows in the lines are doubled, the effects at the stations are doubled. If all flows are tripled, the effects at the stations are tripled. If all flows are reversed in directions, the effects at the stations are reversed, etc.

The additive property $A(q + q') = A(q) + A(q') = p + p'$, where $p = A(q)$ and $p' = A(q')$, means that if two flows are superposed, then the effects at the stations are given by the sum of the effects of the two individual flows. This is what we have called the superposition property.

Example 9.3. If x and y are vectors in the plane, i.e. in V_2, (9.3) has the form

$$
(9.9) \qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 &= y_1, \\ a_{21}x_1 + a_{22}x_2 &= y_2. \end{aligned}
$$

We can think of (x_1, x_2) and (y_1, y_2) either as vectors or as the coordinates of points, the tips of vectors. Then (9.9) defines a transformation of the plane that takes any vector, or point, (x_1, x_2) into the vector, or point, (y_1, y_2). Many special cases of (9.9) have simple geometric meanings.

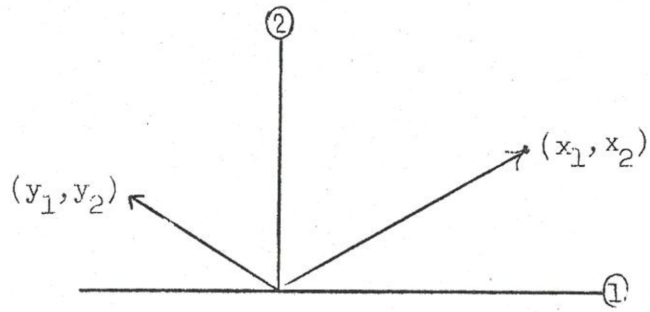


Figure 9.2

(1) Projection on the x-axis: $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

(2) Reflection in the y-axis: $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

(3) Rotation through 90°: $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

(4) Rotation through an angle θ: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

(See Problem 9.2.)

(5) Shear: $A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$. (See Problem 9.4.)

Example 9.4. If $p(t)$ is a polynomial of degree n, its derivative $\frac{dp}{dt}$ or $Dp(t)$ is a polynomial of degree $n-1$. It is well known that the derivative operator is linear, that is,

$$D(ap(t) + bq(t)) = aDp(t) + bDq(t).$$

Hence D is a linear operator from the vector space P^n to the vector space P^{n-1}. If we do not care to worry about the effect of D on the degree of a polynomial we can regard D as an operator from P^{∞} to P^{∞}.

These examples illustrate various types and aspects of functional relations between the elements of two vector spaces (which may coincide, as in Example 9.3). Such relations are generally called <u>transformations</u>, or <u>operators</u>, or <u>mappings</u>. By far the most important of these transformations are the <u>linear</u> ones, those having property (9.7) or its equivalent in whatever notation is used.

We shall be concerned mostly with linear transformations arising from matrices. A more general discussion will be given in Section 15.

Problems

9.1 Compute $A(x)$, if possible, given A and x:

(a)
$$A = \begin{pmatrix} 3 & 0 & 1 \\ 2 & 0 & 3 \\ 1 & 0 & 5 \end{pmatrix}, \qquad x = (1, -1, 0).$$

(b)
$$A = \begin{pmatrix} 3 & 0 & 2 & 5 \\ -1 & 1 & 2 & 7 \\ 0 & -1 & 2 & -5 \end{pmatrix}, \qquad x = (1, 2, 4, -1).$$

(c)
$$A = \begin{pmatrix} 3 & 0 \\ 7 & 1 \\ 6 & 5 \\ 7 & 3 \end{pmatrix}, \qquad x = (2, -5).$$

(d)
$$A = \begin{pmatrix} 3 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad x = (1, 0, -1).$$

(e)
$$A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & -1 & 3 \\ 4 & 5 & 1 \end{pmatrix}, \qquad x = (x_1, x_2, x_3).$$

9.2 Show that a rotation through an angle θ is the matrix transformation

$$x_1 \cos \theta - x_2 \sin \theta = y_1,$$
$$x_1 \sin \theta + x_2 \cos \theta = y_2.$$

[Hint. In Figure 9.3, $x_1 = r \cos \alpha$, $y_1 = r \cos(\theta + \alpha)$, etc.]

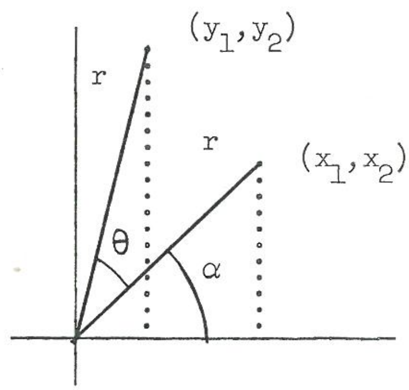


Figure 9.3

9.3 What is the geometric significance of the transformations defined by the following matrices?

(a) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(b) $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$.

(c) $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$.

9.4 Describe in geometric or physical terms the effect of the shear transformation determined by the matrix

$$A = \begin{pmatrix} 1 & 0.1 \\ 0 & 1 \end{pmatrix} .$$

9.5 What must be the form of a linear function $y = f(x)$ where x and y are real varables, i.e. a function such that $f(ax+bx') = af(x) + bf(x')$?

9.6 (a) Show that the derivative operator D defines a linear transformation of C^1 into C^0. Of C^∞ into C^∞.

(b) Given any element $p(t)$ of C^0, we can define an operator L by

$$L(x(t)) = Dx(t) + p(t)\,x(t).$$

Show that L is a linear operator of C^1 into C^0. Relate this to the material in Chapter 1, Section 6.

(c) Define and discuss linear differential operators of second and higher order.

9.7 Prove that if L is a linear operator then

$$L\left(\sum_{i=1}^{k} a_i x_i\right) = \sum_{i=1}^{k} a_i L(x_i).$$

10. Multiplication of Matrices.

We now return to the problems of Gizmo-Dynamics, Inc. (Example 9.1). Suppose that the four types of gizmos (Mark I, Mark II, Mark III and Mark IV) are actually used in five types of missiles. Specifically, suppose missile X-1 uses 20 Mark I gizmos, 10 Mark II's, 5 Mark III's and 0 Mark IV's. We can then represent missile X-1's gizmo requirements by the vector (20, 10, 5, 0). Similarly missiles X-2, X-3, X-4 and X-5 have the gizmo requirements (30, 30, 30, 30), (30, 40, 50, 60), (20, 40, 60, 80,) and (1, 0, 0, 4000).

The president of Gizmo-Dynamics, Inc. has just procured a contract to provide all the gizmos for z_1 of the X-1 missiles, z_2 of the X-2 missiles, z_3 of the X-3 missiles, z_4 of the X-4 missiles and z_5 of the X-5 missiles. How many nuts, screws and bolts should the purchasing department order (as a function of the vector z)?

We recall that in Section 9 we had the relation

$$y = A(x), \tag{10.1}$$

where y is the vector giving the required number of nuts, bolts and screws, x is the gizmo-production-plan vector and A is the matrix given by equation (9.1). This could also be written in the form

$$y_i = \sum_{j=1}^{4} a_{ij}x_j , \qquad i = 1,2,3, \tag{10.2}$$

as in (9.2).

From the data of the present problem we can compute the required gizmo production plan:

$$
\begin{aligned}
x_1 &= 20z_1 + 30z_2 + 30z_3 + 20z_4 + z_5, \\
x_2 &= 10z_1 + 30z_2 + 40z_3 + 40z_4 + 0z_5, \\
x_3 &= 5z_1 + 30z_2 + 50z_3 + 60z_4 + 0z_5, \\
x_4 &= 0z_1 + 30z_2 + 60z_3 + 80z_4 + 4000z_5.
\end{aligned}
\tag{10.3}
$$

This can be written in the form

$$x = B(z), \tag{10.4}$$

where B is the 4×5 matrix

$$B = \begin{pmatrix} 20 & 30 & 30 & 20 & 1 \\ 10 & 30 & 40 & 40 & 0 \\ 5 & 30 & 50 & 60 & 0 \\ 0 & 30 & 60 & 80 & 4000 \end{pmatrix}, \tag{10.5}$$

and z is the vector $(z_1, z_2, z_3, z_4, z_5)$.

Alternately, (10.3) can be written in the form

$$x_j = \sum_{k=1}^{5} b_{jk} z_k , \qquad j = 1,2,3,4, \tag{10.6}$$

where b_{jk} are the elements of B, which is given in (10.5).

We can combine (10.1) and (10.4) to express y as a function of z,

$$y = A(B(z)). \tag{10.7}$$

To write this out explicitly we combine (10.2) and (10.6) to put (10.7) in the form

$$y_i = \sum_{j=1}^{4} a_{ij} \left(\sum_{k=1}^{5} b_{jk} z_k\right) , \qquad i = 1,2,3,$$

or, changing the order of summation,

$$y_i = \sum_{k=1}^{5} \left(\sum_{j=1}^{4} a_{ij} b_{jk}\right) z_k , \qquad i = 1,2,3.$$

Hence we can write

$$(10.8) \qquad y_i = \sum_{k=1}^{5} c_{ik} z_k , \qquad i = 1,2,3,$$

where

$$(10.9) \qquad c_{ik} = \sum_{j=1}^{4} a_{ij} b_{jk} , \qquad i = 1,2,3;\ k = 1,2,3,4,5.$$

We thus see that <u>the successive application of two matrix transformations, B followed by A, is a matrix transformation C defined by equation (10.9)</u>. This important fact is the basis of the algebra of matrices, for it has been found that if we define C to be the <u>product</u> of A and B, $C = AB$, the multiplication of matrices has many of the properties of ordinary multiplication of scalars. We therefore make the following definition.

<u>Definition 10.1</u>. Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{jk})$ be an $n \times r$ matrix. The <u>product</u> $AB = C$ is an $m \times r$ matrix (c_{ik}), where

(10.10) $$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} , \qquad i = 1,2, \ldots ,m; \; k = 1,2, \ldots ,r.$$

Note that the product AB is defined only when the number of columns in A is equal to the number of rows in B.

A scheme for remembering how to compute the product is indicated in Figure 10.1. The scheme is usually stated: "To get AB multiply the rows of A by the columns of B."

$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix}$$

Figure 10.1

<u>Example 10.1.</u> Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} , \qquad B = \begin{pmatrix} 5 & 6 & 7 \\ 8 & 9 & 0 \end{pmatrix} .$$

Then

$$AB = \begin{pmatrix} 1\cdot5 + 2\cdot8 & 1\cdot6 + 2\cdot9 & 1\cdot7 + 2\cdot0 \\ 3\cdot5 + 4\cdot8 & 3\cdot6 + 4\cdot9 & 3\cdot7 + 4\cdot0 \end{pmatrix} = \begin{pmatrix} 21 & 24 & 7 \\ 47 & 54 & 21 \end{pmatrix} .$$

BA is not defined, since the number of columns of B is not equal to the number of rows of A.

Although the matrix product is not <u>commutative</u> (i.e. AB might be defined while BA might not, or, even if both are defined, they might not be equal, as in Problem 10.2), it is nevertheless <u>associative</u>, i.e. $A(BC) = (AB)C$. This may be seen as follows.

We motivated matrix multiplication by pointing out that we may regard the $s \times n$ matrix C as effecting a transformation from V_n to V_s; and that if the $r \times s$ matrix B effects a transformation from V_s to V_r then the product matrix BC effects the composite transformation from V_n to V_r. If also, the $m \times r$ matrix A effects a transformation from V_r to V_m then both $(AB)C$ and $A(BC)$ effect the composite transformation V_n to V_s to V_r to V_m, so that, $(AB)C = A(BC)$.

This is a "conceptual" proof. The reader who prefers a computational proof may proceed as follows. Let A, B, C be $m \times r$, $r \times s$ and $s \times n$ matrices respectively with elements a_{ij}, b_{jk}, c_{kp}. Let $A(BC) = D = (d_{ip})$, and $(AB)C = (e_{ip})$. Then

$$d_{ip} = \sum_{j=1}^{r} a_{ij}\left(\sum_{k=1}^{s} b_{jk}c_{kp}\right) = \sum_{k=1}^{s}\left(\sum_{j=1}^{r} a_{ij}b_{jk}\right)c_{kp} = e_{ip}.$$

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

Let us now return to equations (10.1) and (10.2). If we introduce the 4×1 matrices

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix},$$

(10.2) merely states that

(10.11) $$Y = AX.$$

Thus the functional relation (10.1) can be written in the product form (10.11). Similarly (10.7) becomes $Y = A(BZ)$, and the fact that this can be written as $Y = CZ$, with $C = AB$, is merely a special case of the associative law.

Such considerations are so useful that in working with matrices we shall consider an n-component vector to be simply an $n \times 1$ matrix. If we wish to emphasize this point of view we may refer to a "column vector." Similarly, a $1 \times n$ matrix is called a "row vector." Because of the above considerations the column vector is the more useful.

Since $n \times 1$ matrices (column vectors) form a vector space we might expect that $m \times n$ matrices do so also, provided we give suitable definitions for addition of matrices and multiplication by scalars.

Definition 10.2. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices, and let s be any scalar. We define $C = A + B$ and $D = sA$ as the $m \times n$ matrices (c_{ij}) and (d_{ij}) where $c_{ij} = a_{ij} + b_{ij}$ and $d_{ij} = sa_{ij}$.

We shall also have need of the following concept.

Definition 10.3. The transpose of an $m \times n$ matrix A, designated by A^t, is the $n \times m$ matrix obtained by interchanging rows and columns of A. The element in the i-th row and the j-th column of A^t is the same as the element in the j-th row and i-th column of A; thus, $a^t_{ij} = a_{ji}$.

Definition 10.4. The $m \times n$ matrix whose elements are all equal to zero is called the zero matrix, and is designated by 0.

Example 10.2.

(a) $$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^t = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

(b) The transpose of a column vector is a row vector; e.g. $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}^t = (1 \quad 2 \quad 3)$.

The basic rules of matrix algebra are given in the following theorem.

<u>Theorem 10.1.</u> If a and b are scalars and A, B, C are matrices of appropriate sizes to give meaning to the indicated operations, then

(a) The set of $m \times n$ matrices forms a vector space;

(b) $(AB)C = A(BC)$;

(c) $C(aA + bB) = a(CA) + b(CB)$,

$(aA + bB)C = a(AC) + b(BC)$;

(d) $AO = OA = O$;

(e) $(aA + bB)^t = aA^t + bB^t$;

(f) $(AB)^t = B^tA^t$.

<u>Proof.</u> (a), (c), (d), and (e) are left as exercises for the reader. (For (a), cf. Problem 3.4). (b) has already been proved. To prove (f) let $C = AB$. In terms of components this can be written

$$c_{ji} = \sum_k a_{jk}b_{ki}.$$

Then

$$c^t_{ij} = c_{ji} = \sum_k a_{jk}b_{ki} = \sum_k b^t_{ik}a^t_{kj},$$

which says that $C^t = B^tA^t$.

<u>Example 10.3.</u> To illustrate part (f) of Theorem 10.1 let

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

$$(AB) = \begin{pmatrix} 0 & 1 \\ 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \\ 3 \end{pmatrix}$$

$$(AB)^t = (-1,5,3)$$

$$B^tA^t = (3,-1) \begin{pmatrix} 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} = (-1,5,3) = (AB)^t$$

It is worth noting specifically some algebraic properties of scalars that are not in general true for matrices.

(i) Even when AB and BA both exist and are of the same size, we cannot assume that they are equal. See Problem 10.2.

(ii) We can have $AB = 0$ with neither A nor B equal to zero. See Problem 10.3. As a consequence we cannot conclude from $C \neq 0$ and $AC = BC$ that $A = B$.

The following theorem gives us at least a partial way of overcoming the difficulty mentioned in (ii) above.

Theorem 10.2. If A is an $m \times n$ matrix such that $Av = 0$ for every $n \times 1$ column vector v, then $A = 0$.

Proof. Let e_j designate the $n \times 1$ vector in which the j-th component is 1 and all the other components are 0. Then one finds that the i-th component of Ae_j is the element a_{ij} of A. If $Av = 0$ for every $n \times 1$ vector v then in particular $Ae_j = 0$ for every j, and so $a_{ij} = 0$ for every i and every j. That is, $A = 0$.

Problems

10.1 Compute the indicated matrix products.

(a) $\begin{pmatrix} 1 & 3 & 2 & 0 \\ 0 & -1 & 2 & 1 \\ -1 & 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 2 \\ 0 & 0 \\ 2 & 5 \end{pmatrix}$.

(b) $\begin{pmatrix} 1 & 0 & 6 \\ 2 & 0 & 1 \\ -1 & 1 & -2 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}$.

(c) $\begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & 5 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 6 & 3 & 2 & 1 \\ 2 & -1 & 1 & 1 \end{pmatrix}$.

(d) $(0, 1, 3, 1) \begin{pmatrix} 2 & 1 & 4 \\ 3 & 0 & 2 \\ -1 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix}$.

10.2 Let A be an $m \times n$ matrix and B an $r \times s$ matrix. Note that the product AB is defined if and only if $n = r$, while the product BA is defined if and only if $s = m$. Show that if $AB = BA$ then $m = n = r = s$. Show that even if AB and BA are both defined they might not be equal, by writing down two 2×2 matrices A and B and noting that (unless you are very unlucky) $AB \neq BA$.

10.3 Let $A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix}$. Show that $AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Also compute BA.

10.4 If $A = \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}$, $C = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$,

compute whichever of the following products are possible: AB, BA, BC, CB, $A^2 = AA$, C^2, $A(BC)$, $(AB)C$. Note that the last two are equal.

10.5 Given

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad w = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix},$$

compute whichever of the following are possible: $A + B$, AB, B^2, Bv, v^2, v^tv, vv^t, $Av + Bw$, $Av + w^tB^t$, v^tAv.

10.6 If A and B are $n \times n$ matrices is $(A+B)(A-B) = A^2-B^2$ necessarily true?

10.7 An $n \times n$ matrix A is said to be <u>symmetric</u> if $A^t = A$.

(a) If A and B are symmetric $n \times n$ matrices is AB necessarily symmetric?

(b) If A is symmetric is A^2? A^3? A^m?

10.8 We saw in Problem 10.3 that if A and B are $n \times n$ matrices and $AB = 0$ it does not necessarily follow that $BA = 0$. Prove that it does necessarily follow that $(BA)^2 = 0$.

10.9 If U is an $r \times m$ matrix, and D and F are $r \times 1$ column vectors, express the relation

$$f_h = \sum_{i=1}^{m} \sum_{j=1}^{r} u_{ji} d_j u_{hi} , \qquad h = 1, \ldots ,r,$$

in matrix form. Ans. $F = UU^tD$.

10.10 Using the result of Problem 9.2 show that a rotation through angle θ followed by a rotation through angle ϕ is what it ought to be.

10.11 A rotation followed by a shear is rather complicated, but if we rotate, shear, and then rotate back again we would get something recognizable. Try to predict the answer and then test your prediction by carrying out the algebra.

10.12 Are rotations in 3-space commutative? Test this by setting up the matrices for 90° rotation about the x-axis and about the y-axis, and multiplying them in the two different orders.

10.13 Show that $(ABC)^t = C^tB^tA^t$, and indicate how a similar relation holds for any number of factors. [Hint. $(ABC)^t = ((AB)C)^t = C^t(AB)^t =$ etc.]

10.14 Write subroutines for the following processes:

(a) Given A and v, compute Av.

(b) Given A and B, compute AB.

(c) Given A, compute A^t.

10.15 Prove the following corollaries of Theorem 10.2:

(a) If $\{v_1,v_2, \cdots , v_n\}$ is a basis for n-component column vectors, and if $Av_i = 0$, $i = 1, \cdots , n$, then $A = 0$.

(b) If A and B are $m \times n$ matrices such that $Av = Bv$ for all n-component column vectors v, then $A = B$.

11. Solution of Linear Equations

In this section we return to the study of systems of linear equations. With the theory we have developed we can now get a better understanding of the methods and results of Section 1.

The general system of m linear equations in n variables can be written in the matrix form

$$Ax = b, \tag{11.1}$$

where A is an $m \times n$ matrix and x and b are respectively $n \times 1$ and $m \times 1$ column vectors. Associated with the system (11.1) there is the homogeneous system

$$Ax = 0. \tag{11.2}$$

The basic theorem regarding the solutions of these equations is the following.

Theorem 11.1.

(a) The set of solutions of the homogeneous equation (11.2) forms a vector subspace W of V_n.

(b) If x_p is a particular solution of (11.1) and w is any element of W then $x_p + w$ is a solution of (11.1); conversely, any solution x of (11.1) is expressible in the form $x = x_p + w$ for some w in W.

Proof. (a) The matrix A defines a transformation $A(x) = y$ from V_n to V_m, and we can equally well write (11.2) in the form $A(x) = 0$. We have seen in Section 9 that this transformation is linear, that is,

$$A(x_1 + x_2) = A(x_1) + A(x_2), \qquad A(cx_1) = cA(x_1).$$

It follows that if x_1 and x_2 are both solutions of (11.2) so are $x_1 + x_2$ and cx_1. Hence, by Theorem 5.1, the set of all solutions is a vector subspace of V_n. Let us denote it by W.

(b) If $A(x_p) = b$ and $A(w) = 0$ then, by the linearity, $A(x_p + w) = b + 0 = b$; that is, $x_p + w$ is a solution of (11.1). Conversely, let x be any solution of (11.1), so that $A(x) = b$. Then $A(x - x_p) = b - b = 0$; and so $x - x_p$, being a solution of (11.2), is an element w of W. That is, $x = x_p + w$.

It should be noticed that in this proof we have made no use of special properties of matrices. The proof is still valid if $A(x) = y$ is any linear transformation from a vector space V to a vector space U. We have, in fact, already met this situation in Section 6 of Chapter 1. If $p(t)$ is a given continuous function then the passage from any function $x(t)$ in the space C^1 to the function $q(t)$ in C^0 defined by

$$\frac{d}{dt} x(t) + p(t)x(t) = q(t)$$

is a linear transformation (cf. Problem 9.6). Theorem 11.1 applied to this transformation contains nearly all the results of Theorem 6.1 of Chapter 1.

The general case of Theorem 11.1 still leaves us with two important questions to consider, one for each of the two parts of the theorem:

(A). How do we determine the dimension and construct a basis of the subspace W of solutions of the homogeneous equations?

(B). How do we construct a particular solution, if there is any, of the non-homogeneous equation?

The answers to these questions depend on the nature of the vector spaces and the linear transformation. Chapter 1, Section 6 and Chapter 2, Section 1 were each concerned with these answers, but the methods used were very different in the two cases. The same problem will arise again in Chapter 6, where still other methods will be needed. Indeed, a considerable amount of important mathematics in a variety of fields has developed from attempts to answer these questions for various types of linear operators.

We shall use the elimination method of Section 1 to supply answers to questions (A) and (B) for the matrix equations (11.1) and (11.2). In that section we reduced a system of equations to the echelon and reduced echelon forms by a sequence of elementary operations on the equations. We now define the analogous operations on matrices. For this purpose we regard each row of a matrix as a row vector.

<u>Definition 11.1</u>. There are three <u>elementary</u> <u>row</u> <u>operations</u> on a matrix:

(i) Multiplication of one row vector by a non-zero scalar;

(ii) Addition of a scalar multiple of one row vector to a different row vector;

(iii) Interchange of two row vectors.

<u>Example 11.1</u>. Consider the matrix

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

To illustrate (i) we multiply the first row by the scalar 1/2, to get

$$B = \begin{pmatrix} 1 & 2 & 2 & -1 & 2 & 2 \\ 1 & 2 & 2 & 0 & 1 & 3 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ -2 & -4 & -3 & 2 & -4 & -2 \\ 1 & 2 & 1 & -2 & 3 & 2 \end{pmatrix} .$$

To illustrate (ii) we add -1 times the first row to the second row, 2 times the first row to the fourth row, and -1 times the first row to the fifth row, to get

$$C = \begin{pmatrix} 1 & 2 & 2 & -1 & 2 & 2 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix} .$$

To illustrate (iii) we interchange the second and fourth rows, to get

$$D = \begin{pmatrix} 1 & 2 & 2 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix} .$$

Continuing in this manner we arrive at the echelon form

$$E = \begin{pmatrix} 1 & 2 & 2 & -1 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the reduced echelon form

$$F = \begin{pmatrix} 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Definition 11.2. The row space of an $m \times n$ matrix A is the subspace of V_n spanned by its row vectors; the column space of A is the subspace of V_m spanned by its column vectors. The null space of A is the subspace of V_n consisting of the solutions of $Ax = 0$, (the space designated by W in Theorem 11.1).

There are important relations between these subspaces. To see what these are we shall first examine the row, column, and null spaces of a matrix in reduced echelon form.

Example 11.1, continued.

$$F = \begin{pmatrix} 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The row space is obviously spanned by the set of four non-zero rows $\{v_1, v_2, v_3, v_4\}$. These rows are independent: v_2, for example, cannot be a combination of the others since v_2 has 1 as its third component whereas all the others have 0; and a similar argument holds for v_1, v_3, and v_4. The set $\{v_1, v_2, v_3, v_4\}$ is therefore a basis of the row space, and the dimension of this space is 4.

It is evident that the first, third, fourth, and sixth columns form a basis of the column space. Hence the dimension of the column space is also 4.

We can get the general solution to $Fx = 0$ by solving the first four equations independently for x_1, x_3, x_4, x_6 in terms of the remaining variables x_2, x_5, thus

$$\begin{aligned} x_1 &= -2x_2 - x_5, \\ x_3 &= 0, \\ x_4 &= \phantom{-2x_2 -{}} x_5, \\ x_6 &= 0. \end{aligned}$$

Note that the x's for which we solve correspond to the columns that form the basis of the column space. We can now give the remaining variables arbitrary values, say $x_2 = s_1$ and $x_5 = s_2$, and write the solution in the vector form

$$(11.3) \qquad x = s_1 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s_2 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} .$$

Thus the dimension of the null space is 2.

It is easy to see that any matrix in reduced echelon form can be treated in a similar manner, and so we can make the following general statement.

Theorem 11.2. For a matrix in echelon form the row space and the column space have the same dimension, equal to the number of non-zero rows. The dimension of the null space is equal to the total number of columns minus the dimension of the column space.

We now want to extend these results, as far as possible, to general matrices. This is done by the following theorem.

Theorem 11.3. Row operations on a matrix do not change (a) its row space, (b) its null space, or (c) the dimension of its column space.

Proof. (a) If $v_1, v_2, \dots, v_m$ are the row vectors of A, then the row space of A is spanned by the set

$$S = \{ v_1, v_2, v_3, \dots, v_m \} .$$

Operating on the row vectors by the three elementary operations gives the sets

(i) $\quad S_1 = \{ cv_1, v_2, v_3, \dots , v_m \}$, $\quad c \neq 0,$

(ii) $\quad S_2 = \{ v_1+cv_2, v_2, v_3, \dots , v_m \}$,

(iii) $\quad S_3 = \{ v_2, v_1, v_3, \dots , v_m \}$.

(To avoid complicated notation we have taken the operations as applying to the first one or two rows, but the argument is perfectly general.) It is quite obvious that every vector in S_1 is a combination of vectors in S, and conversely. Hence S_1 and S span the same space. The same conclusion is even more obvious for S_3 and S . Finally, every vector in S_2 is evidently a combination of vectors in S, and by writing $v_1 = (v_1 + cv_2) - cv_2$ we see that the converse is true in this case also.

(b) x is a vector of the null space of A if $Ax = 0$, that is, if

$$\sum_{j=1}^{n} a_{1j}x_j = 0,$$

$$\sum_{j=1}^{n} a_{2j}x_j = 0,$$

$$(11.4) \qquad \dots$$

$$\sum_{j=1}^{n} a_{mj}x_j = 0.$$

If we apply operation (ii) to A the corresponding equations for the new matrix A' are

$$\begin{aligned}
&\sum_{j=1}^{n} (a_{1j} + ca_{2j})x_j = 0, \\
&\sum_{j=1}^{n} a_{2j}x_j = 0, \\
(11.5)\qquad &\quad \cdots \\
&\sum_{j=1}^{n} a_{mj}x_j = 0.
\end{aligned}$$

In writing the first of equations (11.5) as

$$\sum_{j=1}^{n} a_{1j}x_j + c\sum_{j=1}^{n} a_{2j}x_j = 0,$$

it is evident that any solution x of the system (11.4) is a solution of (11.5) and conversely; that is, the null spaces of A and A' are the same. In a similar manner the same conclusion can be drawn for elementary operations (i) and (iii).

(c) The column vectors of A are

$$u_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}, \quad j = 1, \cdots, n.$$

The column space of A is spanned by $\{u_1, \cdots, u_n\}$, and we obtain a basis for this space by discarding those u_j which are linear combinations of their predecessors, as in the proof of Theorem 7.2. Thus u_k is discarded if

$$u_k = \sum_{j<k} b_j u_j. \tag{11.6}$$

If A' is obtained from A by elementary operation (ii) the column vectors of A' are

$$u'_j = \begin{pmatrix} a_{1j} + c\, a_{2j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} ,$$

and it is easy to check that if (11.6) holds then also

$$u'_k = \sum_{j<k} b_j u'_j$$

and conversely. Thus the discarding process for the sets $\{u_1, \cdots, u_n\}$ and $\{u'_1, \cdots, u'_n\}$ are exactly the same, and so the dimensions of the column spaces of A and A' are the same. Operations (i) and (iii) can be treated similarly.

By combining these two theorems, and the fact that any matrix can be changed to reduced echelon form by a series of elementary row operations, we get the following properties of a general matrix.

Corollary 11.1. The row space and the column space of a matrix have the same dimension.

Definition 11.3. The common dimension of the row space and the column space of a matrix is called the rank of the matrix.

Corollary 11.2. A matrix and its transpose have the same rank.

Corollary 11.3. The rank of a matrix is the number of non-zero rows in its echelon (or reduced echelon) form, and these rows are a basis of the row space of the matrix.

Corollary 11.4. The null space of an $m \times n$ matrix of rank r has dimension $n - r$, and a basis is easily obtainable from the reduced echelon form.

Example 11.1, concluded. The rank of A is 4. As a basis of the row space of A we can take either $\{(1,2,2,-1,2,2), (0,0,1,0,0,2), (0,0,0,1,-1,2), (0,0,0,0,0,1)\}$ from E or $\{(1,2,0,0,1,0), (0,0,1,0,0,0), (0,0,0,1,-1,0), (0,0,0,0,0,1)\}$ from F. The general solution, (11.3), of $Fx = 0$ is also the general solution of $Ax = 0$.

Example 11.2. Find the dimension and a basis of the subspace of V_5 spanned by $\{(1,2,3,4,5), (2,4,6,1,-4), (-3,-6,-9,-2,5), (0,1,1,1,0), (1,3,4,3,1)\}$.

We need merely to set up the matrix having these five vectors as rows and reduce it to echelon form (reduced echelon form is not required). One solution (it is not unique) is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

giving the dimension 3 and the basis $\{(1,2,3,4,5), (0,1,1,1,0), (0,0,0,1,2)\}$.

This example should be compared with Example 7.3. The methods are nearly identical, but the earlier case required more care, as it asked for a basis to be chosen from the given set of vectors, whereas here we allow any basis.

Corollary 11.4 answers question (A) for a general matrix. An answer to question (B) is now easy to give.

Corollary 11.5. The equation $Ax = b$ has a solution if and only if the vector b is in the column space of A.

Proof. In the notation of part (c) of the proof of Theorem 11.3 the equation $Ax = b$ can be written in the form

$$(11.7) \qquad \sum_{j=1}^{n} x_j u_j = b.$$

That is, if x is a solution of $Ax = b$ then (11.7) says that b lies in the subspace spanned by the column vectors $\{u_1, \dots, u_n\}$. Conversely, if b lies in this subspace then there must be x_i's such that (11.7) is true, and the vector x with these components is a solution of $Ax = b$.

This result is usually stated in a different form. We define A to be <u>the matrix</u> of the equation $Ax = b$, and the matrix obtained by adjoining b to A as an extra column as the <u>augmented matrix</u> of the equation. The augmented matrix is often designated by (A,b). The following statement is easily seen to be equivalent to Corollary 11.5.

Corollary 11.6. $Ax = b$ has a solution if and only if the rank of the augmented matrix equals the rank of the matrix of the equation.

Example 11.3. Consider the equation $Ax = b$ with

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

and for the two values of b,

$$b_1 = \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} \quad \text{and} \quad b_2 = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} .$$

In the first case the reduction of A to echelon form reduces the augmented matrix (A, b_1) to

$$\begin{pmatrix} 1 & -1 & 1 & 2 \\ 0 & 1 & -2 & -2 \\ 0 & 0 & 0 & 5 \end{pmatrix} .$$

The fourth column is obviously not a combination of the other three, and hence by Corollary 11.5 the equation $Ax = b_1$ has no solution. Equally well one can apply Corollary 11.6, noting that the rank of A is 2 and of (A, b_1) is 3.

On the other hand, (A, b_2) reduces to

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

and by either Corollary 11.5 or 11.6 we see that $Ax = b_2$ has a solution.

A particular solution of $Ax = b$, if one exists, is easily obtainable from the reduced echelon form by setting to zero all those variables whose corresponding columns do not constitute the chosen basis of the column space. Thus in Example 11.1, a particular solution of

$$Fx = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 0 \end{pmatrix} \quad \text{is} \quad x_p = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{pmatrix} .$$

Of course a particular solution is not unique and in special cases there may be easier ways to get one.

We need one final statement, to relate a particular solution of the original equation to that of the echelon form. This is

Theorem 11.4. The set of solutions of $Ax = b$ is unchanged by row operations on the augmented matrix (A,b).

The proof is almost identical with that of Theorem 11.3(b). We need merely put $b_1, b_2, \ldots, b_m$ on the right hand sides of equations (11.4) and then carry through the same argument. (Note that the set of solution of $Ax = b$ is not a vector subspace if $b \neq \vec{0}$, but this does not affect the proof.)

Example 11.3, continued. To get the general solution of $Ax = b_2$, continue to the reduced echelon form

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

A particular solution is

$$x_p = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} .$$

Since the rank of A is 2, and $n = 3$, the dimension of the null space is $3 - 2 = 1$, and so there is one independent solution of the homogeneous equation, namely

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} .$$

Hence by Theorem 11.1 all solutions of $Ax = b_2$ have the form

$$x = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} ,$$

t being any scalar.

Since the vectors x are in V_3 a good pictorial representation of the solutions can be given. (Figure 11.1.) The null space is a line through the origin (a 1-dimensional vector subspace) and the tips of the solution vectors x lie on a line parallel to this, determined by any particular solution x_p.

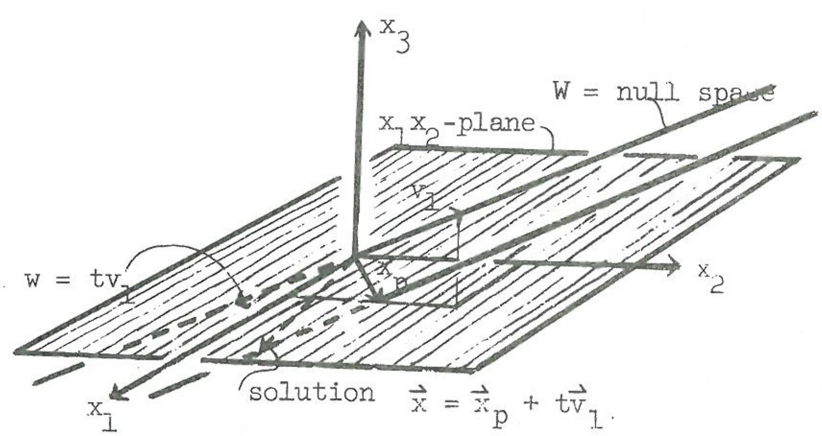


Figure 11.1

Figure 11.2 illustrates the case in V_3 when the null space is a plane (rank of A is 1). Here there are two independent solutions of the homogeneous equation, and the set of solutions of the non-homogeneous equation are the vectors to the points in a plane parallel to the null space.

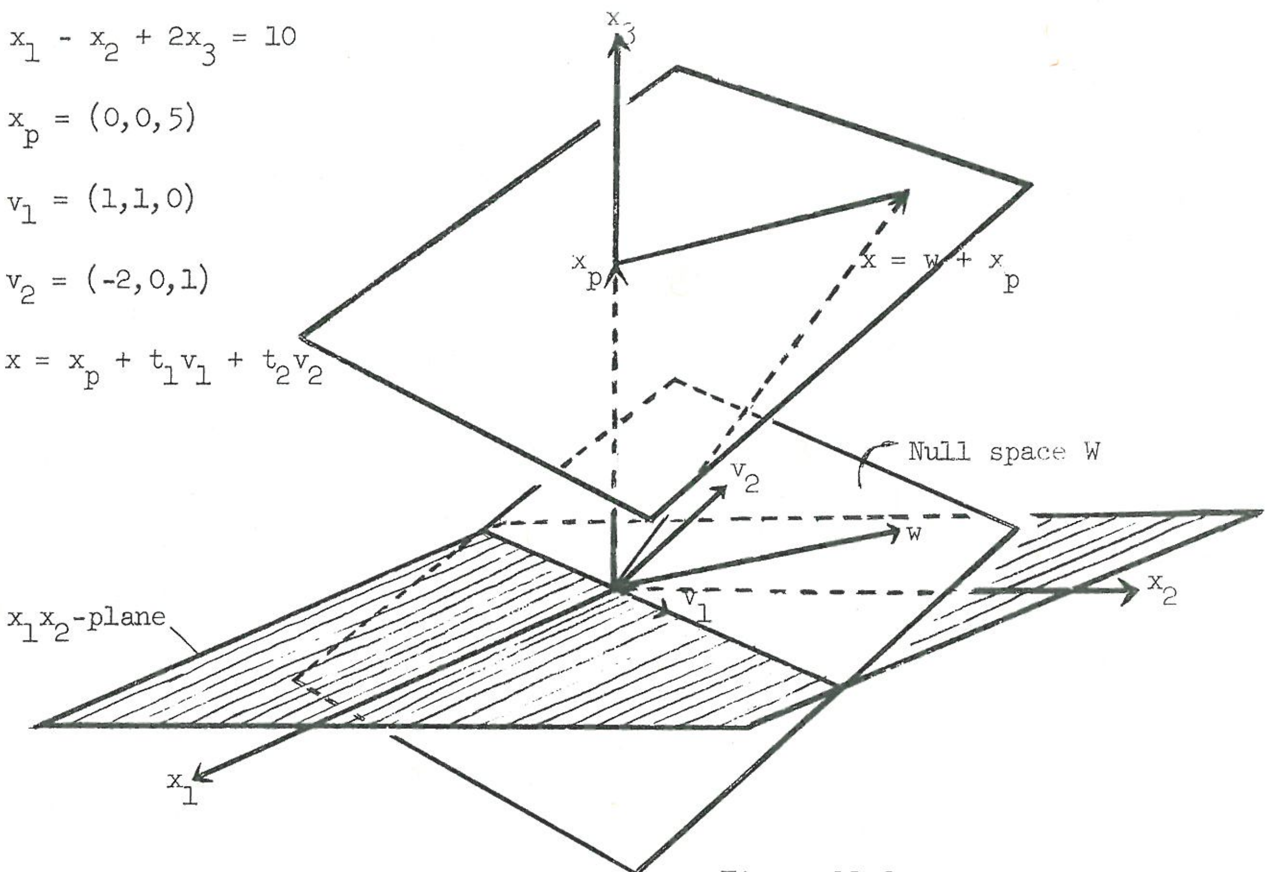


Figure 11.2

Problems

1.1 Find the rank of each of the following matrices.

(a) $\begin{pmatrix} -1 & 0 & 1 & 3 \\ 1 & 2 & 0 & 4 \\ 0 & 1 & 2 & 0 \end{pmatrix}$. (b) $\begin{pmatrix} 1 & 2 & 4 \\ -1 & 2 & 3 \\ -2 & 0 & -1 \end{pmatrix}$.

(c) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$. (d) $\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$.

11.2 Using the data in Problem 7.1(a) and (b), find a basis for each of the subspaces and the dimensions of the subspaces.

11.3 Find the general solution of each of the following systems of equations.

(a)
$$\begin{array}{rcrcrcrl} 2x_1 & + & 2x_2 & - & x_3 & = & 3, & \qquad = 2, \\ x_1 & - & 2x_2 & + & 2x_3 & = & 0, & \qquad = 1, \\ 3x_1 & & & + & x_3 & = & 3, & \qquad = 0, \\ x_1 & - & 4x_2 & - & x_3 & = & -1, & \qquad = 1. \end{array}$$

(b)
$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= 0, \\ 2x_1 + 3x_2 + 4x_3 + 5x_4 &= 0, \\ 3x_1 + 4x_2 + 5x_3 + 6x_4 &= 0, \\ 4x_1 + 5x_2 + 6x_3 + 7x_4 &= 0. \end{aligned}$$

Ans. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = s_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + s_2 \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix}$.

(c)
$$\begin{aligned} 3x_1 + 2x_2 - x_3 + x_4 &= 2, \\ 5x_1 - 3x_2 + x_3 - x_4 &= 1, \\ 2x_1 - 5x_2 + 2x_3 - 2x_4 &= -1. \end{aligned}$$

Ans. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = s_1 \begin{pmatrix} 1 \\ 8 \\ 19 \\ 0 \end{pmatrix} + s_2 \begin{pmatrix} -1 \\ -8 \\ 0 \\ 19 \end{pmatrix} + \frac{1}{19} \begin{pmatrix} 8 \\ 7 \\ 0 \\ 0 \end{pmatrix}$.

(d)
$$\begin{aligned} x_1 + x_2 - 2x_3 \phantom{{}-3x_4} + x_5 &= 4, \\ 2x_1 + x_2 \phantom{{}-2x_3} - 3x_4 - 2x_5 &= -10, \\ x_1 \phantom{{}+x_2} + 2x_3 - x_4 + 2x_5 &= 5, \\ x_1 - x_2 + 6x_3 + 2x_4 + 3x_5 &= 14, \\ 3x_1 + 2x_2 - 2x_3 - x_4 - 2x_5 &= -5. \end{aligned}$$

Ans. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s \begin{pmatrix} -2 \\ 4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 3 \end{pmatrix}$.

11.4 If A is an $m \times n$ matrix of rank r prove the following:

(a) $r \leqq m$, $r \leqq n$.

(b) If $r = m$, $Ax = b$ has a solution for any b.

(c) If $r < n$, $Ax = 0$ has a non-trivial solution.

11.5 Rewrite Theorem 1.3, giving the precise conditions under which each of the cases occurs.

11.6 Write a program that, given an $n \times n$ matrix A and an $n \times 1$ column vector b, will either find the unique solution of $Ax = b$ or state that there is no unique solution.

11.7 Write a program to find the general solution of any equation $Ax = b$.

12. Matrix Equations. Inverse Matrices.

The conclusions of Section 11 can be applied to a somewhat more general type of equation than (11.1), thereby leading to further useful properties of matrices. In equation (11.1) let us replace the one-column matrices x and b by p-column matrices X and B, and ask for solutions of

$$AX = B. \tag{12.1}$$

To investigate this problem we first note the following fact:

If $AB = C$ then the k-th column vector of C is A times the k-th column vector of B.

This follows at once from the definition, equation (10.10), of the product AB, by simply regarding k as fixed.

Hence (12.1) has a solution if and only if each of the equations

$$(12.2) \qquad \mathbf{A}x = z_k , \qquad k = 1, \dots ,p$$

has a solution, where $z_1, \dots ,z_p$ are the column vectors of B. By Corollary 11.5 this is true if and only if each z_k lies in the column space of **A**; hence, by Corollary 11.6, if and only if the rank of the augmented matrix (A,B) equals the rank of A.

The important conclusions to be drawn from these considerations are contained in the following theorem.

Theorem 12.1. If AB = C, then

(a) the column space of C is contained within the column space of **A**;

(b) the row space of C is contained within the row space of B;

(c) the rank of C cannot exceed the rank of either **A** or B.

Proof. (a) If AB = C then the equation AX = C has a solution (namely, X = B), and by the above remarks every column of C is a combination of columns of **A**. Hence every vector of the column space of C lies in the column space of **A** (Theorem 5.2).

(b) If AB = C then $B^tA^t = C^t$. By (a), the column space of C^t is contained within the column space of B^t. Since the row space of a matrix is the column space of its transpose, this proves (b).

(c) From (a) it follows that the vectors of a basis of the column space of C must be combinations of the vectors of a basis of the column space of A, and so the rank of C cannot exceed the rank of A.

Similarly from (b) it follows that the rank of C cannot exceed the rank of B.

Of the various applications of Theorem 12.1 perhaps the most important is to the algebra of $n \times n$ matrices, some of which we will now develop. For the rest of this section the word "matrix" will refer to an $n \times n$ matrix, n having a fixed value throughout, and a "vector" will be an $n \times 1$ column matrix. This restriction has two useful consequences. First, the set of matrices is closed under the operations of multiplication and transpose; i.e. if A and B are matrices so are AB and A^t. Secondly, the linear transformation $Ax = y$ is a transformation from V_n to itself; the full importance of this will become apparent only in Chapter 5.

<u>Definitions 12.1.</u>

(a) The elements a_{ii} form the <u>main diagonal</u> of $A = (a_{ij})$.

(b) A is a <u>diagonal matrix</u> if all the elements not on the main diagonal are zero.

(c) The <u>identity</u> matrix I is the diagonal matrix each of whose diagonal elements is 1. That is, $I = (\delta_{ij})$, where

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j. \end{cases} \tag{12.3}$$

<u>Example 12.1.</u> For $n = 3$:

(a) The main diagonal of $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \\ 3 & 1 & -2 \end{pmatrix}$ is 1,0,-2.

(b) $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is a diagonal matrix.

(c) $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Theorem 12.2. The identity matrix I has the following properties: (a) the rank of I is n; (b) $AI = IA = A$ for every matrix A; (c) $Ix = x$ for every vector x; (d) I is the only matrix having property (b).

Proof. (a) is obvious, and (b) and (c) follow from the definition of the product of two matrices. To prove (d) suppose that some matrix E had the same property, $AE = EA = A$ for every matrix A. Choosing A to be I in the relation gives, in part, $IE = I$; choosing A to be E in (b) gives, in part, $IE = E$. Hence $E = I$.

We can add, subtract, and multiply matrices - can we divide them? The process of division is defined basically in terms of multiplication; to divide scalars, b by a, we seek a solution of the equation $ax = b$. For matrices we are accordingly led right back to equation (12.1), where now A, B, and X are all $n \times n$ matrices, and we have the answer that division is possible provided the column space of B is contained in the column space of A. However, we must be careful in our terminology, for division of B by A could equally well mean solution of the equation $XA = B$, which is by no means the same thing as $AX = B$. By writing $XA = B$ in the form $A^tX^t = B^t$ we find that this equation has a solution provided the row space of B is contained in the row space of A.

Such considerations as these are sometimes unavoidable, but there is one important case when we can dispense with all worry about columns or rows of B. If A is of rank n then the column space and the row space of A each consist of the entire space V_n and so automatically contain all columns and rows of any matrix B. Hence we can always "divide" by such a matrix A. To exploit this situation properly we need a few definitions and some easily proved theorems.

Definitions 12.2. (a) An $n \times n$ matrix is non-singular if its rank is n. If its rank is less than n it is singular.

(b) Given a matrix A, a matrix A_1 is a right (or left) inverse of A if $AA_1 = I$ (or $A_1A = I$). If A_1 is both a right and a left inverse of A it is simply called an inverse of A.

Example 12.2. By multiplying, we find that

$$\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{12.4}$$

Hence $\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$ is a left inverse of $\begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}$ and $\begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}$ is a right inverse of $\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$.

The following basic theorem tells us that we actually do not need to worry about "left" and "right" inverse.

Theorem 12.3. (a) If A is non-singular it has both a right and a left inverse.

(b) If A has a right (or left) inverse it is non-singular, and hence also has a left (or right) inverse.

(c) In either case (a) or (b) there is a unique inverse A^{-1} which is also the unique right inverse and the unique left inverse.

Proof. (a) This is just the argument of the paragraph preceding Definition 12.2, for the case $B = I$.

(b) If $AA_1 = I$ (or $A_1A = I$) then by Theorem 12.1(c) we must have

$$\text{rank of } A \geqq \text{rank of } I = n.$$

Hence A is non-singular.

(c) Let A_1 be any right inverse and A_2 any left inverse. Then $AA_1 = I$ and $A_2A = I$. Multiply the first of these equations by A_2 on the left, and the second by A_1 on the right. We get

$$A_2AA_1 = A_2I = A_2,$$

$$A_2AA_1 = IA_1 = A_1.$$

Hence $A_1 = A_2$, i.e. any right inverse equals any left inverse. It follows that there can be only one right inverse and one left inverse, which are equal and hence are the unique inverse.

Example 12.2, continued. The theorem assures us that from (12.4) automatically follows

$$\begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

Each of the two matrices is the inverse of the other, and, given one of them, the other is uniquely determined.

The importance of the inverse lies in its allowing us to dispense with the process of division and instead use multiplication by the inverse. We state this in the form of a theorem.

Theorem 12.4. If A is non-singular and B is any $n \times p$ matrix then the equation $AX = B$ has the unique solution $X = A^{-1}B$. Similarly, if C is a $p \times n$ matrix the equation $XA = C$ has the unique solution $X = CA^{-1}$.

Proof. $X = A^{-1}B$ is a solution of $AX = B$, since $A(A^{-1}B) = (AA^{-1})B = IB = B$. Conversely, if X is any solution of $AX = B$ then $A^{-1}B = A^{-1}(AX) = (A^{-1}A)X = IX = X$. The other half of the theorem is proved similarly.

Example 12.2, concluded. To solve

$$2x + 3y = 1,$$
$$3x + 5y = -4,$$

we merely compute

$$\begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -4 \end{pmatrix} = \begin{pmatrix} 17 \\ -11 \end{pmatrix},$$

i.e. $x = 17$, $y = -11$.

We have finally to consider how to compute A^{-1}, given A. The most direct method is to use the Gauss elimination method to solve the equation $AX = I$. We saw in Section 1 how this could be done, essentially by starting with the augmented matrix (A, I) and reducing A to reduced echelon form. (See Table 1.1c.) The reduced echelon form for a non-singular matrix A is just I, and so (A, I) reduces to the form (I, Z), corresponding to the equation $IX = Z$. Now Theorem 11.4 tells us that such a reduction does not change the set of solutions, so any solution X of $AX = I$ is a solution of $IX = Z$. (The theorem was proved for column vectors x but it can be applied to the column vectors of X, one at a time.) Since $X = A^{-1}$ is the unique solution of the first equation we must have $Z = A^{-1}$.

Example 12.3. Find A^{-1} if

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} .$$

There is no need to follow the routine Gauss method if deviations make the work easier. Our problem is to reduce **A** to I by <u>any</u> convenient row operations. In this case we start at the bottom and work up, adding each row to the one above.

$$\left(\begin{array}{cccc|cccc} 2 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \end{array}\right)$$

gives

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \end{array}\right) .$$

Now work from the top down, adding each row to the one below.

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 1 & 0 & 1 & 2 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 & 2 & 3 & 4 \end{array}\right) .$$

Hence

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

This example will be used in Chapter 6.

If a single equation $Ax = b$ is to be solved there is no point in computing A^{-1}, for this is a harder job than solving the equation. But frequently one has to solve many such equations with the same A but different b's, and then it may well be more efficient to compute A^{-1} once for all and replace the solution of each equation by the simple matrix multiplication $x = A^{-1}b$.

The next theorem lists some useful properties of inverses and related matrices.

Theorem 12.5.

(a) If $A, B, \ldots, D$ are non-singular then their product $AB\ldots D$ is non-singular and $(AB\cdots D)^{-1} = D^{-1}\cdots B^{-1}A^{-1}$.

(b) If A is non-singular so is A^t, and $(A^t)^{-1} = (A^{-1})^t$.

(c) If A is non-singular and if B and C are any $m \times n$ and $n \times p$ matrices respectively, then $\operatorname{rank} BA = \operatorname{rank} B$ and $\operatorname{rank} AC = \operatorname{rank} C$.

(d) n vectors in V_n are independent (and hence a basis) if and only if the matrix having them as columns, or as rows, is non-singular.

(e) $Ax = b$ has a unique solution if and only if A is non-singular.

(f) $Ax = 0$ has a non-trivial solution if and only if A is singular.

Proof. (a) Let $F = AB\cdots D$ and $G = D^{-1}\cdots B^{-1}A^{-1}$. We need only to prove that $GF = I$. Now

$$GF = (D^{-1}\cdots B^{-1}A^{-1})(AB\cdots D)$$

$$= D^{-1}\cdots B^{-1}(A^{-1}A)B\cdots D.$$

Since $A^{-1}A = I$, and $IB = B$, this reduces to

$$GF = D^{-1}\cdots B^{-1}B\cdots D.$$

Obviously the process can be continued, removing the middle pair each time, until we have left only $GF = D^{-1}D = I$.

(b) From $AA^{-1} = I$ we get, taking transposes, $(A^{-1})^tA^t = I^t = I$. Since $(A^t)^{-1}$ is uniquely defined by the property $(A^t)^{-1}A^t = I$ we have $(A^{-1})^t = (A^t)^{-1}$.

(c) From Theorem 12.1(c) we have rank $BA \leqq$ rank B. But if A is non-singular we also have $(BA)A^{-1} = B$, and applying the same theorem we have rank $B =$ rank $(BA)A^{-1} \leqq$ rank BA. Hence rank $BA =$ rank B. A similar argument works for AC.

(d), (e), and (f) are simple consequences of the corollori s in Section 11, and their proofs are left to the reader. (See Problem 12.4.)

We conclude this section with a few examples of matrix algebra, showing similarities and differences between it and ordinary scalar algebra.

Example 12.4. Solve $AXB = C$.

If A and B are non-singular the method of Theorem 12.4 can be applied; we get rid of A and B as factors of X by multiplying on the left by A^{-1} and on the right by B^{-1}. Thus

$$A^{-1}CB^{-1} = A^{-1}AXBB^{-1} = IXI = X.$$

If both A and B are singular this is a nasty problem that has to be examined from first principles as in Section 11.

Example 12.5. If $AC = BC$ and C is non-singular then $A = B$. We merely multiply on the right by C^{-1}, giving $ACC^{-1} = BCC^{-1}$, or $A = B$. We saw in Section 10 that this cancellation law is not in general true.

Example 12.6. Solve $X^2 = I$. The proof in scalar algebra that $x^2 = 1$ has only two solutions rests on a form of the cancellation law, namely that if $(x - 1)(x + 1) = 0$ then either $x - 1 = 0$ or $x + 1 = 0$. If we try the same procedure here we get as far as $(X - I)(X + I) = 0$ (prove this) but we cannot take the next step. As a matter of fact $X^2 = I$ has an infinite number of solutions (see Problem 12.6).

Problems

12.1 Find the inverse, when possible, of each of the following matrices.

(a) $\begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}$. (b) $\begin{pmatrix} 2 & 0 & 1 \\ -1 & 2 & 1 \\ 1 & 4 & 3 \end{pmatrix}$.

(c) $\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}$.

(d) $\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$.

(e) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

Answer (d):

$$\frac{1}{5}\begin{pmatrix} 4 & 3 & 2 & 1 \\ 3 & 6 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix} .$$

12.2 Use the results of Problem 12.1 and Theorem 12.4, if possible, to solve each of the following.

(a) $$\begin{aligned} 2x + y &= 7, \\ -x + 3y &= -1. \end{aligned}$$

(b) $$\begin{aligned} 2x + z &= 3, \\ -x + 2y + z &= 4, \\ x + 4y + 3z &= -1. \end{aligned}$$

(c) $$\begin{aligned} y + 2z &= -1, \\ 3x + 4y + 5z &= 4, \\ 6x + 7y + 8z &= 9. \end{aligned}$$

(d) $$\begin{aligned} 2x_1 - x_2 &= 4, \\ -x_1 + 2x_2 - x_3 &= -1, \\ - x_2 + 2x_3 - x_4 &= -3, \\ - x_3 + 2x_4 &= 3. \end{aligned}$$

12.3 (a) When is a diagonal matrix non-singular? How do you quickly determine the rank of a diagonal matrix? Describe the inverse of a non-singular diagonal matrix.

(b) Describe the effect on the rows or columns of a matrix of multiplying it by a diagonal matrix on the left. On the right.

12.4 Prove parts (d), (e), and (f) of Theorem 12.5.

12.5 If $ABC = I$:

(a) Prove that A^{-1}, B^{-1}, and C^{-1} exist;

(b) Solve for each of A, B, and C in terms of the other two.

12.6 (cf. Example 12.6.) For $n = 2$ find solutions for $X^2 = I$ by expanding

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 \text{ and equating to } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

12.7 In this problem we are concerned only with the set S of all 2×2 matrices of the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Show the following:

(a) S is a 2-dimensional vector space.

(b) S is closed under multiplication.

(c) Multiplication is commutative.

(d) Every non-zero element has an inverse.

(e) The equation $X^2 = -I$ has exactly two solutions.

12.8 Write a program to compute the inverse of an $n \times n$ matrix.

12.9 In Theorem 12.2 show that I is the only matrix have the property (c).

13. Determinants.

Throughout this section and the next we continue to deal only with square matrices. With each $n \times n$ matrix we can associate a scalar called the <u>determinant</u> of the matrix. The student should be warned against the misconception that the determinant is the matrix. The matrix contains n^2 numbers while the determinant is only one number, so it is clear that in general there is much more data in a matrix than in its determinant.*

The definition of a determinant is somewhat involved, and one rarely has occasion to refer directly to it. Of more use are certain properties that follow from the definition. We shall first present these, and some of their applications, and only later give the definition and proof of the basic properties.

We designate the determinant of the matrix $A = (a_{ij})$ by $\det A$ or $|A|$ or $|(a_{ij})|$ or

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdot & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} .$$

The fundamental properties of determinants can be stated as follows:

(a) If any one row of A is multiplied by a scalar c, then $\det A$ is multiplied by c.

---

*One occasionally uses the phrase "an $n \times n$ determinant." By this one means, of course, the determinant of an $n \times n$ matrix. Similarly, when we speak of a "row," "column," or "element" of a determinant we are referring to the matrix whose determinant is under consideration.

Example: $$\begin{vmatrix} a & b & c \\ kd & ke & kf \\ g & h & i \end{vmatrix} = k \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} .$$

(b) If to any row of $\mathbf{A}$ is added a scalar multiple of another row, then det $\mathbf{A}$ is unchanged.

Example: $$\begin{vmatrix} a & b & c \\ d & e & f \\ g+ka & h+kb & i+kc \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} .$$

(c) Interchanging any two rows of A changes the sign of det A.

Example: $$\begin{vmatrix} g & h & i \\ d & e & f \\ a & b & c \end{vmatrix} = - \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} .$$

(d) If two rows of A are identical, then det A = 0.

(e) If A has a row of zeros, then det A = 0.

(f) If one given row vector u of A is the sum of two vectors v and w, then det $\mathbf{A}$ = det A_1 + det A_2, where A_1 and A_2 are identical with $\mathbf{A}$ except in having the row u replaced by v and w respectively.

Example: $$\begin{vmatrix} a_1+a_2 & b_1+b_2 & c_1+c_2 \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} a_2 & b_2 & c_2 \\ d & e & f \\ g & h & i \end{vmatrix} .$$

(g) Properties (a) to (f) hold for columns as well as for rows.

(h) $\det I = 1$.

(i) $\det A^t = \det A$.

(j) $\det A = 0$ or $\neq 0$ according as A is singular or non-singular.

(k) $\det (AB) = (\det A)(\det B)$; if A^{-1} exists, then $\det (A^{-1}) = (\det A)^{-1}$.

Some comments on these are in order. (a) and (f) together say that $\det A$ is a linear function of each of its row vectors. (a), (b), (c) correspond to the elementary row operations on matrices; they tell us the effect of these operations on $\det A$. Since by these operations we can reduce any $n \times n$ matrix either to I or to a matrix with a row of zeros - cases covered by (h) and (e) - we can in this way compute the value of a determinant.

Some of these properties are "more fundamental" than others. Proofs of (a), (c), (f), (h), and (i) depend explicitly on the definition of a determinant and shall be postponed until the end of this section. All the other properties follow from these five, and we shall derive them at appropriate points. At present we prove (d), (e), (b), and (g) in this order.

(d) If two rows of A are equal, interchanging them does not affect $\det A$. But by (c), interchanging them changes $\det A$ to $-\det A$. Hence $\det A = -\det A$, and so $\det A = 0$.

(e) If A has a row of zeros, multiplying this row by 0 does not affect $\det A$. But by (a), such a multiplication reduces $\det A$ to 0. Hence $\det A = 0$.

(b) How this follows from (f), (a), and (d) is best seen from an example:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g+ka & h+kb & i+kc \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} a & b & c \\ d & e & f \\ ka & kb & kc \end{vmatrix}$$

$$= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} + k \begin{vmatrix} a & b & c \\ d & e & f \\ a & b & c \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} + 0.$$

(g) Property (i) is the critical one in this proof. Suppose, for instance, we want to find the effect on det A of interchanging the first and third columns. We do this in three steps: let $B = A^t$, let C = result of interchanging first and third rows of B, let $D = C^t$. Then det B = det A, by (i); det C = - det B, by (c); det D = det C, by (i). The result is det D = - det A, i.e. the application of (c) with "row" replaced by "column." A similar argument can easily be made for all the other cases covered by (g).

Before proving (j) and (k) let us look at the process mentioned above for evaluating a determinant. We use elementary row or column operations and properties (a), (b), (c) until we can use either (e) or (h). (a) is generally used in the form: If a row (or column) is divided by c and the resulting determinant multiplied by c the result is equal to the original determinant.

Example 13.1.

$$\begin{vmatrix} 1 & 2 & 3 \\ 1 & 5 & -3 \\ -2 & 0 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & 3 & -6 \\ 0 & 4 & 8 \end{vmatrix}$$

adding (-1) times first row to second and 2 times first row to third

$$= 3\cdot 4 \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 0 & 1 & 2 \end{vmatrix}$$

dividing second row by 3 and third row by 4

$$= 12 \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{vmatrix}$$

subtracting the second row from the third

$$= 48 \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{vmatrix}$$

$$= 48 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 48.$$

Example 13.2.

$$A = \begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 6 & 1 & 4 \\ -3 & -6 & -9 & -2 & 5 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 3 & 4 & 3 & 2 \end{vmatrix} .$$

Starting the process of reducing to echelon form gives

$$|A| = \begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & -7 & -6 \\ 0 & 0 & 0 & 10 & 20 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & -1 & -3 \end{vmatrix}$$

$$(13.1) \qquad = 10 \begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & -7 & -6 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -2 & -3 \end{vmatrix} .$$

On interchanging the second and fourth rows we get

$$|A| = -10 \begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -7 & -6 \\ 0 & 0 & 0 & -2 & -3 \end{vmatrix} .$$

It is now evident that the echelon form of A must have a row of zeros, and hence $|A| = 0$. Alternatively, we could have proceeded from (13.1) to reduce the third column to zero by subtracting from it the second column and the first column.

These examples show clearly the relation between determinants and echelon forms. We can use this relation to prove properties (j) and (k).

Proof of (j). Reduce A to reduced echelon form E by row operations. By (a), (b), and (c) the effect is to multiply $\det A$ by certain non-zero factors, and so we have $\det E = f \det A$, where $f \neq 0$. Now if A is singular E has a row of zeros and so $\det E = 0$ by (e); and hence $\det \mathbf{A} = 0$. If A is non-singular, $E = I$, and $\det \mathbf{A} = 1/f \neq 0$.

Proof of (k). Let $\mathbf{AB} = C$. If A is singular then C is also singular by Theorem 12.1(c) and so we have $\det (\mathbf{AB}) = 0 = (\det A)(\det B)$. We therefore have only to consider the case when $\mathbf{A}$ is non-singular. B is then the unique solution to $\mathbf{AX} = C$, and we can find B by starting with the augmented matrix $(\mathbf{A}, C)$ and reducing $\mathbf{A}$ to its reduced echelon form, I, by row operations. When we do this the augmented matrix $(\mathbf{A}, C)$ reduces to (I, B). As above we have $1 = \det I = f \det A$, where f is a non-zero factor determined by the sequence of row operations used in the reduction. But this same sequence of operations reduces C to B, and so we also have $\det B = f \det C$. Eliminating f from these two equations gives $\det C = (\det A)(\det B)$, which is the first part of (k). For the second part, if A^{-1} exists then $A^{-1}A = I$. By what we have just proved, $\det (A^{-1})(\det A) = \det I = 1$, from which follows $\det (A^{-1}) = (\det A)^{-1}$.

As a consequence of (j) we now have seven ways of expressing the same property of an $n \times n$ matrix A:

The rows of $\mathbf{A}$ are independent;

The columns of $\mathbf{A}$ are independent;

The rank of $\mathbf{A}$ is n;

$\mathbf{A}$ is non-singular;

$\mathbf{A}^{-1}$ exists;

$\det \mathbf{A} \neq 0$;

$Ax = 0$ has only the trivial solution.

The number of ways in which this property arises gives some indication of its importance.

We close this section by giving the definition of a determinant and proof of the remaining fundamental properties. We need some preliminary definitions.

A _permutation_ of the integers $(1,2, \dots ,n)$ is a rearrangement of these integers so that each appears once and only once. For example $(2,4,1,3)$ is a permutation of integers $(1,2,3,4)$.

For each integer in a permutation the _number of inversions associated with that integer_ is the number of integers which precede it and are larger than it. For example, in the permutation $(2,4,1,3)$, the number of inversions associated with 1 is two, with 2 is zero, with 3 is one and with 4 is zero. _The number of inversions in a permutation_ is the sum of the number of inversions associated with each of the entries. Thus, the number of inversions in the permutation $(2,4,1,3)$ is $2 + 0 + 1 + 0 = 3$. The number of inversions in the permutation $(5,3,1,4,2)$ is $2 + 3 + 1 + 1 + 0 = 7$.

If the number of inversions in a permutation is even, then the _parity_ of the permutation is said to be 1; if the number of inversions is odd, the parity is said to be -1. If q is the number of inversions the parity is $(-1)^q$.

The important property of parity is that _an interchange of two adjacent numbers in a permutation changes the parity of the permutation._

For example, the permutation $(5,3,1,4,2)$ has parity -1. If we interchange the 3 and the 1 we get the permutation $(5,1,3,4,2)$ which has $1 + 3 + 1 + 1 + 0 = 6$ inversions and parity +1. In general, any interchange of adjacent elements of a permutation either creates or removes exactly one inversion, and so changes the parity.

If $(j_1, j_2, \ldots, j_n)$ is a permutation of $(1,2,\ldots,n)$ the symbol $\epsilon(j_1, j_2, \ldots, j_n)$ denotes the parity of the permutation.

Now we are ready to define a determinant.

Definition 13.1. The determinant of the $n \times n$ matrix $A = (a_{ij})$ is defined by

$$\det A = \sum \epsilon(j_1, j_2, \ldots, j_n) a_{1j_1} a_{2j_2} \cdots a_{nj_n}, \tag{13.2}$$

the sum being taken over all permutations $(j_1, j_2, \ldots, j_n)$ of $(1,2,\ldots,n)$.

Example 13.3. (a) The permutations of $(1,2)$ are $(1,2)$, of parity 1, and $(2,1)$, of parity -1. Hence

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} .$$

(b) There are 6 permutations of $(1,2,3)$, as follows:

$(1,2,3)$, $(2,3,1)$, $(3,1,2)$ of parity 1;

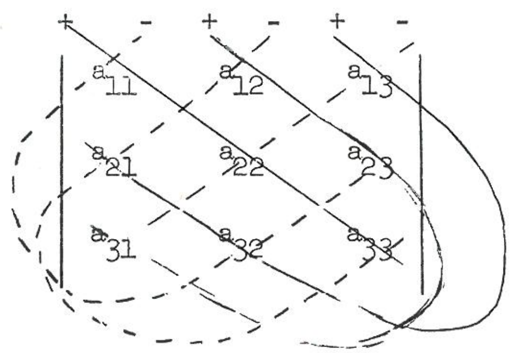
$(1,3,2)$, $(2,1,3)$, $(3,2,1)$ of parity -1.

Hence

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{array}$$

These two cases can be remembered by the pictorial devices

$$\begin{array}{c} + \qquad\qquad - \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{array} = a_{11}a_{22} - a_{12}a_{21}$$



There are no similar schemes for values of n greater than 3.

Definition 13.1 implies that each term in the sum consists of the product of n of the entries a_{ij} taken in such a way that each column and each row of the matrix is represented once and only once. The term has a sign of +1 or -1, equal to the parity of the permutation formed by the column subscripts, when the row subscripts are arranged in natural order.

We are now ready to finish the proofs of properties (a), (f), (c), (h), (i).

(a) Since each term of det A contains as a factor exactly one element of any given row, if we multiply each element of this row by c we multiply each term of det **A** by c. Hence det A is multiplied by c.

(f) **A** similar argument applies here. If, for instance, $a_{1j} = b_{1j} + c_{1j}$ for $j = 1, \ldots, n$, then

$$a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

$$= b_{1j_1} a_{2j_2} \cdots a_{nj_n} + c_{1j_1} a_{2j_2} \cdots a_{nj_n},$$

and by combining corresponding terms the determinant with a's in the first row is the sum of the two determinants with b's and c's in the first row.

(c) Suppose that A' is obtained from A by interchanging the first and second rows. Any term of det A' is then of the form

$$\epsilon(j_1, j_2, \ldots, j_n) a_{2j_1} a_{1j_2} \cdots a_{nj_n},$$

and there will be a corresponding term

$$\epsilon(j_2, j_1, \ldots, j_n) a_{1j_2} a_{2j_1} \cdots a_{nj_n}$$

in det **A**. By the basic property of parity these terms are just negatives of one another, and since the same is true for all terms we have det **A**' = - det **A**.

We have thus proved (c) for an interchange of the first two rows. The same proof will obviously apply to any two <u>adjacent</u> rows. Now suppose that u and v are two rows separated by k other rows. To interchange u and v we first interchange u with each of the k intervening rows, thus bringing it adjacent to v, then interchange u and v, and finally interchange v with each of the k rows, thus bringing it to the original position of u. The total number of adjacent interchanges is then $k + 1 + k = 2k + 1$. The sign of the determinant is therefore changed an odd number of times, and so $\det A' = -\det A$.

(h) For $I = (a_{ij})$ the only non-zero elements are $a_{ii} = 1$. Hence $\det I$ has just one non-zero term $a_{11}a_{22}\ldots a_{nn} = 1$.

(i) Since $a^t_{ij} = a_{ji}$, the terms of $\det A^t$ are of the form

$$(13.3) \qquad \epsilon(j_1, j_2, \ldots, j_n)a_{j_1 1}a_{j_2 2} \cdots a_{j_n n}.$$

Now $(j_1, j_2, \ldots, j_n)$ is a permutation of $(1,2,\ldots,n)$, and so by rearranging the a's we can write (13.3) in the form

$$(13.4) \qquad \epsilon(j_1, j_2, \ldots, j_n)a_{1i_1}a_{2i_2} \cdots a_{ni_n}.$$

Since

$$\epsilon(i_1, i_2, \ldots, i_n)a_{1i_1}a_{2i_2} \cdots a_{ni_n}$$

is a term of $\det A$, we can conclude that $\det A^t = \det A$ if we can show that

$$(13.5) \qquad \epsilon(i_1, i_2, \ldots i_n) = \epsilon(j_1, j_2, \ldots, j_n).$$

To see that (13.5) is true consider a special case, the passage from

$$a_{51}a_{32}a_{13}a_{44}a_{25} \text{ to } a_{13}a_{25}a_{32}a_{44}a_{51}.$$

We do this by successive interchanges of adjacent factors, bringing successively the subscripts 13, 25, 32, ... into the desired positions, thus:

$$\begin{aligned}
51, 32, 13, 44, 25 &\longrightarrow 13, 51, 32, 44, 25 \qquad \text{(2 interchanges)}\\
&\longrightarrow 13, 25, 51, 32, 44 \qquad \text{(3 interchanges)}\\
&\longrightarrow 13, 25, 32, 51, 44 \qquad \text{(1 interchange)}\\
&\longrightarrow 13, 25, 32, 44, 51 \qquad \text{(1 interchange).}
\end{aligned}$$

The total number of interchanges (7) is equal to the number of inversions in the permutation $(5,3,1,4,2)$. From the definition of the number of inversions in a permutation it is evident that this is true in general, that is, if $(j_1, j_2, \ldots, j_n)$ has q inversions then q interchanges of adjacent a's will take us from (13.3) to (13.4). By definition, $\epsilon(j_1, j_2, \ldots, j_n) = (-1)^q$. Also, since $(i_1, i_2, \ldots, i_n)$ has been obtained from $(1,2,\ldots,n)$ by q inversions, we have

$$\epsilon(i_1, i_2, \ldots, i_n) = (-1)^q \epsilon(1,2,\ldots,n) = (-1)^q.$$

Hence (13.5) is true, and (i) is thereby proved.

Problems

13.1 Evaluate the following determinants by reduction to diagonal form.

(a) $\begin{vmatrix} 1 & 0 & -1 \\ 2 & 0 & 1 \\ 3 & 1 & 2 \end{vmatrix}$ (b) $\begin{vmatrix} 4 & 0 & 2 & 5 \\ 3 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 3 & 0 \end{vmatrix}$

(c) $\begin{vmatrix} -3 & 2 & 5 \\ 1 & 0 & 2 \\ 3 & 1 & 1 \end{vmatrix}$ (d) $\begin{vmatrix} 2 & 6 & 6 & 2 \\ 1 & 0 & 4 & 0 \\ 5 & 12 & 16 & 4 \\ 3 & 1 & 6 & 2 \end{vmatrix}$

13.2 Evaluate (a) and (c) above by the method of Example 13.3.

13.3 Prove that the determinant of a diagonal matrix is the product of the elements on the main diagonal.

13.4 A square matrix is <u>upper</u> (or <u>lower</u>) <u>triangular</u> if all the elements below (or above) the main diagonal are zero. What can you say about the determinant of a triangular matrix?

13.5 Check property (j) for the matrices of Problem 13.1.

13.6 If c is a scalar and A an $n \times n$ matrix what is the relation, if any, between $\det(cA)$ and $\det A$?

13.7 By an example, show that it is not generally true that $\det(A+B) = \det A + \det B$.

13.8 What is the determinant of the $n \times n$ matrix having 1's on the diagonal from lower left to upper right and 0's elsewhere?

13.9 A <u>Vandermonde determinant</u> has the form

$$V = \begin{vmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ & & \cdots & & \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{vmatrix} .$$

(a) Prove that if no two of $a_1, a_2, \ldots, a_n$ are equal then $V \neq 0$. [Partial proof. If $V = 0$ the column vectors are dependent. That is, there are scalars $c_0, c_1, \ldots, c_{n-1}$, not all zero, such that

$$c_0 + c_1 a_i + c_2 a_i^2 + \ldots + c_{n-1} a_i^{n-1} = 0, \qquad i = 1, \ldots, n.$$

Can $a_1, a_2, \ldots, a_n$ be roots of the $(n-1)$st degree equation

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-1} x^{n-1} = 0?]$$

(b) Prove that if no two of $a_1, a_2, \ldots, a_n$ are equal, then the system of equations

$$\sum_{i=1}^{n} x_i = 0,$$

$$\sum_{i=1}^{n} a_i x_i = 0,$$

$$\sum_{i=1}^{n} a_i^2 x_i = 0,$$

$$\cdots\cdot$$

$$\sum_{i=1}^{n} a_i^{n-1} x_i = 0,$$

has only the trivial solution. (This fact has many applications.)

13.10 Write a program to evaluate an $n \times n$ determinant.

14. Cofactors. Cramer's Rule.

If from an $n \times n$ matrix (a_{ij}) we delete the i^{th} row and j^{th} column we obtain an $(n-1)x(n-1)$ matrix. The determinant of this matrix is called the minor M_{ij} of a_{ij}. The cofactor A_{ij} of a_{ij}

is defined to be $(-1)^{i+j}M_{ij}$. For example consider the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 13 \\ 14 & 15 & 17 & 18 \end{pmatrix}.$$

The cofactor of the element 6 is

$$A_{22} = (-1)^{2+2} \begin{vmatrix} 1 & 3 & 4 \\ 9 & 11 & 13 \\ 14 & 17 & 18 \end{vmatrix} = 33.$$

The cofactor of the element 2 is

$$A_{12} = (-1)^{1+2} \begin{vmatrix} 5 & 7 & 8 \\ 9 & 11 & 13 \\ 14 & 17 & 18 \end{vmatrix} = -(-17) = 17.$$

In terms of cofactors we can add an important property of determinants to the eleven listed in the previous section, namely:

Laplace's Expansion.* If each element of a given row (or column) of a matrix **A** is multiplied by its cofactor, the sum of these products is det **A**.

Example 14.1.

$$\begin{vmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 9 \end{vmatrix} = 0 \begin{vmatrix} 4 & 5 \\ 7 & 9 \end{vmatrix} - 1 \begin{vmatrix} 3 & 5 \\ 6 & 9 \end{vmatrix} + 2 \begin{vmatrix} 3 & 4 \\ 6 & 7 \end{vmatrix}$$

$$= 0(1) - 1(-3) + 2(-3) = -3$$

---

*This is only a special case of the general Laplace expansion which involves any number of rows (or columns) simultaneously.

This is "expansion by elements of the first row." The "expansion by elements of the second column" is

$$\begin{vmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 9 \end{vmatrix} = -1 \begin{vmatrix} 3 & 5 \\ 6 & 9 \end{vmatrix} + 4 \begin{vmatrix} 0 & 2 \\ 6 & 9 \end{vmatrix} - 7 \begin{vmatrix} 0 & 2 \\ 3 & 5 \end{vmatrix}$$

$$= -1(-3) + 4(-12) - 7(-6) = -3.$$

Evaluation of a determinant by direct use of the Laplace expansion is usually not advisable, being long and complicated. However, if we combine the Laplace expansion with the technique of the preceding section we get usually the best method of evaluating a determinant with small integer elements. (If the elements are more complicated numbers a routine reduction to echelon form is generally preferable.)

Example 14.2. Evaluate

$$|A| = \begin{vmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 6 & 1 & 4 \\ -3 & -6 & -7 & -2 & 5 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 3 & 5 & 3 & 2 \end{vmatrix} .$$

Subtract the second column from the third and from the fourth, to get

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

If we now expand by elements of the fourth row we have only one cofactor to worry about, A_{42}, for the others will be multiplied by zeros and their value need not concern us. Therefore we get

$$|A| = 1 \cdot \begin{vmatrix} 1 & 1 & 2 & 5 \\ 2 & 2 & -3 & 4 \\ -3 & -1 & 4 & 5 \\ 1 & 2 & 0 & 2 \end{vmatrix} .$$

We can get two zeros in the first column by subtracting the second column from the first:

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

Now get another zero in the first column by adding -2 times the fourth row to third:

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

$$= -(-1) \begin{vmatrix} 1 & 2 & 5 \\ 2 & -3 & 4 \\ -5 & 4 & 1 \end{vmatrix} .$$

Adding suitable multiples of the first row to the second and third gives

$$|A| = \begin{vmatrix} 1 & 2 & 5 \\ 0 & -7 & -6 \\ 0 & 14 & 26 \end{vmatrix} = -2 \begin{vmatrix} 7 & 6 \\ 7 & 13 \end{vmatrix} = -2 \begin{vmatrix} 7 & 6 \\ 0 & 7 \end{vmatrix} = -98$$

We frequently encounter determinants whose elements are not scalars but are polynomials in one or more variables or various other mathematical objects. As long as these objects have the familiar properties of addition and subtraction, and satisfy the associative, commutative, and distributive laws of multiplication, such determinants have a meaning and can be evaluated by the methods at our disposal.

Example 14.3. In Chapters 5 and 6 we shall find that determinants like the following are very important:

$$\begin{vmatrix} \lambda & 0 & 1 & 2 \\ 0 & \lambda & 2 & 3 \\ 1 & 2 & \lambda-1 & 2 \\ 2 & 3 & 2 & \lambda-2 \end{vmatrix}$$

We wish to evaluate this as a polynomial in λ. Some simplification is possible before making a Laplace expansion but it may not be advisable to carry this as far as in a numerical case. In this example we get one more zero in the first row and then expand

$$\begin{vmatrix} \lambda & 0 & 1 & 0 \\ 0 & \lambda & 2 & -1 \\ 1 & 2 & \lambda-1 & -2\lambda+4 \\ 2 & 3 & 2 & \lambda-6 \end{vmatrix}$$

$$= \lambda \begin{vmatrix} \lambda & 2 & -1 \\ 2 & \lambda-1 & -2\lambda+4 \\ 3 & 2 & \lambda-6 \end{vmatrix} + \begin{vmatrix} 0 & \lambda & -1 \\ 1 & 2 & -2\lambda+4 \\ 2 & 3 & \lambda-6 \end{vmatrix} .$$

Combine to get more zeros:

$$\lambda \begin{vmatrix} \lambda & 2 & -1 \\ 2 & \lambda-1 & -2\lambda+4 \\ -\lambda+3 & 0 & \lambda-5 \end{vmatrix} + \begin{vmatrix} 0 & \lambda & -1 \\ 1 & 2 & -2\lambda+4 \\ 0 & -1 & 5\lambda-14 \end{vmatrix} .$$

The first determinant we expand by the special method for the 3×3 case:

$$\lambda[\lambda(\lambda-1)(\lambda-5) + 2(-2\lambda+4)(-\lambda+3) + (\lambda-1)(-\lambda+3) - 4(\lambda-5)]$$

$$- \begin{vmatrix} \lambda & -1 \\ -1 & 5\lambda-14 \end{vmatrix}$$

$$= \lambda(\lambda^3-6\lambda^2+5\lambda+4\lambda^2-20\lambda+24-\lambda^2+4\lambda-3-4\lambda+20) - (5\lambda^2-14\lambda-1)$$

$$= \lambda^4 - 3\lambda^3 - 20\lambda^2 + 55\lambda + 1.$$

Expressed in symbols, the Laplace Expansion can be stated:

(14.1) $$\sum_{j=1}^{n} a_{ij}A_{ij} = \det A, \quad \text{for any } i;$$

(14.2) $$\sum_{i=1}^{n} a_{ij}A_{ij} = \det A, \quad \text{for any } j.$$

Suppose that for some fixed i the i-th row vector of A, $(a_{i1}, a_{i2}, \dots, a_{in})$, is replaced by a vector $(b_1, b_2, \dots, b_n)$, thus forming a new matrix A'. This would not change the cofactors of the elements of the i-th row, and so we would have from (14.1),

$$\sum_{j=1}^{n} b_jA_{ij} = \det A'.$$

Of course a similar statement holds for columns. Now apply this to the case where the vector $(b_1, b_2, \dots, b_n)$ is some other row vector of A, that is $b_j = a_{kj}$ for $k \neq i$. Then A' has two equal rows and so $\det A' = 0$. We therefore get

(14.3) $$\sum_{j=1}^{n} a_{ij}A_{kj} = 0 \qquad \text{if } i \neq k.$$

The symbol δ_{ik} was defined in (12.3) to be 1 if $i = k$ and 0 if $i \neq k$. Equations (14.1) and (14.3) can be combined in the form

$$(14.4) \qquad \sum_{j=1}^{n} a_{ij}A_{kj} = \delta_{ik}\det \mathbf{A}, \qquad \begin{array}{l} i = 1,2,\ldots,n, \\ k = 1,2,\ldots,n, \end{array}$$

Similarly, working with rows instead of columns we get

$$(14.5) \qquad \sum_{i=1}^{n} a_{ij}A_{jk} = \delta_{jk}\det \mathbf{A}, \qquad \begin{array}{l} j = 1,2,\ldots,n, \\ k = 1,2,\ldots,n. \end{array}$$

The sums appearing on the left of equations (14.4) and (14.5) are not quite matrix products. To obtain a matrix product we need merely to introduce the transpose, $\tilde{A}$, of the matrix whose elements are the cofactors A_{ij}. Then (14.4) and (14.5) can be written in the form

$$(14.6) \qquad A\tilde{A} = (\det A)I, \qquad \tilde{A}A = (\det A)I.$$

(The transposed matrix of cofactors, $\tilde{A}$, is sometimes called the adjoint of A, but as this word has a different and more important meaning in the general theory of linear transformations its use for $\tilde{A}$ should be avoided.)

The similarity between equation (14.6) and the definition,

$$AA^{-1} = I, \quad A^{-1}A = I,$$

of the inverse of A is striking. In fact we have at once:

If A is non-singular,

$$(14.7) \qquad A^{-1} = \frac{1}{\det A}\tilde{A}.$$

At first glance this equation seems to offer a convenient means of computing A^{-1} but this is true only for $n = 2$ and for some very special cases with $n > 2$. Direct computation of $\tilde{A}$ involves the evaluation of n^2 determinants of size $(n-1)x(n-1)$, a formidable task even for $n = 4$.

One interesting application of equation (14.7) is the derivation of <u>Cramer's Rule</u> for the solution of the equation $Ax = b$. If A is nonsingular the solution is

$$x = A^{-1}b$$
$$= \frac{1}{\det A}\tilde{A}b.$$

In terms of components this can be written

$$(14.8) \qquad x_j = \frac{1}{\det A} \sum_{i=1}^{n} A_{ij}b_i , \qquad j = 1,\ldots,n.$$

Now let A_j be the matrix obtained from A by replacing the j-th column vector by the column vector b. We have seen that

$$\sum_{i=1}^{n} A_{ij}b_i = \det A_j,$$

and so (14.8) can be written

$$x_j = \frac{\det A_j}{\det A} .$$

This is Cramer's Rule.

<u>Example 14.4</u>. Solve for x_2:

$$
\begin{aligned}
2x_1 - x_2 &= 1, \\
-x_1 + 2x_2 - x_3 &= 0, \\
- x_2 + 2x_3 - x_4 &= 0, \\
-x_3 + 2x_4 &= 0.
\end{aligned}
$$

Use Cramer's Rule:

$$
\det A = \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} = \begin{vmatrix} 0 & 3 & -2 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & -2 & 3 & 0 \end{vmatrix} = \begin{vmatrix} 3 & -2 & 0 \\ -1 & 2 & -1 \\ -2 & 3 & 0 \end{vmatrix} = \begin{vmatrix} 3 & -2 \\ -2 & 3 \end{vmatrix} = 5,
$$

$$
\det A_2 = \begin{vmatrix} 2 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} = -\begin{vmatrix} -1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3.
$$

Hence $x_2 = 3/5$.

It must be emphasized that determinants and such related topics as Laplace's Expansion, adjoints and Cramer's Rule, while of considerable use in proving theorems and deriving properties of matrices and vector spaces, are not generally useful computing devices. Except in very special cases, we do not solve n equations in n variables by Cramer's Rule if $n > 3$. Nor do we find the inverse of a matrix by computing all its cofactors, nor a determinant by direct application of Laplace Expansions. Instead, we use the techniques we developed earlier, using echelon forms. These techniques are quicker and more easily adapted to computer programming.

One special case where Cramer's Rule may be useful is in hand computation when the matrix involved is <u>sparse</u>, that is, when it has relatively few non-zero terms, and when we want the values of only one or a few of the unknowns. Problem 16.2 is a good example of a case of this kind.

We now give a proof of the Laplace Expansion. It is convenient to divide the proof into three parts.

<u>Part 1</u>. In the expansion (13.2) of $\det A$, the terms involving a_{11} are $a_{11}A_{11}$.

<u>Proof</u>. Since

$$\det A = \sum \epsilon(j_1, j_2, \ldots, j_n) a_{1j_1} a_{2j_2} \cdots a_{nj_n},$$

summed over all permutations $(j_1, j_2, \ldots, j_n)$ of $(1, 2, \ldots, n)$, the terms involving a_{11} are

$$(14.9) \qquad a_{11} \sum \epsilon(1, j_2, \ldots, j_n) a_{2j_2} \cdots a_{nj_n},$$

summed over all permutations $(j_2, \ldots, j_n)$ of $(2, \ldots, n)$. Also, since $A_{11} = M_{11}$, we have

$$(14.10) \qquad A_{11} = \sum \epsilon(j_2, \ldots, j_n) a_{2j_2} \cdots a_{nj_n},$$

summed over the same set of permutations. Here, of course, the parity $\epsilon(j_2, \ldots, j_n)$ is defined in terms of inversions of $(j_2, \ldots, j_n)$ from the natural order $(2, \ldots, n)$. It is evident that $\epsilon(1, j_2, \ldots, j_n) = \epsilon(j_2, \ldots, j_n)$. The statement of Part 1 then follows from comparison of (14.9) and (14.10).

Part 2. In the expansion of $\det A$, the terms involving a_{ij} are $a_{ij}A_{ij}$.

Proof. We reduce this case to the previous one by interchanging rows and columns so as to bring a_{ij} into the upper left corner. We first interchange the original i-th row successively with each of the $i-1$ rows above it, and then interchange the resultant j-th column with each of the $j-1$ columns to its left. For example, if $i = 2$, $j = 3$,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{21} & a_{23} & a_{22} \\ a_{11} & a_{13} & a_{12} \\ a_{31} & a_{33} & a_{32} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{23} & a_{21} & a_{22} \\ a_{13} & a_{11} & a_{12} \\ a_{33} & a_{31} & a_{32} \end{pmatrix}.$$

This process does not change the relative order of any rows or columns except the i-th row and the j-th column. If primed letters refer to the new matrix we therefore have $M_{ij} = M'_{11}$, and so $A_{ij} = (-1)^{i+j}A'_{11}$. Now in shifting the rows and columns we have made altogether $(i - 1) + (j - 1) = i + j - 2$ changes of sign of the determinant, and therefore $\det A = (-1)^{i+j-2}\det A' = (-1)^{i+j}\det A'$. By Part 1 the terms of $\det A'$ that involve a'_{11} $(= a_{ij})$ are $a'_{11}A'_{11}$, and hence the terms of $\det A$ that involve a_{ij} are

$$(-1)^{i+j}a'_{11}A'_{11} = a_{ij}A_{ij}.$$

Part 3. Laplace's Expansion.

Proof. For a given value of i, each term of the expansion of $\det A$ contains exactly one factor of the type a_{ij}, that is, exactly one factor from the i-th row. We saw in Part 2 that the sum of the terms involving a_{ij} is $a_{ij}A_{ij}$. Hence

$$\det A = \sum_{j=1}^{n} a_{ij}A_{ij}.$$

The proof for expansion by the elements of a column is similar.

Problems

14.1 Evaluate

$$\begin{vmatrix} 2 & 1 & 3 \\ -4 & 2 & -1 \\ 1 & 3 & 2 \end{vmatrix}$$

(a) by expanding by elements of the third row,

(b) by expanding by elements of the first column.

14.2 Verify the following statements.

(a) $$\begin{vmatrix} 0 & 1 & -1 & 2 \\ 1 & 0 & 1 & 2 \\ 1 & 2 & 0 & -1 \\ 2 & 0 & 1 & 0 \end{vmatrix} = 11.$$

(b) $$\begin{vmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{vmatrix} = 6.$$

(c) $$\begin{vmatrix} 0 & 1 & 2 & 3 & 4 \\ -1 & 0 & -1 & -2 & -3 \\ -2 & 1 & 0 & 1 & 2 \\ -3 & 2 & -1 & 0 & -1 \\ -4 & 3 & -2 & 1 & 0 \end{vmatrix} = 0.$$

(d) $$\begin{vmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 4 & 5 & 7 \end{vmatrix} = 2.$$

(e) $$\begin{vmatrix} a+1 & 1 & 1 & 1 \\ 1 & b+1 & 1 & 1 \\ 1 & 1 & c+1 & 1 \\ 1 & 1 & 1 & d+1 \end{vmatrix} = abcd + abc + abd + acd + bcd.$$

14.3 Find all values of λ for which the given matrix is singular.

(a) $$\begin{pmatrix} \lambda+1 & 2 & 3 & 4 \\ 1 & \lambda+2 & 3 & 4 \\ 1 & 2 & \lambda+3 & 4 \\ 1 & 2 & 3 & \lambda+4 \end{pmatrix}.$$ (b) $$\begin{pmatrix} \lambda & 0 & 1 & -1 \\ -1 & \lambda+1 & -1 & 1 \\ -1 & 1 & \lambda-1 & 1 \\ -1 & 1 & 0 & \lambda \end{pmatrix}.$$

14.4 Use equation (14.7) to show that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

provided $ad - bc \neq 0$.

14.5 Use Cramer's Rule to solve the following problems.

(a) Solve for x, y, z.

$$\begin{aligned} 2x - y + z &= 0, \\ x + 2y - 2z &= 3, \\ x - 3y + 2z &= -2. \end{aligned}$$

(b) Solve for x_2.

$$\begin{aligned} x_1 - x_2 \phantom{{}+ x_3} + x_4 &= 1, \\ x_1 + x_3 - 2x_4 &= 0, \\ 2x_1 - x_3 &= 2, \\ x_2 + x_3 - x_4 &= -1. \end{aligned}$$

(c) Solve for x_1.

$$\begin{aligned} x_1 + x_2 - x_3 \qquad\qquad &= 0, \\ x_1 + x_2 + x_3 - x_4 \qquad &= 0, \\ x_2 + x_3 + x_4 - x_5 &= 0, \\ x_3 + x_4 + x_5 &= 0, \\ x_4 + x_5 &= 1. \end{aligned}$$

14.6 If A is non-singular prove that

(a) $\det \tilde{A} = (\det A)^{n-1}$. [Hint. Use (14.7) and Problem 13.6.]

(b) $\tilde{\tilde{A}} = (\det A)^{n-2}A$. (These are also true if A is singular.)

14.7 Prove the following: If the elements of a non-singular matrix A are integers then the elements of A^{-1} are integers if and only if $\det A = \pm 1$.

14.8 (a) Show that

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (b - a)(c - a)(c - b).$$

[Hint. Subtract rows and look for factors.]

(b) Given three distinct numbers a, b, c, and any three numbers p, q, r, find a quadratic polynomial

$$f(t) = x_0 + x_1 t + x_2 t^2$$

such that

$$f(a) = p, \qquad f(b) = q, \qquad f(c) = r.$$

14.9 We have seen that triple diagonal matrices are of frequent occurrence and have useful special properties. The determinant of any triple diagonal matrix can be computed in the following manner. Denote an $n \times n$ triple diagonal determinant by

$$D_n = \begin{vmatrix} a_1 & c_2 & & & & \\ b_2 & a_2 & c_3 & & & \\ & b_3 & a_3 & c_4 & & \\ & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & b_{n-1} & a_{n-1} & c_n \\ & & & & b_n & a_n \end{vmatrix},$$

where all the elements not shown are zeros.

(a) By expanding by elements of the last row and making one additional simplification show that

$$D_n = a_n D_{n-1} - b_n c_n D_{n-2}.$$

[Suggestion. Try a special case first, say for $n = 5$.]

This equation is called a <u>recursion formula</u>. It enables us to compute D_n, step by step, for any value of n, if we have two consecutive D's to start with. Obvious starting values are $D_1 = a_1$, $D_2 = a_1a_2 - b_2c_2$.

(b) Write the recursion formula for a triple diagonal matrix with 2's on the main diagonal and -1's on the adjacent diagonals. Find the value of such an $n \times n$ determinant for $n = 1,2,\ldots,6$. Make a conjecture as to its value for any n. Can you prove your conjecture?

(c) Express

$$\begin{vmatrix} \lambda & 1 & 0 & 0 & 0 \\ 1 & \lambda & 2 & 0 & 0 \\ 0 & 2 & \lambda & 3 & 0 \\ 0 & 0 & 3 & \lambda & 4 \\ 0 & 0 & 0 & 4 & \lambda \end{vmatrix}$$

as a polynomial in λ. Ans. $\lambda^5 - 30\lambda^3 + 89\lambda$.

14.10 The following equations arise from a problem in heat conduction (cf. Problem 16.6). Solve for x_2.

$$\begin{aligned} 4x_1 - x_2 \qquad\qquad\qquad &= 3, \\ -x_1 + 4x_2 - x_3 - x_4 \qquad &= 0, \\ - x_2 + 4x_3 \qquad - x_5 &= 0, \\ - x_2 \qquad + 4x_4 - x_5 &= 5, \\ - x_3 - x_4 + 4x_5 &= 3. \end{aligned}$$

Answer. $x_2 = 73/89$.

15. Linear Transformations.

The general concept of a linear transformation from one vector space to another was introduced in Section 9. Since then our considerations have been confined to the transformations defined by matrices, and to properties of the matrices themselves. We shall now return to the general case, partly for its own sake and partly to show its intimate connection with the special case of matrix transformations.

From the examples in Section 9 we are led to the following definition:

Definition 15.1. Given vector spaces V and W, (which may coincide), a linear function from V to W is a function $L(v) = w$, where v is in V and w is in W, such that

$$L(a_1 v_1 + a_2 v_2) = a_1 L(v_1) + a_2 L(v_2) \tag{15.1}$$

for any scalars a_1, a_2 and any vectors v_1, v_2 in V. A linear function is also called a linear transformation or a linear operator or a linear mapping. $L(v)$ is called the image of v or the transform of v.

The defining property of linearity (15.1) is easily extendable to the more general form

$$L\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i L(v_i). \tag{15.2}$$

For we have

$$\begin{aligned} &L(a_1 v_1 + a_2 v_2 + a_3 v_3 + \dots + a_n v_n) \\ &= a_1 L(v_1) + L(a_2 v_2 + a_3 v_3 + \dots + a_n v_n) \\ &= a_1 L(v_1) + a_2 L(v_2) + L(a_3 v_3 + \dots + a_n v_n) \end{aligned}$$

and so on.

Many of the concepts that we introduced for matrices can be extended to linear transformations in general. Some of this was done in Section 11, where we saw that Theorem 11.1 applied to general linear transformations. Much of the algebra of matrices can be applied to general linear operators by means of the following definitions.

Definition 15.2.

(a) Let L_1 and L_2 be operators from V to W, and let c be any scalar. We define the operators $L_3 = L_1 + L_2$ and $L_4 = cL_1$, from V to W, by

$$L_3(v) = L_1(v) + L_2(v), \qquad L_4(v) = cL_1(v).$$

(b) Let L_1 be an operator from V to W, and L_2 an operator from W to U. We define the operator $L = L_2L_1$, from V to U, by

$$L(v) = L_2(L_1(v)).$$

(c) Let V be contained in W. (An important special case is $W = V$.) We define the identity operator I from V to W by

$$I(v) = v.$$

(d) If L_1 and L_2 are operators from V to V such that $L_1L_2 = L_2L_1 = I$ then each of L_1 and L_2 is the inverse of the other, and we write $L_1 = L_2^{-1}$, $L_2 = L_1^{-1}$.

Some important properties follow at once from these definitions.

Theorem 15.1. In (a) and (b) above, if L_1 and L_2 are each linear operators so are $L_1 + L_2$, cL_1, and L_2L_1. In (c), I is linear. In (d), if L_1 is linear so is L_2.

Proof. Proofs of (a), (b), and (c) are straightforward and are left as exercises. To prove (d), let v_1, v_2 be any vectors in V and let $L_2(v_1) = w_1$, $L_2(v_2) = w_2$. Then $L_1(w_1) = v_1$ and $L_1(w_2) = v_2$. Since L_1 is linear, $L_1(a_1w_1 + a_2w_2) = a_1v_1 + a_2v_2$. Then

$$\begin{aligned} L_2(a_1v_1 + a_2v_2) &= L_2L_1(a_1w_1 + a_2w_2) \\ &= I(a_1w_1 + a_2w_2) \\ &= a_1w_1 + a_2w_2 \\ &= a_1L_2(v_1) + a_2L_2(v_2); \end{aligned}$$

that is, L_2 is linear.

To avoid complications, for the rest of this section we shall assume that $W = V$; that is, we shall consider only linear transformations from V to V. This is analogous to our restricting ourselves to square matrices in Section 12. Many of the things we do can be extended in fairly obvious fashion to the more general case, but we leave such extensions to the interested reader. (See, in particular, Problems 15.7 and 15.8.)

The following theorem is similar to Theorem 10.1 plus some results of Section 11.

Theorem 15.2.

(a) Linear operators from V to V form a vector space under Definition 15.2(a).

(b) $c(L_1L_2) = (cL_1)L_2 = L_1(cL_2)$.

(c) $(L_1L_2)L_3 = L_1(L_2L_3)$.

(d) $L_1(L_2 + L_3) = L_1L_2 + L_1L_3$; $(L_2 + L_3)L_1 = L_2L_1 + L_3L_1$.

(e) $IL = LI = L$.

(f) If L^{-1} exists it is unique.

Sample Proof (d). We wish to show that

$$[L_1(L_2 + L_3)]v = (L_1L_2 + L_1L_3)v$$

for any vector v in V. We have

$$\begin{aligned}
[L_1(L_2 + L_3)]v &= L_1[(L_2 + L_3)v] && \text{by Def. 15.2(b)} \\
&= L_1[L_2v + L_3v] && \text{by Def. 15.2(a)} \\
&= L_1(L_2v) + L_1(L_3v) && \text{by linearity of } L_1 \\
&= (L_1L_2)v + (L_1L_3)v && \text{by Def. 15.2(b)} \\
&= (L_1L_2 + L_1L_3)v && \text{by Def. 15.2(a).}
\end{aligned}$$

The other parts of the proof are similar and are left to the reader.

Some applications of the algebra of linear transformations are shown in the following examples.

Example 15.1. Let $V = P^n$ or P^∞. We define two operators E and Δ, called respectively the translation operator and the difference operator, by

$$E\,p(t) = p(t+1),$$

$$\Delta\,p(t) = p(t+1) - p(t).$$

For example,

$$E(t^2 - 3t + 1) = (t+1)^2 - 3(t+1) + 1 = t^2 - t - 1,$$

$$\begin{aligned}
\Delta\,(t^2 - 3t + 1) &= [(t+1)^2 - 3(t+1) + 1] - [t^2 - 3t + 1] \\
&= 2t - 2.
\end{aligned}$$

These operators are of basic importance in the study of Difference Equations. (See, for instance, S. Goldberg, <u>Introduction to Difference Equations</u>, John Wiley and Sons, New York, 1958.)

Obviously $\Delta = E - I$. It is easy to see that E is linear, hence so is Δ. Now E and I commute, that is, $IE = EI$. It follows that they both commute with Δ, and hence that we can apply the ordinary algebra of polynomials to E, I and Δ. In particular we can expand $E^n = (I + \Delta)^n$ by the binomial theorem:

$$E^n = I^n + n\Delta I^{n-1} + \frac{n(n-1)}{2!}\Delta^2 I^{n-2} + \frac{n(n-1)(n-2)}{3!}\Delta^3 I^{n-3} + \dots ,$$

and so

$$E^n p(t) = I^n p(t) + n\Delta I^{n-1} p(t) + \frac{n(n-1)}{2!}\Delta^2 I^{n-1} p(t) + \dots .$$

Evidently $E^n p(t) = p(t + n)$, and $I^k p(t) = p(t)$ for any value of k. Hence we have

$$p(t + n) = p(t) + n\Delta p(t) + \frac{n(n-1)}{2!}\Delta^2 p(t) + \dots .$$

This is the simplest case of <u>Newton's Interpolation Formula</u>. (See Goldberg, loc. cit., p. 38.)

<u>Example 15.2</u>. In Theorem 12.3 we proved that a right inverse of a matrix is also a left inverse. To see that this statement is not true for all linear operators consider the derivative operator, $D = \frac{d}{dt}$, in P^∞. We define another operator S in P^∞ by

$$S\ p(t) = \int_0^t p(u)\, du.$$

If $p(t)$ is a polynomial so is $S\ p(t)$, and S is easily shown to be linear. Now

$$DS\ p(t) = \frac{d}{dt}\int_0^t p(u)\ du = p(t),$$

by the fundamental theorem of calculus, and so $DS = I$. But $SD \neq I$, as is easily seen from an example:

$$SD(t^2 + 1) = S(2t) = \int_0^t 2u\ du = t^2 \neq t^2 + 1.$$

Hence S is a right inverse of D but not a left inverse.

As a matter of fact D cannot have <u>any</u> left inverse. For suppose we had $TD = I$. Then $TD(p(t)) = p(t)$ and $TD(p(T) + c) = p(t) + c$. But $D(p(t)) = D(p(t) + c)$, and so $TD(p(t)) = TD(p(t) + c)$, contradicting the previous statement if $c \neq 0$.

<u>Example 15.3</u>. Let $V = C^{\infty}$ and consider transformations of the form

$$L\ f(t) = (D + p(t))f(t) = \frac{df}{dt} + p(t)f(t),$$

where $p(t)$ is some fixed element of C^{∞}. As a particular case $p(t)$ can be a constant c.

Consider the product of two such operators, for example $(D + t)(D + t^2)$. To see how to perform this multiplication apply the product operator to a function f. We have

$$\begin{aligned}(D + t)(D + t^2)f &= (D + t)(Df + t^2 f) \\ &= D(Df) + D(t^2 f) + tDf + t(t^2 f) \\ &= D^2 f + (t^2 Df + 2tf) + tDf + t^3 f \\ &= [D^2 + (t^2 + t)D + (2t + t^3)]f.\end{aligned}$$

Hence

$$(D + t)(D + t^2) = D^2 + (t^2 + t)D + (2t + t^3).$$

Similarly we find

$$(D + t^2)(D + t) = D^2 + (t^2 + t)D + (1 + t^3)$$
$$\neq (D + t)(D + t^2).$$

Thus these differential operators do not in general commute, and we cannot handle them as if they were simply polynomials in D. However, if we restrict ourselves to the case where the functions $p(t)$ are constants we find that the operators do commute, and they can be added, multiplied, factored, etc. just like polynomials in a variable D. These comments have important consequences in the theory of linear differential equations.

Our discussion thus far has put no restriction on the space V. We now assume that V has a finite dimension, n. Let $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis of V.

Let L be any linear transformation from V to V. Any vector $\vec{u}$ of V is a linear combination of the basis,

$$\vec{u} = \sum_{j=1}^{n} x_j \vec{v}_j , \tag{15.3}$$

and it follows from equation (15.2) that

$$L(\vec{u}) = \sum_{j=1}^{n} x_j L(\vec{v}_j). \tag{15.4}$$

Hence, <u>a linear transformation is completely determined by its effect on the vectors of a basis</u>.

To specify a linear transformation L we thus need to know what it does to each basis vector. Since the transform of $\vec{v}_j$ is an element of V it is thereby expressible as a linear combination of the basis. That is

$$(15.5) \qquad L(\vec{v}_j) = \sum_{i=1}^{n} a_{ij} \vec{v}_i , \qquad j = 1,\ldots,n.$$

The transformation L is completely determined by (15.5) and (15.4) if the matrix $A = (a_{ij})$ is given. Conversely, the transformation determines the matrix, for since $\{\vec{v}_1, \ldots ,\vec{v}_n\}$ is an independent set the coefficients a_{ij} in (15.5) are unique, by Theorem 6.4(c). Thus a given basis of V determines a one-to-one correspondence between linear transformations from V to V and $n \times n$ matrices.

To view this relationship from another angle let us return to the considerations of Section 8. In the proof of Theorem 8.2 we saw that equation (15.3) could be interpreted as establishing an isomorphism $\vec{u} \longleftrightarrow (x_1, \ldots ,x_n)$ between V and V_n. We wish to see what corresponds to L in this isomorphism. Let $L(\vec{u})$ correspond to $(y_1, \ldots ,y_n)$, that is,

$$L(\vec{u}) = \sum_{i=1}^{n} y_i \vec{v}_i .$$

Combining this equation with (15.4) and (15.5) gives

$$\begin{aligned}\sum_{i=1}^{n} y_i \vec{v}_i &= L(\vec{u}) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij} \vec{v}_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_j \vec{v}_i\end{aligned}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \vec{v}_i \, .$$

Referring again to Theorem 6.4(c) we see that the coefficients of $\vec{v}_i$ on the two sides of this equation must be equal, and so we get

$$(15.6) \qquad y_i = \sum_{j=1}^{n} a_{ij} x_j \, , \qquad i = 1, \ldots, n.$$

If we introduce the column vectors

$$x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \, , \quad y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \, ,$$

equation (15.6) is just the familiar matrix equation $y = Ax$.

We can sum up these conclusions in the following statement: Given a basis in a vector space V of dimension n, any linear transformation from V to V is associated with a unique matrix transformation of the components of the vectors in V with respect to this basis.

Example 15.4. Consider the operators E and Δ of Example 15.1, taking $V = P^3$. A natural basis for P^3 is $\{1, t, t^2, t^3\}$. Following our general notation above we can write for (15.3)

$$p(t) = x_1 + x_2 t + x_3 t^2 + x_4 t^3.$$

For (15.5) we have for the translation operator

$$E(1) = 1$$

$$E(t) = 1 + t$$

$$E(t^2) = 1 + 2t + t^2$$

$$E(t^3) = 1 + 3t + 3t^2 + t^3.$$

Hence (note the way the row and column subscripts are used in (15.5)),

$$A_E = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

If we now wish to find $E(2 + 3t - t^2 - 2t^3)$ we need merely compute

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ -5 \\ -7 \\ -2 \end{pmatrix}$$

to get the answer $2 - 5t - 7t^2 - 2t^3$.

Since $\Delta = E - I$ we can see at once that

$$A_\Delta = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

The matrix representation of a given linear transformation depends on the chosen basis of V. This observation gives rise to two questions: A. How does a change of basis affect the matrix? B. Can we choose a basis for which the matrix has a particularly simple form? We shall consider Question A here and defer discussion of the more difficult Question B to Chapter 5.

Before attacking Question A we must pick up where we left off in Section 8. There we were considering the effect of a change of basis on the components of a vector. With a slight change of notation the results were as follows:

Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\{\vec{v}'_1, \ldots, \vec{v}'_n\}$ be two bases of V. Each $\vec{v}_i$ is then a combination of the $\vec{v}'_i$,

$$(15.7) \qquad \vec{v}_i = \sum_{j=1}^{n} c_{ji}\, \vec{v}'_j, \qquad i = 1,\ldots,n.$$

Let a vector $\vec{u}$ be expressed in terms of the two bases as

$$\vec{u} = \sum_{i=1}^{n} x_i\, \vec{v}_i = \sum_{i=1}^{n} x'_i\, \vec{v}'_i .$$

Then (cf. equation (8.8))

$$(15.8) \qquad x'_j = \sum_{i=1}^{n} c_{ji} x_i, \qquad j = 1, \ldots, n.$$

The uses of the coefficients c_{ji} in (15.7) and (15.8) must be carefully distinguished. They differ in two respects. Equation (15.7) expressed the unprimed basis in terms of the primed, whereas (15.8) expressed the primed components in terms of the unprimed. In equation (15.7) the sum is on the row subscript of c_{ji}, while in (15.8) it is on the column subscript.

Equation (15.8) is, of course, just the matrix equation

$$(15.9) \qquad x' = Cx.$$

Thus a change of basis effects a matrix transformation on the components of a vector.

We might wonder if every matrix transformation can arise in this way, that is, if any given matrix can act as the C in this process. That this is not the case can be seen by interchanging the roles of $\vec{v}_i$ and $\vec{v}'_i$ in the above process. This leads to a result $x = C'x'$. Combining this with (15.9) gives $x = C'Cx$ for every x, from which it follows that $C'C = I$, and hence that C is non-singular. However it is not hard to show that if C is any given non-singular matrix we can find bases related to one another by equation (15.7) for the given C. (See Problem 15.4.)

Now we are in a position to answer Question A. Let $L(\vec{u}) = \vec{w}$ be any linear transformation in V, and let $\vec{w}$ be expressed in terms of the bases as

$$\vec{w} = \sum_{i=1}^{n} y_i \vec{v}_i = \sum_{i=1}^{n} y'_i \vec{v}'_i .$$

Then corresponding to (15.9) we have

$$y' = Cy. \tag{15.10}$$

Let L be represented by the matrix A in the $\vec{v}$-basis and by A' in in the $\vec{v}'$-basis; then

$$y = Ax, \tag{15.11}$$

$$y' = A'x'. \tag{15.12}$$

Writing (15.9) in the form $x = C^{-1}x'$, we combine this with (15.10) and (15.11) to get

$$y' = Cy = CAx = CAC^{-1}x'. \tag{15.13}$$

Since we have seen that, with respect to a given basis, the matrix representation of a transformation is unique, we conclude from this equation and (15.12) that

$$A' = CAC^{-1}.$$

This relationship is easily remembered by the diagram:

$$\begin{tikzcd} x \arrow[r, "A"] \arrow[d, "C"'] & y \arrow[d, "C"] \\ x' \arrow[r, "A'"'] & y' \end{tikzcd}$$

Figure 15.1

From this we can read off various relationships, e.g.

$$y = Ax \quad , \quad x' = Cx \quad , \quad \text{etc.,} \quad x = C^{-1}x', \text{ etc.}$$

(Note that we cannot say $x = A^{-1}y$ unless we know that A^{-1} exists.) The passage from x' to y' can be done directly by A', or indirectly through x and y: $x = C^{-1}x'$, $y = Ax$, $y' = Cy$. Equation (15.13) shows that in combining C^{-1}, A, and C in this order we must multiply them from <u>right</u> to <u>left</u>, thus $A' = CAC^{-1}$.

Matrices A and A' related in this fashion are said to be <u>similar</u>, and the passage from A to A' is called a <u>similarity transformation</u>. We can thus answer Question A by saying that a change of basis effects a similarity transformation on the matrix.

<u>Example 15.4, continued</u>. Consider the basis $\{1, t, t(t-1), t(t-1)(t-2)\}$ of P^3. To get the matrix A'_E of the operator E in this new basis we need the matrix C defined by (15.7) expressing the old basis in terms

of the new one. Here, however, it is easier to express the new basis in terms of the old, thus giving C^{-1}:

$$\begin{aligned} 1 &= 1, \\ t &= \quad t, \\ t(t-1) &= \quad -t + t^2, \\ t(t-1)(t-2) &= 2t - 3t^2 + t^3; \end{aligned}$$

$$C^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Direct computation then gives

$$A'_E = CA_EC^{-1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Let us check this by getting A'_E directly from the transform of the basis by E.

$$\begin{aligned} E(1) &= 1 && = 1, \\ E(t) &= t + 1 && = 1 + t, \\ E(t(t-1)) &= (t+1)t && = \quad 2t + t(t-1), \\ E(t(t-1)(t-2)) &= (t+1)t(t-1) && = \quad 3t(t-1) + t(t-1)(t-2). \end{aligned}$$

This gives the same A'_E as above.

Here again we have $A'_\Delta = A'_E - I$, so

$$A'_\Delta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

It is this simple form of the difference operator that makes this particular basis useful in the study of Difference Equations.

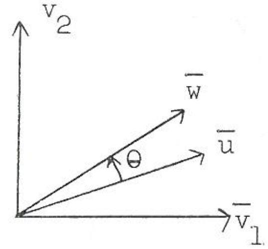
It is important to distinguish between the two uses of a matrix exemplified by (15.9) and (15.11). In (15.9) the n-tuples x and x' are representations in <u>different bases</u> of the <u>same vector</u> $\vec{u}$, whereas in (15.11) the n-tuples x and y are representations in the <u>same basis</u> of <u>different vectors</u> $\vec{u}$ and $\vec{w}$. The following example illustrates the relation between these two aspects of a matrix.

<u>Example 15.5.</u> Consider the matrix

$$R = \begin{pmatrix} \cos\Theta & -\sin\Theta \\ \sin\Theta & \cos\Theta \end{pmatrix},$$

Θ being some fixed angle.

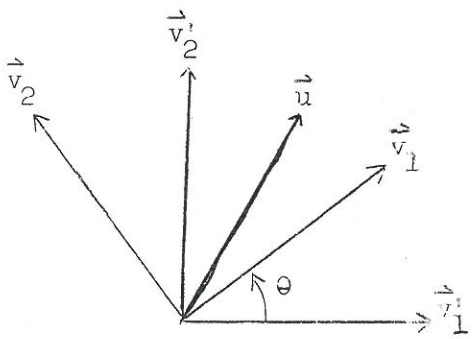
(a) In Problem 9.2 we considered this matrix as defining a transformation $Rx = y$, and showed that this transformation is a rotation through an angle Θ. (Figure 15.2).



$$\vec{u} = x_1\vec{v}_1 + x_2\vec{v}_2,$$
$$\vec{w} = y_1\vec{v}_1 + y_2\vec{v}_2;$$
$$x_1\cos\Theta - x_2\sin\Theta = y_1,$$
$$x_1\sin\Theta + x_2\cos\Theta = y_2.$$

Figure 15.2

(b) In Example 8.3 the same matrix was shown to define a rotation of axes, $x' = Rx$. (Figure 15.3). Here θ must be considered to be the angle from the $\vec{v}'$ basis to the $\vec{v}$ basis.



$$\vec{u} = x_1\vec{v}_1 + x_2\vec{v}_2 = x_1'\vec{v}_1' + x_2'\vec{v}_2';$$

$$\vec{v}_1 = \vec{v}_1'\cos\theta + \vec{v}_2'\sin\theta,$$

$$\vec{v}_2 = -\vec{v}_1'\sin\theta + \vec{v}_2'\cos\theta;$$

$$x_1' = x_1\cos\theta - x_2\sin\theta,$$

$$x_2' = x_1\sin\theta + x_2\cos\theta.$$

Figure 15.3

The application of a similarity transformation in Example 15.4 is not a particularly useful one, for the result is more easily obtained by direct computation with the new basis. The following example indicates how a change of basis can be used to solve a geometric problem whose straightforward solution would be very complicated.

<u>Example 15.6</u>. We wish to find the equation expressing a rotation of 90° in euclidean 3-space about the axis $x_2 = x_3$ in the x_2x_3-plane. The rotation will transform a column vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$ into a column

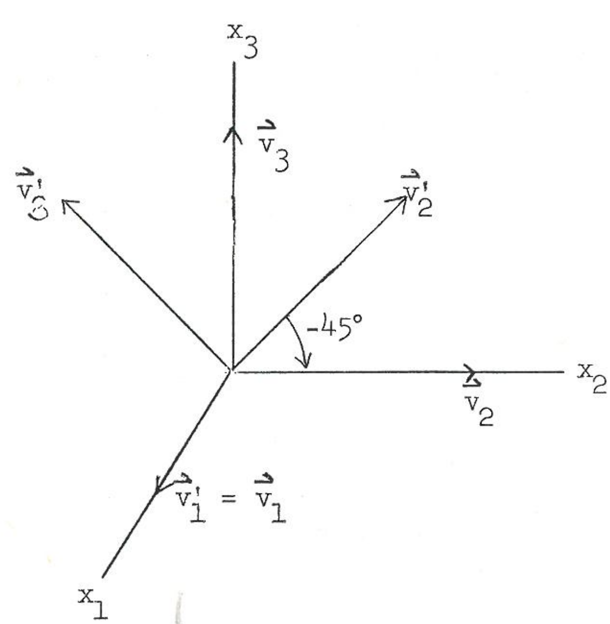


Figure 15.4

vector $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, and our problem is to find the matrix A such that $y = Ax$. To do this we introduce a new set of axes, chosen in a way to facilitate the expression of the rotation in the form $y' = A'x'$. From the relation between the two sets of axes we can then determine A from A' by a similarity transformation.

Let $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be a basis of unit vectors along the three axes, and $\{\vec{v}_1', \vec{v}_2', \vec{v}_3'\}$ another set of perpendicular unit vectors with $\vec{v}_1' = \vec{v}_1$ and $\vec{v}_2'$ along the axis of rotation (Figure 15.4). The relation between $\{\vec{v}_2, \vec{v}_3\}$ and $\{\vec{v}_2', \vec{v}_3'\}$ is the same as the relation in Example 15.5(b) between $\{\vec{v}_1, \vec{v}_2\}$ and $\{\vec{v}_1', \vec{v}_2'\}$ with $\theta = -45°$, while the $\vec{v}_1$, $\vec{v}_1'$ component is unchanged. Hence the matrix associated with this change of basis is

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(-45°) & -\sin(-45°) \\ 0 & \sin(-45°) & \cos(-45°) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Since C corresponds to a rotation through an angle of $-45°$, C^{-1} corresponds to a rotation through $45°$, and so

$$C^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 45° & -\sin 45° \\ 0 & \sin 45° & \cos 45° \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

The rotation we want is one of $90°$ about $\vec{v}_2'$ as axis. In terms of the $\vec{v}'$-basis its matrix is (cf. Example 15.5(a))

$$A' = \begin{pmatrix} \cos 90^\circ & 0 & -\sin 90^\circ \\ 0 & 1 & 0 \\ \sin 90^\circ & 0 & \cos 90^\circ \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} .$$

Hence the same rotation in terms of the $\vec{v}$-basis, is given by

$$A = C^{-1}A'C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & 1/2 & 1/2 \\ 1/\sqrt{2} & 1/2 & 1/2 \end{pmatrix} .$$

In terms of coordinates of a point (x_1, x_2, x_3) and its transform (y_1, y_2, y_3) this is

$$\begin{aligned} y_1 &= \phantom{-x_1/\sqrt{2} +{}} x_2/\sqrt{2} - x_3/\sqrt{2}, \\ y_2 &= -x_1/\sqrt{2} + x_2/2 + x_3/2 , \\ y_3 &= x_1/\sqrt{2} + x_2/2 + x_3/2 . \end{aligned}$$

A similar procedure could obviously be used to get the equations of a rotation through any given angle about any given axis through the origin.

Problems

15.1 As in Example 15.4, find the matrix $A_{E^{-1}}$ representing E^{-1}, defined by $E^{-1}p(t) = p(t - 1)$. Show that $A_{E^{-1}} = A_E^{-1}$.

15.2 Show that the operator S of Example 15.2 is linear.

15.3 Prove that the rank and the determinant of a matrix are not changed by a similarity transformation.

15.4 If $\{\vec{v}_1', \ldots, \vec{v}_n'\}$ is a basis of V and C is any non-singular $n \times n$ matrix show that $\{\vec{v}_1, \ldots, \vec{v}_n\}$ defined by (15.7) is also a basis of V.

15.5 (Refer to Example 15.3).

(a) Show that for differential operators with non-constant coefficients factorization is not unique, by showing that

$$D^2 = DD = (D + \frac{1}{t + c})(D - \frac{1}{t + c})$$

for any constant c.

(b) If L_1 and L_2 are operators we can try to solve for v the operator equation $L_1L_2(v) = w$ by solving $L_1(z) = w$, $L_2(v) = z$. Solve $D^2y = 3t$ in two ways by using the two factorizations $D^2 = DD$ and $D^2 = (D + \frac{1}{t})(D - \frac{1}{t})$. [Hint. In the second case the general method of Chapter 1, Section 6 can be used.] Do the two answers agree?

15.6 The <u>shear</u> transformation considered in Example 9.3 and Problem 9.4 can be defined geometrically as follows: Given a fixed vector $\vec{z}$, each point of the plane is moved in the direction of $\vec{z}$ by an amount proportional to the directed distance from the point to the line of $\vec{z}$. We will call $\vec{z}$ the <u>direction</u> of the shear and the proportionality factor the <u>strength</u> of the shear.

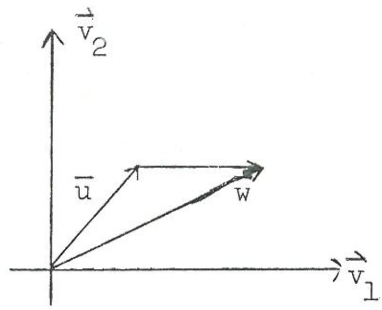


Figure 15.5

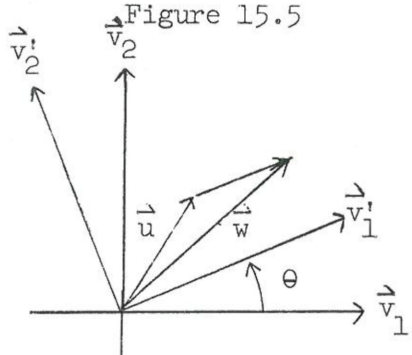


Figure 15.6

From Figure 15.5 it can be seen that a shear of strength c in the direction $\vec{v}_1$ gives the linear transformation $S(\vec{u}) = \vec{w}$ defined in terms of the basis $\{\vec{v}_1, \vec{v}_2\}$ by the matrix $\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$. We now ask: For the same basis, what is the matrix of the shear of strength c in the direction $\vec{v}_1'$ making an angle θ with $\vec{v}_1$? (Figure 15.6.)

Answer. $\begin{pmatrix} 1 - c\sin\theta\cos\theta & c\cos^2\theta \\ -c\sin^2\theta & 1 + c\sin\theta\cos\theta \end{pmatrix}$.

15.7 Let us investigate the result of change of basis on a linear transformation L from a space V to a different space W. Here we have two changes of basis to consider, one in V and one in W.

Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\{\vec{v}_1', \ldots, \vec{v}_n'\}$ be two bases in V related by

$$\vec{v}_i = \sum_{j=1}^{n} b_{ji}\,\vec{v}_j^{\,\prime}, \qquad B \text{ non-singular } n \times n,$$

and $\{\vec{w}_1, \ldots, \vec{w}_m\}$ and $\{\vec{w}_1^{\,\prime}, \ldots, \vec{w}_m^{\,\prime}\}$ two bases in W related by

$$\vec{w}_i = \sum_{j=1}^{m} c_{ji}\,\vec{w}_j^{\,\prime}, \qquad C \text{ non-singular } m \times m.$$

Let L be represented in the two pairs of bases by

$$L(\vec{v}_j) = \sum_{i=1}^{m} a_{ij}\,\vec{w}_i \quad \text{and} \quad L(\vec{v}_j^{\,\prime}) = \sum_{i=1}^{m} a_{ij}^{\prime}\,\vec{w}_i^{\,\prime}.$$

Show that $A' = CAB^{-1}$.

15.8 Prove that for transformations between two different spaces Question B can be answered as follows:

Let L be a linear transformation from a space V to a different space W. Let V and W have dimensions n and m respectively. Then there is a unique integer r, with $0 \leq r \leq \min(m,n)$, and a basis of V and one of W, such that the matrix A representing L with respect to these two bases has the property that $a_{11} = a_{22} = \ldots = a_{rr} = 1$ and all other $a_{ij} = 0$.

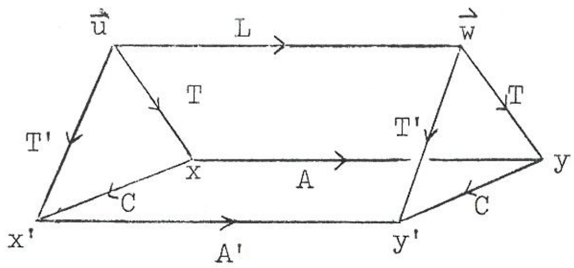
15.9 Let us designate the passage from the vector $\vec{u}$ to the n-tuple x, defined by

$$\vec{u} = \sum_{i=1}^{n} x_i\,\vec{v}_i,$$

by the symbol T, and from $\vec{u}$ to x', defined by

$$\vec{u} = \sum_{i=1}^{n} x_i' \, \vec{v}_i'$$

by T'. Then Figure 15.1 can be enlarged to



Discuss some of the relationships indicated in this figure and derive some identities connecting the various transformations.

16. Miscellaneous Problems and Applications.

16.1 Integral Equations.

(a) An example of an integral equation is

$$\int_a^b K(x,y)f(y)dy = f(x) + g(x) , \qquad \text{for } a \leq x \leq b. \tag{16.1}$$

Here the functions $K(x,y)$ and $g(x)$ are given and it is desired to find a function $f(x)$ such that (16.1) is satisfied. To find an approximation to the solution we approximate the integral by the Trapezoidal Rule to obtain

$$\begin{aligned} h[\tfrac{1}{2} K(x,y_0)f(y_0) + K(x,y_1)f(y_1) + \dots + K(x,y_{n-1})f(y_{n-1}) \\ + \tfrac{1}{2} K(x,y_n)f(y_n)] = f(x) + g(x). \end{aligned} \tag{16.2}$$

In general we cannot hope to satisfy this equation for all values of x from a to b. However, if we restrict x in the same manner as y, using only $n + 1$ equally spaced values from a to b, equation (16.2) is approximated by

$$\sum_{j=0}^{n} K_{ij}f_j = f_i + g_i , \qquad i = 0, \dots, n, \tag{16.3}$$

where $\qquad f_j = f(y_j), \qquad f_i = f(x_i), \qquad g_i = g(x_i),$

$$K_{ij} = \begin{cases} hK(x_i,y_j) & \text{if } j \neq 0,n, \\ \frac{1}{2} hK(x_i,y_j) & \text{if } j = 0,n, \end{cases}$$

(See Figure 16.1.) In matrix notation (16.3) can be written as

$$Kf = f + g \quad \text{or} \quad (K - I)f = g.$$

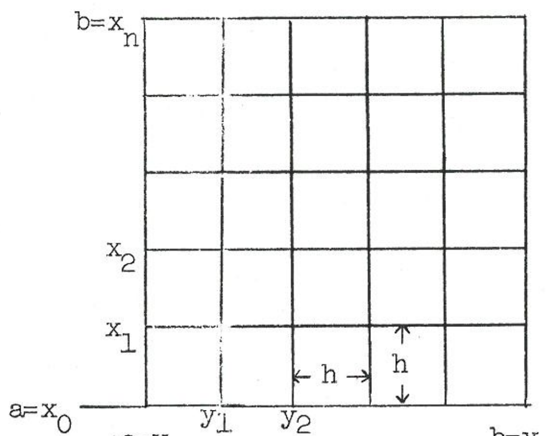


Figure 16.1

Thus the integral equation can be approximated by a system of linear equations.

Using this method find the approximate values of $f(0)$, $f(1/2)$ and $f(1)$, for the solution of the integral equation

$$(16.4) \qquad \int_0^1 (1 + xy)f(y)\,dy = f(x) + 3x^2, \qquad 0 \leq x \leq 1.$$

Answer. $f(0) = 4.7$, $f(1/2) = 5.0$, $f(1) = 4.0$.

(b) If we approximate the integral by Simpson's Rule instead of the Trapezoidal Rule we get the same result (16.3) but with a different definition for K_{ij}. Solve the above problem (16.4) using Simpson's Rule with $h = 1/2$.

Answer. $f(0) = \frac{25}{6}$, $f(1/2) = \frac{53}{12}$, $f(1) = \frac{19}{6}$.

[Note. The exact solution of (16.4) is $f(x) = \frac{25}{6} + 2x - 3x^2$, as can easily be checked.]

16.2 Random Walks.

(a) Figure 16.2 shows a map of a certain college compus. A totally drunken student is staggering along the paths shown. At each branch in the road he is equally likely to take any of the alternative paths, including the one along which he arrived. If he gets to West Street or North Avenue he will be seen by the policemen and will be apprehended. If he gets back to the fraternity house he will escape scot-free.

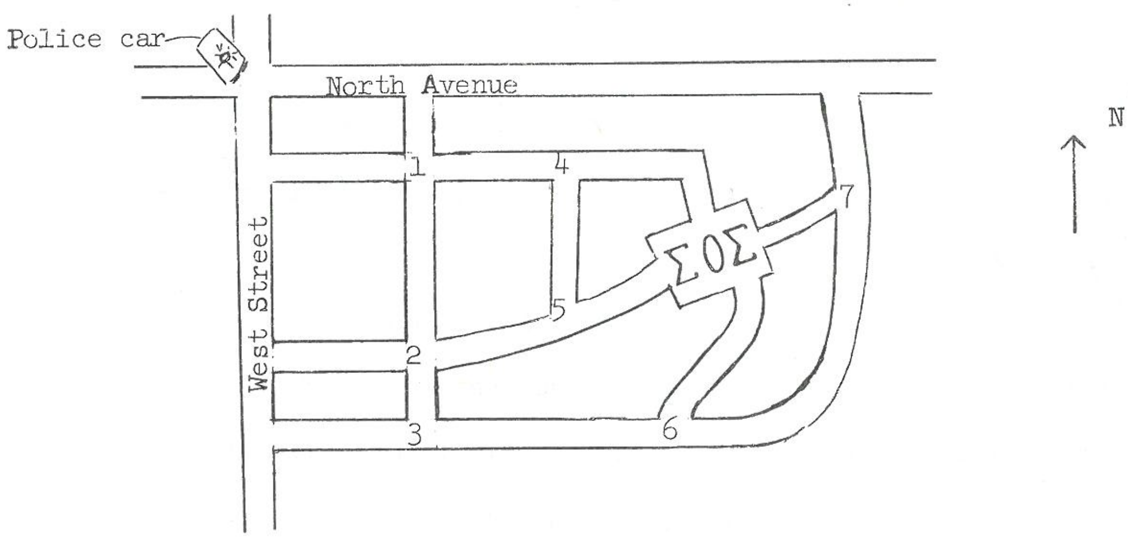


Figure 16.2

Let p_i be the probability that he will escape, given that he is at the i-th junction. Consider, for example, p_7. He has a choice of three paths, each with probability 1/3. If he goes north he is arrested. If he goes west (which he does with probability 1/3) he escapes. If he goes south (which he does with probability 1/3) he then has a probability p_6 of escaping; from probability theory, his probability of going south and then escaping is the product $(1/3)p_6$. Hence the total probability p_7 is $1/3 + 1/3(p_6)$. In this manner we get the following set of equations:

$$p_1 = \frac{1}{4} p_2 + \frac{1}{4} p_4 ,$$

$$p_2 = \frac{1}{4} p_3 + \frac{1}{4} p_5 + \frac{1}{4} p_1 ,$$

$$p_3 = \frac{1}{3} p_2 + \frac{1}{3} p_6 ,$$

$$p_4 = \frac{1}{3} p_1 + \frac{1}{3} p_5 + \frac{1}{3} ,$$

$$p_5 = \frac{1}{3} p_2 + \frac{1}{3} p_4 + \frac{1}{3} ,$$

$$p_6 = \frac{1}{3} p_3 + \frac{1}{3} p_7 + \frac{1}{3} ,$$

$$p_7 = \frac{1}{3} p_6 + \frac{1}{3} .$$

If the student is at junction 2, what is his probability of escaping? Answer. $515/1763 = 0.29$.

(b) Find the probability that a student escapes, starting from each junction of the roadwork of Figure 16.3. On the junctions marked C he is caught and on those marked E he escapes. [Hint. Use the symmetry of the figure to reduce the six equations to three.]

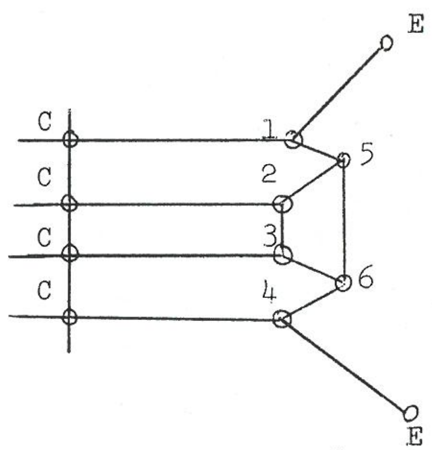


Figure 16.3

Random walk problems arise in many physical situations, notably in the escape of neutrons from a fission pile.

16.3 Quantitative Analysis?

A certain coed had a trying day in the Home Ec. Lab. She had been given four labelled jars and was told to record the weight of the contents of each jar before blending them in a saucepan. Not only did she forget to weigh them, but she also neglected to record what was on the labels. She did remember however that three of the jars were labelled Water, Salt and Sugar respectively. As for the fourth, she couldn't remember whether it was baking soda, cream of tartar, or glucose, but she was certain it was one of these three.

She took the saucepan to a friend who was studying Chemical Engineering. He analyzed the contents and reported that it contained 14 moles of C; 30 moles of H; 20 moles of O; 4 moles of Na and 2 moles of Cl and nothing else. He then wrote down:

Water: H_2O .	Baking soda: $NaHCO_3$.
Salt: NaCl .	Cream of tartar: $KHC_4H_4O_6$.
Sugar: $C_{12}H_{22}O_{11}$.	Glucose: $C_6H_{12}O_6$.

Fourth jar: $C_\alpha H_\beta O_\gamma (Na)_\delta (K)_\epsilon$.

What was in the fourth jar and how much did the contents of each jar weigh?

16.4 Mixtures.

Four commercial brands (Brand A, Brand B, Brand C, Brand D) of "mixed party nuts" have the following distribution of types of nut (in percentage).

Brand / Type of Nut	A	B	C	D
Peanuts	40	40	50	60
Walnut	10	20	20	10
Brazil Nut	20	20	10	10
Cashew	20	20	10	20
Almond	10	0	10	0
Price (¢/lb.)	60	50	70	80

Is it possible to mix together appropriate quantities of each brand so that all three of the following conditions are satisfied?

(1) The total cost is $3.00.

(2) There is one pound of cashews in the mix.

(3) The amount of peanuts, walnuts and brazil nuts in the mix are in the ratios 2:1:1.

16.5 Structures.

(a)

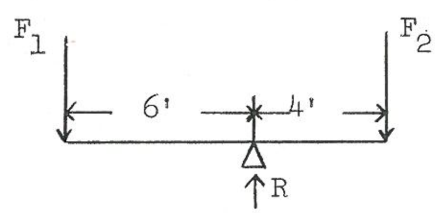


Figure 16.4

If R is not known, what can we say about F_1, F_2 in Figure 16.4? That is, describe the set of all vectors of the form $\begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$ where F_1, F_2 are in equilibrium. Is the set a linear subspace?

(b)

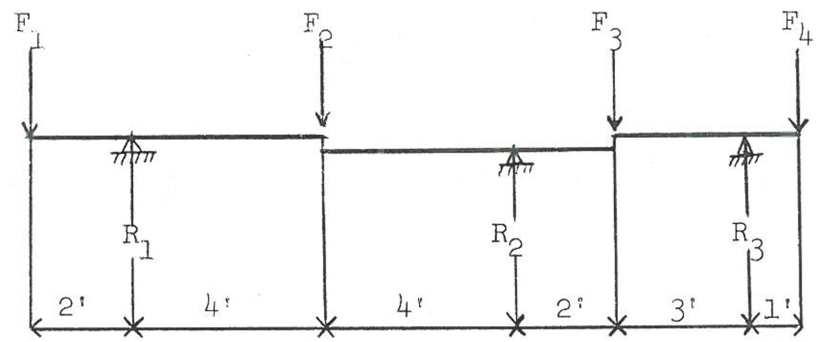


Figure 16.5

If none of the R_i are known in Figure 16.5 describe the set of all vectors of the form $\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{pmatrix}$, where F_1, F_2, F_3, F_4 are in equilibrium.

Is the set a linear subspace?

(c)

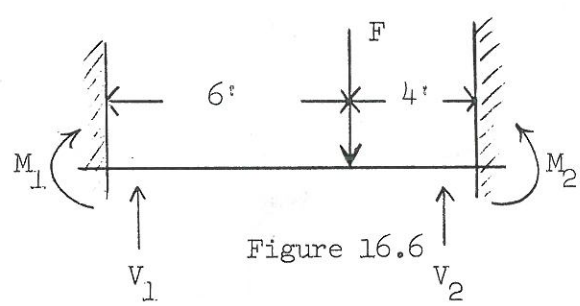


Figure 16.6

Figure 16.6 represents a "statically indeterminate structure." Load F is given. The vertical reactions V_1 and V_2 and the end moments M_1, M_2 cannot be determined by statics alone. However some relations exist among them. Describe the set of all vectors of the form $\begin{pmatrix} V_1 \\ V_2 \\ M_1 \\ M_2 \end{pmatrix}$, where the V_i's and M_i's are the quantities indicated in Figure 16.4. Is the set a linear subspace?

(d) Consider also the possible reactions H_1, H_2 as indicated in Figure 16.7.

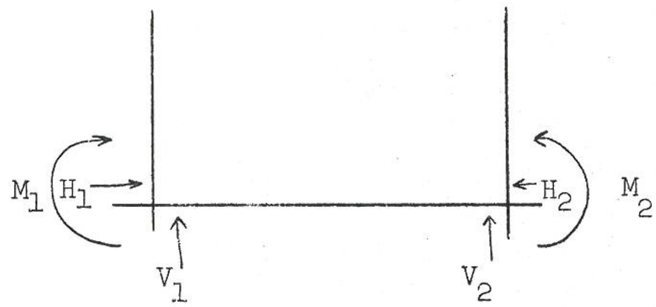


Figure 16.7

Describe the set of all vectors of the form $\begin{pmatrix} V_1 \\ V_2 \\ M_1 \\ M_2 \\ H_1 \\ H_2 \end{pmatrix}$. Is the set a linear subspace?

(e)

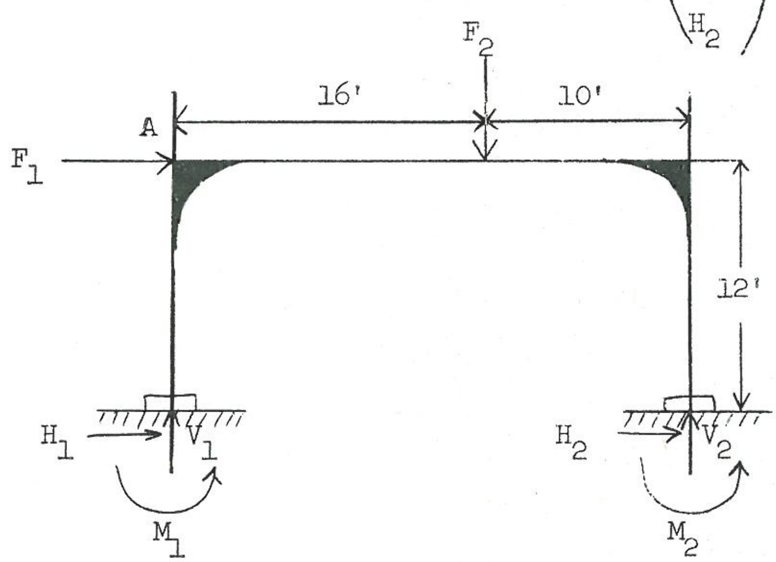


Figure 16.8

In Figure 16.8 loads F_1 and F_2 are known. Describe the set of all vectors of the form $\begin{pmatrix} H_1 \\ H_2 \\ V_1 \\ V_2 \\ M_1 \\ M_2 \end{pmatrix}$, where the entries represent reactions consistent with the laws of statics.

Answer: $\sum H = 0, \qquad F_1 + H_1 + H_2 = 0;$

$\sum V = 0, \qquad F_2 - V_1 - V_2 = 0;$

$\sum M_A = 0, \qquad 12H_1 + M_1 - 16F_2 + 26V_2 + 12H_2 + M_2 = 0;$

$$H_2 = -H_1 - F_1,$$

$$V_2 = F_2 - V_1,$$

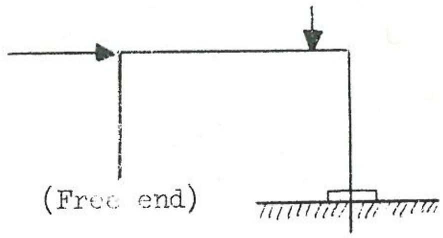
$$M_2 = -12H_1 - M_1 + 16F_2 - 26V_2 - 12H_2 = 12F_1 + 26V_1 - 10F_2 - M_1,$$

$$\begin{pmatrix} H_1 \\ H_2 \\ V_1 \\ V_2 \\ M_1 \\ M_2 \end{pmatrix} = \begin{pmatrix} H_1 \\ -H_1 - F_1 \\ V_1 \\ F_2 - V_1 \\ M_1 \\ 12F_1 - 10F_2 + 26V_1 - M_1 \end{pmatrix} = H_1 \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + V_1 \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 26 \end{pmatrix} + M_1 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ -F_1 \\ 0 \\ F_2 \\ 0 \\ 12F_1 - 10F_2 \end{pmatrix}$$

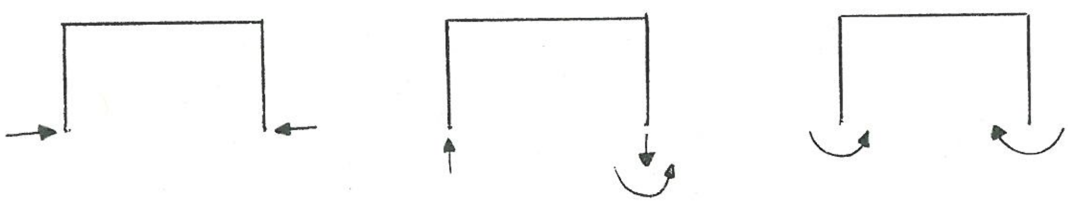
where H_1, V_1 and M_1 are arbitrary.

Intepretation: The last term $\begin{pmatrix} 0 \\ -F_1 \\ 0 \\ F_2 \\ 0 \\ 12F_1 - 10F_2 \end{pmatrix}$

gives the reactions for the structure

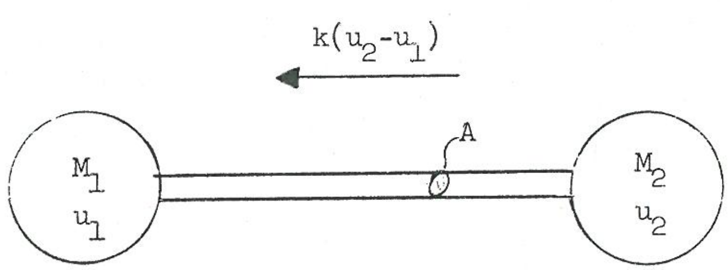


The first three terms give arbitrary multiples of "self-equilibrating" force systems:



16.6 <u>Steady State Temperature Distribution</u>.

(a) If M_1 and M_2 are two masses joined by a metal rod then heat will flow from the hotter to the colder mass, at a rate which is proportional to the temperature difference (see Figure 16.9).



Note: $k = \frac{KA}{\ell}$ where K is the coefficient of thermal conductivity, A is the cross-sectional area of the rod, and ℓ is its length.

Figure 16.9

If u_1 denotes the temperature of M_1, and c_1 denotes its specific heat, then the rate at which the heat content of M_1 increases is

$\frac{d}{dt}(M_1c_1u_1)$. Equating this to the heat arriving from M_2 we obtain the equations

$$M_1c_1\frac{du_1}{dt} = k(u_2 - u_1) ,$$

$$M_2c_2\frac{du_2}{dt} = k(u_1 - u_2) .$$

When the temperatures reach a constant value we have $\frac{du_1}{dt} = \frac{du_2}{dt} = 0$, so that $u_1 = u_2$ in the steady state.

By a similar heat balance carried out at each of the masses m_1, m_2, m_3, find the analogous system of differential equations for u_1, u_2, u_3 for the system shown in Figure 16.10. Assume all the rods and masses are of the same material and size. Find the steady state temperatures u_1, u_2, u_3, given that the temperatures u_4, u_5, u_6, u_7 have the constant values 0, 10, -10, 20.

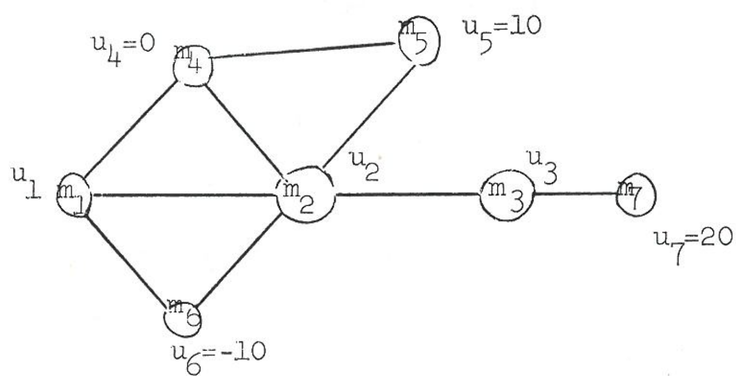


Figure 16.10

Answer. $(u_1,u_2,u_3) = \frac{1}{5}(-14,8,54)$.

(b) From the analysis of part (a) it follows easily that if all the bars are identical then, in the steady state, the temperature of any of the masses is equal to the average of the temperatures of the neighboring masses which connect directly to it. Use this to find the steady state temperature u_4 in the system shown in Figure 16.11 where the other temperatures shown are the constants indicated.

Answer. $u_4 = 426/89 = 4.79$.

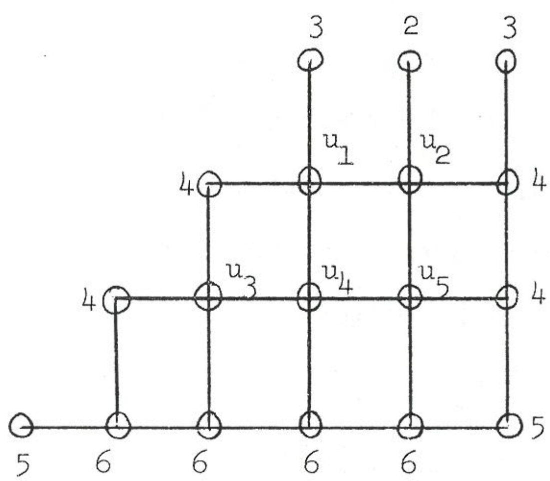


Figure 16.11

(c) If all the bars shown in Figure 16.12 are of the same material and cross section, set up the equations for the steady state temperatures u_1, u_2, u_3, u_4, u_5, u_6. [Note that the rate at which a bar transmits heat is inversely proportional to its length.]

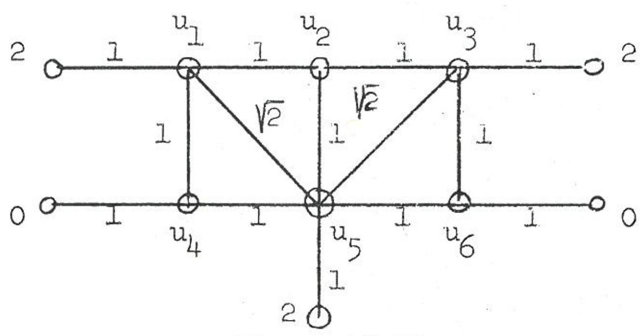


Figure 16.12

16.7 Leontief's Model of the Economy is sketched briefly below. Explain in more detail the steps which follow.

Suppose that there are n industries $I_1, I_2, \ldots, I_n$, turning out one product each, $P_1, P_2, \ldots, P_n$. Each product is sold to consumers and also to the industries. Let x_i denote the amount of product P_i produced by industry I_i. Let d_{ij} denote the amount of product P_i needed by industry I_j and let c_i denote the consumer demand for product P_i. Thus if the production rate is equal to the demand

$$(1) \qquad x_i = \sum_{j=1}^{n} d_{ij} + c_i .$$

Assume that the quantity of product P_i used in the production of P_j is proportional to the amount of P_j produced. Thus

$$(2) \qquad d_{ij} = a_{ij}x_j .$$

The a_{ij} are called "technological coefficients."

Hence

$$x_i = \sum_{j=1}^{n} a_{ij}x_j + c_i ,$$

or

$$(I - A)x = c .$$

Thus if one knows the technological coefficients and consumer demand, he can determine the production rates.

Let p_j denote the price of one unit of product P_j. The cost of materials for making one unit of P_j is $\sum_{i=1}^{n} a_{ij}p_i$. The value added per unit of product P_j by industry I_j is defined as the price minus the cost: $p_j - \sum_{i=1}^{n} a_{ij}p_i = v_j$; $(I - A^t)p = v$ or $(I - A)^t p = v$. The "value added" includes labor costs and profits; if it, and the technological coefficients, are known then the price can be determined.

16.8 It has been observed that chickens have a definite "pecking order", i.e. chicken A will peck chicken B but not conversely. Sometines, however, the following, apparently paradoxical, behavior pattern appears. Chicken A pecks chicken B, chicken B pecks chicken C, but chicken C pecks chicken A. This is called a cycle.

$$a_{ij} = 1 \text{ if chicken } i \text{ pecks chicken } j.$$
$$= 0 \text{ otherwise.}$$

Let $A = (a_{ij})$.

Prove that Trace (A^3) is equal to three times the number of cycles. (The Trace of a matrix is the sum of the diagonal elements).