Chapter 1 LINEAR FUNCTIONS

You probably recall from calculus that a **function** is a rule which associates particular values of one variable quantity to particular values of another variable quantity. **Analysis** is that branch of mathematics devoted to the study, or analysis, of functions. The main kind of analysis that goes on is this: for small changes in the first variable, we try to determine an approximate value to the corresponding change in the second. Now, we ask, for large changes in the first variable to what extent can we predict, from such approximations, the corresponding change in the second? The primary technique involved in this kind of analysis is **simplification** of the problem. That is, we *replace* the given function by a suitable very simple and more easily calculable function and work with this simple function instead (making sure to keep in mind the effect of that replacement).

The simplest possible functions are those which behave linearly. This means that they have a straight line as graph. Such a function has the following property. The increment in the value of the function corresponding to an increment in the variable is a constant multiple of that increment:

$$f(x+t) - f(x) = Ct \tag{1.1}$$

for some C. Now, when one moves to the consideration of functions of several variable quantities the study of even these simplest functions becomes complex enough that it forms a special mathematical discipline, called **linear algebra**. The calculus of one variable, coupled with the concepts and techniques of linear algebra constitute the basic tools of analysis of functions of several variables. It is our purpose in this text to study this subject. First then, we must study the notions and methods of linear algebra.

We begin our study in a familiar context: that of the solution of a system of simultaneous equations in more than one unknown. We shall develop a standard technique for discovering solutions (if they exist), called **row reduction**. This is the foundation pillar of the theory of linear algebra. After a brief section on notational conventions, we look at the system of equations from another point of view. Instead of seeking particular solutions of a system, we analyze the system itself. This leads us to consider the fundamental concept of linear algebra: that of a **linear transformation**. In this larger context we can resolve the question of existence of solutions and effectively describe the totality of all solutions to a given linear problem. We proceed then to analyze the totality of linear transformations as an object of interest in itself. This chapter ends with the study of several important topics allied with linear algebra. We study the plane as the system of complex numbers and the inner and vector products in three space.

1.1 Simultaneous Equations

Let us begin by considering a well-known problem: that of finding solutions to systems of simultaneous linear equations. The simplest nontrivial example is that of two equations in two unknowns.

Examples

1. $$\begin{aligned} 8x + 5y &= 3 \\ 7x - y &= 8 \end{aligned} \tag{1.2}$$

The technique for solution is that of elimination of one of the variables. This is accomplished by multiplying the equations by appropriate nonzero numbers and adding or subtracting the resulting equations. This is quite legitimate, for the set of solutions of the system will not be changed by any such operations. It is our intention to select such operations so that we eventually obtain as equations: $x =$ something, $y =$ something. In the present case this is quite easy: if we add five times the second equation to the first, y will conveniently disappear:

$$\begin{array}{l} 8x + 5y = 3 \\ \underline{35x - 5y = 40} \\ 43x \qquad\; = 43 \end{array}$$

and we obtain the equation $x = 1$. Substituting that in the first equation gives $8 + 5y = 3$, or $y = -1$. Then $x = 1$, $y = -1$ is the solution. Let us try a few more illustrative examples.

2. $$\begin{aligned} 3x - 2y &= 9 \\ -x + 3y &= 11 \end{aligned}$$

We can eliminate x as follows: multiply the second equation by 3 and add:

$$\begin{array}{r} 3x - 2y = 9 \\ -3x + 9y = 33 \\ \hline 7y = 42 \end{array}$$

We obtain $y = 6$ and $x = 7$.

3. $$\begin{aligned} 3x + 4y &= 7 \\ 6x + 8y &= 15 \end{aligned} \tag{1.3}$$

If we subtract twice the first equation from the second, we obtain a mess:

$$\begin{array}{r} 6x + 8y = 15 \\ -6x - 8y = 14 \\ \hline 0 = 1 \end{array} \tag{1.4}$$

Thus there can be no numbers x and y satisfying Equations (1.3), because they imply the Equation (1.4) which is patently false. Notice, if the second equation were

$$6x + 8y = 14$$

then our technique would lead to the equation $0 = 0$ which is true, but hardly offers much new information. We can conclude that our simple technique of elimination does not always produce results. We shall go into the causes for this in Section 1.3.

4. Let us now generalize our technique to systems involving more variables. Consider, for example, the system

$$\begin{aligned} x + y + z &= 5 \\ 3x - 2y + 5z &= -1 \\ 2x + y - z &= 0 \end{aligned} \tag{1.5}$$

The first equation expresses x in terms of y and z; if the second and third equation were free of x we could solve as above for the two unknowns y, z and then use the first to find x. But now it is easy to replace those last two equations by another two which must also be satisfied and which are free of the variable x. We use the first to eliminate x from the latter two. Namely, subtract three times the first from the second:

$$\begin{array}{r} 3x - 2y + 5z = -\ 1 \\ 3x + 3y + 3z = \ \ 15 \\ \hline -5y + 2z = -16 \end{array}$$

and twice the first from the third:

$$\begin{array}{r} 2x + y - z = \ \ 0 \\ 2x + 2y + 2z = \ \ 10 \\ \hline -y - 3z = -10 \end{array}$$

The system (1.5) has been replaced by this new system:

$$\begin{aligned} x + y + z &= 5 \\ -5y + 2z &= -16 \\ -y - 3z &= -10 \end{aligned} \tag{1.6}$$

and we can now see our way clear to the end. We solve the last two as a system in two unknowns:

$$\begin{array}{r} -5y + 2z = -16 \\ 5y + 15z = \ \ 50 \\ \hline 17z = \ \ 34 \\ z = \ \ \ 2 \end{array}$$

Then, substituting this value in the last equation, we obtain $-y - 6 = -10$ or $y = 4$. Finally, substituting these values for y and z in the first equation, we find $x = -1$. Thus the solution is $x = -1$, $y = 4$, $z = 2$.

5. $$\begin{aligned} x - y - z &= 5 \\ 2x + y - 3z &= 0 \\ -4x - y + z &= 10 \end{aligned}$$

Eliminate x from the second and third equations by adding appropriate multiples of the first;

$$\begin{array}{r} 2x + \;\;y - 3z = \quad 0 \\ 2x - 2y - 2z = \quad 10 \\ \hline 3y - \;\;z = -10 \end{array}$$

$$\begin{array}{r} -4x - \;\;y + \;\;z = 10 \\ 4x - 4y - 4z = 20 \\ \hline -5y - 3z = 30 \end{array}$$

The given system has been replaced by these equations:

$$\begin{aligned} x - y - \;\;z &= \quad 5 \\ 3y - \;\;z &= -10 \\ -5y - 3z &= \quad 30 \end{aligned}$$

We solve the last two easily: $y = -30/7$, $z = -20/7$. Substitutions into the first equation completes the solution: $x = -15/7$.

Of course, we can run into difficulties as we did in the two unknown equations of Example 3. We should be prepared for such occurrences and perhaps even more mysterious ones. Nevertheless, our technique is productive: if there is a solution to be had we can locate it by this process of successive eliminations. Furthermore, it easily generalizes to systems with more unknowns. This is the technique stated for the case of n unknowns. Eliminate the first variable from all the equations except the first by adding appropriate multiples of the first. Then, we handle the resulting equations as a system in $n - 1$ unknowns. That is, using the second equation we can eliminate the second variable from all but the second equation, using the new third equation we can eliminate the third variable from the remaining equations, and so forth. Eventually we run out of equations and we ought to be able to find the desired solution by a succession of substitutions.

We shall want to do more than discover solutions if they exist. We want to be able to predict the existence of solutions; we want to be able to compare systems, and we want to know in some sense how many solutions there are. In other words, we should come to understand the nature of a given system of equations. In order to do that we have to analyze this technique and develop a notation and theory which do so. That is where linear algebra begins. Before going into this, we study another pair of more complicated examples.

Examples

6.
$$\begin{aligned}
x + 2y - z - 3w &= 13\\
5x - y - z + 2w &= -14\\
y + z + w &= 4\\
3x + 2y - 2z &= -7
\end{aligned}$$

According to our technique, we replace the last three equations by a new set in which the variable x does not appear. We do this by adding the suitable multiple of the first equation:

$$\begin{aligned}
(-5) \times (\text{first}) + \text{second}: &\quad -11y + 4z + 17w = -79\\
0 \times (\text{first}) + \text{third}: &\quad y + z + w = 4\\
(-3) \times (\text{first}) + \text{fourth}: &\quad -4y + z + 9w = -46
\end{aligned}$$

Now we solve this set by applying the same procedure: we now eliminate y. Of course the order of the equations is not relevant; we could have listed them some other way. Since we can avoid fractions by adding multiples of the second equation to the first and third, let's do it that way.

$$\begin{aligned}
(11) \times (\text{second}) + \text{first}: &\quad 15z + 28w = -35\\
4 \times (\text{second}) + \text{third}: &\quad 5z + 13w = -30
\end{aligned}$$

Finally, of this set, $(-3) \times$ (second) + first gives $-11w = 55$. Thus the original set of four equations is replaced by this set:

$$\begin{aligned}
x + 2y - z - 3w &= 13\\
y + z + w &= 4\\
5z + 13w &= -30\\
-11w &= 55
\end{aligned}$$

The solutions are now easily found,

$$w = -5 \qquad z = 7 \qquad y = 2 \qquad x = 1$$

7. Now, let us consider this set:

$$\begin{aligned}
x + 2y + 3z + u - v &= 2\\
-5x + y + 7z &= -5\\
2y + 4u + 3v &= 18\\
3z - u - v &= -5
\end{aligned} \tag{1.7}$$

x is already eliminated from the last two equations. Using the first to eliminate x from the second, we obtain these three equations in place of the last three above,

$$\begin{aligned} 11y + 22z + 5u - 5v &= 5 \\ 2y + 4u + 3v &= 18 \\ 3z - u - v &= -5 \end{aligned} \tag{1.8}$$

Now y is already eliminated from the last. We eliminate it from the second (without getting involved with fractions) in this way:

$(-2) \times (\text{first}) + (11) \times (\text{second})$: $-44z + 34u + 43v = 188$

Now this equation together with the last of the set (1.8) gives this system

$$\begin{aligned} -44z + 34u + 43v &= 188 \\ 3z - u - v &= -5 \end{aligned}$$

We can eliminate v from the first to obtain $129z - 9u = -27$. Thus the system (1.7) has been transformed into this:

$$\begin{aligned} x + 2y + 3z + u - v &= 2 \\ 11y + 22z + 5u - 5v &= 5 \\ 3z - u - v &= -5 \\ 129z - 9u &= -27 \end{aligned} \tag{1.9}$$

Now we can solve for x by the first equation once we know y, z, u, v; we can solve for y by the second once we know z, u, v; we can solve for v in the third once we know z and u; and we can use any z, u which make the last equation true. For example, if $z = 0$, we must have $u = 3$, and so on up the line: $v = 2$, $y = 0$, $x = 1$. Notice that for any value of z we can always find u, v, x, y that make these equations all hold. Thus in this case there is more than one solution.

Formulation of the Procedure: Row Reduction

Now, let us turn to the abstract formulation of this procedure. In the general case we will have some, say m, equations in n unknowns. Let us refer to the unknowns as $x^1, \ldots, x^n$. These m equations may be written as

$$\begin{aligned} a_1{}^1x^1 + a_2{}^1x^2 + \cdots + a_n{}^1x^n &= b^1 \\ a_1{}^2x^1 + a_2{}^2x^2 + \cdots + a_n{}^2x^n &= b^2 \\ &\cdots \\ a_1{}^mx^1 + a_2{}^mx^2 + \cdots + a_n{}^mx^n &= b^n \end{aligned} \tag{1.10}$$

We proceed to solve this system as follows: multiply the first equation by $-a_1{}^2/a_1{}^1$ and add it to the second equation; multiply the first equation by $-a_1{}^3/a_1{}^1$ and add it to the third and so on. The result will be a new system, which we may write this way:

$$\begin{aligned} a_1{}^1x^1 + a_2{}^1x^2 + \cdots + a_n{}^1x^n &= b^1 \\ \alpha_2{}^2x^2 + \cdots + \alpha_n{}^2x^n &= \beta^2 \\ \cdots \\ \alpha_2{}^mx^2 + \cdots + \alpha_n{}^mx^n &= \beta^n \end{aligned} \tag{1.11}$$

We now continue with the same technique applied to the system of $m-1$ equations in $n-1$ unknowns given by the system (1.11) (except for the first equation). This is an effective reduction of the problem, because x^1 can be computed from the first equation once $x^2, \ldots, x^n$ are known. Of course, if $a_1{}^1 = 0$, this technique must be slightly modified. We just renumber the equations so that the coefficient of x^1 in the first one is nonzero and then proceed as above. If that is impossible then x^1 appears in no equation so we can disregard it and work with x^2 instead.

We now introduce a formalism which allows us to keep track of this procedure. It is clear that the essence of the left side of the system of Equations (1.10) is embodied in the array of numbers.

$$\mathbf{A} = \begin{pmatrix} a_1{}^1 & a_2{}^1 & \cdots & a_n{}^1 \\ a_1{}^2 & a_2{}^2 & \cdots & a_n{}^2 \\ & & \cdots & \\ & & \cdots & \\ & & \cdots & \\ a_1{}^m & a_2{}^m & \cdots & a_n{}^m \end{pmatrix} \tag{1.12}$$

This array is called a **matrix**: the upper index of the general term is the **row index** and the lower index is the **column index**. Thus, $a_5{}^3$ is the number in the third row and fifth column, a_{42}^7 is in the seventh row and forty-second column, $a_k{}^j$ is the number in the jth row and the kth column. Symbol (1.12) is an $m \times n$ **matrix**: it has m rows and n columns. The matrix

$$\mathbf{b} = \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}$$

is an $m \times 1$ matrix. Equations (1.10) can now be written symbolically as

$$\mathbf{Ax} = \mathbf{b} \tag{1.13}$$

Now the technique for solving the equation described above consists of a sequence of such equations with new matrices **A** and **b**, ending with one whose solution is obvious. The step from one equation to the next is performed by a **row operation** (remember, the rows are the separate equations); that is, one of these particular steps:

Step 1. Multiply a row by a nonzero constant.
Step 2. Add one row to another.
Step 3. Interchange two rows.

It is clear (and will be verified in Section 1.3) that any such operation does *not* change the collection of solutions. Finally, the end result desired is a matrix of this form, called a **row-reduced matrix**:

$$\begin{pmatrix} 1 & \alpha_2{}^1 & & \cdots & \alpha_n{}^1 \\ 0 & 1 & \alpha_3{}^2 & \cdots & \alpha_n{}^2 \\ 0 & 0 & \cdots & 1 & \cdots \\ & \vdots & \vdots & \ddots & \end{pmatrix} \tag{1.14}$$

Descriptively: the first nonzero entry of any row is a 1 and this 1 in any row is to the right of the 1 in any previous row. This is the kind of matrix the above procedure leads to; and it is most desirable because the system it represents can be immediately solved. In order to see this, we shall distinguish between two cases by resolving the dotted ambiguity in the lower right corner of (1.14).

Let

$$\mathbf{Ax} = \mathbf{b}$$

be a system of linear equations where **A** is a row-reduced matrix (of the form (1.14)). Let d be the number of nonzero rows of **A**.

Case 1. $d = n$. In this case the system of equations has this form:

$$\begin{aligned} x^1 + a_2{}^1x^2 \qquad\qquad + \cdots + a_n{}^1x^n &= b^1 \\ x^2 + a_3{}^2x^3 + \cdots + a_n{}^2x_n &= b^2 \\ \vdots \qquad\qquad & \quad \vdots \\ x^{n-1} + a_n^{n-1}x^n &= b^{n-1} \\ x^n &= b^n \\ 0 &= b^{n+1} \\ &\ \ \vdots \\ 0 &= b^m \end{aligned}$$

Thus there is a solution if and only if $b^{n+1} = \cdots = b^m = 0$, and the solution is found by successive substitutions: In this case the solution is unique.

Case 2. $d < n$. In this case our system has the form:

$$\begin{aligned}
x^1 + a_2{}^1x^2 \qquad\qquad &+ \cdots + a_n{}^1x^n = b^1 \\
x^2 + a_3{}^2x^3 &+ \cdots + a_n{}^2x^n = b^2 \\
&\vdots \\
x^d + a_{d+1}^d\, x^{d+1} &\cdots + a_n{}^dx^n = b^d \\
&0 = b^{d+1} \\
&\vdots \\
&0 = b^m
\end{aligned}$$

(We may have to reindex the variables in order to get all the leading 1's in a line.) There is a solution if and only if $b^{d+1} = \cdots = b^m = 0$, and all the solutions are obtained by giving $x^{d+1}, \ldots, x^n$ arbitrary values, and finding the values of the remaining variables by substitutions.

We now summarize the factual (rather than the procedural) content of this discussion in a theorem, the proof of which will appear in Section 1.3.

Definition 1. A matrix $\mathbf{A}$ is called a **row-reduced matrix** if

(i) the first nonzero entry in any row is 1,
(ii) in any row this first 1 appears to the right of the first 1 in any preceding row.

The number of nonzero rows of $\mathbf{A}$ is called its **index**.

Theorem 1.1. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix.* $\mathbf{A}$ *can be brought into row-reduced form* $\mathbf{A}'$ *by a succession of row operations. The equation* $\mathbf{Ax} = \mathbf{b}$ *has precisely the same solutions as the equation* $\mathbf{A}'\mathbf{x} = \mathbf{b}'$ *if* $\mathbf{b}'$ *is obtained from* $\mathbf{b}$ *by the same sequence of row operations that led from* $\mathbf{A}$ *to* $\mathbf{A}'$.

• EXERCISES

1. Find solutions for these systems

(a) $$\begin{aligned} 2x - 3y &= 23 \\ 3x + y &= -4 \end{aligned}$$

(b) $$\begin{aligned} \tfrac{1}{2}x + 4y &= 10 \\ -\tfrac{3}{2}x + 8y &= 0 \end{aligned}$$

(c) $$\begin{aligned} x + y + z &= 15 \\ x - y + z &= 3 \\ 2x - 3y - 5z &= -7 \end{aligned}$$

(d)
$$\begin{aligned} x+y+z+w&=4\\ x+y+z-w&=2\\ x-y+z-w&=0\\ -x+2y-3z+4w&=2 \end{aligned}$$

(e)
$$\begin{aligned} -x+y+z&=0\\ 2x+2y+10z&=28\\ x+y+z&=22 \end{aligned}$$

(f)
$$\begin{aligned} 3x+6y+9z&=12\\ x+2y+3z&=4 \end{aligned}$$

(g)
$$\begin{aligned} x+y&=7\\ x-y&=1\\ 3x-4y&=0 \end{aligned}$$

(h)
$$\begin{aligned} x+y&=7\\ x+2y&=9\\ x+3y&=11 \end{aligned}$$

(i)
$$\begin{aligned} x+y+z+w&=4\\ x+y-z-w&=6 \end{aligned}$$

(j)
$$\begin{aligned} x+2y+z&=0\\ x-3y-6z&=4\\ 4x+8y+4z&=11 \end{aligned}$$

2. A **homogeneous system** of linear equations is a system of the form $\mathbf{Ax}=\mathbf{0}$; that is, the right-hand side is zero. Find nonzero solutions (if possible) to these homogeneous systems.

(a)
$$\begin{aligned} x+y+z&=0\\ x-y+z&=0\\ x+2y+z&=0 \end{aligned}$$

(b)
$$\begin{aligned} x+y+z&=0\\ x-y-z&=0 \end{aligned}$$

(c)
$$\begin{aligned} x+y+z+w&=0\\ x-2y+z-2w&=0\\ 2x-y+2z-w&=0 \end{aligned}$$

3. Suppose $(x^1,\ldots,x^n)$ is a solution for a given homogeneous system. Show that for every real number t, $(tx^1,\ldots,tx^n)$ is also a solution.

4. If $(x^1,\ldots,x^n)$, $(y^1,\ldots,y^n)$ are solutions for a given homogeneous system, then so is $(x^1+y^1,\ldots,x^n+y^n)$.

5. Find the row-reduced matrix which corresponds to the given matrix according to Theorem 1.1.

$$\mathbf{A}=\begin{pmatrix} 0 & 7 & 1\\ 3 & 2 & 2\\ -1 & 6 & 4 \end{pmatrix}$$

$$\mathbf{B}=\begin{pmatrix} 1 & 0 & 0 & 6 & 5\\ 2 & 3 & 0 & 0 & 0\\ 1 & 0 & -1 & -1 & 1\\ 0 & 0 & 0 & 2 & 0 \end{pmatrix}$$

$$\mathbf{C}=\begin{pmatrix} 1 & 2 & 6 & 1 \\ -2 & -4 & 0 & 2 \\ 0 & 0 & 8 & 8 \\ 3 & 6 & 9 & 12 \end{pmatrix}$$

6. Let

$$\mathbf{a}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad \mathbf{b}=\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \qquad \mathbf{c}=\begin{pmatrix} 3 \\ 2 \\ 0 \\ 2 \end{pmatrix}$$

Solve these equations:

(a) $\mathbf{Ax}=\mathbf{b}$, (b) $\mathbf{Bx}=\mathbf{a}$, (c) $\mathbf{Bx}=\mathbf{c}$, (d) $\mathbf{Cx}=\mathbf{a}$,
(e) $\mathbf{Cx}=\mathbf{c}$, where **A**, **B**, **C** are given in Exercise 5.

• PROBLEMS

1. Show that the system

$$ax+by=\alpha$$
$$cx+dy=\beta$$

has a solution no matter what α, β are if $ad-bc \neq 0$, and there is only one such solution.

2. Can you suggest an explanation of the ugly phenomenon illustrated by Example 3?

3. Is there only one row-reduced matrix to which a given matrix may be reduced by row operations? If $\mathbf{A}'$ and $\mathbf{A}''$ are two such row-reduced matrices, coming from a given matrix **A** show that they must have the same index.

4. Suppose we have a system of n equations in n unknowns, $\mathbf{Ax}=\mathbf{b}$. After row reduction the index of the row-reduced matrix is also n. Show that in this case the equation $\mathbf{Ax}=\mathbf{b}$ always has one and only one solution for every **b**.

5. Suppose that you have a system of m equations in n unknowns to solve. What should you expect in the way of existence and uniqueness of solutions in the cases $m<n$, $m>n$?

6. Suppose we are given the $n \times n$ system $\mathbf{Ax}=\mathbf{b}$, and all the rows of **A** are multiples of the first row; that is, there are $s^1, \ldots, s^n$ such that $a_j{}^i=s^i a_j{}^1$ for all j and $i=1, \ldots, n$. Under what conditions will the given system have a solution?

7. Suppose instead that the columns of **A** are multiples of the first column. Can you make any assertions?

8. Verify that if the columns of a 3×3 matrix are multiples of the first column, then the rows are multiples of one of the rows.

1.2 Numbers, Notation, and Geometry

We now interrupt our discussion of simultaneous equations in order to introduce certain facts and notational conventions which shall be in use throughout this text. We shall also describe the geometry of the plane from the linear, or vector point of view as an alternative introduction to linear algebra.

First of all, the collection $\{1, 2, 3, \ldots\}$ of positive integers (the "counting numbers") will be denoted by P. Every integer n has an immediate successor, denoted by $n + 1$. If a fact is true for the integer 1 and also holds for the successor of every integer for which it is true, then it is true for all integers. This is the Principle of Mathematical Induction, which we shall take as an axiom, or defining property of the integers. We shall formulate it this way.

Principle of Mathematical Induction. Let S be a subset of P with these properties:

(i) 1 is a member of S,
(ii) whenever a particular integer n is in S, so also is its successor $n + 1$ in S.

Then S must be the set P of all positive integers.

This assertion is intuitively clear. You can see, for example that 2 is in S. For by (i) 1 is in S, and thus by (ii) $1 + 1 = 2$ is also in S. Continuing, $3 = 2 + 1$ is in S, again by (ii). By applying (ii) another time 4 is in S. Applying (ii) another 32 times we see that all the integers up to 36 are also in S. No positive integer can escape: since 1 is in S we need only apply (ii) n times to verify that the integer n is in S. In fact, the assertion of the principle of mathematical induction is that there are no integers other than those that can be captured in this way, and in this sense the principle is a defining property of the integers.

The principle of mathematical induction provides us with a tool for writing proofs of assertions for all positive integers which avoids the phrases: "continuing in this way," "and so forth," "...," .... We shall find this a helpful device in verifying assertions concerning problems with an unspecified number, n, of unknowns. Let us illustrate this method by proving a few propositions about integers.

Proposition 1. *The sum of the first n integers is* $(1/2)n(n+1)$.

Proof. Let S be the set of integers for which Proposition 1 is true. Certainly 1 is in S:

$$1 = \tfrac{1}{2} \cdot 1(1+1)$$

Now, assuming the assertion of Proposition 1 for any integer n, we show that it also holds for $n+1$.

$$\begin{aligned} 1 + \cdots + n + 1 &= 1 + \cdots + n + n + 1 = \tfrac{1}{2}n(n+1) + n + 1 \\ &= (n+1)(\tfrac{1}{2}n + 1) = \tfrac{1}{2}(n+1)(n+2) \end{aligned}$$

which is the appropriate conclusion. Thus by the principle of mathematical induction, Proposition 1 is proven.

Proposition 2. *The sum of the first n odd integers is* n^2.

Proof. $1 = 1^2$ surely. We now assume the proposition for any n, and show that it follows for $n+1$:

$$\begin{aligned} 1 + 3 + \cdots + 2(n+1) - 1 &= 1 + 3 + \cdots + 2n - 1 + 2n + 1 \\ &= n^2 + 2n + 1 = (n+1)^2 \end{aligned}$$

Proposition 3. *Let* K *be a given positive integer. Then for any integer* n *we can write*

$$n = QK + R \tag{1.15}$$

with $0 \le R < K$ *in one and only one way.*

Proof. We may of course immediately discard the case $K = 1$ for in that case (1.15) is just the trivial comment that $n = n \cdot 1$ for all n. Thus take $K > 1$, and now proceed by mathematical induction. The proposition is true for $n = 1$:

$$1 = 0 \cdot K + 1$$

Now we assume that the proposition is true for any given integer n. Thus

$$n = QK + R$$

for some Q and R, $0 \le R < K$. Then $R + 1 \le K$. If $R + 1 < K$, we have

$$n + 1 = QK + (R + 1)$$

with $0 \leq R+1 < K$, as desired. Otherwise, $R+1=K$, in which case

$$n+1 = QK + K = (Q+1)K + 0$$

as desired. Thus, by mathematical induction, (1.15) is possible for every integer n. This representation is unique, for if

$$n = Q'K + R'$$

is also possible with $0 \leq R' < K$, then we have

$$Q'K + R' = QK + R$$

or

$$(Q' - Q)K = R - R'$$

and $R - R'$ is between $-K$ and K. Now the only multiple of K between $-K$ and K is 0, so $(Q' - Q)K = R - R' = 0$ from which we conclude $R = R'$, $Q' = Q$.

Set Notation

The set of positive integers forms a subset of a larger number system, the set Z of all integers. Z consists of all positive integers, their negatives and 0. The collection of all quotients of members of Z is the set of **rational numbers**, denotedy by Q. Q is a very large subset of the set of all **real numbers** R. For the purposes of geometric interpretation we will conceive of the real number system R as being in one-to-one correspondence with the points on a straight line. That is, given a straight line, we fix two points on it, one is the origin O, and the other denotes the unit of measurement. All other points P on the line are given a numerical value: it is the displacement from O as measured on the given scale (negative if O is between P and 1 and positive otherwise). (See Figure 1.1.)

There are certain ideas and notations in connection with sets which we

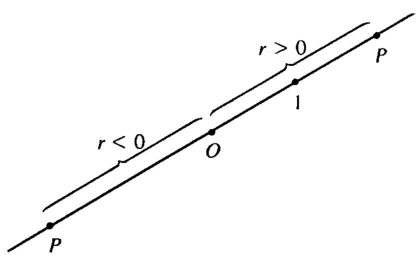


Figure 1.1

shall standardize before proceeding. Customarily, in any given context there is one largest set of objects under consideration, called the **universe** (it may be the positive integers P, or the rabbits in Texas, or the people on the moon) and all sets are actually subsets of this universe. If X is a set and x is an object we shall write $x \in X$ to mean: x **is a member of** X. $x \notin X$ means that x **is not a member of** X. Thus, for example, $-7 \in Z$, but $-7 \notin P$. The set with no elements is called the **empty set**, and is designated $\varnothing$. Most specific sets are defined by a property: the set in question is the set of all elements of the universe that have that property. We use the following short-hand form to represent that phrase

$$\{x \in U: x \text{ has that property}\}$$

For example, the set of all positive real numbers is $\{x \in R: x > 0\}$. The set of all Englishmen who drink coffee is $\{x \in \text{England}: x \text{ drinks coffee}\}$. The set of all integers between 8 and 18 is $\{x \in Z: 8 \leq x \leq 18\}$. This is the same as $\{x \in P: 8 \leq x \leq 18\}$ and $\{x \in Z: |x - 13| \leq 5\}$.

If X and Y are two sets, and every element of X is an element of Y we shall say that X **is contained in** Y, written $X \subset Y$. Notice that $\varnothing \subset X$ for every set X. We shall consider also these operations on sets:

$-X$: the set of all x not in X
$X \cup Y$: the set of all x in either X or Y (or both)
$X \cap Y$: the set of all x in both X and Y
$X - Y$: the set of all x in X, and not in Y

(Consult Figure 1.2 for a pictorial interpretation.) Notice that $X - Y$ is the same as $X \cap -Y$. There are many other identities: $--X = X$, $-X \cup -Y = -(X \cap Y)$, $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$, and so on, so don't be surprised when two different collections of symbols identify the same set. A final operation is that of forming the Cartesian product. If U is a given universe, then $U \times U$ is **the set of all ordered pairs** of elements in U. $U \times U$ is often denoted by U^2. By extension we can define U^3 as **the set of all ordered triples** (x^1, x^2, x^3) of elements of U; and more generally U^n is **the set of all ordered** n**-tuples** of elements of U.

If $X^1, \ldots, X^n$ are subsets of U, the set of all ordered n-tuples $(x^1, \ldots, x^n)$ with $x^1 \in X^1, \ldots, x^n \in X^n$ is denoted $X^1 \times \cdots \times X^n$. Not every subset of U^n is of the form $X^1 \times \cdots \times X^n$, those which are of this form are called **rectangles**.

Thus the space of n-tuples of real numbers is denoted R^n. If $I^1, \ldots, I^n$ are intervals in R, then $I^1 \times \cdots \times I^n$ is indeed a rectangle. A point $(x^1, \ldots, x^n)$ in R^n will be denoted, when specific reference to its elements is

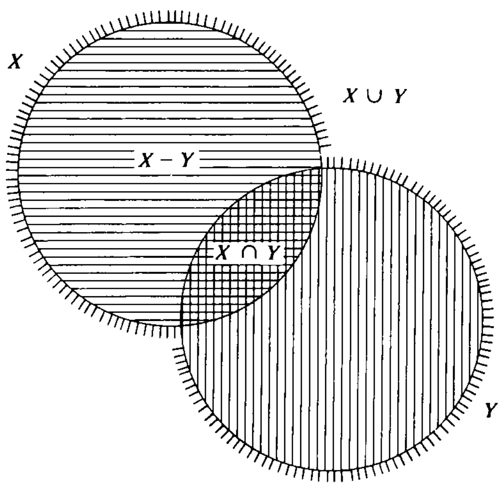


Figure 1.2

not required, by a single boldfaced letter $\mathbf{x} = (x^1, \ldots, x^n)$. If $\mathbf{a} = (a^1, \ldots, a^n)$, $\mathbf{b} = (b^1, \ldots, b^n)$ are two points in R^n with $b^i > a^i$, $1 \leq i \leq n$ we shall use the notation $[\mathbf{a}, \mathbf{b}]$ to denote the rectangle.

$$\{(x^1, \ldots, x^n) \colon a^1 \leq x^1 \leq b^1, \ldots, a^n \leq x^n \leq b^n\}$$

If the inequalities are strict we shall denote the rectangle by $(\mathbf{a}, \mathbf{b})$:

$$(\mathbf{a}, \mathbf{b}) = \{(x^1, \ldots, x^n) \colon a^1 < x^1 < b^1, \ldots, a^n < x^n < b^n\}$$

A function from a set X to another set Y is a rule which associates to each point of x a uniquely determined point y in Y. It is customary to avoid the use of the new word *rule* by defining a function as a certain kind of subset of $X \times Y$. Namely, a **function** is a set of ordered pairs (x, y) with $x \in X$, $y \in Y$, with each $x \in X$ appearing precisely once as a first member. If (x, y) is such a pair we denote y by $f(x)$: $y = f(x)$. We shall use the notation $f\colon X \to Y$ to indicate that f is to mean a function from X to Y. X is called the **domain of** f; the **range of** f is the set $\{f(x) \colon x \in X\}$ of values of f. If every point of y appears as a value of f we say that f maps X **onto** Y. If every point of y is the value of f at at most one x in X, we say that f is **one-to-one**. More precisely, f is one-to-one if $x \neq x'$ implies $f(x) \neq f(x')$. Now, if f is a one-to-one function from X onto Y, then for each $y \in Y$, there is one and only one $x \in X$ such that $f(x) = y$. This defines a function $g\colon Y \to X$ which is also one-to-one and onto and has this property: $g(y) = x$ if and only if $f(x) = y$. In this case we shall say that f **is invertible** and g is its **inverse**, denoted

$g = f^{-1}$. Finally, if $f_1 : X \to Y, f_2 : Y \to Z$ are two functions, we can compose them to form a third function, denoted $f_2 \circ f_1 : X \to Z$ defined by

$$f_2 \circ f_1(x) = f_2(f_1(x))$$

Plane Geometry

We now turn to the geometric study of the plane, as an alternative introduction to linear algebra. According to the notion of the Cartesian coordinate system we can make a correspondence between a plane, supposed to be of infinite extent in all directions, and the collection R^2 of ordered pairs of real numbers. This is done in the following way: first a point on the plane is chosen, to be called **the origin** and denoted O (Figure 1.3). Then two distinct lines intersecting at O are drawn (it is ordinarily supposed that these lines are perpendicular, but it is hardly necessary). These lines are called **the coordinate axes**; they are sometimes referred to more specifically as the x and y axes, ordered counterclockwise (Figure 1.4). Now a point is chosen on each of these axes; we call these E_1 and E_2 (Figure 1.5). These are the "unit lengths" in each of the directions of the coordinate axes. Having chosen a unit on these lines, we can put each of them in one-to-one correspondence with the real numbers. Now, letting P be any point in the plane, we associate a pair of real numbers to P in this way. Draw the lines through P which are parallel to the coordinate axes and let x be the intersection with the line through E_1 and y the intersection with the line through

Figure 1.3

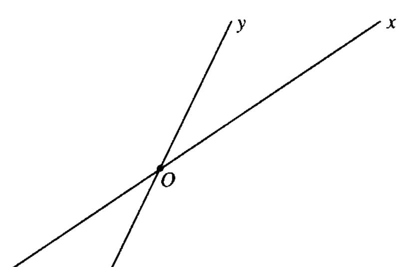


Figure 1.4

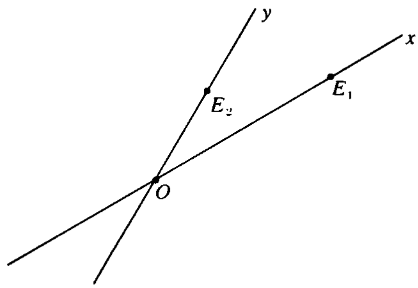


Figure 1.5

E_2. Then we identify P with the pair of real numbers (x, y) (Figure 1.6.) In this way to every point in the plane there corresponds a point in R^2 (called its **coordinates** relative to the choice O, E_1, E_2). Clearly, for any pair of real numbers (x, y) we have a point with those coordinates, namely the fourth vertex of the parallelogram of side lengths x and y along the coordinates axes with one vertex at O (Figure 1.6).

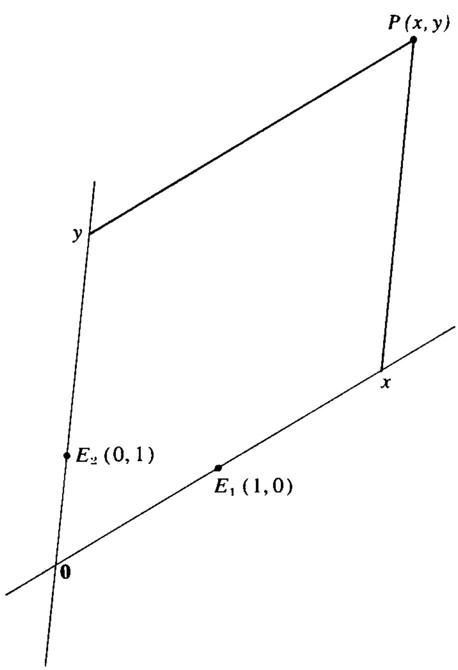


Figure 1.6

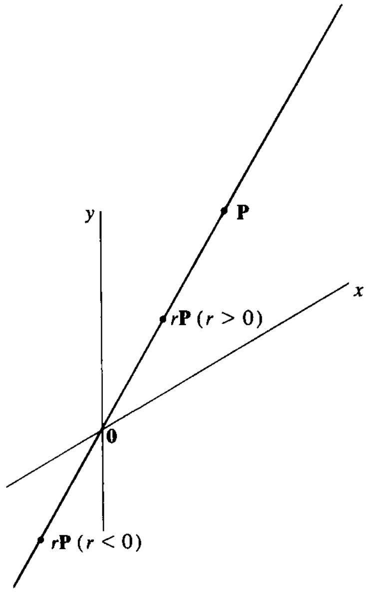


Figure 1.7

Once a particular point in the plane is fixed as the origin, there can be defined two operations on the points of the plane, and these operations form the tools of linear algebra. Since they cannot be defined on the points until an origin is chosen, we are forced to distinguish between the point set **the plane** and **the plane with chosen point**. This distinction gives rise to the notion of **vector**: a **vector** is a point in the plane-with-origin. The vector can be physically realized as the directed line segment from the origin to the given point; such a visualization is nothing more than a heuristic aid. It is important to realize that as sets, the set of vectors in the plane is the same as the set of points in the plane. The difference is that the set of vectors has additional structure: a particular point has been designated the origin. We shall denote vectors by boldface letters; thus the point P becomes the vector $\mathbf{P}$, the origin O becomes the vector $\mathbf{0}$. We shall now describe these two operations geometrically and then compute them in coordinates.

1. **Scalar Multiplication.** Let $\mathbf{P}$ be a vector in the plane, and r a real number. Consider the line through $\mathbf{0}$ and $\mathbf{P}$. Considering $\mathbf{P}$ now as a unit length, we can put that line into one-to-one correspondence with R. Using this scale, $r\mathbf{P}$ is the point corresponding to the real number r. Said differently, $r\mathbf{P}$ is one of the points on that line whose distance from $\mathbf{0}$ is $|r|$ times the distance of $\mathbf{P}$ from $\mathbf{0}$ (Figure 1.7). Now, if $\mathbf{P}$ has the coordinates (x, y) we shall see that $r\mathbf{P}$ has coordinates (rx, ry). First, suppose $r > 0$. Draw the

triangle formed by the line through **0**, **P** and $r\mathbf{P}$, and the $\mathbf{E}_1$ axis and the lines parallel to the $\mathbf{E}_2$ axis (Figure 1.8). Triangles I and II are similar. Thus, referring to the lengths as denoted in Figure 1.8,

$$\frac{|\mathbf{P}|}{|r\mathbf{P}|} = \frac{x}{s}$$

By definition $|\mathbf{P}|/|r\mathbf{P}| = 1/r$, thus the first coordinate of $r\mathbf{P}$ (here denoted by s), is rx. The second coordinate is similarly seen to be ry. Thus $r\mathbf{P}$ has the coordinates (rx, ry). The case $r < 0$ is only slightly more complicated.

2. **Addition.** Let **P**, **Q** be vectors in R^2. Then **0**, **P**, **Q** are three vertices of a uniquely determined parallelogram. We define $\mathbf{P} + \mathbf{Q}$ to be the fourth vertex. The description of this operation in terms of coordinates is extremely simple: if **P** has coordinates (x, y), and **Q** has coordinates (s, t), then $\mathbf{P} + \mathbf{Q}$ has $(x + s, y + t)$ as coordinates. There is nothing profound to be learned from the verification of this fact, so we shall not go through it in detail. After all, it is not our purpose here to logically incorporate plane geometry

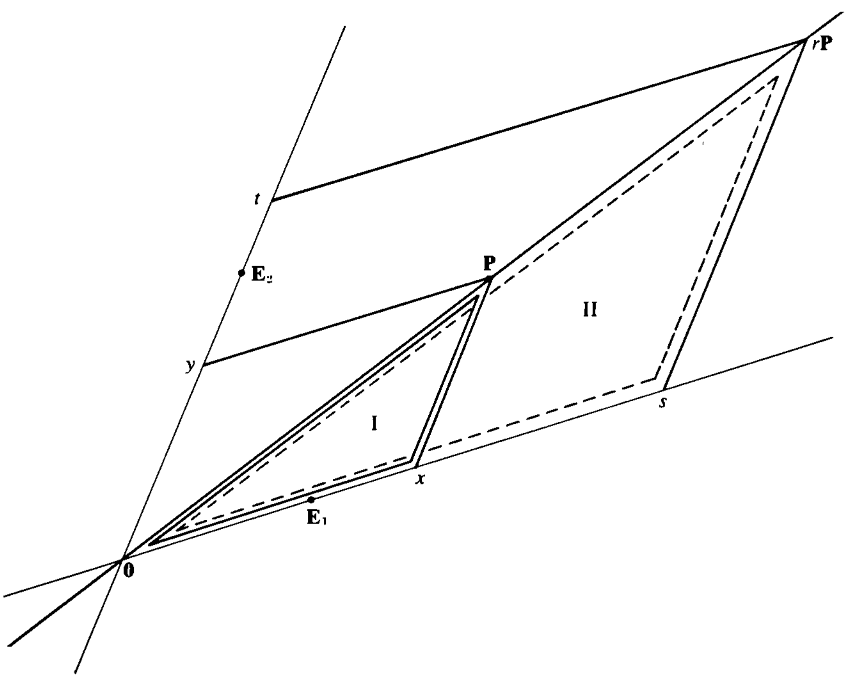


Figure 1.8

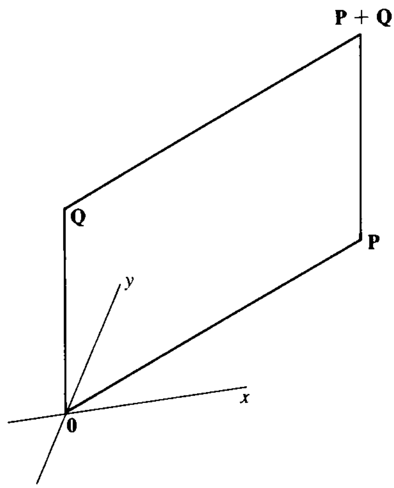


Figure 1.9

into our mathematics, but rather to use it as an intuitive tool for comprehension. For those who are suspicious of our assurances we include the verification of a special case. Consider Figure 1.9 and include the data relating to the coordinates (Figure 1.10): To show that the length of the line segment **0B** is $s + x$, we must verify that the length of **AB** is s. Draw the line through P and parallel to $\mathbf{0E}_1$, and let **C** be the intersection of that line with the line through **P** + **Q** and **B**. The quadrilateral **ABCP** has parallel sides and thus is a parallelogram. Hence, **AB** and **PC** have the same length. Now triangles $\mathbf{0}s\mathbf{Q}$ and **PC(P** + **Q)** have pairwise parallel sides (as shown in Figure 1.10) and further **0Q** and **P(P** + **Q)** have the same length. Thus these triangles are congruent so the length of **PC** is the same as the length of $\mathbf{0}s$, namely s. Thus the length of **AB** is also s, and so **0B** has length $s + x$. Notice that this is a special case since it refers to an explicit picture which does not cover all possibilities.

The operation inverse to addition is **subtraction**: **P** − **Q** is that vector which must be added to **Q** in order to obtain **P** (Figure 1.11). The best way to visualize **P** − **Q** is as the directed line segment running from the head of **Q** to the head of **P** (denoted **L** in Figure 1.11). In actuality, **P** − **Q** is the vector obtained by translating **L** to the origin; in practice, it is customary not to do this but to systematically confuse a vector with any of its translates. We shall do this only for purposes of pictorial representation.

Notice that, having chosen the vectors $\mathbf{E}_1$ and $\mathbf{E}_2$, we can express any vector in the plane uniquely in terms of them and the operations of addition

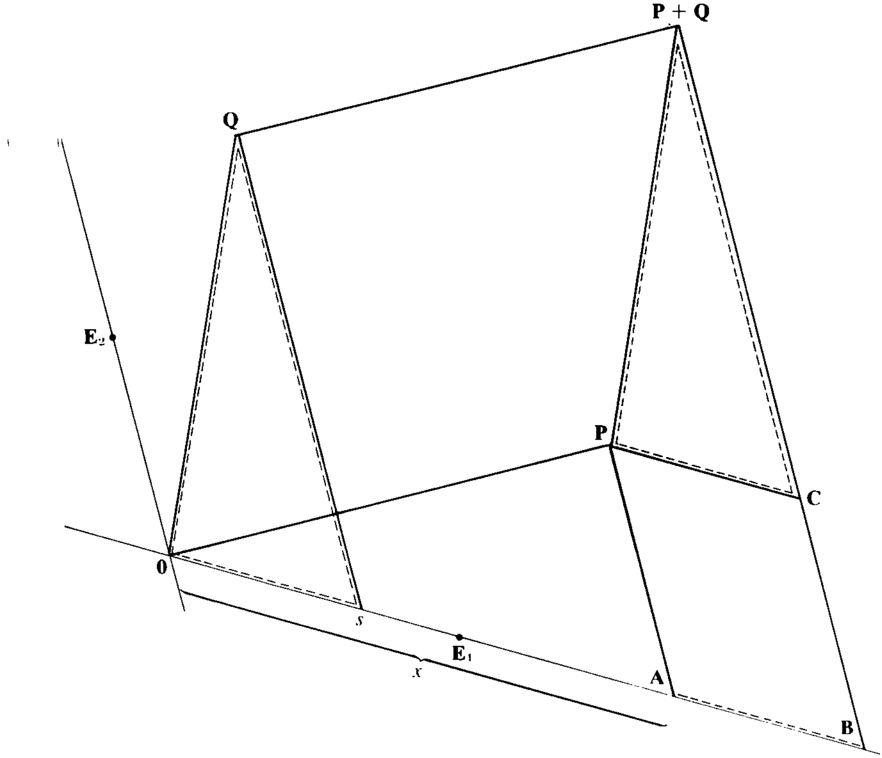


Figure 1.10

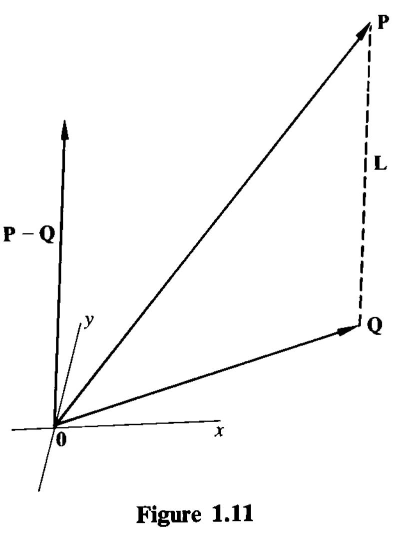


Figure 1.11

and scalar multiplication:

$$(x, y) = (x, 0) + (0, y) = x(1, 0) + y(0, 1) = x\mathbf{E}_1 + y\mathbf{E}_2$$

This is true no matter how $\mathbf{E}_1$ and $\mathbf{E}_2$ are chosen, so long as the points $\mathbf{0}, \mathbf{E}_1, \mathbf{E}_2$ do not lie on the same line (we say the *vectors* $\mathbf{E}_1$ *and* $\mathbf{E}_2$ *are not collinear*). Thus we have this important fact.

Proposition 4. *Let $\mathbf{E}_1$ and $\mathbf{E}_2$ be any two noncollinear vectors in the plane. Then we can write any vector $\mathbf{Q}$ uniquely as*

$$\mathbf{Q} = x^1\mathbf{E}_1 + x^2\mathbf{E}_2$$

x^1 and x^2 are the coordinates of $\mathbf{Q}$ relative to the choice of origin $\mathbf{0}$ and the vectors $\mathbf{E}_1$ and $\mathbf{E}_2$.

If we state this geometric fact purely as a fact about R^2, it turns out to be a theoretical assertion about the solvability of a pair of linear equations. Thus, let us suppose $\mathbf{E}_1 = (a_1{}^1, a_1{}^2)$, $\mathbf{E}_2 = (a_2{}^1, a_2{}^2)$ relative to some standard coordinate system (for example, the usual rectangular coordinates). First of all, how do we express algebraically the assertion that $\mathbf{E}_1$ and $\mathbf{E}_2$ do not lie on the same line? We need an algebraic description of a *straight line* through the origin.

Proposition 5.

(i) *A set L is a straight line through $\mathbf{0}$ if and only if there exists $(a, b) \in R^2$ such that*

$$L = \{(x, y) : ax + by = 0\}$$

(ii) *The points (x, y), (x', y') lie on the same line through the origin if and only if*

$$\frac{x}{x'} = \frac{y}{y'}$$

You certainly recall these facts from analytic geometry—we leave the verification to the exercises. Returning to the vectors $\mathbf{E}_1$ and $\mathbf{E}_2$, these geometric facts become the following algebraic fact.

Proposition 6. *Let*

$$\mathbf{A} = \begin{pmatrix} a_1{}^1 & a_2{}^1 \\ a_1{}^2 & a_2{}^2 \end{pmatrix}$$

be a 2×2 matrix with nonzero columns.

(i) If $a_1{}^1a_2{}^2 - a_1{}^2a_2{}^1 \neq 0$, then the equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution for every $b \in R^2$.

(ii) The equation $\mathbf{Ax} = \mathbf{0}$ has a nonzero solution if and only if $a_1{}^1a_2{}^2 = a_1{}^2a_2{}^1$.

(iii) If $a_1{}^1a_2{}^2 = a_1{}^2a_2{}^1$, the equation $\mathbf{Ax} = \mathbf{b}$ has a solution if and only if $a_1{}^1b^2 = a_1{}^2b^1$.

Proof.

(i) This condition is according to Proposition 5 (ii) precisely the assertion that the vectors $(a_1{}^1, a_1{}^2)$, $(a_2{}^1, a_2{}^2)$ are not collinear. Then, according to Proposition 4, for any (b^1, b^2) there is a unique pair (x^1, x^2) such that

$$(b^1, b^2) = x^1(a_1{}^1, a_1{}^2) + x^2(a_2{}^1, a_2{}^2)$$

This is the same as the pair of equations $\mathbf{b} = \mathbf{Ax}$.

(ii) By the above, if $a_1{}^1a_2{}^2 = a_1{}^2a_2{}^1$, then the only solution of $\mathbf{Ax} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. On the other hand, if $a_1{}^1a_2{}^2 = a_1{}^2a_2{}^1$, then either $(a_2{}^2, -a_1{}^2)$ or $(-a_2{}^1, a_2{}^2)$ is a nonzero solution of $\mathbf{Ax} = \mathbf{0}$, or all the entries of $\mathbf{A}$ are zero, in which case everything solves $\mathbf{Ax} = \mathbf{0}$.

(iii) If $a_1{}^1a_2{}^2 = a_1{}^2a_2{}^1$, then $(a_1{}^1, a_1{}^2)$, $(a_2{}^1, a_2{}^2)$ lie on the same line through the origin by Proposition 5(ii). Any combination $x^1(a_1{}^1, a_1{}^2) + x^2(a_2{}^1, a_2{}^2)$ will have to lie on that line, and conversely, any point on that line must be such a combination. Thus $\mathbf{Ax} = \mathbf{b}$ has a solution if and only if (b^1, b^2) lies on the line through $\mathbf{0}$ determined by $(a_1{}^1, a_1{}^2)$. The equation for this is, by Proposition 5(ii),

$$\frac{b^1}{a_1{}^1} = \frac{b^2}{a_1{}^2} \quad \text{or} \quad b^2a_1{}^1 = b^1a_1{}^2$$

Examples

8. Let L_1 be the line through (0, 0) and (3, 2) and L_2 the line through (1, 1) and (0, 6). Find the point of intersection of L_1 and L_2.

L_1 has the equation $2x - 3y = 0$, and L_2 the equation $5x + y = 6$. The point of intersection must lie on both lines, and thus is the pair (x, y) solving

$$\begin{aligned} 2x - 3y &= 0 \\ 5x + y &= 6 \end{aligned}$$

We find $x = 18/17$, $y = 12/17$.

9. Find the line L through the point (7, 3) that is parallel to the line L': $8x + 2y = 17$. L will be given by an equation of the form

$ax + by = c$. In order to be parallel to L', L and L' must have no point of intersection, so the equations

$$\begin{aligned} 8x + 2y &= 17 \\ ax + by &= c \end{aligned}$$

can have no common solution. Thus we must have

$$\frac{8}{a} = \frac{2}{b}$$

Furthermore, since (7, 3) is on L, we must also have

$$7a + 3b = c$$

This pair of equations in three unknowns has for a solution $a = 4$, $b = 1$, $c = 31$. Thus L is given by the equation

$$4x + y = 31$$

• EXERCISES

7. Show that for every integer n,

$$1^2 + 2^2 + \cdots + n^2 = \tfrac{1}{6}n(n+1)(2n+1)$$

8. Show that for every integer n,

$$2 + 2^2 + \cdots + 2^n = 2^{n+1} - 2$$

9. Show that $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ and $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$.

10. Give an example of a subset of a Cartesian product which is not a rectangle.

11. Find the point of intersection of these pairs of lines in R^2:

(a) $\begin{aligned} 3x + y &= 7 \\ x - 17y &= 1 \end{aligned}$ (c) $\begin{aligned} 2x + 2y &= -1 \\ x + 12y &= 14 \end{aligned}$

(b) $\begin{aligned} x - 2y &= 4 \\ 2x + y &= 0 \end{aligned}$ (d) $\begin{aligned} y &= 2x + 7 \\ x &= 3y + 18 \end{aligned}$

12. Find the line through $\mathbf{P}$ which is parallel to L:

(a) $\mathbf{P} = (2, -1)$, L: $3x + 7y = 4$

(b) $\mathbf{P} = (8, 1)$, L: $x - y = -1$

(c) $\mathbf{P} = (0, -7)$, L: $y - 2x = 3$

• PROBLEMS

9. We can define the line through $\mathbf{P}$ and $\mathbf{Q}$ as the set of all $\mathbf{X}$ such that the vector $\mathbf{X} - \mathbf{P}$ is parallel to the vector $\mathbf{P} - \mathbf{Q}$. Show that two vectors are parallel if and only if one is a multiple of the other. Conclude that the line

through **P** and **Q** is the set

$$\{\mathbf{P} + t(\mathbf{P} - \mathbf{Q}) : t \in R\}$$

10. Using the definition in Exercise 9 show that a straight line in the plane is, in terms of coordinates, given as

$$\{(x, y) \in R^2 : ax + by + c = 0\}$$

for suitable a, b, c.

11. Suppose L is a line given by the equation $bx + ay + c = 0$
    (a) Show that the tangent of the angle this line makes with the horizontal is $-b/a$.
    (b) Show that the vector (a, b) is perpendicular to L.
    (c) Find the point on L which is closest to the origin.
12. Find the line through the point **P** and perpendicular to L:
    (a) $\mathbf{P} = (3, 7)$, L: $x - 3y = 2$.
    (b) $\mathbf{P} = (-1, 1)$, L: $2x + 3y = 0$.
    (c) $\mathbf{P} = (0, 2)$, L: $5x = 2y$.
13. Suppose coordinates have been chosen in the plane. Let $\mathbf{E}_1$, $\mathbf{E}_2$ be two vectors in the plane which are not collinear. (That is, $\mathbf{0}$, $\mathbf{E}_1$, $\mathbf{E}_2$ do not lie on a straight line.) Then we can recoordinatize the plane relative to this choice of principal directions. Give formulas which relate these two coordinatizations in terms of the given coordinates of $\mathbf{E}_1$, $\mathbf{E}_2$ (see Figure 1.12).

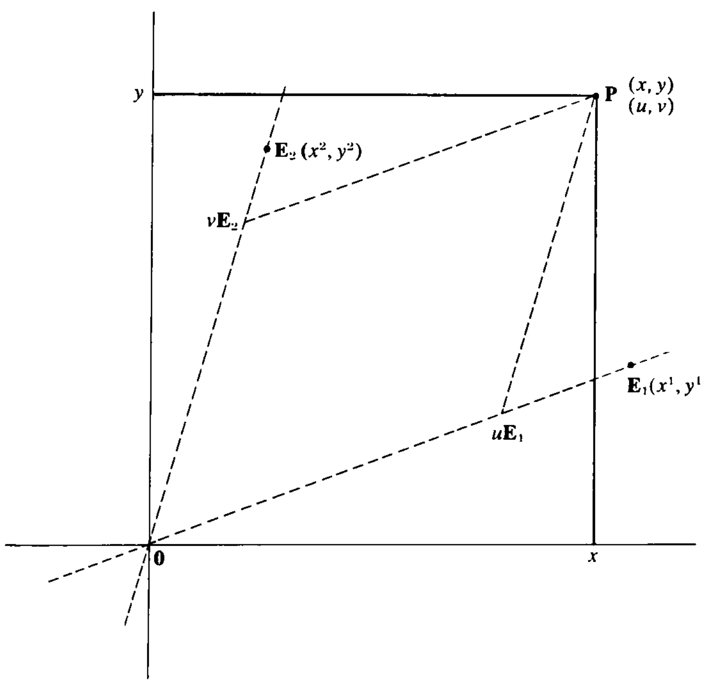


Figure 1.12

14. In the text, the equation $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$ was verified by induction. There is another way of doing this. An $n \times n$ matrix has n^2 entries. There are n of these entries on the diagonal and $1 + 2 + \cdots + n - 1$ entries both above and below that diagonal. Thus

$$2(1 + 2 + \cdots + n - 1) + n = n^2$$

1.3 Linear Transformations

We now return to the problem of analyzing systems of simultaneous linear equations, with a broader question in mind: given the $m \times n$ matrix **A**, for which **b** is there a solution of the equation $\mathbf{Ax} = \mathbf{b}$? In order to study this, we associate to **A** the function from R^m to R^n:

$$f_A(x^1, \ldots, x^n) = (a_1{}^1x^1 + \cdots + a_n{}^1x^n, \ldots, a_1{}^mx^1, \ldots, a_n{}^mx^n) \qquad (1.16)$$

Thus the set of **b**, such that $\mathbf{Ax} = \mathbf{b}$, is precisely the range of f_A.

Let us begin by introducing the two fundamental operations on R^n (just as in the case $n = 2$ studied in the previous section):

1. **Scalar Multiplication**: for $r \in R$, $x = (x^1, \ldots, x^n) \in R^n$, define

$$rx = (rx^1, \ldots, rx^n)$$

2. **Addition**: for $x = (x^1, \ldots, x^n)$, $y = (y^1, \ldots, y^n) \in R^n$, define

$$x + y = (x^1 + y^1, \ldots, x^n + y^n)$$

Definition 2. A function f from R^n to R^m is a **linear transformation** if it preserves these two operations, that is, if

$$\begin{aligned} f(r\mathbf{x}) &= rf(\mathbf{x}) \\ f(\mathbf{x} + \mathbf{y}) &= f(\mathbf{x}) + f(\mathbf{y}) \end{aligned}$$

The function f_A defined above for the $m \times n$ matrix **A** is linear:

$$\begin{aligned} f_A(r\mathbf{x}) &= (a_1{}^1rx^1 + \cdots + a_n{}^1rx^n, \ldots, a_1{}^mrx^1 + \cdots + a_n{}^mrx^n) \\ &= (r(a_1{}^1x^1 + \cdots a_n{}^1x^n), \ldots, r(a_1{}^mx^1 + \cdots + a_n{}^mx_n)) \\ &= rf_A(\mathbf{x}) \end{aligned}$$

$$\begin{aligned}f_A(\mathbf{x}+\mathbf{y}) &= (a_1{}^1(x^1+y^1)+\cdots+a_n{}^1(x^n+y^n),\ldots,a_1{}^m(x^1+y^1)\\ &\quad+\cdots+a_n{}^m(x^n+y^n))\\ &= (a_1{}^1x^1+\cdots+a_n{}^1x^n+a_1{}^1y^1+\cdots+a_n{}^1y^n,\ldots,\\ &\quad a_1{}^mx^1+\cdots+a_n{}^mx^n+a_1{}^my^1+\cdots+a_n{}^my^n)\\ &= f_A(\mathbf{x})+f_A(\mathbf{y})\end{aligned}$$

The significance of the introduction of linear transformations, from the point of view of systems of equations, is that it provides a context in which to consistently interpret the technique of row reduction. For, the application of a row operation to a system of equations amounts to composition of the associated linear transformation with a particular linear transformation corresponding to the row operation. Once we have seen that we can analyze the given system by studying these successive compositions. Looking ahead, it is even more important to recognize row reduction as a tool for analyzing linear transformations. Let us now interpret the row operations as linear transformations.

Type I. Multiply a row by a nonzero constant. Consider the multiplication of the rth row by $c \neq 0$. Let P_1 be the transformation on R^m:

$$P_1(b^1,\ldots,b^m) = (b^1,\ldots,cb^r,\ldots,b^m)$$

(multiplication of the rth entry by c). The effect of this row operation is that of composing the transformation f_A: $R^n \to R^m$ with P_1, and changing the equation $\mathbf{Ax} = \mathbf{b}$ into the equation $\mathbf{P}_1\mathbf{Ax} = \mathbf{P}_1\mathbf{b}$. These two equations have the same set of solutions since the transformation P_1 can be reversed (it is **invertible**). Precisely, its inverse is given by multiplying the rth entry by $1/c$.

Type II. Add one row to another. Adding the rth row to the sth row corresponds to this transformation on R^m:

$$P_2(b^1,\ldots,b^m) = (b_1,\ldots,b^r,\ldots,b^s+b^r,\ldots,b^m)$$

Again, this step in the solution of the equations amounts to transforming the equation $\mathbf{Ax} = \mathbf{b}$ to $\mathbf{P}_2\mathbf{Ax} = \mathbf{P}_2\mathbf{b}$. Since P_2 is invertible (what is its inverse?) we cannot have affected the solutions.

Type III. Interchange two rows. Interchanging the rth and sth rows corresponds to the transformation

$$P_3(b^1,\ldots,b^r,\ldots,b^s,\ldots,b^m) = (b^1,\ldots,b^s,\ldots,b^r,\ldots,b^m)$$

The importance of these observations is this: the row operations correspond to linear transformations which in turn are representable by matrices. The solution of the system of equations $\mathbf{Ax} = \mathbf{b}$ thus can be accomplished *completely in terms of manipulations with the matrix corresponding to the system.* It is our purpose now to study the representation of linear transformations by matrices and the representation of composition of transformations.

In R^n the n vectors $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ play a fundamental role. We shall refer to them as $\mathbf{E}_1, \ldots, \mathbf{E}_n$, respectively. Thus $\mathbf{E}_i$ has all entries zero, but the ith, which is 1.

Proposition 7. *Any vector in R^n has a unique representation as a linear combination of $\mathbf{E}_1, \ldots, \mathbf{E}_n$.*

Proof. Obviously,

$$(b^1, \ldots, b^n) = b^1\mathbf{E}_1 + \cdots + b^n\mathbf{E}_n$$

We shall refer to the set of vectors $\mathbf{E}_1, \ldots, \mathbf{E}_n$ as the *standard basis for R^n*. Out of Proposition 7 comes this more illuminating fact.

Proposition 8. *Corresponding to any linear transformation $L\colon R^n \to R^m$ there is a unique $m \times n$ matrix $(a_j{}^i)$ such that*

$$L(x^1, \ldots, x^n) = \left(\sum_{j=1}^n a_j{}^1 x^j, \ldots, \sum_{j=1}^n a_j{}^m x^j\right) \tag{1.17}$$

Proof. It is clear, by the way, that, given the matrix $(a_j{}^i)$, Equation (1.17) does define a linear transformation. Now, given L, since it is linear, we can write

$$L((x^1, \ldots, x^n)) = L(x^1\mathbf{E}_1 + \cdots + x^n\mathbf{E}_n) = x^1L(\mathbf{E}_1) + \cdots + x^nL(\mathbf{E}_n) \tag{1.18}$$

Thus a linear transformation is completely determined by what it does to the standard basis. Let

$$L(\mathbf{E}_1) = (a_1{}^1, \ldots, a_1{}^m), \ldots, L(\mathbf{E}_n) = (a_n{}^1, \ldots, a_n{}^m)$$

Then Equation (1.18) becomes

$$\begin{aligned}
L((x^1, \ldots, x^n)) &= x^1(a_1{}^1, \ldots, a_1{}^m) + \cdots + x^n(a_n{}^1, \ldots, a_n{}^m) \\
&= (x^1a_1{}^1, \ldots, x^1a_1{}^m) + \cdots + (x^na_n{}^1, \ldots, x^na_n{}^m) \\
&= (x^1a_1{}^1 + \cdots + x^na_n{}^1, \ldots, x^1a_1{}^m + \cdots + x^na_n{}^m)
\end{aligned}$$

which is just Equation (1.17).

Matrix Multiplication

Now we must discover how to represent the composition of two linear transformations by an operation on matrices. There is only one way to make this discovery: compute. Suppose then that $T\colon R^n \to R^m$ and $S\colon R^m \to R^p$ are linear transformations represented by the matrices $(a_j{}^i)$ and $(b_j{}^i)$, respectively. Then, we can compute the composition ST as follows:

$$T(x^1, \ldots, x^n) = \left(\sum_{j=1}^{n} a_j{}^1 x^j, \ldots, \sum_{j=1}^{n} a_j{}^m x^j\right)$$

$$ST(x^1, \ldots, x^n) = \left(\sum_{k=1}^{m} b_k{}^1\left(\sum_{j=1}^{n} a_j{}^k x^j\right), \ldots, \sum_{k=1}^{m} b_k{}^p\left(\sum_{j=1}^{n} a_j{}^k x^j\right)\right)$$

$$= \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{m} b_k{}^1 a_j{}^k\right)x^j, \ldots, \sum_{j=1}^{n}\left(\sum_{k=1}^{m} b_k{}^p a_j{}^k\right)x^j\right)$$

Thus ST is represented by the $p \times n$ matrix

$$\left(\sum_{k=1}^{m} b_k{}^i a_j{}^k\right) \tag{1.19}$$

Definition 3. Let $\mathbf{A} = (a_j{}^i)$ be an $m \times n$ matrix and $\mathbf{B} = (b_j{}^i)$ a $p \times m$ matrix. Then the **product BA** is defined as the $p \times n$ matrix whose (i, j)th entry is given by (1.19).

The preceding discussion thus provides the verification of

Proposition 9. *If $T\colon R^n \to R^m$, $S\colon R^m \to R^p$ are represented by the matrices* **A**, **B**, *respectively, then ST is represented by the product* **BA**.

The product operation may seem a bit obscure at first sight; but it is easily described in this way: the (i, j)th entry of **BA** is found by multiplying entry by entry the ith row of **B** to the jth column of **A**, and adding.

Examples

10. $$\mathbf{A} = \begin{pmatrix} 5 & 3 & 7 \\ 6 & 5 & 1 \\ 8 & 11 & -4 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 6 & 1 & 0 \\ -3 & 2 & 5 \\ 4 & 4 & 4 \end{pmatrix}$$

Let $\mathbf{AB} = (c_j{}^i)$. Then

$$
\begin{array}{lllll}
c_1{}^1 = 5.6 + & 3(-3) + & 7.4 & = & 49 \\
c_1{}^2 = 6.6 + & 5(-3) + & 1.4 & = & 25 \\
c_1{}^3 = 8.6 + & 11(-3) + & (-4)4 & = & -1 \\
\vdots & \vdots & & & \\
c_3{}^2 = 6.0 + & 5.5 + & 1.4 & = & 29 \\
\vdots & \vdots & & &
\end{array}
$$

$$
\mathbf{AB} = \begin{pmatrix} 49 & 39 & 43 \\ 25 & 20 & 29 \\ -1 & 14 & 39 \end{pmatrix}
$$

11.
$$
(2 \quad 5 \quad 1)\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = (-1 \cdot 2 + 0 \cdot 5 + 1 \cdot 1) = (-1)
$$

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

12.
$$
\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 7 & 0 \\ -1 & 4 & -2 \end{pmatrix} = \begin{pmatrix} -1 & -7 & 0 \\ -1 & 4 & -2 \end{pmatrix}
$$

$$
\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 7 & 0 \\ -1 & 4 & -2 \end{pmatrix} = \begin{pmatrix} -1 & 4 & -2 \\ 1 & 7 & 0 \end{pmatrix}
$$

$$
\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 7 & 0 \\ -1 & 4 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 11 & -2 \\ -1 & 4 & -2 \end{pmatrix}
$$

Now, let us recapitulate the discussion of this section so far. The problem of systems of m linear equations in n unknowns amounts to describing the range of a linear transformation $T: R^n \to R^m$. The technique of row reduction corresponds to composing T by a succession of invertible transformations on R^m. These transformations are those which provide the row operations; we shall call them elementary transformations. Linear transformations can be represented by means of the standard basis by matrices, and composition of the transformations corresponds to matrix multiplication. Thus, we solve a system of linear equations as follows: Multiply the matrix on the left by a succession of elementary matrices in order to obtain a row-reduced matrix. Then we can easily read off the solutions. Since multiplication by an elementary matrix is the same as applying the corresponding row operation to the matrix it is easy to keep track of this process.

Examples

13. Let us consider the system of four equations in three unknowns corresponding to the matrix

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

We shall record the process of row reduction in two columns. In the first we shall list the succession of transformations which A undergoes and in the second we shall accumulate the products of the corresponding elementary matrices.

(a) Multiply the third row by -1 and interchange it with the first,

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

(b) Multiply the first row by 3 and subtract it from the second; multiply the first row by 4 and subtract it from the third.

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

(c) Divide the second row by 2 and the third row by 5.

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

(d) Subtract the second row from and add one-half the third row to the fourth.

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 5/2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 1/2 & 3/2 & 0 \\ 1/5 & 0 & 4/5 & 0 \\ 1/10 & -1/2 & -11/10 & 1 \end{pmatrix}$$

Let us denote the product of the elementary matrices by $\mathbf{P}$; thus $\mathbf{P}$ is the last matrix on the right and the matrix on the left is $\mathbf{PA}$. Now, it is easy to see that if $\mathbf{PAx} = \mathbf{y}$ has a solution, the fourth entry of $\mathbf{y}$ must be zero. Now our original problem

$$\mathbf{Ax} = \mathbf{b}$$

has a solution if and only if $\mathbf{PAx} = \mathbf{Pb}$ is solvable (since $\mathbf{P}$ is invertible). Thus $\mathbf{b}$ is in the range of $\mathbf{A}$ if and only if the fourth entry of $\mathbf{Pb}$ is zero: $\mathbf{Ax} = \mathbf{b}$ can be solved if and only if

$$\tfrac{1}{10}b^1 - \tfrac{1}{2}b^2 - \tfrac{1}{10}b^3 + b^4 = 0$$

If $\mathbf{b}$ satisfies that condition, there is an $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$; we find it by solving $\mathbf{PAx} = \mathbf{Pb}$:

$$\begin{aligned} &x^1 - x^3 = -b^3 \\ &x^2 + \tfrac{5}{2}x^3 = \tfrac{1}{2}b^2 + \tfrac{3}{2}b^3 \\ &x^3 = \tfrac{1}{5}b^1 + \tfrac{4}{5}b^3 \end{aligned}$$

14. Consider now the system in three unknowns given by

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

We row reduce as above.

(a) Multiply row 1 by 4 and subtract it from row 2; multiply row 1 by 3 and subtract it from row 3.

$$\begin{pmatrix} 1 & 3 & -2 \\ 0 & -10 & 6 \\ 0 & -5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$$

(b) Subtract 1/2 of row 2 from row 3; divide row 2 by -10

$$\begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & -3/5 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2/5 & -1/10 & 0 \\ -1 & 1/2 & 1 \end{pmatrix}$$

The system $\mathbf{Ax} = \mathbf{b}$ thus has a solution if and only if

$$-b^1 + \tfrac{1}{2}b^2 + b^3 = 0 \tag{1.20}$$

In that case the solution is given by

$$\begin{aligned} x^1 + 3x^2 - 2x^3 &= b^1 \\ x^2 - \tfrac{3}{5}x^3 &= \tfrac{2}{5}b^1 - \tfrac{1}{10}b^2 \end{aligned}$$

Any arbitrarily chosen value of x^3 will provide a solution (granted the condition (1.20) is satisfied).

15.
$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 & 2 \\ 3 & -1 & 1 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix}$$

(a) Divide row 1 by 2.

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 3 & -1 & 1 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(b) Subtract 3 times row 1 from row 2; subtract twice row 1 from row 3.

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & -3 \\ 0 & 2 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

(c) Multiply row 2 by -1, and subtract twice the result from row 3.

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 2 & -8 \end{pmatrix} \begin{pmatrix} 1/2 & 0 & 0 \\ 3 & -1 & 0 \\ -8 & 2 & 1 \end{pmatrix}$$

Here there is no condition for the equation $\mathbf{PAx} = \mathbf{y}$ to be solvable, thus every problem $\mathbf{Ax} = \mathbf{b}$ is also solvable. The solution is found by writing the system $\mathbf{PAx} = \mathbf{Pb}$:

$$\begin{aligned} x^1 + \qquad\qquad x^4 &= \tfrac{1}{2}b^1 \\ x^2 - x^3 + 3x^4 &= 3b^1 - b^2 \\ 2x^3 - 8x^4 &= -8b^1 + 2b^2 + b^3 \end{aligned}$$

Clearly, the value of x^4 can be freely chosen, and x^1, x^2, x^3 are easily found by the equations.

Validity of Row Reduction

The basic point behind the present discussion is that the study of m simultaneous linear equations in n unknowns is the same as the study of linear transformations of R^n into R^m, which is the same as the study of $m \times n$ matrices under multiplication by $m \times m$ matrices. The matrix version of this story is the easiest to work, if only because it imposes the minimum amount of notation. However, the linear transformation interpretation is the most significant, and in the next section we will follow that line of thought. But first, let us record a proof of the main result of Section 1.1 in terms of matrices.

Theorem 1.2. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. There is a finite collection* $\mathbf{E}_0, \ldots, \mathbf{E}_s$ *of elementary* $m \times m$ *matrices such that the product* $\mathbf{E}_s \cdots \mathbf{E}_0\mathbf{A}$ *is in row-reduced form. Let* $\mathbf{P} = \mathbf{E}_s \cdots \mathbf{E}_0$ *and let* d *be the index of* $\mathbf{PA}$.

(i) *The system* $\mathbf{Ax} = \mathbf{b}$ *has a solution if and only if* $\mathbf{PAx} = \mathbf{Pb}$ *has a solution.*

(ii) *The system* $\mathbf{Ax} = \mathbf{b}$ *has a solution if and only if the last* $m - d$ *entries of* $\mathbf{Pb}$ *vanish.*

(iii) $n - d$ *unknowns can be freely chosen in any solution of* $\mathbf{Ax} = \mathbf{b}$.

Proof. First of all, we may, by a sequence of row operations, replace $\mathbf{A}$ with a matrix whose first nonzero column is

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1.21}$$

namely, supposing the jth column is the first nonzero column. Thus some entry in that column, say $a_j{}^i$, is nonzero. Interchange the first and jth rows. This is accomplished by multiplication on the left by an elementary matrix of Type III, call it $\mathbf{E}_0$. Now, $\mathbf{E}_0\mathbf{A} = (\alpha_j{}^i)$ with $\alpha_j{}^1 \neq 0$. Multiply the first row by $(\alpha_j{}^1)^{-1}$; this makes the $(1, j)$ entry 1 and is accomplished by means of an elementary matrix, say $\mathbf{E}_1$. Now, let $\mathbf{E}_k$ be the elementary matrix representing the operation of adding $-\alpha_j{}^k(\alpha_j{}^1)^{-1}$ times the first row to the jth row (this makes the (i, j) entry zero). Then $\mathbf{E}_m \cdots \mathbf{E}_0\mathbf{A}$ has its first nonzero column (1.21).

The proof now proceeds by induction on m. If $m = 1$ the proof is concluded: the $1 \times n$ matrix $(0, \ldots, 0, 1, a_{j+1}^1, \ldots, a_n{}^1)$ is in row-reduced form. For $m > 1$, the matrix $\mathbf{E}_m \cdots \mathbf{E}_0\mathbf{A}$ has the form

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & a_{j+1}^1 & \cdots & a_n{}^1 \\ & \mathbf{0} & & & & \mathbf{B} & \end{pmatrix} \tag{1.22}$$

where $\mathbf{B}$ is an $(m-1) \times (n-j)$ matrix. The induction assumption thus applies to $\mathbf{B}$. There is a collection $\mathbf{F}_0, \ldots, \mathbf{F}_s$ of $(m-1) \times (m-1)$ elementary matrices such that $\mathbf{F}_s \cdots \mathbf{F}_0\mathbf{B}$ is in row-reduced form. Now let

$$\mathbf{E}_{m+j} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & \mathbf{F}_j & \end{pmatrix}$$

Then, it is easy to compute that further multiplication of (1.22) by these matrices does not affect the first row, and in fact,

$$E_{m+s} \quad \cdots \quad \mathbf{E}_0\mathbf{A} = \begin{pmatrix} 0 & 1 & a_{j+1}^1 & \cdots & a_n^1 \\ 0 & 0 & \mathbf{F}_s & \cdots & \mathbf{F}_0\mathbf{B} \end{pmatrix}$$

which is in row-reduced form.

(i) Suppose there are $\mathbf{x}$ and $\mathbf{b}$ such that $\mathbf{Ax} = \mathbf{b}$. Then multiplication by $\mathbf{P}$ preserves the equality, so $\mathbf{PAx} = \mathbf{Pb}$. On the other hand, supopse $\mathbf{x}$, $\mathbf{b}$ are given such that $\mathbf{PAx} = \mathbf{Pb}$. Let f_P be the transformation on R^m corresponding to $\mathbf{P}$. f_P is a composition of row operations which are invertible, thus f_P is invertible. In particular, f_P is one-to-one, so since $f_P(\mathbf{Ax}) = f_P(\mathbf{b})$ we must have $\mathbf{Ax} = \mathbf{b}$.

(ii) If d is the index of the $m \times n$ matrix $\mathbf{PA}$, its last $m-d$ rows vanish. Thus for $\mathbf{PAx} = \mathbf{Pb}$ to hold for some $\mathbf{x}$, the last $m-d$ rows of $\mathbf{Pb}$ must vanish. By (i) this is also the condition for $\mathbf{Ax} = \mathbf{b}$ to have a solution.

(iii) The solutions of $\mathbf{Ax} = \mathbf{b}$ are the same as those of $\mathbf{PAx} = \mathbf{Pb}$. This latter system has the form

$$\begin{aligned}
x^1 + a_2{}^1x^2 + \quad &\cdots \quad + a_n{}^1x^n = \sum p_j{}^1b^j \\
x^2 + a_3{}^2x^3 + \quad &\cdots \quad + a_n{}^2x^n = \sum p_j{}^2b^j \\
&\cdots \\
x^d + a_{d+1}^d x^{d+1} + \cdots + a_n{}^d x^n \quad &\qquad = \sum p_j{}^d b^j
\end{aligned}$$

Clearly, $x^1, \ldots, x^d$ are uniquely determined once $x^{d+1}, \ldots, x^n, b^1, \ldots, b^n$ are known. The b's are restricted by the last $m-d$ equations of $\mathbf{Pb} = \mathbf{0}$, but $x^{d+1}, \ldots, x^n$ are free to take any values.

● EXERCISES

13. Compute the products $\mathbf{AB}$:

(a) $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

(b) $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 2 & 5 & 1 \\ 1 & 0 & -2 & 0 \\ 6 & -3 & 3 & -2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & -2 & 3 & 0 \\ 2 & 3 & 0 & -1 \\ 0 & 1 & 2 & 3 \end{pmatrix}$

(c) $\mathbf{A} = (6, 6, 3, 2, 1,)$ $\quad \mathbf{B} = \begin{pmatrix} 4 \\ 2 \\ -1 \\ 8 \\ 7 \end{pmatrix}$

(d) $\mathbf{A} = \begin{pmatrix} 4 & 2 & 8 & 6 \\ 1 & 2 & -3 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & -1 & -1 \end{pmatrix}$ $\quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}$

14. Compute the products **BA** for the matrices **A**, **B** of Exercise 13.

15. Compute the matrix corresponding to the sequence of row operations which row reduce the matrices of Exercise 5.

16. For the given $m \times n$ matrix **A** find conditions on the vector **b** in R^m under which the equation $\mathbf{Ax} = \mathbf{b}$ has a solution.

(a) **A** as given in Exercise 13(a).

(b) **A** as given in Exercise 13(b).

(c)

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

(d)

$$\mathbf{A} = \begin{pmatrix} 2 & 6 & 0 & 1 \\ 4 & 2 & 0 & 0 \\ 2 & 1 & 0 & 1 \\ 8 & 6 & 0 & -1 \end{pmatrix}$$

17. Show that if **A** is an $m \times n$ matrix with $m > n$, then there are always **b** for which the equation $\mathbf{Ax} = \mathbf{b}$ has no solution.

18. Verify that the composition of two linear transformations is again linear.

19. Suppose that T: $R^m \to R^m$ and has this property:

$$T(\mathbf{E}_1) = 0, \ldots, T(\mathbf{E}_m) = 0.$$

Show that $T(\mathbf{x}) = 0$ for every $\mathbf{x} \in R^m$.

20. Show that there is only one linear function on R^n with this property:

$$f(\mathbf{E}_1) = \mathbf{E}_2, f(\mathbf{E}_2) = \mathbf{E}_3, \ldots, f(\mathbf{E}_n) = \mathbf{E}_1$$

• PROBLEMS

15. Let f: $R^n \to R$ be defined by $f(x^1, \ldots, x^n) = \sum_{i=1}^n x^i$. Show that f is a linear function. Is the function

$$g(x^1, \ldots, x^n) = \sum_{i=1}^n (x^i)^2$$

linear? Is the function $h(x^1, x^2) = x^1x^2$ linear?

16. Suppose that S, T are linear transformations of R^n to R^m. Show that $S+T$, defined by

$$(S+T)(x) = S(x) + T(x)$$

is also linear. Show that the matrix representing $S+T$ is the entry by entry sum of the matrices representing S, T, respectively.

17. Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ be $n \times n$ matrices. Show that

$$\begin{aligned} (\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \\ (\mathbf{A}+\mathbf{B})\mathbf{C} &= \mathbf{AC}+\mathbf{BC} \\ \mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC} \end{aligned}$$

Show that $\mathbf{AB} = \mathbf{BA}$ need not be true.

18. Write down the products of the elementary matrices which row reduce these matrices:

$$\begin{pmatrix} 0 & 4 & 7 & 3 & 1 \\ 1 & 0 & -5 & 6 & 2 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 5 & 1 & 2 \\ 0 & -2 & -1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 6 & 3 & 2 & 1 \\ -2 & -5 & 4 & 0 \\ 0 & 3 & 2 & 1 \\ 3 & 3 & 6 & 3 \end{pmatrix}$$

19. Is it possible to apply further operations to the matrices of Exercise 18 in order to bring them to the identity? Notice that when this is possible for a given matrix $\mathbf{A}$, the product $\mathbf{P}$ of the elementary matrices corresponding to these operations has the property $\mathbf{PA} = \mathbf{I}$. That is, $\mathbf{P}$ is an inverse to $\mathbf{A}$. Using this suggestion compute inverses to these matrices also:

$$\begin{pmatrix} 3 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 8 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & -1 & -1 \end{pmatrix}$$

20. Find a 2×2 matrix $\mathbf{A}$, different from the identity such that $\mathbf{A}^2 = \mathbf{I}$. Find a 2×2 matrix such that $\mathbf{A}^2 = -\mathbf{I}$.

21. Is the equation $(\mathbf{I}+\mathbf{A})(\mathbf{I}+\mathbf{B}) = \mathbf{I}+\mathbf{A}+\mathbf{B}$ possible (with nonzero $\mathbf{A}$ and $\mathbf{B}$)?

22. An $n \times n$ matrix $\mathbf{A} = (a_j{}^i)$ is said to be diagonal if $a_j{}^i = 0$ for $i \neq j$. Show that diagonal matrices commute; that is, if $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices, $\mathbf{AB} = \mathbf{BA}$. Give necessary and sufficient conditions for a diagonal matrix to have an inverse.

1.4 Linear Subspaces of R^n

In the last section we saw that the equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

can be solved just for $\mathbf{b}$'s restricted by certain linear equations and that the set of solutions of that equation might have some degrees of freedom. In both cases these sets are determined by some linear equations; such sets are called linear subspaces of R^n. We will begin with an intrinsic definition of linear subspace and the notion of its dimension. In the next section we shall find a simple relation between the dimensions of the sets related to the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Definition 4.

(i) A set V in R^n is a **linear subspace** if it is closed under the operations of addition and scalar multiplication. That is, these conditions must be satisfied:

(1) $v_1, v_2 \in V$ implies $v_1 + v_2 \in V$.
(2) $r \in R$, $\mathbf{v} \in V$ implies $r\mathbf{v} \in V$.

(ii) If S is a set of vectors in R^n, the **linear span of** S, denoted $[S]$ is the set of all vectors of the form

$$c^1\mathbf{v}_1 + \cdots + c^k\mathbf{v}_k$$

with $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$.

(iii) The **dimension** of a linear subspace V of R^n is the minimum number of vectors whose linear span is V.

Linear Span

Having now given the intuitively loaded word "dimension" a definition, we had better hope that it suits our preconception of that notion. It does just that in R^3: a line is one dimensional since it is the linear span of but one vector; and a plane is two dimensional because we need that many vectors to span it. In fact, it is precisely those observations which have motivated the above definition. We should also ask that the above definition makes this assertion true: R^n has dimension n. You may need a little convincing that this is not immediately obvious, since you do know of n vectors (the standard basis) whose linear span is R^n. But how can we be sure that *we cannot find* less than n vectors with the same properties? Consider this restatement of the notion of "spanning": If the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span R^n, then the system of n simultaneous linear equations

$$\sum_{j=1}^{k} x^i \mathbf{v}_i = \mathbf{b}$$

has a solution for every $\mathbf{b} \in R^n$. We already know from the preceding section that this cannot be if $k < n$, and that gives us a proof that R^n has dimension n. We now repeat the arguments in the present context.

Theorem 1.3. *If the set S of vectors in R^n spans R^n, then S has at least n members. Thus, the dimension of R^n is n.*

Proof. The proof is by induction on n and goes like this. Supposing that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span R^n, one of them must have a nonzero first entry. Subtracting an appropriate multiple of that from each of the others, we may suppose that the remaining $k - 1$ vectors all have first entry equal to zero. Then they are the same as vectors in R^{n-1}, and since the original $\mathbf{v}_1, \ldots, \mathbf{v}_k$ spanned R^n we can show that these must span R^{n-1}. Now, by induction $k - 1 \geq n - 1$, and we have it. (Notice that this is the same as the first step in the proof of Theorem 1.2.) Here now is a more precise argument.

If none of the $\mathbf{v}_1, \ldots, \mathbf{v}_k$ has a nonzero first entry then $\mathbf{E}_1 = (1, 0, \ldots, 0)$ could hardly be in their linear span. Letting a_j be the first entry of $\mathbf{v}_j$, we may suppose (by reordering) that $a_1 \neq 0$. Now let $\mathbf{w}_1 = \mathbf{v}_1$ and $\mathbf{w}_j = \mathbf{v}_j - a_j a_1^{-1} \mathbf{v}_1$ for $j = 2, \ldots, k$. The vectors $\mathbf{w}_1, \ldots, \mathbf{w}_k$ have the same linear span as the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ (see Problem 18); the difference is that only $\mathbf{w}_1$ has a nonzero first entry. Let $\mathbf{w}_1 = (a_1, \mathbf{b}_1)$, $\mathbf{w}_2 = (0, \mathbf{b}_2), \ldots, \mathbf{w}_k = (0, \mathbf{b}_k)$, where $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are in R^{n-1}. Now, $\mathbf{b}_2, \ldots, \mathbf{b}_k$ span R^{n-1}. For let $\mathbf{c} \in R^{n-1}$. Then $(0, \mathbf{c}) \in R^n$, and since $\mathbf{w}_1, \ldots, \mathbf{w}_k$ span R^n, there are $x^1, \ldots, x^k \in R$ such that

$$\sum_{i=1}^{k} x^i \mathbf{w}_i = (0, \mathbf{c})$$

Thus $x^1 a_1 + x^2 \cdot 0 + \cdots + x^n \cdot 0 = 0$, $x^1 \mathbf{b}_1 + \cdots + x^k \mathbf{b}_k = \mathbf{c}$. Since $a_1 \neq 0$, the first equation implies $x^1 = 0$, so the second equation becomes $x^2 \mathbf{b}_2 + \cdots + x^k \mathbf{b}_k = \mathbf{c}$. Thus, $\mathbf{b}_2, \ldots, \mathbf{b}_k$ span R^{n-1}, so by induction $k - 1 \geq n - 1$; that is, $k \geq n$. Thus, $\dim R^n \geq n$. On the other hand, the standard basis $\mathbf{E}_1, \ldots, \mathbf{E}_n$ clearly spans, so in fact $\dim R^n = n$.

Examples

16. Let

$$\mathbf{v}_1 = (0, 1, 0, 3)$$
$$\mathbf{v}_2 = (2, 2, 2, 2)$$
$$\mathbf{v}_3 = (3, 3, 3, 3)$$

be three vectors in R^4, and let S be their linear span. Then clearly $\dim S \leq 3$. But it is also clear that $\mathbf{v}_3$ is superfluous, since $\mathbf{v}_3 = 3/2(\mathbf{v}_2)$. Thus S is also the linear span of $\mathbf{v}_1$, $\mathbf{v}_2$: if

$$\mathbf{v} = a^1\mathbf{v}_1 + a^2\mathbf{v}_2 + a^3\mathbf{v}_3$$

then we can also write

$$\mathbf{v} = a^1\mathbf{v}_1 + (a^2 + 3/2(a^3))\mathbf{v}_2$$

Thus, $\dim S \leq 2$. In fact, S has precisely dim 2. For suppose there were a vector $\mathbf{w} = (a^1, a^2, a^3, a^4)$ which spanned S. Then we would have numbers c_1, c_2 such that $\mathbf{v}_1 = c_1\mathbf{w}$, $\mathbf{v}_2 = c_2\mathbf{w}$. Explicitly this becomes

$$\begin{aligned} 0 &= c_1 a^1 & \quad 2 &= c_2 a^1 \\ 1 &= c_1 a^2 & \quad 2 &= c_2 a^2 \\ 0 &= c_1 a^3 & \quad 2 &= c_2 a^3 \\ 3 &= c_1 a^4 & \quad 2 &= c_2 a^4 \end{aligned}$$

But this is clearly impossible. By the second equation we must have $c_1 \neq 0$, so by the first we must have $a^1 = 0$. But $2 = c_2 a^1$, which could not be. Thus, $\dim S = 2$.

17. Let V be the subset of R^4 given by

$$V = \{\mathbf{v} : v^1 + v^2 + v^3 - v^4 = 0\}$$

V is certainly a linear subspace of R^4. We will shortly have the theoretical tools to deduce that V has dimension 3; with a little work we can show it now. First of all, let $\mathbf{A}_1 = (1, 0, 0, 1)$, $\mathbf{A}_2 = (0, 1, 0, 1)$, $\mathbf{A}_3 = (0, 0, 1, 1)$. Then $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ are all in V, and if $v = (v^1, v^2, v^3, v^4)$, since $v^4 = v^1 + v^2 + v^3$ we have

$$\mathbf{v} = v^1\mathbf{A}_1 + v^2\mathbf{A}_2 + v^3\mathbf{A}_3$$

Thus V is the linear span of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, so dim $V \leq 3$. On the other hand, if dim $V < 3$, then $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ can all be expanded in terms of some pair of vectors $\mathbf{B}_1$, $\mathbf{B}_2$. If we delete the fourth entry in all these vectors this amounts to saying that the standard basis vectors in R^3 can be spanned by a pair of vectors. But dim $R^3 = 3$, so this is impossible. Thus dim $V = 3$ also.

Independence

Repeating the definition once again, dimension is the minimum number of vectors it takes to span a linear space. There is another closely allied intuitive concept: that of "degrees of freedom" or "independent directions." In such phrases as "there is a four parameter family of curves," "two independently varying quantities are involved," allusion is being made to a dimension-like notion. Now, if we try to pin down this notion mathematically and specify the concept of independence in the linear space context, it turns out to be precisely the requirement for a spanning set of vectors to be minimal. In other words, the dimension of a linear space is also the **maximum number of degrees of freedom**, or **indpendent vectors** in the space.

Definition 5. Let S be a set of vectors in R^n. We say that S is a **set of independent vectors** if the equation

$$x^1\mathbf{v}_1 + \cdots + x^k\mathbf{v}_k = \mathbf{0}$$

with $x^1, \ldots, x^k \in R$ and $\mathbf{v}_1, \ldots, \mathbf{v}_k$ distinct elements of S implies $x^1 = 0, \ldots, x^k = 0$.

The standard basis of R^n is an independent set, as is very easy to verify. We now verify that R^n has in fact no more than n degrees of freedom in this sense.

Proposition 10. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be an independent set in R^n. Then $k \leq n$.*

Proof. The proof is by induction on k. The case $k = 1$ is automatically true, since $n \geq 1$ always. Now let us proceed to the induction step ($k > 1$). Let a_j be the first entry of $\mathbf{v}_j$; we can thus write $\mathbf{v}_j = (a_j, \mathbf{b}_j)$, where $\mathbf{b}_j \in R^{n-1}$. If all the a_j are zero, then $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are an independent set in R^{n-1}. By the induction assumption then, $k \leq n - 1$, so $k \leq n$. Now suppose instead that some a_j is nonzero. We can reorder the given vectors so that $a_1 \neq 0$. Let $\mathbf{w}_i = \mathbf{v}_1 - a_i a_1^{-1}\mathbf{v}_1$ for $i \geq 2$. Then the first entry of $\mathbf{w}_i$ is 0, so $\mathbf{w}_i = (0, \boldsymbol{\beta}_i)$ with $\boldsymbol{\beta}_i \in R^{n-1}$. $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k$ are an independent set in R^{n-1}. For if $\sum_{i=2}^{k} c^i \boldsymbol{\beta}_i = 0$, then also $\sum_{i=2}^{k} c^i \mathbf{w}_i = 0$, so

$$\left(-\sum_{i=2}^{k} c^i a_i\right) a_1^{-1} \mathbf{v}_1 + \sum_{i=2}^{k} c^i \mathbf{v}_i = \mathbf{0}$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are an independent set, $c^2 = \cdots = c^k = 0$, so $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k$ are also independent. Thus, by induction, once again $k - 1 \leq n - 1$. Thus in every case $k \leq n$, and the proposition is proved.

Examples

18. Let

$$\begin{aligned} \mathbf{v}_1 &= (0, 3, 0, 2) \\ \mathbf{v}_2 &= (5, 1, 1, 2) \\ \mathbf{v}_3 &= (1, 0, 2, 2) \end{aligned}$$

In order to show that these vectors are independent we must show that the system of equations

$$x^1 \mathbf{v}_1 + x^2 \mathbf{v}_2 + x^3 \mathbf{v}_3 = 0 \tag{1.23}$$

has only the zero solution. But this system is the same as the system corresponding to the matrix whose columns are $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$:

$$\mathbf{A} = \begin{pmatrix} 0 & 5 & 1 \\ 3 & 1 & 0 \\ 0 & 1 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

If we row reduce this matrix we obtain

$$\mathbf{PA} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Now the system $\mathbf{PAx} = \mathbf{0}$ obviously has only the zero solution: if

$$\begin{aligned} x^1 + x^2 + x^3 &= 0 \\ x^2 + 2x^3 &= 0 \\ x^3 &= 0 \\ 0 &= 0 \end{aligned} \tag{1.24}$$

we find, reading upward that $x^3 = 0$, $x^2 = 0$, $x^1 = 0$. Since $\mathbf{P}$ is invertible then the system $\mathbf{Ax} = \mathbf{0}$ has only the zero solution. What is the same, if (1.23) holds, so must (1.24), so $x^3 = x^2 = x^1 = 0$. Thus the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are independent.

19. Now let

$$\mathbf{v}_1 = (3, 2, 1, 0)$$
$$\mathbf{v}_2 = (1, 2, 3, 1)$$
$$\mathbf{v}_3 = (2, 0, -2, -1)$$

Again, let $\mathbf{A}$ be the matrix whose columns are $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$:

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

$\mathbf{A}$ row reduces to

$$\mathbf{PA} = \begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The system $\mathbf{PAx} = \mathbf{0}$ has the solutions

$$x^1 = -3x^2 + 2x^3$$
$$x^2 = x^3$$

The system $\mathbf{Ax} = \mathbf{0}$ has the same solutions. Taking $x^3 = 1$ we have the particular solution $(-1, 1, 1)$. Thus

$$-\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = 0$$

20. Four vectors in R^3 cannot be independent. Let

$$\mathbf{v}_1 = (2, 1, 2)$$
$$\mathbf{v}_2 = (0, 3, 0)$$
$$\mathbf{v}_3 = (1, 0, 4)$$
$$\mathbf{v}_4 = (0, 1, 2)$$

Find a linear relation which these vectors must satisfy. If we row reduce the matrix whose columns are the **v**'s, we obtain the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 0 & 1 \\ 0 & 1 & -1/6 & 1/3 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Now, corresponding to any value of x^4 we obtain a solution of $\mathbf{Ax} = \mathbf{0}$, and thus of $\sum x^i\mathbf{v}_i = 0$. Take $x^4 = 1$. Then

$$\begin{aligned} x^3 &= x^4 = 1 \\ x^2 &= \tfrac{1}{6}x^3 - \tfrac{1}{3}x^4 = \tfrac{1}{6} \\ x^1 &= -3x^2 - x^4 = -\tfrac{1}{2} \end{aligned}$$

Thus

$$-\tfrac{1}{2}\mathbf{v}_1 - \tfrac{1}{6}\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 = 0$$

Now, the equivalent form of these two propositions about R^n, that any spanning set of vectors has at least n members, and any independent set has at most n members, holds for any linear subspace of R^n as well.

Proposition 11. *Let V be a linear subspace of R^n of dimensions d.*

(i) *A spanning set has no less than d elements.*
(ii) *An independent set has no more than d elements.*

Proof. Part (i) is of course just the definition, so we need only consider part (ii). The proof amounts to a reduction to the case where V is R^d, and an application of Proposition 10.

Let $\mathbf{w}_1, \ldots, \mathbf{w}_d$ span V; since V has dimension d there exists such vectors. Suppose, as in (ii), that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are independent vectors in V. Then we can write each $\mathbf{v}_j$ as a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_d$;

$$\mathbf{v}_j = \sum_{j=1} a_j{}^i\mathbf{w}_i \qquad 1 \le j \le k$$

for suitable numbers $a_j{}^i$. The vectors $(a_j{}^1, \ldots, a_j{}^d)$ for $j = 1, \ldots, k$ are vectors in R^d corresponding to the vectors $\{\mathbf{v}_j\}$; we shall now show that they are likewise independent. For if,

$$\sum_{j=1}^{k} c^j(a_j{}^1, \ldots, a_j{}^d) = 0$$

then also $\sum_{j=1}^{k} c^j \mathbf{v}_j = 0$ by this computation:

$$\sum_{j=1}^{k} c^j \mathbf{v}_j = \sum_{j=1}^{k} c^j \sum_{i=1}^{d} a_j{}^i \mathbf{w}_i = \sum_{i=1}^{d} \left(\sum_{j=1}^{k} c^j a_j{}^i \right) \mathbf{w}_i = 0 \cdot \mathbf{w}_1 + \cdots + 0 \cdot \mathbf{w}_d$$

Thus, by the independence of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, we must have $c^1 = 0, \ldots, c^k = 0$. Thus the k vectors $(a_j{}^1, \ldots, a_j{}^d)$ are independent in R^d, so by Proposition 10 $d \geq k$.

Definition 6. Let V be a linear space. A **basis** of V is a set S of vectors such that each $\mathbf{v} \in V$ can be written in the form

$$\mathbf{v} = \sum_{i=1}^{k} c^i \mathbf{v}_i \qquad \textit{with } c^i \in R, \ \mathbf{v}_i \in S$$

in one and only one way.

Another way of putting this is: a basis for a linear subspace V is a set of independent and spanning vectors in V.

Proposition 12. *S is a basis for the linear space V if and only if both these conditions hold:*

(i) *S is an independent set,*
(ii) *the linear span of S is V.*

Proof. Suppose that S is a basis of V. Since every vector in V can be written as a linear combination of vectors in S, certainly (ii) is true: V is the linear span of S. Since $\mathbf{0}$ can be written in *only one way* as a linear combination of vectors in V, any time we have

$$c^1\mathbf{v}_1 + \cdots + c^k\mathbf{v}_k = \mathbf{0}$$

with $c^1, \ldots, c^k$ in R and $\mathbf{v}_1, \ldots, \mathbf{v}_k$ distinct members of S, we must have $c^1 = 0, \ldots, c^k = 0$ (since $\mathbf{0} = 0 \cdot \mathbf{v}_1 + \cdots + 0 \cdot \mathbf{v}_k$ also). Thus (i) holds: S is an independent set.

Conversely, suppose now that (i) and (ii) are true for the set S. Then (by (ii)) any vector $\mathbf{v}$ in V can be written

$$\mathbf{v} = c^1\mathbf{v}_1 + \cdots + c^k\mathbf{v}_k \tag{1.25}$$

with $c^i \in R$, $\mathbf{v}_i \in S$. This can be done in only one way because of the independence. In fact, suppose (1.25) holds, and also

$$\mathbf{v} = a^1\mathbf{v}_1 + \cdots + a^k\mathbf{v}_k \tag{1.26}$$

is true, with $(c^1 - a^1)\mathbf{v}_1 + \cdots + (c^k - a^k)\mathbf{v}_k = 0$, so $c^i = a^i$ since the $\mathbf{v}_j$ are independent.

Dimension and Basis

The important facts to know about dimension of linear subspaces of R^n are these: such a space V always has a basis with a finite number of elements. That number is the same for all bases and is the dimension of V, and is not greater than n. We summarize this as follows:

Theorem 1.4. *Let V be a linear subspace of R^n.*

(i) *There is an integer $d \leq n$ such that V has dimension d.*
(ii) *Any basis of V has precisely d elements.*
(iii) *d independent vectors in V form a basis.*
(iv) *d spanning vectors in V form a basis.*

Proof. (i) The proof of this part of the theorem is by mathematical induction on n. If $n = 1$, either $V = \{0\}$ or V has a nonzero vector, in which case $V = R$. Thus either dim $V = 0$ or 1, so dim $V \leq 1$. Now we proceed to the induction step. Let us describe how it goes. We assume the assertion (i) for $n - 1$, and consider R^{n-1} as the set of n-tuples in R^n with zero first entry. If V is a subspace of R^n, it intersects this space in some subspace of R^{n-1} which is, by induction spanned by some δ vectors, with $\delta \leq n - 1$. Now, choosing any other vector in V with a nonzero first entry, this together with the vectors referred to above will span V. Now we make this argument precise.

Let V be a subspace of R^n. If $V = \{0\}$, then dim $V = 0$; if not, V has a nonzero vector $\mathbf{v}_0 = (a^1, \ldots, a^n)$. One of the entries is nonzero; we may, by reordering the coordinates assume that $a^1 \neq 0$. Let now

$$W = \{\mathbf{w} \in R^{n-1} : (0, \mathbf{w}) \in V\}$$

W is a linear subspace of R^{n-1}. For if $\mathbf{w}_1, \mathbf{w}_2 \in W$ and $c^1, c^2 \in R$, we also have

$$c^1(0, \mathbf{w}_1) + c^2(0, \mathbf{w}_2) = (0, c^1\mathbf{w}_1 + c^2\mathbf{w}_2)$$

in V, so $c^1\mathbf{w}_1 + c^2\mathbf{w}_2 \in W$. Now, by the induction hypothesis, W has dimension $\delta \leq n - 1$. Let $\mathbf{w}_1, \ldots, \mathbf{w}_\delta$ span W. By definition of W, $\mathbf{v}_1 = (0, \mathbf{w}_1), \ldots, \mathbf{v}_\delta = (0, \mathbf{w}_\delta)$ are in V. Now we need only show that $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_\delta$ span V. Let $\mathbf{v} \in V$, and let c be its first entry. Then $\mathbf{v} - c(a^1)^{-1}\mathbf{v}_0$ is also in V and its first entry is 0. Thus this vector is of the form $(0, \mathbf{w})$ with $\mathbf{w} \in W$. Then there are $c^1, \ldots, c^\delta$ such that

$$\mathbf{w} = c^1\mathbf{w}_1 + \cdots + c^\delta\mathbf{w}_\delta$$

Thus

$$\mathbf{v} - c(a^1)^{-1}\mathbf{v}_0 = (0, \mathbf{w}) = c^1(0, \mathbf{w}_1) + \cdots + c^\delta(0, \mathbf{w}_\delta)$$

or,

$$\mathbf{v} = c(a^1)^{-1}\mathbf{v}_0 + c^1\mathbf{v}_1 + \cdots + c^\delta\mathbf{v}_\delta$$

Thus, there are $\delta + 1$ vectors which span V, so V has dimension d with $d \leq \delta + 1 \leq (n-1) + 1 = n$.

(ii) This follows easily from Proposition 12. If S is a basis for V (dim $V = d$), then since S spans, it has at least d elements, and since S is independent it has at most d elements.

(iii) Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_d$ are independent vectors in V; we must show that they span. Let $\mathbf{v}_0 \in V$. By Proposition 12(ii), since dim $V = d$, $\mathbf{v}_0, \ldots, \mathbf{v}_d$ are dependent, so there exist $(c^0, \ldots, c^d) \neq 0$ such that

$$c^0\mathbf{v}_0 + \cdots + c^a\mathbf{v}_d = 0$$

If $c^0 = 0$, since $\mathbf{v}_1, \ldots, \mathbf{v}_d$ are independent we must also have $c^1 = \cdots = c^d = 0$, a contradiction. Thus $c^0 \neq 0$, so $\mathbf{v}_0 = (-c^0)^{-1}(c^1\mathbf{v}_1 + \cdots + c^d\mathbf{v}_d)$ as desired.

(iv) Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_d$ span V. If they are dependent, then the equation

$$c^1\mathbf{v}_1 + \cdots + c^d\mathbf{v}_d = 0$$

holds with at least one $c^i \neq 0$. If say $c^r \neq 0$, then

$$\mathbf{v}_r = (-c^r)^{-1}(c^1\mathbf{v}_1 + \cdots + c^{r-1}v_{r-1} + c^{r+1}v_{r+1} + \cdots + c^d\mathbf{v}_d)$$

so $\mathbf{v}_1, \ldots, \mathbf{v}_d$, with $\mathbf{v}_r$ excluded, also span V. Hence, V has dimension at most $d-1$, a contradiction, so we must have had $\mathbf{v}_1, \ldots, \mathbf{v}_d$ independent and thus a basis.

This final proposition, whose proof is left as an exercise, is an indication of the (theoretical) ease in finding bases.

Proposition 13. *Let V be a linear subspace of R^n of dimension d.*

(i) *Any set of vectors whose span is V contains d vectors which form a basis.*
(ii) *Any set of independent vectors in V is part of a basis for V.*

Examples

21. Find a basis for the linear span V of the vectors

$$\begin{aligned}\mathbf{v}_1 &= (4, 3, 2, 1)\\ \mathbf{v}_2 &= (5, 2, 2, 1)\\ \mathbf{v}_3 &= (0, 1, 0, 1)\\ \mathbf{v}_4 &= (1, 0, 0, 1)\end{aligned}$$

and express V by a linear equation.

We want to find all vectors **b** of the form

$$\sum x^i \mathbf{v}_i = \mathbf{b} \tag{1.27}$$

and we want to find a basis for such vectors. Now (1.27) is the system corresponding to the matrix **A** whose columns are the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$. The span of these vectors is just the range of **A**. If **P** is a product of elementary matrices row reducing **A**, then any vector **b** is in the range of **A** if and only if **Pb** is in the range of **PA**. Thus by row reduction we should easily be able to solve our problem.

$$\mathbf{A} = \begin{pmatrix} 4 & 5 & 0 & 1 \\ 3 & 2 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

The end result of row reduction produces

$$\mathbf{PA} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & -4 \\ 0 & 0 & -1/2 & 1 \\ 1 & 1 & -3/2 & -4 \end{pmatrix}$$

Thus, the range of **A** is obtained by setting the fourth entry of **Pb** to zero:

$$V = \text{range of } \mathbf{A} = \{(b^1, b^2, b^3, b^4) \colon b^1 + b^2 - \tfrac{3}{2}b^3 - 4b^4 = 0\}$$

V has dimension at least three since it contains the independent vectors $(4, 0, 0, 1)$, $(0, 4, 0, 1)$, $(0, 0, 2/3, -1/4)$. On the other hand, $V \neq R^4$, so $\dim V \leq 3$. Thus, $\dim V = 3$ and these three vectors are a basis.

22. Find a basis for the linear subspace V of R^5 given by the equations

$$\begin{aligned} 5x^1 + 8x^2 + 3x^3 + x^4 + x^5 &= 0 \\ x^1 - x^3 - x^5 &= 0 \\ x^2 + 2x^4 &= 0 \end{aligned}$$

We are seeking the solution space of $\mathbf{Ax} = \mathbf{0}$, where

$$\mathbf{A} = \begin{pmatrix} 5 & 8 & 3 & 1 & 1 \\ 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 2 & 0 \end{pmatrix}$$

Row reduction leads to

$$\mathbf{PA} = \begin{pmatrix} 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & -2 & -15 & 6 \end{pmatrix}$$

and V is the set of $\mathbf{x}$ such that $\mathbf{PAx} = \mathbf{0}$. According to these equations x^4 and x^5 are to be freely chosen and x^1, x^2, x^3 determined by this choice. Thus, $\dim V = 2$. Choosing $(x^4, x^5) = (1, 0)$, $(0, 1)$, respectively, we obtain as a basis

$$(-\tfrac{15}{2}, -2, -\tfrac{15}{2}, 1, 0) \qquad (4, 0, 3, 0, 1)$$

• EXERCISES

21. What is the dimension of the linear span of these vectors?
    (a) $\mathbf{v}_1 = (-1, 2, -1, 0)$
    $\mathbf{v}_2 = (2, 5, 7, 2)$
    $\mathbf{v}_3 = (0, 2, 1, 1)$
    $\mathbf{v}_4 = (3, 5, 7, 1)$
    (b) $\mathbf{v}_1 = (-1, 0, 2, 1)$
    $\mathbf{v}_2 = (2, 2, -2, 2)$
    $\mathbf{v}_3 = (1, 1, 1, 1)$
    (c) $\mathbf{v}_1 = (0, 2, 1, 1)$
    $\mathbf{v}_2 = (1, 7, 3, 3)$
    $\mathbf{v}_3 = (0, 0, 0, 1)$
    $\mathbf{v}_4 = (1, 3, 1, 2)$
    $\mathbf{v}_5 = (1, 5, 2, 2)$
    (d) $\mathbf{v}_1 = (0, 0, 1, 1, 1)$
    $\mathbf{v}_2 = (1, 0, 0, 1, 1)$
    $\mathbf{v}_3 = (0, 1, 0, 1, 0)$
22. What is the dimension of the space S given by these equations:
    (a) $S = \{\mathbf{x} \in R^5 : x^1 + x^2 - x^3 - x^4 = 0, x^1 + x^3 = 0\}$
    (b) $S = \{\mathbf{x} \in R^5 : x^2 + x^4 + x^5 = 0, x^1 - x^3 + x^4 = 0$
    $x^1 - x^2 - x^3 - x^5 = 0\}$
    (c) $S = \{\mathbf{x} \in R^4 : x^1 + x^2 + x^3 = x^3 - x^2 - x^1 + x^4\}$
23. Determine the linear span of these vectors by a system of equations
    (a) $\mathbf{v}_1 = (1, 0, 0, 1)$
    $\mathbf{v}_2 = (0, 1, 1, 0)$
    $\mathbf{v}_3 = (0, 1, 0, 1)$
    (b) $\mathbf{v}_1 = (2, 2, 6, 2)$
    $\mathbf{v}_2 = (1, 2, 3, 0)$
    $\mathbf{v}_3 = (0, 1, 0, -1)$
    (c) $\mathbf{v}_1 = (1, 0, 1)$
    $\mathbf{v}_2 = (-1, 1, 1)$
    (d) $\mathbf{v}_1 = (1, 0, 0, 0, 0)$
    $\mathbf{v}_2 = (2, 0, 1, 0, 1)$

24. Are these vectors independent?

(a) $\mathbf{v}_1, \ldots, \mathbf{v}_5$ as given in Exercise 21(c).

(b) $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ as given in Exercise 23(a).

(c) $\mathbf{v}_1 = (0, 2, 0, 2, 0, 6)$
$\mathbf{v}_2 = (1, 1, -1, -1, 1, 1)$
$\mathbf{v}_3 = (2, 4, 6, 8, 10, 12)$
$\mathbf{v}_4 = (0, 0, -2, -2, 0, 0)$
$\mathbf{v}_5 = (0, 1, 0, 0, 1, 0)$
$\mathbf{v}_6 = (1, 1, 1, 1, 1, 1)$

25. Find all linear relations involving these sets of vectors.

(a) $\mathbf{v}_1 = (0, 1, 1)$
$\mathbf{v}_2 = (5, 3, 1)$
$\mathbf{v}_3 = (0, 2, 0)$
$\mathbf{v}_4 = (1, -1, 1)$

(b) $\mathbf{v}_1 = (0, 2, 0, 2)$
$\mathbf{v}_2 = (0, 1, 0, 0)$
$\mathbf{v}_3 = (0, 1, 0, 1)$
$\mathbf{v}_4 = (0, 0, 0, 1)$
$\mathbf{v}_5 = (1, 0, -1, 0)$

(c) $\mathbf{v}_1 = (0, 0, 0, 0)$
$\mathbf{v}_2 = (1, 1, 1, 1)$
$\mathbf{v}_3 = (1, 1, 0, 0)$
$\mathbf{v}_4 = (0, 0, -2, -2)$

26. Find a basis for the linear subspace of R^5 spanned by $(0, 0, 0, 1, 1)$, $(0, 1, 0, 0, 0)$, $(1, 0, 0, 0, 1)$, $(1, 1, 0, 0, 1)$, $(2, 1, 0, 1, 2)$

27. Find a basis for these linear spaces:

(a) $\{(x^1, \ldots, x^5) \in R^5 : x^1 + 2x^2 + x^3 = 0,\ x^1 + 2x^4 + x^5 = 0,$
$x^1 + x^5 = 0\}$

(b) $\{(x^1, \ldots, x^4) \in R^4 : x^1 - x^2 + x^3 - x^4 = 0,\ x^1 - x^3 = 0\}$

28. If the given vectors on R^5 are independent, extend them to a basis:

(a) $(0, 0, 0, 0, 1)$, $(0, 0, 0, 1, 1)$, $(0, 0, 1, 1, 1)$

(b) $(1, 5, 2, 0, -3)$, $(6, 7, 0, 2, 1)$, $(1, 0, -1, -2, 0)$, $(1, 1, 1, 1, 1)$

(c) $(4, 4, 3, 2, 1)$, $(3, 3, 3, 2, 1)$, $(2, 2, 2, 2, 1)$

• PROBLEMS

23. Suppose we are given k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in R^n. Let $\mathbf{w}_1 = \mathbf{v}_1$, $\mathbf{w}_2 = \mathbf{v}_2 - \beta_2 \mathbf{v}_1, \ldots, \mathbf{w}_k = \mathbf{v}_k - \beta_k \mathbf{v}_1$ for some numbers $\beta_2, \ldots, \beta_k$. Show that the sets $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ have the same linear span.

24. The proof of Theorem 1.3 proceeds by assuming that the set S consists of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. What of the case where S has infinitely many elements?

25. Prove Proposition 13.

26. Show that if V, W are subspaces of R^n, so is $V \cap W$.

27. Show that if **A** is obtained from **B** by a row operation, the linear span of the rows of **A** is the linear span of the rows of **B**.

28. Show that if **A** is a row-reduced matrix the dimension of the linear span of the rows of **A** is the same as its index.

1.5 Rank + Nullity = Dimension

Now let us apply the propositions of the preceding section about linear spaces, and in particular the notion of dimension, to the subject of linear transformations. There are certain obvious linear spaces to be associated to a given transformation.

Definition 7. Let T: $R^n \to R^m$ be a linear transformation.

(i) The set

$$K(T) = \{\mathbf{v} \in R^n : T(\mathbf{v}) = 0\}$$

is a linear subspace of R^n, called the **kernel of** T. Its dimension is the **nullity of** T, denoted $\nu(T)$.

(ii) The set

$$R(T) = \{T(\mathbf{v}) : v \in R^n\}$$

is a linear subspace of R^m, called the **range of** T. Its dimension is the **rank of** T, denoted $\rho(T)$.

Theorem 1.5. *Let* T: $R^n \to R^m$ *be a linear transformation. We have*

$$n = \nu(T) + \rho(T)$$

that is, dimension = nullity + rank.

Proof. For short, write $\nu(T)$ as ν. Let $\mathbf{v}_1, \ldots, \mathbf{v}_\nu$ be a basis for the kernel of T. Let $\mathbf{v}_{\nu+1}, \ldots, \mathbf{v}_n$ be the rest of a basis for R^n: $\mathbf{v}_1, \ldots, \mathbf{v}_\nu, \mathbf{v}_{\nu+1}, \ldots, \mathbf{v}_n$ thus span R^n. Let $\mathbf{w}_j = T(\mathbf{v}_j)$ for $j = \nu + 1, \ldots, n$. Now the crux of the matter is this: $\mathbf{w}_{\nu+1}, \ldots, \mathbf{w}_n$ form a basis for the range of T. Once this is shown, we will have $\rho(T) = n - \nu$, which is the desired equation.

(i) Let $\mathbf{w} \in R(T)$. Then there is a $\mathbf{v} \in R^n$ such that $\mathbf{w} = T(\mathbf{v})$. Expand $\mathbf{v}$ in the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$: $\mathbf{v} = c^1\mathbf{v}_1 + \cdots + c^n\mathbf{v}_n$. Then

$$\begin{aligned}\mathbf{w} = T(\mathbf{v}) &= T(c^1\mathbf{v}_1 + \cdots + c^n\mathbf{v}_n)\\ &= c^1T(\mathbf{v}_1) + \cdots + c^\nu T(\mathbf{v}_\nu) + c^{\nu+1}T(\mathbf{v}_{\nu+1}) + \cdots + c^nT(\mathbf{v}_n)\\ &= c^{\nu+1}\mathbf{w}_{\nu+1} + \cdots + c^n\mathbf{w}_n\end{aligned}$$

The second line is justified since T is linear and the third follows since $\mathbf{v}_1, \ldots, \mathbf{v}_\nu$ are in the kernel of T and $T(\mathbf{v}_{\nu+1}) = \mathbf{w}_{\nu+1}, \ldots, T(\mathbf{v}_n) = \mathbf{w}_n$. Thus these last vectors span $R(T)$.

(ii) $\mathbf{w}_{\nu+1}, \ldots, \mathbf{w}_n$ are independent. Suppose

$$c^{\nu+1}\mathbf{w}_{\nu+1} + \cdots + c^n\mathbf{w}_n = \mathbf{0} \tag{1.28}$$

We must show that the $\{c^j\}$ are all zero. In any event, from (1.28) we have

$$T(c^{\nu+1}\mathbf{v}_{\nu+1} + \cdots + c^n\mathbf{v}_n) = c^{\nu+1}T(\mathbf{v}_{\nu+1}) + \cdots + c^nT(\mathbf{v}_n) = 0$$

so $c^{\nu+1}\mathbf{v}_{\nu+1} + \cdots + c^n\mathbf{v}_n \in K(T)$. $\mathbf{v}_1, \ldots, \mathbf{v}_\nu$ span $K(T)$ so there are $c^1, \ldots, c^\nu$ such that

$$c^{\nu+1}\mathbf{v}_{\nu+1} + \cdots + c^n\mathbf{v}_n = c^1\mathbf{v}_1 + \cdots + c^\nu\mathbf{v}_\nu$$

or

$$(-c^1)\mathbf{v}_1 + \cdots + (-c^\nu)\mathbf{v}_\nu + c^{\nu+1}\mathbf{v}_{\nu+1} + \cdots + c^n\mathbf{v}_n = 0$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are independent, all the c^j are zero, as required. The theorem is proven.

Examples

23. Let T: $R^4 \to R^3$ be given by the matrix

$$\mathbf{A} = \begin{pmatrix}1 & 3 & 2 & 7\\ 0 & 0 & 1 & 1\\ 0 & 1 & 0 & 0\end{pmatrix} \tag{1.29}$$

We can completely analyze this transformation by row reduction. $\mathbf{A}$ easily row reduces to

$$\begin{pmatrix}1 & 3 & 2 & 7\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\end{pmatrix} \tag{1.30}$$

merely by interchanging the last two rows. Thus, letting P be the transformation corresponding to

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

we know that PT is the linear transformation corresponding to (1.30). Now, the range of PT is easily seen to be all of R^3, and the range of T is P^{-1} (range of PT), which is again all of R^3, so $\rho(T) = 3$. The kernel of T is the same as the kernel of PT, which has the equations given by (1.30):

$$\begin{aligned} x^1 + 3x^2 + 2x^3 + 7x^4 &= 0 \\ x^2 &= 0 \\ x^3 + x^4 &= 0 \end{aligned} \tag{1.31}$$

The set of all such solutions is found by letting x^4 take on all real values and solving for the remaining coordinates by (1.31). Thus $K(T) = \{(-5t, 0, -t, t) \colon t \in R\}$, which is one dimensional.

24. Let $T\colon R^4 \to R^4$ be given by the matrix

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

Let us row reduce this matrix, keeping track of our row operations:

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

$$\mathbf{PA} = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -3 & 4 & 1 & 0 \\ -2 & 5 & 0 & 1 \end{pmatrix}$$

Now, the kernel of T is easy to find; it is the same as the kernel of the transformation S corresponding to the last matrix **PA** (because

$S =$ the composition of T by an invertible transformation). Now the kernel of S, and thus also of T, has, corresponding to the matrix **PA**, the form:

$$\begin{aligned} x^1 + x^2 + x^3 + 2x^4 &= 0 \\ x^2 \quad + \ x^4 &= 0 \\ 0 &= 0 \\ 0 &= 0 \end{aligned}$$

or $x^2 = -x^4$, $x^1 = -x^3 - x^4$. Thus,

$$K(T) = \{(-(u+v),\ -v,\ u,\ v) \colon (u, v) \in R^2\}$$

so $\nu(T) = 2$. The range of T is a little harder to find. If R is the transformation corresponding to the product of the elementary matrices on the left, then $S = R \circ T$, so the range of T is R^{-1} of the range of S, which has the equations $x^3 = x^4 = 0$. (That is, the vector $(b^1, \ldots, b^4)$ is in the range of S if and only if there exist $(x^1, \ldots, x^4)$ such that

$$\begin{aligned} x^1 + x^2 + x^3 + 2x^4 &= b^1 \\ x^2 \quad + \ x^4 &= b^2 \\ 0 &= b^3 \\ 0 &= b^4 \end{aligned}$$

The necessary and sufficient condition is $b^3 = b^4 = 0$.) Thus the necessary and sufficient condition for $\mathbf{v}$ to be in the range of T is that $\mathbf{Pv}$ be in the range of S; that is, the third and fourth coordinates of $\mathbf{Pv}$ must vanish:

$$\begin{aligned} -3x^1 + 4x^2 + x^3 &= 0 \\ -2x^1 - 5x^2 + x^4 &= 0 \end{aligned}$$

Thus, $\rho(T) = 2$.

25. Let us do one more example briefly. Suppose that $T\colon R^3 \to R^5$ corresponds to the matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \\ 4 & 3 & 7 \end{pmatrix}$$

This matrix can be row reduced to

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

by multiplication on the left by this matrix

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

The kernel of T can be found by looking at the row-reduced form $\mathbf{A}$; it is the set of $x = (x^1, x^2, x^3)$ in R^3 such that $\mathbf{Ax} = \mathbf{0}$. Precisely, we must have $x^1 + x^3 = 0$, $x^2 + x^3 = 0$. Thus a vector is in $K(T)$ if its first and second coordinates are the negative of the third; that is, $K(T) = \{(-t, -t, t) \colon t \in R\}$. Thus $\nu(T) = 1$. The range of T is the set of $\mathbf{x} = (x^1, \ldots, x^5)$ such that $\mathbf{Px}$ is in the range of $\mathbf{A}$ (since $\mathbf{A} = \mathbf{PT}$). The 5-tuples in the range of $\mathbf{A}$ are precisely those with third, fourth, and fifth coordinates zero. Thus the third through fifth coordinates of $\mathbf{Px}$ must be zero for $\mathbf{x}$ to be in the range of T. Specifically $R(T)$ is the set of simultaneous solutions of

$$\begin{aligned} 2x^1 - x^2 + x^3 \qquad\qquad\quad &= 0 \\ x^1 - x^2 \qquad + x^4 \qquad &= 0 \\ -x^1 - x^2 - x^3 - x^4 + x^5 &= 0 \end{aligned}$$

We can take x^1, x^2 as free variables and use these equations to define x^3, x^4, x^5; thus

$$R(T) = \{(u, v - 2u, v - u, 3v - 2u) \colon (u, v) \in R^2\}$$

so $\rho(T) = 2$.

These examples illustrate the fact that Theorem 1.5 can be formulated purely in terms of matrices. We now do just that.

Proposition 14. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix, representing the linear transformation* $T\colon R^n \to R^m$. *Then*

$$\begin{aligned}\rho(T) &= \textit{number of independent columns of } \mathbf{A}\\ &= \textit{number of independent rows of } \mathbf{A}\\ &= \textit{index of the row-reduced matrix to which } \mathbf{A} \textit{ can be reduced.}\end{aligned}$$

Finally, we can also reformulate Theorem 1.5 as a conclusion for systems of linear equations, thus bringing us to the ultimate version of Theorems 1.1 and 1.2.

Theorem 1.6. *Suppose given a system of* m *linear equations in* n *unknowns, and suppose* d *is the index, or rank, of the corresponding matrix* $\mathbf{A}$. *Then*

(i) $d \le m$, $d \le n$.
(ii) $\{\mathbf{x}\colon \mathbf{A}\mathbf{x} = \mathbf{0}\}$ *is a vector space of dimension* $n - d$.
(iii) $\{b\colon$ *there exists a solution of* $\mathbf{A}\mathbf{x} = \mathbf{b}\}$ *is a vector space of dimension* d.

● EXERCISES

29. Describe by linear equations the range and kernel of the linear transformations given by these matrices in terms of the standard basis:

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

(b) $\begin{pmatrix} 0 & 6 \\ 2 & 3 \\ 4 & 1 \\ 1 & 7 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & 1 & 1 & 2 \\ 8 & 0 & 0 & 0 \\ 4 & -1 & 2 & 2 \\ -1 & 0 & 3 & 4 \end{pmatrix}$

(d) $\begin{pmatrix} 8 & 0 & 0 & 1 & 6 \\ 1 & 3 & 0 & 2 & 6 \\ 0 & 0 & 1 & 2 & 2 \end{pmatrix}$

30. Find bases for $K(T)$, $R(T)$ for each T given by the matrices (a)–(d) of Exercise 29.

31. Let $f\colon R^n \to R$ be a nonzero linear function. Show that the kernel of f is a linear subspace of R^n of dimension $n - 1$.

32. Let $f(x^1, \ldots, x^n) = \sum_{i=1}^n x^i$. Find a spanning set of vectors for the kernel of f.

PROBLEMS

29. Let $T: R^n \to R^m$ be a linear transformation. Then $K(T)$ and $R(T)$ are linear subspaces of R^n, R^m, respectively.

30. Let T be the transformation represented by the $m \times n$ matrix $\mathbf{A}$. Show that $R(T)$ is spanned by the columns of $\mathbf{A}$. Show that

$$K(T) = \{(x^1, \ldots, x^n): \sum_{j=1}^{n} \mathbf{C}_j x^j = 0\}$$

where $\mathbf{C}_1, \ldots, \mathbf{C}_n$ are the columns of $\mathbf{A}$.

31. Let $\mathbf{w} \in R^n$. Define $\perp(\mathbf{w})$ as the set of $\mathbf{v}$ such that $\sum_{i=1}^n v^i w^i = 0$. Show that for $\mathbf{w} \neq \mathbf{0}$, $\perp(\mathbf{w})$ is a subspace of R^n of dimension $n - 1$.

32. Let $S \subset R^n$. Define $\perp(S)$ as the set of $\mathbf{v}$ such that $\sum_{i=1}^n v^i w^i = 0$ for all $\mathbf{w} \in S$. Show that $\perp(S)$ is a subspace of R^n, and $\dim \perp(S) + \dim[S] = n$.

1.6 Invertible Matrices

In this section we shall pay particular attention to the collection of linear transformations of R^n into R^n—or, what is the same, the $n \times n$ matrices. From the point of view of linear equations this is reasonable; for it is usually the case that a given problem will have as many equations as unknowns.

First of all; it is clear that there are certain operations which are defined on the collection of all linear transformations of R^n, thus making of this set an algebraic object of some sort. We collect together all these notions in the following definition.

Definition 8. The algebra of linear operators on R^n denoted by E^n, is the collection of linear transformations provided with these operations:

(i) if f is in E^n, and c is a real number,

$$(cf)(x) = cf(x)$$

(ii) if f, g are in E^n, $f + g$ is defined by

$$(f + g)(x) = f(x) + g(x)$$

(iii) $f \circ g$ is defined by

$$(f \circ g)(x) = f(g(x))$$

It is important to think of the elements of E^n as functions taking n-tuples of numbers into n-tuples of numbers; but in working with them it is convenient to represent them in terms of the standard basis by matrices. Thus, we are led to consider also the algebra M^n of real $n \times n$ matrices with the operations of scalar multiplication, addition and multiplication, the definitions of which we now recapitulate.

Definition 9. The **algebra** M^n is the collection of $n \times n$ matrices provided with these operations:

(i) If $\mathbf{A} = (a_j^{\ i})$ is in M^n, and c is a real number,

$$c\mathbf{A} = (ca_j^{\ i})$$

(ii) If $\mathbf{A} = (a_j^{\ i})$ and $\mathbf{B} = (b_j^{\ i})$ are in M^n, then

$$\mathbf{A} + \mathbf{B} = (a_j^{\ i} + b_j^{\ i})$$

$$\mathbf{AB} = \left(\sum_{k=1}^{n} a_k^{\ i} b_j^{\ k}\right)$$

The two algebras E^n, M^n are completely interchangeable, for M^n is just the explicit representation of E^n relative to the standard basis.

Now the operations on M^n obey certain laws, some of which we have already observed in previous sections. Let us list some important ones.

Proposition 15. *These equations hold for all $n \times n$ matrices* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *and all real numbers k.*

(i) $k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$
(ii) $\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$
(iii) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
(iv) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

If $\mathbf{A}$ is a given matrix, we shall let $\mathbf{A}^2$ denote $\mathbf{A} \cdot \mathbf{A}$, $\mathbf{A}^3 = \mathbf{A} \cdot \mathbf{A} \cdot \mathbf{A}$, and in general $\mathbf{A}^n$ is the n-fold product of $\mathbf{A}$ with itself. Since we may also add matrices, and multiply by real numbers, we may consider polynomials in a given matrix. That is, $\mathbf{A}^2 + 3\mathbf{A} + \mathbf{A}$, $\mathbf{A}^7 + 3\pi\mathbf{A}^3 + \mathbf{A}^6, \ldots$. In fact, if we adopt the usual convention that $\mathbf{A}^0 = I$, then for any polynomial $p(X) = \sum_{i=0}^n c_i X^i$ in the indeterminate X, we may consider the matrix $p(\mathbf{A}) = \sum_{i=0}^n c_i \mathbf{A}^i$. A most remarkable observation can now be made, by noticing that the collection M^n of $n \times n$ matrices is the same as the collection R^{n^2} of n^2-tuples of real numbers.

Proposition 16. *Given any $n \times n$ matrix* $\mathbf{A}$, *there is a nonzero polynomial p of degree at most n^2 such that $p(\mathbf{A}) = \mathbf{0}$.*

Proof. An element of M^n is a rectangular array of n^2 real numbers, thus corresponds to an element of R^{n^2}. We may make this correspondence explicit, by, say, placing the rows one after another. That is, the matrix (a_j^i) corresponds to the vector $(a_1^1, \ldots, a_n^1, a_1^2, \ldots, a_n^2, a_1^3, \ldots, a_{n-1}^n, a_n^n)$ in R^{n^2}. In any event, the notions of sum and scalar multiplication is the same in the two interpretations. Now consider the matrices $\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \ldots, \mathbf{A}^{n^2}$. These $n^2 + 1$ vectors in R^{n^2} cannot be independent so there are real numbers $c_0, c_1, \ldots, c_{n^2}$, *not all zero*, such that

$$c_{n^2} \mathbf{A}^{n^2} + \cdots + c_2 \mathbf{A}^2 + c_1 \mathbf{A} + c_0 \mathbf{I} = \mathbf{0}$$

Thus the proposition is verified with p the polynomial

$$p(X) = c_{n^2} X^{n^2} + \cdots + c_2 X^2 + c_1 X + c_0$$

We may rephrase this proposition in this way: Every matrix is a root of some nonzero polynomial equation with real coefficients. From the purely algebraic point of view this formulation is of some interest and raises the converse speculation: given a polynomial with real coefficients, does it have some $n \times n$ matrix as a root? We shall verify this fact, and with n no greater than two. More precisely, we shall, in a later section, introduce the system of complex numbers as a certain collection of 2×2 matrices, and later verify that every real polynomial has a root in the system of complex numbers. This is known as the fundamental theorem of algebra.

Now, a linear transformation in E^n is *invertible* if it has an inverse as a function from R^n to R^n. For this it must be one-to-one and onto; that is, it must have zero nullity and rank n. We have seen (n = rank + nullity) that either of these assertions implies the other. Now it is clear that these assertions must be expressible in terms of matrices; we now do that.

Definition 10. The $n \times n$ matrix **A is invertible** if there is a matrix **B** such that $\mathbf{BA} = \mathbf{I} = \mathbf{AB}$. In this case **B** is said to be an inverse for **A**.

Proposition 17. *An invertible matrix has a unique inverse.*

Proof. This is clear: if **B**, **C** are inverses to **A**, then all these equations hold:

$$\mathbf{BA} = \mathbf{I} = \mathbf{AB} \qquad \mathbf{CA} = \mathbf{I} = \mathbf{AC}$$

Then

$$\mathbf{B} = \mathbf{BI} = \mathbf{B(AC)} = \mathbf{(BA)C} = \mathbf{IC} = \mathbf{C}$$

We shall denote the inverse of a matrix $\mathbf{A}$, if it exists, by $\mathbf{A}^{-1}$. The relationship between matrices and linear transformations gives us this propostion:

Proposition 18. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix. These assertions are equivalent*:

(i) $\mathbf{A}$ *is invertible.*
(ii) $\mathbf{A}$ *represents an invertible transformation.*
(iii) *There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.
(iv) *There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.
(v) $\mathbf{A}$ *has index* n.

Proof. We have already seen (in discussing systems of linear equations) that (ii) and (v) are equivalent. By definition (i) implies both (iii) and (iv). Thus we have left to prove that (i) and (ii) are equivalent, (iii) implies (i), and (iv) implies (i).

(i) implies (ii). Let $\mathbf{A}$ be the given invertible matrix, and T the transformation it represents. Let S be the transformation represented by the inverse, $\mathbf{A}^{-1}$, of $\mathbf{A}$. Since $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$, we have $T \cdot S = I = S \cdot T$. Thus S is inverse to T, so (ii) holds.

(ii) implies (i) by the same kind of reasoning with the roles of *matrix* and *transformation* interchanged.

(iii) implies (i). If T is the linear transformation represented by $\mathbf{A}$, then by (iii), there is a transformation S such that $S \circ T = I$. Thus, if $T(\mathbf{x}) = \mathbf{0}$ we must also have $\mathbf{x} = S(T(\mathbf{x})) = S(\mathbf{0}) = \mathbf{0}$, so T has nullity zero and is thus invertible. Thus (iii) implies (ii), so also implies (i).

(iv) implies (i). If again, T is the transformation represented by $\mathbf{A}$, by (iv) there is a transformation S such that $T \circ S = I$. This implies that T has rank n and thus is invertible.

Computing the Inverse

Now, it is clear that the question of invertibility for a given matrix is important and that the problems arise of effectively deciding this question and of effectively computing the inverse, if it exists. To ask that the rows (or columns) be independent, or span R^n, while responsive to this question hardly provides a procedure for determining invertibility. We shall now introduce two such procedures: one is a continuation of row reduction and the second is based on the notion of the determinant. The determinant is a real-valued function defined on the algebra M^n of $n \times n$ matrices; its basic property is that it is nonzero only on the invertible matrices. We shall depend heavily on the determinant in the study of eigenvectors (Section 1.7). In Section 1.9 we shall explore the connection between the determinant and the notion of volume in R^3.

In order to verify the critical properties of the determinant function it is necessary to return to the *elementary matrices*, for they provide a technique for decomposing an invertible matrix into a product of simple ones, and as a result, a technique for computing inverses. We recall these facts: the elementary matrices are the matrices which represent the row operations. Since the row operations are invertible, so are the elementary matrices invertible. For any matrix $\mathbf{A}$ there is a sequence $\mathbf{P}_s, \ldots, \mathbf{P}_0$ of elementary matrices such that $\mathbf{B} = \mathbf{P}_s\mathbf{P}_{s-1} \cdots \mathbf{P}_0\mathbf{A}$ is in row-reduced form. The index of $\mathbf{A}$ is the number of nonzero rows of $\mathbf{B}$. We augment these facts by this further observation:

Proposition 19. *Suppose that* $\mathbf{A}$ *is an invertible* $n \times n$ *matrix. There is a sequence* $\mathbf{P}_t, \ldots, \mathbf{P}_0$ *of elementary* $n \times n$ *matrices such that* $\mathbf{P}_t \cdots \mathbf{P}_0\mathbf{A}$ *is the identity matrix:*

$$\mathbf{P}_t \cdots \mathbf{P}_0\mathbf{A} = \mathbf{I}$$

Proof. The proof will be by induction on n. It is a slight modification of Theorem 1.2. The first column of $\mathbf{A}$ is nonzero since the columns of $\mathbf{A}$ must be independent ($\mathbf{A}$ is invertible). As we have seen in the proof of Theorem 1.2, there exist elementary matrices $\mathbf{P}_0, \ldots, \mathbf{P}_k$ such that the first column of $\mathbf{P}_k \cdots \mathbf{P}_0\mathbf{A}$ is $\mathbf{E}_1$. Thus,

$$\mathbf{P}_k \cdots \mathbf{P}_0\mathbf{A} = \begin{array}{cc} \begin{array}{cc} 1 & n-1 \end{array} & \\ \begin{pmatrix} 1 & \mathbf{A}^1 \\ 0 & \mathbf{A}_2{}^2 \end{pmatrix} & \begin{array}{c} 1 \\ n-1 \end{array} \end{array}$$

Since $\mathbf{P}_k \cdots \mathbf{P}_0\mathbf{A}$ is invertible so is $\mathbf{A}_2{}^2$ (see Problem 37). Thus the proposition applies to $\mathbf{A}_2{}^2$. There is a sequence $\mathbf{Q}_s, \ldots, \mathbf{Q}_{k+1}$ of elementary $(n-1) \times (n-1)$ matrices such that $\mathbf{Q}_s \cdots \mathbf{Q}_{k+1}\mathbf{A}_2{}^2 = \mathbf{I}$. Let

$$\mathbf{P}_j = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{Q}_j \end{pmatrix} \qquad \text{for } j = k+1, \ldots, s$$

$$\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{A} = \begin{pmatrix} 1 & \mathbf{A}^1 & & \\ 0 & \mathbf{Q}_s & \cdots & \mathbf{Q}_{k+1}\mathbf{A}_2{}^2 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{A}^1 \\ 0 & \mathbf{I} \end{pmatrix}$$

Now the matrix

$$\begin{pmatrix} 1 & -\mathbf{A}^1 \\ 0 & \mathbf{I} \end{pmatrix}$$

is the product $\mathbf{P}_{s+n-1} \cdots \mathbf{P}_{s+1}$ of elementary matrices corresponding to these row operations: subtract a_j^1 times the jth row from the first row, $j = 2, \ldots, n$. Finally,

$$\mathbf{P}_{s+n-1} \cdots \mathbf{P}_0\mathbf{A} = \begin{pmatrix} 1 & -\mathbf{A}^1 \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{A}^1 \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \mathbf{I}$$

as required.

This proposition provides us with an effective way for computing inverses; we just continue the process of row reduction until we obtain the identity. Then the corresponding product of elementary matrices is the inverse.

Examples

26.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 \\ -3 & 4 & 2 \\ 1 & 0 & 8 \end{pmatrix}$$

Row reduction	Product of elementary matrices
$\begin{pmatrix} 1 & 2 & 2 \\ 0 & 10 & 8 \\ 0 & -2 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & \frac{4}{10} \\ 0 & 1 & \frac{8}{10} \\ 0 & 0 & \frac{76}{10} \end{pmatrix}$	$\begin{pmatrix} \frac{7}{10} & -\frac{1}{10} & 0 \\ \frac{3}{10} & \frac{1}{10} & 0 \\ -\frac{4}{10} & \frac{2}{10} & 1 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} \frac{137}{190} & -\frac{21}{190} & -\frac{1}{19} \\ \frac{65}{190} & \frac{15}{190} & -\frac{2}{19} \\ -\frac{4}{76} & \frac{2}{76} & \frac{10}{76} \end{pmatrix}$

Thus

$$\mathbf{A}^{-1} = \tfrac{1}{190}\begin{pmatrix} 137 & -21 & -10 \\ 65 & 15 & -20 \\ -10 & 5 & 25 \end{pmatrix}$$

27.

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ -1 & -1 & 0 & 1 \\ -1 & 4 & 1 & -24 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 0 & -1 & 1 & 3 \\ 0 & 4 & 2 & -22 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -10 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ -4 & 9 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -10 \end{pmatrix} \begin{pmatrix} -1 & 2 & -1 & 0 \\ 1 & -2 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ -6 & 11 & -2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\frac{22}{10} & \frac{42}{10} & -\frac{12}{10} & \frac{2}{10} \\ \frac{28}{10} & -\frac{53}{10} & \frac{6}{10} & -\frac{3}{10} \\ 1 & -1 & 1 & 0 \\ \frac{6}{10} & -\frac{11}{10} & \frac{2}{10} & -\frac{1}{10} \end{pmatrix}$$

Thus,

$$\mathbf{A}^{-1} = \frac{1}{10} \begin{pmatrix} -22 & 42 & -12 & 2 \\ 28 & -53 & 6 & -3 \\ 10 & -10 & 10 & 0 \\ 6 & -11 & 2 & -1 \end{pmatrix}$$

The Determinant Function

The determinant of a matrix is a pretty complicated concept; before going into a study of it and its properties, we shall first see how to compute it. Looking ahead, the method of computation comes from Equations (1.35) and (1.36), but we shall not use those equations to derive it. Instead we shall simply describe the technique for finding determinants.

The determinant of a 2×2 matrix is defined by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

The determinant of a 3×3 matrix is found as follows. First, select a row. The determinant will be a sum of the products of the elements of that row with j numbers, called their **cofactors**. The cofactor of the (i, j)th entry is $(-1)^{i+j}$ times the determinant of the 2×2 matrix remaining when the ith row and jth column are deleted.

Examples

28. Compute the determinant of

$$A = \begin{pmatrix} 1 & 3 & 2 \\ -1 & 4 & 0 \\ 7 & -2 & 1 \end{pmatrix}$$

If we select the first row we find

$$\begin{aligned} \det A &= 1[4(1) - (-2)0] - 3[(-1)(1) - 7(0)] + 2[(-1)(-2) - 7(4)] \\ &= -45 \end{aligned}$$

Selecting the second row:

$$\begin{aligned} \det A &= -(-1)[3(1) - (-2)2] + 4[1(1) - 2(7)] + 0[\ldots] \\ &= -45 \end{aligned}$$

Selecting the third row:

$$\begin{aligned} \det A &= 7[3(0) - 4(2)] - (-2)[1(0) - 2(-1)] + [1(4) - 3(-1)] \\ &= -45 \end{aligned}$$

Now, we could also have selected a column first, and proceeded in the same way. For example, selecting the second column:

$$\begin{aligned} \det A &= -3[(-1)1 - 0(7)] + 4[1(1) - 7(0)] - 2[1(0) - 2(-1)] \\ &= -45 \end{aligned}$$

Now, in general, the determinant of the $n \times n$ matrix is found in the same way. Select a row (or column). The determinant is the sum of the products of the entries in that row (or column) with their cofactors. The cofactor of the (i, j)th entry is $(-1)^{i+j}$ times the determinant of the $(n-1) \times (n-1)$ matrix remaining when the ith row and jth column are deleted.

29. Let

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

Select the first row

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

$$+ \det\begin{pmatrix} 2 & 6 & -1 \\ 1 & 0 & 4 \\ 2 & 1 & -1 \end{pmatrix} - 0\det\begin{pmatrix} 2 & 6 & 0 \\ 1 & 0 & 0 \\ 2 & 1 & 1 \end{pmatrix}$$

We now compute the determinants of the 3×3 matrices by taking advantage of the location of the 0's. Select the second column in the first three, and don't bother with the last since its factor is 0:

$$\begin{aligned} \det \mathbf{A} &= 4(-1)[6(4) - 0(-1)] - 3(-1)[2(4) - 1(-1)] \\ &\quad +(-6)[1(-1) - 4(2)] - [2(4) - 1(-1)] \\ &= -24. \end{aligned}$$

30.
$$\mathbf{A} = \begin{pmatrix} 6 & 2 & 1 & 0 & -1 \\ 4 & 3 & 8 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 8 & 1 & 4 & 0 & 1 \\ 2 & 1 & 4 & -1 & 1 \end{pmatrix}$$

Select the third row:

$$\det \mathbf{A} = (-1)^{3+3}(2)\det\begin{pmatrix} 6 & 2 & 0 & -1 \\ 4 & 3 & 1 & 0 \\ 8 & 1 & 0 & 1 \\ 2 & 1 & -1 & 1 \end{pmatrix}$$

Select the third column:

$$\det \mathbf{A} = 2(-1)^{2+3}\det\begin{pmatrix} 6 & 2 & -1 \\ 8 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} + (-1)^{4+3}(-1)\det\begin{pmatrix} 6 & 2 & -1 \\ 4 & 3 & 0 \\ 8 & 1 & 1 \end{pmatrix}$$

$$= 96$$

We turn now to the theory of determinants. We begin with a definition of the determinant function which is appropriate to the theoretical discussion and then verify that it has the multiplicative property:

$$\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$$

The formulas (1.35) and (1.36) below which form the basis for the preceding computations will result from a rewriting of the formula for the determinant.

The determinant of an $n \times n$ matrix can be described in this way: it is the sum of all products of precisely one element from each row and column, *with appropriate signs.* Our first business is to determine this appropriate sign. A selection of precisely one element from each row and column is described as follows: In the first row we select a certain element, say in the $\pi(1)$ column. In the second row we select an element, *coming from a different column,* say $\pi(2)$. We have $\pi(2) \neq \pi(1)$, and so forth. We select the element $a^i_{\pi(i)}$ in the ith row and $\pi(i)$th column, making sure that the numbers $\pi(1), \ldots, \pi(n)$ are all distinct. These numbers then form a rearrangement, or *permutation* of the numbers $1, \ldots, n$. To form the determinant then, we consider all products

$$a^1_{\pi(1)} \cdots a^n_{\pi(n)}$$

as π ranges over all permutations of the numbers $1, \ldots, n$. A particular kind of permutation is an **interchange of two successive integers**:

$$\begin{array}{rcl}
1 & \to & 1 \\
2 & \to & 2 \\
\vdots & & \vdots \\
i & \to & i+1 \\
i+1 & \to & i \\
\vdots & & \vdots \\
n & \to & n
\end{array}$$

(We consider the integers as arranged in a circle, so that 1 *is the successor to* n.) Now it is a fact about permutations, that any permutation consists of a succession of such interchanges. There may be many ways to build up a given permutation by these simple interchanges, but the *parity of the number involved is always the same.* That is, if we can write a given permutation as a succession of an even number of interchanges, then every way of writing that permutation as a succession of interchanges will involve an even number. For example, consider the permutation on four integers

$$1\ 2\ 3\ 4 \to 3\ 1\ 4\ 2$$

This is obtained by this succession of interchanges:

$$\begin{array}{cccc}
1 & 2 & 3 & 4 \\
2 & 1 & 3 & 4 \\
2 & 3 & 1 & 4 \\
3 & 2 & 1 & 4 \\
3 & 1 & 2 & 4 \\
3 & 1 & 4 & 2
\end{array}$$

Here is a better way of doing it:

1 2 3 4
1 3 2 4
3 1 2 4
3 1 4 2

Either way, there is an odd number of interchanges involved. We shall not verify these facts about permutations; the verification would be tangential to our present study. However, we shall use these facts. We shall say that a given permutation is **even** if it can be formed by an even numbered succession of interchanges; the permutation is **odd** if an odd numbered succession of interchanges is required. For any permutation π, its **sign**, denoted $\varepsilon(\pi)$ will be $+1$ if π is even, and -1 if π is odd. There is another way of defining the sign function on permutations which is described in Problem 36. This description does not involve the notion of interchange.

Definition 11. If $\mathbf{A} = (a_j{}^i)$ is an $n \times n$ matrix its **determinant** is

$$det\ \mathbf{A} = \sum_{\text{all permutations } \pi} \varepsilon(\pi) \prod_{i=1}^{n} a^i_{\pi(i)} \tag{1.32}$$

We shall now show that $\det \mathbf{A} \neq 0$ if and only if $\mathbf{A}$ is invertible, by showing in fact a stronger statement: $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$.

Lemma 1.
(i) $\det \mathbf{I} = 1$.
(ii) *If* $\mathbf{A}$ *has a zero row*, $\det \mathbf{A} = 0$.

Proof.
(i) Writing $\mathbf{I} = (a_j{}^i)$, we have $a^i_{\pi(i)} = 0$, unless $\pi(i) = i$. Thus, the sum (1.32) has only one nonzero term, that corresponding to the identity permutation. Since each $a_i{}^i = 1$, $\det \mathbf{I} = 1 \cdot 1 \cdots 1 = 1$.
(ii) If the jth row of $\mathbf{A}$ is zero, each term of the sum (1.32) has a factor $a^j_{\pi(j)} = 0$, so is zero. Then $\det \mathbf{A} = 0$.

Lemma 2. *If* $\mathbf{P}$ *is an elementary matrix, and* $\mathbf{A}$ *any matrix,*

$$\det(\mathbf{PA}) = \det \mathbf{P} \cdot \det \mathbf{A} \tag{1.33}$$

Proof. Let $\mathbf{A} = (a_j{}^i)$, $\mathbf{PA} = (b_j{}^i)$.

Type I. If $\mathbf{P}$ multiplies the rth row by c, then

$$\det \mathbf{PA} = \sum \varepsilon(\pi) \prod_{i=1}^{n} b^i_{\pi(i)} = \sum \varepsilon(\pi) a^1_{\pi(1)} \cdots c a^r_{\pi(r)} \cdots a^n_{\pi(n)}$$

$$= \sum \varepsilon(\pi) c \prod_{i=1}^{n} a^i_{\pi(i)} = c \det \mathbf{A}$$

In the special case $\mathbf{A} = \mathbf{I}$, we have $\det \mathbf{P} = \det(\mathbf{PI}) = c \det \mathbf{I} = c$. Thus (1.33) holds in this case.

Type II. Suppose now $\mathbf{P}$ interchanges the rth and sth rows. Let η represent the permutation which interchanges r and s. Thus, $b_j{}^i = a_j^{\eta(i)}$. Now we compute:

$$\det \mathbf{PA} = \sum \varepsilon(\pi) \prod_{i=1}^{n} b^i_{\pi(i)} = \sum \varepsilon(\pi) \prod a^{\eta(i)}_{\pi(i)}$$

Now we change the index of summation. Let $\pi = \tau \cdot \eta$, and sum over τ.

$$\det \mathbf{PA} = \sum \varepsilon(\tau \cdot \eta) \prod a^{\eta(i)}_{\tau(\eta(i))} = -\sum \varepsilon(\tau) \prod_{i=1}^{n} a^{\eta(i)}_{\tau(\eta(i))}$$

The sign changes since η is an interchange; thus, if τ is even, $\tau \cdot \eta$ is odd. Now the product $\prod_{i=1}^{n} a^{\eta(i)}_{\tau(\eta(i))}$ is the same as the product $\prod_{i=1}^{n} a^i_{\tau(i)}$ (another change of index) so

$$\det \mathbf{PA} = \sum \varepsilon(\tau) \prod_{j=1} a^i_{\tau(i)} = -\det \mathbf{A}$$

In particular, $\det \mathbf{P} = \det(\mathbf{PI}) = -1$, so (1.33) holds in this case.

Type III. Suppose that $\mathbf{P}$ adds α times row r to row s. Then $b_j{}^i = a_j{}^i$ if $i \neq s$ and $b_j{}^s = a_j{}^i + \alpha a_j{}^r$. We now compute

$$\begin{aligned} \det(\mathbf{PA}) &= \sum \varepsilon(\pi) \prod_{i=1}^{n} b^i_{\pi(i)} \\ &= \sum \varepsilon(\pi) \prod_{i=1}^{n} a^i_{\pi(i)} + \alpha \sum \varepsilon(\pi) \prod_{\substack{i=1 \\ i \neq s \\ i \neq r}}^{n} a^i_{\pi(i)} a^r_{\pi(r)} a^r_{\pi(s)} \end{aligned} \tag{1.34}$$

The first term on the right is $\det \mathbf{A}$. The second term is zero. We can see that by splitting up the sum into odd and even permutations. Let η represent the interchange of r and s. It is important to note that the odd permutations are just those of the form $\pi \cdot \eta$, where π is even. Thus the last term in Equation (1.34) is

$$\begin{aligned} &\sum_{\pi \text{ even}} \prod_{i \neq r,s} a^i_{\pi(i)} \cdot a^r_{\pi(r)} \cdot a^r_{\pi(s)} - \sum_{\pi \text{ odd}} \prod_{i \neq r,s} a^i_{\pi(i)} \cdot a^r_{\pi(r)} \cdot a^r_{\pi(s)} \\ = {}& \sum_{\pi \text{ even}} \prod_{i \neq r,s} a^i_{\pi(i)} \cdot a^r_{\pi(r)} \cdot a^r_{\pi(s)} - \sum_{\pi \text{ even}} \prod_{i \neq r,s} a^i_{\pi(\eta(i))} a^r_{\pi(\eta(r))} a^r_{\pi(\eta(s))} \\ = {}& \sum_{\pi \text{ even}} \prod_{i \neq r,s} a^i_{\pi(i)} (a^r_{\pi(r)} a^r_{\pi(s)} - a^r_{\pi(s)} a^r_{\pi(r)}) = 0 \end{aligned}$$

Thus, $\det \mathbf{PA} = \det \mathbf{A}$. In particular, $\det \mathbf{P} = 1$, so (1.33) is verified also for Type III elementary matrices.

Now, *lemma* is a word denoting a logical particle of no particular intrinsic interest, but of crucial importance in the verification of a theorem. Here now is the main theorem concerning determinants.

Theorem 1.7. *A matrix* $\mathbf{M}$ *is invertible if and only if* $\det \mathbf{M} \neq 0$. $\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}$ *for any two matrices.*

Proof. Suppose $\mathbf{M}$ is an $n \times n$ matrix which is not invertible. Then there are elementary matrices $\mathbf{P}_s, \ldots, \mathbf{P}_0$ such that $\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{M}$ is row reduced and has zero rows. Thus, by the above lemma

$$0 = \det(\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{M}) = \det \mathbf{P}_s \cdot \det \mathbf{P}_{s-1} \cdots \det \mathbf{P}_0 \cdot \det \mathbf{M}$$

Since the determinant of an elementary matrix is nonzero, we must have $\det \mathbf{M} = 0$. On the other hand, if $\mathbf{M}$ is invertible, there are elementary matrices $\mathbf{P}_s, \ldots, \mathbf{P}_0$ such that $\mathbf{I} = \mathbf{P}_s \cdots \mathbf{P}_0\mathbf{M}$. Then

$$1 = \det \mathbf{I} = \det \mathbf{P}_s \cdot \det \mathbf{P}_{s-1} \cdots \det \mathbf{P}_0 \cdot \det \mathbf{M}$$

Thus $\det \mathbf{M} \neq 0$.

Now let $\mathbf{A}$, $\mathbf{B}$ be two $n \times n$ matrices. If one of $\mathbf{A}$ or $\mathbf{B}$ is not invertible, neither is $\mathbf{AB}$, so $\det \mathbf{AB} = 0$ and either $\det \mathbf{A} = 0$ or $\det \mathbf{B} = 0$. In any case

$$\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}$$

is true. If A and B are invertible, there are elementary matrices $\mathbf{P}_s \cdots \mathbf{P}_0$, $\mathbf{Q}_\sigma \cdots \mathbf{Q}_0$ such that

$$\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{A} = \mathbf{I} = \mathbf{Q}_\sigma \cdots \mathbf{Q}_0\mathbf{B}$$

Then

$$\mathbf{Q}_\sigma \cdots \mathbf{Q}_0\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{AB} = \mathbf{Q}_\sigma \cdots \mathbf{Q}_0(\mathbf{P}_s \cdots \mathbf{P}_0\mathbf{A})\mathbf{B} = \mathbf{Q}_\sigma \cdots \mathbf{Q}_0\mathbf{B} = \mathbf{I}$$

Thus

$$\begin{aligned}
&\det \mathbf{Q}_\sigma \cdots \det \mathbf{Q}_0 \cdot \det \mathbf{P}_s \cdots \det \mathbf{P}_0 \cdot \det(\mathbf{AB}) = 1 \\
&\det \mathbf{Q}_\sigma \cdots \det \mathbf{Q}_0 \cdot \det \mathbf{B} = 1 \\
&\det \mathbf{P}_s \cdots \det \mathbf{P}_0 \cdot \det \mathbf{A} = 1
\end{aligned}$$

Thus again $\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$.

Notice that the formula $\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}$ is far from transparent on the basis of the definition above. In fact, it is not at all derivable without some information regarding the structure of $n \times n$ matrices. We have a means of computing $\mathbf{A}^{-1}$ for a given invertible matrix $\mathbf{A}$; namely, the process of row reduction. But we have not given explicitly any formula for the inverse. Such is provided by the **cofactor expansion of a determinant.** This formula is of theoretical interest, but not of any great computational value. As far as computations are concerned, the surest and quickest route to the inverse is the process of row reduction.

Let $\mathbf{A}$ be an $n \times n$ matrix. The **adjoint matrix of an entry of A** is the $(n-1) \times (n-1)$ matrix obtained by deleting the row and column of the given entry (see Figure 1.13). Let $\mathbf{A}_i^j$ be the adjoint matrix of the entry a_j^i. Then the inverse to the matrix $\mathbf{A}$ (if it is invertible) is easily given by the determinants of the adjoints: the (i, j)th entry of $\mathbf{A}^{-1}$ is

$$(-1)^{i+j} \frac{det\ A^i}{det\ A}$$

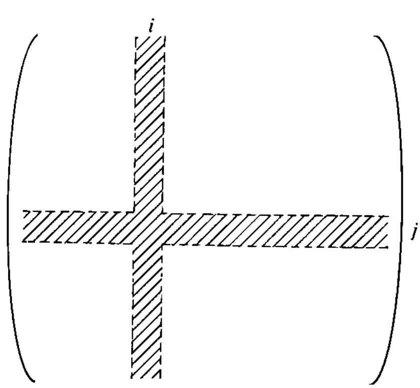
More precisely we have these formulas (the explicit version of $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = I$) known as Cramer's rule:

$$\det \mathbf{A} = \sum_{j=1}^{n} (-1)^{i+j} a_j^i \det \mathbf{A}_i^j \qquad \text{for all } i \tag{1.35}$$

$$\det \mathbf{A} = \sum_{i=1}^{n} (-1)^{i+j} a_j^i \det \mathbf{A}_i^j \qquad \text{for all } j \tag{1.36}$$

$$0 = \sum_{j=1}^{n} (-1)^{i+j} a_j^i \det \mathbf{A}_k^j \qquad \text{for all } i \neq k \tag{1.37}$$

$$0 = \sum_{i=1}^{n} (-1)^{i+j} a_j^i \det \mathbf{A}_i^k \qquad \text{for all } j \neq k \tag{1.38}$$



$\mathbf{A}_j^i$

Figure 1.13

The verifications of these formulas are simpler than it may seem; they can be based directly on formula (1.32). For example, let us verify (1.35). First fix a row index i. We shall break up the sum in (1.32) into n parts: those permutations taking $i \to 1$, $i \to 2, \ldots, i \to n$. Consider, for a fixed column index the permutations taking $i \to j$. (That is, those π for which $\pi(i) = j$.) These are precisely the same as all permutations on *the indices of the matrix adjoint to* $a_j^{\,i}$ (those permutations which take the integers $1, \ldots, n$, except for i, into the integers $1, \ldots, n$, except for j). Thus the terms appearing in the sum (1.32) which have $a_j^{\,i}$ as a factor, are the same as those in (1.35): we must now verify that the signs agree. Let τ be a permutation on the indices of the adjoint to $a_j^{\,i}$. The corresponding permutation π of $(1, \ldots, n)$ does the same as τ and *takes i into j.* The number of interchanges involved in building this permutation is just that for τ, with the interchanges required to send i to j. The last number is $j - i$, which has the same parity as $i + j$. Thus, $\varepsilon(\pi) = (-1)^{i+j}\varepsilon(\tau)$, so the signs of corresponding terms in (1.32) and (1.35) also agree. Thus (1.35) is true. We shall leave the verifications of the other formulas to the exercises. (Equations (1.37) and (1.38) require a small trick.)

Cramer's rule allows for a simple description for solving the equation $\mathbf{Ax} = \mathbf{b}$ when $\mathbf{A}$ is an invertible $n \times n$ matrix. Let $\mathbf{A}^{(i)}$ be the matrix obtained by replacing the ith column of $\mathbf{A}$ with the column b. Then the equation $\mathbf{Ax} = \mathbf{b}$, which is the same as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, turns out, according to Cramer's rule to read

$$x^i = \frac{\det \mathbf{A}^{(i)}}{\det \mathbf{A}} \qquad 1 \le i \le n$$

This is checked out by unraveling all the definitions and applying the formulas of Cramer's rule: since $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$,

$$x^i = \sum_{j=1}^{n} (\mathbf{A}^{-1})_j^{\,i} b^j = \frac{1}{\det \mathbf{A}} \sum_{j=1}^{n} (-1)^{i+j} \det \mathbf{A}_i^{\,j} b^j$$

But the summation is just the determinant of the matrix obtained by replacing the ith column of $\mathbf{A}$ with the column vector $\mathbf{b}$! Thus we can solve by taking quotients of determinants.

Example

31. Solve the equations

$$\begin{aligned} x^1 + 2x^2 - x^3 &= 2 \\ x^1 + x^2 + 3x^3 &= 0 \\ 2x^1 + 2x^2 + x^3 &= 1 \end{aligned}$$

The determinant of the matrix

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

is easily found by cofactor expansion along the first row:

$$\det \mathbf{A} = 1(1-6) - 2(1-6) - 1(2-2) = 5$$

By Cramer's rule

$$x^1 = \tfrac{1}{5}\det\begin{pmatrix} 2 & 2 & -1 \\ 0 & 1 & 3 \\ 1 & 2 & 1 \end{pmatrix} = \tfrac{1}{5}[2(-5) + 1(6+1)] = -\tfrac{3}{5}$$

$$x^2 = \tfrac{1}{5}\det\begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 3 \\ 2 & 1 & 1 \end{pmatrix} = \tfrac{1}{5}[-2(-5) - (1+2)] = \tfrac{7}{5}$$

$$x^3 = \tfrac{1}{5}\det\begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 0 \\ 2 & 2 & 1 \end{pmatrix} = \tfrac{1}{5}[2(0) + 1(1-2)] = -\tfrac{1}{5}$$

(the determinants are computed by column cofactor expansion).

• EXERCISES

33. Find the inverse of these matrices

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

(b) $\begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 4 & 1 & 1 \end{pmatrix}$

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

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

34. Solve the equation

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

where $\mathbf{A}$ is given by

(a) the matrix in Exercise 33(a)

(b) the matrix in Exercise 33(b)

(c)
$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 2 \\ 3 & -1 & 1 \end{pmatrix}$$

(d)
$$\mathbf{A} = \begin{pmatrix} 4 & 6 & 1 \\ 3 & 1 & 2 \\ 0 & 4 & 0 \end{pmatrix}$$

35. Suppose that the $n \times n$ matrix $\mathbf{A} = (a_j{}^i)$ has this property:

$$a_j{}^i = 0 \qquad \text{if } i \leq j$$

Show that $\mathbf{A}^n = 0$.

36. If $\mathbf{A}$ is a matrix such that $\mathbf{A}^n = \mathbf{0}$ show that $\mathbf{I} + \mathbf{A}$ is invertible.

• PROBLEMS

33. Show that if a linear transformation T has rank n, it is invertible. Show that if there is a transformation S such that $T \circ S = I$, T is invertible.

34. Derive Equations (1.35)–(1.38) using the definition of the determinant.

35. Assume this fact about polynomials: A polynomial of degree d has no more than d roots. Prove the following assertions:

(a) Let $\mathbf{A}$ be an $n \times n$ matrix. There are at most n numbers s such that $\mathbf{A} + s\mathbf{I}$ is not invertible.

(b) The $m \times n$ matrix

$$\mathbf{V} = \begin{pmatrix} 1 & r_1 & r_1{}^2 & \cdots & r_1^{n-1} \\ 1 & r_2 & r_2{}^2 & \cdots & r_2^{n-1} \\ 1 & r_n & r_n{}^2 & \cdots & r_n^{n-1} \end{pmatrix}$$

has a nonzero determinant if and only if the r_i are all distinct. (*Hint:* If $\det \mathbf{V} = 0$, there is a nonvanishing linear relation among the columns.)

36. Let

$$f(x^1, \ldots, x^n) = \prod_{i<j} (x^i - x^j)$$

where $x^1, \ldots, x^n$ are distinct numbers. Show that the permutation π is even if and only if

$$f(x^{\pi(1)}, \ldots, x^{\pi(n)}) = f(x^1, \ldots, x^n)$$

and similarly π is odd if and only if

$$f(x^{\pi(1)}, \ldots, x^{\pi(n)}) = -f(x^1, \ldots, x^n)$$

37. Let $\mathbf{A}$ be an invertible $n \times n$ matrix. For $m < n$, let $\mathbf{B}$ be an $(n-m) \times (n-m)$ matrix formed from $\mathbf{A}$ by deleting any m rows and m columns. Show that $\mathbf{B}$ is also invertible. (*Hint:* You need only take $m = 1$, and proceed by induction.)

38. Let
$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
Verify that

$$\mathbf{A}^2 - (a+d)\mathbf{A} + (ad - bc)\mathbf{I} = 0$$

that is, that $\mathbf{A}$ is a zero of a polynomial of degree 2.

39. The same fact is true for all n, that is an $n \times n$ matrix is the zero of a polynomial of degree n. This is part of a famous theorem of algebra, which goes like this: If $\mathbf{A}$ is any matrix, the polynomial

$$P_A(x) = \det(\mathbf{A} - x\mathbf{I})$$

is the **characteristic polynomial** of $\mathbf{A}$. $\mathbf{A}$ is a root of the polynomial equation $P_A(x) = 0$ (Cayley-Hamilton). That is,

$$P_A(\mathbf{A}) = \mathbf{0}$$

Verify the Cayley-Hamilton theorem for (i) a diagonal matrix, (ii) a triangular matrix.

1.7 Eigenvectors and Change of Basis

One fruitful way of studying linear transformations on R^n is to find directions along which they act merely by stretching the vector. For example, if a transformation T is represented by a diagonal matrix

$$\mathbf{A} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & d_n \end{pmatrix} \tag{1.39}$$

then $T(\mathbf{E}_i) = d_i\mathbf{E}_i$, where $\mathbf{E}_1, \ldots, \mathbf{E}_n$ are the standard basis vectors. Thus T acts by stretching by a factor d_i along the ith direction:

$$T(x^1, \ldots, x^n) = d_1 x^1 \mathbf{E}_1 + d_2 x^2 \mathbf{E}_2 + \cdots + d_n x^n \mathbf{E}_n$$

More generally, suppose we can find a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of vectors in R^n such that T acts by stretching along the direction of $\mathbf{v}_i$ for each i:

$$T(\mathbf{v}_i) = d_i \mathbf{v}_i$$

Then, if $\mathbf{v}$ is any vector, the action of T is easily computed by referring $\mathbf{v}$ to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$: if $\mathbf{v} = \sum s^i \mathbf{v}_i$, then $T(\mathbf{v}) = \sum d_i s^i \mathbf{v}_i$. T is represented by the diagonal matrix (1.39) relative to this basis. The process of finding a basis of vectors along which T acts by stretching is called **diagonalization.**

Unfortunately, not all transformations can be so diagonalized and this presents a major difficulty in this line of investigation. For example, a rotation in the plane clearly does not have any such directions in which it acts as a stretch. More precisely, let T be represented by the matrix

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

Then $T(x, y) = (y, -x)$. (T is a clockwise rotation through a right angle.) If $\mathbf{v} = (a, b)$ is such that $T(a, b) = d(a, b)$, we must have

$$da = b \qquad db = -a$$

Then $d^2 a = db = -a$, and there are no real numbers d, a making this equation true (except 0).

Nevertheless, there are many transformations which can be analyzed in this way, and it is our purpose in this section to study the techniques for doing so.

Definition 12. Let $T\colon R^n \to R^n$ be a linear transformation. An **eigenvalue** of T is a number d for which there exists a nonzero vector $\mathbf{v}$ such that $T\mathbf{v} = d\mathbf{v}$. An **eigenvector of T with eigenvalue** d is a nonzero vector $\mathbf{v}$ such that $T\mathbf{v} = d\mathbf{v}$.

Proposition 20. *If $T\colon R^n \to R^n$ is a linear transformation for which there is a basis of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with eigenvalues $d_1, \ldots, d_n$, respectively, then for any vector $\mathbf{v} = \sum s^i \mathbf{v}_i$, $T(\mathbf{v}) = \sum d_i s^i \mathbf{v}_i$.*

Proof. Compute $T(\mathbf{v})$ using the fact that T is linear.

Now we find the eigenvalues of a linear transformation T by making use of this remark: d is an eigenvalue of T if and only if $T - dI$ is singular (not invertible). If $\mathbf{A}$ is the matrix representing T in terms of the standard basis, this condition is verified precisely when $\det(\mathbf{A} - d\mathbf{I}) = 0$. Thus the eigenvalues of T are just the roots of this equation. Notice that when T is rotation by a right angle

$$\det\left[\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} - d\mathbf{I}\right] = d^2 + 1$$

which has no real roots, thus explaining in another way why this transformation has no eigenvectors. We shall see that when we extend the real number system to a system in which every polynomial has a root (the complex numbers), then T can be represented in terms of (complex) eigenvectors. This is one of the important reasons (particularly in the study of differential equations, as we shall see) for so extending the number system. Let us now collect these observations.

Proposition 21. *Let T be a transformation on R^n represented by the matrix* $\mathbf{A}$. *d is an eigenvalue of T if and only if d is a root of the equation*

$$\det(\mathbf{A} - t\mathbf{I}) = 0$$

If d is an eigenvalue, the set of eigenvectors corresponding to d is the kernel of $T - dI$.

Proof. Suppose d is an eigenvalue of T. Then there is a $\mathbf{v} \neq \mathbf{0}$ such that $T\mathbf{v} = d\mathbf{v}$, or $(T - dI)\mathbf{v} = \mathbf{0}$. Thus the nullity of $T - dI$ is positive, so $T - dI$ is not invertible. Thus, $\det(\mathbf{A} - d\mathbf{I}) = 0$. On the other hand, if $\det(\mathbf{A} - d\mathbf{I}) = 0$, then $T - dI$ is not invertible, so has a positive dimensional kernel. If $\mathbf{v} \neq \mathbf{0}$ is in the kernel, $(T - dI)(\mathbf{v}) = 0$, or $T\mathbf{v} = d\mathbf{v}$; thus d is an eigenvector of T.

Examples

32. Let T be represented by the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

Then

$$\mathbf{A} - t\mathbf{I} = \begin{pmatrix} 2-t & 0 \\ 1 & 1-t \end{pmatrix}$$

and $\det(\mathbf{A} - t\mathbf{I}) = t^2 - 3t + 2$. The roots are $t = 2, 1$. The space of eigenvectors corresponding to $t = 2$ is the kernel of

$$\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$$

that is, the space of all vectors (x, y) such that $x - y = 0$. Thus $(1, 1)$ is an eigenvector with eigenvalue 2. The eigenvectors corresponding to $t = 1$ lie in the kernel of

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

that is, in the space of vectors (x, y) such that $x = 0$. $(0, 1)$ is such an eigenvector. Since $(1, 1)$ and $(0, 1)$ are a basis for R^2, we have diagonalized T. Relative to this basis T is represented by the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

33. Consider the transformation given, relative to the standard basis by the matrix

$$\begin{pmatrix} 2 & -4 \\ 1 & 6 \end{pmatrix}$$

Then $\det(\mathbf{A} - t\mathbf{I}) = t^2 - 8t + 16 = (t - 4)^2$. Thus 4 is the only eigenvalue of T.

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

has as kernel $\{(x, y): x + 2y = 0\}$, which is one dimensional. Thus there cannot be a basis of eigenvectors for the only eigenvectors lie on the line $x = -2y$.

Notice that this example differs from that of a rotation, for there is no problem with the roots; the difficulty lies with the transformation itself.

34. Let $T: R^3 \to R^3$ be given by the matrix

$$\mathbf{A} = \begin{pmatrix} -7 & 0 & -18 \\ 2 & 2 & 4 \\ 3 & 0 & 8 \end{pmatrix}$$

Then $\det(\mathbf{A} - t\mathbf{I}) = -t^3 + 3t^2 - 4$. The roots of

$$\det(\mathbf{A} - t\mathbf{I}) = 0$$

are $2, -1$.

Eigenvalue 2:

$$\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} -9 & 0 & -18 \\ 2 & 0 & 4 \\ 3 & 0 & 6 \end{pmatrix}$$

The kernel is the set of vectors (x, y, z) such that $x + 2z = 0$. This space is two dimensional, so we can find two independent eigenvectors with eigenvalue 2; for example, $\mathbf{v}_1 = (0, 1, 0)$, $\mathbf{v}_2 = (-2, 0, 1)$.

Eigenvalue -1:

$$\mathbf{A} - (-1)\mathbf{I} = \begin{pmatrix} -6 & 0 & -18 \\ 2 & 1 & 4 \\ 3 & 0 & 9 \end{pmatrix}$$

The kernel is the set of vectors (x, y, z) such that

$$\begin{aligned} x + 3z &= 0 \quad \text{or} \quad x = -3z \\ 2x + y + 4z &= 0 \quad \text{or} \quad y = 2z \end{aligned}$$

which is one dimensional. An eigenvector is $\mathbf{v}_3 = (-3, 2, 1)$. These $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ thus form a basis of eigenvectors, and T is represented by the matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{1.40}$$

relative to the basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

Jordan Canonical Form

Notice that in general there are two difficulties with the procedure described above. The polynomial $\det(\mathbf{A} - t\mathbf{I})$ may not have many real roots, and it may have multiple roots. As we shall see in the next section the first difficulty can be overcome by transferring to the complex number system. Example 34 above demonstrates that the second possibility, that of multiple roots, may not be severe, whereas Example 33 shows that it can seriously handicap the diagonalization procedure. Continued study of this situation becomes quite difficult and we shall not enter into it. The conclusion is that the typical matrix which cannot be diagonalized is of this form

$$\begin{pmatrix} d & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & d & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & d & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & d & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdots & 0 & d \end{pmatrix} \tag{1.41}$$

representing the transformation

$$T(x^1, \ldots, x^n) = (dx^1 + x^2, dx^2 + x^3, \ldots, dx^n)$$

Given any matrix, we can find a basis of vectors (which includes all possible eigenspaces) relative to which T decomposes into pieces, each of which has the form (1.41). This is called the **Jordan canonical form.**

Change of Basis

Before leaving this subject, let us compute explicitly the formulas which allow us to change bases in R^n. If $\{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$ is a basis for R^n, then any $\mathbf{x}$ in R^n can be written

$$\mathbf{x} = x^1\mathbf{E}_1 + \cdots + x^n\mathbf{E}_n$$

uniquely. We shall refer to the n-tuple $(x^1, \ldots, x^n)$ as the *coordinate of* $\mathbf{x}$ *relative to the basis* $\mathbf{E}$: $\{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$ denoted $\mathbf{x}_E$.

Let $\mathbf{F}$: $\{\mathbf{F}_1, \ldots, \mathbf{F}_n\}$ be another basis for R^n. Let $\mathbf{x}_F$ be the coordinates of $\mathbf{x}$ relative to this new basis. To each set of $\mathbf{E}$ coordinates $\mathbf{x}_E$ we can associate the $\mathbf{F}$ coordinates $\mathbf{x}_F$ of the point corresponding to $\mathbf{x}_E$. In this way we can write $\mathbf{x}_F$ as a function of $\mathbf{x}_E$. The precise relation is this.

Proposition 22. *Let* $\mathbf{E}$: $\{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$, $\mathbf{F}$: $\{\mathbf{F}_1, \ldots, \mathbf{F}_n\}$ *be two different bases for* R^n. *Write the* $\mathbf{E}$*'s in terms of the* $\mathbf{F}$*'s:*

$$\mathbf{E}_j = \sum_{i=1}^{n} a_j{}^i \mathbf{F}_i$$

The matrix $(a_j{}^i)$ *is called the* **change of basis matrix,** *and is denoted* $\mathbf{A}_F{}^E$. *For any point* $\mathbf{x}$ *in* R^n *we have this relation between its* $\mathbf{E}$ *and* $\mathbf{F}$ *coordinates:*

$$\mathbf{x}_F = \mathbf{A}_F{}^E \mathbf{x}_E \tag{1.42}$$

Proof. Let $\mathbf{x}_E = (x^1, \ldots, x^n)$, $\mathbf{x}_F = (y^1, \ldots, y^n)$. Then

$$\begin{aligned}
\mathbf{x} = \sum_{j=1}^{n} x^j \mathbf{E}_j &= \sum_{j=1}^{n} x^j \left(\sum_{i=1}^{n} a_j{}^i \mathbf{F}_i \right) \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_j{}^i x^j \right) \mathbf{F}_i \\
&= \sum_{i=1}^{n} y^i \mathbf{F}_i
\end{aligned}$$

Thus for each i, $y^i = \sum_{j=1}^{n} a_j{}^i x^j$, which is the same as (1.42).

Notice that it follows from (1.42) that $(\mathbf{A}_F{}^E)^{-1} = \mathbf{A}_E{}^F$. For, given any $\mathbf{x}_F$

$$\mathbf{x}_F = \mathbf{A}_F{}^E \mathbf{x}_E = \mathbf{A}_F{}^E \mathbf{A}_E{}^F \mathbf{x}_F$$

Thus $\mathbf{A}_F{}^E \mathbf{A}_E{}^F = \mathbf{I}$.

Now, if T is any linear transformation on R^n, it can be represented by a a matrix, relative to any basis $\mathbf{E}$: $\{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$. Let us denote that matrix by $\mathbf{T}_E$:

$$T(\mathbf{x})_E = \mathbf{T}_E \mathbf{x}_E$$

Proposition 23. *If* $\mathbf{E}$: $\{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$, $\mathbf{F}$: $\{\mathbf{F}_1, \ldots, \mathbf{F}_n\}$ *are two bases of* R^n, *and* T: $R^n \to R^n$ *is a linear transformation, we have*

$$\mathbf{T}_F = (\mathbf{A}_E{}^F)^{-1} \mathbf{T}_E \mathbf{A}_E{}^F$$

Proof.

$$T(\mathbf{x})_F = \mathbf{A}_F{}^E\mathbf{T}(\mathbf{x})_E = \mathbf{A}_F{}^E\mathbf{T}_E\mathbf{x}_E = \mathbf{A}_F{}^E\mathbf{T}_E\mathbf{A}_E{}^F\mathbf{x}_F$$

On the other hand, by definition

$$T(\mathbf{x})_F = \mathbf{T}_F\mathbf{x}_F$$

Thus $\mathbf{T}_F = \mathbf{A}_F{}^E\mathbf{T}_E\mathbf{A}_E{}^F = (\mathbf{A}_E{}^F)^{-1}\mathbf{T}_E\mathbf{A}_E{}^F$

Examples

35. Let T: $R^2 \to R^2$ be represented, relative to the standard basis **E** by

$$\mathbf{T}_E = \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}$$

Let **F**: $\{(1, 1),\ (2, -1)\}$ be another basis. Find the matrix $\mathbf{T}_F$. Now,

$$\mathbf{A}_E{}^F = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$

$$\mathbf{A}_F{}^E = (\mathbf{A}_E{}^F)^{-1} = \tfrac{1}{3}\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$

Thus

$$\mathbf{T}_F = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}\tfrac{1}{3}\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 7/3 & -1 \\ 4/3 & 2/3 \end{pmatrix}$$

36. Let T be given, relative to the standard basis **E** by

$$\mathbf{T}_E = \begin{pmatrix} -7 & 0 & -18 \\ 2 & 2 & 4 \\ 3 & 0 & 8 \end{pmatrix}$$

and let **F**: $\{(0, 1, 0),\ (-2, 0, 1),\ (-3, 2, 1)\}$. We have already seen that **F** is a basis of eigenvectors for T, with eigenvalues $2, 2, -1$, respectively. Thus we may conclude that $\mathbf{T}_F$ is given by (1.40).

• EXERCISES

37. Find a basis of eigenvectors, if possible, for the transformation represented in terms of the standard basis by the matrix **A**:

(a)
$$\mathbf{A} = \begin{pmatrix} 3 & 4 & 2 \\ 2 & 7 & 5/2 \\ -4 & -16 & -6 \end{pmatrix} \qquad \text{eigenvalues: } 2,\ 3,\ -1$$

(b)
$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 2 & 2 \\ -1 & 0 & -3 & -3 \end{pmatrix} \qquad \text{eigenvalues: } 1,\ -1,\ 0,\ 2$$

(c)
$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & -4 & -3 \end{pmatrix} \qquad \text{eigenvalues: } 1,\ 4$$

(d)
$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 2 & 2 & 0 \end{pmatrix}$$

38. Show that for $\mathbf{F}: \{\mathbf{F}_1, \ldots, \mathbf{F}_n\}$ a basis for R^n, and **E** the standard basis, the matrix $\mathbf{A}_E{}^F$ is just the matrix whose columns are $\mathbf{F}_1, \ldots, \mathbf{F}_n$.

39. Find the matrix $\mathbf{A}_E{}^F$ for these pairs of bases in R^n.

(a) $\mathbf{F}$: (1, 0, 1), (0, 1, 1), (1, 0, 0)
$\mathbf{E}$: (0, 1, 2), (2, 0, 1), (1, 2, 0).

(b) $\mathbf{F}$: (1, 0, 0), (2, 0, 1), (0, 1, 0)
$\mathbf{E}$: (3, 1, 5), (0, 2, 3), (−1, −1, 0).

(c) $\mathbf{F}$: (1, 0, 1, 0), (0, 1, 1, 0), (0, 0, 2, 0), (0, 0, 1, 1)
$\mathbf{E}$: (0, 2, 0, 2), (2, 0, 0, 0), (2, 0, 2, 0), (0, 2, 2, 0).

40. Let $T\colon R^3 \to R^3$ be a linear transformation represented by one of

(a)
$$\mathbf{T}_E = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 0 & 3 \\ 1 & 0 & 1 \end{pmatrix}$$
(b)
$$\mathbf{T}_E = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 4 \\ 2 & 0 & -1 \end{pmatrix}$$

relative to the standard basis **E**. Find $\mathbf{T}_F$, where **F** is one of these bases (**F** as in Exercises 39(a) and 39(b)).

41. If $T\colon R^2 \to R^2$ has two independent eigenvectors with the same eigenvalue, then T is represented by a diagonal matrix in any basis.

• PROBLEMS

40. Prove Proposition 20.

41. If T is a linear transformation on R^n represented by the matrix **A** which has n distinct eigenvalues $d_1, \ldots, d_n$, then

$$P_A(x) = (-1)^n(x - d_1)(x - d_2) \cdots (x - d_n)$$

and $P_A(\mathbf{A}) = \mathbf{0}$ (P_A is defined in Problem 39).

42. Let T be a linear transformation on R^n. Let $E(r) = \{\mathbf{v} \in R^n : T\mathbf{v} = r\mathbf{v}\}$. Show that $E(r)$ is a linear subspace of R^n (called the r **eigenspace of** T). Show that if $r \neq s$, then $E(r) \cap E(s) = \{\mathbf{0}\}$.

43. Suppose $\mathbf{A}$ represents a linear transformation on R^n with this property: if $r_1, \ldots, r_k$ are the eigenvalues of $\mathbf{A}$, then $n = \sum_{j=1}^{k} \dim E(r_j)$. Then

$$P_A(x) = (-1)^n (x - r_1)^{\dim E(r_1)} \cdots (x - r_k)^{\dim E(r_k)}$$

Verify the Cayley-Hamilton theorem for $\mathbf{A}$.

44. Find a matrix with no nontrivial eigenspaces. How would you expect to prove the Cayley-Hamilton theorem for such a matrix?

1.8 Complex Numbers

Pythagoras' discovery, that $\sqrt{2}$ is not the quotient of two integers, was considered in his day to be a geometric mystery. His conception of numbers was limited to rational numbers and his desire to measure lengths (to associate numbers to line segments) led to this unhappy realization: there are some lengths which are not measurable! (as the hypotenuse of an isosceles right triangle of leg length 1). It took a long time for mathematicians to realize that the solution to this situation was to expand the notion of number. The general liberation of thought that was the Renaissance led in mathematics to the possibility of expressing the value of certain lengths by never-ending decimals, or continued fractions, or other types of infinite expressions. It was during those days that mathematicians formulated the view that such expressions represented numbers and served to determine all lengths. Earlier, Middle Eastern mathematicians were led from certain algebraic problems to envision extension of the number concept in another direction. As they observed, quite clearly -1 has no square root; some bold adventurer then suggested that we contemplate, in our minds, some purely imaginary quantity whose square would be -1 and treat it as if it were another number. As this supposition did not contradict any of the known facts concerning the number system, it could do no harm—and might do a great deal of good (at least in our minds).

Today we need not be so mysterious or cunning in our ways. We need only recall that there is a 2×2 matrix (see Problem 20) whose square is the negative of the identity. We can thus say quite factually that in the set of 2×2 matrices, -1 does indeed have a square root. Well, there is also a 5×5 matrix, and an $n \times n$ matrix for any n whose square is $-\mathbf{I}$, so we should ask for *the smallest* algebraic system in which -1 has a square root. The

complex number system is this system and we shall later derive the remarkable fact (the fundamental theorem of algebra): Every polynomial has a root in the complex number system.

Now, to be explicit, the matrix

$$\mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{1.43}$$

has the property that $\mathbf{i}^2 = -\mathbf{I}$. The complex number system is the collection of all 2×2 matrices of the form $a\mathbf{I} + b\mathbf{i}$, where a, b are real numbers.

Definition 13. C, **the set of complex numbers** is the collection of all 2×2 matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

Proposition 24.
(i) *The operations of addition and multiplication are defined on C.*
(ii) *Every nonzero complex number has an inverse.*
(iii) *C is in one-to-one correspondence with R^2.*

Proof.
(i) $$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} a+c & -(b+d) \\ b+d & a+c \end{pmatrix}$$

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} ac-bd & -(ad+bc) \\ ad+bc & ac-bd \end{pmatrix}$$

(ii) If

$$\mathbf{M} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

is nonzero, then one of a or b is nonzero, so $\det \mathbf{M} = a^2 + b^2 \neq 0$, and thus $\mathbf{M}$ has an inverse. By Cramer's rule

$$\mathbf{M}^{-1} = \frac{1}{a^2 + b^2}\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

(iii) is obvious, since every complex number is given by a pair of real numbers and conversely every pair (a, b) of real numbers gives rise to a complex number.

Cartesian Form of a Complex Number

We need now a notation which is more convenient than the matrix notation, and we get our cue from (iii) above. The matrices **I** and **i** correspond to the points (1, 0), (0, 1) of the plane and thus form a basis for C. More explicitly

$$\begin{aligned}\begin{pmatrix} a & -b \\ b & a \end{pmatrix} &= a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= a\mathbf{I} + b\mathbf{i} \end{aligned} \tag{1.44}$$

If we identify the real number 1 with the identity matrix **I**; and more generally the real number r with the complex number $r\mathbf{I} + 0\mathbf{i}$, then we can say that every real number is also a complex number. In fact, the complex system is just the real number system with a square root of -1 tacked on. (This takes us full circle back to the original conception of that Arabian adventurer. The difference here is that we now know what we mean by this procedure and that it produces no inconsistencies.)

Thus, we can suppress the identity matrix in the expression and write a complex number in the form $a + bi$. We now recapitulate the relevant facts.

C is the set of all 2×2 matrices $c = a + bi$ with a, b real numbers. a is the **real part** of c, written $a = \operatorname{Re} c$, and b is the **imaginary part**, written $b = \operatorname{Im} c$. And these following rules hold:

$$\begin{aligned} i^2 &= -1 \\ (a + bi) + (c + di) &= (a + c) + (b + d)i \\ (a + bi)(c + di) &= ac - bd + (bd + ac)i \\ (a + bi)^{-1} &= \frac{a - bi}{a^2 + b^2} \qquad \text{when } a^2 + b^2 \neq 0 \end{aligned}$$

Polar Form of a Complex Number

Since C is in one-to-one correspondence with R^2, we can represent complex numbers by points in the plane (see Figure 1.14). Addition of complex numbers is the same as addition of vectors in the plane. We now seek a geometric description of multiplication of complex numbers. For this purpose it is convenient to move to polar coordinates.

Definition 14. Let $z = x + yi$. **The modulus** of z, written $|z|$, is its distance from the origin:

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

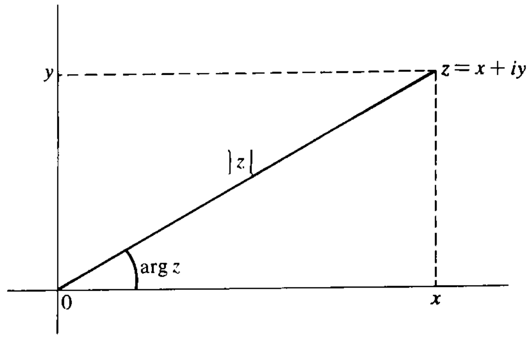


Figure 1.14

The argument of z, written arg z, is defined for $z \neq 0$; it is the angle defining the ray on which z lies:

$$\arg z = \tan^{-1} \frac{y}{x}$$

We can write complex numbers in polar form: If $a = x + yi$ has the polar coordinates (r, θ) then, since $x = r \cos \theta$, $y = r \sin \theta$, we have

$$z = r(\cos \theta + i \sin \theta)$$

(We have moved the i in front of $\sin \theta$ for the obvious notational convenience which results.) The set of points of modulus 1 is the unit circle centered at the origin. It is the set of all points of the form $\cos \theta + i \sin \theta$. We shall sometimes abbreviate this to cis θ. Precisely, cis θ is the point of the unit circle lying on the ray of angle θ. Now, let z, w be two complex numbers,

$$z = r \operatorname{cis} \theta \qquad w = \rho \operatorname{cis} \phi$$

Then

$$\begin{aligned} zw &= r \operatorname{cis}(\theta) \rho \operatorname{cis}(\phi) \\ &= (r \cos \theta + ir \sin \theta)(\rho \cos \phi + i\rho \sin \phi) \\ &= r\rho(\cos \theta \cos \phi - \sin \theta \sin \phi) + ir\rho(\cos \theta \sin \phi + \cos \phi \sin \theta) \\ &= r\rho \operatorname{cis}(\theta + \phi) \end{aligned}$$

Thus we form the product of two complex numbers by *multiplying the modulii and adding the arguments.* (This does not make sense if one of the numbers is zero, but that case is trivial anyway.)

Notice then, if $z = \rho \text{ cis } \theta$, then $z^2 = \rho^2 \text{ cis } 2\theta$ and more generally

$$z^n = \rho^n \text{ cis } n\theta \tag{1.45}$$

This observation leads to the fact that it is easy to extract roots. For the converse of (1.45) is

$$z^{1/k} = \rho^{1/k} \text{ cis } \frac{\theta}{k}$$

Proposition 25. *Let c be a complex number, and k an integer. There are precisely k distinct solutions to the equation $X^k = c$.*

Proof. Write c in polar form: $c = r \text{ cis } \theta$. If $z = \rho \text{ cis } \phi$ is a solution, then

$$r \text{ cis } \theta = c = z^k = \rho^k \text{ cis } k\phi$$

Thus the modulus of z is the kth root of the modulus of c, and the argument of a is an angle such that k times it is θ. Well, $(1/k)\theta$ is such an angle, but so is $(1/k)(\theta + 2\pi)$. In fact, each of the angles

$$\frac{1}{k}(\theta), \frac{1}{k}(\theta + 2\pi), \frac{1}{k}(\theta + 4\pi), \ldots, \frac{1}{k}(\theta + 2(k-1)\pi)$$

have the property that k times it is θ. All these angles are distinct, so $c = r \text{ cis } \theta$ has precisely these k roots:

$$r^{1/k} \text{ cis } \theta, r^{1/k} \text{ cis } \frac{\theta + 2\pi}{k}, \ldots, r^{1/k} \text{ cis } \frac{\theta + 2\pi(k-1)}{k}$$

Complex Eigenvalues

We shall work extensively with the complex number system in this text. In fact, we shall discover many situations besides the algebraic one above where study within the system of complex numbers is beneficial. In particular, let us return to the eigenvalue problems of the preceding section. We consider C^n, the space of n-tuples of complex numbers. We can define linear transformations on C^n just as we did on R^n. In fact, the entire theory of linear algebra through Section 1.7 holds over C^n as well as R^n. Let

$$\mathbf{E}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0) \qquad (1 \text{ in the } i\text{th place})$$

be the **standard basis vectors** for C^n. Again, any linear transformation on C^n is given by a matrix $\mathbf{A} = (a_j{}^i)$ of complex numbers relative to the standard basis:

$$T(z^1, \ldots, z^n) = (\sum a_j{}^1 z^j, \ldots, \sum a_j{}^n z^j)$$

for all $(z^1, \ldots, z^n) \in C^n$.

Examples

37. Consider the matrix

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

as representing a transformation T on C^2 relative to the standard basis. Its eigenvalues are the roots of $\det(\mathbf{A} - t\mathbf{I}) = 0$. But $\det(\mathbf{A} - t\mathbf{I}) = t^2 + 1$, so the roots are $i, -i$.

Eigenvalue i:

$$\mathbf{A} - i\mathbf{I} = \begin{pmatrix} -i & 1 \\ -1 & i \end{pmatrix}$$

The second row is $-i$ times the first, so the kernel of $\mathbf{A} - i\mathbf{I}$ is given by the single equation $-ix + y = 0$. An eigenvector is $(1, i)$.

The eigenvalue $-i$ has the eigenvector $(1, -i)$. Now, $\mathbf{F}$: $\{(1, i), (1, -i)\}$ are a basis of eigenvectors for C^2, so T becomes diagonalized relative to this basis:

$$\mathbf{T}_F = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

if

$$\mathbf{z} = z^1\begin{pmatrix} 1 \\ i \end{pmatrix} + z^2\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

then

$$T\mathbf{z} = iz^1\begin{pmatrix} 1 \\ i \end{pmatrix} - iz^2\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

38. Consider the matrix

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

representing a transformation T on C^3 relative to the standard basis. $\det(\mathbf{A} - t\mathbf{I}) = -t^3 + t^2 - t + 1$. This polynomial has the roots $1, i, -i$. Since the roots are distinct and each must have a corresponding eigenvector, there is a basis of eigenvectors. We now find such a basis.

Eigenvalue 1:

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

The kernel of $\mathbf{A} - \mathbf{I}$ is found as a linear relation among the columns (recall Example 18). Such a relation is

$$\mathbf{C}_1 - 2\mathbf{C}_2 - 3\mathbf{C}_3 = 0$$

Thus $(1, -2, -3)$ is an eigenvector with eigenvalue 1.

Eigenvalue i:

$$\mathbf{A} - i\mathbf{I} = \begin{pmatrix} -i & 1 & -1 \\ \frac{1}{4} & -i & \frac{3}{4} \\ 2 & 1 & 1-i \end{pmatrix}$$

In order to find a relation among the columns we must row reduce. The result of row reduction is

$$\begin{pmatrix} 1 & i & -i \\ 0 & 1 & -\frac{1}{5} + \frac{3}{5}i \\ 0 & 0 & 0 \end{pmatrix}$$

A solution of the corresponding homogeneous system is found by taking $z_3 = 5$, then we obtain $z_2 = 1 - 3i$, $z_1 = -3 + 4i$. Thus, $(-3 + 4i, 1 - 3i, 5)$ is an eigenvector with eigenvalue i. Similarly, we find the eigenvector $(-3 - 4i, 1 + 3i, 5)$ corresponding to the eigenvalue $-i$. Thus, T is represented by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$$

relative to the basis

$$(1, -2, -3), \quad (-3+4i, 1-3i, 5), \quad (-3-4i, 1+3i, 5).$$

• EXERCISES

42. Find the inverse of these complex numbers:
(a) $5-3i$ (d) $4 \operatorname{cis}(2/3)$
(b) $(1-i)/2$ (e) $\operatorname{cis} 7$
(c) $3+i$

43. Show that $z^{-1} = \bar{z}$ if and only if z is on the unit circle.

44. Show that the complex number $\operatorname{cis} \theta$ represents rotation in the plane through the angle θ, when considered as a 2×2 matrix.

45. Find all kth roots of z:
(a) $k=2, z=-i$. (d) $k=3, z=i$.
(b) $k=5, z=-1$ (e) $k=2, z=3i-4$
(c) $k=4, z=1+i$ (f) $k=3, z=15+5i$

46. Find, if possible, a basis of possibly complex eigenvectors for the transformations represented by these matrices

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

(b) $\begin{pmatrix} 17 & -13 \\ 5 & -1 \end{pmatrix}$ (d) $\begin{pmatrix} 0 & 1 & 1 & -1 \\ 2 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 2 \end{pmatrix}$

• PROBLEMS

45. Compute that the matrix
$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
has square equal to $-\mathbf{I}$. We have chosen $\mathbf{i}$ to be the 2×2 matrix
$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
so that the correspondence between complex numbers and operations on R^2 will be correct. More precisely, we conceive a complex number in two ways: as a certain transformation on the plane, and as a vector on the plane. Given two complex numbers z, w we may interpret their product in two

ways: composition of the transformations, or the application of the transformation corresponding to z to the vector w. We would like these two interpretations to have the same result. If $z = a + ib$, $w = c + id$, show that zw,

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c & -d \\ d & c \end{pmatrix}$$

and

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix}$$

are all the same under that correspondence.

46. Show that the complex numbers z, $\bar{z}$, when considered as vectors in R^2 are independent (unless they are real or pure imaginary).

47. Why do the complex eigenvalues of a real matrix come in conjugate pairs?

1.9 Space Geometry

In this section we shall introduce the basic notions of three-dimensional geometry, using vector notation. First of all, as in the plane, we select a particular point in space, called the origin and denoted $\mathbf{0}$. That being done we may refer to the points of space as vectors and think heuristically of the directed line segment from the origin to the point as a vector. The operations of scalar multiplication and addition can be defined as on the plane—and expressed in terms of coordinates in much the same way:

(i) If $\mathbf{P}$ is a vector and r a real number, $r\mathbf{P}$ is the vector lying on the line through $\mathbf{0}$ and $\mathbf{P}$ and of distance from $\mathbf{0}$ equal to $|r|$ times the length of the segment $\mathbf{0P}$. If $r > 0$, $r\mathbf{P}$ lies on the same side of $\mathbf{0}$ as $\mathbf{P}$; if $r < 0$, $r\mathbf{P}$ lies on the opposite side.

(ii) if $\mathbf{P}$, $\mathbf{Q}$ are two vectors in space, there is a unique parallelogram lying in the plane determined by $\mathbf{P}$ and $\mathbf{Q}$, three of whose vertices are $\mathbf{0}$, $\mathbf{P}$, $\mathbf{Q}$. We define $\mathbf{P} + \mathbf{Q}$ to be the fourth vertex.

Now, we turn to the coordinatization of space. Having chosen a point as origin, let $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ be three new points with the property that $\mathbf{0}$, $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ do not all lie on the same plane (we say **the vectors** $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ **are not coplanar**). The three lines determined by the vectors $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ are called the **coordinate axes**. Just as in two dimensions the choices of the vectors $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ enables us to put each line in one-to-one correspondence with the real numbers.

In three dimensions two lines determine a plane. We shall call the planes

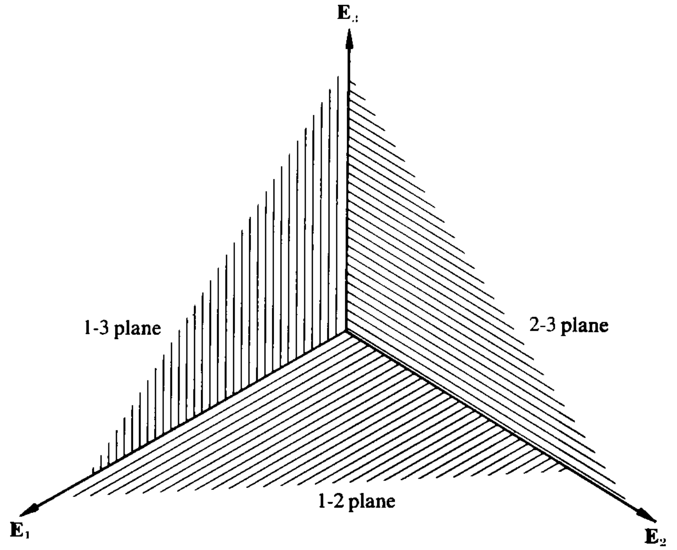


Figure 1.15

through **0** determined by $\mathbf{E}_1$ and $\mathbf{E}_2$ the 1-2 plane, by $\mathbf{E}_1$ and $\mathbf{E}_3$ the 1-3 plane, and the plane determined by $\mathbf{E}_2$ and $\mathbf{E}_3$ is the 2-3 plane (Figure 1.15). These three planes are called the **coordinate planes**. Each of these planes can be put into one-to-one correspondence with R^2 just as in the case of two dimensions. Now, to each point in space we can associate a triple of numbers relative to these choices in the following way. Let **P** be any such point. There is a unique plane through **P** which is parallel to the 2-3 plane; and this plane intersects the 1 axis in a unique point. This point has the coordinate x^1 relative to the scale determined by $\mathbf{E}_1$. We shall call x^1 the **first coordinate** of **P**. The second, x^2, is found in the same way: by intersecting the plane through **P** and parallel to the 1-3 plane with the 2 axis. Finally we find the third coordinate x^3 similarly, and associate the triple (x^1, x^2, x^3) to **P**. In this way we put all of space into one-to-one correspondence with R^3, dependent upon the choice of vectors $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$, called a **basis** for space.

The expression in terms of coordinates of the operations of addition and scalar multiplication are precisely the same as in R^2 (no matter what basis is chosen):

$$r(x^1, x^2, x^3) = (rx^1, rx^2, rx^3)$$

$$(x^1, x^2, x^3) + (y^1, y^2, y^3) = (x^1 + y^1, x^2 + y^2, x^3 + y^3)$$

There is no need to check that these formulas correspond to the geometric descriptions given above; we need only refer to the computation in the plane.

When we are interested in the pictorial representation of problems of three-dimensional Eculidean geometry it is best if we consistently use a particular coordinatization. For this purpose we select the "right-handed rectangular coordinate system"; where the coordinate axes are mutually perpendicular and the order $1 \to 2 \to 3$ is that of a right-handed screw (see Figure 1.16). It is common in particular problems to refer to the coordinates by the letters (x, y, z) rather than (x^1, x^2, x^3). We shall use the numbered coordinates when it is more convenient to do so.

Inner Product

Now, the basic notions of Euclidean geometry are length and angle. It will be of importance to us to derive expressions for these in terms of coordinates. Consider first the length of the line segment $\mathbf{0P}$ between the origin and the point $\mathbf{P}$ with coordinates (x, y, z). This can be easily computed by use of the Pythagorean theorem (consult Figure 1.17). Let $\mathbf{P}'$ be the point of intersection with the xz plane of the line through $\mathbf{P}$ and parallel to the y axis. Then $\mathbf{0PP}'$ is a right triangle, so

$$|\mathbf{0P}|^2 = |\mathbf{0P}'|^2 + |\mathbf{P}'\mathbf{P}|^2$$

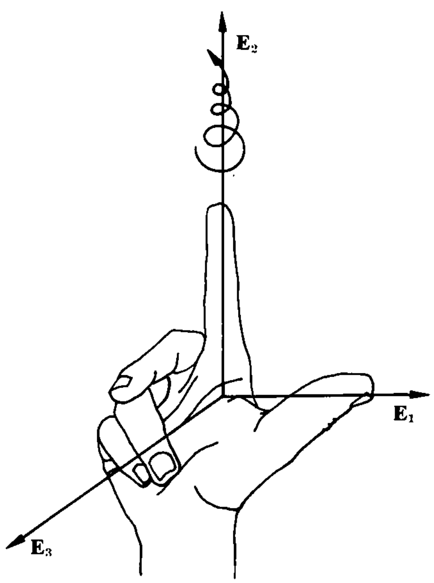


Figure 1.16

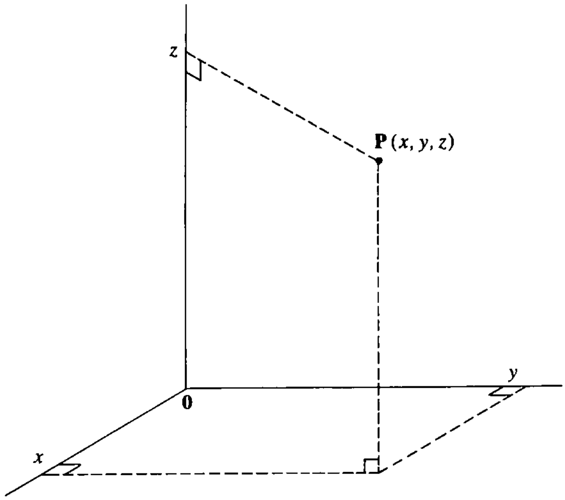


Figure 1.17

Letting **P**″ be the point of intersection with the x axis of the line through **P**′ and parallel to the z axis, we obtain

$$|\mathbf{0P}|^2 = |\mathbf{0P}''|^2 + |\mathbf{P}''\mathbf{P}'|^2 + |\mathbf{P}'\mathbf{P}|^2$$

But now $|\mathbf{0P}''|^2 = x^2$, $|\mathbf{P}''\mathbf{P}'|^2 = z^2$, $|\mathbf{P}'\mathbf{P}|^2 = y^2$, so

$$|\mathbf{0P}| = [x^2 + y^2 + z^2]^{1/2}$$

Now, suppose $\mathbf{P}(x, y, z)$, $\mathbf{Q}(a, b, c)$ are any two points in space. By definition of addition, **P** is the fourth vertex of the parallelogram three of whose vertices are **0**, $\mathbf{P} - \mathbf{Q}$ and **Q**. Thus, the side through **P** and **Q** has the same length as the side through $\mathbf{P} - \mathbf{Q}$ and **0**, so

$$|\mathbf{PQ}| = |(\mathbf{P} - \mathbf{Q})\mathbf{0}| = [(x - a)^2 + (y - b)^2 + (z - c)^2]^{1/2} \tag{1.46}$$

Finally, we can compute the angle between **P** and **Q** by the law of cosines (consult Figure 1.18); if θ is that angle, then

$$|\mathbf{PQ}|^2 = |\mathbf{0P}|^2 + |\mathbf{0Q}|^2 - 2|\mathbf{0P}|\,|\mathbf{0Q}| \cos\theta$$

In coordinates,

$$\begin{aligned}(x - a)^2 + (y - b)^2 + (z - c)^2 = x^2 + y^2 + z^2 + a^2 + b^2 + c^2 \\ -2(x^2 + y^2 + z^2)^{1/2} \\ \times (a^2 + b^2 + c^2)^{1/2} \cos\theta\end{aligned}$$

which reduces to

$$\cos\theta = \frac{xa + yb + cz}{(x^2 + y^2 + z^2)^{1/2}(a^2 + b^2 + c^2)^{1/2}} \tag{1.47}$$

The form in the numerator thus has some special importance: it together with the notion of length determines angles. It is called the **inner product** of the two vectors **P**, **Q**.

Definition 15. Let **P**, **Q** be two vectors in space. Their **Euclidean inner product**, denoted $\langle \mathbf{P}, \mathbf{Q} \rangle$, is defined as $|\mathbf{P}|\,|\mathbf{Q}|\cos\theta$, where θ is the angle between **P** and **Q**. In coordinates, $\mathbf{P} = (x_1, y_1, z_1)$, $\mathbf{Q} = (x_2, y_2, z_2)$,

$$\langle \mathbf{P}, \mathbf{Q} \rangle = x_1x_2 + y_1y_2 + z_1z_2$$

Propositions 26. *The nonzero vectors* **P** *and* **Q** *are perpendicular if and only if* $\langle \mathbf{P}, \mathbf{Q} \rangle = 0$.

Proof. **P** and **Q** are perpendicular if and only if the angle θ between them is a right angle. θ is a right angle if and only if $\cos\theta = 0$, and this holds precisely when $\langle \mathbf{P}, \mathbf{Q} \rangle = 0$.

A **plane through the origin** is the linear span of two vectors. If **N** is a vector perpendicular to such a plane $\prod$, then $\prod$ is given by the equation

$$\prod : \langle \mathbf{x}, \mathbf{N} \rangle = 0$$

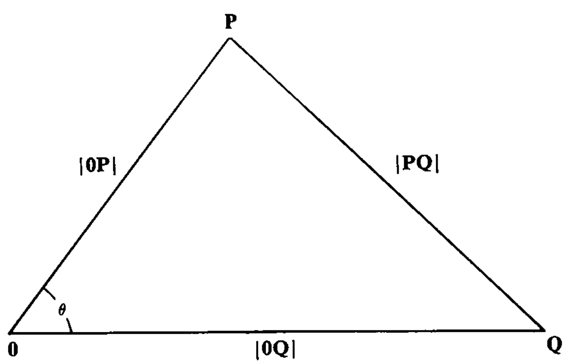


Figure 1.18

More generally, if $\mathbf{p}$ is a point on a plane (not necessarily through the origin) and N is orthogonal to $\prod$, $\prod$ is given by the equation

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{N} \rangle = 0$$

A line through the origin is the linear span of a single vector, and can be expressed by two linear equations (since a line is the intersection of two planes).

Examples

39. Find the equation of the plane through (1, 2, 0) spanned by the two vectors (1, 0, 1), (3, 1, 2). If $\mathbf{N} = (n^1, n^2, n^3)$ perpendicular to this plane we must have

$$\langle \mathbf{N}, (1, 0, 1) \rangle = n^1 + n^3 = 0$$
$$\langle \mathbf{N}, (3, 0, 2) \rangle = 3n^1 + n^2 + 2n^3 = 0$$

A solution of this system is $(1, -1, -1)$, so we may take $\mathbf{N} = (1, -1, -1)$. Then the equation of the plane is

$$\langle \mathbf{x} - (1, 2, 0), (1, -1, -1) \rangle = 0 \quad \text{or} \quad x - y - z + 1 = 0$$

40. Find the equation of the plane through $\mathbf{P} = (1, 0, -1)$, $\mathbf{Q} = (2, 2, 2)$, $\mathbf{R} = (3, 1, 1)$. If $\mathbf{N}$ is perpendicular to the plane, we have

$$\langle \mathbf{N}, \mathbf{P} - \mathbf{R} \rangle = 0 \qquad \langle \mathbf{N}, \mathbf{Q} - \mathbf{R} \rangle = 0$$

Letting $\mathbf{N} = (n^1, n^2, n^3)$, we obtain this system of equations:

$$-2n^1 - n^2 - 2n^3 = 0$$
$$-\ n^1 + n^2 + n^3 = 0$$

which has a solution $\mathbf{N} = (7, 4. 3)$. Thus the equation we seek is $\langle \mathbf{x} - \mathbf{N}, \mathbf{R} \rangle = 0$, or $7(x - 3) + 4(y - 1) + 3(z - 1) = 0$ or

$$7x + 4y + 3z = 28$$

41. Find the equations of the line L through (4, 0, 0) and perpendicular to the plane in Example 40. If x is on L we must have

$$\langle \mathbf{x} - (4, 0, 0), \mathbf{P} - \mathbf{R} \rangle = 0 \qquad \langle \mathbf{x} - (4, 0, 0), \mathbf{Q} - \mathbf{R} \rangle = 0$$

so we may take these as the equations:

$$\begin{aligned} 2x + y + 2z &= 8 \\ -x + y + z &= -4 \end{aligned}$$

Vector Product

Given two noncollinear vectors $\mathbf{v}_1$, $\mathbf{v}_2$ in space, the set of vectors perpendicular to $\mathbf{v}_1$, $\mathbf{v}_2$ is a line. We shall now develop a useful formula for selecting a particular vector on that line, called the vector product $\mathbf{v}_1 \times \mathbf{v}_2$. If $\mathbf{N}$ is on that line, and $\mathbf{x}$ is in the linear span of $\mathbf{v}_1$ and $\mathbf{v}_2$, we have

$$\langle \mathbf{x}, \mathbf{N} \rangle = 0$$

On the other hand, since $\mathbf{x}$, $\mathbf{v}_1$, $\mathbf{v}_2$ are coplanar we have

$$\det\begin{pmatrix} \mathbf{x} \\ \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = 0$$

Now there is a uniquely determined vector $\mathbf{N}$ such that

$$\langle \mathbf{x}, \mathbf{N} \rangle = \det\begin{pmatrix} \mathbf{x} \\ \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}$$

for all $\mathbf{x} \in R^3$. This is easily seen using coordinates. Write

$$\mathbf{v}_1 = ({v_1}^1, {v_1}^2, {v_1}^3), \quad \mathbf{v}_2 = ({v_2}^1, {v_2}^2, {v_2}^3), \ \mathbf{x} = (x^1, x^2, x^3)$$

Then

$$\begin{aligned} \det\begin{pmatrix} \mathbf{x} \\ \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} &= \det\begin{pmatrix} x^1 & x^2 & x^3 \\ {v_1}^1 & {v_1}^2 & {v_1}^3 \\ {v_2}^1 & {v_2}^2 & {v_2}^3 \end{pmatrix} \\ &= x^1({v_1}^2{v_2}^3 - {v_1}^3{v_2}^2) - x^2({v_1}^1{v_2}^3 - {v_1}^3{v_2}^1) \\ &\quad + x^3({v_1}^1{v_2}^2 - {v_1}^2{v_2}^1) \\ &= \langle (x^1, x^2, x^3), (({v_1}^2{v_2}^3 - {v_1}^3{v_2}^2), ({v_1}^3{v_2}^1 - {v_1}^1{v_2}^3), \\ &\quad ({v_1}^1{v_2}^2 - {v_1}^2{v_2}^1)) \rangle \end{aligned}$$

Definition 16. Let $\mathbf{v} = (v^1, v^2, v^3)$, $\mathbf{w} = (w^1, w^2, w^3)$ be two vectors in R^3. The **vector product** $\mathbf{v} \times \mathbf{w}$ is defined by

$$\mathbf{v} \times \mathbf{w} = (v^2w^3 - v^3w^2, v^3w^1 - v^1w^3, v^1w^2 - v^2w^1)$$

Proposition 27.

(i) $\langle \mathbf{x}, \mathbf{v} \times \mathbf{w} \rangle = \det\begin{pmatrix} \mathbf{x} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix}$ *for all* $\mathbf{x} \in R^3$.

(ii) $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$.

(iii) $\mathbf{v} \times \mathbf{w}$ *is orthogonal to* $\mathbf{v}$ *and* $\mathbf{w}$.

(iv) *The equation of the plane through the origin spanned by* $\mathbf{v}$ *and* $\mathbf{w}$ *has the equation* $\langle \mathbf{x}, \mathbf{v} \times \mathbf{w} \rangle = 0$.

The proof of this proposition is completely contained in the preceding discussion. The basic property of the vector product is the first; it follows, for example, that for any three vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$

$$\langle \mathbf{u}, \mathbf{v} \times \mathbf{w} \rangle = \langle \mathbf{u} \times \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \times \mathbf{u} \rangle = \langle \mathbf{v} \times \mathbf{w}, \mathbf{u} \rangle$$

Notice that if $\mathbf{v}$, $\mathbf{w}$ are collinear, $\mathbf{v} \times \mathbf{w} = 0$. If they are not collinear, the ordered basis $\mathbf{u} \to \mathbf{v} \to \mathbf{v} \times \mathbf{w}$ is right handed (see Figure 1.19). The following proposition gives an important geometric interpretation of the magnitude of $\mathbf{v} \times \mathbf{w}$.

Proposition 28. *Let* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ *be three noncollinear vectors.*

(i) *The area of the parallelogram spanned by* $\mathbf{u}$ *and* $\mathbf{v}$ *is* $\|\mathbf{u} \times \mathbf{v}\|$.

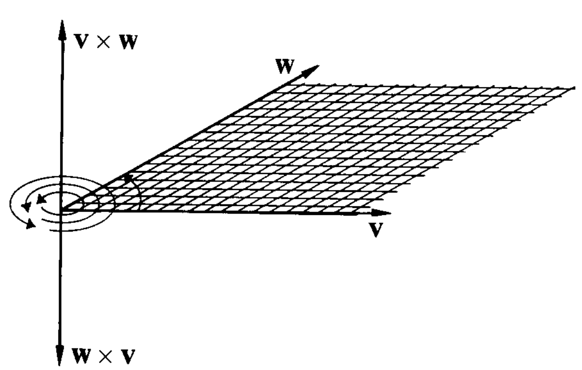


Figure 1.19

(ii) *The volume of the parallelepiped spanned by* **u**, **v**, *and* **w** *is*

$$\left|\det\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{w}\end{pmatrix}\right|$$

Proof.

Step (a). The first step is to verify (ii) in case the vectors **u**, **v**, **w** are mutually perpendicular. In that case we must show that

$$\det\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{w}\end{pmatrix} = \|\mathbf{u}\|\cdot\|\mathbf{v}\|\cdot\|\mathbf{w}\|$$

This follows easily from the multiplicative property of the determinant. First we note that

$$\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{w}\end{pmatrix}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \begin{pmatrix}\langle\mathbf{u}, \mathbf{u}\rangle & 0 & 0\\ 0 & \langle\mathbf{v}, \mathbf{v}\rangle & 0\\ 0 & 0 & \langle\mathbf{w}, \mathbf{w}\rangle\end{pmatrix}$$

since the (i, j)th entry is the inner product of the ith row of the first matrix with the jth row of the second (see Problem 49). Thus,

$$\det\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{w}\end{pmatrix}^2 = \det\begin{pmatrix}\mathbf{u}\\ \mathbf{v}\\ \mathbf{w}\end{pmatrix}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \|\mathbf{u}\|^2\|\mathbf{v}\|^2\|\mathbf{w}\|^2$$

Step (b). In particular, if **u**, **v** are perpendicular, then **u**, **v**, **u** $\times$ **v** are mutually perpendicular, so

$$\|\mathbf{u}\times\mathbf{v}\|^2 = \langle\mathbf{u}\times\mathbf{v}, \mathbf{u}\times\mathbf{v}\rangle = \det\begin{pmatrix}\mathbf{u}\times\mathbf{v}\\ \mathbf{u}\\ \mathbf{v}\end{pmatrix}$$

$$= \|\mathbf{u}\times\mathbf{v}\|\cdot\|\mathbf{u}\|\cdot\|\mathbf{v}\|$$

so $\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\cdot\|\mathbf{v}\|$, when **u** is perpendicular to **v**.

Step (c). Now we prove part (i) in general. Let θ be the the angle between **u** and **v** (see Figure 1.20). Then the area of the parallelogram spanned by **u** and **v** is the product of the base and the height of the base:

$$\text{area} = \|\mathbf{u}\|a = \|\mathbf{u}\|\cdot\|\mathbf{v}\|\sin\theta$$

Now the vector $\mathbf{u} \times (\mathbf{u} \times \mathbf{v})$ is orthogonal to $\mathbf{u}$ and $\mathbf{u} \times \mathbf{v}$, so lies in the plane spanned by $\mathbf{u}$ and $\mathbf{v}$ and is orthogonal to $\mathbf{u}$. We have

$$\sin \theta = \cos\left(\frac{\pi}{2} - \theta\right) = \frac{\langle \mathbf{v}, \mathbf{u} \times (\mathbf{u} \times \mathbf{v})\rangle}{\|\mathbf{v}\| \cdot \|\mathbf{u} \times (\mathbf{u} \times \mathbf{v})\|}$$

Since $\mathbf{u}$ and $\mathbf{u} \times \mathbf{v}$ are orthogonal, by Step (b) we have $\|\mathbf{u} \times (\mathbf{u} \times \mathbf{v})\| = \|\mathbf{u}\| \cdot \|\mathbf{u} \times \mathbf{v}\|$. Thus

$$\begin{aligned}\text{area} &= \|\mathbf{u}\| \cdot \|\mathbf{v}\| \frac{\langle \mathbf{v}, \mathbf{u} \times (\mathbf{u} \times \mathbf{v})\rangle}{\|\mathbf{v}\| \cdot \|\mathbf{u} \times (\mathbf{u} \times \mathbf{v})\|} \\ &= \|\mathbf{u}\| \frac{\langle \mathbf{u} \times \mathbf{v}, \mathbf{u} \times \mathbf{v}\rangle}{\|\mathbf{u}\|\ \|\mathbf{u} \times \mathbf{v}\|} = \|\mathbf{u} \times \mathbf{v}\|\end{aligned}$$

Step (*d*). To prove part (ii) we refer to Figure 1.21. The volume of the parallelepiped spanned by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is the product of the area of the base and the altitude:

$$\text{volume} = \|\mathbf{u} \times \mathbf{v}\| b = \|\mathbf{u} \times \mathbf{v}\| \cdot \|\mathbf{w}\| \sin \phi$$

Since $\mathbf{u} \times \mathbf{v}$ is orthogonal to the $\mathbf{u}$, $\mathbf{v}$ plane,

$$\sin \phi = \cos\left(\frac{\pi}{2} - \phi\right) = \frac{|\langle \mathbf{w}, \mathbf{u} \times \mathbf{v}\rangle|}{\|\mathbf{w}\| \cdot \|\mathbf{u} \times \mathbf{v}\|}$$

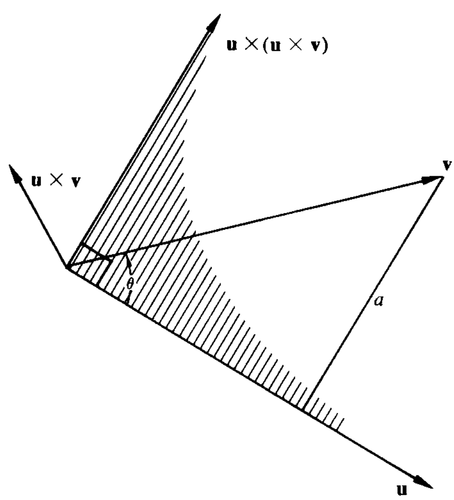


Figure 1.20

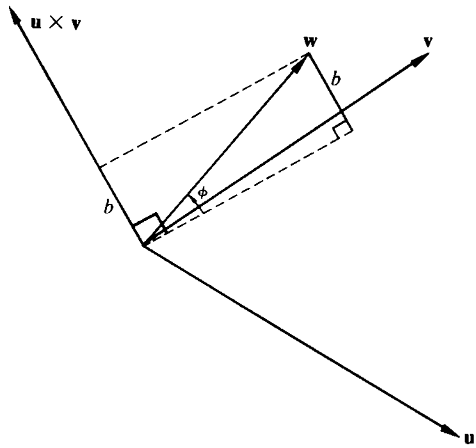


Figure 1.21

Thus

$$\text{volume} = \|\mathbf{u} \times \mathbf{v}\| \cdot \|\mathbf{w}\| \frac{|\langle \mathbf{w}, \mathbf{u} \times \mathbf{v} \rangle|}{\|\mathbf{w}\| \|\mathbf{u} \times \mathbf{v}\|} = \left|\det \begin{pmatrix} \mathbf{w} \\ \mathbf{u} \\ \mathbf{v} \end{pmatrix}\right|$$

A final equality which will prove useful is this:

$$\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \langle \mathbf{u}, \mathbf{v} \rangle^2$$

This follows easily from the above arguments:

$$\begin{aligned} \|\mathbf{u} \times \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2 \phi \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2 \phi) \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \langle \mathbf{u}, \mathbf{v} \rangle^2 \end{aligned}$$

since the angle between $\mathbf{u}$ and $\mathbf{v}$ is ϕ.

• EXERCISES

47. Which pairs of the following vectors are orthogonal?

$\mathbf{v}_1 = (2, 1, 2)$, $\mathbf{v}_2 = (3, -1, 4)$, $\mathbf{v}_3 = (7, 0, 5)$, $\mathbf{v}_4 = (6, -2, 5)$
$\mathbf{v}_5 = (1, 3, 0)$, $\mathbf{v}_6 = (0, 0, 1)$, $\mathbf{v}_7 = (-15, 5, 21)$

48. Find the vector products $\mathbf{v}_i \times \mathbf{v}_j$ for all pairs of vectors given in Exercise 47.

49. Find a vector $\mathbf{v}$ such that

$$\langle \mathbf{v}, \mathbf{v}_1 \rangle = 2 \qquad \langle \mathbf{v}, \mathbf{v}_2 \rangle = -1 \qquad \langle \mathbf{v}, \mathbf{v}_3 \rangle = 7$$

where $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are given in Exercise 47.

50. Find the equation of the plane spanned by the vectors (a) $\mathbf{v}_1, \mathbf{v}_6$, (b) $\mathbf{v}_2, \mathbf{v}_5$, (c) $\mathbf{v}_5, \mathbf{v}_6$, (d) $\mathbf{v}_2, \mathbf{v}_4$, where the $\mathbf{v}_i$ are given in Exercise 47.

51. Find the equation of the line spanned by the vectors given in Exercise 47.

52. Find the equation of the plane through (3, 2, 1) and orthogonal to the vector (−7, 1, 2).

53. Find the equation of the line through (0, 2, 0) and orthogonal to the plane spanned by (1, −1, 1) and (0, 3, 1).

54. Find the equation of the line through the origin and perpendicular to the plane through the points

(a) $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$

(b) (1, 1, 5), (0, 0, 2), (−1, −1, 0)

(c) (0, 0, 0), (0, 0, 1), (0, 1, 0)

55. Find the equation of the line of intersection of the two planes,

(a) determined by (a), (b) of Exercise 54.

(b) determined by (a), (c) of Exercise 54.

(c) determined by (b), (c) of Exercise 54.

56. Find the plane of vectors perpendicular to each of the lines determined in Exercise 55.

57. Let $\mathbf{A}$ be a 3×3 matrix. Show that

(a) if the rows of $\mathbf{A}$ lie on a plane (but not on a line), the set of solutions of $\mathbf{Ax} = \mathbf{b}$ forms a line, or is empty.

(b) if the rows of $\mathbf{A}$ lie on a line, the set of solutions of $\mathbf{Ax} = \mathbf{b}$ forms a plane, or is empty.

58. Show that $\|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\| \cdot \|\mathbf{w}\| \sin \theta$, where θ is the angle between the two vectors $\mathbf{v}$ and $\mathbf{w}$.

59. Is the vector product associative; that is, is

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$$

always true?

60. If $\mathbf{v}$, $\mathbf{w}$ are two noncollinear vectors show that the three vectors $\mathbf{v}$, $\mathbf{v} \times \mathbf{w}$, $\mathbf{v} \times (\mathbf{v} \times \mathbf{w})$ are pairwise orthogonal.

• PROBLEMS

48. Prove the identities of Proposition 27.

49. Let $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ be three vectors in R^3. Let $\mathbf{A}$ be the matrix whose

rows are $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{B}$ the matrix with columns $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$. Show that

(a) the (i, j)th entry of $\mathbf{AB}$ is $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$.

(b) $\det \mathbf{A} = \det \mathbf{B}$.

50. Let $\mathbf{P}$ be given by the coordinates (x, y, z) relative to a choice $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ of basis for space. Show that the point of intersection of the line through $\mathbf{P}$ parallel to the $\mathbf{E}_1$ axis with the 2-3 plane has coordinates $(0, y, z)$.

1.10 Abstract Notions of Linearity

There are many collections of mathematical objects which are endowed with a natural algebraic structure which is very reminiscent of R^n. To be less vague, there is defined, within these collections, the operations of addition and multiplication by real numbers. Furthermore, the problems that naturally arise in these other contexts are reminiscent of the problems on R^n which we have been studying. The question to ask then, is this: does the same theory hold, and will the same techniques work in this more general context? We shall see in this section that for a large class of such objects (the finite-dimensional vector spaces) the theory is the same. We shall see later on that in many other cases, the techniques we have developed can be modified to provide solutions to problems in the more general context. First, let us consider some examples.

Examples

42. If f and g are continuous real-valued functions on the interval $[0, 1]$, then we can define the functions $f + g$, cf as follows:

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) \\ (cf)(x) &= cf(x) \end{aligned}$$

Clearly, $f + g$ and cf are also continuous. Thus we see that operations of addition and scalar multiplication are defined on the collection $C([0, 1])$ of all continuous functions on the interval $[0, 1]$.

43. In the above example, if f and g are differentiable, so are $f + g$ and cf. Thus the space $C^1([0, 1])$ of functions on the interval $[0, 1]$ with continuous derivatives also has the operations of addition and scalar multiplication. Notice that the operation of differentiation takes functions in $C^1([0, 1])$ into $C([0, 1])$: if f is in $C^1([0, 1])$ it has a continuous derivative, so f' is in $C([0, 1])$. Furthermore,

differentiation could be described as a linear transformation:

$$(f+g)' = f' + g'$$
$$(cf)' = cf'$$

So is, by the way, integration a linear transformation:

$$\int(f+g) = \int f + \int g$$
$$\int(cf) = c\int f$$

The fundamental theorem of calculus says that differentiation is the inverse operation for integration:

$$\left(\int f\right)' = f$$

These remarks may strike you as merely a curious way of describing the well-known phenomena, but the implied point of view has led to a wide range of mathematical discoveries. The subject of functional analysis which was developed early in the 20th century came out of this geometric-algebraic approach to long standing problems of analysis.

Examples

44. If S and T are linear transformations of R^n to R^m, then so is the function $S + T$ defined by:

$$(S + T)(x) = S(x) + T(x)$$

We can also multiply a linear transformation by a scalar:

$$(cS)(x) = cS(x)$$

Thus the space $L(R^n, R^m)$ of linear transformations of R^n to R^m has defined on it operations of addition and scalar multiplication.

45. We have already observed (Section 1.6) that the collection M^n of $n \times n$ matrices has defined on it these two important operations. In fact, we used, in an essential way, the fact that when we viewed M^n this way it was just the same as R^{n^2}.

These examples, together with R^n, lead to the notion of an abstract vector space: a set together with the operations of addition and scalar multiplication. We include in the definition the algebraic laws governing these operations.

Definition 17. An **abstract vector space** is a set V with a distinguished element, 0, called the **origin**, on which are defined two operations:

Addition. If v and w are elements of V, then $v + w$ is a well-defined element of V.

Scalar multiplication. If v is in V and c is a real number, cv is a well-defined element of V. These operations must behave in accordance with these laws:

(i) $v + (w + x) = (v + w) + x$,
(ii) $v + w = w + v$,
(iii) $v + 0 = v$,
(iv) $c(v + w) = cv + cw$,
(v) $c_1(c_2 w) = (c_1 c_2)w$,
(vi) $1w = w$.

The preceding examples are all abstract vector spaces; the verifications of the required laws are easily performed. We now want to investigate the extent to which the ideas and facts discussed in the case of R^n carry over to abstract vector spaces. First of all, all the definitions carry over sensibly to the abstract case if we just replace the word R^n by the words **an abstract vector space** V. Thus we take these notions as defined also in the abstract case: **linear transformation, linear subspace, span, independent, basis, dimension.**

Now there is one bit of amplification necessary in the case of **dimension.** We have until now encountered spaces of only finite dimension.

Example

46. Let R^∞ be the collection of all sequences of real numbers. Thus an element of R^∞ is an ordered ∞-tuple,

$$(x^1, x^2, \ldots, x^n, \ldots)$$

R^∞ is an abstract vector space with these operations:

$$(x^1, x^2, \ldots, x^n, \ldots) + (y^1, y^2, \ldots, y^n, \ldots) \\ = (x^1 + y^1, x^2 + y^2, \ldots, x^n + y^n, \ldots)$$
$$c(x^1, x^2, \ldots, x^n, \ldots) = (cx^1, cx^2, \ldots, cx^n, \ldots)$$

Now R^∞ has an infinite set of independent vectors. Let E_n be the sequence all of whose entries are zero but for the nth, which is 1. This entire collection $\{E_1, \ldots, E_n, \ldots\}$ is an independent set. For if there is a relation among some finite subset of these, it must be of the form

$$c^1 E_1 + \cdots + c^k E_k = 0$$

(of course, many of the c's may be zero). But

$$c^1E_1 + \cdots + c^kE_k = (c^1, c^2, \ldots, c^k, 0, 0, \ldots)$$

so if this vector is zero we must have $c^1 = c^2 = \cdots = c^k = 0$. Thus indeed the set $\{E_1, \ldots, E_n, \ldots\}$ is an infinite independent set on R^∞.

We now make the following restriction to the so-called finite-dimensional vector space; and we shall see that all of the preceding information about R^n holds also in this more general case.

Definition 18. A vector space V is **finite dimensional** if there is a finite set of vectors $v_1, \ldots, v_k$ which span V. That R^∞ is not finite dimensional follows from some of the observations to be made below. It can also be verified in the terms of the above definition (see Problem 53). The important result about finite-dimensional vector spaces is that they are no different from the spaces R^n.

Proposition 29. *Let V be a finite-dimensional vector space of dimension d. If $v_1, \ldots, v_d$ is a basis for V, every vector in V can be expressed uniquely as a combination of $v_1, \ldots, v_d$:*

$$v = x^1v_1 + \cdots + x^dv_d$$

$(x^1, \ldots, x^d)$ is called the **coordinate** of v relative to the basis $v_1, \ldots, v_d$. The correspondence $v \to (x^1, \ldots, x^d)$ is a one-to-one linear transformation of V onto R^d.

Proof. The definition of basis (Definition 6) makes this proposition quite clear. We leave the verifications to the reader (Problem 54).

What is not so clear is that every finite-dimensional vector space has a basis, and that every basis has the same number of elements. However, once these facts are established the above proposition serves to reduce the general finite-dimensional space to one of the R^n, and the results of Section 1.3 through 1.6 carry over.

Proposition 30. *Every finite-dimensional vector space V has a finite basis, and every basis has the same number of elements, the dimension of V.*

Proof. Suppose V is finite dimensional. Then V has a finite spanning set. Let $\{v_1, \ldots, v_d\}$ be a spanning set with the minimal number of vectors; by definition V has dimension d. We shall show that $\{v_1, \ldots, v_d\}$ is a basis.

Since $\{v_1, \ldots, v_d\}$ span, every vector in V can be written as a linear combination of these vectors. We have to show that there is only one way in which this can be done. Suppose for some vector v we have two different such ways:

$$v = x^1 v_1 + \cdots + x^d v_d = y^1 v_1 + \cdots + y^d v_d \tag{1.48}$$

Then

$$(x^1 - y^1)v_1 + \cdots + (x^d - y^d)v_d = 0$$

Since these two expressions differ we must have $x^j \neq y^j$ for some j. Thus

$$v_j = \frac{1}{x^j - y^j} \sum_{i \neq j} (x^i - y^i)v_i$$

Now this equation says that v_j is in the linear span of the $d - 1$ elements $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_d$, so these elements serve to span all of V also. But this contradicts the minimal assumption about d. Thus it must be impossible to express v in terms of $v_1, \ldots, v_d$ in two different ways. Hence $\{v_1, \ldots, v_d\}$ is a basis.

That any two bases have the same number of elements follows easily from Proposition 28 (see also Problem 55). Let $T: V \to R^d$ be the linear transformation associating to each vector its coordinate relative to the above basis $\{v_1, \ldots, v_d\}$. If $\{w_1, \ldots, w_\delta\}$ is another basis, let $S: V \to R$ be the same coordinate mapping relative to this basis. Then $L = S \cdot T^{-1}$ is a one-to-one linear mapping of R^d onto R^δ, so $\rho(L) = \delta$, $\nu(L) = 0$. Thus (rank + nullity = dimension): $\delta = d$.

● PROBLEMS

51. Show that for any finite set of vectors $S = \{v_1, \ldots, v_k\}$ in R^∞, there is a vector $w \in R$ which does not lie in their linear span $[S]$. (*Hint*: Let $\mathbf{v}'$ represent the first $(k + 1)$-tuple of entries in v. Since $\mathbf{v}_1', \ldots, \mathbf{v}_k'$ cannot span R^{k+1}, there is a vector $\mathbf{w}'$ in R^{k+1} which cannot be written as a combination of $\mathbf{v}_1', \ldots, \mathbf{v}_k'$. Let $w = (\mathbf{w}', 0, \ldots)$.)

52. Are the vectors $\mathbf{E}_1, \ldots, \mathbf{E}_n, \ldots$ in R described in Example 43 a basis for R^∞?

53. Let R_0^∞ be the collection of those sequences of real numbers $(x^1, x^2, \ldots, x^n, \ldots)$ such that $x^n = 0$ for all but finitely many n. Then R_0^∞ is a linear subspace of R^∞. Show that the vectors $\mathbf{E}_1, \ldots, \mathbf{E}_n, \ldots$ are a basis for R_0^∞.

54. Prove Proposition 29.

55. Prove, by following the arguments in Section 1.4, that any two bases of a finite-dimensional vector space have the same number of elements.

56. Let V, W be two vector spaces. Show that the collection $L(V, W)$ of linear transformations from V to W is a vector space under the two operations:

(a) if $c \in R, L \in L(V, W), (cL)(x) = cL(x)$,
(b) if $L, L' \in L(V, W), (L + L')(x) = L(x) + L'(x)$.

57. What is the dimension of $L(R^n, R^m)$?

58. Show that a vector space V is finite dimensional if there is a *one-to-one* linear transformation of V_0 into R^n for some n.

59. Show that a vector space V is finite dimensional if there is a linear transformation T of R^n *onto* V for some n.

60. Verify that the collection P of polynomials is an abstract vector space. For a positive integer n, let P_n be the collection of polynomials of degree not more than n. Show that P_n is a linear subspace of P. Show that P is not finite dimensional, whereas P_n is. What is the dimension of P_n?

61. Let $x_0, \ldots, x_n$ be distinct real numbers and $c_0, \ldots, c_n$ another collection of real numbers. Show that there is one and only one polynomial p in P_n such that

$$p(x_i) = c_i \qquad 0 \leq i \leq n$$

(*Hint*: Let $L: P_n \to R^{n+1}$ be defined by $L(p) = (p(x_0), \ldots, p(x_n))$. Show that L has rank $n + 1$.)

62. Let g be a polynomial, and define the function $G: P \to P$:

$$G(p) = pg$$

Show that G is a linear function. Describe the range and kernel of G.

63. Define $D_k: P \to P$: $D_k(p) = d^kp/dx^k$. What are the range and kernel of D_k?

64. Let $x_0 \in R$, and let $c_0, \ldots, c_k$ be given numbers. Show that there is one and only one polynomial p in P_n such that

$$p(x_0) = d_0 \qquad \frac{dp}{dt}(x_0) = c_1, \ldots, \frac{d^kp}{dx^k}(x_0) = c_k$$

(*Hint:* Use the same idea as in Exercise 61.)

65. Does $D_k: P \to P$ have any eigenvalues?

66. Show that $C([0, 1])$ is not a finite-dimensional vector space.

1.11 Inner Products

The notion of *length*, or *distance*, is important in the geometric study of planar and spatial configurations. In Section 1.3 we studied these concepts and related them to an algebraic concept, the inner product. From the point of view of analysis also it is true that these concepts are significant:

it is in terms of distance that we can express "closeness" and in particular "convergence." By analogy with R^3 we define the inner product in R^n, and in terms of it, distance. While we are here we shall, in this section, introduce some topological terms.

Definition 19. The **inner product** of two vectors $\mathbf{v} = (v^1, \ldots, v^n)$, $\mathbf{w} = (w^1, \ldots, w^n)$, denoted by $\langle \mathbf{v}, \mathbf{w} \rangle$ is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum v^i w^i$$

We shall say that $\mathbf{v}$ **is orthogonal to** $\mathbf{w}$ if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. The **distance** $d(\mathbf{v}, \mathbf{w})$ **between** $\mathbf{v}$ **and** $\mathbf{w}$ is defined by

$$d(\mathbf{v}, \mathbf{w}) = [\sum (v^i - w^i)^2]^{1/2}$$

The **modulus** $|\mathbf{v}|$ of a vector $\mathbf{v}$ is the distance between $\mathbf{v}$ and $\mathbf{0}$,

$$|\mathbf{v}| = d(\mathbf{v}, \mathbf{0}) = [\sum (v^i)^2]^{1/2}$$

Distance in R^n behaves much as it does in R^2 and R^3; in particular, the Pythagorean theorem holds:

$$d(\mathbf{v}, \mathbf{w})^2 = d(\mathbf{v}, \mathbf{x})^2 + d(\mathbf{x}, \mathbf{w})^2 \tag{1.49}$$

when $\langle \mathbf{v} - \mathbf{w}, \mathbf{w} - \mathbf{x} \rangle = 0$. In any event, two points are no further apart than the sum of the distances from a third,

$$d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{x}) + d(\mathbf{x}, \mathbf{w}) \tag{1.50}$$

These facts will be verified in the problems.

Topological Notions

Definition 20. The **ball in R^n of radius $R > 0$ and center c**, denoted $B(\mathbf{c}, R)$, is the set of all points whose distance from $\mathbf{c}$ is less than R:

$$B(\mathbf{c}, R) = \{\mathbf{x} \in R^n : d(\mathbf{x}, \mathbf{c}) < R\}$$

A set S is said to be a **neighborhood** of a point $\mathbf{c}$ if it contains some ball centered at $\mathbf{c}$. A set U is said to be **open** if it contains a neighborhood of each of its points.

Thus, a set S is a neighborhood of $\mathbf{c}$ if there is *some* R (presumably very small) such that

$$d(\mathbf{x}, \mathbf{c}) < R \text{ implies } \mathbf{x} \in S$$

A set U is open if for every $\mathbf{c} \in U$, there is an R such that $U \supset B(\mathbf{c}, R)$. Notice that any ball is open. For suppose $\mathbf{x} \in B(\mathbf{c}, R)$. Then $d(\mathbf{x}, \mathbf{c}) < R$, so $R - d(\mathbf{x}, \mathbf{c}) > 0$. Now $B(\mathbf{c}, R)$ contains the ball of radius $R - d(\mathbf{x}, \mathbf{c})$ centered at $\mathbf{x}$. For if $\mathbf{y}$ is a point in that ball, then by (1.50),

$$d(\mathbf{y}, \mathbf{c}) \leq d(\mathbf{y}, \mathbf{x}) + d(\mathbf{x}, \mathbf{c}) < R - d(\mathbf{x}, \mathbf{c}) + d(\mathbf{y}, \mathbf{c}) = R$$

Here is a collection of formal properties of the collection of open sets.

Proposition 31.
(i) *R^n is open.*
(ii) *If $U_1, \ldots, U_n$ are open, so is $U_1 \cap \cdots \cap U_n$.*
(iii) *If C is any collection of open sets, then the set of all points belonging to any of the sets in C is open. (This set is denoted $\bigcup U$).*

Proof.
(i) Clearly, R^n contains a ball centered at every one of its points.
(ii) Suppose $U_1, \ldots, U_n$ are open, and $\mathbf{x}$ is in every U_i. Then there are $R_1, \ldots, R_n$ such that $U_1 \supset B(\mathbf{x}, R_1), \ldots, U_n \supset B(\mathbf{x}, R_n)$. Let $R = \min[R_1, \ldots, R_n]$. Then if $d(\mathbf{y}, \mathbf{x}) < R$, $\mathbf{y}$ is in each $B(\mathbf{x}, R_i)$ so is in each U_i. Thus $\mathbf{y}$ is in $U_1 \cap \cdots \cap U_n$. In particular, $U_1 \cap \cdots \cap U_n \supset B(\mathbf{x}, R)$. Thus $U_1 \cap \cdots \cap U_n$ is a neighborhood of any one of its points $\mathbf{x}$, and is thus open.
(iii) Suppose C is a collection of open sets. If $\mathbf{x}$ is in any one of them, say U, then since U is open there is an R such that $U \supset B(\mathbf{x}, R)$. Thus, $\bigcup_{U \in C} U \supset B(\mathbf{x}, R)$. Thus $\bigcup_{U \in C} U$ is a neighborhood of any one of its points, so is open.

Many of the concepts a mathematician studies are so-called **local** concepts: They happen in a neighborhood of a point, or are determined by what goes on near a point; far behavior being irrelevant. Differentiation is thus local, whereas integration is not. The importance of open sets is that it is precisely on such sets that we should study these local concepts, since their definition at a point depends on behavior in some neighborhood of the point.

If a set is open its complement, the set of all points not in the given set, is said to be **closed**. Thus, S is a closed subset of R^n if $R^n - S = \{x \in R^n : \mathbf{x} \notin S\}$ is open. Corresponding to Proposition 31 we have this proposition about closed sets.

Proposition 32.
(i) *R^n is closed.*
(ii) *If $S_1, \ldots, S_n$ are closed, so is $S_1 \cup \cdots \cup S_n$.*

(iii) *If C is a collection of closed sets, then the set of all points common to all the sets of C is closed.* (*This set is denoted* $\bigcap_{S \in C} S$).

Proof. Problem 67.

Notice that there are sets which are both open and closed. There are not many of them. R^n and $\varnothing$ are the only ones. There are also sets which are neither open nor closed, and there are many of them. For example, an interval is open in R^1 if it contains neither end point, closed if it contains both, and neither open nor closed if it contains only one end point.

We are acquainted with the notion of "dropping a perpendicular" in the plane. That is, if l is a line and $\mathbf{p}$ is a point not on the line, then we can drop a perpendicular from $\mathbf{p}$ to l as in Figure 1.22. The point $\mathbf{p}_0$ of intersection of the perpendicular with l is the point on l which is closest to $\mathbf{p}$. A more sophisticated way of describing this situation is to say that $\mathbf{p}_0$ is the **orthogonal projection of p on** l. The concept of orthogonal projection generalizes to R^n and will prove quite useful there. In order to discuss this problem, we shall generalize even further.

Definition 21. A **Euclidean vector space** is an abstract vector space V on which is defined a real-valued function of pairs of vectors, called the **inner product**, and denoted $\langle\,,\,\rangle$. The inner product must obey these laws:

(i) $\langle v, v\rangle \geq 0$. If $\langle v, v\rangle = 0$, then $v = 0$.
(ii) $\langle v, w\rangle = \langle w, v\rangle$.
(iii) $\langle av, w\rangle = a\langle v, w\rangle$.
(iv) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$.

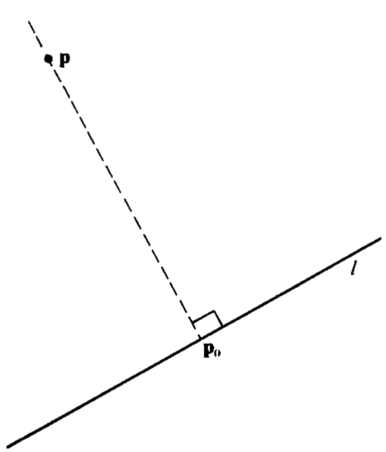


Figure 1.22

It is clear that R^n is a Euclidean vector space when endowed with its inner product. The space $C[0, 1]$ of continuous functions on the unit interval is a Euclidean vector space with this inner product:

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\, dt$$

We leave it to the reader to verify that the laws (i)–(iv) are obeyed. It is interesting that the laws (i)–(iv) are all that is essential to the notion of inner product; that is, any such function behaving in accordance with those laws will have all the properties of an inner product. Despite the inherent interest in this "metamathematical" point, we shall not pursue it further, but take it for granted that the above definition has indeed abstracted the essence of this notion.

In terms of an inner product on a vector space we can define the notions of **length** and **orthogonality**:

$$\|v\| = [\langle v, v\rangle]^{1/2}$$

$$v \perp w \quad \text{if and only if} \quad \langle v, w\rangle = 0$$

The important bases in a Euclidean vector space are those bases whose vectors are mutually orthogonal. More specifically, we shall call a set $\{E_1, \ldots, E_n\}$ in a Euclidean vector space V an **orthonormal set** if

$$\|E_i\| = 1 \qquad \text{for all } i$$

$$E_i \perp E_j \qquad \text{for all } i \neq j$$

If the vectors $E_1, \ldots, E_n$ span V we shall call them an orthonormal basis. (Any orthonormal set of vectors is independent—Problem 68.) The basic geometric fact concerning orthonormal sets is the following:

Proposition 33. *Let V be a Euclidean vector space and $\{E_1, \ldots, E_n\}$ an orthonormal set in V. For any vector v in V, the vector $v_0 = v - \sum_{i=1}^n \langle v, E_i\rangle E_i$ is orthogonal to the linear span S of $\{E_1, \ldots, E_n\}$.*

Proof. Let $w = \sum_{i=1}^n c_i E_i$ be in S. Then

$$\langle v, w\rangle = \langle v - \sum_{i=1}^n \langle v, E_i\rangle \langle E_i, w\rangle = \langle v, w\rangle - \sum_{i=1}^n \langle v, E_i\rangle \langle E_i, w\rangle$$

Now

$$\langle E_i, w\rangle = \langle E_i, \sum_{j=1}^{n} c_j E_j\rangle = \sum_{j=1}^{n} c_j \langle E_i, E_j\rangle = c_i$$

$$\langle v, w\rangle = \langle v, \sum_{i=1}^{n} c_i E_i\rangle = \sum_{i=1}^{n} c_i \langle v, E_i\rangle$$

Thus

$$\langle v, w\rangle = \sum_{i=1}^{n} c_i \langle v, E_i\rangle - \sum_{i=1}^{n} \langle v, E_i\rangle c_i = 0$$

Theorem 1.8. *Let V be a Euclidean vector space, and let $\{E_1, \ldots, E_n\}$ be an orthonormal set in V. For any vector v, let*

$$v_0 = \sum_{i=1}^{n} \langle v, E_i\rangle E_i$$

Then

(i) $\|v\|^2 = \|v - v_0\|^2 + \|v_0\|^2$;
(ii) for any w in the linear span of $\{E_1, \ldots, E_n\}$,

$$\|v - v_0\|^2 \leq \|v - w\|^2$$

Proof.

(i)
$$\begin{aligned} \|v\|^2 &= \langle v, v\rangle = \langle (v - v_0) + v_0, (v - v_0) + v_0\rangle \\ &= \|v - v_0\|^2 + \|v_0\|^2 + \langle v_0, v - v_0\rangle + \langle v - v_0, v_0\rangle \end{aligned}$$

The last two terms are zero by the preceding proposition, since v_0 is in the linear span of $\{E_1, \ldots, E_n\}$.

(ii)
$$\begin{aligned} \|v - w\|^2 &= \langle v - w, v - w\rangle \\ &= \langle v - v_0 + v_0 - w, v - v_0 + v_0 - w\rangle \\ &= \|v - v_0\|^2 + \|v_0 - w\|^2 + \langle v - v_0, v_0 - w\rangle + \langle v_0 - w, v - v_0\rangle \end{aligned}$$

Again, the last two terms are zero for both v_0, w and thus also $v_0 - w$ is in the linear span of $\{E_1, \ldots, E_n\}$. Thus

$$\|v - w\|^2 = \|v - v_0\|^2 + \|v_0 - w\|^2 \geq \|v - v_0\|^2$$

so (ii) is proven.

Gram-Schmidt Process

Notice that $v - v_0 = v'$ is orthogonal to the linear span S of $\{E_1, \ldots, E_n\}$. v_0 is the vector in S which is closest to v; it is called the orthogonal projection of v into S. It seems, by Theorem 1.8 that one needs an orthonormal basis in order to find orthogonal projections; the following proposition gives a procedure for obtaining orthonormal basis for finite-dimensional vector spaces, and thus with it, orthogonal projections.

Proposition 34. *Let $F_1, \ldots, F_n$ be a basis for a Euclidean vector space V. We can find an orthonormal basis $E_1, \ldots, E_n$ so that the linear span of $E_1, \ldots, E_j$ is the same as the linear span of $F_1, \ldots, F_j$ for all j.*

Proof. The proof is by induction on n. If $n = 1$, we need only take $E_1 = \|F_1\|^{-1}F_1$.

Now in general, let $F_1, \ldots, F_n$ be a basis for a Euclidean vector space V. Then the linear span W of $F_1, \ldots, F_{n-1}$ is a Euclidean vector space also, and we can apply the proposition to W by the inductive hypothesis. Let $E_1, \ldots, E_{n-1}$ be an orthonormal basis with the required properties. Now, we must find a vector E_n such that

$$\|E_n\| = 1$$
$$(E_n, E_i) = 0 \qquad \text{all } i \neq n$$
$$F_n \text{ is in the linear span of } E_1, \ldots, E_n$$

If E_n is a vector that fulfills the last two conditions, then we can take $E_n = \|E_n\|^{-1}E_n$. Thus we need only find a vector filling the last two conditions. That is easy; take

$$E_n = F_n - \sum_{j<n} (F_n, E_j)E_j$$

Then, for $i < n$,

$$\begin{aligned}(E_n, E_i) &= (F_n, E_i) - \sum_{j<n} (F_n, E_j)(E_j, E_i)\\ &= (F_n, E_i) - (F_n, E_i)(E_i, E_i) = 0\end{aligned}$$

Furthermore,

$$F_n = E_n + \sum_{j<n} (F_n, E_j)E_j$$

so the last two conditions are fulfilled and the proposition is proven.

The proof of this proposition provides a procedure for finding orthonormal bases in an Euclidean vector space, known as the **Gram–Schmidt** process. It goes like this:

First, pick any basis $F_1, \ldots, F_n$ of V. Take

$$E_1 = \|F_1\|^{-1}F_1$$

Then choose $E_2 = F_2 - (F_2, E_1)E_1$, and divide by the length to find E_2, and so forth. If $E_1, \ldots, E_j$ are found, take

$$E_{j+1}^0 = F_{j+1} - (F_{j+1}, E_1)E_1 - (F_{j+1}, E_2)E_2 - \cdots - (F_{j+1}, E_j)E_j$$

and let E_{j+1} be the vector of length one collinear with E_{j+1}^0.

Examples

47. Apply the Gram–Schmidt process to this basis of R^3:

$$\mathbf{F}_1 = (1, 0, 1)$$
$$\mathbf{F}_2 = (3, -1, 2)$$
$$\mathbf{F}_3 = (0, 0, 1)$$

Take

$$\mathbf{E}_1 = \frac{\mathbf{F}_1}{\|\mathbf{F}_1\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

$$\mathbf{E}_2 = (3, -1, 2) - \left(3 \cdot \frac{1}{\sqrt{2}} + (-1) \cdot 0 + 2 \cdot \frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

$$= (3, -1, 2) - \left(\frac{5}{2}, 0, \frac{5}{2}\right) = \left(\frac{1}{2}, -1, \frac{-1}{2}\right)$$

Then

$$\mathbf{E}_2 = \left(\frac{2}{7}\right)^{1/2}\left(\frac{1}{2}, -1, \frac{-1}{2}\right) = \left(\frac{1}{(14)^{1/2}}, -\left(\frac{2}{7}\right)^{1/2}, \frac{-1}{(14)^{1/2}}\right)$$

$$\mathbf{E}_3{}^0 = (0, 0, 1) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) + \frac{3}{(14)^{1/2}}\left(\frac{1}{(14)^{1/2}},\right.$$

$$\left. -\left(\frac{2}{7}\right)^{1/2}, \frac{-1}{(14)^{1/2}}\right)$$

and finally

$$\mathbf{E}_3 = \left(\frac{-2}{(17)^{1/2}}, \frac{-3}{(17)^{1/2}}, \frac{2}{(17)^{1/2}}\right)$$

48. Find an orthonormal basis for the kernel of $\lambda: R^4 \to R$, $\lambda(x^1, x^2, x^3, x^4) = x^1 + x^2 + x^3 + 2x^4$.

First of all, let us pick a suitable basis for $K(\lambda)$; that is, $(1, 0, 0, -1/2)$, $(0, 1, 0, -1/2)$, $(0, 0, 1, -1/2)$. Applying the Gram–Schmidt process, we obtain

$$\mathbf{E}_1 = \frac{2}{\sqrt{5}}\left(1, 0, 0, \frac{-1}{2}\right)$$

$$\mathbf{E}_2{}^0 = \left(0, 1, 0, \frac{-1}{2}\right) - \frac{1}{2\sqrt{5}} \cdot \frac{2}{\sqrt{5}}\left(1, 0, 0, \frac{-1}{2}\right)$$

$$\mathbf{E}_2 = \left(\frac{-1}{(30)^{1/2}}, \left(\frac{5}{6}\right)^{1/2}, 0, \frac{-2}{(30)^{1/2}}\right)$$

$$\mathbf{E}_3{}^0 = \left(0, 0, 1, \frac{-1}{2}\right) - \frac{2}{\sqrt{5}} \cdot \frac{2}{\sqrt{5}}\left(1, 0, 0, \frac{-1}{2}\right)$$

$$- \frac{1}{(30)^{1/2}}\left(\frac{-1}{(30)^{1/2}}, \left(\frac{5}{6}\right)^{1/2}, 0, \frac{-2}{(30)^{1/2}}\right)$$

$$\mathbf{E}_3 = \frac{(1094)^{1/2}}{30}\left(\frac{1}{6}, \frac{1}{6}, 1, \frac{2}{5}\right)$$

49. Find the orthogonal projection of $(3, 1, 2)$ into the kernel of $T: R^3 \to R$:

$$T(x, y, z) = x + 2y + z$$

Now the kernel of T is spanned by $\mathbf{F}_1 = (2, -1, 0)$, $\mathbf{F}_2 = (0, -1, 2)$. Applying the Gram–Schmidt process, we obtain the orthonormal basis

$$\mathbf{E}_1 = (\tfrac{1}{5})^{1/2}(2, -1, 0) \qquad \mathbf{E}_2 = (\tfrac{6}{5})^{1/2}(-\tfrac{1}{5}, -\tfrac{2}{5}, 1)$$

Thus the orthogonal projection of $(3, 1, 2)$ into this plane is

$$(\tfrac{1}{5})^{1/2}5(\tfrac{1}{5})^{1/2}(2, -1, 0) + (\tfrac{6}{5})^{1/2}1(\tfrac{6}{5})^{1/2}(-\tfrac{1}{5}, -\tfrac{2}{5}, 1) = (\tfrac{44}{25}, -\tfrac{37}{25}, \tfrac{6}{5})$$

50. Find the point on the line

$$L: \begin{aligned} x + \ y - z &= 0 \\ 3y + z &= 0 \end{aligned}$$

which is closest to $(7, 1, 0)$. L is the linear span of the vector $(-4, -1, 3)$. Thus the orthogonal projection of $(7, 1, 0)$ on this line (the closest point) is

$$\left\langle (7, 1, 0), \frac{(-4, -1, 3)}{(26)^{1/2}} \right\rangle \frac{(-4, -1, 3)}{(26)^{1/2}} = \frac{27}{26}(4, -1, 3)$$

• EXERCISES

61. Which of the following sets are open; closed; or neither.
    (a) $\{x \in R: 2 < |x - 5| < 13\}$.
    (b) $\{x \in R: 0 < x \leq 4\}$.
    (c) $\{x \in R: x \geq 32\}$.
    (d) $\{\mathbf{x} \in R^n: \langle \mathbf{x}, \mathbf{x} \rangle = 4\}$.
    (e) $\{\mathbf{x} \in R^3: \langle \mathbf{x}, (0, 2, 1) \rangle = 0\}$.
    (f) $\{\mathbf{x} \in R^3: 2 < \|\mathbf{x} - (3, 0, 3)\| < 14\}$.
    (g) $\{\mathbf{x} \in R^n: x^1 > 0, \ldots, x^n > 0\}$.
    (h) The set of integers (considered as a subset of R).
    (i) $\{\mathbf{x} \in R^n: \sum_{i=1}^{n} x^i a^i < \varepsilon\}$.
    (j) $\{\mathbf{x} \in R^n: \sum_{i=1}^{n} x^i a^i \neq 1\}$.
    (k) $\{\mathbf{x} \in R^n: \sum (x^i)^3 < \sum (x^i)^2\}$.
62. Find the point on the plane

$$x + 3y + 2z = 4$$

closest to the point $(1, 0, 1)$.

63. Find the point on the line

$$\begin{aligned} x + 7y + z &= 2 \\ x \qquad - z &= 0 \end{aligned}$$

closest to the point $(-7, 1, 0)$.

64. Find an orthonormal basis for the linear span of
    (a) $\mathbf{v}_1 = (0, 2, 2)$, $\mathbf{v}_2 = (1, 0, 2)$, $\mathbf{v}_3 = (1, 2, 4)$.
    (b) $\mathbf{v}_1 = (0, 1, 0, 1)$, $\mathbf{v}_2 = (1, 0, 1, 0)$, $\mathbf{v}_3 = (1, 1, 2, 3)$.
    (c) $\mathbf{v}_1 = (0, 3, 0, 0, 0)$, $\mathbf{v}_2 = (0, 6, 0, 3, 0)$, $\mathbf{v}_3 = (0, 0, 2, -1, 1)$.
    (d) $\mathbf{v}_1 = (1, 2, 3, 4)$, $\mathbf{v}_2 = (4, 3, 2, 1)$, $\mathbf{v}_3 = (2, 1, 4, 3)$.

65. Find orthonormal bases for the linear span and kernel of these transformations on R^4:

(a) $\begin{pmatrix} 8 & 6 & 1 & 0 \\ 1 & 2 & 0 & 2 \\ 0 & 3 & 3 & 0 \\ 7 & 4 & 1 & -2 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ -1 & 1 & -1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

• PROBLEMS

67. Prove Proposition 32.

68. Show that an orthogonal set of vectors is independent.

69. Give an example of a sequence $\{U_n\}$ of open sets such that $\bigcap_{n=1}^{\infty} U_n$ is not open.

70. Give an example of a sequence $\{C_n\}$ of closed sets such that $\bigcup_{n=1}^{\infty} C_n$ is open.

71. Find an orthonormal basis for the linear span of 1, x, x^2, x^3 in the vector space $C([0, 1])$ with the inner product $\langle f, g\rangle = \int fg$.

In the next four problems V represents a vector space endowed with an inner product, denoted $\langle\ ,\ \rangle$.

72. Let v, w, x be three points in V such that $v - x$ is orthogonal to $w - x$. Show that the Pythagorean theorem is valid:

$$\|v - w\|^2 = \|v - x\|^2 + \|x - w\|^2$$

73. Let v, w be two vectors in V. Show that the vector in the linear span of w which is closest to v is

$$v_0 = \frac{\langle v, w\rangle}{\|w\|^2} w \tag{1.51}$$

(You can verify this by minimizing the function $f(t) = \|v - tw\|^2$ by calculus.)

74. Prove **Schwarz's inequality:**

$$|\langle v, w\rangle| \leq \|v\| \cdot \|w\|$$

for any two vectors in V. (*Hint*: $\|v - v_0\|^2 \geq 0$ where v_0 is given by (1.50).)

75. Prove the **triangle inequality:**

$$\|v - x\| \leq \|v - w\| + \|w - x\|$$

for any three vectors in V (use Schwarz's inequality).

76. Let V be a vector space with an inner product. Suppose that W is a subspace of V. Let $\perp(W) = \{v: \langle v, w\rangle = 0 \text{ for all } w \in W\}$. This is called the orthogonal complement of W. Show that $\perp(W)$ is a linear subspace of V and (if V is finite dimensional) that W and $\perp(W)$ together span V.

77. Let T: $R^n \to R^m$ be a linear transformation represented by the matrix **A**. Show that the rows of **A** span $\perp(K(T))$.

78. Show that a linear transformation is one-to-one on the orthogonal complement of its kernel.

• FURTHER READING

R. E. Johnson, *Linear Algebra*, Prindle, Weber & Schmidt, Boston, 1968. This book covers the same material and includes a derivation of the Jordan canonical form.

K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1961. This book is more thorough and abstract, and has a full discussion of canonical forms.

L. Fox, *An Introduction to Numerical Linear Algebra*, Oxford University Press, 1965. This is a detailed treatment of computational problems in matrix theory.

H. K. Nickerson, D. C. Spencer, and N. Steenrod, *Advanced Calculus*, Van Nostrand, Princeton, N.J., 1957. This set of notes has a full treatment of all the abstract linear algebra required in modern analysis.

• MISCELLANEOUS PROBLEMS

79. Show that if $\mathbf{A}'$ is obtained from $\mathbf{A}$ by a sequence of row operations then these equations have the same solutions: $\mathbf{Ax} = \mathbf{0}$, $\mathbf{A}'\mathbf{x} = \mathbf{0}$.

80. Show that every nonempty set of positive integers has a least element.

81. Show that a set with n elements has precisely 2^n subsets.

82. Show that the n-fold Cartesian product of a set with k elements has k^n elements.

83. Can you interpret the case $k = 2$ in Problem 82 so as to deduce the assertion of Exercise 3?

84. Let $\mathbf{A} = (a_j{}^i)$ be an $n \times n$ matrix such that $a_j{}^i = 0$ if $i - j > r$ for some $r \geq 0$. Show that $\mathbf{A}^{n-r} = \mathbf{0}$. Show that the same conclusion follows from the assumption $j - i > r$ for some $r \geq 0$. Will the hypothesis $|i - j| > r$ do as well?

85. Let T: $R^n \to R^m$ be a linear transformation of rank r. Show that there are linear transformations S_1: $R^m \to R^{m-r}$, S_2: $R^{n-r} \to R^n$ such that

(a) S_1 has rank $m - r$ and $\mathbf{b} \in R(T)$ if and only if $S_1\mathbf{b} = \mathbf{0}$.

(b) S_2 has rank $n - r$ and $\mathbf{x} \in K(T)$ if and only if $\mathbf{x} \in R(S_2)$.

86. Suppose that T: $R^n \to R^n$ and $T^k = I$. Show that T is invertible.

87. Let S be a subset of R^n. Show that the linear span $[S]$ of S is the intersection of all linear subspaces of R^n containing S.

88. Let S, T be subsets of R^n. Show that

$$\dim([S \cup T]) \leq \dim([S]) + \dim([T]),$$

and equality holds if and only if $[S] \cap [T] = \{0\}$.

89. Let V and W be subspaces of R^n. Let X be the set of all sums $\mathbf{v} + \mathbf{w}$ with $\mathbf{v} \in V$, $\mathbf{w} \in W$. Show that X is a linear subspace of R^n. The relationship between X and V and W is indicated by writing $X = V + W$. If in addition $V \cap W = \{0\}$, then every $\mathbf{x} \in X$ can be written in the form $\mathbf{v} + \mathbf{w}$ in only one way. In this case, $X = V + W$ with $V \cap W = 0$, we say that X is the **direct sum of** V **and** W and write $X = V \oplus W$.

90. Suppose $X = V \oplus W$. Then $\dim X = \dim V + \dim W$.

91. Show that if λ: $R^n \to R$ is a linear function, there exists a $\mathbf{w} \in R^n$ such that $\lambda(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{v} \in R^n$.

92. If S is a subset of R^n define

$$\perp(S) = \{\mathbf{v} \in R^n: \langle \mathbf{v}, \mathbf{s} \rangle = 0 \text{ for all } \mathbf{s} \in S\}.$$

(a) Show that $\perp(S)$ is a subspace of R^n and that $S \cap \perp(S) = \{0\}$.
(b) Show that $[S] = \perp(\perp(S))$.
(c) If V is a linear subspace of R^n, $R^n = V \oplus \perp(V)$.

93. Suppose that T: $V \to W$ is a linear transformation and V is not finite dimensional. Show that either the rank or the nullity of T must be infinite.

94. Let V be an abstract vector space. **A bilinear function** p on V is a function of two variables in V with these properties:

$$p(cv, w) = cp(v, w) \qquad p(v, cw) = cp(v, w)$$
$$p(v_1 + v_2, w) = p(v_1, w) + p(v_2, w) \qquad p(v, w_1 + w_2) = p(v, w_1) + p(v, w_2)$$

Show that the sum of two bilinear functions is bilinear. In fact, the space B_V of all bilinear functions is an abstract vector space. If V is finite dimensional, what is the dimension of B_V? (*Hint:* See the next problem.)

95. Let p be a bilinear function on R^n. Let

$$a_i;_j = p(\mathbf{E}_i, \mathbf{E}_j)$$

Show that p is completely determined by the matrix $(a_i;_j)$.

96. Let V be an abstract vector space.

(a) Show that the space V^* of linear functions on V is a vector space under addition and scalar multiplication.

(b) If $\dim V = d$, show that $\dim V^* = d$ also.

(c) Show that to every $\lambda \in R^{n*}$ there is a $\mathbf{w} \in R^n$ such that $\lambda(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $v \in R^n$. (Recall Problem 91.)

97. Suppose that V is a linear subspace of W. We define the **annihilator** of V, denoted $\operatorname{ann}(V)$, to be the set of $\lambda \in W^*$ such that $\lambda(v) = 0$ if $v \in V$.

Show that ann(V) is a linear subspace of W^*. If dim $W = n$, dim $V = d$, show that ann(V) has dimension $n - d$.

98. Let V be a linear subspace of R^n, and suppose that T: $V \to R^m$ is a linear transformation. Show that there is a linear transformation T': $R^n \to R^m$ defined on all of R^n which extends T.

99. The **closure** of a set S, denoted $\bar{S}$, is the set of all points **x** such that every neighborhood of **x** contains points of S. Find the closure of all the sets in Problem 61.

100. Show that the closure of a set S is the smallest closed set containing S.

101. The **boundary** of a set S, denoted ∂S, is the set of all points **x** such that every neighborhood of **x** contains points of both S and the complement of S. Find the boundary of all the sets in Problem 61.

102. Show that the boundary of a set is a closed set.

103. Show that the boundary of a set S is also the boundary of its complement $R^n - S$. In fact, show that $\partial S = \bar{S} \cap \overline{(R^n - S)}$.

104. Let T: $V \to W$ be a linear transformation of a vector space with an inner product. The **adjoint** of T is the transformation T^*: $W \to V$ defined in this way

$$\langle T^*(w), v\rangle = \langle w, Tv\rangle \qquad \text{for all } v \in V$$

(a) Show that T^* is a well-defined linear transformation.

(b) If T: $R^n \to R^m$ is represented by the matrix $\mathbf{A} = (a_j{}^i)$, then T^*: $R^m \to R^n$ is represented by the matrix $\mathbf{A}^* = (a^*{}_i^j)$, where $a^*{}_i^j = a_i{}^j$. (This matrix is called the **adjoint** or **transpose** of **A**.)

(c) Show that $R(T^*)$ is complementary to $K(T)$.

(d) In fact, $\rho(T^*) = \nu(T)$, $\nu(T^*) = \rho(T)$.

105. A bilinear form p on a vector space V is called **symmetric** if it obeys the law: $p(v, w) = p(w, v)$ for all v and w. An inner product is a symmetric bilinear form and much of the formal manipulations with inner products remains valid for symmetric bilinear forms. For example, the Gram-Schmidt process (Proposition 32) gives rise to this fact (see if you can work the proof of Proposition 32 to give it):

Proposition. *Let p be a symmetric bilinear form on V. Suppose $F_1, \ldots, F_n$ is a basis for V. We can find another basis, $E_1, \ldots, E_n$ of V such that the linear span of $E_1, \ldots, E_j$ is the same as that of $F_1, \ldots, F_j$ for all j, and $p(E_i, E_j) = 0$ if $i \neq j$.*

We shall call such a basis $E_1, \ldots, E_n$ **p-orthogonal.**

106. Let p be a symmetric bilinear form on a vector space V, and suppose $E_1, \ldots, E_n$ is a p-orthogonal basis.

(a) Show that $p(v, w)$ can be computed in terms of this basis as follows: if $\mathbf{v} = \sum v^i E_i$, $w = \sum w^i E_i$, then

$$p(v, w) = \sum_{i=1}^{n} v^i w^i p(E_i, E_i) \tag{1.52}$$

(b) Show that p is an inner product on the linear span of the E_i such that $p(E_i, E_i) > 0$.

(c) Similarly, $-p$ is an inner product on the linear span of the E_i such that $p(E_i, E_i) < 0$.

107. Prove this fact: Let p be a symmetric bilinear form on a finite-dimensional vector space V. There is a basis $E_1, \ldots, E_n$, integers r, s such that $r + s \leq n$ and such that if $v = \sum v^i E_i$, then

$$p(v, v) = \sum_{i \leq r} (v^i)^2 - \sum_{r \leq i \leq r+s} (v^i)^2 \tag{1.53}$$

(*Hint:* Modify the basis $\{E_i\}$ in Problem 106 so that (1.52) becomes (1.53).)

108. The integers r, s of Problem 107 are determined by p alone, and are independent of the basis. Here is a sketch of how a proof would go. Suppose $F_1, \ldots, F_n$ is another p-orthogonal basis and ρ is the number of F_i's such that $p(F_i, F_i) > 0$. We have to show $\rho = r$. Let W be the linear span of these F's. Expressing points of W in terms of the basis $E_1, \ldots, E_n$ we may consider the transformation T: $W \to R^n$ given by

$$T(\sum v^i E_i) = (v^1, \ldots, v^r)$$

T is one-to-one on W, for if $w \in W$, and $w \neq 0$,

$$0 < p(w, w) = \sum_{i \leq r} (v^i)^2 - \sum_{r \leq i \leq r+s} (v^i)^2$$

so we must have

$$\sum_{i \leq r} (v^i)^2 > 0$$

on W. Since T is one-to-one, it follows that $r \geq \rho$. The inequality $\rho \geq r$ follows from the same argument with the roles of $E_1, \ldots, E_n$ and $F_1, \ldots, F_n$ interchanged.

109. Let $\mathbf{A} = (a_{i;j})$ be a symmetric $n \times n$ matrix, that is, $a_{i;j} = a_{j;i}$ Then $\mathbf{A}$ determines a symmetric bilinear form on R^n as follows:

$$p_A(\mathbf{v}, \mathbf{w}) = \sum_{i,j} a_{i;j} v^i w^j$$

If $\mathbf{P}$ is the matrix corresponding to the change of basis from the standard basis to that described in Problem 105, then $\mathbf{P}^*\mathbf{A}\mathbf{P}$ is diagonal. Verify that assertion.

110. Find the p-orthogonal basis and the representation (1.53) of Problem 107 for the symmetric bilinear forms given by these matrices:

(a) $\begin{pmatrix} 4 & 3 & 0 & 1 \\ 3 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & 0 \\ -1 & 0 & 0 & -1 \end{pmatrix}$

111. Describe the sets $p(\mathbf{v}, \mathbf{v}) > 0$, $=0$, <0 in R^4 where p is given by

$$p(\mathbf{v}, \mathbf{v}) = (v^1)^2 + (v^2)^2 + (v^3)^2 - (v^4)^2$$

112. A transformation $T\colon V \to V$ is called **self-adjoint** if it is self-adjoint ($\langle T\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, T\mathbf{w}\rangle$ for all $\mathbf{v}$, $\mathbf{w} \in V$). Show that if T is a self-adjoint transformation on R^n, then

$$R^n = K(T) \oplus R(T)$$

113. Suppose that $\mathbf{v}$, $\mathbf{w}$ are eigenvectors of a self-adjoint transformation T on V with different eigenvalues. Show that $\langle \mathbf{v}, \mathbf{w}\rangle = 0$.

114. If T is a self-adjoint transformation on R^n, and $\mathbf{v}_0 \in R^n$ is such that $\sum (v_0{}^i)^2 = 1$ and

$$\langle T\mathbf{v}_0, \mathbf{v}_0\rangle = \max\{\langle T\mathbf{v}, \mathbf{v}\rangle ; \sum (v^i)^2 = 1\}$$

then $\mathbf{v}_0$ is an eigenvector for T.

115. Use Problems 113 and 114 to prove the Spectral theorem for self-adjoint operators on R^n:

Theorem. *There is an orthonormal basis* $\mathbf{E}_1, \ldots, \mathbf{E}_n$ *of eigenvectors of* T. *T can be computed in terms of this basis by*

$$T(\sum x^i \mathbf{E}_i) = \sum x^i c_i \mathbf{E}_i$$

116. Find a basis of eigenvectors in R^4 for the self-adjoint transformations given by the matrices (a), (b) of Problem 110.

117. Orthonormalize these bases of R^4:

(a) $(1, 0, 0, 0), (0, 1, 1, 1), (0, 0, 2, 2), (3, 0, 0, 3)$.

(b) $(-1, -1, -1, -1), (0, -1, -1, -1), (0, 0, -1, -1), (0, 0, 0, -1)$.

(c) $(0, 1, 0, 1), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0)$.

118. Find the orthogonal projection of R^5 onto these spaces:

(a) The span of $(0, 1, 0, 0, 1)$.

(b) The span of $(1, 1, 0, 0, 0)$, $(1, 0, 1, 0, 0)$.

(c) The span of $(1, 0, 0, 0, 1)$, $(0, 1, 0, 0, 1)$, $(0, 0, 1, 0, 1)$.

(d) The span of the vectors given in (c) and the vector $(0, 0, 0, 1, 1)$.