

Larry Smith

# Linear Algebra

Second Edition

With 21 Figures



Springer-Verlag  
New York Berlin Heidelberg Tokyo

Larry Smith  
Mathematisches Institut  
Universität Göttingen  
D3400 Göttingen  
West Germany

*Editorial Board*

F. W. Gehring  
Department of Mathematics  
University of Michigan  
Ann Arbor, MI 48109  
U.S.A.

P. R. Halmos  
Department of Mathematics  
Indiana University  
Bloomington, IN 47405  
U.S.A.

---

AMS Subject Classification: 15-01

---

Library of Congress Cataloging in Publication Data  
Smith, Larry.

Linear algebra.

(Undergraduate texts in mathematics)

Includes index.

1. Algebras, Linear. I. Title. II. Series.

QA184.S63 1984 512'.5 84-5419

© 1978, 1984 by Springer-Verlag New York Inc.

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

Typeset by Composition House Ltd., Salisbury, England.

Printed and bound by R. R. Donnelley & Sons, Harrisonburg, Virginia.

Printed in the United States of America.

9 8 7 6 5 4 3 2 (Second printing, 1985)

ISBN 0-387-96015-5 Springer-Verlag New York Berlin Heidelberg Tokyo  
ISBN 3-540-96015-5 Springer-Verlag Berlin Heidelberg New York Tokyo

This text is written for a course in linear algebra at the (U.S.) sophomore undergraduate level, preferably directly following a one-variable calculus course, so that linear algebra can be used in a course on multidimensional calculus. Realizing that students at this level have had little contact with complex numbers or abstract mathematics, the book deals almost exclusively with real finite-dimensional vector spaces in a setting and formulation that permits easy generalization to abstract vector spaces. The parallel complex theory is developed in the exercises.

The book has as a goal the principal axis theorem for real symmetric transformations, and a more or less direct path is followed. As a consequence there are many subjects that are not developed, and this is intentional.

However, a wide selection of examples of vector spaces and linear transformations is developed, in the hope that they will serve as a testing ground for the theory. The book is meant as an *introduction* to linear algebra and the theory developed contains the essentials for this goal. Students with a need to learn more linear algebra can do so in a course in abstract algebra, which is the appropriate setting. Through this book they will be taken on an excursion to the algebraic/analytic zoo, and introduced to some of the animals for the first time. Further excursions can teach them more about the curious habits of some of these remarkable creatures.

For the second edition of the book I have added, amongst other things, a safari into the wilderness of canonical forms, where the hardy student can pursue the Jordan form with the tools developed in the preceding chapters.

Göttingen,  
June 1984

LARRY SMITH





1	Vectors in the plane and space	1
2	Vector spaces	13
3	Subspaces	21
4	Examples of vector spaces	27
5	Linear independence and dependence	35
6	Bases and finite-dimensional vector spaces	42
7	The elements of vector spaces: a summing up	56
8	Linear transformations	65
9	Linear transformations: some numerical examples	90
10	Matrices and linear transformations	104
11	Matrices	112
12	Representing linear transformations by matrices	130
12 <sup>bis</sup>	More on representing linear transformations by matrices	153
13	Systems of linear equations	165
14	The elements of eigenvalue and eigenvector theory	191
14 <sup>bis</sup>	Multilinear algebra: determinants	225
15	Inner product spaces	244
16	The spectral theorem and quadratic forms	274
17	Jordan canonical form	303
18	Applications to linear differential equations	334
	Index	357
	List of notations	361

In physics certain quantities such as *force*, *displacement*, *velocity*, and *acceleration* possess both a magnitude and a direction and they are most usually represented geometrically by drawing an arrow with the magnitude and direction of the quantity in question. Physicists refer to the arrow as a *vector*, and call the quantities so represented *vector quantities*. In the study of the calculus the student has no doubt encountered vectors, and their algebra, particularly in connection with the study of lines and planes and the differential geometry of space curves. Vectors can be described as *ordered pairs* of points ( $P$ ,  $Q$ ) which we call the *vector from  $P$  to  $Q$*  and often denote by  $\overrightarrow{PQ}$ . This is substantially the same as the physics definition, since all it amounts to is a technical description of the word "arrow."  $P$  is called the *initial point* and  $Q$  the *terminal point*.

For our purposes it will be convenient to regard two vectors as being equal if they have the same length and the same magnitude. In other words we will regard  $\overrightarrow{PQ}$  and  $\overrightarrow{RS}$  as determining the *same vector* if  $\overrightarrow{RS}$  results by moving  $\overrightarrow{PQ}$  parallel to itself.

(*N.B.* Vectors that conform to this definition are called *free vectors*, since we are "free to pick" their initial point. Not all "vectors" that occur in nature conform to this convention. If the vector quantity depends not only on its direction and magnitude but its initial point it is called a *bound vector*. For example, torque is a bound vector. In the force-vector diagram represented by Figure 1.1  $\overrightarrow{PQ}$  does not have the same effect as  $\overrightarrow{RS}$  in pivoting a bar. In this book we will consider only free vectors.)

With this convention of equality of vectors in mind it is clear that if we fix a point  $O$  in space called the *origin*, then we may regard all our vectors as having their initial point at  $O$ . The vector  $\overrightarrow{OP}$  will very often be abbreviated to  $\vec{P}$ , if the point  $O$  which serves as the origin of all vectors is clear from



Figure 1.1

context. The vector  $\vec{P}$  is called the *position vector* of the point  $P$  relative to the origin  $O$ .

In physics vector quantities such as force vectors are often added together to obtain a resultant force vector. This process may be described as follows. Suppose an origin  $O$  has been fixed. Given vectors  $\vec{P}$  and  $\vec{Q}$  their sum is defined by the Figure 1.2. That is, draw the parallelogram determined by the three points  $P$ ,  $O$  and  $Q$ . Let  $R$  be the fourth vertex and set  $\vec{P} + \vec{Q} = \vec{R}$ .

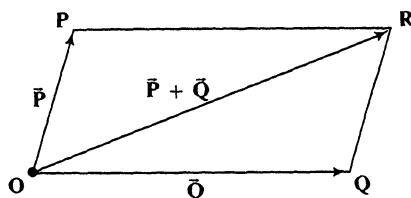


Figure 1.2

The following basic rules of vector algebra may be easily verified by elementary Euclidean geometry.

- (1)  $\vec{P} + \vec{Q} = \vec{Q} + \vec{P}$ .
- (2)  $(\vec{P} + \vec{Q}) + \vec{R} = \vec{P} + (\vec{Q} + \vec{R})$ .
- (3)  $\vec{P} + \vec{O} = \vec{P} = \vec{O} + \vec{P}$ .

It is also possible to define the operation of multiplying a vector by a number. Suppose we are given a vector  $\vec{P}$  and a number  $a$ . If  $a > 0$  we let  $a\vec{P}$  be the vector with the same direction as  $\vec{P}$  only  $a$  times as long (see Figure 1.3). If  $a < 0$  we set  $a\vec{P}$  equal to the vector of magnitude  $a$  times the magnitude of  $\vec{P}$  but having direction *opposite* of  $\vec{P}$  (see Figure 1.4). If  $a = 0$  we set  $a\vec{P}$

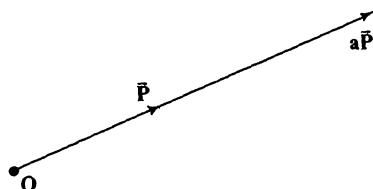


Figure 1.3

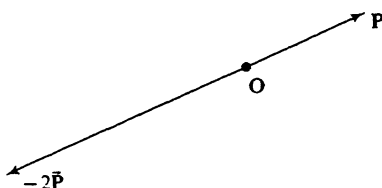


Figure 1.4

equal to  $\vec{0}$ . It is then easy to show that vector algebra satisfies the following additional rules:

- (4)  $\vec{P} + (-1\vec{P}) = \vec{0}$
- (5)  $a(\vec{P} + \vec{Q}) = a\vec{P} + a\vec{Q}$
- (6)  $(a + b)\vec{P} = a\vec{P} + b\vec{P}$
- (7)  $(ab)\vec{P} = a(b\vec{P})$
- (8)  $0\vec{P} = \vec{0}, 1\vec{P} = \vec{P}$

Note that Rule 6 involves two types of addition, namely addition of numbers and addition of vectors.

Vectors are particularly useful in studying lines and planes in space. Suppose that an origin  $O$  has been fixed and  $L$  is the line through the two points  $P$  and  $Q$  as in Figure 1.5. Suppose that  $R$  is any other point on  $L$ .

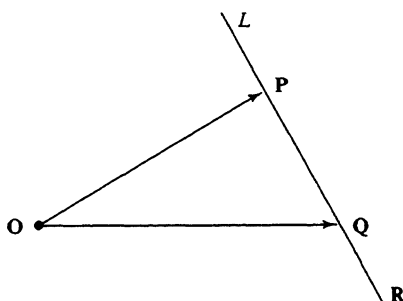


Figure 1.5

Consider the position vector  $\vec{R}$ . Since the two points  $P, Q$  completely determine the line  $L$ , it is quite reasonable to look for some relation between the vectors  $\vec{P}, \vec{Q}$ , and  $\vec{R}$ . One such relation is provided by Figure 1.6. Observe that

$$\vec{S} + \vec{P} = \vec{Q}$$

Therefore if we write  $-\vec{P}$  for  $(-1)\vec{P}$  we see that

$$\vec{S} = \vec{Q} - \vec{P}.$$

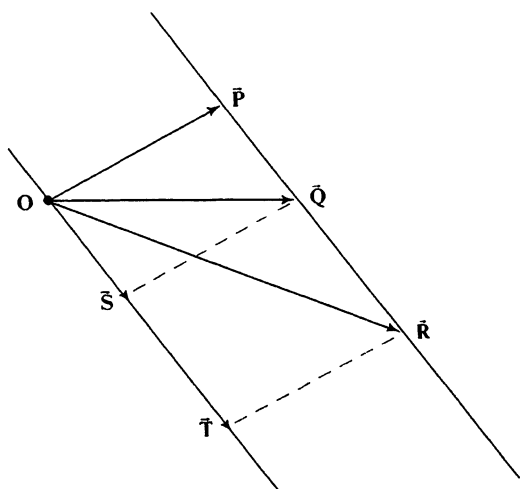


Figure 1.6

Notice that there is a number  $t$  such that

$$\vec{T} = t\vec{S}.$$

Moreover

$$\vec{R} = \vec{P} + \vec{T}$$

and hence we find

$$(*) \quad \vec{R} = \vec{P} + t(\vec{Q} - \vec{P}).$$

Equation (\*) is called the *vector equation* of the line  $L$ . To make practical computations with this equation it is convenient to introduce in addition to the origin  $O$  a cartesian coordinate system as in Figure 1.7. Every point  $P$  then has coordinates  $(x, y, z)$ , and if we have two points  $P$  and  $Q$  with coordinates  $(x_P, y_P, z_P)$  and  $(x_Q, y_Q, z_Q)$  then it is quite easy to check that

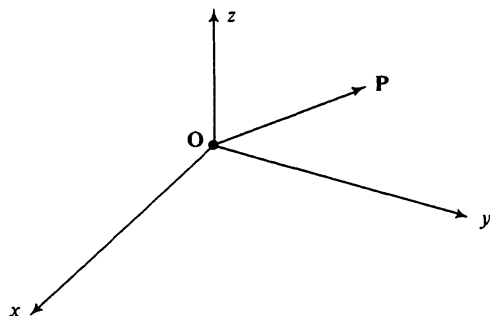


Figure 1.7

$\vec{P} + \vec{Q}$  is the position vector of the point with components  $(x_P + x_Q, y_P + y_Q, z_P + z_Q)$ . Likewise for a number  $a$  the vector  $a\vec{P}$  is the position vector of the point with coordinates  $(ax_P, ay_P, az_P)$ . Thus we find by considering the coordinates of the points represented Equation (\*) that  $(x, y, z)$  lies on the line  $L$  through  $P, Q$  iff

$$\begin{aligned}
 (**) \quad x &= x_P + t(x_Q - x_P), \\
 y &= y_P + t(y_Q - y_P), \\
 z &= z_P + t(z_Q - z_P).
 \end{aligned}$$

**EXAMPLE 1.** Does the point  $(1, 2, 3)$  lie on the line passing through the points  $(4, 4, 4)$  and  $(1, 0, 1)$ ?

*Solution.* Let  $L$  be the line through  $P = (4, 4, 4)$  and  $Q = (1, 0, 1)$ . Then the points of  $L$  must satisfy the equations

$$\begin{aligned}
 x &= 4 + t(1 - 4) = 4 - 3t, \\
 y &= 4 + t(0 - 4) = 4 - 4t, \\
 z &= 4 + t(1 - 4) = 4 - 3t,
 \end{aligned}$$

where  $t$  is a number. Let us check if this is possible:

$$\begin{aligned}
 1 &= 4 - 3t, \\
 2 &= 4 - 4t, \\
 3 &= 4 - 3t.
 \end{aligned}$$

The first equation gives

$$-3 = -3t \quad t = 1.$$

Putting this in the last equation gives

$$3 = 4 - 3 = 1$$

which is impossible. Therefore  $(1, 2, 3)$  does not lie on the line through  $(4, 4, 4)$  and  $(1, 0, 1)$ .

**EXAMPLE 2.** Let  $L_1$  be the line through the points  $(1, 0, 1)$  and  $(1, 1, 1)$ . Let  $L_2$  be the line through the points  $(0, 1, 0)$  and  $(1, 2, 1)$ . Determine if the lines  $L_1$  and  $L_2$  intersect. If so find their point of intersection.

*Solution.* The equations of  $L_1$  are

$$\begin{aligned}
 x &= 1 + t_1(1 - 1) = 1, \\
 y &= 0 + t_1(1 - 0) = t_1, \\
 z &= 1 + t_1(1 - 1) = 1.
 \end{aligned}$$

The equations of  $L_2$  are

$$\begin{aligned}
 x &= 0 + (1 - 0)t_2 = t_2, \\
 y &= 1 + (2 - 1)t_2 = 1 + t_2, \\
 z &= 0 + (1 - 0)t_2 = t_2.
 \end{aligned}$$

If a point lies on both of these lines we must have

$$\begin{aligned}1 &= t_2, \\ t_1 &= 1 + t_2, \\ 1 &= t_2.\end{aligned}$$

Therefore  $t_2 = 1$  and  $t_1 = 2$ . Hence  $(1, 2, 1)$  is the only point these lines have in common.

**EXAMPLE 3.** Determine if the lines  $L_1$  and  $L_2$  with equations

$$\begin{aligned}L_1 \quad x &= 1 - 3t, \\ y &= 1 + 3t, \\ z &= t, \\ L_2 \quad x &= -2 - 3t, \\ y &= 4 + 3t, \\ z &= 1 + t,\end{aligned}$$

have a point in common.

*Solution.* If a point  $(x, y, z)$  lies on both lines it must satisfy both sets of equations, so there is a number  $t_1$  such that

$$\begin{aligned}x &= 1 - 3t_1, \\ y &= 1 + 3t_1, \\ z &= t_1,\end{aligned}$$

and a number  $t_2$  with

$$\begin{aligned}x &= -2 - 3t_2, \\ y &= 4 + 3t_2, \\ z &= 1 + t_2,\end{aligned}$$

and the answer to the problem is reduced to determining if in fact two such numbers can be found, that is if the simultaneous equations

$$\begin{aligned}1 - 3t_1 &= -2 - 3t_2, \\ 1 + 3t_1 &= 4 + 3t_2, \\ t_1 &= 1 + t_2,\end{aligned} \tag{*}$$

have any solutions. Writing these equations in the more usual form they become

$$\begin{aligned}3 &= 3t_1 - 3t_2, \\ -3 &= -3t_1 + 3t_2, \\ -1 &= -t_1 + t_2.\end{aligned}$$

By dividing the first equation by 3, the second by  $-3$ , and multiplying the third by  $-1$  we get

$$1 = t_1 - t_2,$$

$$1 = t_1 - t_2,$$

$$1 = t_1 - t_2,$$

giving

$$t_1 = 1 + t_2.$$

What does this mean? It means that no matter what value of  $t_2$  we choose there is a value of  $t_1$ , namely  $t_1 = 1 + t_2$ , which satisfies Equations (\*). By varying the values of  $t_2$  we get all the points on the line  $L_2$ . For each such value of  $t_2$  the fact that there is a (corresponding) value of  $t_1$  solving Equations (\*) shows that every point of the line  $L_2$  lies on the line  $L_1$ . Therefore these lines must be the same!

The lesson to be learned from this example is that the equations of a line are not unique. This should be geometrically clear since we only used two points of the line to determine the equations, and there are many such possible pairs of points.

**EXAMPLE 4.** Determine if the lines  $L_1$  and  $L_2$  with equations

$$x = 1 + t,$$

$$L_1 \quad y = 1 + t,$$

$$z = 1 - t,$$

$$x = 2 + t,$$

$$L_2 \quad y = 2 - t,$$

$$z = 2 - t,$$

have a point in common.

**Solution.** As in Example 3 our task is to determine if the simultaneous equations

$$(*) \quad \begin{aligned} 1 + t_1 &= 2 + t_2, \\ 1 + t_1 &= 2 - t_2, \\ 1 - t_1 &= 2 - t_2, \end{aligned}$$

has any solutions. In more usual form these equations become

$$-1 = -t_1 + t_2,$$

$$-1 = -t_1 - t_2,$$

$$-1 = t_1 - t_2.$$



Adding the first two equations gives

$$-2 = -2t_1,$$

so  $t_1$  must equal 1. Putting this into the last equation we get

$$-1 = 1 - t_2,$$

so  $t_2$  must equal 2. But substituting these values of  $t_1$  and  $t_2$  into either of the first two equations leads to a contradiction, namely

$$-1 = -1 + 2 = 1,$$

$$-1 = -1 - 2 = -3,$$

therefore no values of  $t_1$  and  $t_2$  can simultaneously satisfy Equations (\*) so the lines have no point in common.

In Chapter 13 we will take up the study of solving simultaneous linear equations in detail. There we will explain various techniques and "tests" that will make the problems encountered in Examples 3 and 4 routine.

Suppose now that  $P$ ,  $Q$ , and  $R$  are three noncolinear points. Then they determine a unique plane  $\Pi$ . If we introduce a fixed origin  $O$  then it is possible to deduce an equation that is satisfied by the position vectors of points of  $\Pi$ . Considering Figure 1.8 shows that

$$\vec{A} - \vec{Q} = s(\vec{P} - \vec{Q}) + t(\vec{R} - \vec{Q})$$

that is

$$(*) \quad \vec{A} = s(\vec{P} - \vec{Q}) + t(\vec{R} - \vec{Q}) + \vec{Q}.$$

Equation (\*) is called the *vector equation* of the plane  $\Pi$ . Compare it to the vector equation of a line. Note the presence of the two parameters  $s$  and  $t$  instead of the single parameter  $t$

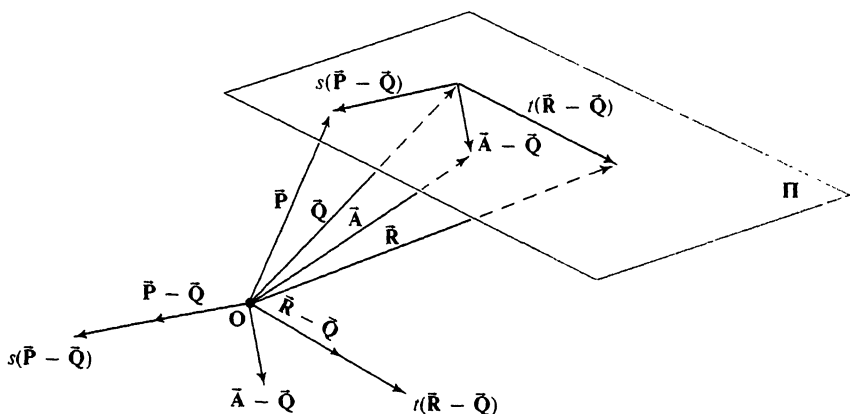


Figure 1.8

If we now introduce a coordinate system and pass to components in Equation (\*) we obtain:

$$\begin{aligned}
 x &= s(x_P - x_Q) + t(x_R - x_Q) + x_Q, \\
 y &= s(y_P - y_Q) + t(y_R - y_Q) + y_Q, \\
 z &= s(z_P - z_Q) + t(z_R - z_Q) + z_Q.
 \end{aligned}$$

(\*\*)

We may regard Equation (\*\*) as the equation of the plane  $\Pi$  or we may regard it as a system of three equations in the two unknowns  $s, t$  which we may formally eliminate and obtain the more familiar equation

$$ax + by + cz + d = 0$$

(\*\*)

where we may take (or twice these values, or  $-7$  times, etc.)

$$\begin{aligned}
 a &= (y_R - y_Q)(z_P - z_Q) - (z_R - z_Q)(y_P - y_Q), \\
 b &= (z_R - z_Q)(x_P - x_Q) - (x_R - x_Q)(z_P - z_Q), \\
 c &= (x_R - x_Q)(y_P - y_Q) - (y_R - y_Q)(x_P - x_Q), \\
 d &= -(ax_P + by_P + cz_P).
 \end{aligned}$$

Equation (\*\*) is also called the equation of the plane  $\Pi$ .

EXAMPLE 5. Find the equation of the plane through the points

$$(1, 0, 1), \quad (0, 1, 0), \quad (1, 1, 1).$$

Determine if the point  $(0, 0, 0)$  lies in this plane.

*Solution.* We know that the equation has the form

$$ax + by + cz + d = 0$$

and all we must do is crank out values for  $a, b, c, d$ . (Remember they are not unique.) We must have

$$\begin{aligned}
 a + c + d &= 0, \\
 b + d &= 0, \\
 a + b + c + d &= 0,
 \end{aligned}$$

since the points  $(1, 0, 1)$ ,  $(0, 1, 0)$ , and  $(1, 1, 1)$  lie in this plane. Thus

$$a + c = 0, \quad d = 0, \quad b = 0, \quad a = -c.$$

So the plane has the equation

$$x - z = 0$$

and  $(0, 0, 0)$  lies in it.

EXAMPLE 6. Determine the equation of the line of intersection of the planes

$$\begin{aligned}
 x - z &= 0, \\
 x + y + z + 1 &= 0.
 \end{aligned}$$

## 1: Vectors in the plane and space

*Solution.* The line in question has an equation of the form

$$x = a + ut,$$

$$y = b + vt,$$

$$z = c + wt,$$

for suitable numbers  $a, b, c, u, v, w$ . Since such points must lie in both planes we have

$$a + ut - (c + wt) = 0,$$

$$a + ut + b + vt + c + wt + 1 = 0,$$

for all values of  $t$ . Put  $t = 0$ . Then

$$a - c = 0,$$

$$a + b + c + 1 = 0.$$

The first equation yields  $a = c$ . Combining this with the second equation and setting  $b = 1$  yields  $2a + 2 = 0$ . Hence  $a = -1 = c$ . Next put  $t = 1$ . Then

$$0 = a + ut - (c + wt) = -1 + u + 1 - w,$$

$$0 = a + ut + b + vt + c + wt + 1$$

$$= -1 + u + 1 + v - 1 + w + 1.$$

The first equation yields  $u = w$ . Combining this with the second equation and setting  $u = 1$  yields  $w = u = 1$  and  $v = -2$ . Then

$$x = -1 + t,$$

$$y = 1 - 2t,$$

$$z = -1 + t,$$

are the equations of a line containing the two points  $(-1, 1, -1)$  and  $(0, -1, 0)$  which lie in both planes and hence must be the line of intersection.

## EXERCISES

1. Suppose that an origin  $\mathbf{O}$  and a coordinate system have been fixed. Let  $\mathbf{P}$  be a point. Define vectors  $\bar{\mathbf{E}}_1$ ,  $\bar{\mathbf{E}}_2$ , and  $\bar{\mathbf{E}}_3$  by requiring that they be the position vectors of the points  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(0, 0, 1)$ , respectively. Let the coordinates of  $\mathbf{P}$  be  $(x_{\mathbf{P}}, y_{\mathbf{P}}, z_{\mathbf{P}})$ . Show that

$$\bar{\mathbf{P}} = x_{\mathbf{P}}\bar{\mathbf{E}}_1 + y_{\mathbf{P}}\bar{\mathbf{E}}_2 + z_{\mathbf{P}}\bar{\mathbf{E}}_3$$

The vectors

$$x_{\mathbf{P}}\bar{\mathbf{E}}_1, \quad y_{\mathbf{P}}\bar{\mathbf{E}}_2, \quad z_{\mathbf{P}}\bar{\mathbf{E}}_3$$

are called the *component vectors* of  $\bar{\mathbf{P}}$  relative to the given coordinate system.

2. Find the equation of the line through the two points  $(1, 0, -1)$ ,  $(2, 3, -1)$ . Does the point  $(0, 1, -1)$  lie on this line?
3. Does the point  $(1, 1, 1)$  lie in the plane through the points  $(1, 1, 0)$ ,  $(0, 1, 1)$ ,  $(1, 0, 1)$ ?
4. Does the line through the points  $(1, 1, 1)$ ,  $(1, -1, 1)$  lie in the plane through the points  $(1, -1, 0)$ ,  $(1, 0, -1)$ ,  $(-1, 1, 1)$ ?
5. Show that the point  $(1, -2, 1)$  lies on the line through the two points  $(0, 1, -1)$  and  $(2, -5, 3)$ .
6. Let  $\mathbf{P} = (x_1, y_1, z_1)$ ,  $\mathbf{Q} = (x_2, y_2, z_2)$  be two points. Show that the midpoint of the line segment  $\overline{\mathbf{PQ}}$  is

$$\left( \frac{x_1 + x_2}{2}, \quad \frac{y_1 + y_2}{2}, \quad \frac{z_1 + z_2}{2} \right).$$

7. Find the equation of the line through the origin bisecting the angle formed by  $\mathbf{A}$   $\mathbf{O}$   $\mathbf{B}$ , where  $\mathbf{A} = (1, 0, 0)$ ,  $\mathbf{B} = (0, 0, 1)$ .
8. Verify that vectors  $\overline{\mathbf{PQ}}$  and  $\overline{\mathbf{RS}}$  represent the same vector  $\overline{\mathbf{T}}$  where  $\mathbf{P} = (0, 1, 1)$ ,  $\mathbf{Q} = (1, 3, 4)$ ,  $\mathbf{R} = (1, 0, -1)$ ,  $\mathbf{S} = (2, 2, 2)$ . Find the coordinates of  $\mathbf{T}$ .
9. Find the sum of the vectors  $\overline{\mathbf{PQ}}$  and  $\overline{\mathbf{RS}}$  where  $\mathbf{P} = (0, 1, 1)$ ,  $\mathbf{Q} = (1, 0, 0)$ ,  $\mathbf{R} = (1, 0, -1)$ ,  $\mathbf{S} = (2, 2, 2)$ .
10. Let  $\mathbf{P} = (1, 1)$ ,  $\mathbf{Q} = (2, 3)$ ,  $\mathbf{R} = (-2, 3)$ ,  $\mathbf{S} = (1, -1)$  Find  $\overline{\mathbf{PQ}} - \overline{\mathbf{RS}}$ ,  $\overline{\mathbf{PQ}} + \overline{\mathbf{RS}}$ .
11. Show that the points  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ ,  $\mathbf{D}$  with the following coordinates form a parallelogram in a plane:  $\mathbf{A} = (1, 1)$ ,  $\mathbf{B} = (3, 2)$ ,  $\mathbf{C} = (2, 3)$ ,  $\mathbf{D} = (0, 2)$ .
12. Let  $\mathbf{P} = (1, 0, 1)$ ,  $\mathbf{Q} = (1, 1, 1)$ , and  $\mathbf{R} = (-1, 1, -1)$ . Find the coordinates of  $\overline{\mathbf{T}}$  where
  - (a)  $\overline{\mathbf{T}} = 2\overline{\mathbf{P}} - \overline{\mathbf{Q}}$
  - (b)  $\overline{\mathbf{T}} = \overline{\mathbf{PQ}}$
  - (c)  $\overline{\mathbf{T}} = 2\overline{\mathbf{R}}$
  - (d)  $\overline{\mathbf{T}} = -\overline{\mathbf{R}}$
  - (e)  $\overline{\mathbf{T}} = \overline{\mathbf{PQ}} + \overline{\mathbf{PR}}$
  - (f)  $\overline{\mathbf{T}} = a\overline{\mathbf{P}} + b\overline{\mathbf{Q}} + c\overline{\mathbf{R}}$ , where  $a, b, c$  are given constants.

13. In each of (a)–(g) find a vector equation of the line satisfying following conditions:

- (a) passing through the point  $\mathbf{P} = (-2, 1)$  and having slope  $\frac{1}{2}$
- (b) passing through the point  $(0, 3)$  and parallel to the  $x$ -axis
- (c) the tangent line to  $y = x^2$  at  $(2, 4)$
- (d) the line parallel to the line of (c) passing through the origin
- (e) the line passing through points  $(1, 0, 1)$  and  $(1, 1, 1)$
- (f) the line passing through the origin and the midpoint of the line segment  $\overline{\mathbf{PQ}}$  where  $\mathbf{P} = (\frac{1}{2}, \frac{1}{2}, 0)$ ,  $\mathbf{Q} = (0, 0, 1)$
- (g) the line on  $xy$ -plane passing through  $(1, 1, 0)$  and  $(0, 1, 0)$ .

**14.** In each of (a)–(g) determine a vector equation of the plane satisfying the given conditions:

- (a) the plane determined by  $(0, 0)$ ,  $(1, 0)$ , and  $(1, 1)$
- (b) the plane determined by  $(0, 0, 1)$ ,  $(1, 0, 1)$ , and  $(1, 1, 1)$
- (c) The plane determined by  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(1, 1, 1)$  (Does the origin lie on this plane?)
- (d) the plane parallel to the  $xy$ -plane and containing the point  $(1, 1, 1)$
- (e) the plane through the origin and containing the points  $\mathbf{P} = (1, 0, 0)$ ,  $\mathbf{Q} = (0, 1, 0)$
- (f) the plane through three points  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , where  $\mathbf{A} = (1, 0, 1)$ ,  $\mathbf{B} = (-1, 2, 3)$ , and  $\mathbf{C} = (2, 6, 1)$  (Does the origin lie on this plane?)
- (g) the plane parallel to  $yz$ -plane passing through the point  $(1, 1, 1)$ .

# Vector spaces

# 2

In the previous chapter we reviewed the basic notions of vectors in space and their elementary application to the study of lines and planes. We derived elementary vector equations for lines and planes and saw how once a coordinate system was chosen these vector equations lead to the familiar equations of analytic geometry. However, particularly in application to physics, it is often very important to know the relation between the equations for the same plane (or line) in different coordinate systems. This leads us to the notion of a *coordinate transformation*. The appropriate domain in which to study such transformations are the abstract vector spaces to be introduced now.

**Definition.** A *vector space* is a set, whose elements are called vectors, together with two operations. The first operation, called *vector addition*, assigns to each pair of vectors  $\mathbf{A}$  and  $\mathbf{B}$  a vector denoted by  $\mathbf{A} + \mathbf{B}$ , called their sum. The second operation, called *scalar multiplication*, assigns to each vector  $\mathbf{A}$  and each number<sup>1</sup>  $r$  a vector denoted by  $r\mathbf{A}$ . The two operations are required to have the following properties:

**Axiom 1.**  $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$  for each pair of vectors  $\mathbf{A}$  and  $\mathbf{B}$  (Commutative law of vector addition).

**Axiom 2.**  $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$  for each triple of vectors,  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$ .

**Axiom 3.** There is a unique vector  $\mathbf{0}$ , called the zero vector, such that  $\mathbf{A} + \mathbf{0} = \mathbf{A}$  for every vector  $\mathbf{A}$ .

<sup>1</sup> For the moment we agree that the number  $r$  is a real number

## 2: Vector spaces

**Axiom 4.** For each vector  $\mathbf{A}$  there corresponds a unique vector  $-\mathbf{A}$  such that  $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$ .

**Axiom 5.**  $r(\mathbf{A} + \mathbf{B}) = r\mathbf{A} + r\mathbf{B}$  for each real number  $r$  and each pair of vectors  $\mathbf{A}$  and  $\mathbf{B}$ .

**Axiom 6.**  $(r + s)\mathbf{A} = r\mathbf{A} + s\mathbf{A}$  for each pair of real numbers  $r$  and  $s$  and each vector  $\mathbf{A}$ .

**Axiom 7.**  $(rs)\mathbf{A} = r(s\mathbf{A})$  for each pair  $r, s$  of real numbers and each vector  $\mathbf{A}$ .

**Axiom 8.** For each vector  $\mathbf{A}$ ,  $1\mathbf{A} = \mathbf{A}$ .

In developing the mathematical theory of linear algebra we are going to follow the *axiomatic method*. That is a *vector*, *vector addition*, and *scalar multiplication* constitute the basic terms of the theory. They are not defined but rather our study of linear algebra will be based on the properties of these terms as specified by the preceding eight axioms. In the axiomatic treatment, what vectors, vector addition and scalar multiplication are is immaterial, rather what is important is the properties these quantities have as consequences of the axioms. Thus in our development of the theory we may not use properties of vectors that are not stated in or are consequences of the preceding axioms. We may use any properties of vectors, etc. that are stated in the axioms: for example, that the vector  $\mathbf{0}$  is unique, or that  $\mathbf{A} = 1\mathbf{A}$  for any vector  $\mathbf{A}$ . On the other hand we may *not* say that a vector is an arrow with a specified head and tail.

The advantage of the axiomatic approach is that results so obtained will apply to any special case or example that we wish to consider. The converse is definitely false. Presently we will see an enormous number of examples of vector spaces. Let us first begin with some elementary consequences of the axioms.

**Proposition 2.1.**  $0\mathbf{A} = \mathbf{0}$ .

PROOF. We have

$$\mathbf{A} = 1\mathbf{A} = (1 + 0)\mathbf{A} = 1\mathbf{A} + 0\mathbf{A} = \mathbf{A} + 0\mathbf{A}$$

by using Axioms 8, 6, and 8 again. To the equation

$$\mathbf{A} = \mathbf{A} + 0\mathbf{A}$$

we apply Axiom 1 getting

$$\mathbf{A} = 0\mathbf{A} + \mathbf{A}.$$

Now apply Axiom 4 and we obtain by Axiom 2:

$$\begin{aligned} \mathbf{0} &= \mathbf{A} + (-\mathbf{A}) = (0\mathbf{A} + \mathbf{A}) + (-\mathbf{A}) = 0\mathbf{A} + (\mathbf{A} + (-\mathbf{A})) \\ &= 0\mathbf{A} + \mathbf{0} = 0\mathbf{A} \end{aligned}$$

by Axiom 3.

That is

$$\mathbf{0} = 0\mathbf{A}$$

which is the desired conclusion.  $\square$

**Notational Convention:** It should be clear by now that we will reserve capital letters for vectors and small letters for numbers.

The proof of (2.1) was given in considerable detail to illustrate how results are deduced by the axiomatic method. In the sequel we will not be so detailed in our proofs, leaving to the reader the task of providing as much detail as he feels needed.

**Proposition 2.2.**  $(-1)\mathbf{A} = -\mathbf{A}$ .

PROOF. We have by (2.1)

$$\mathbf{0} = 0\mathbf{A} = (1 - 1)\mathbf{A} = 1\mathbf{A} + (-1)\mathbf{A} = \mathbf{A} + (-1)\mathbf{A}.$$

Now add  $-\mathbf{A}$  to both sides giving

$$\begin{aligned} -\mathbf{A} &= -\mathbf{A} + (\mathbf{A} + (-1)\mathbf{A}) = (-\mathbf{A} + \mathbf{A}) + (-1)\mathbf{A} \\ &= (\mathbf{A} - \mathbf{A}) + (-1)\mathbf{A} = \mathbf{0} + (-1)\mathbf{A} = (-1)\mathbf{A} + \mathbf{0} \\ &= (-1)\mathbf{A} \end{aligned}$$

as required.  $\square$

**Proposition 2.3.**  $\mathbf{0} + \mathbf{A} = \mathbf{A}$ .

PROOF. Exercise.  $\square$

These formal deductions may seem like a sterile intellectual exercise—an indication of the absurdity of too much reliance on abstraction and formalism. On the contrary, they help to point up the advantages of an abstract formulation of a mathematical theory. For if the basic terms are not defined, the possibility is opened of assigning to them content in new and unforeseen ways. If in this way the axioms become true statements when the meanings assigned to the basic terms vector, vector addition, and scalar multiplication are specified, we have constructed a **model** for the abstract theory. That is, if we can assign a meaning to the terms vector, vector addition and scalar multiplication such that Axioms 1–8 become true statements about this assignment then we say we have constructed a *model* or example of the axioms. Here then is one standard such model.



## Cartesian or Euclidean spaces

**Definition.** Let  $k$  be a positive integer. The *Cartesian  $k$ -space* denoted by  $\mathbb{R}^k$ , is the set of all sequences  $(a_1, a_2, \dots, a_k)$  of  $k$  real numbers together with the two operations

$$(a_1, a_2, \dots, a_k) + (b_1, b_2, \dots, b_k) = (a_1 + b_1, a_2 + b_2, \dots, a_k + b_k)$$

and

$$r(a_1, \dots, a_k) = (ra_1, ra_2, \dots, ra_k).$$

(In particular  $\mathbb{R}^1 = \mathbb{R}$  is the set of real numbers with their usual addition and multiplication.) The number  $a_i$  is called the  *$i$ th component* of  $(a_1, \dots, a_k)$ .

**Theorem 2.4.** For each positive integer  $k$ ,  $\mathbb{R}^k$  is a vector space.

Before beginning the proof of (2.4) let us consider exactly what it is that we are trying to prove. We are going to assign meanings to the three basic terms of the axioms for a vector space. Namely, by a vector we will mean a  $k$ -tuple  $(a_1, \dots, a_k)$ . For vectors  $\mathbf{A} = (a_1, \dots, a_k)$  and  $\mathbf{B} = (b_1, \dots, b_k)$ , the equality of the two vectors  $\mathbf{A} = \mathbf{B}$  means that  $a_1 = b_1, a_2 = b_2, \dots, a_k = b_k$ . By addition of  $\mathbf{A}$  and  $\mathbf{B}$ , we shall mean the vector  $(a_1 + b_1, \dots, a_k + b_k)$ , that is we define

$$\mathbf{A} + \mathbf{B} = (a_1 + b_1, \dots, a_k + b_k).$$

Likewise we define

$$r\mathbf{A} = (ra_1, \dots, ra_k).$$

Axioms 1–8 for a vector space then become statements about  $k$ -tuples and we must verify that they are true statements.

**PROOF OF (2.4).** We will verify the axioms in turn.

**Axiom 1.** Let  $\mathbf{A} = (a_1, \dots, a_k)$ ,  $\mathbf{B} = (b_1, \dots, b_k)$ . Then

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= (a_1 + b_1, \dots, a_k + b_k) \\ &= (b_1 + a_1, \dots, b_k + a_k) = \mathbf{B} + \mathbf{A}\end{aligned}$$

so Axiom 1 is true.

**Axiom 2.** Let  $\mathbf{A} = (a_1, \dots, a_k)$ ,  $\mathbf{B} = (b_1, \dots, b_k)$ , and  $\mathbf{C} = (c_1, \dots, c_k)$ . Then

$$\begin{aligned}(\mathbf{A} + \mathbf{B}) + \mathbf{C} &= (a_1 + b_1, \dots, a_k + b_k) + (c_1, \dots, c_k) \\ &= (a_1 + b_1 + c_1, \dots, a_k + b_k + c_k) \\ &= (a_1, \dots, a_k) + (b_1 + c_1, \dots, b_k + c_k) \\ &= \mathbf{A} + (\mathbf{B} + \mathbf{C})\end{aligned}$$

so Axiom 2 holds.

**Axiom 3.** We let  $\mathbf{0} = (0, \dots, 0)$ . Then for any  $\mathbf{A} = (a_1, \dots, a_k)$  we will have

$$\mathbf{A} + \mathbf{0} = (a_1 + 0, \dots, a_k + 0) = (a_1, \dots, a_k) = \mathbf{A}.$$

Moreover if  $\mathbf{B} = (b_1, \dots, b_k)$  is any vector such that

$$\mathbf{A} + \mathbf{B} = \mathbf{A}$$

then

$$(a_1 + b_1, \dots, a_k + b_k) = (a_1, \dots, a_k)$$

and therefore

$$\begin{array}{rcl} a_1 + b_1 = a_1 & \Rightarrow & b_1 = 0 \\ a_2 + b_2 = a_2 & \Rightarrow & b_2 = 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_k + b_k = a_k & \Rightarrow & b_k = 0 \end{array}$$

i.e.  $\mathbf{B} = \mathbf{0}$ . Thus  $\mathbf{0}$  is the unique vector with the property that  $\mathbf{A} + \mathbf{0} = \mathbf{A}$ , and Axiom 3 holds.

**Axiom 4.** Let  $\mathbf{A} = (a_1, \dots, a_k)$  and set  $-\mathbf{A} = (-a_1, \dots, -a_k)$ . Then

$$\begin{aligned} \mathbf{A} + (-\mathbf{A}) &= (a_1, \dots, a_k) + (-a_1, \dots, -a_k) \\ &= (a_1 - a_1, \dots, a_k - a_k) = (0, \dots, 0) = \mathbf{0}. \end{aligned}$$

Moreover if  $\mathbf{C} = (c_1, \dots, c_k)$  is any vector such that

$$\mathbf{A} + \mathbf{C} = \mathbf{0}$$

then

$$(a_1 + c_1, \dots, a_k + c_k) = (0, \dots, 0)$$

and therefore

$$\begin{array}{rcl} a_1 + c_1 = 0 & \Rightarrow & c_1 = -a_1 \\ a_2 + c_2 = 0 & \Rightarrow & c_2 = -a_2 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_k + c_k = 0 & \Rightarrow & c_k = -a_k \end{array}$$

i.e.  $\mathbf{C} = -\mathbf{A}$ . Thus  $-\mathbf{A}$  is the unique vector with the property that  $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$  and Axiom 4 holds.

## 2: Vector spaces

**Axiom 5.** Let  $r$  be a real number and  $\mathbf{A} = (a_1, \dots, a_k)$  and  $\mathbf{B} = (b_1, \dots, b_k)$  be vectors. Then

$$\begin{aligned} r(\mathbf{A} + \mathbf{B}) &= r(a_1 + b_1, \dots, a_k + b_k) = (r(a_1 + b_1), \dots, r(a_k + b_k)) \\ &= (ra_1 + rb_1, \dots, ra_k + rb_k) = (ra_1, \dots, ra_k) + (rb_1, \dots, rb_k) \\ &= r(a_1, \dots, a_k) + r(b_1, \dots, b_k) = r\mathbf{A} + r\mathbf{B} \end{aligned}$$

so that Axiom 5 is satisfied.

**Axiom 6.** Let  $r, s$  be numbers and  $\mathbf{A} = (a_1, \dots, a_k)$ . Then

$$\begin{aligned} (r + s)\mathbf{A} &= ((r + s)a_1, \dots, (r + s)a_k) \\ &= (ra_1 + sa_1, \dots, ra_k + sa_k) \\ &= (ra_1, \dots, ra_k) + (sa_1, \dots, sa_k) \\ &= r(a_1, \dots, a_k) + s(a_1, \dots, a_k) \\ &= r\mathbf{A} + s\mathbf{A} \end{aligned}$$

so Axiom 6 holds.

**Axiom 7.** Let  $r, s$  be numbers and  $\mathbf{A} = (a_1, \dots, a_k)$ . Then

$$\begin{aligned} (rs)\mathbf{A} &= (rsa_1, \dots, rsa_k) = r(sa_1, \dots, sa_k) \\ &= r(s(a_1, \dots, a_k)) = r(s(\mathbf{A})) \end{aligned}$$

so Axiom 7 holds.

**Axiom 8.** Instant.

Therefore  $\mathbb{R}^k$  is a vector space. □

Before turning to a few additional examples of vector spaces let us deduce some more elementary consequences of the axioms. One of these is the general associative law.

**Proposition 2.5.** *Let  $n$  be an integer  $n \geq 3$ . Then any two ways of associating a sum*

$$\mathbf{A}_1 + \dots + \mathbf{A}_n$$

*of  $n$ -vectors give the same vector. Consequently sums may be written without parentheses.*

The proof of this proposition is elementary and may be carried out by induction on  $n$ . Similarly we have:

**Proposition 2.6.** *Let  $n$  be any integer  $\geq 2$ . Then the sum of any  $n$ -vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is independent of the order in which the sum is taken.*

**Notations.** We will use the symbol  $\in$  as an abbreviation for “is an element of.” Thus  $x \in S$  should read:  $x$  is an element of the set  $S$ .

The symbol  $\subset$  is an abbreviation for “is contained in.” Thus  $S \subset T$  should be read: the set  $S$  is contained in the set  $T$ .

If  $S$  and  $T$  are sets then the collection of elements contained in either set is denoted by  $S \cup T$ . Thus  $x \in S \cup T$  is equivalent to  $x \in S$  or  $x \in T$ . The collection of all elements common to both sets is denoted  $S \cap T$ . Thus  $x \in S \cap T$  is equivalent to  $x \in S$  and  $x \in T$ . The set  $S \cup T$  is called the **union** of  $S$  and  $T$  and  $S \cap T$  the **intersection** of  $S$  and  $T$ . We denote by  $\emptyset$  the empty set.

The axioms for a vector space that we have given are the axioms for a *real* vector space, that is a vector space whose scalars are the *real numbers*, which we denoted by  $\mathbb{R}$ . It is also possible and often important to study vector spaces whose scalars are the *complex numbers*, which we denote by  $\mathbb{C}$ . A vector space with complex scalars is called a *complex vector space*. The axioms for a complex vector space are exactly as for a real vector space except that the **numbers** (= scalars) are to be complex. The generic example of a complex vector space is the complex Cartesian space  $\mathbb{C}^k$  of  $k$ -tuples  $\mathbf{A} = (a_1, \dots, a_k)$  of *complex numbers* where, for vectors  $\mathbf{A} = (a_1, \dots, a_k)$  and  $\mathbf{B} = (b_1, \dots, b_k)$  and for scalars  $r \in \mathbb{C}$ , vector addition and scalar multiplication are given by

$$\mathbf{A} + \mathbf{B} = (a_1 + b_1, \dots, a_k + b_k)$$

and

$$r\mathbf{A} = (ra_1, \dots, ra_k).$$

For a while at least we will study only real vector spaces, indicating where necessary the modifications required in the complex case.

## EXERCISES

1. Assume that the plane is equipped with a coordinate system. The set  $\mathcal{V}$  of all vectors  $\bar{\mathbf{P}}$  with initial point at the origin and terminal point at  $\mathbf{P} = (x, y)$  of Chapter 1 is a vector space with the operation described in Chapter 1.
2. Let  $A = \{(2a, a) | a \in \mathbb{R}\}$ ,  $B = \{(b, b) | b \in \mathbb{R}\}$ . Find  $A \cup B$  and  $A \cap B$ .
3.  $A = \{(2n, n) | n \in \text{integers}\}$ ,  $B = \{(k + 1, k) | k \in \text{integers}\}$ . Find  $A \cup B$ ,  $A \cap B$ .
4. If  $A \subset B$ ,  $B \subset C$  then  $A \subset C$ .
5. If  $A \subset B$ ,  $A \subset C$  then  $A \subset B \cap C$ .
6. If  $A \supset B$ ,  $A \supset C$  then  $A \supset B \cup C$ .
7. Show  $A \cap (B \cap C) = (A \cap B) \cap (A \cap C)$  and  $A \cap (B \cup C) = (A \cap B) \cap (A \cap C)$ , where  $A, B, C$  are sets.
8. Let  $A = \{x \in \mathbb{R} | |x| > 1\}$ ,  $B = \{x \in \mathbb{R} | -2 < x < 3\}$ . Find  $A \cup B$  and  $A \cap B$ .

## 2: Vector spaces

9. Let  $\mathcal{V}$  be a vector space. Prove each of the following statements:

- (a) If  $\mathbf{A} \in \mathcal{V}$  and  $a$  is a number then  $a\mathbf{A} = \mathbf{0}$  iff  $a = 0$  or  $\mathbf{A} = \mathbf{0}$ , or both.
- (b) If  $\mathbf{A} \in \mathcal{V}$  and  $a$  is a number then  $a\mathbf{A} = \mathbf{A}$  iff  $a = 1$  or  $\mathbf{A} = \mathbf{0}$ , or both.

10. Let  $V$  be the set of all ordered pairs of real numbers  $(a, b)$ . Define an addition for the elements of  $V$  by the rule

$$(a, b) \oplus (c, d) = (a + c, b + d).$$

and a multiplication of elements of  $V$  by numbers by the rule

$$a \cdot (c, d) = (ac, ad).$$

Is  $V$ , with these two operations, a vector space? Justify your answer.

11. A *translation* of the plane  $\Pi$  is a function  $T: \Pi \rightarrow \Pi$  with the following two properties:

- (1) There is a constant  $k$ , called the **length** of the translation, such that for every point  $p \in \Pi$  the distance from  $p$  to  $T(p)$  is equal to  $k$ .
- (2) For any two points  $p, q \in \Pi$  the distance from  $p$  to  $q$  is the same as the distance from  $T(p)$  to  $T(q)$ .

Introduce cartesian coordinates in  $\Pi$  and show:

- (a) If  $T$  is a translation, then for every  $p = (x, y) \in \Pi$ ,  $T(p) = (x + l_1, y + l_2)$  where  $T(0, 0) = (l_1, l_2)$ .  $T$  is said to be the translation by  $l = (l_1, l_2)$ .
- (b) If  $l = (l_1, l_2) \in \Pi$  show that

$$T(p) = (x + l_1, y + l_2), \quad p = (x, y) \in \Pi$$

defines a translation of  $\Pi$ .

If  $T, S$  are translations define their sum  $T \oplus S$  by

$$(T \oplus S)(p) = T(S(p)).$$

- (c) Show that  $T \oplus S$  is again a translation. If  $T$  is the translation by  $l = (l_1, l_2)$  and  $a$  is a number define  $a \cdot T$  to be the translation by  $(al_1, al_2)$ .
- (d) Show that the set of translations with the addition  $\oplus$  and scalar multiplication  $\cdot$  form a vector space.

12. Show that if a vector space contains two elements then it contains infinitely many.

**Definition.** A nonempty subset  $\mathcal{U}$  of a vector space  $\mathcal{V}$  is called a *linear subspace* of  $\mathcal{V}$  iff the following two conditions are satisfied:

- (1) if  $\mathbf{A} \in \mathcal{U}$  and  $\mathbf{B} \in \mathcal{U}$  then  $\mathbf{A} + \mathbf{B} \in \mathcal{U}$
- (2) if  $\mathbf{A} \in \mathcal{U}$  and  $r \in \mathbb{R}$  then  $r\mathbf{A} \in \mathcal{U}$ .

These two conditions assert that applying the two basic vector operations to elements of the collection  $\mathcal{U}$  give again elements of the collection  $\mathcal{U}$ . If the vector space  $\mathcal{V}$  is complex then Condition (2) should be replaced by (2C) if  $\mathbf{A} \in \mathcal{U}$  and  $c \in \mathbb{C}$  then  $c\mathbf{A} \in \mathcal{U}$ , and likewise in the sequel.

**Proposition 3.1.** If  $\mathcal{U}$  is a linear subspace of the vector space  $\mathcal{V}$  then  $\mathcal{U}$  is itself a vector space if we define vector addition and scalar multiplication as in  $\mathcal{V}$ .

**PROOF.** Notice that Conditions (1), (2) assure us that we have operations on  $\mathcal{U}$ , i.e., if  $\mathbf{A}$  and  $\mathbf{B}$  belong to  $\mathcal{U}$  so do  $\mathbf{A} + \mathbf{B}$  and if  $r$  belongs to  $\mathbb{R}$ ,  $r\mathbf{A}$  also belongs to  $\mathcal{U}$ . The properties expressed by Axioms 1, 2, 5, 6, 7, and 8 are valid for vectors in  $\mathcal{V}$  and hence for vectors in the smaller set  $\mathcal{U}$ . To verify Axiom 3 we first show that  $\mathbf{0} \in \mathcal{U}$ . Since  $\mathcal{U}$  is nonempty there exists at least one vector  $\mathbf{A} \in \mathcal{U}$ . By (2.1) and Condition (2) for a subspace  $\mathbf{0} = 0\mathbf{A} \in \mathcal{U}$ . Axiom 3 is now immediate since it holds in  $\mathcal{V}$ . To verify Axiom 4 suppose that  $\mathbf{A} \in \mathcal{U}$ . Then  $(-1)\mathbf{A} \in \mathcal{U}$ , but by (2.2),  $(-1)\mathbf{A} = -\mathbf{A}$  and therefore  $-\mathbf{A} \in \mathcal{U}$  and Axiom 4 holds.  $\square$

## EXAMPLES

- (1)  $\mathcal{V}$  is always a subspace of  $\mathcal{V}$ .
- (2) The set consisting of the zero vector alone  $\{\mathbf{0}\}$  is always a subspace of  $\mathcal{V}$ . We often abuse notation and write  $\mathbf{0}$  for this subspace.

### 3. Subspaces

**Definition.** If  $A_1, \dots, A_n$  are vectors of  $\mathcal{V}$ , then a *linear combination* of  $A_1, \dots, A_n$  is a vector of the form

$$A = a_1 A_1 + \dots + a_n A_n$$

where  $a_1, \dots, a_n$  are numbers.

**Definition.** If the vectors  $A_1, \dots, A_n$  are fixed, the *linear span* of  $A_1, \dots, A_n$ , denoted  $\mathcal{L}(A_1, \dots, A_n)$ , is the set of all vectors of  $\mathcal{V}$  which are linear combinations of  $A_1, \dots, A_n$ .

**Proposition 3.2.** Suppose that  $A_1, \dots, A_n$  are vectors of  $\mathcal{V}$ , then  $\mathcal{L}(A_1, \dots, A_n)$  is a linear subspace of  $\mathcal{V}$ .

**PROOF.** We must verify that the two conditions of the definition of linear subspace are satisfied by the linear combinations of  $A_1, \dots, A_n$ . So suppose

$$A', A'' \in \mathcal{L}(A_1, \dots, A_n).$$

Then

$$A' = a'_1 A_1 + \dots + a'_n A_n$$

$$A'' = a''_1 A_1 + \dots + a''_n A_n$$

for suitable numbers  $a'_1, \dots, a'_n, a''_1, \dots, a''_n$ . Then using the generalized associative and commutative laws we find

$$\begin{aligned} A' + A'' &= a'_1 A_1 + \dots + a'_n A_n + a''_1 A_1 + \dots + a''_n A_n \\ &= a'_1 A_1 + a''_1 A_1 + \dots + a'_n A_n + a''_n A_n \\ &= (a'_1 + a''_1) A_1 + \dots + (a'_n + a''_n) A_n \end{aligned}$$

which shows that  $A' + A''$  is again a linear combination of  $A_1, \dots, A_n$ , that is,  $A' + A'' \in \mathcal{L}(A_1, \dots, A_n)$ . Similarly if  $r \in \mathbb{R}$  and  $A \in \mathcal{L}(A_1, \dots, A_n)$  then

$$A = a_1 A_1 + \dots + a_n A_n$$

for suitable numbers  $a_1, \dots, a_n$ , so

$$\begin{aligned} rA &= r(a_1 A_1 + \dots + a_n A_n) \\ &= ra_1 A_1 + \dots + ra_n A_n \end{aligned}$$

showing  $rA \in \mathcal{L}(A_1, \dots, A_n)$ . Therefore  $\mathcal{L}(A_1, \dots, A_n)$  is a subspace of  $\mathcal{V}$ .  $\square$

The idea of the linear span is not restricted to finite sets of vectors, but may be extended to arbitrary sets of vectors as follows.

**Definition.** Let  $\mathcal{V}$  be a vector space and  $E \subset \mathcal{V}$ , that is,  $E$  is a collection of vectors in  $\mathcal{V}$ . A *linear combination* of vectors in  $E$  is a vector in  $\mathcal{V}$  of the form

$$a_1 A_1 + a_2 A_2 + \dots + a_n A_n$$

where  $A_1, \dots, A_n \in E$ . The *linear span* of  $E$ , denoted by  $\mathcal{L}(E)$ , is the set of all vectors that are linear combinations of vectors of  $E$ .

**Proposition 3.3.** *Let  $\mathcal{V}$  be a vector space and  $E \subset \mathcal{V}$ . Then  $\mathcal{L}(E)$  is a linear subspace of  $\mathcal{V}$ .*

The proof of (3.3) follows closely the proof of (3.2) and will be left to the diligent student. Note that the linear span allows us to assign to each subset of  $\mathcal{V}$  a subspace of  $\mathcal{V}$ . Note  $E \subset \mathcal{L}(E)$ .

**Proposition 3.4.** *Let  $\mathcal{V}$  be a vector space and  $E \subset \mathcal{V}$ . Then  $E = \mathcal{L}(E)$  iff  $E$  is a linear subspace of  $\mathcal{V}$ .*

PROOF. Suppose that  $E$  is a linear subspace of  $\mathcal{V}$ . Then if  $A_1, \dots, A_n \in E$  and  $a_1, \dots, a_n$  are numbers the vector  $a_1 A_1 + \dots + a_n A_n$  belongs to  $E$  because  $E$  is closed under the operation of scalar multiplication and vector addition. Therefore  $\mathcal{L}(E) \subset E$ . Since  $E \subset \mathcal{L}(E)$  we must conclude that  $E = \mathcal{L}(E)$ .

Conversely, suppose that  $E = \mathcal{L}(E)$ . If  $A, B \in E$  then  $A + B$  is certainly a linear combination of vectors in  $E$  and hence  $A + B$  belongs to  $\mathcal{L}(E)$ , which since  $\mathcal{L}(E) = E$  leads us to conclude  $A + B \in E$ . Likewise  $aA$  is a linear combination of vectors of  $E$  and hence belongs to  $\mathcal{L}(E) = E$ . Therefore  $E$  is closed under vector addition and scalar multiplication, and hence  $E$  is a linear subspace of  $\mathcal{V}$ .  $\square$

The preceding propositions show that in general a vector space has an abundance of subspaces.

EXAMPLE. In  $\mathbb{R}^3$  consider the subspace spanned by the two vectors  $A = (1, 0, 1)$  and  $B = (0, 1, 0)$ . (See Figure 3.1.) Note that this is just the plane through the origin,  $x - z = 0$ . That is the vectors in  $\mathcal{L}(A, B)$  are those vectors  $(x, y, z) \in \mathbb{R}^3$  whose coordinates satisfy the equation  $x - z = 0$ .

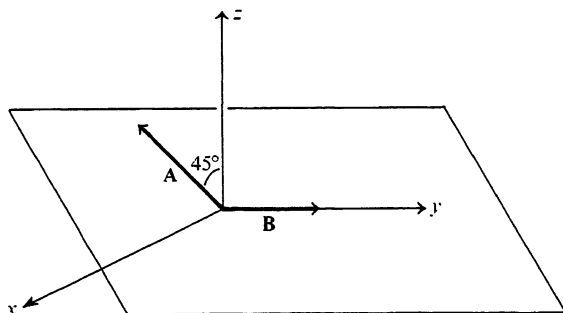


Figure 3.1



### 3: Subspaces

**Proposition 3.5.** Let  $\mathcal{S}$  and  $\mathcal{T}$  be subspaces of  $\mathcal{V}$ . Then  $\mathcal{S} \cap \mathcal{T}$  is also a subspace of  $\mathcal{V}$ .

PROOF. Suppose that  $\mathbf{A} \in \mathcal{S} \cap \mathcal{T}$  and  $\mathbf{B} \in \mathcal{S} \cap \mathcal{T}$ . Then  $\mathbf{A} \in \mathcal{S}$  and  $\mathbf{B} \in \mathcal{S}$ . Since  $\mathcal{S}$  is a subspace  $\mathbf{A} + \mathbf{B} \in \mathcal{S}$ . Likewise  $\mathbf{A} \in \mathcal{T}$  and  $\mathbf{B} \in \mathcal{T}$  and since  $\mathcal{T}$  is a subspace  $\mathbf{A} + \mathbf{B} \in \mathcal{T}$ . Therefore  $\mathbf{A} + \mathbf{B} \in \mathcal{S} \cap \mathcal{T}$ . If  $r$  is a number, then since  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces  $r\mathbf{A} \in \mathcal{S}$  and  $r\mathbf{A} \in \mathcal{T}$  so  $r\mathbf{A} \in \mathcal{S} \cap \mathcal{T}$  showing that  $\mathcal{S} \cap \mathcal{T}$  is again a subspace of  $\mathcal{V}$ .  $\square$

**Definition.** If  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces of  $\mathcal{V}$ , their *sum*, denoted by  $\mathcal{S} + \mathcal{T}$ , is defined to be the set of all vectors  $\mathbf{C}$  in  $\mathcal{V}$  of the form

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

where  $\mathbf{A} \in \mathcal{S}$  and  $\mathbf{B} \in \mathcal{T}$ .

**Proposition 3.6.** If  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces of the vector space  $\mathcal{V}$  then so is  $\mathcal{S} + \mathcal{T}$ .

PROOF. Suppose that  $\mathbf{C}_1, \mathbf{C}_2 \in \mathcal{S} + \mathcal{T}$ . Write

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{A}_1 + \mathbf{B}_1 & \mathbf{A}_1 \in \mathcal{S}, \mathbf{B}_1 \in \mathcal{T} \\ \mathbf{C}_2 &= \mathbf{A}_2 + \mathbf{B}_2 & \mathbf{A}_2 \in \mathcal{S}, \mathbf{B}_2 \in \mathcal{T}. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{C}_1 + \mathbf{C}_2 &= \mathbf{A}_1 + \mathbf{B}_1 + \mathbf{A}_2 + \mathbf{B}_2 \\ &= (\mathbf{A}_1 + \mathbf{A}_2) + (\mathbf{B}_1 + \mathbf{B}_2) \end{aligned}$$

Let  $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ ,  $\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2$ . Since  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces  $\mathbf{A} \in \mathcal{S}$  and  $\mathbf{B} \in \mathcal{T}$  while since

$$\mathbf{C}_1 + \mathbf{C}_2 = \mathbf{A} + \mathbf{B}$$

it follows that  $\mathbf{C}_1 + \mathbf{C}_2 \in \mathcal{S} + \mathcal{T}$ .

Next suppose that  $\mathbf{C} \in \mathcal{S} + \mathcal{T}$  and  $r \in \mathbb{R}$ . Then we may write

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \quad \mathbf{A} \in \mathcal{S}, \mathbf{B} \in \mathcal{T}$$

and hence

$$r\mathbf{C} = r(\mathbf{A} + \mathbf{B}) = (r\mathbf{A}) + (r\mathbf{B}).$$

Since  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces  $r\mathbf{A} \in \mathcal{S}$ ,  $r\mathbf{B} \in \mathcal{T}$  and hence  $r\mathbf{C} \in \mathcal{S} + \mathcal{T}$ .  $\square$

**PROBLEM.** Suppose that  $\mathcal{S}, \mathcal{T}$  are subspaces of  $\mathcal{V}$ . When is  $\mathcal{S} \cup \mathcal{T}$  again a subspace of  $\mathcal{V}$ ?

*Answer.* If  $\mathcal{S} \subset \mathcal{T}$  or  $\mathcal{T} \subset \mathcal{S}$ .

## EXERCISES

- Which of the following collections of vectors in  $\mathbb{R}^3$  are subspaces?
  - $\mathcal{U} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 = 0\}$
  - $\mathcal{U} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 = 0\}$
  - $\mathcal{U} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 = 0\}$
  - $\mathcal{U} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 = 1\}$
  - $\mathcal{U} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 \geq 0\}$ .
- Determine the subspace of  $\mathbb{R}^3$  which is the linear span of the three vectors  $(1, 0, 1)$ ,  $(0, 1, 0)$ ,  $(0, 1, 1)$ .
- Repeat Exercise 2 for  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(1, 1, 1)$ .
- Suppose  $\mathcal{S}$ ,  $\mathcal{T}$  are subspaces of  $\mathcal{V}$  and  $\mathcal{S} \cap \mathcal{T} = \mathbf{0}$ . Show that every vector in  $\mathcal{S} + \mathcal{T}$  can be written *uniquely* in the form  $\mathbf{A} + \mathbf{B}$ ,  $\mathbf{A} \in \mathcal{S}$ ,  $\mathbf{B} \in \mathcal{T}$ . Construct an example to show that this is false if  $\mathcal{S} \cap \mathcal{T} \neq \mathbf{0}$ .
- Show that any nonzero vector spans  $\mathbb{R}^1$ .
- Show that the two sets of vectors

$$\{\mathbf{A} = (1, 1, 0), \mathbf{B} = (0, 0, 1)\}$$

and

$$\{\mathbf{C} = (1, 1, 1), \mathbf{D} = (-1, -1, 1)\}$$

span the same subspace of  $\mathbb{R}^3$ .

- Let  $\mathcal{V}$  be the set of pairs of numbers  $\mathbf{A} = (a_1, a_2)$ . If  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$  define  $(\mathbf{B} = (b_1, b_2))$

$$\mathbf{A} + \mathbf{B} = (a_1 + b_1, a_2 + b_2).$$

If  $a$  is a number define

$$a\mathbf{A} = (aa_1, 0)$$

Is  $\mathcal{V}$  a vector space? Why?

- Suppose  $\mathcal{V}$  is a vector space and  $E, F$  are subsets of  $\mathcal{V}$ . Show
  - $E \subset F \Rightarrow \mathcal{L}(E) \subset \mathcal{L}(F)$ .
  - $\mathcal{L}(E \cup F) = \mathcal{L}(E) + \mathcal{L}(F)$ .
  - $\mathcal{L}(E \cap F) \subset \mathcal{L}(E) \cap \mathcal{L}(F)$ .
- Let  $\mathcal{V}$  be a vector space and  $E, F \subset \mathcal{V}$ . Suppose  $\mathcal{L}(E) \subset \mathcal{L}(F)$ . Is it true that  $E \subset F$ ?
- Suppose that  $\mathcal{V}$  is a vector space and  $E \subset \mathcal{V}$ . If  $\mathcal{U}$  is a subspace containing  $E$  then  $\mathcal{U}$  contains  $\mathcal{L}(E)$ .
- Suppose  $\mathcal{V}$  is a vector space and  $E \subset \mathcal{V}$ . Show that  $\mathcal{L}(E) = \cap \{\mathcal{U} \mid \mathcal{U} \text{ is a subspace of } \mathcal{V} \text{ and } \mathcal{U} \text{ contains } E\}$ .
- For any subspace  $\mathcal{U}$  of  $\mathcal{V}$  show that  $\mathcal{U} + \mathcal{U} = \mathcal{U}$ .
- Find all the linear subspaces of  $\mathbb{R}^2$ .

### 3: Subspaces

14. Find all the linear subspaces of  $\mathbb{R}^3$ .
15. Show that a subset  $E$  of a vector space  $\mathcal{V}$  which does not contain  $\mathbf{0}$  is not a subspace of  $\mathcal{V}$ .
16. Let  $E$  be the subset of  $\mathbb{R}^2$  defined by  $E = \{(x, y) | x \geq 0, y \in \mathbb{R}\}$ . Is  $E$  a subspace of  $\mathbb{R}^2$ ?
17. Let  $E = \{(x, 2x + 1) | x \in \mathbb{R}\}$ .  $E$  is a subset of  $\mathbb{R}^2$ . Is  $E$  a subspace of  $\mathbb{R}^2$ ?
18. (a) Let  $E = \{(2a, a) | a \in \mathbb{R}\}$ . Is  $E$  a subspace of  $\mathbb{R}^2$ ?  
 (b) Let  $B = \{(b, b) | b \in \mathbb{R}\}$ . Is  $B$  a subspace of  $\mathbb{R}^2$ ?  
 (c) What is  $E \cap B$ ?  
 (d) Is  $E \cup B$  a subspace of  $\mathbb{R}^2$ ?  
 (e) What is  $E + B$ ?
19. Let  $\mathcal{S}$  and  $\mathcal{T}$  be subspaces of  $\mathcal{V}$ . Prove:
  - (a)  $\mathcal{S} + \mathcal{T} = \mathcal{L}(\mathcal{S} \cup \mathcal{T})$ .
  - (b)  $\mathcal{S} \cap (\mathcal{S} + \mathcal{T}) = \mathcal{S}$ .
  - (c)  $\mathcal{S} + \mathcal{T} = \mathcal{T} + \mathcal{S}$ .
  - (d) If  $\mathcal{S} \subset \mathcal{T}$ , then  $\mathcal{S} + \mathcal{T} = \mathcal{T}$ .
20. Let  $\mathcal{V}$  be a vector space,  $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{V}$ . Suppose  $\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{0}$ . Show that  $\mathcal{L}(\mathbf{A}, \mathbf{B}) = \mathcal{L}(\mathbf{B}, \mathbf{C})$ .
21. Let  $\mathcal{V}$  be a vector space and  $\mathcal{W}$  a subspace of  $\mathcal{V}$ . Show that
 
$$\{\mathbf{A} \in \mathcal{V} | \mathbf{A} \notin \mathcal{W}\}$$
 is not a vector subspace of  $\mathcal{V}$ .
22. Let  $\mathcal{V}$  be a vector space,  $\mathcal{W}$  a subspace of  $\mathcal{V}$ , and  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$ . Assume that
 
$$\mathbf{A} \notin \mathcal{W} \quad \text{but} \quad \mathbf{A} \in \mathcal{L}(\mathcal{W} \cup \{\mathbf{B}\}).$$
 Show that  $\mathbf{B} \in \mathcal{L}(\mathcal{W} \cup \{\mathbf{A}\})$ .
23. Let  $\mathcal{S}, \mathcal{T}, \mathcal{U}$  be subspaces of a vector space  $\mathcal{V}$ . Is it always true that
 
$$\mathcal{S} \cap (\mathcal{T} + \mathcal{U}) = (\mathcal{S} \cap \mathcal{T}) + (\mathcal{S} \cap \mathcal{U})?$$
24. Can it happen for two subspaces  $\mathcal{S}$  and  $\mathcal{T}$  of  $\mathcal{V}$  that  $\mathcal{S} \cap \mathcal{T} = \emptyset$ ?
25. Can you find two vectors  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^3$  such that  $\mathcal{L}(\mathbf{A}, \mathbf{B}) = \mathbb{R}^3$ ?

Before continuing with our study of the elementary properties of vector spaces and their linear subspaces let us collect a list of examples of vector spaces. We have already encountered the cartesian  $k$ -space  $\mathbb{R}^k$  and so for the sake of completeness let us begin by listing this example:

EXAMPLE 1.  $\mathbb{R}^k$

The first new example that we have in this chapter is primarily designed to destroy the belief that a vector is a quantity with both direction and magnitude and to give meaning to the phrase in our comments on axiomatics in Chapter 2, that “the possibility is opened of assigning to them (the axioms of a vector space) content in new and unforeseen ways.”

EXAMPLE 2.  $\mathcal{P}_n(\mathbb{R})$

The vectors in  $\mathcal{P}_n(\mathbb{R})$  are polynomials

$$p(x) = a_0 + a_1x + \cdots + a_mx^m$$

of degree less than or equal to  $n$ , that is,  $m \leq n$ . Addition of vectors is to be ordinary addition of polynomials and multiplication of a polynomial by a number, the ordinary product of a polynomial by a number. With these interpretations of the basic terms:

vector  $\leftrightarrow$  polynomial of degree  $\leq n$

vector addition  $\leftrightarrow$  addition of polynomials

scalar multiplication  $\leftrightarrow$  multiplication of a polynomial by a number,

we obtain an example of a vector space. To verify that  $\mathcal{P}_n(\mathbb{R})$  is indeed a vector space we must check that the eight declarative sentences obtained

#### 4· Examples of vector spaces

from these interpretations of the basic terms vector, vector addition, scalar multiplication, are true sentences. This is a straightforward deduction from the (assumed) properties of real numbers following the pattern of (2.4) and will be left to the diligent reader.

Note that in this example it is very difficult to say what direction or length a vector has.

##### EXAMPLE 3. $\mathbb{C}$

Let us denote by  $\mathbb{C}$  the *complex numbers*. (We will not here be concerned with the technical details of constructing the complex numbers, but will take them as we learned them in grammar school.) Recall that a complex number looks like

$$a + bi$$

where  $a, b$  are real numbers, and  $i$  is a number with  $i^2 = -1$ .

The vectors in our vector space will be complex numbers. Addition of vectors is to be the ordinary addition of complex numbers, and scalar multiplication the familiar process of multiplying a complex number by a real number. With these interpretations of the basic terms

vector  $\leftrightarrow$  complex number

vector addition  $\leftrightarrow$  addition of complex numbers

scalar multiplication  $\leftrightarrow$  multiplication of a complex number  
by a real number,

we obtain an example of a vector space. The verifications are again routine.

Note that in Example 3 we are not using all of the structure that we have, for it is possible to multiply two complex numbers, that is, *in this example* we may multiply two vectors, something it is not always possible to do in a vector space. This is a possibility worthy of further study, and we will do just that when we study spaces of linear transformation and *linear algebras*.

*Note.* the product of two polynomials of degree at most  $n$  will have degree at most  $2n$ , so you cannot multiply elements of  $\mathcal{P}_n(\mathbb{R})$  in any obvious way.

Example 2 is a very important example, and a prototype for many others of the same type. These examples are characterized by the fact that their “vectors” are actually functions of some type or other.

##### EXAMPLE 4. $\mathcal{P}(\mathbb{R})$

The simplest way to obtain a space akin to but different from  $\mathcal{P}_n(\mathbb{R})$  is simply to remove the restriction that the polynomials have degree at most  $n$ . In this way we obtain the vector space  $\mathcal{P}(\mathbb{R})$  whose vectors are the polynomials

$$p(x) = a_0 + a_1x + \cdots + a_mx^m$$

with no restriction on  $m$ . The interpretation of the basic terms we propose in this example is:

vector  $\leftrightarrow$  polynomial

vector addition  $\leftrightarrow$  addition of polynomials

scalar multiplication  $\leftrightarrow$  multiplication of a polynomial by a number.

It is again a routine verification that the vector space axioms are satisfied.

#### EXAMPLE 5. $\mathcal{C}(a, b)$

This is a very fancy example, and included only to indicate the wealth of possible examples of vector spaces.

Let  $a$  and  $b$  be numbers with  $a < b$ . The vectors of  $\mathcal{C}(a, b)$  are the continuous functions defined for  $a \leq x \leq b$ . Addition of vectors is to be addition of functions. That is if  $f$  and  $g$  are functions defined and continuous for  $a \leq x \leq b$  then  $f + g$  is the function defined by

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

for all  $a \leq x \leq b$ . It is an important theorem of the calculus that  $f + g$  is again a continuous function for  $a \leq x \leq b$ . Scalar multiplication is to be defined as the ordinary product of a function by a number. If the function is continuous then its product by a number is also continuous. The basic terms of a vector space are to be interpreted as follows in this example:

vector  $\leftrightarrow$  continuous function on  $a \leq x \leq b$

vector addition  $\leftrightarrow$  addition of functions

scalar multiplication  $\leftrightarrow$  multiplication of a function by a number.

Again the vector space axioms are easily verified.

EXAMPLE 6. *Linear Homogenous equations.* A linear homogenous equation in the variables  $x_1, \dots, x_n$  is an equation of the form

$$(*) \quad a_1x_1 + a_2x_2 + \dots + a_nx_n = 0.$$

A solution to this equation is a sequence of  $n$ -numbers  $(s_1, \dots, s_n)$  such that

$$a_1s_1 + \dots + a_ns_n = 0.$$

If  $\mathbf{A} = (s_1, \dots, s_n)$  and  $\mathbf{B} = (t_1, \dots, t_n)$  are solutions to  $(*)$  define

$$\mathbf{A} + \mathbf{B} = (s_1 + t_1, \dots, s_n + t_n).$$

We claim that  $\mathbf{A} + \mathbf{B}$  is again a solution to  $(*)$ . For we have

$$\begin{aligned} a_1(s_1 + t_1) + a_2(s_2 + t_2) + \dots + a_n(s_n + t_n) \\ &= a_1s_1 + a_1t_1 + a_2s_2 + a_2t_2 + \dots + a_ns_n + a_nt_n \\ &= a_1s_1 + a_2s_2 + \dots + a_ns_n + a_1t_1 + a_2t_2 + \dots + a_nt_n \\ &= 0 + 0 = 0 \end{aligned}$$

#### 4: Examples of vector spaces

as we claimed. Next define  $a\mathbf{A}$  for a number  $a$  to be

$$a\mathbf{A} = (as_1, \dots, as_n).$$

Simple manipulation shows

$$\begin{aligned} a_1(as_1) + \dots + a_n(as_n) &= aa_1s_1 + \dots + aa_ns_n \\ &= a(a_1s_1 + \dots + a_ns_n) \\ &= a(0) = 0 \end{aligned}$$

so that  $a\mathbf{A}$  is again a solution to (\*). We now define a vector space  $\mathcal{V}$  by the interpretation

vector  $\leftrightarrow$  solution to (\*)

vector addition  $\leftrightarrow$  as defined above

scalar multiplication  $\leftrightarrow$  as defined above.

To show that  $\mathcal{V}$  is a vector space we will show that it is actually a linear subspace of  $\mathbb{R}^n$ . For by definition the vectors of  $\mathcal{V}$  are sequences  $(s_1, \dots, s_n)$  of numbers and hence are vectors in  $\mathbb{R}^n$ . The process of adding solutions and multiplying solutions by scalars is exactly the process of adding vectors in  $\mathbb{R}^n$  and multiplying a vector of  $\mathbb{R}^n$  by a number. In our preceding discussion we checked

- (1) If  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$  then  $\mathbf{A} + \mathbf{B} \in \mathcal{V}$ .
- (2) If  $\mathbf{A} \in \mathcal{V}$  then  $a\mathbf{A} \in \mathcal{V}$  for any number  $a$ .

Thus we may apply (3.1) to conclude that  $\mathcal{V}$  is a vector space. But wait! In order to apply (3.1) to  $\mathcal{V}$  we must know that  $\mathcal{V}$  is nonempty, that is, that (\*) has at least one solution. Happily this is a simple point, because  $(0, \dots, 0)$  is a solution to (\*) as one easily sees, since

$$a_1 0 + a_2 0 + \dots + a_n 0 = 0 + \dots + 0 = 0.$$

Thus  $\mathcal{V}$  is a linear subspace of  $\mathbb{R}^n$ .

The preceding example may be extended from one equation to many, but this is a topic for future study. (Await Chapter 13.)

Continuing our list of examples we introduce:

**EXAMPLE 7.** Let  $S$  be a set and  $\mathcal{F}(S)$  the set of all functions  $\mathbf{f}: S \rightarrow \mathbb{R}$ . If  $\mathbf{f}, \mathbf{g} \in \mathcal{F}(S)$  define

$$\mathbf{f} + \mathbf{g}: S \rightarrow \mathbb{R}$$

by

$$(\mathbf{f} + \mathbf{g})(s) = \mathbf{f}(s) + \mathbf{g}(s)$$

and for a real number  $r$  define

$$r\mathbf{f}: S \rightarrow \mathbb{R}$$

by

$$(r\mathbf{f})(s) = r(\mathbf{f}(s)),$$

for all  $s \in S$ .

Equipped with this vector addition and scalar multiplication the set  $\mathcal{F}(S)$  becomes a vector space. The zero vector of  $\mathcal{F}(S)$  is the function

$$\mathbf{0} : S \rightarrow \mathbb{R}$$

defined by

$$\mathbf{0}(s) = 0$$

for all  $s \in S$ ; that is,  $\mathbf{0}$  is the constant function which takes the value 0 for all  $s \in S$ . The negative of  $\mathbf{f} \in \mathcal{F}(S)$  is the function

$$-\mathbf{f} : S \rightarrow \mathbb{R}$$

defined by

$$(-\mathbf{f})(s) = -\mathbf{f}(s).$$

It is now routine to verify that the axioms of a real vector space are satisfied for  $\mathcal{F}(S)$ .

We can also make the set of all complex-valued functions  $\mathcal{F}_{\mathbb{C}}(S) = \{\mathbf{f} : S \rightarrow \mathbb{C}\}$  into a complex vector space by setting

$$\begin{aligned}(\mathbf{f} + \mathbf{g})(s) &= \mathbf{f}(s) + \mathbf{g}(s) \\ (c\mathbf{f})(s) &= c(\mathbf{f}(s))\end{aligned}$$

for all  $\mathbf{f}, \mathbf{g} \in \mathcal{F}_{\mathbb{C}}(S)$ ,  $s \in S$ , and  $c \in \mathbb{C}$ .

If  $T \subset S$  then we denote by  $\mathcal{F}(S, T)$  the set of all functions

$$\mathbf{f} : S \rightarrow \mathbb{R}$$

such that

$$\mathbf{f}(s) = 0 \quad \text{for all } s \in T.$$

The set  $\mathcal{F}(S, T) \subset \mathcal{F}(S)$  is in fact a subspace. For if  $\mathbf{f}, \mathbf{g} \in \mathcal{F}(S, T)$  and  $s \in T$  then

$$(\mathbf{f} + \mathbf{g})(s) = \mathbf{f}(s) + \mathbf{g}(s) = 0 + 0 = 0$$

and if  $r \in \mathbb{R}$  then

$$(r\mathbf{f})(s) = r\mathbf{f}(s) = r \cdot 0 = 0,$$

so that  $\mathbf{f} + \mathbf{g} \in \mathcal{F}(S, T)$ ,  $r\mathbf{f} \in \mathcal{F}(S, T)$ , and finally since  $\mathbf{0} \in \mathcal{F}(S, T)$  we see that  $\mathcal{F}(S, T)$  is a subspace of  $\mathcal{F}(S)$ .

Finally we will close the **introduction** to examples of vector spaces by describing a rather artificial example.



#### 4: Examples of vector spaces

EXAMPLE 8. Let  $\mathcal{V}$  be the set of all *positive* real numbers and define for  $A, B \in \mathcal{V}$  a vector sum by

$$A + B = A \cdot B$$

where the product on the right is the usual product of numbers. If  $a$  is a number and  $A \in \mathcal{V}$  define

$$a \cdot A = A^a$$

that is the number  $A$  raised to the  $a$  power. Note that since  $A > 0$  the  $(1/a)$ th root of  $A$  will always exist. For example, with these definitions

$$\begin{aligned}2 + 3 &= 6 \\2 \cdot 3 &= 9.\end{aligned}$$

We claim that with these definitions of vector, vector addition and scalar multiplication  $\mathcal{V}$  becomes a vector space. The details of verification are left to you.

The preceding list only barely scratches the surface of the enormous variety of examples of vector spaces. More examples will appear as we progress through the book and will by no means exhaust the possibilities.

#### EXERCISES

1. Show that  $\mathcal{P}_r(\mathbb{R})$  is a linear subspace of  $\mathcal{P}_s(\mathbb{R})$  whenever  $r \leq s$ .
2. Show that  $\mathcal{P}_r(\mathbb{R})$  is always a subspace of  $\mathcal{P}(\mathbb{R})$ .
3. What is the span of  $\{1 + x, 1 - x\}$  in  $\mathcal{P}(\mathbb{R})$ ?
4. What is the span of  $\{1, x^2, x^4\}$  in  $\mathcal{P}_4(\mathbb{R})$ ?
5. Find a vector that spans the subspace  $2x - 3y = 0$  of  $\mathbb{R}^2$ .
6. Find a pair of vectors that span the subspace  $x + y - 2z = 0$  of  $\mathbb{R}^3$ .
7. Verify that example 7 is indeed a vector space. What is the zero vector in this example?
8. Let  $\mathcal{E}$  be the subset of  $\mathcal{P}_r(\mathbb{R})$  defined by

$$\mathcal{E} = \{p(x) \mid p(x) \in \mathcal{P}_r(\mathbb{R}) \text{ and } p(-x) = p(x)\}$$

Show that  $\mathcal{E}$  is a linear subspace of  $\mathcal{P}_r(\mathbb{R})$ .

9. The set of all continuous functions  $y = f(x)$ ,  $-\infty < x < \infty$  satisfying the differential equation

$$y'' - y' - 2y = 0$$

is a vector space. (In fact any solution of this differential equation is a linear combination of  $y = e^{-x}$  and  $y = e^{2x}$ .)

10. The set of all continuous solutions of a linear differential equation

$$a_0(x)y^{(n)} + a_1(x)y^{(n-1)} + \cdots + a_{n-1}(x)y' + a_n(x)y = 0$$

where  $a_0(x)$  is not zero on  $[a, b]$ , and  $a_i(x)$  are continuous on  $[a, b]$ ,  $i = 1, 2, \dots, n$ , is a vector space.

11. Let  $\mathcal{D} = \{f \in \mathcal{C}(a, b) \mid f \text{ is differentiable on } (a, b)\}$ . Show that  $\mathcal{D}$  is a subspace of  $\mathcal{C}(a, b)$ .
12. Let  $\mathcal{P}(a, b) = \{\text{Polynomials of } x, a \leq x \leq b\}$ . Then  $\mathcal{P}(a, b) \subset \mathcal{D} \subset \mathcal{C}(a, b)$ . Show  $\mathcal{P}(a, b)$  is a subspace of  $\mathcal{D}$  and also a subspace of  $\mathcal{C}(a, b)$ .
13. Consider the set  $\mathcal{D}^{(n)}$  of all  $n$ -times differentiable functions on the interval  $[a, b]$ .  $\mathcal{D}^{(n)}$  is also a vector space. Is  $\mathcal{P}(a, b) \subset \mathcal{D}^{(n)}$ ?
14. Let  $S$  be a set and  $\mathcal{F}(S)$  the vector space of real-valued functions on  $S$ . If  $A, B \subset S$  show that

$$\mathcal{F}(S, A) \cap \mathcal{F}(S, B) = \mathcal{F}(S, A \cup B).$$

If  $A \subset B$  show that  $\mathcal{F}(S, A)$  contains  $\mathcal{F}(S, B)$  as a subspace.

15. Let  $S$  be a set and  $T \subset S$ . Is it true that

$$\mathcal{F}(S) = \mathcal{F}(S, T) + \mathcal{F}(S, S - T)?$$

16. Let  $\mathcal{V}$  be a vector space and  $S$  a set. Define  $\text{Fun}(S, \mathcal{V}) = \{f: S \rightarrow \mathcal{V}\}$ , that is, the set of all functions from  $S$  into  $\mathcal{V}$ . (In this notation  $\mathcal{F}(S) = \text{Fun}(S, \mathbb{R})$ .) If  $f, g \in \text{Fun}(S, \mathcal{V})$  define  $f + g: S \rightarrow \mathcal{V}$  by

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

If  $f \in \text{Fun}(S, \mathcal{V})$  and  $r \in \mathbb{R}$  define  $r \cdot f: S \rightarrow \mathcal{V}$  by

$$(r \cdot f)(s) = r \cdot f(s).$$

- (a) Show that  $\text{Fun}(S, \mathcal{V})$  equipped with these two operations becomes a vector space.
- (b) If  $\mathcal{V}$  is a subspace of  $\mathcal{W}$  show  $\text{Fun}(S, \mathcal{V})$  is a subspace of  $\text{Fun}(S, \mathcal{W})$ .
- (c) If  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces of  $\mathcal{V}$  show

$$\text{Fun}(S, \mathcal{S}) + \text{Fun}(S, \mathcal{T}) = \text{Fun}(S, \mathcal{S} + \mathcal{T}).$$

17. Let  $\mathcal{V}$  be a vector space and

$$\varphi: \mathcal{V} \rightarrow \mathbb{R}$$

such that

$$(LF 1) \quad \varphi(A + B) = \varphi(A) + \varphi(B), \quad \forall A, B \in \mathcal{V},$$

$$(LF 2) \quad \varphi(rA) = r\varphi(A), \quad \forall A \in \mathcal{V}, \forall r \in \mathbb{R}.$$

Show that  $\mathcal{N} = \{A \in \mathcal{V} \mid \varphi(A) = 0\}$  is a vector subspace of  $\mathcal{V}$ .

18. Let  $\mathcal{V}$  be a vector space and  $S$  be a set and  $s \in S$  a fixed element. Define

$$\epsilon_s: \text{Fun}(S, \mathcal{V}) \rightarrow \mathcal{V}$$

#### 4: Examples of vector spaces

by

$$e_s(\mathbf{f}) = \mathbf{f}(s).$$

Show that

$$\mathcal{N} = \{\mathbf{f} \in \text{Fun}(S, \mathcal{V}) \mid e_s(\mathbf{f}) = 0\}$$

is a subspace of  $\text{Fun}(S, \mathcal{V})$ .

19. Let  $\mathcal{V}$  and  $\mathcal{W}$  be vector spaces. Show that the cartesian product  $\mathcal{V} \times \mathcal{W}$  consisting of all the ordered pairs  $(\mathbf{A}, \mathbf{B})$ ,  $\mathbf{A} \in \mathcal{V}$ ,  $\mathbf{B} \in \mathcal{W}$  becomes a vector space if we define vector addition and scalar multiplication componentwise, that is

$$(\mathbf{A}', \mathbf{B}') + (\mathbf{A}'', \mathbf{B}'') = (\mathbf{A}' + \mathbf{A}'', \mathbf{B}' + \mathbf{B}''),$$

$$r(\mathbf{A}, \mathbf{B}) = (r\mathbf{A}, r\mathbf{B}).$$

20. Let  $\mathcal{V}$  be a vector space,  $S$  a set and  $s, t \in S$ . Define

$$\varphi: \text{Fun}(S, \mathcal{V}) \rightarrow \mathcal{V} \times \mathcal{V}$$

by

$$\varphi(\mathbf{f}) = (\mathbf{f}(s), \mathbf{f}(t)).$$

Show that

$$\mathcal{N} = \{\mathbf{f} \in \text{Fun}(S, \mathcal{V}) \mid \varphi(\mathbf{f}) = 0\}$$

is a vector subspace of  $\text{Fun}(S, \mathcal{V})$ .

21. Let  $\mathcal{V}$  be a vector space,  $\mathcal{W}$  a subspace of  $\mathcal{V}$ ,  $S$  a set and  $s \in S$ . Show that

$$\mathcal{U} = \{\mathbf{f} \in \text{Fun}(S, \mathcal{V}) \mid \mathbf{f}(s) \in \mathcal{W}\}$$

is a subspace of  $\text{Fun}(S, \mathcal{V})$ .

**Definition.** A set of vectors  $E$  is said to be *linearly dependent* if there exist distinct vectors  $\mathbf{A}_1, \dots, \mathbf{A}_k$  in  $E$  and numbers  $a_1, \dots, a_k$ , not all zero, such that

$$(*) \quad a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \dots + a_k \mathbf{A}_k = \mathbf{0}.$$

Equation  $(*)$  is called a *linear relation* between  $\mathbf{A}_1, \dots, \mathbf{A}_k$ .

EXAMPLE 1. Let  $E$  be the set of vectors

$$E = \{(0, 1, 0), (0, 1, 1), (0, 0, 1)\}$$

in  $\mathbb{R}^3$ . Then  $E$  is a linearly dependent set of vectors because

$$1(0, 1, 0) + 1(0, 0, 1) + (-1)(0, 1, 1) = (0, 0, 0).$$

EXAMPLE 2. Let  $E$  be the set of vectors

$$E = \{p(x) \mid \text{such that degree } p(x) \text{ is at most } 1\}$$

in  $\mathcal{P}_3(\mathbb{R})$ . Then  $E$  is a linearly dependent set of vectors because the vectors  $1, x, 1 + x$  belong to  $E$  and

$$1(1 + x) + (-1)x + (-1)(1) = 0.$$

**Definition.** A set of vectors  $E$  that is *not* linearly dependent is said to be *linearly independent*.

EXAMPLE 3. Let  $E$  be the vectors

$$\{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}$$

in  $\mathbb{R}^3$ . Then  $E$  is a linearly independent set of vectors. In order to prove this suppose to the contrary that  $E$  is linearly dependent. Then there must exist numbers  $a_1, a_2, a_3$ , not all zero, such that

$$a_1(1, 1, 1) + a_2(0, 1, 1) + a_3(0, 0, 1) = (0, 0, 0).$$

But if these were so, then since

$$a_1(1, 1, 1) + a_2(0, 1, 1) + a_3(0, 0, 1) = (a_1, a_1 + a_2, a_1 + a_2 + a_3)$$

we would have

$$(0, 0, 0) = (a_1, a_1 + a_2, a_1 + a_2 + a_3)$$

and hence that

$$a_1 = 0 \quad a_1 + a_2 = 0 \quad a_1 + a_2 + a_3 = 0$$

from which we see that of necessity

$$a_1 = 0, \quad a_2 = 0, \quad a_3 = 0.$$

But this contradicts our original assumption that not all of  $a_1, a_2$ , and  $a_3$  are zero. Therefore the set  $E$  cannot be linear dependent, and hence must be linearly independent.

**EXAMPLE 4.** Let  $E$  be the set of vectors  $\{1, i\}$  in  $\mathbb{C}$ . Then  $E$  is linearly independent. For if we suppose that  $\{1, i\}$  is dependent, then there are real numbers  $a_1, a_2$ , not both zero, such that

$$a_1(1) + a_2(i) = \mathbf{0}.$$

Now let  $a_1 - a_2 i$  be the conjugate complex number. Then  $\mathbf{0} = a_1 + a_2 i$  implies

$$\mathbf{0} = (a_1 + a_2 i)(a_1 - a_2 i) = a_1^2 + a_2^2$$

which is impossible. Therefore  $E$  cannot be linearly dependent and hence must be linearly independent.

*Remark.* The proof above uses the conjugate complex number. There is an alternative proof as follows. Since we consider only real vector spaces now, suppose there are two real numbers  $a_1, a_2$  so that

$$a_1 1 + a_2 i = \mathbf{0}.$$

Then

$$a_1 = -a_2 i \quad (*).$$

$a_1$  is real from assumption, and since  $a_2$  is real  $(*)$  says that  $a_1$  is also purely imaginary. The only way to avoid a contradiction is for  $a_1 = 0$  and  $a_2 = 0$ . Thus  $\{1, i\}$  is a set of linearly independent vectors.

EXAMPLE 5. Let  $E$  be the set of vectors  $\{1 + x, 1 - x\}$  in  $\mathcal{P}_1(\mathbb{R})$ . Then  $E$  is linearly independent. For suppose to the contrary that  $\{1 + x, 1 - x\}$  is linear dependent. Then there exist numbers  $a_1, a_2$ , not both zero, such that

$$a_1(1 + x) + a_2(1 - x) = \mathbf{0}.$$

Then we will have

$$\begin{aligned} \mathbf{0} &= a_1(1 + x) + a_2(1 - x) = a_1 + a_1x + a_2 - a_2x \\ &= (a_1 + a_2) + (a_1 - a_2)x. \end{aligned}$$

Remember that a polynomial is identically zero iff all its coefficients are zero. Therefore we have

$$a_1 + a_2 = 0, \quad a_1 - a_2 = 0.$$

Solving these equations we find

$$a_1 = 0, \quad a_2 = 0$$

which is a contradiction to the assumption that  $\{1 + x, 1 - x\}$  is a linearly dependent set of vectors. Therefore it is linearly independent.

EXAMPLE 6. Let  $S$  be a set. For each  $s \in S$  the *characteristic function* of  $s$  is the function

$$\chi_s : S \rightarrow \mathbb{R}$$

defined by

$$\chi_s(t) = \begin{cases} 1 & \text{if } t = s, \\ 0 & \text{if } t \neq s. \end{cases}$$

If  $s_1, \dots, s_k \in S$  are distinct points then their characteristic functions  $\chi_{s_1}, \dots, \chi_{s_k} \in \mathcal{F}(S)$  are linearly independent. To see this suppose that

$$a_1 \chi_{s_1} + \dots + a_k \chi_{s_k} = \mathbf{0}$$

is a linear relation between  $\chi_{s_1}, \dots, \chi_{s_k}$ . Then

$$\begin{aligned} 0 &= (a_1 \chi_{s_1} + \dots + a_k \chi_{s_k})(s_i) \\ &= a_1 \chi_{s_1}(s_i) + \dots + a_k \chi_{s_k}(s_i) \\ &= a_1 0 + \dots + a_{i-1} 0 + a_i 1 + a_{i+1} 0 + \dots + a_k 0 = a_i \end{aligned}$$

so  $a_1 = 0, a_2 = 0, \dots, a_k = 0$  and  $\chi_{s_1}, \dots, \chi_{s_k}$  are linearly independent.

A very quick test for a linearly dependent set is the following:

**Proposition 5.1.** *If a set of vectors  $E$  contains the vector  $\mathbf{0}$ , it is linear dependent.*

PROOF. Clearly

$$1 \cdot \mathbf{0} = \mathbf{0}$$

so letting  $A_1 = \mathbf{0} \in E$  and  $a_1 = 1, k = 1$  we satisfy the condition of linear dependence.  $\square$

## 5: Linear independence and dependence

**Corollary 5.2.** *If  $E$  is a linear subspace of  $\mathcal{V}$  then  $E$  is a linearly dependent set of vectors.*

**PROOF.** A linear subspace always contains  $\mathbf{0}$ . Apply (5.1).  $\square$

**Definition.** A vector  $A$  is said to be *linearly dependent on a set of vectors  $E$*  iff  $A \in \mathcal{L}(E)$ .

**Proposition 5.3.** *A set of vectors  $E$  is linearly dependent iff there is a vector  $A$  in  $E$  linearly dependent on the remaining vectors of  $E$ .*

**PROOF.** Suppose that  $E$  is linearly dependent. Then we may find distinct vectors  $A_1, A_2, \dots, A_k$  in  $E$ , and numbers  $a_1, \dots, a_k$ , not all zero, such that

$$a_1 A_1 + a_2 A_2 + \dots + a_k A_k = \mathbf{0}.$$

Since not all the numbers  $a_1, \dots, a_k$  are zero, we can by changing the order arrange so that  $a_1 \neq 0$ . Then we have

$$a_1 A_1 = -a_2 A_2 - a_3 A_3 - \dots - a_k A_k$$

and since  $a_1 \neq 0$ ,

$$A_1 = \frac{-a_2}{a_1} A_2 + \frac{-a_3}{a_1} A_3 + \dots + \frac{-a_k}{a_1} A_k$$

and hence  $A_1 \in \mathcal{L}(A_2, \dots, A_k)$  which shows (since  $A_1, \dots, A_k$  are distinct) that  $A_1$  is linearly dependent on the remaining vectors of  $E$ .

Conversely, if there is a vector  $A$  in  $E$  which is linearly dependent on the remaining vectors of  $E$  we may find distinct vectors  $A_2, \dots, A_k$ , different from  $A$ , such that

$$A = a_2 A_2 + \dots + a_k A_k.$$

Then

$$\mathbf{0} = (-1)A_1 + a_2 A_2 + \dots + a_k A_k$$

is a linear relation between  $A_1 = A, A_2, \dots, A_k$  showing that  $E$  is a linearly dependent set.  $\square$

**Theorem 5.4.** *If  $E$  is a finite set of vectors spanning the linear subspace  $\mathcal{U}$  of  $\mathcal{V}$ , that is  $\mathcal{L}(E) = \mathcal{U}$ , then there exists a subset  $F$  of  $E$  such that  $F$  is a linearly independent set of vectors and  $\mathcal{L}(F) = \mathcal{U} = \mathcal{L}(E)$ .*

**PROOF.** If  $E$  is linearly independent there is nothing to prove. So suppose that  $E$  is a linearly dependent set of vectors. By (5.3) there exists a vector  $A$  that is linearly dependent on the remaining vectors of  $E$ . Denote this set of remaining vectors by  $E'$ . Thus  $A \in \mathcal{L}(E')$ . Therefore  $\mathcal{L}(E') = \mathcal{L}(E)$  because (see Chapter 3, Exercises 8 and 12)  $\mathcal{L}(E') \subset \mathcal{L}(E)$  and

$$\begin{aligned} \mathcal{L}(E) &= \mathcal{L}(\{A\} \cup E') \subset \mathcal{L}(A) + \mathcal{L}(E') \subset \mathcal{L}(E') + \mathcal{L}(E') \\ &= \mathcal{L}(E'). \end{aligned}$$

Now we can repeat our argument on  $E'$ . That is, either  $E'$  is linearly independent, in which case we are done, or it is linearly dependent and we can use the preceding argument to reduce the size of  $E'$  by one vector. Since the set  $E$  of vectors that we began with is finite the theorem will follow by repeating the argument a finite number of times.  $\square$

**EXAMPLE 7.** Let  $E = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$  be vectors in  $\mathbb{R}^3$ . Find a linearly independent set  $F$  which is a subset of  $E$  such that  $\mathcal{L}(F) = \mathcal{L}(E)$ .

**Solution 1.** The proof of Theorem (5.5) suggests that we look for a vector in  $E$  linearly dependent on the remaining vectors of  $E$  and throw it away. If that doesn't work do it again, etc. Now we observe that

$$(1, 1, 1) = 1(1, 0, 0) + 1(0, 1, 0) + 1(0, 0, 1)$$

so that setting  $F = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$  we have  $\mathcal{L}(F) = \mathcal{L}(E)$ . The set  $F$ , however, is a linearly independent set of vectors and so we are done.

**Solution 2.** Proceeding as in Solution 1 we note that  $(1, 0, 0) = 1(1, 1, 1) + (-1)(0, 1, 0) + (-1)(0, 0, 1)$  and setting  $H = \{(1, 1, 1), (0, 1, 0), (0, 0, 1)\}$  we have  $\mathcal{L}(H) = \mathcal{L}(E)$ . The set  $H$ , however, is a linearly independent set of vectors and so we are done.

*Moral.* If  $E$  is a finite set of vectors in  $\mathcal{V}$  then there does *not* exist a unique subset  $F \subset E$  with  $F$  linearly independent and  $\mathcal{L}(F) = \mathcal{L}(E)$ .

**EXAMPLE 8.** Let  $S$  be a finite set and  $X = \{\chi_s | s \in S\} \subset \mathcal{F}(S)$  the set of characteristic functions of the elements of  $S$ . Then  $X$  spans  $\mathcal{F}(S)$ , because for any  $\mathbf{f} \in \mathcal{F}(S)$  we have the equality

$$\mathbf{f} = \sum_{s \in S} \mathbf{f}(s) \chi_s$$

as is seen by evaluating the right-hand side at an arbitrary element  $t \in S$  giving

$$\begin{aligned} \left[ \sum_{s \in S} \mathbf{f}(s) \chi_s \right](t) &= \sum_{s \in S} \mathbf{f}(s) \cdot \chi_s(t) \\ &= \left[ \sum_{s \neq t} \mathbf{f}(s) \cdot 0 \right] + \mathbf{f}(t) = \mathbf{f}(t) \end{aligned}$$

as required.

*Remark.* Example 8 cannot be extended to infinite sets. For if  $S$  is an infinite set then the function

$$\mathbf{f}: S \rightarrow \mathbb{R}$$

defined by  $\mathbf{f}(s) = 1, s \in S$  is not a (finite) linear combination of characteristic functions.



## 5: Linear independence and dependence

### EXERCISES

1. Which of the following sets of vectors in  $\mathbb{R}^3$  are linearly dependent and which are linearly independent?

$$E = \{(1, 1, 1), (0, 1, 0), (1, 0, 1)\}$$

$$F = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$$

$$G = \{(1, 1, 1), (1, 1, 0), (1, 0, 1)\}$$

$$H = \{(1, 0, 0), (0, 1, 0), (1, 1, 1)\}$$

$$K = \{(1, 1, 1), (0, 1, 0), (0, 0, 1)\}$$

2. Which of the following sets of vectors in  $\mathcal{P}_2(\mathbb{R})$  are linearly dependent and which are linearly independent?

$$E = \{1, x, x^2\}$$

$$F = \{1 + x, 1 - x, x^2, 1\}$$

$$G = \{x^2 - 1, x + 1, x^2 - x, x^2 + x\}$$

$$H = \{x - x^2, x^2 - x\}$$

$$K = \{1, 1 - x, 1 - x^2\}$$

3. Show that the set of vectors in  $\mathbb{R}^3$

$$E = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$$

is linearly dependent, but that any set of three of them is linearly independent.

4. Let  $\mathcal{U}$  be the subspace of  $\mathbb{R}^5$  spanned by the vectors

$$E = \{(1, 1, 0, 0, 1), (1, 1, 0, 1, 1), (0, 1, 1, 1, 1), (2, 1, -1, 0, 1)\}$$

Find a linearly independent subset  $F$  of  $E$  with  $\mathcal{L}(F) = \mathcal{U}$ .

5. Let  $\mathcal{U}$  be the subspace of  $\mathcal{P}_3(\mathbb{R})$  spanned by

$$E = \{x^3, x^3 - x^2, x^3 + x^2, x^3 - 1\}$$

find a linearly independent subset  $F$  of  $E$  spanning  $\mathcal{U}$ .

6. Under what conditions on the numbers  $a$  and  $b$  are the vectors  $(1, a), (1, b)$  linearly independent in  $\mathbb{R}^2$ .
7. Suppose that  $E$  and  $F$  are sets of vectors in  $\mathcal{V}$  with  $E \subset F$ . Prove that if  $E$  is linearly dependent then so is  $F$ .
8. Suppose that  $E$  and  $F$  are sets of vectors in  $\mathcal{V}$  with  $E \subset F$ . Prove that if  $F$  is linearly independent then so is  $E$ .
9. Show that functions  $e^x$  and  $e^{2x}$  form a set of linearly independent vectors in  $\mathcal{C}(-\infty, \infty)$ .
10. Show that

$$\left\{ \cos x, \sin x, \sin\left(x + \frac{\pi}{3}\right) \right\}$$

is a set of linearly independent vectors in  $\mathcal{C}(-\infty, \infty)$ .

11. Is the pair of complex numbers  $\alpha + \beta i, \alpha - \beta i$ , a set of linearly independent vectors in  $\mathbb{C}$  where  $\alpha, \beta$  are any nonzero real numbers?

12. Show that the set of polynomials  $E = \{x^2, 1 + x^2\}$  is a set of linearly independent vectors in  $\mathcal{P}_2(\mathbb{R})$ . What is the space spanned by  $E$ ? Is  $\mathcal{L}(E) = \mathcal{P}_2(\mathbb{R})$ ? If not, find a vector in  $\mathcal{P}_2(\mathbb{R})$  which does not belong to  $\mathcal{L}(E)$ , and show together with  $E$  they form a set of linearly independent vectors in  $\mathcal{P}_2(\mathbb{R})$ .
13. Let  $E$  be a set of vectors in the vector space  $\mathcal{V}$ . Show that  $E$  is linearly independent iff for every proper subset  $E'$  of  $E$ ,  $\mathcal{L}(E') \neq \mathcal{L}(E)$ . Equivalently, show that  $E$  is linearly dependent iff there exists a proper subset  $E'$  of  $E$  such that  $\mathcal{L}(E') = \mathcal{L}(E)$ .
14. Let  $S = \{x_1, x_2, x_3\}$  and consider the functions  $f, g, h \in \mathcal{F}(S)$  defined by
- $$\begin{aligned} f(x_1) &= 0, & f(x_2) &= 1, & f(x_3) &= 1, \\ g(x_1) &= 1, & g(x_2) &= 0, & g(x_3) &= 1, \\ h(x_1) &= 1, & h(x_2) &= 1, & h(x_3) &= 0. \end{aligned}$$

Do  $\{f, g, h\}$  form a linearly independent set?

15. Let  $E', E''$  be linearly independent sets of vectors in  $\mathcal{V}$ . Show that:
- $E' \cap E''$  is linearly independent.
  - $E' \cup E''$  is linearly independent iff  $\mathcal{L}(E') \cap \mathcal{L}(E'') = \{0\}$ .
16. Let  $A, B, C \in \mathcal{V}$  be linearly independent vectors. Show that  $\{A + B, B + C, C + A\}$  are linearly independent.
17. Let  $\mathcal{S}, \mathcal{T}$  be subspaces of a vector space  $\mathcal{V}$ ,  $S \in \mathcal{S}, T \in \mathcal{T}, S \neq 0, T \neq 0$ . Assume  $\mathcal{S} \cap \mathcal{T} = \{0\}$ . Show that  $\{S, T\}$  is a linearly independent set of vectors.
18. Show that any four vectors in  $\mathbb{R}^3$  must be linearly dependent.
19. Let  $\mathcal{S}, \mathcal{T}$  be subspaces of  $\mathcal{V}$  and  $S$  a set. Suppose  $f \in \text{Fun}(S, \mathcal{S}), g \in \text{Fun}(S, \mathcal{T})$ , and  $\mathcal{S} \cap \mathcal{T} = 0$ . Show that, regarded as vectors in  $\text{Fun}(S, \mathcal{V})$ , the vectors  $f$  and  $g$  are linearly independent.
20. Let  $A, B, C \in \mathcal{V}$ . If  $C \notin \mathcal{L}(A, B)$  show that  $A$  and  $B$  are linearly independent iff  $A + C$  and  $B + C$  are linearly independent.
21. Let  $f_1, \dots, f_k \in \mathcal{P}(\mathbb{R})$  be polynomials of positive degree. Suppose  $f_1, \dots, f_k$  are linearly independent. Does it follow that  $(d/dt)f_1, \dots, (d/dt)f_k$  are linearly independent?
22. Let  $S$  be a set. A permutation  $\sigma$  of  $S$  is a bijective (= one-to-one and onto) mapping  $\sigma: S \rightarrow S$ . For a permutation  $\sigma$  of  $S$  and  $f \in \mathcal{F}(S)$  define

$$f_\sigma: S \rightarrow \mathbb{R} \mid f_\sigma(s) = f(\sigma(s)).$$

Suppose  $f_1, \dots, f_n \in \mathcal{F}(S)$  are linearly independent. Show that  $(f_1)_\sigma, \dots, (f_n)_\sigma$  are also linearly independent.

23. If  $f_1, \dots, f_n \in \mathcal{P}_m(\mathbb{R})$  are linearly dependent then so are  $(d/dt)f_1, \dots, (d/dt)f_n$ .
24. Let  $\mathcal{V}$  be the set of all polynomials  $p(x) \in \mathcal{P}_m(\mathbb{R}), m \geq 2$  such that

$$p(0) = 0 \quad \text{and} \quad \left. \frac{d}{dx} p(x) \right|_{x=0} = 0.$$

Show that  $\mathcal{V}$  is a subspace of  $\mathcal{P}_m(\mathbb{R})$  and find a basis for this subspace. (See page 43 for the definition of basis.)

# 6 Bases and finite-dimensional vector spaces

**Definition.** A vector space  $\mathcal{V}$  is said to be *finite dimensional* iff there exists a *finite* set of vectors  $E$  with  $\mathcal{L}(E) = \mathcal{V}$ .

EXAMPLE 1.  $\mathbb{R}^k$  is finite dimensional.

To see this we introduce the vectors

$$\mathbf{E}_i = (0, 0, \dots, 1, 0, \dots, 0), \quad i = 1, \dots, k.$$

↑  
ith place

For example, if  $k = 4$  then

$$\mathbf{E}_1 = (1, 0, 0, 0),$$

$$\mathbf{E}_2 = (0, 1, 0, 0),$$

$$\mathbf{E}_3 = (0, 0, 1, 0),$$

$$\mathbf{E}_4 = (0, 0, 0, 1).$$

If  $\mathbf{A}$  is any vector in  $\mathbb{R}^k$  then  $\mathbf{A} = (a_1, \dots, a_k)$  and hence we find

$$\mathbf{A} = a_1 \mathbf{E}_1 + \dots + a_k \mathbf{E}_k.$$

For example in  $\mathbb{R}^4$  we have

$$(1, 2, 3, 4) = 1\mathbf{E}_1 + 2\mathbf{E}_2 + 3\mathbf{E}_3 + 4\mathbf{E}_4.$$

Thus  $\mathbb{R}^k = \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_k)$  and since  $E = \{\mathbf{E}_1, \dots, \mathbf{E}_k\}$  is a finite set,  $\mathbb{R}^k$  is finite dimensional.

EXAMPLE 2.  $\mathcal{P}_k(\mathbb{R})$  is finite dimensional.

To see this let  $E = \{1, x, x^2, \dots, x^k\}$ . If  $p(x)$  is any vector in  $\mathcal{P}_k(\mathbb{R})$  then

$$p(x) = a_0 1 + a_1 x + \dots + a_k x^k$$

and hence  $\mathcal{P}_k(\mathbb{R}) = \mathcal{L}(1, x, x^2, \dots, x^k)$ . Since the set  $E = \{1, x, \dots, x^k\}$  is finite  $\mathcal{P}_k(\mathbb{R})$  is finite dimensional.

**EXAMPLE 3.** Let  $S$  be a finite set then  $\mathcal{F}(S)$  is finite dimensional because the vectors (characteristic functions)  $X = \{\chi_s | s \in S\}$  span  $\mathcal{F}(S)$  (Chapter 5, Example 8).

**Theorem 6.1.** Let  $\mathcal{V}$  be a finite-dimensional vector space. Then there exists a finite set of linearly independent vectors  $F$  that spans  $\mathcal{V}$ , that is  $\mathcal{L}(F) = \mathcal{V}$ .

**PROOF.** Since  $\mathcal{V}$  is finite dimensional there is a finite set of vector  $E$  with  $\mathcal{L}(E) = \mathcal{V}$ . By (5.5) we may find a linearly independent set  $F \subset E$  with  $\mathcal{L}(F) = \mathcal{L}(E)$ . Then  $F$  is the required set of vectors.  $\square$

The property that a set of vectors be both linearly independent and span  $\mathcal{V}$  is most important and fundamental to further developments. We therefore introduce:

**Definition.** A set of vectors  $E$  in a vector space  $\mathcal{V}$  is called a *basis* for  $\mathcal{V}$  iff  $E$  is linearly independent and  $\mathcal{L}(E) = \mathcal{V}$ .

**EXAMPLE 4.** The vectors  $(1, 1), (1, 0)$  are a basis for  $\mathbb{R}^2$ .

To show this we must check that  $(1, 1), (1, 0)$  are linearly independent and span  $\mathbb{R}^2$ . To see they span, suppose  $(x, y) \in \mathbb{R}^2$ . We wish to find numbers  $a, b$  such that

$$(x, y) = a(1, 1) + b(1, 0) = (a + b, a)$$

therefore

$$y = a, \quad x = a + b,$$

so solving for  $a$  and  $b$  we get

$$a = y, \quad b = x - y,$$

and therefore

$$(x, y) = y(1, 1) + (x - y)(1, 0)$$

showing  $(x, y) \in \mathcal{L}\{(1, 1), (1, 0)\}$ , and therefore that  $\{(1, 1), (1, 0)\}$  spans  $\mathbb{R}^2$ . To show  $(1, 1), (1, 0)$  linearly independent, we suppose that

$$a(1, 1) + b(1, 0) = (0, 0)$$

is a linear relation between them. Then, as above we get

$$0 = a, \quad 0 = a + b$$

whence  $a = 0, b = 0$  as required.

EXAMPLE 5. The vectors

$$1, x - 1, (x - 2)(x - 1)$$

form a basis for  $\mathcal{P}_2(\mathbb{R})$ .

We must again check linear independence and spanning. To check linear independence suppose

$$0 = a_1 \cdot 1 + a_2(x - 1) + a_3(x - 2)(x - 1)$$

is a linear relation. Multiplying out gives

$$\begin{aligned} 0 &= a_1 + a_2(x - 1) + a_3(x^2 - 3x + 2) \\ &= a_1 + a_2x - a_2 + a_3x^2 - 3a_3x + 2a_3 \\ &= (a_1 - a_2 + 2a_3) + (a_2 - 3a_3)x + a_3x^2 \end{aligned}$$

so we must have (a polynomial is zero iff all its coefficients are zero)

$$\begin{aligned} 0 &= a_1 - a_2 + 2a_3 \\ 0 &= a_2 - 3a_3 \\ 0 &= a_3 \end{aligned}$$

which, solving in reverse order, gives

$$a_3 = 0, \quad a_2 = 0, \quad a_1 = 0$$

as required.

To show  $\{1, x - 1, (x - 2)(x - 1)\}$  is a spanning set let  $p(x) = a_0 + a_1x + a_2x^2 \in \mathcal{P}_2(\mathbb{R})$ . We are looking for numbers  $b_0, b_1, b_2$  such that

$$a_0 + a_1x + a_2x^2 = b_0 + b_1(x - 1) + b_2(x - 2)(x - 1).$$

Multiplying out and equating coefficients we obtain the simultaneous equations

$$\begin{aligned} a_0 &= b_0 - b_1 + 2b_2 \\ a_1 &= b_1 - 3b_2 \\ a_2 &= b_2 \end{aligned}$$

which may be solved to yield

$$\begin{aligned} b_2 &= a_2 \\ b_1 &= a_1 + 3a_2 \\ b_0 &= a_0 + a_1 + a_2 \end{aligned}$$

so

$$p(x) = (a_0 + a_1 + a_2)1 + (a_1 + 3a_2)(x - 1) + a_2(x - 2)(x - 1)$$

and therefore  $\{1, x - 1, (x - 2)(x - 1)\}$  spans  $\mathcal{P}_2(\mathbb{R})$ .

EXAMPLE 6. In  $\mathbb{R}^3$  the vectors  $(1, -1, 0)$ ,  $(0, 1, -1)$  are a basis for the subspace

$$\mathcal{V} = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}.$$

To see this we first show that  $(1, -1, 0)$ ,  $(0, 1, -1)$  span  $\mathcal{V}$ . So let  $(x, y, z) \in \mathcal{V}$ . Then

$$(*) \quad x + y + z = 0.$$

We wish to write

$$\begin{aligned}(x, y, z) &= a(1, -1, 0) + b(0, 1, -1) \\ &= (a, b - a, -b)\end{aligned}$$

so we must have

$$x = a, \quad y = b - a, \quad z = -b.$$

To solve for  $a$  and  $b$  use the first and last equations giving

$$a = x, \quad b = -z.$$

But is this consistent with the middle equation? Substituting gives

$$y = -z - x$$

which recalling  $(*)$  we see is valid precisely for the vectors of  $\mathcal{V}$ . Thus if  $(x, y, z) \in \mathcal{V}$  then

$$(x, y, z) = x(1, -1, 0) - z(0, 1, -1),$$

and therefore the vectors  $(1, -1, 0)$ ,  $(0, 1, -1)$  span  $\mathcal{V}$ . The check for linear independence of  $\{(1, -1, 0), (0, 1, -1)\}$  is easy and we omit it.

Example 6 is illustrated in Figure 6.1.

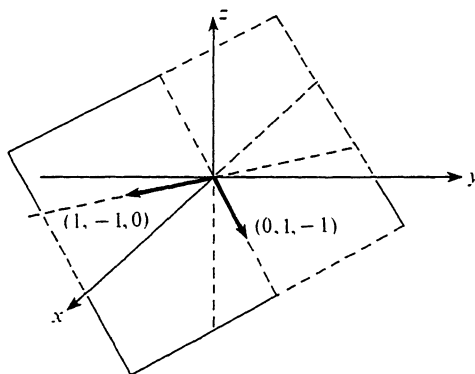


Figure 6.1

The preceding theorem shows that every finite-dimensional vector space has a basis. Actually more is true, namely any two bases for a finite-dimensional vector space contain the same number of vectors. To prove this important result we will need a few preliminary steps, the first of which is:

**Proposition 6.2.** Let  $E = \{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  be a finite set of vectors. Then  $E$  is linearly dependent iff

$$\mathbf{A}_m \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_{m-1})$$

for some  $m \leq k$ .

**PROOF.** If for some  $m \leq k$  we have  $\mathbf{A}_m \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_{m-1})$  then  $\mathbf{A}_m$  is linearly dependent on the remaining vectors of  $E$  and hence by (5.2)  $E$  is a linear dependent set of vectors.

On the other hand suppose that  $E$  is a linearly dependent set of vectors. Then there exists a linear relation

$$a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \dots + a_k \mathbf{A}_k = 0.$$

Choose  $m$  to be the largest integer between 1 and  $k$  for which  $a_m \neq 0$ . Then since  $a_{m+1} = \dots = a_k = 0$  we may write our linear relation as

$$a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \dots + a_m \mathbf{A}_m = 0$$

and since  $a_m \neq 0$  we may solve for  $\mathbf{A}_m$  obtaining

$$\mathbf{A}_m = \frac{-a_1}{a_m} \mathbf{A}_1 + \frac{-a_2}{a_m} \mathbf{A}_2 + \dots + \frac{-a_{m-1}}{a_m} \mathbf{A}_{m-1}$$

which shows that  $\mathbf{A}_m \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_{m-1})$ . □

**Proposition 6.3.** Let  $E = \{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  be a finite set of vectors in  $\mathcal{V}$ . If  $F$  is any linearly independent set of vectors in  $\mathcal{L}(E)$  then  $F$  is finite and the number of elements in  $F$  is at most  $k$ .

**PROOF.** Suppose that  $H$  is a finite subset of  $F$ . Let the vectors of  $H$  be  $\mathbf{B}_1, \dots, \mathbf{B}_s$ . Note that  $\mathbf{B}_1, \dots, \mathbf{B}_s$  are linearly independent. We must show that  $s \leq k$ . To do this we consider the set of vectors

$$G_1 = \{\mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_k\}.$$

Note that this set is linearly dependent by (5.3) since

$$\mathbf{B}_s \in \mathcal{L}(E) = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k).$$

Therefore we may apply (6.2) to conclude that  $G_1$  is linearly dependent in another way, namely there exists a vector in the set  $G_1$  linearly dependent on the preceding vectors. It cannot be  $\mathbf{B}_s$  because  $\mathbf{B}_s \neq 0$  (Why?). Therefore it must be an  $\mathbf{A}_i$ , and by rearranging terms we may assume that it is  $\mathbf{A}_k$ . That is

$$\mathbf{A}_k \in \mathcal{L}(\mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-1}).$$

Now let  $E_1 = \{\mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-1}\}$  and note  $\mathcal{L}(\mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-1}) = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k)$ . Reasoning as before we see that some vector in the set

$$G_2 = \{\mathbf{B}_{s-1}, \mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-1}\}$$

must be linearly dependent on the vectors that precede it. It cannot be  $\mathbf{B}_{s-1}$  since  $\mathbf{B}_{s-1} \neq 0$ . Nor can it be  $\mathbf{B}_s$  since  $\{\mathbf{B}_{s-1}, \mathbf{B}_s\}$  are linearly independent (Why?). Therefore it must be an  $\mathbf{A}_j$  and again by shuffling we may assume that it is  $\mathbf{A}_{k-1}$ . Thus

$$\mathbf{A}_{k-1} \in \mathcal{L}(\mathbf{B}_{s-1}, \mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-2})$$

and hence that

$$\begin{aligned}\mathcal{L}(\mathbf{B}_{s-1}, \mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-2}) &= \mathcal{L}(\mathbf{B}_s, \mathbf{A}_1, \dots, \mathbf{A}_{k-1}) \\ &= \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k)\end{aligned}$$

Suppose now that  $s > k$ . Then we may continue the above process to show that

$$\mathcal{L}(\mathbf{B}_{s-k}, \mathbf{B}_{s-k+1}, \dots, \mathbf{B}_s) = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k)$$

by repeating the above argument  $k$  times. But then

$$\mathbf{B}_{s-k-1} \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k) = \mathcal{L}(\mathbf{B}_{s-k}, \dots, \mathbf{B}_s)$$

and hence  $\{\mathbf{B}_1, \dots, \mathbf{B}_s\}$  is a linearly dependent set by (5.3) contrary to our hypothesis. Therefore  $s \leq k$  as required.  $\square$

**Theorem 6.4.** *A finite-dimensional vector space  $\mathcal{V}$  has a basis. Any two bases contain the same number of elements.*

**PROOF.** We proved the first statement in (6.1). To prove the second statement suppose that

$$E = \{\mathbf{A}_1, \dots, \mathbf{A}_n\}, \quad F = \{\mathbf{B}_1, \dots, \mathbf{B}_m\}$$

are bases for  $\mathcal{V}$ . Then  $F$  is a linearly independent set in  $\mathcal{L}(E) = \mathcal{V}$  so that  $m \leq n$  by (6.3). On the other hand  $E$  is a linearly independent set in  $\mathcal{V} = \mathcal{L}(F)$  so  $n \leq m$ , also by (6.3). Therefore  $m = n$ .  $\square$

**EXAMPLE 7.**  $\mathcal{P}(\mathbb{R})$  is not finite dimensional.

To see this consider the set  $F = \{1, x, x^2, \dots, x^n, \dots\}$  of vectors in  $\mathcal{P}(\mathbb{R})$ . This set is linearly independent. If  $\mathcal{P}(\mathbb{R})$  were finite dimensional, say  $\mathcal{P}(\mathbb{R}) = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k)$  then applying (6.4) we would conclude that  $F$  is a finite set which it is not. Therefore  $\mathcal{P}(\mathbb{R})$  is not finite dimensional.

If a vector space  $\mathcal{V}$  is not finite dimensional, then we say it is *infinite dimensional*. It is possible to combine (6.4) with the ideas of the preceding example to characterize finite- and infinite-dimensional vector spaces in terms of the sets of linearly independent vectors that they contain.

**Theorem 6.5.** *A vector space  $\mathcal{V}$  is finite dimensional iff every linearly independent set of vectors in  $\mathcal{V}$  is finite. A vector space  $\mathcal{W}$  is infinite dimensional iff there exists in  $\mathcal{W}$  an infinite linearly independent set of vectors.*



**PROOF.** If  $\mathcal{V}$  is finite dimensional then (6.3) says that every linear independent set of vectors in  $\mathcal{V}$  is finite. On the other hand suppose that every linearly independent set in  $\mathcal{V}$  is finite but the vector space  $\mathcal{V}$  is not finite dimensional. (Remember that this means  $\mathcal{V}$  is not spanned by any finite set of vectors in  $\mathcal{V}$ .) Let  $\mathbf{A}_1 \neq \mathbf{0} \in \mathcal{V}$ . Then  $\{\mathbf{A}_1\}$  is a linearly independent set of vectors. Since  $\mathcal{V}$  is not finite dimensional  $\mathcal{L}(\mathbf{A}_1) \neq \mathcal{V}$ . Therefore we may select a vector  $\mathbf{A}_2$  in  $\mathcal{V}$  that is not in  $\mathcal{L}(\mathbf{A}_1)$ . From (6.2) it follows that  $\{\mathbf{A}_1, \mathbf{A}_2\}$  is a linearly independent set. Let us repeat this process. In this way we obtain vectors  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots$  such that  $\mathbf{A}_{i+1} \notin \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_i)$  and hence by (6.2) the infinite set  $\{\mathbf{A}_1, \mathbf{A}_2, \dots\}$  of vectors in  $\mathcal{V}$  is linearly independent. This is a contradiction of the fact that every linearly independent set of vectors in  $\mathcal{V}$  is finite. Therefore the assumption that  $\mathcal{V}$  is not finite dimensional must be false, so  $\mathcal{V}$  is finite dimensional.  $\square$

**EXAMPLE 8.** If  $S$  is an infinite set then  $\mathcal{F}(S)$  is infinite dimensional.

To see this notice that in Example 8 of Chapter 5 it was shown that the set  $X = \{\chi_s | s \in S\}$  of characteristic functions is a linearly independent set of vectors in  $\mathcal{F}(S)$ . If  $S$  is infinite, this set is infinite.

Notice that (6.5) says that the concept of finite dimensionality, which we defined in terms of spanning properties of vectors, is equivalent to certain linear independence properties of vectors. It is for this reason that bases are so important because they combine the spanning and linear independence concepts.

**Definition.** Let  $\mathcal{V}$  be a finite-dimensional vector space. Then the number of vectors in any basis for  $\mathcal{V}$  is called the *dimension* of  $\mathcal{V}$ , and is written  $\dim \mathcal{V}$ .

**EXAMPLE 9.**  $\dim \mathbb{R}^k = k$ .

We have already seen that the set  $E = \{\mathbf{E}_1, \dots, \mathbf{E}_k\}$  is a basis for  $\mathbb{R}^k$ .

**EXAMPLE 10.**  $\dim \mathcal{P}_n(\mathbb{R}) = n + 1$ .

We have already seen that the set  $E = \{1, x, x^2, \dots, x^n\}$  is a basis for  $\mathcal{P}_n(\mathbb{R})$ .

**Notation.** If  $S$  is any set we write  $|S|$  for the number of elements of  $S$  when  $S$  is finite and let  $|S| = \infty$  when  $S$  is not finite.

By combining Examples 3 and 8 we obtain:

**EXAMPLE 11.**  $\dim \mathcal{F}(S) = |S|$ .

**Theorem 6.6.** Let  $\mathcal{V}$  be a finite-dimensional vector space with basis  $\mathbf{A}_1, \dots, \mathbf{A}_n$ . Then any vector  $\mathbf{A} \in \mathcal{V}$  may be written uniquely as a linear combination

$$\mathbf{A} = a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \dots + a_n \mathbf{A}_n.$$

The numbers in the sequence  $(a_1, \dots, a_n)$  are called the coordinates (or components) of  $\mathbf{A}$  relative to the ordered basis  $\mathbf{A}_1, \dots, \mathbf{A}_n$ .

PROOF. Since  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is a basis for  $\mathcal{V}$  we have  $\mathcal{V} = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n)$  so since  $\mathbf{A} \in \mathcal{V} = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n)$  there is at least one way to write  $\mathbf{A}$  as a linear combination

$$\mathbf{A} = a_1\mathbf{A}_1 + a_2\mathbf{A}_2 + \cdots + a_n\mathbf{A}_n.$$

Suppose that

$$\mathbf{A} = b_1\mathbf{A}_1 + b_2\mathbf{A}_2 + \cdots + b_n\mathbf{A}_n$$

were another way to write  $\mathbf{A}$  as a linear combination of  $\mathbf{A}_1, \dots, \mathbf{A}_n$ . Then

$$\mathbf{0} = \mathbf{A} - \mathbf{A} = (a_1 - b_1)\mathbf{A}_1 + (a_2 - b_2)\mathbf{A}_2 + \cdots + (a_n - b_n)\mathbf{A}_n.$$

Since  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is a basis it is linearly independent, hence none of the coefficients in the preceding equation can be nonzero. That is

$$a_1 - b_1 = 0, \quad a_2 - b_2 = 0, \quad \dots, \quad a_n - b_n = 0$$

and hence

$$a_1 = b_1, \quad a_2 = b_2, \quad \dots, \quad a_n = b_n$$

which establishes uniqueness. □

EXAMPLE 12. Find the coordinates of  $2 - x$  relative to the basis  $\{(1 - x), (1 + x)\}$  for  $\mathcal{P}_1(\mathbb{R})$ .

Solution. We have by (6.6)

$$2 - x = a(1 - x) + b(1 + x)$$

for suitable numbers  $a$  and  $b$ . Multiplying out gives

$$2 - x = a + b + (b - a)x$$

and equating coefficients gives

$$2 = a + b, \quad -1 = -a + b.$$

Therefore  $b = \frac{1}{2}$  and  $a = \frac{3}{2}$ . So the answer is  $(\frac{3}{2}, \frac{1}{2})$ .

EXAMPLE 13. Find the coordinates of  $(1, 1, 1)$  relative to the basis  $\{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$  for  $\mathbb{R}^3$ .

Answer.  $(1, 0, 0)$ . Think about this one!

EXAMPLE 14. Find the coordinates of  $1 + x + x^2$  relative to the basis  $\{1, x - 1, (x - 2)(x - 1)\}$  for  $\mathcal{P}_2(\mathbb{R})$ .

Solution. We require numbers  $a, b, c$  such that

$$1 + x + x^2 = a + b(x - 1) + c(x - 2)(x - 1).$$

Multiplying out gives

$$\begin{aligned} 1 + x + x^2 &= a + b(x - 1) + c(x^2 - 3x + 2) \\ &= a + bx - b + cx^2 - 3cx + 2c \\ &= (a - b + 2c) + (b - 3c)x + cx^2 \end{aligned}$$

and so equating coefficients gives

$$1 = a - b + 2c$$

$$1 = b - 3c$$

$$1 = c$$

which solve to give

$$c = 1, \quad b = 4, \quad a = 3$$

whence

$$1 + x + x^2 = 3 \cdot 1 + 4(x - 1) + (x - 2)(x - 1),$$

so that the coordinates of  $1 + x + x^2$  relative to the basis  $\{1, x - 1, (x - 2)(x - 1)\}$  are  $(3, 4, 1)$ .

**Theorem 6.7.** *Let  $\mathcal{V}$  be a finite-dimensional vector space and  $A_1, \dots, A_m$  linearly independent vectors in  $\mathcal{V}$ . Then there exist vectors  $B_1, \dots, B_n$  in  $\mathcal{V}$  such that the set  $\{A_1, \dots, A_m, B_1, \dots, B_n\}$  is a basis for  $\mathcal{V}$ .*

**PROOF.** Suppose that  $\mathcal{L}(A_1, \dots, A_m) = \mathcal{V}$ . Then there is nothing to prove. So we may suppose that  $\mathcal{L}(A_1, \dots, A_m) \neq \mathcal{V}$ . Let  $B_1$  be a vector in  $\mathcal{V}$  that is not in  $\mathcal{L}(A_1, \dots, A_m)$ . Then the set  $\{A_1, \dots, A_m, B_1\}$  is linearly independent. For if it were linearly dependent then by (6.2) some vector would be linearly dependent on the preceding ones. The vector cannot be an  $A$  because  $A_1, \dots, A_m$  are linearly independent. It cannot be  $B_1$  because we chose  $B_1$  so that  $B_1 \notin \mathcal{L}(A_1, \dots, A_m)$ . Therefore

$$\{A_1, \dots, A_m, B_1\}$$

is a linearly independent set. If  $\mathcal{L}(A_1, \dots, A_m, B_1) = \mathcal{V}$  we are done. If not we may repeat the argument starting with the vectors,  $A_1, \dots, A_m, B_1$ . In this way we obtain a set

$$\{A_1, \dots, A_m, B_1, \dots, B_n\}$$

of linearly independent vectors in  $\mathcal{V}$ . By (6.4) this process must stop when  $m + n = \dim \mathcal{V}$ , in which case

$$\mathcal{L}(A_1, \dots, A_m, B_1, \dots, B_n) = \mathcal{V}$$

as required. □

**Theorem 6.8.** *Let  $\mathcal{V}$  be a finite-dimensional vector space and  $\mathcal{U}$  a linear subspace of  $\mathcal{V}$ . Then  $\mathcal{U}$  is finite dimensional and  $\dim \mathcal{U} \leq \dim \mathcal{V}$ .*

PROOF. Suppose that  $\mathcal{U}$  is not finite dimensional. Then by (6.5) there is an infinite linearly independent set  $E$  of vectors in  $\mathcal{U}$ . But then  $E$  is certainly an infinite linearly independent set of vectors in  $\mathcal{V}$ , and hence again by (6.5)  $\mathcal{V}$  is infinite dimensional, contradiction! Therefore  $\mathcal{U}$  is finite dimensional. Let  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  be a basis for  $\mathcal{U}$  and  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  a basis for  $\mathcal{V}$ . Then by (6.3)  $m \leq n$ .  $\square$

Very often in dealing with problems of a concrete nature, we will be working in one of the finite-dimensional vector spaces  $\mathbb{R}^n$ ,  $\mathcal{P}_n(\mathbb{R})$  whose dimension we already know. If we wish to check that some set  $\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  of vectors is a basis in one of these spaces it turns out that we must only check that the set contains the correct number of vectors to be a basis and has *one* of the two characteristic properties; *linear independence*, *spanning*, of a basis. More precisely we have:

**Theorem 6.9.** *Let  $\mathcal{V}$  be a finite-dimensional vector space of dimension  $n$ . If the vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  in  $\mathcal{V}$  are linearly independent then they are a basis.*

PROOF. By (6.7) if  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is not a basis we may find vectors  $\mathbf{B}_1, \dots, \mathbf{B}_m$  so that  $\mathbf{A}_1, \dots, \mathbf{A}_n, \mathbf{B}_1, \dots, \mathbf{B}_m$  is a basis for  $\mathcal{V}$ . But then  $\dim \mathcal{V} = m + n \neq n$  a contradiction.  $\square$

**Theorem 6.10.** *Let  $\mathcal{V}$  be a finite-dimensional vector space of dimension  $n$ . If the vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  span  $\mathcal{V}$  they are a basis.*

PROOF. Since the vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  span  $\mathcal{V}$  some subset of  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is a basis for  $\mathcal{V}$  by (5.4). By rearranging the ordering we may assume that  $\{\mathbf{A}_1, \dots, \mathbf{A}_m\}$  is a basis for  $\mathcal{V}$  for some  $m \leq n$ . Then  $m = \dim \mathcal{V} = n$  by (6.5) which shows  $m = n$  and hence that  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is a basis for  $\mathcal{V}$ .  $\square$

**Corollary 6.11.** *Let  $\mathcal{V}$  be a finite-dimensional vector space of dimension  $n$ . Suppose that  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$  and  $\dim \mathcal{U} = n$ . Then  $\mathcal{U} = \mathcal{V}$ .*

PROOF. Since  $\dim \mathcal{U} = n$  there is a basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  for  $\mathcal{U}$ . By (6.9)  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  must be a basis for  $\mathcal{V}$ . Thus

$$\mathcal{U} = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n) = \mathcal{V}$$

as required.  $\square$

EXAMPLE 15. The vectors  $(1, 1, 1)$ ,  $(1, 1, 0)$ ,  $(1, 0, 0)$  are a basis for  $\mathbb{R}^3$ .

PROOF. There are three vectors above and  $\mathbb{R}^3$  has dimension 3, so by (6.9) we need only check that the three vectors are linearly independent. If

$$a_1(1, 1, 1) + a_2(1, 1, 0) + a_3(1, 0, 0) = (0, 0, 0)$$

then

$$(a_1 + a_2 + a_3, a_1 + a_2, a_1) = (0, 0, 0)$$

so

$$\begin{aligned} a_1 + a_2 + a_3 &= 0 \\ a_1 + a_2 &= 0 \\ a_1 &= 0 \end{aligned}$$

and  $a_1 = a_2 = a_3 = 0$ . So they are linearly independent.

EXAMPLE 16. The vectors  $(1, 1, -2)$ ,  $(0, 3, -3)$  are a basis for the subspace

$$\mathcal{V} = \{(x, y, z) | x + y + z = 0\}.$$

PROOF. We already know, by Example 6, that  $\dim \mathcal{V} = 2$ . So it will suffice to check

(0)  $(1, 1, -2)$  and  $(0, 3, -3)$  belong to  $\mathcal{V}$ ,

(1)  $(1, 1, -2)$  and  $(0, 3, -3)$  are linearly independent.

Condition (0) is easy since

$$1 + 1 - 2 = 0 \quad \text{and} \quad 0 + 3 - 3 = 0.$$

To check (1), suppose

$$a(1, 1, -2) + b(0, 3, -3) = \mathbf{0}$$

is a linear relation. Then

$$\begin{aligned} (0, 0, 0) &= (a, a, -2a) + (0, 3b, -3b) \\ &= (a, a + 3b, -2a - 3b) \end{aligned}$$

so equating coefficients gives

$$\begin{aligned} 0 &= a \\ 0 &= a + 3b \\ 0 &= -2a - 3b \end{aligned}$$

and  $a = 0, b = 0$  as required.

EXAMPLE 17. For what values of  $r$  are the vectors  $(r, 1, 1)$ ,  $(1, r, 1)$ ,  $(1, 1, r)$  a basis for  $\mathbb{R}^3$ ?

Solution. Since  $\dim \mathbb{R}^3 = 3$  it will suffice to find when the vectors in question are linearly independent. So suppose that

$$a(r, 1, 1) + b(1, r, 1) + c(1, 1, r) = \mathbf{0}$$

is a linear relation. Then

$$(ar + b + c, a + br + c, a + b + cr) = \mathbf{0} = (0, 0, 0),$$

so taking coordinates gives

$$\begin{aligned}ar + b + c &= 0 \\a + br + c &= 0 \\a + b + cr &= 0.\end{aligned}$$

From the last equation we get

$$(*) \quad cr = -(a + b).$$

Multiplying the second equation by  $r$  and substituting gives

$$\begin{aligned}0 &= ar + br^2 - (a + b) = a(r - 1) + b(r^2 - 1) \\&= (r - 1)[a + b(r + 1)].\end{aligned}$$

So if  $r - 1 \neq 0$  we may divide by  $r - 1$  and solve for  $a$  to get

$$(**) \quad a = -(r + 1)b.$$

Multiplying the first equation by  $r$  and substituting in  $(*)$  and  $(**)$  we obtain

$$\begin{aligned}0 &= [-(r + 1)b]r^2 + br - (a + b) \\&= -r^2(r + 1)b + rb - [-(r + 1)b + b] \\&= [-r^2(r + 1) + r + (r + 1) - 1]b \\&= (-r^3 - r^2 + r + r)b = -(r^3 + r^2 - 2r)b \\&= -r(r^2 + r - 2)b = -r(r + 2)(r - 1)b.\end{aligned}$$

So assuming  $r \neq 0, -2, 1$  we may divide by  $-r(r + 2)(r - 1)$  and conclude  $b = 0$ , where  $a = 0$  by  $(**)$  and  $c = 0$  by  $(*)$ . Therefore the vectors are independent when  $r \neq 0, 1, -2$ . For  $r = 0$ , the original system simplifies to

$$\begin{aligned}0 &= b + c, \\0 &= a + c, \\0 &= a + b.\end{aligned}$$

Substituting the first equation into the second gives

$$0 = a - b$$

and adding to the third gives

$$0 = 2a$$

giving  $a = 0$  and  $b = 0, c = 0$ , so again the vectors are independent. For  $r = 1$ , the three vectors are the same vector three times, so cannot be a basis for  $\mathbb{R}^3$ . And finally for  $r = -2$ , there is the linear relation

$$(-2, 1, 1) + (1, -2, 1) + (1, 1, -2) = 0$$

so the vectors are not linearly independent. In conclusion, the vectors  $(r, 1, 1), (1, r, 1), (1, 1, r)$  are a basis for  $\mathbb{R}^3$  if and only if  $r \neq 1, -2$ .

## EXERCISES

1. Which of the following sets of vectors are bases for
- $\mathbb{R}^3$
- ?

$$E = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$$

$$F = \{(1, 0, 1), (0, 1, 0), (1, 1, 1)\}$$

$$G = \{(-1, 1, 1), (1, -1, 1), (1, 1, -1)\}$$

$$H = \{(1, -1, 0), (1, 0, -1), (-1, 0, 1)\}$$

2. Which of the following sets of vectors are bases for
- $\mathcal{P}_2(\mathbb{R})$
- ?

$$E = \{1, 1 + x, 1 + x + x^2\}$$

$$F = \{1, (x - 1), (x - 1)(x - 2)\}$$

$$G = \{x^2 + x + 1, x^2 - x + 1, x^2 - x - 1\}$$

$$H = \{x^2 + x, x^2 + 1, x + 1\}.$$

3. Find a basis for each of the following subspaces of
- $\mathbb{R}^3$
- .

$$\mathcal{U} = \{(x, y, z) | x - z = 0\}$$

$$\mathcal{S} = \{(x, y, z) | x + y + z = 0\}$$

$$\mathcal{T} = \{(x, y, z) | x - y + z = 0\}$$

$$\mathcal{W} = \{(x, y, z) | x = 0 \text{ and } y + z = 0\}$$

4. Find a basis for the subspaces of
- $\mathcal{P}_3(\mathbb{R})$
- given by

$$\mathcal{V} = \{p(x) | p(0) = 0\}$$

$$\mathcal{S} = \left\{ p(x) \left| \frac{d}{dx} p(x) = 0 \right. \right\}$$

$$\mathcal{T} = \{p(x) | p(x) = a_0 + a_1x + a_3x^3\}$$

$$\mathcal{W} = \{p(x) | p(-x) = p(x)\}.$$

5. Calculate the dimension of each of the subspaces in (3) and (4).
6. Show that if  $\mathcal{V}$  is a vector space that is *not* finite dimensional then there exists in  $\mathcal{V}$  an infinite sequence  $\mathbf{A}_1, \mathbf{A}_2, \dots$ , of linearly independent vectors.
7. Suppose that  $\mathcal{V}$  is a finite-dimensional vector space and  $\mathcal{S}$  is a linear subspace of  $\mathcal{V}$ . Show that there exists a linear subspace  $\mathcal{T}$  of  $\mathcal{V}$  such that  $\mathcal{S} \cap \mathcal{T} = \{\mathbf{0}\}$  and  $\mathcal{S} + \mathcal{T} = \mathcal{V}$ . (Hint: Study the proof of (6.7).)
8. Suppose that  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces of a finite-dimensional vector space  $\mathcal{V}$ , and  $\mathcal{S} \cap \mathcal{T} = \{\mathbf{0}\}$ . Show that

$$\dim(\mathcal{S} + \mathcal{T}) = \dim \mathcal{S} + \dim \mathcal{T}.$$

9. Suppose that
- $\mathcal{S}$
- and
- $\mathcal{T}$
- are subspaces of a finite-dimensional vector space
- $\mathcal{V}$
- . show that

$$\dim(\mathcal{S} + \mathcal{T}) = \dim \mathcal{S} + \dim \mathcal{T} - \dim(\mathcal{S} \cap \mathcal{T})$$

(Hint: Choose a basis for  $\mathcal{S} \cap \mathcal{T}$ , extend it to a basis for  $\mathcal{S}$ , extend it to a basis for  $\mathcal{T}$ , and count.)

10. Suppose that  $\mathcal{U}$  is a subspace of a finite-dimensional vector space  $\mathcal{V}$ . Show  $\dim \mathcal{U} = \dim \mathcal{V}$  iff  $\dim \mathcal{U} = \dim \mathcal{V}$ .
11. What is the dimension of  $\mathbb{C}$  as a vector space over  $\mathbb{R}$ ?
12. Let  $\mathcal{S} = \{(a, 0, 0)\}$  and  $\mathcal{T} = \{(0, b, b)\}$  be sets of vectors in  $\mathbb{R}^3$ . Show  $\mathcal{S}$  and  $\mathcal{T}$  are subspaces. Find a basis for  $\mathcal{S} + \mathcal{T}$ .
13. Let  $\mathcal{S}$  be the subspace of  $\mathbb{R}^3$  given by

$$\mathcal{S} = \{(x, y, z) \mid y - z = 0\}.$$

Find a subspace  $\mathcal{T}$  of  $\mathbb{R}^3$  such that  $\mathcal{S} \cap \mathcal{T} = \{0\}$  and  $\mathcal{S} + \mathcal{T} = \mathbb{R}^3$ .

14. Under what conditions on the number  $a$  will the vectors  $(a, 1, 0)$ ,  $(1, a, 1)$ ,  $(0, 1, a)$  be a basis for  $\mathbb{R}^3$ ?
15. Let  $A_1, \dots, A_n$  be vectors in  $\mathcal{V}$ . Suppose that  $n = \dim \mathcal{V}$ . Show that  $\{A_1, \dots, A_n\}$  is linearly independent iff  $\dim \mathcal{L}(A_1, \dots, A_n) = n$ .
16. The equation  $y = 3x$  defines a straight line in the  $xy$ -plane. Show that if  $A, B$  are on this line then  $\vec{A}, \vec{B}$  are linearly dependent vectors.
17. Let  $A, B, C, D$  be four distinct points on a plane  $\Pi$ . Show that the vectors  $\vec{AB}, \vec{AC}, \vec{AD}$  form a set of linearly dependent vectors.
18. Let  $A, B, C$  be three points in a space. Show that  $A, B, C$  are not colinear iff  $\{\vec{AB}, \vec{AC}\}$  is a set of linearly independent vectors.
19. Let  $S = \{u, v, w\}$ . Show that the functions  $\mathbf{f}, \mathbf{g}, \mathbf{h} \in \mathcal{F}(S)$  defined by

$$\begin{array}{lll} \mathbf{f}(u) = 1 & \mathbf{g}(u) = 1 & \mathbf{h}(u) = 1 \\ \mathbf{f}(v) = 1 & \mathbf{g}(v) = 1 & \mathbf{h}(v) = 0 \\ \mathbf{f}(w) = 1 & \mathbf{g}(w) = 0 & \mathbf{h}(w) = 0 \end{array}$$

are a basis for  $\mathcal{F}(S)$ . Find the coordinates of the characteristic functions  $\chi_u, \chi_v, \chi_w$  relative to this basis.

20. Let  $S$  be a finite set and  $\mathcal{V}$  a finite-dimensional vector space. Show that  $\text{Fun}(S, \mathcal{V})$  is finite dimensional and find its dimension.
21. Let  $S$  be a finite set and  $T \subset S$ . Find the dimension of  $\mathcal{F}(S, T)$ .
22. Let  $\mathcal{V}$  be a finite-dimensional vector space over  $\mathbb{C}$ . By forgetting a part of the structure,  $\mathcal{V}$  becomes a real vector space. Show  $\dim_{\mathbb{R}} \mathcal{V} = 2 \dim_{\mathbb{C}} \mathcal{V}$  where  $\dim_{\mathbb{C}} \mathcal{V}$  is the dimension of  $\mathcal{V}$  as a complex vector space, and  $\dim_{\mathbb{R}} \mathcal{V}$  is its dimension regarded as a real vector space.



# 7 The elements of vector spaces: a summing up

Our objective in this section is to work out a number of numerical examples to illustrate and illuminate the theory of vector spaces we have developed so far.

EXAMPLE 1. Determine whether or not the vector

$$\mathbf{A} = (1, -2, 0, 3)$$

is a linear combination of the vectors

$$\mathbf{B}_1 = (3, 9, -4, -2), \quad \mathbf{B}_2 = (2, 3, 0, -1), \quad \mathbf{B}_3 = (2, -1, 2, 1).$$

That is, does  $\mathbf{A}$  belong to the linear span of  $\mathbf{B}_1$ ,  $\mathbf{B}_2$ , and  $\mathbf{B}_3$ , or in symbols is  $\mathbf{A} \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3)$ ?

*Solution.* Suppose that  $\mathbf{A} \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3)$ . Then there are numbers  $b_1, b_2, b_3$  such that

$$\mathbf{A} = b_1\mathbf{B}_1 + b_2\mathbf{B}_2 + b_3\mathbf{B}_3.$$

Therefore

$$\begin{aligned}(1, -2, 0, 3) &= b_1(3, 9, -4, -2) + b_2(2, 3, 0, -1) + b_3(2, -1, 2, 1) \\ &= (3b_1 + 2b_2 + 2b_3, 9b_1 + 3b_2 - b_3, \\ &\quad -4b_1 + 2b_3, -2b_1 - b_2 + b_3)\end{aligned}$$

and therefore

$$\begin{aligned}(1) \quad & 1 = 3b_1 + 2b_2 + 2b_3, \\ (2) \quad & -2 = 9b_1 + 3b_2 - b_3, \\ (3) \quad & 0 = -4b_1 + 0b_2 + 2b_3, \\ (4) \quad & 3 = -2b_1 - b_2 + b_3.\end{aligned}$$

Add (4) to (2) to get

$$(5) \quad 1 = 7b_1 + 2b_2.$$

(5) yields

$$(6) \quad \begin{aligned} 2b_2 &= 1 - 7b_1 \\ b_2 &= \frac{1 - 7b_1}{2}. \end{aligned}$$

(3) gives

$$(7) \quad \begin{aligned} 2b_3 &= 4b_1 \\ b_3 &= 2b_1. \end{aligned}$$

Putting (6), (7) into (4) gives

$$\begin{aligned} 3 &= -2b_1 + \frac{7b_1 - 1}{2} + 2b_1 \\ 6 &= -4b_1 + 7b_1 - 1 + 4b_1 \\ 7 &= 7b_1 \\ 1 &= b_1. \end{aligned}$$

So

$$b_1 = 1, \quad b_2 = -3, \quad b_3 = 2$$

and hence

$$\mathbf{A} = \mathbf{B}_1 - 3\mathbf{B}_2 + 2\mathbf{B}_3$$

so  $\mathbf{A}$  is a linear combination of  $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$  and hence  $\mathbf{A}$  belongs to  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3)$ .

EXAMPLE 2. Determine whether or not the vector

$$\mathbf{A} = 1 + x - 2x^2 + 4x^3$$

belongs to the subspace  $\mathcal{P}_3(\mathbb{R})$  spanned by the vectors

$$\mathbf{B}_1 = 1 - x, \quad \mathbf{B}_2 = 1 - x^2, \quad \mathbf{B}_3 = 1 - x^3.$$

*Solution.* Assume that  $\mathbf{A}$  belongs to  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3)$ . Then there are numbers  $b_1, b_2, b_3$  such that

$$\mathbf{A} = b_1\mathbf{B}_1 + b_2\mathbf{B}_2 + b_3\mathbf{B}_3$$

that is

$$\begin{aligned} 1 + x - 2x^2 + 4x^3 &= b_1(1 - x) + b_2(1 - x^2) + b_3(1 - x^3) \\ &= b_1 + b_2 + b_3 - b_1x - b_2x^2 - b_3x^3 \end{aligned}$$

and hence

$$\begin{aligned}1 &= b_1 + b_2 + b_3, \\1 &= -b_1, \\-2 &= -b_2, \\4 &= -b_3.\end{aligned}$$

Therefore the only possibility is:

$$b_1 = -1, \quad b_2 = 2, \quad b_3 = -4.$$

But then we receive the impossible equation

$$1 = -1 + 2 - 4 = -3.$$

Therefore there are no numbers  $b_1, b_2, b_3$  such that

$$\mathbf{A} = b_1 \mathbf{B}_1 + b_2 \mathbf{B}_2 + b_3 \mathbf{B}_3$$

and hence  $\mathbf{A}$  does not belong to the subspace spanned by  $\mathbf{B}_1, \mathbf{B}_2$ , and  $\mathbf{B}_3$ .

EXAMPLE 3. In  $\mathbb{R}^4$  let  $\mathcal{S}$  and  $\mathcal{T}$  be the subspaces defined by

$$\begin{aligned}\mathcal{S} &= \{\mathbf{A} = (a_1, a_2, a_3, a_4) \mid a_1 - a_2 + a_3 - a_4 = 0\} \\ \mathcal{T} &= \{\mathbf{A} = (a_1, a_2, a_3, a_4) \mid a_1 + a_2 + a_3 + a_4 = 0\}.\end{aligned}$$

Find a basis for  $\mathcal{S} \cap \mathcal{T}$ .

*Solution.* A vector  $\mathbf{A}$  in  $\mathbb{R}^4$  belongs to  $\mathcal{S} \cap \mathcal{T}$  iff

$$\begin{aligned}a_1 - a_2 + a_3 - a_4 &= 0 \\ a_1 + a_2 + a_3 + a_4 &= 0\end{aligned}$$

or

$$\begin{aligned}a_1 + a_3 &= 0 \\ a_2 + a_4 &= 0\end{aligned}$$

or

$$\begin{aligned}a_1 &= -a_3 \\ a_2 &= -a_4.\end{aligned}$$

Let

$$\begin{aligned}\mathbf{B}_1 &= (1, 0, -1, 0) \\ \mathbf{B}_2 &= (0, -1, 0, 1).\end{aligned}$$

Note that  $\mathbf{B}_1$  is a vector in  $\mathcal{S} \cap \mathcal{T}$  because

$$\begin{aligned}1 &= -(-1) \\ 0 &= -0,\end{aligned}$$

and  $\mathbf{B}_2$  is a vector in  $\mathcal{S} \cap \mathcal{T}$  because

$$\begin{aligned} 0 &= -0 \\ -1 &= -(1). \end{aligned}$$

We claim that  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is a basis for  $\mathcal{S} \cap \mathcal{T}$ . To prove this we must demonstrate two facts. First that  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is linearly independent and second that  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2) = \mathcal{S} \cap \mathcal{T}$ .

(1)  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is linearly independent. Suppose to the contrary that  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is linearly dependent. Then there are numbers  $b_1, b_2$ , not both zero, such that

$$\begin{aligned} 0 &= b_1 \mathbf{B}_1 + b_2 \mathbf{B}_2 \\ (0, 0, 0, 0) &= b_1(1, 0, -1, 0) + b_2(0, -1, 0, 1) \\ &= (b_1, -b_2, -b_1, b_2) \end{aligned}$$

and hence

$$0 = b_1, \quad 0 = -b_2$$

so  $b_1 = 0 = b_2$ , a contradiction. Hence  $\{\mathbf{B}_1, \mathbf{B}_2\}$  must be linearly independent.

(2)  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2) = \mathcal{S} \cap \mathcal{T}$ . Suppose that  $\mathbf{A} \in \mathcal{S} \cap \mathcal{T}$ . Then we have seen

$$\mathbf{A} = (a_1, a_2, -a_1, -a_2).$$

Therefore

$$\mathbf{A} = a_1(1, 0, -1, 0) - a_2(0, -1, 0, 1)$$

and hence

$$\mathbf{A} = a_1 \mathbf{B}_1 - a_2 \mathbf{B}_2$$

so  $\mathbf{A} \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$ . Therefore  $\mathcal{S} \cap \mathcal{T}$  is contained in  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$ . But since  $\mathbf{B}_1, \mathbf{B}_2$  belong to  $\mathcal{S} \cap \mathcal{T}$ ,  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$  is contained in  $\mathcal{S} \cap \mathcal{T}$ . The only conclusion possible is therefore that  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2) = \mathcal{S} \cap \mathcal{T}$ .

Hence we have found that the vectors

$$\mathbf{B}_1 = (1, 0, -1, 0), \quad \mathbf{B}_2 = (0, -1, 0, 1)$$

are a basis for  $\mathcal{S} \cap \mathcal{T}$ . Note that  $\mathcal{S} \cap \mathcal{T}$  therefore has dimension 2.

EXAMPLE 4. Show that the vectors

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

and

$$\mathbf{B}_1 = (1, 0, 0, 1), \quad \mathbf{B}_2 = (0, -1, -1, 0)$$

span the same linear subspace of  $\mathbb{R}^4$ . Find a basis for this subspace.

## 7: The elements of vector spaces: a summing up

*Solution.* If you think about it for a moment, you will see that we must prove the following facts:

$$\mathbf{A}_1 \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$$

$$\mathbf{A}_2 \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$$

$$\mathbf{A}_3 \in \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$$

$$\mathbf{B}_1 \in \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$$

$$\mathbf{B}_2 \in \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3).$$

To do this we note the following

$$\mathbf{A}_1 = \mathbf{B}_1 + \mathbf{B}_2$$

$$\mathbf{A}_2 = \mathbf{B}_1 + 2\mathbf{B}_2$$

$$\mathbf{A}_3 = -\mathbf{B}_2$$

$$\mathbf{B}_1 = \mathbf{A}_1 + \mathbf{A}_3$$

$$\mathbf{B}_2 = -\mathbf{A}_3.$$

These equations are obtained as in Example 7.1.

If we next note that  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is linearly independent we see that  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is a basis for  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2) = \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$  and hence  $\dim \mathcal{L}(\mathbf{B}_1, \mathbf{B}_2) = 2 = \dim \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$ .

EXAMPLE 5. Let  $\mathcal{S}$  and  $\mathcal{T}$  be the subspaces of  $\mathbb{R}^3$  defined by

$$\mathcal{S} = \{(x, y, z) | x = y = z\}$$

$$\mathcal{T} = \{(x, y, z) | x = 0\}.$$

Show that  $\mathcal{S} + \mathcal{T} = \mathbb{R}^3$ .

*Solution.* We must show that any vector  $\mathbf{A}$  in  $\mathbb{R}^3$  may be written in the form

$$\mathbf{A} = \mathbf{S} + \mathbf{T}$$

where  $\mathbf{S}$  is a vector in  $\mathcal{S}$  and  $\mathbf{T}$  is a vector in  $\mathcal{T}$ . So let  $\mathbf{A} = (a_1, a_2, a_3)$  and note

$$\mathbf{A} = (a_1, a_1, a_1) + (0, a_2 - a_1, a_3 - a_1).$$

The vector  $(a_1, a_1, a_1)$  belongs to  $\mathcal{S}$  and the vector  $(0, a_2 - a_1, a_3 - a_1)$  belongs to  $\mathcal{T}$ . Therefore

$$\mathbf{A} = \mathbf{S} + \mathbf{T}$$

as required.

EXAMPLE 6. Show the polynomials

$$\mathbf{A}_1 = 1, \quad \mathbf{A}_2 = t - 1, \quad \mathbf{A}_3 = (t - 1)^2 = t^2 - 2t + 1$$

form a basis for  $\mathcal{P}_2(\mathbb{R})$ . Find the coordinates of the vector

$$\mathbf{B} = 2t^2 - 5t + 6$$

relative to this ordered basis.

*Solution.* To show that  $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$  is a basis we may apply Example 5 of Chapter 6 and (6.9) to conclude that it suffices to show  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  are linearly independent. This is easy. For if

$$\mathbf{0} = a_1\mathbf{A}_1 + a_2\mathbf{A}_2 + a_3\mathbf{A}_3 = (a_1 - a_2 + a_3) + (a_2 - 2a_3)t + a_3t^2$$

then

$$a_3 = 0$$

$$a_2 - 2a_3 = 0$$

$$a_1 - a_2 + a_3 = 0$$

or

$$a_3 = 0, \quad a_2 = 0, \quad a_1 = 0$$

so  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  cannot be linearly dependent. Hence they form a basis for  $\mathcal{P}_2(\mathbb{R})$  by (6.9) and Example 5 of Chapter 6.

To find the coordinates of  $\mathbf{B}$  we write

$$\mathbf{B} = b_1\mathbf{A}_1 + b_2\mathbf{A}_2 + b_3\mathbf{A}_3$$

$$2t^2 - 5t + 6 = b_1 + b_2(t - 1) + b_3(t - 1)^2$$

$$6 - 5t + 2t^2 = b_1 - b_2 + b_3 + (b_2 - 2b_3)t + b_3t^2$$

so

$$6 = b_1 - b_2 + b_3$$

$$-5 = b_2 - 2b_3$$

$$2 = b_3.$$

Therefore

$$b_3 = 2$$

$$b_2 = -5 + 2b_3 = -1$$

$$b_1 = 6 + b_2 - b_3 = 6 - 1 - 2 = 3.$$

Therefore the coordinates of  $6 - 5t + 2t^2$  relative to the ordered basis  $\{1, (1 - t), (1 - t)^2\}$  are  $(3, -1, 2)$ .

**EXAMPLE 7.** Find the dimension of the subspace of  $\mathbb{R}^3$  given by

$$\mathcal{V} = \{(x, y, z) | x + 2y + z = 0, x + y + 2z = 0, 2x + y + z = 0\}.$$

# 7: The elements of vector spaces : a summing up

*Solution.* A vector  $\mathbf{A} = (x, y, z)$  belongs to  $\mathcal{V}$  iff

$$x + 2y + z = 0$$

$$x + y + 2z = 0$$

$$2x + y + z = 0$$

or

$$-x + y = 0$$

$$-x + z = 0$$

$$2x + y - z = 0$$

or

$$y = x, \quad z = x, \quad 2x + x - x = 0$$

or

$$y = x, \quad z = x, \quad 2x = 0$$

or

$$y = 0, \quad z = 0, \quad x = 0$$

so  $\mathcal{V} = \{\mathbf{0}\}$  and  $\dim \mathcal{V} = 0$ .

**EXAMPLE 8.** Let  $S = \{u, v, w\}$  and  $\mathcal{V} = \{\mathbf{f} \in \mathcal{F}(S) \mid \mathbf{f}(u) = \mathbf{f}(w)\}$ . Show that  $\mathcal{V}$  is a subspace of  $\mathcal{F}(S)$  and find its dimension.

*Solution.* To check that  $\mathcal{V}$  is a subspace note first of all that it is nonempty since  $\mathbf{0} \in \mathcal{V}$ . If  $\mathbf{f}, \mathbf{g} \in \mathcal{V}$  then

$$(\mathbf{f} + \mathbf{g})(u) = \mathbf{f}(u) + \mathbf{g}(u) = \mathbf{f}(w) + \mathbf{g}(w) = (\mathbf{f} + \mathbf{g})(w)$$

so  $\mathbf{f} + \mathbf{g} \in \mathcal{V}$ . Likewise for a number  $r$

$$(r\mathbf{f})(u) = r\mathbf{f}(u) = r\mathbf{f}(w) = (r\mathbf{f})(w),$$

so  $r\mathbf{f} \in \mathcal{V}$ , and therefore  $\mathcal{V}$  is a subspace of  $\mathcal{F}(S)$ .

To find the dimension of  $\mathcal{V}$  recall that  $\dim \mathcal{F}(S) = 3$ . Moreover,  $\mathcal{V} \neq \mathcal{F}(S)$  because  $\chi_u \notin \mathcal{V}$ . Therefore  $\dim \mathcal{V} < 3$ , i.e.  $\dim \mathcal{V} = 0, 1$ , or  $2$ . A little guessing shows that the two vectors  $\chi_u - \chi_w, \chi_v$  belong to  $\mathcal{V}$ . They are also linearly independent, for if

$$a(\chi_u - \chi_w) + b\chi_v = \mathbf{0}$$

is a linear relation, then

$$a\chi_u + b\chi_v - a\chi_w = \mathbf{0}$$

and evaluating at  $u$  and  $v$  gives

$$0 = a\chi_u(u) + b\chi_v(u) - a\chi_w(u) = a$$

$$0 = a\chi_u(v) + b\chi_v(v) - a\chi_w(v) = b$$

so  $a = 0, b = 0$  as required. Therefore  $\dim \mathcal{V} \geq 2$ , which since  $\dim \mathcal{V} \leq 2$ , shows  $\dim \mathcal{V} = 2$ .

## EXERCISES

1. Let  $\mathbf{A} = (1, 0, 1)$ ,  $\mathbf{B} = (-1, 1, 0)$ ,  $\mathbf{C} = (0, 1, 1)$  be three vectors of  $\mathbb{R}^3$ . Show  $\mathcal{L}(\mathbf{A}, \mathbf{B}) = \mathcal{L}(\mathbf{B}, \mathbf{C})$ .
2.  $\{\mathbf{B}_1, \mathbf{B}_2\}$  is a basis for  $\mathcal{L}(\mathbf{B}_1, \mathbf{B}_2)$  in Example 4 of this chapter. Find the coordinates (in components) of  $\mathbf{A}_1$  relative to this basis  $\{\mathbf{B}_1, \mathbf{B}_2\}$ . Do the same problem for  $\mathbf{A}_2, \mathbf{A}_3$ .
3. Let  $\mathbf{P} = (a, b)$  be a vector in  $\mathbb{R}^2$ .
  - (a) Show that  $a, b$  are the coordinates of  $\mathbf{P}$  relative to the basis  $\mathbf{E}_1 = (1, 0)$ ,  $\mathbf{E}_2 = (0, 1)$ .
  - (b) Find the coordinates of  $\mathbf{P} = (a, b)$  relative to the basis  $\mathbf{E}_1 = (1, 0)$  and  $\mathbf{F}_2 = (0, 2)$ .
  - (c) Find the coordinates of  $\mathbf{P}$  relative to the basis  $\mathbf{G} = (1, 1)$  and  $\mathbf{H} = (-1, 2)$ .
4. Are the polynomials  $\{x + x^3, 1 + x^2\}$  a set of linearly independent vectors of  $\mathcal{P}_3(\mathbb{R})$ ? If so, is  $\{x + x^3, 1 + x^2\}$  a basis for  $\mathcal{P}_3(\mathbb{R})$ ? What is the dimension of  $\mathcal{L}(x + x^3, 1 + x^2)$ ?
5. The solution space  $\mathcal{S}$  of  $x - 2y + 3z = 0$  is a subspace of  $\mathbb{R}^3$ . Show that  $\dim \mathcal{S} = 2$ . Find a basis for  $\mathcal{S}$ .
6. Let  $\mathcal{S}$  be a set. A map

$$\tau: \mathcal{S} \rightarrow \mathcal{S}$$

is called an *involution* iff

$$\tau(\tau(s)) = s, \quad \forall s \in \mathcal{S}.$$

Let  $\tau$  be an involution on the set  $\mathcal{S}$  and define

$$\mathcal{F}_+(S) = \{\mathbf{f} \in \mathcal{F}(S) \mid \mathbf{f}(\tau(s)) = \mathbf{f}(s), \forall s \in S\},$$

$$\mathcal{F}_-(S) = \{\mathbf{f} \in \mathcal{F}(S) \mid \mathbf{f}(\tau(s)) = -\mathbf{f}(s), \forall s \in S\}.$$

- (a) Show that  $\mathcal{F}_+(S)$  and  $\mathcal{F}_-(S)$  are subspaces of  $\mathcal{F}(S)$ .
- (b) Show that  $\mathcal{F}_+(S) \cap \mathcal{F}_-(S) = \{0\}$ .
- (c) Let  $\mathbf{f} \in \mathcal{F}(S)$  and define

$$\mathbf{f}_+: S \rightarrow \mathbb{R} \quad \text{by} \quad \mathbf{f}_+(s) = \mathbf{f}(s) + \mathbf{f}(\tau(s))$$

$$\mathbf{f}_-: S \rightarrow \mathbb{R} \quad \text{by} \quad \mathbf{f}_-(s) = \mathbf{f}(s) - \mathbf{f}(\tau(s)).$$

Show that  $\mathbf{f}_+ \in \mathcal{F}_+(S)$  and  $\mathbf{f}_- \in \mathcal{F}_-(S)$ .

- (d) Show that  $\mathcal{F}(S) = \mathcal{F}_+(S) + \mathcal{F}_-(S)$ .
- (e) If  $S$  is finite show that

$$\dim \mathcal{F}_+(S) + \dim \mathcal{F}_-(S) = |S|.$$

7. Let  $\mathcal{V}$  be a vector space, and  $\mathcal{S}, \mathcal{T}$  finite-dimensional subspaces of  $\mathcal{V}$ . If  $\mathcal{S}$  and  $\mathcal{T}$  have the same dimension, and  $\mathcal{S} \subset \mathcal{T}$  show that  $\mathcal{S} = \mathcal{T}$ .



7: The elements of vector spaces: a summing up

8. Let  $\mathcal{S}, \mathcal{T}, \mathcal{U}$  be subspaces of a vector space  $V$ . Is it always true that

$$(\mathcal{S} + \mathcal{T}) \cap \mathcal{U} = (\mathcal{S} \cap \mathcal{U}) + (\mathcal{T} \cap \mathcal{U})?$$

9. Let  $S = \{s, t, u\}$  and let  $\tau: S \rightarrow S$  be the involution obtained by interchanging  $s$  and  $u$  and leaving  $t$  alone, that is

$$\tau(s) = u, \quad \tau(t) = t, \quad \tau(u) = s.$$

Find bases for  $\mathcal{F}_+(S)$  and  $\mathcal{F}_-(S)$ .

10. Find a basis for the subspace of  $\mathcal{P}_4(\mathbb{R})$  spanned by the polynomials  $f \in \mathcal{P}_4(\mathbb{R})$  that satisfy the equation

$$f(z) = z^4 f\left(\frac{1}{z}\right).$$

In physics when vectors are used by fixing a coordinate system and employing components, it is very important to understand how the components will change if the coordinate system is changed. A change of coordinates is best expressed in terms of transformations of vectors. The simplest such transformations are the linear ones.

In the calculus when the study of curves in the plane and space are introduced the student was confronted with "vector-valued functions of a scalar." The simplest such functions were the linear ones.

In this chapter we will begin the study of linear transformations, the generalization of these linear functions to higher dimensions. A typical linear transformation from  $\mathbb{R}^2$  to  $\mathbb{R}^3$  is given by a system of linear equations

$$y_1 = a_{11}x_1 + a_{12}x_2$$

$$y_2 = a_{21}x_1 + a_{22}x_2$$

$$y_3 = a_{31}x_1 + a_{32}x_2$$

where  $\mathbf{X} = (x_1, x_2)$  are the components of a vector  $\mathbf{X}$  of  $\mathbb{R}^2$  and  $\mathbf{Y} = (y_1, y_2, y_3)$  are the components of the vector of  $\mathbb{R}^3$  that corresponds to  $\mathbf{X}$  under the particular transformation. The numbers  $a_{11}, \dots$  are fixed and determine the transformation. The rectangular array of numbers

$$a_{11}, \quad a_{12}$$

$$a_{21}, \quad a_{22}$$

$$a_{31}, \quad a_{32}$$

is called the matrix of the transformation (relative to the usual bases for  $\mathbb{R}^2$  and  $\mathbb{R}^3$ ). It is cumbersome and inconvenient to deal with linear transforma-

tions in terms of a system of linear equations. Instead we shall define them as functions between vector spaces with two simple properties.

**Definition.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be vector spaces. A *linear transformation*  $T$  from  $\mathcal{V}$  to  $\mathcal{W}$ , written  $T: \mathcal{V} \rightarrow \mathcal{W}$ , is a function that assigns to each vector  $A$  in  $\mathcal{V}$  a vector  $T(A)$  in  $\mathcal{W}$  such that the following two properties hold:

- (1)  $T(A + B) = T(A) + T(B)$  for all  $A, B \in \mathcal{V}$ .
- (2)  $T(aA) = aT(A)$  for all numbers  $a$  and all  $A \in \mathcal{V}$ .

We might paraphrase this definition by saying that a linear transformation from  $\mathcal{V}$  to  $\mathcal{W}$  is a function that preserves vector addition and scalar multiplication.

It is to be emphasized that as soon as the dimension of  $\mathcal{V}$  and  $\mathcal{W}$  exceed 1 there are an enormous number of linear transformations from  $\mathcal{V}$  to  $\mathcal{W}$ . Their variety is what makes the subject of linear algebra really interesting. The classification of linear transformations is what will concern us throughout the remainder of this course, although we will only scratch the surface of classification theory!

Let us begin by examining some of the boundless variety of linear transformations.

**EXAMPLE 1.** In  $\mathbb{R}^2$  consider the function

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

given by

$$T(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$$

where  $\theta$  is a fixed number  $0 \leq \theta < 2\pi$ . We claim that  $T$  is a linear transformation. To prove this we must verify that  $T$  possesses the two properties characteristic of linear transformations. So suppose that  $A = (a_1, a_2)$ ,  $B = (b_1, b_2)$  are vectors in  $\mathbb{R}^2$ . Then

$$\begin{aligned} T(A + B) &= T(a_1 + b_1, a_2 + b_2) \\ &= ((a_1 + b_1)\cos \theta - (a_2 + b_2)\sin \theta, \\ &\quad (a_1 + b_1)\sin \theta + (a_2 + b_2)\cos \theta) \\ &= (a_1 \cos \theta - a_2 \sin \theta + b_1 \cos \theta - b_2 \sin \theta, \\ &\quad a_1 \sin \theta + a_2 \cos \theta + b_1 \sin \theta + b_2 \cos \theta) \\ &= (a_1 \cos \theta - a_2 \sin \theta, a_1 \sin \theta + a_2 \cos \theta) \\ &\quad + (b_1 \cos \theta - b_2 \sin \theta, b_1 \sin \theta + b_2 \cos \theta) \\ &= T(A) + T(B), \end{aligned}$$

and hence the first characteristic property of a linear transformation is satisfied. To verify the second property we suppose that  $a$  is a number and compute

$$\begin{aligned}
 T(a\mathbf{A}) &= T(aa_1, aa_2) \\
 &= (aa_1 \cos \theta - aa_2 \sin \theta, aa_1 \sin \theta + aa_2 \cos \theta) \\
 &= (a(a_1 \cos \theta - a_2 \sin \theta), a(a_1 \sin \theta + a_2 \cos \theta)) \\
 &= a(a_1 \cos \theta - a_2 \sin \theta, a_1 \sin \theta + a_2 \cos \theta) \\
 &= aT(\mathbf{A}),
 \end{aligned}$$

so that the second characteristic property of a linear transformation holds for  $T$ .

Therefore  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a linear transformation.

Geometrically  $T$  has a very simple interpretation. It is a rotation of the plane  $\mathbb{R}^2$  through the angle  $\theta$  (in radian measure) in the counterclockwise direction as shown in Figure 8.1. In studying less concretely given linear transformations we will often seek such simple geometric interpretations.

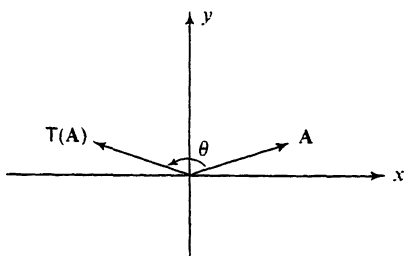


Figure 8.1

EXAMPLE 2. Consider the function  $\mathbb{R}^3$  to  $\mathbb{R}^2$  given by

$$T(x, y, z) = (x, y)$$

We claim that  $T$  is a linear transformation. To prove this we must verify that  $T$  has the two characteristic properties of a linear transformation. This is done in the following computations ( $\mathbf{A} = (a_1, a_2, a_3)$ ,  $\mathbf{B} = (b_1, b_2, b_3)$ )

$$\begin{aligned}
 T(\mathbf{A} + \mathbf{B}) &= T(a_1 + b_1, a_2 + b_2, a_3 + b_3) \\
 &= (a_1 + b_1, a_2 + b_2) = (a_1, a_2) + (b_1, b_2) \\
 &= T(\mathbf{A}) + T(\mathbf{B})
 \end{aligned}$$

and

$$T(a\mathbf{A}) = T(aa_1, aa_2, aa_3) = (aa_1, aa_2) = a(a_1, a_2) = aT(\mathbf{A})$$

(where  $a$  is a number).

## 8: Linear transformations

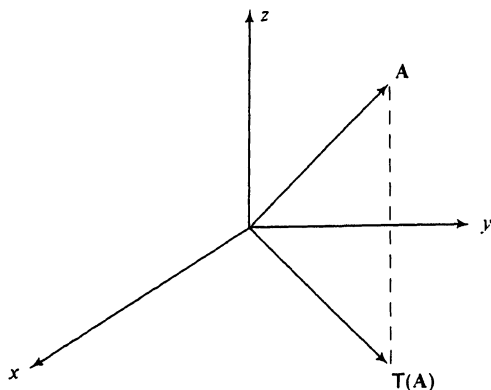


Figure 8.2

EXAMPLE 3. A slight variation of Example 2 is given by the function  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  such that

$$T(x, y, z) = (x, y, 0).$$

It is routine to check that  $T$  is a linear transformation. Geometrically,  $T$  is projection onto the  $xy$  plane, as in Figure 8.2. It is possible to combine the projection of Example 3 with the rotation of Example 1 to obtain a more complex linear transformation. Many of the linear transformations from  $\mathbb{R}^3$  to  $\mathbb{R}^2$  are of this composite form.

EXAMPLE 4. Define a function  $T$  from  $\mathbb{R}^3$  to  $\mathcal{P}_2(\mathbb{R})$  by

$$T(a_1, a_2, a_3) = a_1 + a_2x + a_3x^2$$

It is routine to verify that  $T$  is a linear transformation. Presently we will be able to use it to show that  $\mathbb{R}^3$  and  $\mathcal{P}_2(\mathbb{R})$  are in some sense the “same” vector space.

Before turning to some more examples of linear transformations let us see what additional properties they must have.

**Proposition 8.1.** *Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be a linear transformation. Then  $T(\mathbf{0}) = \mathbf{0}$ . (Note that we have used the same symbol  $\mathbf{0}$  to denote the zero vector of  $\mathcal{V}$  and of  $\mathcal{W}$ .)*

PROOF. We have

$$T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0}$$

by (2.1). □

**Proposition 8.2.** *Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be a linear transformation. If  $A_1, \dots, A_n$  are vectors in  $\mathcal{V}$  and  $a_1, \dots, a_n$  are numbers, then*

$$T(a_1A_1 + a_2A_2 + \dots + a_nA_n) = a_1T(A_1) + a_2T(A_2) + \dots + a_nT(A_n)$$

PROOF. We apply the fact that  $T$  preserves vector sums to write

$$T(a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n) = T(a_1 \mathbf{A}_1) + \cdots + T(a_n \mathbf{A}_n)$$

and now we apply the fact that  $T$  preserves scalar products to write

$$T(a_1 \mathbf{A}_1) + \cdots + T(a_n \mathbf{A}_n) = a_1 T(\mathbf{A}_1) + a_2 T(\mathbf{A}_2) + \cdots + a_n T(\mathbf{A}_n)$$

which is the desired conclusion.  $\square$

**Proposition 8.3.** *Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be a linear transformation. Suppose that  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$  and set*

$$T(\mathcal{U}) = \{T(\mathbf{A}) \in \mathcal{W} \mid \mathbf{A} \in \mathcal{U}\}$$

*(that is  $T(\mathcal{U})$  is the set of all vectors in  $\mathcal{W}$  that are of the form  $T(\mathbf{A})$  for some vector  $\mathbf{A}$  in  $\mathcal{U}$ ). Then  $T(\mathcal{U})$  is a linear subspace of  $\mathcal{W}$ .*

PROOF. Suppose that  $\mathbf{A}', \mathbf{B}'$  are vectors in  $T(\mathcal{U})$ . Then there are vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{U}$  such that

$$\mathbf{A}' = T(\mathbf{A}) \quad \text{and} \quad \mathbf{B}' = T(\mathbf{B}).$$

Then  $\mathbf{A} + \mathbf{B}$  is in  $\mathcal{U}$  since  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$ . Therefore  $\mathbf{C} = T(\mathbf{A} + \mathbf{B})$  is in  $T(\mathcal{U})$ . But since  $T$  is a linear transformation

$$\mathbf{C} = T(\mathbf{A} + \mathbf{B}) = T(\mathbf{A}) + T(\mathbf{B}) = \mathbf{A}' + \mathbf{B}'$$

so we see that  $\mathbf{A}' + \mathbf{B}'$  is in  $T(\mathcal{U})$ .

Next we suppose that  $a$  is a number. Then since  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$ ,  $a\mathbf{A} \in \mathcal{U}$ . Therefore  $\mathbf{C} = T(a\mathbf{A}) \in T(\mathcal{U})$ . But since  $T$  is a linear transformation

$$\mathbf{C} = T(a\mathbf{A}) = aT(\mathbf{A}) = a\mathbf{A}'$$

so that  $a\mathbf{A}'$  is in  $T(\mathcal{U})$  also.  $\square$

**Proposition 8.4.** *Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation and  $D$  is a set of vectors in  $\mathcal{V}$ . Then*

$$T(\mathcal{L}(D)) = \mathcal{L}(T(D))$$

*where  $T(D)$  denotes the set of vectors in  $\mathcal{W}$  of the form  $T(\mathbf{A})$  for some vector  $\mathbf{A}$  in  $D$ .*

PROOF. The proof is similar to that of (8.3). We must show that each vector in  $T(\mathcal{L}(D))$  is in  $\mathcal{L}(T(D))$  and conversely. A vector in  $T(\mathcal{L}(D))$  is a vector  $T(\mathbf{A})$  for some  $\mathbf{A}$  in  $\mathcal{L}(D)$ . Then there are vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  in  $D$  so that

$$\mathbf{A} = a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n$$

for suitable numbers  $a_1, \dots, a_n$ . Therefore by (8.2)

$$T(\mathbf{A}) = a_1 T(\mathbf{A}_1) + \cdots + a_n T(\mathbf{A}_n)$$

and since the right hand side belongs to  $\mathcal{L}(T(D))$  we see that  $T(\mathcal{L}(D)) \subset \mathcal{L}(T(D))$ . To prove the converse suppose that  $A'$  belongs to  $\mathcal{L}(T(D))$ . Then there are vectors  $A'_1, \dots, A'_n$  in  $T(D)$ , and numbers  $a_1, \dots, a_n$  such that

$$A' = a_1 A'_1 + \dots + a_n A'_n.$$

There are also vectors  $A_1, \dots, A_n$  in  $D$  with

$$A'_1 = T(A_1), \quad A'_2 = T(A_2), \quad \dots, \quad A'_n = T(A_n).$$

Let

$$A = a_1 A_1 + a_2 A_2 + \dots + a_n A_n.$$

Then  $A$  is a vector in  $\mathcal{L}(D)$  and by (8.2)

$$\begin{aligned} T(A) &= T(a_1 A_1 + \dots + a_n A_n) \\ &= a_1 T(A_1) + \dots + a_n T(A_n) \\ &= a_1 A'_1 + \dots + a_n A'_n = A' \end{aligned}$$

so that  $A'$  belongs to  $T(\mathcal{L}(D))$ , that is  $\mathcal{L}(T(D)) \subset T(\mathcal{L}(D))$  as was to be shown.  $\square$

**Proposition 8.5.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  and  $S: \mathcal{W} \rightarrow \mathcal{U}$  are linear transformations. Then

$$S \cdot T: \mathcal{V} \rightarrow \mathcal{U} \quad \text{defined by} \quad (S \cdot T)(A) = S(T(A))$$

is a linear transformation.

**PROOF.** We must verify that  $S \cdot T$  has the two characteristic properties of a linear transformation. So suppose that  $A, B$  belong to  $\mathcal{V}$ . Then using the facts that  $S$  and  $T$  are linear transformations we find

$$\begin{aligned} S \cdot T(A + B) &= S(T(A + B)) \\ &= S(T(A) + T(B)) \\ &= S(T(A)) + S(T(B)) \\ &= S \cdot T(A) + S \cdot T(B). \end{aligned}$$

Likewise for any number  $a$  we find

$$\begin{aligned} S \cdot T(aA) &= S(T(aA)) = S(aT(A)) \\ &= aS(T(A)) = aS \cdot T(A) \end{aligned}$$

and hence  $S \cdot T$  is a linear transformation.  $\square$

**Definition.** Suppose that  $S, T: \mathcal{V} \rightarrow \mathcal{W}$  are both linear transformations.

Let

$$S + T: \mathcal{V} \rightarrow \mathcal{W}$$

be the function defined by

$$(S + T)(A) = S(A) + T(A).$$

**Proposition 8.6.** *If  $S, T: \mathcal{V} \rightarrow \mathcal{W}$  are linear transformations then so is  $S + T: \mathcal{V} \rightarrow \mathcal{W}$ .*

PROOF. We suppose that  $A, B$  belong to  $\mathcal{V}$ . Then we compute as follows

$$\begin{aligned}(S + T)(A + B) &= S(A + B) + T(A + B) \\ &= S(A) + S(B) + T(A) + T(B) \\ &= S(A) + T(A) + S(B) + T(B) \\ &= (S + T)(A) + (S + T)(B).\end{aligned}$$

For any number  $a$  we have

$$\begin{aligned}(S + T)(aA) &= S(aA) + T(aA) \\ &= aS(A) + aT(A) \\ &= a(S(A) + T(A)) \\ &= a((S + T)(A)).\end{aligned}$$

Hence the function  $S + T: \mathcal{V} \rightarrow \mathcal{W}$  possesses the two characteristic properties of a linear transformation.  $\square$

**Definition.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation and  $a$  is a number. Define the function

$$aT: \mathcal{V} \rightarrow \mathcal{W}$$

by

$$(aT)(A) = aT(A).$$

**Proposition 8.7.** *If  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation and  $a$  is a number, then  $aT: \mathcal{V} \rightarrow \mathcal{W}$  is also a linear transformation.*

PROOF. The proof is similar to that of (8.6) and is left to the reader.  $\square$

**Definition.** If  $\mathcal{V}$  and  $\mathcal{W}$  are vector spaces, the zero transformation, denoted by  $0: \mathcal{V} \rightarrow \mathcal{W}$ , is defined by  $0(A) = 0$  for all  $A$  in  $\mathcal{V}$ .

Note that taken together (8.6) and (8.7) say that we are on our way to showing that the set of all linear transformations from  $\mathcal{V}$  to  $\mathcal{W}$  is **again a vector space**. This is indeed true, and a very important theorem.

**Theorem 8.8.** *Let  $\mathcal{V}$  and  $\mathcal{W}$  be vector spaces and  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  the set of all linear transformations from  $\mathcal{V}$  to  $\mathcal{W}$ . Then  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  is a vector space.*

PROOF. We must verify that Axioms 1–8 of Chapter 2 are satisfied for the procedure of adding linear transformations and multiplying them by numbers as defined above. The zero transformation defined above is the zero vector in Axiom 3. The negative of the transformation  $T$  is  $(-1)T$  in Axiom 4. The verifications of the details are routine.  $\square$



Note that (8.8) should finally lay to rest the idea that a vector is a quantity with direction and magnitude. The vector spaces  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  are enormously important and we will see more and more of them.

**Definition.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation. The *image* of  $T$ , denoted by  $\text{Im } T$ , is defined by  $\text{Im } T = T(\mathcal{V})$ . The *kernel* of  $T$  denoted by  $\ker T$  is defined by  $\ker T = \{A \in \mathcal{V} \mid T(A) = \mathbf{0}\}$ .

EXAMPLE 5. Let  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^3$  be the linear transformation given by

$$T(x, y, z, w) = (x + y, z + w, 0).$$

Find the image and kernel of  $T$ .

*Solution.* To find the image of  $T$  is easy. It is the  $xy$ -plane, that is

$$\text{Im } T = \{(u, v, 0) \in \mathbb{R}^3\} = \mathcal{V}_{xy}.$$

Because, given any vector  $(u, v, 0)$  one checks

$$T(u, 0, v, 0) = (u, v, 0),$$

showing  $\text{Im } T \supset \mathcal{V}_{xy}$ .

The last coordinate of any vector in  $\text{Im } T$  must be zero, so  $\text{Im } T \subset \mathcal{V}_{xy}$  and it follows that  $\text{Im } T = \mathcal{V}_{xy}$ .

The kernel of  $T$  consists of all vectors  $(x, y, z, w) \in \mathbb{R}^4$  such that

$$(0, 0, 0) = (x + y, z + w, 0).$$

So one quickly sees

$$\ker T = \{(a, -a, b, -b) \in \mathbb{R}^4 \mid a, b \in \mathbb{R}\}.$$

Note that both  $\text{Im } T$  and  $\ker T$  are subspaces of  $\mathbb{R}^3$  and  $\mathbb{R}^4$  respectively. This is not a special property of this example as the next result shows.

**Proposition 8.9.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation. Then  $\text{Im } T$  is a linear subspace of  $\mathcal{W}$  and  $\ker T$  is a linear subspace of  $\mathcal{V}$ .

PROOF. The fact that  $\text{Im } T$  is a linear subspace of  $\mathcal{W}$  is the special case of (8.3) where  $\mathcal{U} = \mathcal{V}$ . To show that  $\ker T$  is a linear subspace of  $\mathcal{V}$  we suppose that  $A, B \in \ker T$ . Then

$$T(A + B) = T(A) + T(B) = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

so that  $A + B \in \ker T$ . Likewise

$$T(aA) = aT(A) = a \cdot \mathbf{0} = \mathbf{0}$$

for any number  $a$ , and hence  $\ker T$  is a linear subspace of  $\mathcal{V}$ . □

**Proposition 8.10.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation. If  $\mathcal{V}$  is finite dimensional, then  $\ker T$  and  $\operatorname{Im} T$  are finite dimensional and

$$\dim \mathcal{V} = \dim \operatorname{Im} T + \dim \ker T.$$

*Remark.* The proof of (8.10) is important because it indicates the process whereby one solves a system of linear homogeneous equations.

**PROOF.** Since  $\mathcal{V}$  is finite dimensional  $\ker T$  is finite dimensional by (6.8). Choose a basis  $A_1, \dots, A_s$  for  $\ker T$ . By (6.7) we may find vectors  $B_1, \dots, B_t$  so that  $A_1, \dots, A_s, B_1, \dots, B_t$  is a basis for  $\mathcal{V}$ . By (8.4) we have

$$\begin{aligned} \operatorname{Im} T &= T(\mathcal{V}) = T(\mathcal{L}(A_1, \dots, A_s, B_1, \dots, B_t)) \\ &= \mathcal{L}(T(A_1), \dots, T(A_s), T(B_1), \dots, T(B_t)) \\ &= \mathcal{L}(0, \dots, 0, T(B_1), \dots, T(B_t)) \\ &= \mathcal{L}(T(B_1), \dots, T(B_t)) \end{aligned}$$

and hence  $\operatorname{Im} T$  is finite dimensional. We claim that the vectors  $T(B_1), \dots, T(B_t)$  are a basis for  $\operatorname{Im} T$ . The preceding equation shows that  $T(B_1), \dots, T(B_t)$  spans  $\operatorname{Im} T$  and so we must show that they are linearly independent. To this end suppose that there are numbers  $b_1, \dots, b_t$ , not all zero, so that

$$b_1 T(B_1) + b_2 T(B_2) + \dots + b_t T(B_t) = 0.$$

Let

$$B = b_1 B_1 + \dots + b_t B_t.$$

Then  $B$  is a vector in  $\mathcal{V}$ . By (8.2)

$$T(B) = b_1 T(B_1) + \dots + b_t T(B_t) = 0$$

so  $B \in \ker T$ . Since the vectors  $A_1, \dots, A_s$  are a basis for  $\ker T$  we may find numbers  $a_1, \dots, a_s$  so that

$$B = a_1 A_1 + \dots + a_s A_s$$

or

$$b_1 B_1 + \dots + b_t B_t = a_1 A_1 + \dots + a_s A_s$$

which may be written

$$0 = a_1 A_1 + \dots + a_s A_s - b_1 B_1 - \dots - b_t B_t$$

which is a linear relation between the vectors  $A_1, \dots, A_s, B_1, \dots, B_t$ . Since not all of  $b_1, \dots, b_t$  are zero this contradicts the fact that  $A_1, \dots, A_s, B_1, \dots, B_t$  are a basis for  $\mathcal{V}$ . Hence  $T(B_1), \dots, T(B_t)$  is a basis for  $\operatorname{Im} T$ . Thus

$$\dim \ker T = s$$

$$\dim \operatorname{Im} T = t$$

$$\dim \mathcal{V} = s + t$$

as was to be shown. □

EXAMPLE 6. Let  $D: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  be the linear transformation defined by

$$D(p(x)) = \frac{d}{dx} p(x).$$

The fact that  $D$  is a linear transformation is a consequence of elementary facts from the calculus.

Note that

$$D(\mathcal{P}_n(\mathbb{R})) = \mathcal{P}_{n-1}(\mathbb{R}).$$

The kernel of  $D$  consists of all those polynomials with  $(d/dx)p(x) = 0$ . The only such polynomials are the constant polynomials. Thus

$$\ker D = \mathcal{P}_0(\mathbb{R}).$$

Let  $L: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  be the linear transformation defined by

$$L(p(x)) = \int_0^x p(t) dt.$$

The fact that  $L$  is a linear transformation is also a consequence of elementary facts from the calculus. Note that

$$L(\mathcal{P}_n(\mathbb{R})) = \{p(x) \in \mathcal{P}_{n+1}(\mathbb{R}) \mid p(0) = 0\},$$

that is  $L(\mathcal{P}_n(\mathbb{R}))$  consists of those polynomials of degree  $n + 1$  whose constant term is zero. Thus

$$\text{Im } L = \{p(x) \mid p(0) = 0\},$$

that is the image of  $L$  consists of all the polynomials with zero constant term.

To calculate the kernel of  $L$  we recall that

$$D(L(p(x))) = \frac{d}{dx} \int_0^x p(t) dt = p(x).$$

Therefore, in our notation for composition of linear transformations

$$(*) \quad D \cdot L(p(x)) = p(x).$$

If, therefore,  $L(p(x)) = 0$  then

$$p(x) = DL(p(x)) = D(0) = 0.$$

Hence  $p(x) \in \ker L$  iff  $p(x) = 0$ , that is

$$\ker L = \{0\}.$$

Likewise Equation  $(*)$  shows that

$$\text{Im } D = \mathcal{P}(\mathbb{R}).$$

For if  $p(x) \in \mathcal{P}(\mathbb{R})$  then

$$p(x) = D(L(x))$$

showing  $p(x) \in D(\mathcal{P}(\mathbb{R}))$ .

EXAMPLE 7. A slight variation of the preceding example may be constructed by considering

$$D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$$

given again by

$$D(p(x)) = \frac{d}{dx} p(x).$$

This is possible because  $\deg D(p(x)) \leq \deg p(x)$ . (This is not possible for  $L$  since  $\deg L(p(x)) = 1 + \deg p(x)$ .) We may therefore compose  $D$  with itself to form

$$\begin{aligned} D^2 &= D \cdot D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R}) \\ D^3 &= D \cdot D^2: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R}) \\ &\vdots \\ D^m &= D \cdot D^{m-1}: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R}). \end{aligned}$$

Notice that

$$\begin{aligned} \ker D &= \mathcal{P}_0(\mathbb{R}) \\ \operatorname{Im} D &= \mathcal{P}_{n-1}(\mathbb{R}) \\ \ker D^2 &= \mathcal{P}_1(\mathbb{R}) \\ \operatorname{Im} D^2 &= \mathcal{P}_{n-2}(\mathbb{R}) \end{aligned}$$

and more generally

$$\begin{aligned} \ker D^m &= \mathcal{P}_{m-1}(\mathbb{R}) \\ \operatorname{Im} D^m &= \mathcal{P}_{n-m}(\mathbb{R}). \end{aligned}$$

In particular

$$\begin{aligned} \ker D^{n+1} &= \mathcal{P}_n(\mathbb{R}) \\ \operatorname{Im} D^{n+1} &= \{0\}. \end{aligned}$$

That is

$$D^{n+1} = 0: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R}).$$

The linear transformation  $D$  is called **nilpotent** because of this latter property. The nilpotent transformations play a central role in the more advanced theory of linear algebra.

**Definition.** A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  is said to be an *isomorphism* iff there exists a linear transformation  $S: \mathcal{W} \rightarrow \mathcal{V}$  such that

$$\begin{aligned} S \cdot T(A) &= A \quad \text{for all } A \in \mathcal{V} \\ T \cdot S(B) &= B \quad \text{for all } B \in \mathcal{W}. \end{aligned}$$

If there is a linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  that is an isomorphism, then we say that  $\mathcal{V}$  and  $\mathcal{W}$  are *isomorphic*.

If  $\mathcal{V}$  and  $\mathcal{W}$  are isomorphic vector spaces then in some sense they are the same. More precisely an isomorphism  $T$  will translate true theorems in  $\mathcal{V}$  into true theorems in  $\mathcal{W}$  and conversely.

**EXAMPLE 8.** The vector spaces  $\mathbb{R}^{n+1}$  and  $\mathcal{P}_n(\mathbb{R})$  are isomorphic for all non-negative integers  $n$ .

To see this we must define a linear transformation

$$T: \mathbb{R}^{n+1} \rightarrow \mathcal{P}_n(\mathbb{R})$$

and a linear transformation

$$S: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathbb{R}^{n+1}$$

such that

$$\begin{aligned} ST(A) &= A \quad \text{for all } A \in \mathbb{R}^{n+1} \\ TS(p(x)) &= p(x) \quad \text{for all } p(x) \in \mathcal{P}_n(\mathbb{R}). \end{aligned}$$

To do this we set

$$T(a_1, \dots, a_{n+1}) = a_1 + a_2 x + a_3 x^2 + \dots + a_{n+1} x^n$$

and

$$S(b_0 + b_1 x + \dots + b_n x^n) = (b_0, b_1, \dots, b_n).$$

It is routine to verify that  $S, T$  are linear transformations and that

$$\begin{aligned} ST(A) &= A \\ TS(p(x)) &= p(x) \end{aligned}$$

for all  $A$  in  $\mathbb{R}^{n+1}$  and  $p(x)$  in  $\mathcal{P}_n(\mathbb{R})$ .

In view of Example 8 and the theorems to follow, the reader might be puzzled why, if the two vector spaces  $\mathbb{R}^{n+1}$  and  $\mathcal{P}_n(\mathbb{R})$  are isomorphic, we bothered to introduce both of them rather than stick to good old  $\mathbb{R}^{n+1}$ . The answer is not simple and involves in part the "style" of the mathematics of the second half of this century. However to be more specific, one reason to introduce both examples is that each suggests certain natural phenomena. For example,  $\mathbb{R}^{n+1}$  lends itself quite nicely to the rotation and projection type linear transformations of Examples 1 and 3. These make little geometric

sense in  $\mathcal{P}_n(\mathbb{R})$  directly. On the other hand,  $\mathcal{P}_n(\mathbb{R})$  suggests the linear transformation  $D$ , of Example 6 which in the context of  $\mathbb{R}^{n+1}$  is more than a little forced and artificial. As remarked previously the nilpotent transformations, of which  $D$  is the most natural example, are destined to play a central role in the further study of linear transformations. Finally, there is no “natural” way to construct an isomorphism from  $\mathbb{R}^{n+1}$  to  $\mathcal{P}_n(\mathbb{R})$ . For example we leave to the reader the verification that each of the following linear transformations

$$T_1(a_1, \dots, a_{n+1}) = a_2 + a_3x + \cdots + a_{n+1}x^{n-1} + a_1x^n$$

$$T_2(a_1, \dots, a_{n+1}) = a_1 + (a_1 + a_2)x + \cdots + (a_1 + a_2 + \cdots + a_{n+1})x^n$$

are isomorphisms of  $\mathbb{R}^{n+1}$  to  $\mathcal{P}_n(\mathbb{R})$ .

**Proposition 8.11.** *A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  is an isomorphism iff  $\ker T = \{0\}$  and  $\text{Im } T = \mathcal{W}$ .*

*Note.* We are required to construct a linear transformation

$$S: \mathcal{W} \rightarrow \mathcal{V}$$

such that

$$ST(A) = A \quad \text{for all } A$$

$$TS(B) = B \quad \text{for all } B.$$

In order to do this we will first show that the linear transformation  $T$  has a special property, namely for each vector  $B$  in  $\mathcal{W}$  there is exactly one vector  $A$  in  $\mathcal{V}$  with  $T(A) = B$ . This result is important enough to be of separate interest.

**Proposition 8.12.** *Suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is a linear transformation with  $\ker T = \{0\}$ . Then for each vector  $B$  in  $\text{Im } T$  there is exactly one vector  $A$  in  $\mathcal{V}$  such that  $T(A) = B$ .*

**PROOF.** Since  $B$  is a vector in  $\text{Im } T$  there is certainly at least one vector  $A$  such that  $T(A) = B$ . If there were another vector  $C$  with  $T(C) = B$  then we would have

$$T(A - C) = T(A) - T(C) = B - B = 0.$$

Therefore  $A - C \in \ker T = \{0\}$  and hence

$$A - C = 0 \quad \text{or} \quad A = C$$

so  $C$  could not have been different from  $A$  at all. □

**EXAMPLE 9.** Consider the linear transformation of Example 3

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

defined by

$$T(x, y, z) = (x, y, 0).$$

The vector  $(1, 1, 0)$  belongs to  $\text{Im } T$  and

$$\begin{aligned}(1, 1, 0) &= T(1, 1, 1) \\ (1, 1, 0) &= T(1, 1, -1).\end{aligned}$$

Of course

$$\ker T = \{(0, 0, z)\} \neq \{0\}.$$

The preceding result is in a sense quite surprising. For if  $T: \mathcal{V} \rightarrow \mathcal{W}$  has the property that for each  $\mathbf{B} \in \text{Im } T$  there is exactly one  $\mathbf{A}$  in  $\mathcal{V}$  with  $T(\mathbf{A}) = \mathbf{B}$  then  $\ker T = \{0\}$ . For  $0 \in \text{Im } T$ , since  $\text{Im } T$  is a subspace of  $\mathcal{W}$ , and  $T(0) = 0$  since  $T$  is a linear transformation. Therefore if  $\mathbf{A}$  is any vector in  $\mathcal{V}$  with  $T(\mathbf{A}) = 0$  we must have  $\mathbf{A} = 0$ . What is surprising is that the converse holds, and this is the content of (8.12). It is a reflection of the *homogeneity* of linear transformations.

**PROOF OF (8.11).** Since  $\ker T = \{0\}$  and  $\text{Im } T = \mathcal{W}$  there is for each vector  $\mathbf{A}$  in  $\mathcal{W}$  exactly one vector  $\mathbf{A}$  in  $\mathcal{V}$  such that  $T(\mathbf{A}) = \mathbf{B}$ . Define a function

$$S: \mathcal{W} \rightarrow \mathcal{V}$$

by setting

$$S(\mathbf{B}) = \mathbf{A},$$

that is

$$S(\mathbf{B}) = \mathbf{A} \quad \text{iff} \quad T(\mathbf{A}) = \mathbf{B}$$

Note that since there is exactly one  $\mathbf{A}$  in  $\mathcal{V}$  with  $T(\mathbf{A}) = \mathbf{B}$  this definition yields a well-defined function

$$S: \mathcal{W} \rightarrow \mathcal{V}.$$

We claim that  $S$  is a linear transformation. To prove this we suppose  $\mathbf{B}, \mathbf{C}$  are vectors in  $\mathcal{W}$ . Let  $\mathbf{A}$  and  $\mathbf{D}$  be the unique vectors of  $\mathcal{V}$  with

$$T(\mathbf{A}) = \mathbf{B}, \quad T(\mathbf{D}) = \mathbf{C}.$$

Then since  $T$  is a linear transformation

$$T(\mathbf{A} + \mathbf{D}) = T(\mathbf{A}) + T(\mathbf{D}) = \mathbf{B} + \mathbf{C}.$$

Therefore  $\mathbf{A} + \mathbf{D}$  is the unique vector of  $\mathcal{V}$  with

$$T(\mathbf{A} + \mathbf{D}) = \mathbf{B} + \mathbf{C}.$$

In terms of our function  $S$  this says

$$S(\mathbf{B} + \mathbf{C}) = \mathbf{A} + \mathbf{D} = S(\mathbf{B}) + S(\mathbf{C}).$$

If now  $r$  is a number, then since  $T$  is a linear transformation

$$T(rA) = rT(A) = rB.$$

Therefore

$$S(rB) = rA = rS(A).$$

Thus we have shown that the function

$$S: \mathcal{W} \rightarrow \mathcal{V}$$

is a linear transformation. It is immediate from the definitions that

$$S(T(A)) = A \quad \text{for all } A \text{ in } \mathcal{V}$$

$$T(S(B)) = B \quad \text{for all } B \text{ in } \mathcal{W}.$$

Therefore  $T$  is an isomorphism.

To prove the converse direction, that is, if  $T$  is an isomorphism then  $\ker T = \{0\}$  and  $\text{Im } T = \mathcal{V}$ , is routine and is left to the reader as an exercise.  $\square$

Let us look at some examples:

EXAMPLE 10. Let  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$  be the linear transformation

$$T(a_1, a_2, a_3, a_4) = (a_1, a_1 + a_2, a_1 + a_2 + a_3, a_1 + a_2 + a_3 + a_4)$$

Then we claim that  $T$  is an isomorphism. To see this we *do not* have to construct a linear transformation

$$S: \mathbb{R}^4 \rightarrow \mathbb{R}^4$$

such that

$$ST(A) = A \quad \text{for all } A \text{ in } \mathbb{R}^4$$

$$TS(A) = A \quad \text{for all } A \text{ in } \mathbb{R}^4$$

although such an  $S$  will certainly exist. Rather we will use (8.11). First notice that  $\ker T = \{0\}$ . For if

$$T(a_1, a_2, a_3, a_4) = 0$$

then

$$(a_1, a_1 + a_2, a_1 + a_2 + a_3, a_1 + a_2 + a_3 + a_4) = (0, 0, 0, 0)$$

so

$$a_1 = 0$$

$$a_1 + a_2 = 0$$

$$a_1 + a_2 + a_3 = 0$$

$$a_1 + a_2 + a_3 + a_4 = 0$$

or

$$a_1 = 0, \quad a_2 = 0, \quad a_3 = 0, \quad a_4 = 0.$$



## 8: Linear transformations

Next note that by (8.10) we have

$$4 = \dim \mathbb{R}^4 = \dim \operatorname{Im} T + \dim \ker T = \dim \operatorname{Im} T + 0.$$

Therefore  $\dim \operatorname{Im} T = 4$ . Thus by (6.10)  $\operatorname{Im} T = \mathbb{R}^4$ . To summarize

$$\begin{aligned}\operatorname{Im} T &= \mathbb{R}^4 \\ \ker T &= \{\mathbf{0}\}\end{aligned}$$

so  $T$  is an isomorphism by (8.11).

EXAMPLE 11. Let  $S$  be a set and  $T \subset S$ . If

$$f : T \rightarrow \mathbb{R}$$

is a function in  $\mathcal{F}(T)$  we define a new function

$$L(f) : S \rightarrow \mathbb{R}$$

by

$$L(f)(s) = \begin{cases} f(s) & \text{if } s \in T, \\ 0 & \text{if } s \notin T. \end{cases}$$

The assignment  $f \mapsto L(f)$  defines a function

$$L : \mathcal{F}(T) \rightarrow \mathcal{F}(S),$$

which is a linear transformation. To see this we simply check the definitions:

$$\begin{aligned}L(f + g)(s) &= (f + g)(s) = f(s) + g(s) \\ &= (L(f))(s) + (L(g))(s), \quad s \in T\end{aligned}$$

and  $L(f + g)(s) = 0 = (L(f))(s) + (L(g))(s)$ ,  $s \notin T$ , while

$$L(rf)(s) = \begin{cases} rf(s) = r(L(f))(s) & \text{if } s \in T \\ 0 \cdot r(L(f))(s) & \text{if } s \notin T. \end{cases}$$

Let us compute the image of  $L$ . Note that for all  $f \in \mathcal{F}(T)$  we have

$$(L(f))(s) = 0 \quad \text{if } s \in S - T$$

where  $S - T = \{s \in S \mid s \notin T\}$ . Therefore  $\operatorname{Im} L \subset \mathcal{F}(S, S - T)$ . Moreover if  $g \in \mathcal{F}(S, S - T)$

$$g : S \rightarrow \mathbb{R}$$

so by restricting  $g$  to  $T$  we get

$$g|_T : T \rightarrow \mathbb{R}$$

defined by

$$g|_T(s) = g(s), \quad s \in T.$$

Note that by definition of  $L$

$$\begin{aligned} L(g|_T)(s) &= \begin{cases} g(s) & \text{if } s \in T \\ 0 & \text{if } s \notin T \end{cases} \\ &= g(s) \end{aligned}$$

because  $g \in \mathcal{F}(S, S - T)$ . Therefore  $g \in \text{Im } L$ , and hence  $\mathcal{F}(S, S - T) \subset \text{Im } L$ . Since we already have established the reverse inclusion we conclude  $\text{Im } L = \mathcal{F}(S, S - T)$ .

The kernel of  $L$  may be computed quite easily. If  $f \in \ker L$  then

$$0 = L(f)(s) = \begin{cases} f(s) & \text{if } s \in T \\ 0 & \text{if } s \notin T \end{cases}$$

so that

$$f = 0: T \rightarrow \mathbb{R}$$

and thus

$$\ker L = \{0\}.$$

**EXAMPLE 12.** Let  $S$  be a set and  $T \subset S$ . If  $f \in \mathcal{F}(S)$  then by restricting the domain of  $f$  to  $T$  we obtain a function

$$R_T(f): T \rightarrow \mathbb{R}$$

defined by

$$R_T(f)(s) = f(s).$$

The function

$$R_T: \mathcal{F}(S) \rightarrow \mathcal{F}(T)$$

given by

$$f \mapsto R_T(f)$$

is easily checked to be a linear transformation.

To compute the kernel of  $R_T$  observe that

$$R_T f = 0 \in \mathcal{F}(T) \quad \text{iff} \quad f(s) = 0, \quad s \in T.$$

So we see

$$\ker R_T = \mathcal{F}(S, T).$$

Let  $L$  be the linear transformation of Example 11. By definition we have

$$R_T L(f)(s) = f(s), \quad s \in T$$

so  $R_T L(f) = f \in \mathcal{F}(T)$ , therefore  $R_T L = I$ , and  $\text{Im } R_T = \mathcal{F}(T)$ . (The transformation denoted by  $I$  here is the identity transformation. Precisely,  $I: \mathcal{V} \rightarrow \mathcal{V}$  and  $I(A) = A$  for all  $A \in \mathcal{V}$  is called the **identity transformation** (of the vector space  $\mathcal{V}$ ).)

Finally, by restricting  $R_T$  to the subspace  $\mathcal{F}(S, S - T)$  we obtain

$$\begin{aligned} R_T: \mathcal{F}(S, S - T) &\rightarrow \mathcal{F}(T) \\ L: \mathcal{F}(T) &\rightarrow \mathcal{F}(S, S - T). \end{aligned}$$

It is easy to check that these linear transformations are inverse isomorphisms, so  $\mathcal{F}(T)$  is isomorphic to  $\mathcal{F}(S, S - T)$ .

At this point it is worthwhile to describe a very general process for constructing linear transformations.

## Linear extension construction

Suppose given the following data

- (1)  $\mathcal{V}, \mathcal{W}$  vector spaces
- (2)  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  a basis for  $\mathcal{V}$
- (3)  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  vectors in  $\mathcal{W}$ , they need not be a basis, **they need not even be distinct**.

We define a function  $T: \mathcal{V} \rightarrow \mathcal{W}$  as follows. Let  $\mathbf{A}$  be a vector in  $\mathcal{V}$ . By (6.6) there are unique numbers  $a_1, \dots, a_n$  (the coordinates of  $\mathbf{A}$  relative to  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ) such that

$$\mathbf{A} = a_1\mathbf{A}_1 + \dots + a_n\mathbf{A}_n.$$

Now define

$$T(\mathbf{A}) = a_1\mathbf{B}_1 + \dots + a_n\mathbf{B}_n.$$

Note that

$$T(\mathbf{A}_i) = \mathbf{B}_i \quad i = 1, 2, \dots, n.$$

*Note.* The reason that  $T$  is well defined is because  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is a basis for  $\mathcal{V}$ . For in order to define  $T$  we first must write

$$\mathbf{A} = a_1\mathbf{A}_1 + \dots + a_n\mathbf{A}_n,$$

and we are using the fact that  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  spans  $\mathcal{V}$ . The fact that  $T(\mathbf{A})$  is unambiguously defined in  $\mathcal{W}$  uses the fact that the coordinates of  $\mathbf{A}$  are unique, which is true since  $\mathbf{A}_1, \dots, \mathbf{A}_n$  are linearly independent.

Note also, that the linear transformation  $T$  depends not only on the sets of vectors  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  but also on their *order*. To change the order of the  $\mathbf{A}$ s and/or  $\mathbf{B}$ s will change the transformation  $T$ .

We call the function  $T$  the *linear extension* of

$$T(\mathbf{A}_i) = \mathbf{B}_i \quad i = 1, \dots, n.$$

The adjective **linear** is justified by:

**Proposition 8.13.** *Suppose that  $\mathcal{V}$  and  $\mathcal{W}$  are vector spaces.  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  a basis for  $\mathcal{V}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  vectors in  $\mathcal{W}$ . Then the linear extension of*

$$T(\mathbf{A}_i) = \mathbf{B}_i \quad i = 1, \dots, n$$

*is a linear transformation*

$$T: \mathcal{V} \rightarrow \mathcal{W}.$$

**PROOF.** We must show that

$$T(\mathbf{A} + \mathbf{C}) = T(\mathbf{A}) + T(\mathbf{C})$$

$$T(r\mathbf{A}) = rT(\mathbf{A})$$

for all vectors  $\mathbf{A}$  and  $\mathbf{C}$  in  $\mathcal{V}$  and all numbers  $r$ . So suppose that

$$\mathbf{A} = a_1\mathbf{A}_1 + \dots + a_n\mathbf{A}_n$$

$$\mathbf{C} = c_1\mathbf{A}_1 + \dots + c_n\mathbf{A}_n.$$

Then of course

$$\mathbf{A} + \mathbf{C} = (a_1 + c_1)\mathbf{A}_1 + (a_2 + c_2)\mathbf{A}_2 + \dots + (a_n + c_n)\mathbf{A}_n$$

by the usual sort of argument involving shuffling vectors around. Thus according to our definition of  $T$  we have

$$T(\mathbf{A} + \mathbf{C}) = (a_1 + c_1)\mathbf{B}_1 + \dots + (a_n + c_n)\mathbf{B}_n$$

hence

$$\begin{aligned} T(\mathbf{A} + \mathbf{C}) &= a_1\mathbf{B}_1 + c_1\mathbf{B}_1 + \dots + a_n\mathbf{B}_n + c_n\mathbf{B}_n \\ &= (a_1\mathbf{B}_1 + a_2\mathbf{B}_2 + \dots + a_n\mathbf{B}_n) + (c_1\mathbf{B}_1 + \dots + c_n\mathbf{B}_n) \\ &= T(\mathbf{A}) + T(\mathbf{C}). \end{aligned}$$

In a similar manner we may show

$$T(r\mathbf{A}) = rT(\mathbf{A})$$

for all vectors  $\mathbf{A}$  in  $\mathcal{V}$  and numbers  $r$ . Therefore  $T$  is a linear transformation.  $\square$

The linear extension construction has a number of surprising consequences. Here are some of them:

**Theorem 8.14.** *Suppose that  $\mathcal{V}$  and  $\mathcal{W}$  are finite-dimensional vector spaces. If  $\dim \mathcal{V} = \dim \mathcal{W}$  then  $\mathcal{V}$  and  $\mathcal{W}$  are isomorphic.*

**PROOF.** Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  be a basis for  $\mathcal{V}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  a basis for  $\mathcal{W}$ . Note that they have the same number of elements. Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be the linear transformation that is the linear extension of

$$T(\mathbf{A}_i) = \mathbf{B}_i \quad i = 1, 2, \dots, n.$$

Then by (8.4) and the fact that  $\{A_1, \dots, A_n\}$  is a basis for

$$\begin{aligned}\text{Im } T &= T(\mathcal{L}(A_1, \dots, A_n)) = \mathcal{L}(T(A_1), \dots, T(A_n)) \\ &= \mathcal{L}(B_1, \dots, B_n) = \mathcal{W},\end{aligned}$$

since  $\{B_1, \dots, B_n\}$  is a basis for  $\mathcal{W}$ . Therefore by (8.10)

$$\begin{aligned}n &= \dim \mathcal{V} = \dim \text{Im } T + \dim \ker T = \dim \mathcal{W} + \dim \ker T \\ &= n + \dim \ker T\end{aligned}$$

and hence

$$\dim \ker T = n - n = 0.$$

Therefore  $\ker T = \{0\}$  and hence  $T$  is an isomorphism by (8.11).  $\square$

Notice that the converse of (8.14) is also true. That is if  $\mathcal{V}$  and  $\mathcal{W}$  are finite-dimensional vector spaces that are isomorphic then they must have the same dimension. To see that this is so, suppose

$$T: \mathcal{V} \rightarrow \mathcal{W}$$

is a linear transformation that is an isomorphism. Let  $\{A_1, \dots, A_n\}$  be a basis for  $\mathcal{V}$ , so that  $\dim \mathcal{V} = n$ . Then

$$\begin{aligned}\mathcal{W} &= \text{Im } T = T(\mathcal{V}) = T(\mathcal{L}(A_1, \dots, A_n)) \\ &= \mathcal{L}(T(A_1), \dots, T(A_n))\end{aligned}$$

so that the vectors  $T(A_1), \dots, T(A_n)$  span  $\mathcal{W}$  and hence satisfy half of the basis condition. It remains to show that  $\{T(A_1), \dots, T(A_n)\}$  are linearly independent, for then they are a basis and hence  $\dim \mathcal{W} = n$  also. So we suppose  $\{T(A_1), \dots, T(A_n)\}$  are dependent. Then there are numbers  $a_1, \dots, a_n$ , not all zero, such that

$$a_1 T(A_1) + a_2 T(A_2) + \dots + a_n T(A_n) = 0.$$

We then have

$$\begin{aligned}0 &= a_1 T(A_1) + a_2 T(A_2) + \dots + a_n T(A_n) \\ &= T(a_1 A_1) + T(a_2 A_2) + \dots + T(a_n A_n) \\ &= T(a_1 A_1 + a_2 A_2 + \dots + a_n A_n).\end{aligned}$$

Let  $S: \mathcal{W} \rightarrow \mathcal{V}$  be the inverse of  $T$ , that is the linear transformation which must exist since  $T$  is an isomorphism. Apply  $S$  to the equation

$$0 = T(a_1 A_1 + a_2 A_2 + \dots + a_n A_n)$$

and obtain

$$\begin{aligned}S(0) &= S T(a_1 A_1 + \dots + a_n A_n) \\ 0 &= a_1 A_1 + a_2 A_2 + \dots + a_n A_n,\end{aligned}$$

which says that the vectors  $\{A_1, \dots, A_n\}$  are linearly dependent, contrary to the fact they are a basis for  $\mathcal{V}$ . Thus  $\dim \mathcal{W} = n$  also.

To summarize: Two finite-dimensional vector spaces  $\mathcal{V}$  and  $\mathcal{W}$  are isomorphic iff they have the same dimension.

**Corollary 8.15.** *If a vector space  $\mathcal{V}$  has finite dimension  $n$  then  $\mathcal{V}$  is isomorphic to  $\mathbb{R}^n$ .*  $\square$

In view of (8.15) it is again tempting to ask if it was really worthwhile introducing abstract vector spaces if indeed up to isomorphism there are only (at least for finite-dimensional ones) the spaces  $\mathbb{R}^n$ . The answer lies, as in the discussion after Example 7, in the aesthetics of mathematics and the naturality of certain constructions in some examples and not in others. In particular the operator (i.e. linear transformation)

$$D: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$$

is quite natural in  $\mathcal{P}(\mathbb{R})$  but not in  $\mathbb{R}^n$ , while the rotations (see Example 1)

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

make little sense in  $\mathcal{P}(\mathbb{R})$ , or  $\mathcal{L}(\mathcal{V}, \mathcal{W})$ .

Another reason is that there is no natural isomorphism between  $\mathcal{V}$  and  $\mathbb{R}^n$ . Highly simple and natural constructions, geometric or otherwise, in  $\mathcal{V}$  may go over into complex and unnatural configurations in  $\mathbb{R}^n$ .

**Notation.** Let  $E_i \in \mathbb{R}^n$  be the vector

$$E_i = (0, \dots, 1, 0, \dots).$$

$\uparrow$   
ith place

We have already seen that the vectors  $\{E_1, \dots, E_n\}$  are a basis for  $\mathbb{R}^n$ .

**Proposition 8.16.** *Let  $A_1, \dots, A_n$  be vectors in the vector space  $\mathcal{V}$ . Let*

$$T: \mathbb{R}^n \rightarrow \mathcal{V}$$

*be the linear extension of*

$$T(E_i) = A_i \quad i = 1, 2, \dots, n.$$

*Then  $\{A_1, \dots, A_n\}$  is linearly independent iff  $\ker T = \{0\}$ .*

**PROOF.** Suppose that  $\{A_1, \dots, A_n\}$  is linearly independent. If  $B = (b_1, \dots, b_n) \in \mathbb{R}^n$  then

$$T(b_1, \dots, b_n) = b_1 A_1 + b_2 A_2 + \dots + b_n A_n.$$

Suppose that

$$T(b_1, \dots, b_n) = 0.$$

Then

$$b_1 \mathbf{A}_1 + \cdots + b_n \mathbf{A}_n = \mathbf{0}$$

which since  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is linearly independent implies

$$(b_1, \dots, b_n) = (0, \dots, 0).$$

That is  $\ker T = \{\mathbf{0}\}$ .

On the other hand suppose that  $\ker T = \{\mathbf{0}\}$ . If

$$a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n = \mathbf{0}$$

is a linear relation between  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  then

$$T(a_1, \dots, a_n) = a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n = \mathbf{0}.$$

So  $(a_1, \dots, a_n) = (0, \dots, 0)$  since  $\ker T = \{\mathbf{0}\}$ . Therefore the linear relation is the trivial one, so  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is linearly independent.  $\square$

Similarly we have:

**Proposition 8.17.** Let  $\mathbf{A}_1, \dots, \mathbf{A}_n$  be vectors in the vector space  $\mathcal{V}$ . Let

$$T: \mathbb{R}^n \rightarrow \mathcal{V}$$

be the linear extension of

$$T(\mathbf{E}_i) = \mathbf{A}_i \quad i = 1, 2, \dots, n.$$

Then  $\mathbf{A}_1, \dots, \mathbf{A}_n$  span  $\mathcal{V}$  iff  $\text{Im } T = \mathcal{V}$ .

PROOF. Exercise.  $\square$

EXAMPLE 13. Let  $S', S''$  be finite sets and  $\varphi: S' \rightarrow S''$  be a function. Let

$$T_\varphi: \mathcal{F}(S') \rightarrow \mathcal{F}(S'')$$

be the linear extension of

$$T_\varphi(\chi_{s'}) = \chi_{\varphi(s')} \quad s' \in S'.$$

Then  $T_\varphi$  is a linear transformation. If  $f \in \mathcal{F}(S')$  then

$$\begin{aligned} T_\varphi(f) &= T_\varphi\left(\sum_{s' \in S'} f(s') \chi_{s'}\right) \\ &= \sum_{s' \in S'} f(s') \chi_{\varphi(s')}. \end{aligned}$$

From this formula we easily see that

$$\text{Im } T = L(\mathcal{F}(\varphi(S')))$$

where

$$L: \mathcal{F}(\varphi(S')) \rightarrow \mathcal{F}(S'')$$

is the linear transformation of Example 11 for  $\varphi(S') \subset S''$ . In general it is not easy to compute  $\ker T_\varphi$ , see Examples 9 and 10 in the next chapter.

## EXERCISES

1. Show that each of the following is a linear transformation:

- (a)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $T(x, y) = (2x - y, x)$
- (b)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  defined by  $T(x, y, z) = (z, x + y)$
- (c)  $T: \mathbb{R} \rightarrow \mathbb{R}^2$  defined by  $T(x) = (2x, -x)$
- (d)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$  defined by  $T(x, y) = (x + y, y, x)$

2. Show that each of the following is *not* a linear transformation:

- (a)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $T(x, y) = (x^2, y^2)$
- (b)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  defined by  $T(x, y, z) = (x + y + z, 1)$
- (c)  $T: \mathbb{R} \rightarrow \mathbb{R}^2$  defined by  $T(x) = (1, -1)$
- (d)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$  defined by  $T(x, y) = (xy, y, x)$

3. Let  $D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$  be the differentiation operator. That is  $D(p(x)) = (d/dx)p(x)$ . Show that

$$D^{n+1} = \underbrace{D \cdots D}_{(n+1 \text{ times})} = 0: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R}).$$

4. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}$  be defined by  $T(x, y, z) = x - 3y + 2z$ . Show that  $T$  is linear. Find a basis for the kernel of  $T$ . What are  $\dim \ker T$  and  $\dim \operatorname{Im} T$ ?

5. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by  $T(x, y, z) = (y, 0, z)$ . Show  $T$  is linear. Find a basis for  $\ker T$  and  $\operatorname{Im} T$ . What are their dimensions? Let  $\mathcal{U} \subset \mathbb{R}^3$  be the linear subspace of vectors with  $y = 0$  (that is the  $xz$  plane). Find a basis for  $T(\mathcal{U})$ .

6. Let  $T: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_2(\mathbb{R})$  be the linear transformation defined by

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = a_1 + a_2x + a_3x^2.$$

Find a basis for  $\ker T$  and  $\operatorname{Im} T$ . What are their dimensions?

7. (a) Let  $T: \mathbb{R} \rightarrow \mathbb{R}$  be a linear transformation. Show that there exists a number  $t$ , depending only on  $T$ , such that  $T(x) = tx$  for all  $x \in \mathbb{R}$ .

(b) Suppose that  $T: \mathbb{R} \rightarrow \mathbb{R}$  is a linear transformation such that  $T(3) = -4$ . Calculate  $T(-7)$ . (*Hint*: See part (a) again!)

8. (a) Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}$  be a linear transformation. Show that there exist two numbers  $a$  and  $b$ , depending only on  $T$ , such that  $T(x, y) = ax + by$  for all  $(x, y) \in \mathbb{R}^2$ .

(b) Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}$  be a linear transformation. Suppose  $T(1, 1) = 3$ ,  $T(1, 0) = 4$ . Calculate  $T(2, 1)$ .

9. Let  $S: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the *shift operator* which is defined by  $S(a_1, \dots, a_n) = (0, a_1, a_2, \dots, a_{n-1})$ . Compute the dimension of the kernel and image of  $S$ . Do the same for

$$S^k = \underbrace{S \cdots S}_{k \text{ times}}.$$

Show  $S^n = 0$ .



10. A linear transformation  $S: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  is defined by

$$S(p(x)) = xp(x)$$

that is

$$S(a_0 + a_1x + \cdots + a_nx^n) = a_0x + a_1x^2 + \cdots + a_nx^{n+1}.$$

Find  $\ker S$ ,  $\text{Im } S$ . Does  $S^n = 0$  for any  $n$ ?

11. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  is said to be *injective* iff for each pair of vectors  $A, B$  in  $\mathcal{V}$  with  $A \neq B$  we have  $T(A) \neq T(B)$ . Show that  $T$  is injective iff  $T(C) = 0 \Rightarrow C = 0$ , that is  $\ker T = \{0\}$ .

12. Let  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^4$ ,  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^2$  be the linear transformations defined by

$$\begin{aligned} S(a_1, a_2, a_3) &= (a_1 + a_2, a_1 + a_3, a_2 + a_3, a_1 + a_2 + a_3) \\ T(b_1, b_2, b_3, b_4) &= (b_1 + b_2, b_3 + b_4) \end{aligned}$$

Compute  $T \cdot S(1, 1, 1)$ .

13. Let  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  and  $P: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformations defined by

$$\begin{aligned} S(a_1, a_2, a_3) &= (0, a_1, a_2) \\ P(a_1, a_2, a_3) &= (a_1, a_2, 0). \end{aligned}$$

Calculate  $\dim \text{Im } S$ ,  $\dim \ker S$ ,  $\dim \text{Im } P$ ,  $\dim \ker P$ . Find  $SP(1, 0, 1)$ ,  $PS(1, 0, 1)$ ,  $(S + P)(1, 0, 1)$ , and  $(P - 2S)(1, 0, 1)$ .

14. Suppose that  $\mathcal{V}$  is a finite-dimensional vector space and  $S, T: \mathcal{V} \rightarrow \mathcal{V}$  are linear transformations. Let  $\{A_1, \dots, A_n\}$  be a basis for  $\mathcal{V}$ . Suppose further that

$$S(A_i) = T(A_i) \quad i = 1, 2, \dots, n.$$

Show that  $S = T$ .

15. Suppose that  $\mathcal{V}$  is a finite-dimensional vector space and  $T: \mathcal{V} \rightarrow \mathcal{V}$  a linear transformation. Show the following are equivalent:

- (a)  $T$  is an isomorphism
- (b)  $\ker T = \{0\}$
- (c)  $\text{Im } T = \mathcal{V}$ .

(Hint: Use (8.10) and (8.11).)

16. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear extension of  $T(E_i) = A_i$ ,  $i = 1, 2, 3$  where

$$A_1 = (0, 1, 1), \quad A_2 = (-1, 0, 1), \quad A_3 = (0, 1, 2).$$

Find  $\text{Im } T$  and  $\ker T$ . Also find  $T(1, 2, 3)$ .

17. Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the linear extension of

$$\begin{aligned} T(E_1) &= (0, 0, \dots, 0) \\ T(E_2) &= (1, 0, \dots, 0) \\ T(E_3) &= (0, 2, 0, \dots, 0) \\ &\vdots \\ T(E_n) &= (0, 0, \dots, n-1, 0). \end{aligned}$$

Show that  $T$  is a nilpotent operator and more specifically that  $T^n = 0$ . Find  $\text{Im } T$  and  $\ker T$ .

18. Let  $S$  and  $T$  be finite sets. Show that  $\mathcal{F}(S)$  and  $\mathcal{F}(T)$  are isomorphic iff  $S$  and  $T$  have the same number of elements.

19. Let  $S = \{a, b, c\}$  and let  $T: \mathcal{F}(S) \rightarrow \mathbb{R}^2$  be the linear extension of:

$$T(\chi_a) = (0, 1)$$

$$T(\chi_b) = (1, 1)$$

$$T(\chi_c) = (1, 0).$$

Show that the kernel of  $T$  is the subspace spanned by the function

$$f: S \rightarrow \mathbb{R}$$

defined by  $f(a) = 1, f(b) = -1, f(c) = 1$ .

20. Let  $\mathcal{V}$  and  $\mathcal{W}$  be vector spaces and let  $E \subset \mathcal{V}$  be a subset. Define

$$L: \text{Fun}(E, \mathcal{W}) \rightarrow \mathcal{L}(\mathcal{L}(E), \mathcal{W})$$

by

$$L(f)(a_1 E_1 + \cdots + a_n E_n) = a_1 f(E_1) + \cdots + a_n f(E_n).$$

Show that  $L$  is an isomorphism of vector spaces.

21. Let  $S$  be a set and define  $\varphi: \mathcal{L}(\mathcal{F}(S), \mathbb{R}) \rightarrow \mathcal{F}(S)$  by  $\varphi(L)(s) = L(\chi_s)$ , where  $\chi_s$  is the characteristic function of  $s$ . Show that  $\varphi$  is a linear transformation. Show it is an isomorphism if  $S$  is finite. What happens when  $S$  is infinite?

22. Let  $S, T$  be sets and  $\varphi: S \rightarrow T$  a function. Define  $\varphi^*: \mathcal{F}(T) \rightarrow \mathcal{F}(S)$  by  $\varphi^*(f)(s) = f(\varphi(s))$ .

- Show that  $\varphi^*$  is a linear transformation.
- Show that when  $\varphi$  is epic that  $\varphi^*$  is monic.
- In general show that

$$\text{Ker } \varphi^* = \mathcal{F}(T, T - \varphi(S)).$$

- Show that when  $\varphi$  is injective  $\varphi^*$  is epic.

23. Let  $\mathcal{V}', \mathcal{V}''$  be vector subspaces of the finite-dimensional vector space  $\mathcal{V}$ . Show that there is an isomorphism  $T: \mathcal{V}' \rightarrow \mathcal{V}''$  such that  $T$  maps  $\mathcal{V}'$  isomorphically onto  $\mathcal{V}''$  iff  $\dim \mathcal{V}' = \dim \mathcal{V}''$ .

24. Consider pairs of subspaces  $(\mathcal{V}'_1, \mathcal{V}''_2)$  in  $\mathcal{V}$ . We say that two such pairs  $(\mathcal{V}'_1, \mathcal{V}''_2)$  and  $(\mathcal{V}'_1, \mathcal{V}''_2)$  are *equivalent* iff there is an isomorphism  $T: \mathcal{V}' \rightarrow \mathcal{V}''$  such that  $T$  maps  $\mathcal{V}'_i$  isomorphically onto  $\mathcal{V}''_i$  for  $i = 1, 2$ . Try to find necessary and sufficient conditions that two pairs of subspaces are equivalent. (Hints: 1. Think about the formula

$$\dim(\mathcal{V}'_1 + \mathcal{V}''_2) = \dim \mathcal{V}'_1 + \dim \mathcal{V}''_2 - \dim(\mathcal{V}'_1 \cap \mathcal{V}''_2)$$

and how it is proved.

- Think about how many different dimensions you can associate to  $(\mathcal{V}'_1, \mathcal{V}''_2)$ . For example, do not forget  $\dim \mathcal{V}'_1 - \dim \mathcal{V}'_1 \cap \mathcal{V}''_1 = \dim \mathcal{V}'_1 / \mathcal{V}'_1 \cap \mathcal{V}''_1$ , etc.)

# 9 Linear transformations: some numerical examples

Our objective in this section is to present several numerical examples to illustrate the preceding discussion of linear transformations.

EXAMPLE 1. Which of the following are linear transformations

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

and which are not?

- (1)  $T(x, y, z) = (x + y + z, 0, 0)$
- (2)  $T(x, y, z) = (y, z, x)$
- (3)  $T(x, y, z) = (x + y + z, 1, -1)$
- (4)  $T(x, y, z) = (xyz, 0, 0)$

*Solution.* The first and second of these are easily seen to be linear transformations. For example we will check (2).

Let  $\mathbf{A} = (a_1, a_2, a_3)$ ,  $\mathbf{B} = (b_1, b_2, b_3)$ . Then we have

$$\begin{aligned} T(\mathbf{A} + \mathbf{B}) &= T(a_1 + b_1, a_2 + b_2, a_3 + b_3) \\ &= (a_2 + b_2, a_3 + b_3, a_1 + b_1) \\ &= (a_2, a_3, a_1) + (b_2, b_3, b_1) \\ &= T(\mathbf{A}) + T(\mathbf{B}). \end{aligned}$$

$$\begin{aligned} T(r\mathbf{A}) &= T(ra_1, ra_2, ra_3) \\ &= (ra_2, ra_3, ra_1) \\ &= r(a_2, a_3, a_1) \\ &= rT(\mathbf{A}). \end{aligned}$$

The first may be checked similarly.

Suppose you tried to check the third transformation by this method. We get

$$\begin{aligned}T(\mathbf{A} + \mathbf{B}) &= T(a_1 + b_1, a_2 + b_2, a_3 + b_3) \\&= (a_1 + b_1 + a_2 + b_2 + a_3 + b_3, 1, -1) \\&= (a_1 + a_2 + a_3 + b_1 + b_2 + b_3, 1, -1).\end{aligned}$$

while

$$\begin{aligned}T(\mathbf{A}) + T(\mathbf{B}) &= (a_1 + a_2 + a_3, 1, -1) + (b_1 + b_2 + b_3, 1, -1) \\&= (a_1 + a_2 + a_3 + b_1 + b_2 + b_3, 2, -2).\end{aligned}$$

and we note that

$$\begin{aligned}(a_1 + a_2 + a_3 + b_1 + b_2 + b_3, 1, -1) \\ \neq (a_1 + a_2 + a_3 + b_1 + b_2 + b_3, 2, -2)\end{aligned}$$

because  $1 \neq 2$ . Therefore

$$T(\mathbf{A} + \mathbf{B}) \neq T(\mathbf{A}) + T(\mathbf{B})$$

and hence  $T$  is *not* a linear transformation.

Now actually with a little more familiarity with linear transformations it should be immediately “obvious” that  $T$  is not a linear transformation, all that is needed is a *reason why* it isn’t. Here is one simple reason:

$$T(0, 0, 0) = (0, 1, -1) \neq (0, 0, 0),$$

and according to (8.1)

$$T(0, 0, 0) = (0, 0, 0)$$

if  $T$  is linear. Therefore  $T$  is not linear.

Let us look at the last example now. If it is a linear transformation then we must have

$$T(r\mathbf{A}) = rT(\mathbf{A})$$

for all numbers  $r$  and vectors  $\mathbf{A} = (a_1, a_2, a_3)$ . Let us therefore calculate both sides above.

$$\begin{aligned}T(r\mathbf{A}) &= T(ra_1, ra_2, ra_3) \\&= (ra_1ra_2ra_3, 0, 0) \\&= (r^3a_1a_2a_3, 0, 0) \\&= r^3(a_1a_2a_3, 0, 0)\end{aligned}$$

while

$$rT(\mathbf{A}) = r(a_1a_2a_3, 0, 0).$$

So we must ask ourselves if it is true that

$$r^3(a_1a_2a_3, 0, 0) = r(a_1a_2a_3, 0, 0)$$

## 9: Linear transformations: some numerical examples

for *all* numbers  $r$  and vectors  $A = (a_1, a_2, a_3)$ . The answer is clearly no. For if we set  $a_1 = a_2 = a_3 = 1$  and  $r = 2$ , then

$$r^3(a_1a_2a_3, 0, 0) = 8(1, 0, 0) = (8, 0, 0)$$

$$r(a_1a_2a_3, 0, 0) = 2(1, 0, 0) = (2, 0, 0)$$

and these vectors are clearly not equal, so  $T$  cannot be a linear transformation in this case.

EXAMPLE 2. Let  $T, S$  be the linear transformations

$$T, S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

given by the formulas

$$T(x, y, z) = (x + y + z, 0, 0)$$

$$S(x, y, z) = (y, z, x).$$

Calculate

$$(a) (T \cdot S)(1, 0, 1)$$

$$(d) (S - T)(1, 0, 1)$$

$$(b) (S \cdot T)(1, 0, 1)$$

$$(e) (S(S + T)T)(1, 0, 1).$$

$$(c) (S + T)(1, 0, 1)$$

*Solution.* This requires nothing but determination:

$$\begin{aligned} (a) (T \cdot S)(1, 0, 1) &= T(S(1, 0, 1)) && \text{(by definition of composition)} \\ &= T(0, 1, 1) && \text{(by definition of } S) \\ &= (2, 0, 0). && \text{(by definition of } T) \end{aligned}$$

$$\begin{aligned} (b) S \cdot T(1, 0, 1) &= S(T(1, 0, 1)) \\ &= S(2, 0, 0) \\ &= (0, 0, 2). \end{aligned}$$

$$\begin{aligned} (c) (S + T)(1, 0, 1) &= S(1, 0, 1) + T(1, 0, 1) \\ &= (0, 1, 1) + (2, 0, 0) \\ &= (2, 1, 1). \end{aligned}$$

$$\begin{aligned} (d) (S - T)(1, 0, 1) &= S(1, 0, 1) - T(1, 0, 1) \\ &= (0, 1, 1) - (2, 0, 0) \\ &= (-2, 1, 1). \end{aligned}$$

$$\begin{aligned} (e) (S(S + T)T)(1, 0, 1) &= (S(S + T))(T(1, 0, 1)) \\ &= (S(S + T))(2, 0, 0) \\ &= S((S + T)(2, 0, 0)) \\ &= S(S(2, 0, 0) + T(2, 0, 0)) \\ &= S((0, 0, 2) + (2, 0, 0)) \\ &= S(2, 0, 2) = (0, 2, 2). \end{aligned}$$

EXAMPLE 3. Let

$$L: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

be the linear transformation given by the formula

$$L(x, y, z) = (0, x, y).$$

Find the bases for the kernel and image of  $L$ ,  $L^2$ , and  $L^3$ .

*Solution.* Let us work with  $L$  first. The kernel of  $L$  consists of all the vectors  $\mathbf{A} = (a_1, a_2, a_3)$  such that  $L(\mathbf{A}) = \mathbf{0}$ . That means  $\mathbf{A} \in \ker L$  iff

$$(0, 0, 0) = L(\mathbf{A}) = (0, a_1, a_2)$$

that is

$$0 = 0, \quad 0 = a_1, \quad 0 = a_2,$$

and *no restriction* on  $a_3$ . Thus

$$\ker L = \{(0, 0, z) \in \mathbb{R}^3\}$$

or in geometric terms the kernel of  $L$  is the  $z$ -axis. Therefore a basis for the kernel of  $L$  consists of the single vector  $\{(0, 0, 1)\}$ . (There are many other bases, for example,  $\{(0, 0, -99/100)\}$  is also a basis for  $\ker L$ .)

To find a basis for the image of  $L$  we note that  $\mathbf{B} \in \text{Im } L$  iff  $\mathbf{B} = L(\mathbf{A})$  for some vector  $\mathbf{A} \in \mathbb{R}^3$ . Therefore  $\mathbf{B} \in \text{Im } L$  iff

$$(b_1, b_2, b_3) = \mathbf{B} = L(\mathbf{A}) = (0, a_1, a_2)$$

that is<sup>1</sup>

$$b_1 = 0, \quad b_2 = a_1, \quad b_3 = a_2$$

which means

$$\text{Im } L = \{(0, b_2, b_3)\}$$

or put geometrically,  $\text{Im } L$  is the  $y, z$  plane. A basis for  $\text{Im } L$  is therefore the set  $\{(0, 1, 0), (0, 0, 1)\}$ . Another basis is  $\{(0, 1, -1), (0, 1, 0)\}$ .

We must now apply the same analysis to the linear transformation  $L^2$ . First let us find a "formula" for  $L^2$ . We have

$$\begin{aligned} L^2(x, y, z) &= L(L(x, y, z)) = L(0, x, y) \\ &= (0, 0, x) \end{aligned}$$

(Remember, that  $L$  shifts the coordinates to the right one unit, inserts a 0 in the first place and strikes off the third coordinate.) So a vector  $\mathbf{A}$  belongs to the kernel of  $L^2$  iff

$$(0, 0, 0) = L^2(x, y, z) = (0, 0, a_1),$$

<sup>1</sup> For example  $(0, 2, 3) = L(2, 3, 0)$  and  $(0, 2, 3) = L(2, 3, 1)$  etc.

that is

$$0 = 0, \quad 0 = 0, \quad 0 = a_1,$$

so that  $\mathbf{A} \in \ker L^2$  iff  $a_1 = 0$ , while  $a_2$  and  $a_3$  are arbitrary. Thus

$$\ker L^2 = \{(0, a_2, a_3) \in \mathbb{R}^3\}$$

or geometrically,  $\ker L^2$  is the  $y, z$  plane. Hence a basis for  $\ker L^2$  is  $\{(0, 1, 0), (0, 0, 1)\}$ .

To find a basis for the image of  $L^2$  we note that  $\mathbf{B} \in \text{Im } L^2$  iff  $\mathbf{B} = L^2(\mathbf{A})$  for some vector  $\mathbf{A}$  in  $\mathbb{R}^3$ . That is

$$(b_1, b_2, b_3) = \mathbf{B} = L^2(\mathbf{A}) = (0, 0, a_1)$$

or

$$b_1 = 0, \quad b_2 = 0, \quad b_3 = a_1.$$

Therefore we find

$$\text{Im } L^2 = \{(0, 0, z) \in \mathbb{R}^3\}$$

or that  $\text{Im } L^2$  is the  $z$  axis, which has basis  $\{(0, 0, 1)\}$ .

Finally we note that

$$L^3(x, y, z) = (0, 0, 0)$$

so that  $\text{Im } L^3 = \{(0, 0, 0)\}$  and  $\ker L^3 = \mathbb{R}^3$ . A basis for the kernel of  $L^3$  is thus any basis for  $\mathbb{R}^3$ , such as  $\{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$ , and a basis for the image of  $L^3$  is  $\emptyset$ , the empty set.

**EXAMPLE 4.** Let  $S, T$  be the linear transformations  $S, T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$S(x, y, z) = (y, z, x)$$

$$T(x, y, z) = (x + y + z, 0, 0).$$

Find a basis for the kernel of  $S + T$ .

*Solution.* Let us find a formula for  $S + T$ . Notice that

$$\begin{aligned} (S + T)(x, y, z) &= S(x, y, z) + T(x, y, z) \\ &= (y, z, x) + (x + y + z, 0, 0) \\ &= (x + 2y + z, z, x). \end{aligned}$$

That is

$$(S + T)(x, y, z) = (x + 2y + z, z, x).$$

Therefore  $\mathbf{A} \in \ker (S + T)$  iff

$$(0, 0, 0) = (S + T)(\mathbf{A}) = (a_1 + 2a_2 + a_3, a_3, a_1)$$

or

$$0 = a_1 + 2a_2 + a_3, \quad 0 = a_3, \quad 0 = a_1,$$

iff

$$0 = a_2, \quad 0 = a_3, \quad 0 = a_1,$$

and therefore  $\ker(\mathbf{S} + \mathbf{T}) = \{(0, 0, 0)\}$  so the empty set is a basis for  $\ker(\mathbf{S} + \mathbf{T})$ .

EXERCISE. Show that in the preceding example  $\text{Im}(\mathbf{S} + \mathbf{T}) = \mathbb{R}^3$ .

EXAMPLE 5. Which of the following linear transformations are isomorphisms

(a)  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathcal{P}_2(\mathbb{R})$  defined by

$$\mathbf{T}(a_1, a_2, a_3) = a_1 + (a_1 + a_2)x + (a_1 + a_2 + a_3)x^2,$$

(b)  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathcal{P}_2(\mathbb{R})$  defined by  $\mathbf{T}(a_1, a_2, a_3) = a_1 + (a_1 + a_2)x$ ,

(c)  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathbb{R}^4$  defined by

$$\mathbf{T}(a_1, a_2, a_3) = (a_1 + a_2, a_2 + a_3, a_3 + a_1, a_1 + a_2 + a_3),$$

(d)  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$\mathbf{T}(a_1, a_2, a_3) = (a_1 + a_2 + a_3, a_1 + a_2, a_1 + a_2)?$$

*Solution.* Remember to check that  $\mathbf{T}$  is an isomorphism you do *not* want to use the definition of an isomorphism if you can avoid it, because that requires a construction of another linear transformations  $\mathbf{S}$  that may be quite hard. Rather we will try to use properties of linear transformations such as (8.10) and (8.11).

(a) Note that  $\ker \mathbf{T} = \{\mathbf{0}\}$ . Because

$$\mathbf{0} = \mathbf{T}(\mathbf{A}) = \mathbf{T}(a_1, a_2, a_3) = a_1 + (a_1 + a_2)x + (a_1 + a_2 + a_3)x^2$$

iff

$$0 = a_1, \quad 0 = a_1 + a_2, \quad 0 = a_1 + a_2 + a_3$$

or

$$a_1 = 0, \quad a_2 = 0, \quad a_3 = 0.$$

Next we apply (8.10) which gives

$$3 = \dim \mathbb{R}^3 = \dim \ker \mathbf{T} + \dim \text{Im } \mathbf{T} = 0 + \dim \text{Im } \mathbf{T}$$

so

$$\dim \text{Im } \mathbf{T} = 3.$$

Recall that  $\dim \mathcal{P}_2(\mathbb{R}) = 3$ . Therefore  $\text{Im } \mathbf{T}$ , which is a linear subspace of  $\mathcal{P}_2(\mathbb{R})$  by (8.9), has the same dimension as  $\mathcal{P}_2(\mathbb{R})$  and hence  $\text{Im } \mathbf{T} = \mathcal{P}_2(\mathbb{R})$ . Therefore  $\mathbf{T}$  is an isomorphism by (8.11).



## 9: Linear transformations: some numerical examples

For those who are interested, the linear transformation

$$S: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathbb{R}^3$$

given by the formula

$$S(b_0 + b_1x + b_2x^2) = (b_0, b_1 - b_0, b_2 - b_1 - b_0)$$

is actually the transformation required by the definition, and whose existence is assured by (8.11).

- (b) This transformation is *not* an isomorphism. We may see this by applying (8.11). Because

$$T(0, 0, 1) = \mathbf{0}$$

and hence  $\ker T \neq \{\mathbf{0}\}$ .

- (c) This is *not* an isomorphism, because if it were, then by the discussion following (8.14)

$$3 = \dim \mathbb{R}^3 = \dim \mathbb{R}^4 = 4$$

which is impossible.

- (d) This transformation is also *not* an isomorphism. To see this we note that  $\mathbf{B} \in \text{Im } T$  iff there is an  $\mathbf{A}$  with  $\mathbf{B} = T(\mathbf{A})$ , so in this case

$$(b_1, b_2, b_3) = \mathbf{B} = T(\mathbf{A}) = (a_1 + a_2 + a_3, a_1 + a_2, a_1 + a_2)$$

and therefore

$$b_1 = a_1 + a_2 + a_3$$

$$b_2 = a_1 + a_2$$

$$b_3 = a_1 + a_2$$

and in particular  $b_2 = b_3$ . Therefore the vector  $(0, 1, 2)$  does not belong to  $\text{Im } T$  so by (8.11)  $T$  cannot be an isomorphism since  $\text{Im } T \neq \mathbb{R}^3$ .

**EXAMPLE 6.** Let  $T: \mathbb{R}^3 \rightarrow \mathcal{P}_5(\mathbb{R})$  be the linear transformation that is the linear extension of

$$T(1, 1, 1) = x^2 + x^4,$$

$$T(1, 1, 0) = x + x^3 + x^5,$$

$$T(1, 0, 0) = 1.$$

Calculate  $T(0, 0, 1)$ .

**Solution.** First of all we ought to note that  $\{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$  is a basis for  $\mathbb{R}^3$  so that the transformation  $T$  is well defined by the linear extension construction. To calculate  $T(0, 0, 1)$  we must first write

$$(0, 0, 1) = a_1(1, 1, 1) + a_2(1, 1, 0) + a_3(1, 0, 0),$$

that is, we must find the coordinates of  $(0, 0, 1)$  relative to the basis (*ordered* basis)  $(1, 1, 1), (1, 1, 0), (1, 0, 0)$ . This is just some high school algebra.

Suppose

$$(0, 0, 1) = a_1(1, 1, 1) + a_2(1, 1, 0) + a_3(1, 0, 0).$$

Then

$$\begin{aligned}(0, 0, 1) &= (a_1, a_1, a_1) + (a_2, a_2, 0) + (a_3, 0, 0) \\ &= (a_1 + a_2 + a_3, a_1 + a_2, a_1)\end{aligned}$$

so

$$0 = a_1 + a_2 + a_3, \quad 0 = a_1 + a_2, \quad 1 = a_1$$

or

$$a_1 = 1, \quad a_2 = -1, \quad a_3 = 0.$$

So

$$(0, 0, 1) = 1(1, 1, 1) - 1(1, 1, 0) + 0(1, 0, 0)$$

and therefore

$$\begin{aligned}T(0, 0, 1) &= 1(x^2 + x^4) - 1(x + x^3 + x^5) \\ &= -x + x^2 - x^3 + x^4 - x^5.\end{aligned}$$

Notice that since the vectors

$$\{x^2 + x^4, x + x^3 + x^5, 1\}$$

are linearly independent in  $\mathcal{P}_5(\mathbb{R})$  the kernel of the transformation  $T$  of the preceding example is  $\{0\}$ . (Why? Because of (8.15).)

Here is another example involving linear extensions.

**EXAMPLE 7.** Let  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^2$  be the linear transformation that is the linear extension of

$$\begin{aligned}T(1, 0, 0, 0) &= (1, 1), \\ T(1, 1, 0, 0) &= (0, 1), \\ T(1, 1, 1, 0) &= (1, 0), \\ T(1, 1, 1, 1) &= (-1, -1).\end{aligned}$$

Calculate

$$T(4, 3, 2, 1).$$

**Solution.** Again, as in (6), the first step will be to find the coordinates of  $(4, 3, 2, 1)$  relative to the (ordered) basis

$$\{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$$

for  $\mathbb{R}^4$ . O.K. So write

$$(4, 3, 2, 1) = a_1(1, 0, 0, 0) + a_2(1, 1, 0, 0) + a_3(1, 1, 1, 0) + a_4(1, 1, 1, 1)$$

and use high school algebra to find  $a_1, a_2, a_3, a_4$ . We have

$$\begin{aligned}(4, 3, 2, 1) &= (a_1, 0, 0, 0) + (a_2, a_2, 0, 0) + (a_3, a_3, a_3, 0) + (a_4, a_4, a_4, a_4) \\ &= (a_1 + a_2 + a_3 + a_4, a_2 + a_3 + a_4, a_3 + a_4, a_4)\end{aligned}$$

so

$$4 = a_1 + a_2 + a_3 + a_4$$

$$3 = a_2 + a_3 + a_4$$

$$2 = a_3 + a_4$$

$$1 = a_4$$

which implies

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = 1, \quad a_4 = 1.$$

Thus

$$(4, 3, 2, 1) = 1(1, 0, 0, 0) + 1(1, 1, 0, 0) + 1(1, 1, 1, 0) + 1(1, 1, 1, 1)$$

so

$$\begin{aligned}T(4, 3, 2, 1) &= (1, 1) + (0, 1) + (1, 0) + (-1, -1) \\ &= (1 + 0 + 1 - 1, 1 + 1 + 0 - 1) \\ &= (1, 1).\end{aligned}$$

EXAMPLE 8. Let  $\mathcal{V}$  be the subspace

$$\mathcal{V} = \{(x, y, z) \mid x + y + z = 0\}$$

of  $\mathbb{R}^3$ . Is  $\mathcal{V}$  isomorphic to  $\mathbb{R}^2$ ?

*Solution.* Yes. Note  $\dim \mathcal{V} = 2$  since  $\{(1, -1, 0), (0, 1, -1)\}$  is a basis for  $\mathcal{V}$ . Now apply (8.14).

EXAMPLE 9. Let  $S$  be the set  $\{u, v, w\}$  and define

$$E : \mathcal{F}(S) \rightarrow \mathbb{R}$$

by

$$E(\mathbf{f}) = \mathbf{f}(u) + \mathbf{f}(v) + \mathbf{f}(w).$$

Then  $E$  is a linear transformation. To compute the kernel of  $E$  note that

$$E(\mathbf{f}) = 0 \quad \text{iff} \quad \mathbf{f}(u) + \mathbf{f}(v) + \mathbf{f}(w) = 0.$$

Let us define  $\mathbf{g}_1, \mathbf{g}_2 \in \mathcal{F}(S)$ , that is,  $g_1, g_2 : S \rightarrow \mathbb{R}$  by

$$\mathbf{g}_1(u) = 1, \quad \mathbf{g}_1(v) = -1, \quad \mathbf{g}_1(w) = 0,$$

and

$$\mathbf{g}_2(u) = 0, \quad \mathbf{g}_2(v) = -1, \quad \mathbf{g}_2(w) = 1.$$

Then  $\mathbf{g}_1, \mathbf{g}_2 \in \ker \mathbf{E}$ , for

$$\mathbf{E}(\mathbf{g}_1) = \mathbf{g}_1(u) + \mathbf{g}_1(v) + \mathbf{g}_1(w) = 1 + (-1) + 0 = 0$$

$$\mathbf{E}(\mathbf{g}_2) = \mathbf{g}_2(u) + \mathbf{g}_2(v) + \mathbf{g}_2(w) = 0 + (-1) + 1 = 0.$$

The vectors  $\mathbf{g}_1$  and  $\mathbf{g}_2$  are linearly independent because

$$a_1\mathbf{g}_1 + a_2\mathbf{g}_2 = \mathbf{0}$$

implies

$$\begin{aligned} 0 &= (a_1\mathbf{g}_1 + a_2\mathbf{g}_2)(u) = a_1\mathbf{g}_1(u) + a_2\mathbf{g}_2(u) \\ &= a_1 + a_2 \cdot 0 = a_1, \end{aligned}$$

$$\begin{aligned} 0 &= (a_1\mathbf{g}_1 + a_2\mathbf{g}_2)(w) = a_1\mathbf{g}_1(w) + a_2\mathbf{g}_2(w) \\ &= a_1 \cdot 0 + a_2 = a_2, \end{aligned}$$

so  $a_1 = 0$  and  $a_2 = 0$ .

Finally if  $\mathbf{f} \in \ker \mathbf{E}$  then

$$\mathbf{f}(u) + \mathbf{f}(v) + \mathbf{f}(w) = 0$$

so

$$-\mathbf{f}(w) = \mathbf{f}(u) + \mathbf{f}(v).$$

Therefore

$$\begin{aligned} (\mathbf{f} - \mathbf{f}(u)\mathbf{g}_1)(u) &= 0 = (\mathbf{f}(w)\mathbf{g}_2)(u) \\ (\mathbf{f} - \mathbf{f}(u)\mathbf{g}_1)(v) &= \mathbf{f}(v) + \mathbf{f}(u) = -\mathbf{f}(w) = (\mathbf{f}(w)\mathbf{g}_2)(v) \\ (\mathbf{f} - \mathbf{f}(u)\mathbf{g}_1)(w) &= \mathbf{f}(w) = (\mathbf{f}(w)\mathbf{g}_2)(w) \end{aligned}$$

and so

$$\mathbf{f} - \mathbf{f}(u)\mathbf{g}_1 = \mathbf{f}(w)\mathbf{g}_2,$$

whence

$$\mathbf{f} = \mathbf{f}(u)\mathbf{g}_1 + \mathbf{f}(w)\mathbf{g}_2$$

so that  $\mathbf{g}_1, \mathbf{g}_2$  span  $\ker \mathbf{E}$ ; being linearly independent they are a basis for  $\ker \mathbf{E}$ .

EXAMPLE 10. Let  $S'$  and  $S''$  be the sets  $S' = \{u, v, w\}$ ,  $S'' = \{a, b\}$ , and let  $\varphi: S' \rightarrow S''$  be defined by

$$\varphi(u) = a, \quad \varphi(v) = a, \quad \varphi(w) = b.$$

Then  $\varphi(S') = S''$  so by Example 13 of the preceding chapter the transformation

$$\mathbf{T}_\varphi: \mathcal{F}(S') \rightarrow \mathcal{F}(S'')$$

is onto. To compute  $\ker \mathbf{T}_\varphi$  note that by (8.10)

$$\dim \mathcal{F}(S') = \dim \ker \mathbf{T}_\varphi + \dim \operatorname{Im} \mathbf{T}_\varphi = \dim \ker \mathbf{T}_\varphi + \dim \mathcal{F}(S'')$$

## 9: Linear transformations: some numerical examples

so  $\dim \ker T_\varphi = 1$  since  $\dim \mathcal{F}(S') = 3$  and  $\dim \mathcal{F}(S'') = 2$ . Therefore all we need to do to obtain a basis for  $\ker T_\varphi$  is to find a single nonzero vector in  $\ker T_\varphi$ . Clearly  $\chi_u - \chi_v$  is such a vector, so  $\ker T_\varphi$  is spanned by the function  $\mathbf{k} = \chi_u - \chi_v$ .

### EXERCISES

Let  $k(x)$  be a *fixed* polynomial. Define functions

$$M_{k(x)}: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$$

$$L_{k(x)}: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$$

by

$$M_{k(x)}(p(x)) = k(x)p(x)$$

$$L_{k(x)}(p(x)) = \int_0^x p(t)k(t)dt.$$

1. Show that  $M_{k(x)}$  and  $L_{k(x)}$  are linear transformations for all polynomials  $k(x)$ . Calculate

$$M_{x^2+2x^3}(x+x^4) \quad \text{and} \quad L_{x+x^2}(1+x^3).$$

2. Calculate  $DM_x - M_x D: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  where  $D$  is defined in Example 6 of Chapter 8. That is find a "formula" for

$$(DM_x - M_x D)(p(x)).$$

3. Show that  $LM_{k(x)} = L_{k(x)}: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$ .

4. Show that  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by

$$T(x, y) = (ax + by, cx + dy),$$

$a, b, c, d$  fixed real numbers, is a linear transformation and that  $T$  is an isomorphism iff  $ad \neq bc$ .

5. Show that  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$T(x, y, z) = (a_1x + a_2y + a_3z, b_1x + b_2y + b_3z, c_1x + c_2y + c_3z)$$

is a linear transformation. Any linear transformation from  $\mathbb{R}^3$  to  $\mathbb{R}^3$  takes this form, where  $a_i, b_i, c_i$  are scalars.

6. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be defined by  $T(x, y) = (x - y, 2x + y)$  and let  $S: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be defined by  $S(x, y) = (y - 2x, x + y)$ .

(a) Find  $T \cdot S(1, 0)$ ,  $(T + S)(1, 0)$ ,  $S \cdot T(1, 0)$ ,  $(2T - S)(1, 0)$  and  $T^2(1, 0)$ .

(b) Find  $T \cdot S(x, y)$  and  $S \cdot T(x, y)$ .

(c) What are the vectors  $(x, y)$  satisfying  $T(x, y) = (1, 0)$ ?

7. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be defined by  $T(x, y) = (x, x)$ . What is the kernel of  $T$ , and the image of  $T$ ?

## 9. Linear transformations: some numerical examples

8. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by  $T(x, y, z) = (x + y, y + z, z + x)$ . Is  $T$  an isomorphism? Find  $T(1, 0, 0)$ ,  $T(0, 1, 0)$ , and  $T(0, 0, 1)$ .

9. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a linear transformation satisfying the condition

$$T(x, y, z) = (0, 0, 0) \quad \text{whenever } 2x - y + z = 0.$$

Let  $T(0, 0, 1) = (1, 2, 3)$ . Find  $T(1, 0, 0)$  and  $T(0, 1, 0)$ . What is  $\dim \operatorname{Im} T$ ?

10. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by the formula

$$T(x, y, z) = (y + z, x + z, x + y).$$

Show that  $T$  is an isomorphism and find an inverse for  $T$ .

11. Let  $\mathcal{V}$  be a finite-dimensional vector space over  $\mathbb{R}$ , and  $E_1, \dots, E_n$  be a basis for  $\mathcal{V}$ . Define for each  $i$ ,  $1 \leq i \leq n$  a function

$$E_i^*: \mathcal{V} \rightarrow \mathbb{R}$$

by the formula

$$E_i^*(A) = a_i, \quad 1 \leq i \leq n,$$

where

$$A = a_1 E_1 + \dots + a_n E_n.$$

(a) Show that  $E_1^*, \dots, E_n^*$  are linear transformations.

(b) Find a basis for  $\operatorname{Ker} E_i$ .

(c) Let  $S: \mathcal{V} \rightarrow \mathcal{L}(\mathcal{V}, \mathbb{R})$  be the linear extension of the map

$$E_i \rightarrow E_i^*, \quad i = 1, \dots, n.$$

Show that  $S$  is an isomorphism.

(d) Compute  $\dim_{\mathbb{R}} \mathcal{L}(\mathcal{V}, \mathbb{R})$ .

12. Let  $\mathcal{V}$  be a vector space and define  $\mathcal{V}^* = \mathcal{L}(\mathcal{V}; \mathbb{R})$

$$B: \mathcal{V} \rightarrow \mathcal{V}^{**} \quad (= \mathcal{L}(\mathcal{V}^*, \mathbb{R}) = \mathcal{L}(\mathcal{L}(\mathcal{V}; \mathbb{R}), \mathbb{R}))$$

by

$$B(A)(f) = f(A).$$

(a) Show that  $B$  is a linear transformation.

(b) Show that  $\operatorname{Ker} B = \{0\}$ .

(c) Show that  $B$  is an isomorphism if  $\mathcal{V}$  is finite dimensional.

(d) Show by an example that  $\operatorname{Im} B \neq \mathcal{V}^{**}$  when  $\mathcal{V}$  is not finite dimensional.

13. Let  $\mathcal{V}$  be a finite-dimensional vector space with basis  $E = \{E_1, \dots, E_n\}$ . Show that the linear extension construction gives an isomorphism

$$L: \mathcal{F}(E) \rightarrow \mathcal{V}$$

by

$$L(f) = f(E_1) \cdot E_1 + \dots + f(E_n) \cdot E_n.$$

9: Linear transformations: some numerical examples

14. Let  $\mathcal{V}$  be a finite-dimensional vector space with basis  $\mathbf{E} = \{\mathbf{E}_1, \dots, \mathbf{E}_n\}$ . Show that the map (remember  $\mathcal{V}^* = \mathcal{L}(\mathcal{V}; \mathbb{R})$ )

$$\mathbf{C}: \mathcal{V}^* \rightarrow \mathcal{F}(E)$$

defined by

$$\mathbf{C}(\mathbf{T})(\mathbf{E}_i) = \mathbf{T}(\mathbf{E}_i), \quad i = 1, \dots, n$$

is a vector space isomorphism.

15. Define a function

$$l: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$$

by

$$l(p(x)) = p(x + 1).$$

Determine if  $l$  is a linear transformation or not.

16. Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional vector spaces. Suppose  $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$  is a basis for  $\mathcal{V}$  and  $\{\mathbf{F}_1, \dots, \mathbf{F}_m\}$  a basis for  $\mathcal{W}$ . For each ordered pair  $(i, j)$  show that

$$\mathbf{M}_{i,j}: \mathcal{V} \rightarrow \mathcal{W},$$

$$\mathbf{M}_{i,j}(\mathbf{A}) = a_j \mathbf{F}_i, \quad \mathbf{A} = a_1 \mathbf{E}_1 + \dots + a_n \mathbf{E}_n$$

is a linear transformation.

(a) Show  $\{\mathbf{M}_{i,j} | 1 \leq i \leq n, 1 \leq j \leq m\}$  is a basis for  $\mathcal{L}(\mathcal{V}, \mathcal{W})$ .

(b) Show  $\dim \mathcal{L}(\mathcal{V}, \mathcal{W}) = nm$ .

17. Let  $\mathcal{V}$  be a vector space and suppose that  $\mathbf{S}, \mathbf{T} \in \mathcal{V}^* = \mathcal{L}(\mathcal{V}; \mathbb{R})$ . Define

$$\mathbf{L}: \mathcal{V} \rightarrow \mathbb{R}^2$$

by

$$\mathbf{L}(\mathbf{A}) = (\mathbf{S}(\mathbf{A}), \mathbf{T}(\mathbf{A})).$$

(a) Show that  $\mathbf{L}$  is a linear transformation.

(b) Show that  $\text{Ker } \mathbf{L} = \text{Ker } \mathbf{S} \cap \text{Ker } \mathbf{T}$ .

(c) Show that  $\text{Im } \mathbf{L} = \mathbb{R}^2$  iff  $\mathbf{S} \neq 0 \neq \mathbf{T}$ .

18. Let  $\mathcal{V}$  be a vector space and define

$$\mathbf{C}_i: \mathcal{L}(\mathcal{V}; \mathbb{R}) \rightarrow \mathcal{L}(\mathcal{V}; \mathbb{R}^2), \quad i = 1, 2$$

by

$$\left. \begin{aligned} \mathbf{C}_1(\mathbf{T})(\mathbf{A}) &= (\mathbf{T}(\mathbf{A}), 0), \\ \mathbf{C}_2(\mathbf{T})(\mathbf{A}) &= (0, \mathbf{T}(\mathbf{A})), \end{aligned} \right\} \quad \mathbf{A} \in \mathcal{V}.$$

(a) Show that  $\mathbf{C}_i$  is a linear transformation for  $i = 1, 2$ .

(b) Show that  $\text{Ker } \mathbf{C}_i = \{0\}$ ,  $i = 1, 2$ .

(c) Show that  $\text{Im } \mathbf{C}_1 \cap \text{Im } \mathbf{C}_2 = \{0\}$ .

(d) Let  $\mathcal{L}_i = \text{Im } \mathbf{C}_i$ ,  $i = 1, 2$  and show that

$$\mathcal{L}(\mathcal{V}; \mathbb{R}^2) = \mathcal{L}_1 + \mathcal{L}_2.$$

19. Let  $\mathcal{V}$  be a vector space over  $\mathbb{R}$ . A *complex structure* on  $\mathcal{V}$  is a linear transformation

$$J: \mathcal{V} \rightarrow \mathcal{V}$$

such that for all  $A$  in  $\mathcal{V}$

$$J^2(A) = -A.$$

- (a) Show that

$$J(x, y) = (-y, x)$$

is a complex structure on  $\mathbb{R}^2$ .

- (b) Show that

$$J(x, y) = (y, -x)$$

is a different complex structure on  $\mathbb{R}^2$ .

For a real vector space  $\mathcal{V}$  with complex structure  $J$  define a multiplication of vectors in  $\mathcal{V}$  by *complex* scalars by the formula

$$(a + bi) \cdot A = aA + bJ(A)$$

- (c) Show that this definition of complex scalar multiplication makes  $\mathcal{V}$  into a complex vector space.  
 (d) Show that on  $\mathbb{R}^3$  there is no complex structure.

20. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is called *invertible* iff it is an isomorphism.

- (a) If  $\mathcal{V}$  is finite dimensional show that  $T$  is invertible iff  $\text{Ker } T = \{0\}$ .  
 (b) If  $S, T$  are invertible show  $S \cdot T$  is also. What is  $(S \cdot T)^{-1}$ ?



# 10

## Matrices and linear transformations

We come now to the connecting link between linear transformations and matrices. Our approach will be to consider first the case of a linear transformation

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

in some detail and then abstract the salient features to the general case. Let us therefore suppose given a fixed linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . As usual we will denote by  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  the standard basis vectors  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$  in  $\mathbb{R}^3$ . Let

$$T(\mathbf{E}_1) = (a_{11}, a_{21}, a_{31})$$

$$T(\mathbf{E}_2) = (a_{12}, a_{22}, a_{32})$$

$$T(\mathbf{E}_3) = (a_{13}, a_{23}, a_{33}).$$

Notice that the subscript  $ij$  on an  $a$  means that it is the  $i$ th coordinate of  $T(\mathbf{E}_j)$ , and that  $i, j$  assume the values 1, 2, 3. Of course using the vectors  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  we can write these equations as

$$T(\mathbf{E}_1) = a_{11}\mathbf{E}_1 + a_{21}\mathbf{E}_2 + a_{31}\mathbf{E}_3$$

$$T(\mathbf{E}_2) = a_{12}\mathbf{E}_1 + a_{22}\mathbf{E}_2 + a_{32}\mathbf{E}_3$$

$$T(\mathbf{E}_3) = a_{13}\mathbf{E}_1 + a_{23}\mathbf{E}_2 + a_{33}\mathbf{E}_3.$$

If we now recall the “ $\sum$  notation” for sums from the calculus we see that we may sum up these three equations by the one equation

$$T(\mathbf{E}_j) = \sum_{i=1}^3 a_{ij}\mathbf{E}_i \quad j = 1, 2, 3.$$

Suppose now that  $\mathbf{V} = (v_1, v_2, v_3)$  is any vector in  $\mathbb{R}^3$ . Then of course

$$\mathbf{V} = v_1 \mathbf{E}_1 + v_2 \mathbf{E}_2 + v_3 \mathbf{E}_3.$$

Therefore

$$\begin{aligned} \mathbf{T}(\mathbf{V}) &= \mathbf{T}(v_1 \mathbf{E}_1 + v_2 \mathbf{E}_2 + v_3 \mathbf{E}_3) \\ &= v_1 \mathbf{T}(\mathbf{E}_1) + v_2 \mathbf{T}(\mathbf{E}_2) + v_3 \mathbf{T}(\mathbf{E}_3) \end{aligned}$$

because  $\mathbf{T}$  is a linear transformation. Now referring to the preceding list of equations for  $\mathbf{T}(\mathbf{E}_j)$ ,  $j = 1, 2, 3$ , we find

$$\begin{aligned} \mathbf{T}(\mathbf{V}) &= v_1(a_{11}\mathbf{E}_1 + a_{21}\mathbf{E}_2 + a_{31}\mathbf{E}_3) + v_2(a_{12}\mathbf{E}_1 + a_{22}\mathbf{E}_2 + a_{32}\mathbf{E}_3) \\ &\quad + v_3(a_{13}\mathbf{E}_1 + a_{23}\mathbf{E}_2 + a_{33}\mathbf{E}_3) \end{aligned}$$

or after some algebraic manipulations, that

$$\begin{aligned} \mathbf{T}(\mathbf{V}) &= (a_{11}v_1 + a_{12}v_2 + a_{13}v_3)\mathbf{E}_1 + (a_{21}v_1 + a_{22}v_2 + a_{33}v_3)\mathbf{E}_2 \\ &\quad + (a_{31}v_1 + a_{32}v_2 + a_{33}v_3)\mathbf{E}_3, \end{aligned}$$

which in the more compact  $\sum$  notation may be written

$$\mathbf{T}(\mathbf{V}) = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij} v_j \mathbf{E}_i.$$

Thus we see that from the array of 9 numbers

$$(a_{ij})_{\substack{i=1,2,3 \\ j=1,2,3}}$$

we may calculate the value of the linear transformation  $\mathbf{T}$  on any vector  $\mathbf{V}$  whatsoever. Notice that it is not just the collection of 9 numbers that matters but the pattern (which is  $3 \times 3$ ) that they fit into. Notice also that we have tacitly agreed to *fix* the ordered basis  $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3\}$  for  $\mathbb{R}^3$  and calculate all coordinates with respect to this ordered basis. Let us formalize these ideas with a definition.

**Definition.** Let  $\mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a linear transformation. Suppose that

$$\mathbf{T}(\mathbf{E}_j) = \sum_{i=1}^3 a_{ij} \mathbf{E}_i \quad j = 1, 2, 3,$$

for the unique numbers  $(a_{ij})_{i,j=1,2,3}$ . The *matrix of  $\mathbf{T}$  relative to the ordered basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$*  is the  $3 \times 3$  array of numbers

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

Notice that the number  $a_{ij}$  occurs as the entry in the  $i$ th row and  $j$ th column of the matrix of  $\mathbf{T}$ . In order to write down the matrix of  $\mathbf{T}$  we must calculate the coordinates of  $\mathbf{T}(\mathbf{E}_j)$  relative to the ordered basis  $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3\}$  and write the resulting numbers as the  $j$ th *column* of the matrix of  $\mathbf{T}$ .

EXAMPLE 1. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by

$$T(x, y, z) = (x + 2z, y - x, z + y).$$

Calculate the matrix of  $T$  relative to the standard basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  of  $\mathbb{R}^3$ .

*Solution.* We must first calculate the coordinates of  $T(\mathbf{E}_1)$ ,  $T(\mathbf{E}_2)$ ,  $T(\mathbf{E}_3)$  and write them down according to the rule explained above. We have

$$T(\mathbf{E}_1) = T(1, 0, 0) = (1, -1, 0) = 1\mathbf{E}_1 + (-1)\mathbf{E}_2 + 0\mathbf{E}_3$$

$$T(\mathbf{E}_2) = T(0, 1, 0) = (0, 1, 1) = 0\mathbf{E}_1 + (1)\mathbf{E}_2 + (1)\mathbf{E}_3$$

$$T(\mathbf{E}_3) = T(0, 0, 1) = (2, 0, 1) = 2\mathbf{E}_1 + 0\mathbf{E}_2 + 1\mathbf{E}_3.$$

Therefore the matrix of  $T$  relative to the basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  is

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

EXAMPLE 2. Let  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by<sup>1</sup>

$$S(x, y, z) = (0, x, y).$$

Calculate the matrix of  $S$  relative to the standard basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  of  $\mathbb{R}^3$ .

*Solution.* We have

$$S(\mathbf{E}_1) = S(1, 0, 0) = (0, 1, 0) = 0\mathbf{E}_1 + \mathbf{E}_2 + 0\mathbf{E}_3$$

$$S(\mathbf{E}_2) = S(0, 1, 0) = (0, 0, 1) = 0\mathbf{E}_1 + 0\mathbf{E}_2 + \mathbf{E}_3$$

$$S(\mathbf{E}_3) = S(0, 0, 1) = (0, 0, 0) = 0\mathbf{E}_1 + 0\mathbf{E}_2 + 0\mathbf{E}_3$$

so the matrix we seek is

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

In order for the matrix of a linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  to be of any real use to us it should be possible to answer questions about  $T$  in terms of its matrix, and for each operation on  $T$  to have a corresponding operation of the matrix of  $T$ . To illustrate this latter point let us suppose given two linear transformations

$$T, S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

which have matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

<sup>1</sup> The transformation  $S$  is called the *shift* operator.

relative to the standard basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  of  $\mathbb{R}^3$ . There is then also the linear transformation

$$\mathbf{S} \cdot \mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

that is the composition of  $\mathbf{S}$  and  $\mathbf{T}$ . The question we pose is what is the matrix of  $\mathbf{S} \cdot \mathbf{T}$  relative to the standard basis of  $\mathbb{R}^3$ ? To answer this question we simply follow the procedure of the two preceding examples, taking care not to become too entangled in the notations. Let us therefore calculate  $\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_1)$ ,  $\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_2)$  and  $\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_3)$  in terms of what has been given to us, namely the matrices  $\mathbf{A}$  and  $\mathbf{B}$ . We have

$$\begin{aligned}\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_1) &= \mathbf{S}(\mathbf{T}(\mathbf{E}_1)) \\ &= \mathbf{S}(a_{11}\mathbf{E}_1 + a_{21}\mathbf{E}_2 + a_{31}\mathbf{E}_3) \\ &= a_{11}\mathbf{S}(\mathbf{E}_1) + a_{21}\mathbf{S}(\mathbf{E}_2) + a_{31}\mathbf{S}(\mathbf{E}_3) \\ &= a_{11}(b_{11}\mathbf{E}_1 + b_{21}\mathbf{E}_2 + b_{31}\mathbf{E}_3) \\ &\quad + a_{21}(b_{12}\mathbf{E}_1 + b_{22}\mathbf{E}_2 + b_{32}\mathbf{E}_3) \\ &\quad + a_{31}(b_{13}\mathbf{E}_1 + b_{23}\mathbf{E}_2 + b_{33}\mathbf{E}_3)\end{aligned}$$

which after a little algebraic manipulation gives

$$\begin{aligned}\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_1) &= (b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31})\mathbf{E}_1 \\ &\quad + (b_{21}a_{11} + b_{22}a_{21} + b_{23}a_{31})\mathbf{E}_2 \\ &\quad + (b_{31}a_{11} + b_{32}a_{21} + b_{33}a_{31})\mathbf{E}_3.\end{aligned}$$

Thus the first column of the matrix we are seeking is

$$\begin{aligned}&b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31} \\ &b_{21}a_{11} + b_{22}a_{21} + b_{23}a_{31} \\ &b_{31}a_{11} + b_{32}a_{21} + b_{33}a_{31}.\end{aligned}$$

In a similar manner we compute  $\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_2)$ ,  $\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_3)$  and we find that

$$\begin{aligned}\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_2) &= (b_{11}a_{12} + b_{12}a_{22} + b_{13}a_{32})\mathbf{E}_1 \\ &\quad + (b_{21}a_{12} + b_{22}a_{22} + b_{23}a_{32})\mathbf{E}_2 \\ &\quad + (b_{31}a_{12} + b_{32}a_{22} + b_{33}a_{32})\mathbf{E}_3,\end{aligned}$$

and

$$\begin{aligned}\mathbf{S} \cdot \mathbf{T}(\mathbf{E}_3) &= (b_{11}a_{13} + b_{12}a_{23} + b_{13}a_{33})\mathbf{E}_1 \\ &\quad + (b_{21}a_{13} + b_{22}a_{23} + b_{23}a_{33})\mathbf{E}_2 \\ &\quad + (b_{31}a_{13} + b_{32}a_{23} + b_{33}a_{33})\mathbf{E}_3.\end{aligned}$$

Therefore the matrix of  $\mathbf{S} \cdot \mathbf{T}$  relative to the standard basis of  $\mathbb{R}^3$  is given by (the commas have been inserted so that the columns may be easily distinguished)

$$\mathbf{C} = \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31}, & b_{11}a_{12} + b_{12}a_{22} + b_{13}a_{32}, & b_{11}a_{13} + b_{12}a_{23} + b_{13}a_{33} \\ b_{21}a_{11} + b_{22}a_{21} + b_{23}a_{31}, & b_{21}a_{12} + b_{22}a_{22} + b_{23}a_{32}, & b_{21}a_{13} + b_{22}a_{23} + b_{23}a_{33} \\ b_{31}a_{11} + b_{32}a_{21} + b_{33}a_{31}, & b_{31}a_{12} + b_{32}a_{22} + b_{33}a_{32}, & b_{31}a_{13} + b_{32}a_{23} + b_{33}a_{33} \end{pmatrix}.$$

While the above formulas are quite formidable to contemplate there is really a very simple rule underlying the computation of the matrix  $\mathbf{C}$ . Write

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}.$$

The entry  $c_{ij}$  is calculated by taking the  $i$ th row of  $\mathbf{B}$

$$b_{i1}, b_{i2}, b_{i3}$$

and the  $j$ th column of  $\mathbf{A}$

$$a_{1j}$$

$$a_{2j}$$

$$a_{3j}$$

multiplying the elements in the corresponding positions and adding them up, that is

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

This row-by-column method of multiplication for matrices that perhaps some of you have seen before is not artificial, but a natural consequence of the attempt to solve the problem we posed; namely, to calculate the matrix of  $\mathbf{S} \cdot \mathbf{T}$  from  $\mathbf{A}$  and  $\mathbf{B}$ .

**EXAMPLE 3.** Let  $\mathbf{S}$  and  $\mathbf{T}$  be the linear transformations introduced in Examples 1 and 2. Calculate the matrix of the linear transformation  $\mathbf{S} \cdot \mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to the standard basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  of  $\mathbb{R}^3$ .

**Solution.** Recall that we found the matrices

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

to be the matrices of  $\mathbf{S}$  and  $\mathbf{T}$  respectively. Therefore using the row by column multiplication procedure described above we find the required matrix is:

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

**EXAMPLE 4.** With  $\mathbf{S}$  and  $\mathbf{T}$  as in Examples 1 and 2 calculate the matrix of  $\mathbf{T} \cdot \mathbf{S} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to the standard basis of  $\mathbb{R}^3$ .

**Solution.** We have to employ row-by-column multiplication in the opposite order from Example 3. We write

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

and so the required matrix is:

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

which is quite a bit different from Example 3.

The matrix of a linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  gives a very convenient way to specify the transformation. This is particularly so if the matrix has a large number of zero entries. However this may not be the case for a particular  $T$  if we only allow ourselves to use the standard basis  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$  for  $\mathbb{R}^3$ . This suggests that we introduce a matrix for  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to *any* ordered basis  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  for  $\mathbb{R}^3$ , or more generally still, relative to a *pair* of ordered bases  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  and  $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$  for  $\mathbb{R}^3$ . Of course there is nothing sacred about  $\mathbb{R}^3$  and so we might as well consider the general case of a linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  where  $\mathcal{V}$  and  $\mathcal{W}$  are finite-dimensional vector spaces. We will take up this subject in Chapter 12 after some elementary matrix notions in the next chapter.

## EXERCISES

- For each of the following linear transformations of  $\mathbb{R}^3$  to  $\mathbb{R}^3$  calculate the matrix relative to the standard basis:
  - $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  by  $T(x, y, z) = (x + y + z, 0, 0)$
  - $Q: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  by  $Q(x, y, z) = (x, x + y, x + y + z)$
  - $F: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  by  $F(x, y, z) = (x + 2y + 3z, 2x + y, z - x)$
  - $G: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  by  $G(x, y, z) = (y - z, x + y, z - 2x)$ .

- Calculate the matrix of the linear transformations

$$\begin{aligned} T \cdot Q: \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ F \cdot G: \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ Q \cdot F \cdot G: \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \end{aligned}$$

where  $T, Q, F$  and  $G$  are as in Exercise (1).

- Suppose that the matrix of the linear transformation  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is

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

relative to the standard basis of  $\mathbb{R}^3$ . Calculate  $S(1, 2, 3)$ .

## 10: Matrices and linear transformations

4. Let  $P: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by

$$P(x, y, z) = (x, y, 0).$$

Calculate the matrix for  $P$  and for  $P^2$  relative to the standard basis of  $\mathbb{R}^3$ .

5. Let  $S, T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be linear transformations with matrices  $A$  and  $B$  relative to the standard basis of  $\mathbb{R}^3$ . What is the matrix of  $S + T$  relative to this basis?
6. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear extension of

$$T(\mathbf{E}_1) = (-1, 0, 2)$$

$$T(\mathbf{E}_2) = (1, 1, -1)$$

$$T(\mathbf{E}_3) = (1, -3, 4).$$

Calculate the matrix of  $T$  relative to the standard basis.

7. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear extension of

$$T(1, 1) = (1, -1)$$

$$T(1, -2) = (2, 2).$$

Calculate the matrix of the linear transformation  $T$  relative to the standard basis.

8. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation defined by

$$T(x, y, z) = (0, y, z).$$

Calculate the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$ .

9. Let  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$  be the linear transformation defined by

$$T(x, y, z, w) = (0, x, y, z).$$

Calculate the matrix of  $T$  relative to the standard basis. Calculate the matrix of  $T^2, T^3, T^4$  relative to the standard basis.

10. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix relative to the standard basis is

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

(a) Calculate  $T(\mathbf{E}_1), T(\mathbf{E}_2), T(\mathbf{E}_3)$ .

(b) Let  $r$  be a real number, calculate the matrix of  $rT$ , relative to the standard basis.

11. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation whose matrix relative to the standard basis is

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

(a) Calculate  $T(\mathbf{E}_1)$  and  $T(\mathbf{E}_2)$ .

(b) Find  $A_1, A_2$  satisfying  $T(A_i) = \mathbf{E}_i, i = 1, 2$ .

12. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix relative to the standard basis is

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

- (a) Calculate  $T(\mathbf{E}_1)$ ,  $T(\mathbf{E}_2)$ ,  $T(\mathbf{E}_3)$ , and  $T(1, 2, 3)$ .  
 (b) Is  $T$  surjective? (Check if  $(1, 0, 0)$  belongs to  $\text{Im } T$ .)

13. Let  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \in \mathbb{R}^3$  and write

$$\mathbf{A}_j = (a_{1j}, a_{2j}, a_{3j}), \quad j = 1, 2, 3.$$

Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear extension of the function  $T(\mathbf{E}_j) = \mathbf{A}_j$ ,  $j = 1, 2, 3$ . Show that the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$  is  $(a_{ij})$ .

14. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a linear transformation, and  $\mathbf{A} = (a_{ij})$  its matrix with respect to the standard basis of  $\mathbb{R}^3$ . Regard the columns of  $\mathbf{A}$  as vectors in  $\mathbb{R}^3$ . Call these three vectors  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ . Show  $\text{Im } T = \mathcal{L}\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$ .
15. If  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear transformation with matrix

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

with respect to the standard basis of  $\mathbb{R}^3$  show that  $T$  is an isomorphism.



# 11 Matrices

In the last chapter we saw that a linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  could be represented (that is, was completely determined by) 9 numbers arranged in a  $3 \times 3$  array. In this chapter we will study such arrays, which are called matrices. We will return to the connection between matrices and linear transformations in the next chapter.

**Definition.** A rectangular array of numbers composed of  $m$  rows and  $n$  columns

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is called an  $m \times n$  matrix (read  $m$  by  $n$  matrix<sup>1</sup>). If there is a possibility of confusing entries from two adjacent columns as a product we will insert commas between the entries of a given row to carefully distinguish which entry belongs to which column.

The elements  $a_{i1}, a_{i2}, \dots, a_{in}$  form the  $i$ th row of  $A$  and the elements

$$\begin{aligned} &a_{1j}, \\ &a_{2j}, \\ &\vdots \\ &a_{mj}, \end{aligned}$$

<sup>1</sup> We also say that the matrix  $A$  is of, or has, size  $m \times n$ .

form the  $j$ th column of  $A$ . We will often write

$$\mathbf{A} = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

for  $\mathbf{A}$ , or simply

$$\mathbf{A} = (a_{ij})$$

when  $m$  and  $n$  are understood from context. Note that the order of the subscripts is important; the first subscript denotes the row and the second subscript the column to which an entry belongs.

Just as with vectors in  $\mathbb{R}^n$ , two matrices are equal iff they have the same entries. That is:

**Definition.** If  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$  are  $m \times n$  matrices, then  $\mathbf{A} = \mathbf{B}$  iff  $a_{ij} = b_{ij}$  for  $i = 1, 2, \dots, m$  and  $j = 1, \dots, n$ .

Our study of linear transformations suggests the following definitions.

**Definition.** If  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$  are two  $m \times n$  matrices their *sum*,  $\mathbf{A} + \mathbf{B}$ , is the matrix  $\mathbf{C} = (c_{ij})$  where  $c_{ij} = a_{ij} + b_{ij}$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ .

**Definition.** If  $\mathbf{A} = (a_{ij})$  is an  $m \times n$  matrix and  $r$  is a number then  $r\mathbf{A}$ , the *scalar multiple of  $\mathbf{A}$  by  $r$* , is the matrix  $\mathbf{C} = (c_{ij})$  where  $c_{ij} = ra_{ij}$ ,  $i = 1, \dots, m$  and  $j = 1, \dots, n$ .

The following result is a routine verification of definitions:

**Proposition 11.1.** *The matrices of size  $m \times n$  form a vector space under the operations of matrix addition and scalar multiplication. We denote this vector space by  $\mathcal{M}_{mn}$ .*  $\square$

The dimension of the vector space  $\mathcal{M}_{mn}$  is not hard to compute. We take our lead from the method we used to show that  $\dim \mathbb{R}^n = n$ . Introduce the  $m \times n$  matrix  $\mathbf{E}_{rs} = (e_{ij})$  by the requirement

$$e_{ij} = \begin{cases} 0 & \text{if } i \neq r, j \neq s, \\ 1 & \text{if } i = r, j = s. \end{cases}$$

For example the  $6 \times 4$  matrix  $\mathbf{E}_{32}$  is

$$\mathbf{E}_{32} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is then a routine verification to prove:

**Proposition 11.2.** *The vectors  $\{\mathbf{E}_{rs} | r = 1, 2, \dots, m, s = 1, 2, \dots, n\}$  form a basis for  $\mathcal{M}_{mn}$ . Therefore  $\dim \mathcal{M}_{mn} = mn$ .*  $\square$

Recall now that if  $\mathcal{V}$  and  $\mathcal{W}$  are vector spaces we introduced the vector space  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  of linear transformations from  $\mathcal{V}$  to  $\mathcal{W}$ . If  $\mathcal{V}$  and  $\mathcal{W}$  are finite dimensional then we will see that the innocent looking (11.2) implies the useful fact  $\dim \mathcal{L}(\mathcal{V}, \mathcal{W}) = \dim \mathcal{V} \dim \mathcal{W}$ .

EXAMPLE 1.

$$\begin{pmatrix} 3 & 1 & 4 \\ 2 & 0 & -1 \\ -2 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 3 \\ 2 & -9 & 4 \\ 7 & 6 & 1 \end{pmatrix} \\ = \begin{pmatrix} 3+0 & 1-1 & 4+3 \\ 2+2 & 0-9 & -1+4 \\ -2+7 & -1+6 & 0+1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 7 \\ 4 & -9 & 3 \\ 5 & 5 & 1 \end{pmatrix}.$$

EXAMPLE 2.

$$\begin{pmatrix} 1 & 4 & 6 & 0 & 1 \\ 2 & 0 & -1 & 7 & 9 \end{pmatrix} + \begin{pmatrix} 0 & 1 & -1 & 4 & 7 \\ -1 & -2 & -3 & 7 & 9 \end{pmatrix} \\ = \begin{pmatrix} 1+0 & 4+1 & 6-1 & 0+4 & 1+7 \\ 2-1 & 0-2 & -1-3 & 7+7 & 9+9 \end{pmatrix} \\ = \begin{pmatrix} 1 & 5 & 5 & 4 & 8 \\ 1 & -2 & -4 & 14 & 18 \end{pmatrix}.$$

EXAMPLE 3.

$$4 \begin{pmatrix} 3 & 1 \\ 7 & 4 \\ 6 & -4 \end{pmatrix} = \begin{pmatrix} 12 & 4 \\ 28 & 16 \\ 24 & -16 \end{pmatrix}.$$

The discussion of Chapter 10 suggests the following:

**Definition.** If  $\mathbf{A} = (a_{ij})$  is an  $m \times n$  matrix and  $\mathbf{B} = (b_{ij})$  is an  $n \times p$  matrix their *matrix product*  $\mathbf{A} \cdot \mathbf{B}$  is the  $m \times p$  matrix  $\mathbf{AB} = (c_{ij})$  where

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

where  $i = 1, \dots, m, j = 1, \dots, p$ .

Thus the entry of the  $i$ th row and  $j$ th column of the product  $\mathbf{A} \cdot \mathbf{B}$  is obtained by taking the  $i$ th row

$$a_{i1}, a_{i2}, \dots, a_{in}$$

of the matrix  $\mathbf{A}$  and the  $j$ th column of the matrix  $\mathbf{B}$

$$\begin{array}{c} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj}, \end{array}$$

multiplying the corresponding entries together and adding the resulting products, i.e.

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} + \cdots + a_{in}b_{nj},$$

$$i = 1, 2, \dots, m, j = 1, 2, \dots, p.$$

*Note that for the product of  $\mathbf{A}$  and  $\mathbf{B}$  to be defined the number of columns of  $\mathbf{A}$  must be equal to the number of rows of  $\mathbf{B}$ . Thus the order in which the product of  $\mathbf{A}$  and  $\mathbf{B}$  is taken is very important, for  $\mathbf{A} \cdot \mathbf{B}$  can be defined without  $\mathbf{B} \cdot \mathbf{A}$  being defined.*

**EXAMPLE 4.** Compute the matrix product

$$(1 \quad 2 \quad 3) \cdot \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}.$$

*Solution.* Note the answer is a  $1 \times 1$  matrix.

$$(1 \quad 2 \quad 3) \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = (4 + 10 + 18) = (32).$$

Note that the product in the reverse order

$$\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \cdot (1 \quad 2 \quad 3) = \begin{pmatrix} 4 & 8 & 12 \\ 5 & 10 & 15 \\ 6 & 12 & 18 \end{pmatrix}$$

does not even have the same size. So the products in different orders are certainly not equal.

**EXAMPLE 5.** Compute the matrix product

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

*Answer*

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

EXAMPLE 6. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

and

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

Calculate the product  $\mathbf{A} \cdot \mathbf{B}$ .*Solution.* We have

$$\begin{aligned} & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 1+2+3 & 2+4+6 & 3+6+9 & 4+8+12 \\ 4+5+6 & 8+10+12 & 12+15+18 & 16+20+24 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 12 & 18 & 24 \\ 15 & 30 & 45 & 60 \end{pmatrix}. \end{aligned}$$

Note that the product  $\mathbf{B} \cdot \mathbf{A}$  is *not* defined.

**Definition.** A matrix  $\mathbf{A}$  is said to be a *square matrix of size  $n$*  iff it has  $n$  rows and  $n$  columns (that is the number of rows equals the number of columns equals  $n$ ).

If  $\mathbf{A}$  and  $\mathbf{B}$  are *square matrices of size  $n$*  then the products  $\mathbf{AB}$  and  $\mathbf{BA}$  are both defined, and are square matrices of size  $n$ . However they may not be equal.

EXAMPLE 7. Let

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

Compute the matrix products  $\mathbf{AB}$  and  $\mathbf{BA}$ .

*Solution.* We have

$$\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 6 & 3 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 2 & 3 \end{pmatrix}$$

and so we see that  $\mathbf{AB} \neq \mathbf{BA}$ .

As the preceding example shows even if  $\mathbf{AB}$  and  $\mathbf{BA}$  are defined we should not expect that  $\mathbf{AB} = \mathbf{BA}$ .

**Notation.** If  $\mathbf{A}$  is a square matrix then  $\mathbf{AA}$  is defined and is denoted by  $\mathbf{A}^2$ . Similarly

$$\begin{array}{c} \mathbf{A} \cdots \mathbf{A} \\ n \text{ times} \end{array}$$

is defined and denoted  $\mathbf{A}^n$ .

**EXAMPLE 8.** Let

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

Calculate  $\mathbf{A}^2$ .

*Solution.* We have

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

Thus not only does matrix multiplication behave strangely in that it is not commutative, it is also possible for the square of a matrix with nonzero entries to have only zero entries. We may summarize the basic rules of matrix operations in the following formulas: (assume that the indicated operations are defined, that is, that the sizes are correct for the operations to make sense).

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\ \mathbf{A} + (\mathbf{B} + \mathbf{C}) &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\ r(\mathbf{A} + \mathbf{B}) &= r\mathbf{A} + r\mathbf{B} \\ (r + s)\mathbf{A} &= r\mathbf{A} + s\mathbf{A} \\ 0\mathbf{A} &= \mathbf{0} \quad (\text{where } \mathbf{0} = (0_{ij}) \text{ and } 0_{ij} = 0) \\ \mathbf{A} + (-1)\mathbf{A} &= \mathbf{0} \\ \mathbf{A} + \mathbf{0} &= \mathbf{A} \\ (\mathbf{A} + \mathbf{B}) \cdot \mathbf{C} &= \mathbf{A} \cdot \mathbf{C} + \mathbf{B} \cdot \mathbf{C} \\ \mathbf{C} \cdot (\mathbf{A} + \mathbf{B}) &= \mathbf{C} \cdot \mathbf{A} + \mathbf{C} \cdot \mathbf{B} \\ \mathbf{0} \cdot \mathbf{A} &= \mathbf{0} = \mathbf{A} \cdot \mathbf{0} \\ \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{C}) &= (\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{C}. \end{aligned}$$

Of course the first few of these rules were needed to prove (11.1).

In discussing matrices it is convenient to distinguish certain special types of matrices.

### Special types of matrices

*The identity matrix.* The identity matrix of size  $n$  is the square  $n \times n$  matrix denoted by  $\mathbf{I}$ , where

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & & 1 \end{pmatrix} = (i_{ij}) \quad \text{where } i_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

For example, the identity matrices of size 1, 2, 3, and 4 are

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

The following important facts are easily verified:

$$\mathbf{IB} = \mathbf{B} \quad \text{for any } n \times p \text{ matrix } \mathbf{B},$$

$$\mathbf{AI} = \mathbf{A} \quad \text{for any } m \times n \text{ matrix } \mathbf{A}.$$

*Scalar matrices.* A square matrix  $\mathbf{A} = (a_{ij})$  is called a *scalar matrix* iff  $\mathbf{A} = r\mathbf{I}$  for some number  $r$ . For example

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} = 3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is a scalar matrix but

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

is not a scalar matrix. The following formulas are easily checked

$$(a\mathbf{I})\mathbf{B} = a\mathbf{B} \quad \text{for any } n \times p \text{ matrix } \mathbf{B}.$$

$$\mathbf{A}(a\mathbf{I}) = a\mathbf{A} \quad \text{for any } m \times n \text{ matrix } \mathbf{A}.$$

For example

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

*Diagonal matrices.* For any square matrix  $A = (a_{ij})$  of size  $n$  the entries

$$a_{11}, a_{22}, \dots, a_{nn}$$

are called the *diagonal entries* of  $A$ . For example, the diagonal entries of

$$\begin{pmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \\ 9 & 8 & 7 \end{pmatrix}$$

are 3, 5, 7. A square matrix is said to be a *diagonal matrix* iff its only nonzero entries are on the diagonal. That is  $A = (a_{ij})$  is a diagonal matrix iff  $a_i = 0$  for  $i \neq j$ .

For example  $I$  and  $aI$  are diagonal matrices as is

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

Note that the diagonal entries themselves need not be nonzero. For example

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

are also diagonal matrices.

In general a diagonal matrix looks like

$$A = \begin{pmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{pmatrix}$$

where the giant 0s mean that all other entries are zero. If  $A$  and  $B$  are diagonal matrices of size  $n$  then so are  $AB$  and  $BA$ . Indeed if

$$A = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & & 0 \\ & \ddots & \\ 0 & & b_{nn} \end{pmatrix}$$

then

$$AB = \begin{pmatrix} a_{11}b_{11} & & 0 \\ & a_{22}b_{22} & \\ & & \ddots & \\ 0 & & & a_{nn}b_{nn} \end{pmatrix} = BA.$$



**Triangular matrices.** A square matrix  $A$  is said to be *lower triangular* iff  $A = (a_{ij})$  where  $a_{ij} = 0$  if  $j > i$ . For example

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

is a lower triangular matrix. A triangular matrix  $A = (a_{ij})$  for which  $a_{ii} = 0$ ,  $i = 1, \dots, n$  (that is, all of whose diagonal entries are 0) is said to be *strictly triangular*. An example of a strictly triangular matrix is

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

**The Zero matrix.** The *zero matrix* of size  $m \times n$  is the  $m \times n$  matrix  $0$  all of whose entries are 0.

**Idempotent matrices.** A square matrix  $A$  is said to be *idempotent* iff  $A^2 = A$ . There are lots of idempotent matrices. Here are a few examples

$$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

as may be easily checked by explicit computation.

**Nilpotent matrices.** A square matrix  $A$  is said to be *nilpotent* iff there is an integer  $q$  such  $A^q = 0$ . (The smallest such integer  $q$  is called the *index of nilpotence* of  $A$ .) For example if  $A$  is the matrix of the shift operator on  $\mathbb{R}^3$ , that is

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

then

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

and

$$A^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so that  $A$  is nilpotent of index 3.

As a matter of fact there is a very simple geometric explanation for why  $A^3 = 0$ . Indeed any strictly lower triangular matrix

$$B = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ b & c & 0 \end{pmatrix}$$

will be nilpotent of index at most 3. Of course we could prove this by an orgy of calculation but that is not what we want to do. Rather we are going to exploit the relation between matrices and linear transformations that we started to discuss in the preceding section. Let us therefore construct a linear transformation

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

whose matrix is **B**. To do this we must recall that if there is such a **T** then it must satisfy

$$T(1, 0, 0) = (0, a, b)$$

$$T(0, 1, 0) = (0, 0, c)$$

$$T(0, 0, 1) = (0, 0, 0)$$

if its matrix is going to be **B**. (Remember how to compute the columns of **B**!) Therefore

$$\begin{aligned} T(x, y, z) &= T(x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)) \\ &= xT(1, 0, 0) + yT(0, 1, 0) + zT(0, 0, 1) \\ &= x(0, a, b) + y(0, 0, c) + z(0, 0, 0) \\ &= (0, ax, bx + cy), \end{aligned}$$

that is,

$$T(x, y, z) = (0, ax, bx + cy)$$

is the only linear transformation

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

whose matrix is **B**. Let us now compute

$$T^2: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad T^3: \mathbb{R}^3 \rightarrow \mathbb{R}^3.$$

We have

$$\begin{aligned} T^2(x, y, z) &= T(T(x, y, z)) = T(0, ax, bx + cy) \\ &= (0, 0, cax) \end{aligned}$$

and

$$T^3(x, y, z) = T(T^2(x, y, z)) = T(0, 0, cax) = (0, 0, 0).$$

Therefore

$$T^3 = 0: \mathbb{R}^3 \rightarrow \mathbb{R}^3.$$

Hence the matrix of  $T^3$  relative to the standard basis of  $\mathbb{R}^3$  is

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

But now remember that our definition of matrix multiplication was rigged up so that  $\mathbf{B}^3$  would be the matrix of  $T^3$  relative to the standard basis of  $\mathbb{R}^3$  and therefore we have shown that

$$\mathbf{B}^3 = \mathbf{0},$$

and hence that  $\mathbf{B}$  is nilpotent of index at most 3.

Thus we see the geometric reason behind the fact that  $\mathbf{B}^3 = \mathbf{0}$  is that  $\mathbf{B}$  is the matrix of a linear transformation that is basically a shift on  $\mathbb{R}^3$ . (Actually  $T$  is what is called a *weighted* shift, the weights being  $a, b, c$ .) This discussion should show that it is sometimes possible, and indeed advantageous to discover properties of a matrix by examining a linear transformation that it represents. We will see more examples of this later.

*Nonsingular matrices.* A square matrix  $\mathbf{A}$  is said to be *invertible* or *nonsingular* iff there exists a matrix  $\mathbf{B}$  such that

$$\mathbf{AB} = \mathbf{I} \quad \text{and} \quad \mathbf{BA} = \mathbf{I}.$$

If  $\mathbf{A}$  is nonsingular then the matrix  $\mathbf{B}$  with  $\mathbf{AB} = \mathbf{I} = \mathbf{BA}$  is called the *inverse matrix* of  $\mathbf{A}$  and is denoted by  $\mathbf{A}^{-1}$ .

It is a theorem that we will prove in the next chapter that if there exists a matrix  $\mathbf{B}$  such that

$$\mathbf{AB} = \mathbf{I}$$

then also

$$\mathbf{BA} = \mathbf{I}.$$

Thus to check that  $\mathbf{B} = \mathbf{A}^{-1}$  we need only calculate one of the two products  $\mathbf{AB}$  and  $\mathbf{BA}$  and see if it is  $\mathbf{I}$ . For example if

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

then

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

For we have

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

and therefore  $\mathbf{A} = \mathbf{A}^{-1}$ . There is actually a simple geometric explanation for why  $\mathbf{A} = \mathbf{A}^{-1}$ . For a moment's reflection on our discussion of matrices of linear transformations shows that  $\mathbf{A}$  is the matrix of the linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$T(x, y, z) = (y, x, z)$$

that switches the first two coordinates, relative to the standard bases for  $\mathbb{R}^3$ . Clearly switching the first two coordinates twice will change nothing, that is

$$T^2(x, y, z) = T(T(x, y, z)) = T(y, x, z) = (x, y, z).$$

Therefore the matrix of

$$T^2: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

relative to the standard basis of  $\mathbb{R}^3$  is

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

But now remember that the matrix of  $T^2$  is also  $\mathbf{A}^2$ , because that is how we defined matrix multiplication. Thus  $\mathbf{A}^2 = \mathbf{I}$  for the simple geometric reason that it is the matrix of the transformation  $T$  that switches the first two coordinates.

The preceding discussion illustrates again that it is sometimes possible to extract information about a matrix by examining the corresponding linear transformation.

A matrix  $\mathbf{A}$  with the property that  $\mathbf{A} = \mathbf{A}^{-1}$  is called *involutory* or an *involution*. The example above shows that there are nontrivial involutions.

Another example of a nonsingular matrix is

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

The inverse of  $\mathbf{B}$  is the matrix

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

as we may certainly compute

$$\mathbf{BC} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{I}.$$

An example of a matrix that is *not* invertible is

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

and more generally we have:

A nilpotent matrix is not invertible. For suppose that  $A$  is a nilpotent matrix that is invertible. Let  $B$  be an inverse for  $A$ . Since  $A$  is nilpotent there is an integer  $q$  such that

$$A^q = 0.$$

Then

$$0 = A^q B = A^{q-1} AB = A^{q-1} I = A^{q-1}$$

so

$$A^{q-1} = 0.$$

We may then repeat the above trick to show

$$A^{q-2} = 0.$$

If we repeat this trick  $q - 1$  times we will get

$$A = 0.$$

But then

$$I = AB = 0B = 0$$

which is impossible.

We may also show:

The only invertible idempotent matrix is  $I$ . For if  $A$  is an idempotent matrix then  $A^2 = A$ . If in addition  $A$  is invertible with inverse  $B$  then

$$A^2 = A$$

implies

$$A = IA = BAA = BA^2 = BA = I$$

so  $A = I$  as claimed.

*Symmetric and skew-symmetric matrices.* A square matrix  $A = (a_{ij})$  is said to be *symmetric* iff  $a_{ij} = a_{ji}$  for  $i, j = 1, \dots, n$ ; it is said to be *skew-symmetric* iff  $a_{ij} = -a_{ji}$  for  $i, j = 1, \dots, n$ . For example

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 4 & 5 & 6 \\ 2 & 5 & 7 & 8 \\ 3 & 6 & 8 & 9 \end{pmatrix}$$

are symmetric matrices, and

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

are skew-symmetric matrices. Notice that the matrix

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

is not skew-symmetric because

$$a_{11} = 1 \neq -1 = -a_{11}.$$

That is to say, if a matrix  $\mathbf{A} = (a_{ij})$  is skew-symmetric then the equations  $a_{11} = -a_{11}$ ,  $a_{22} = -a_{22}$ ,  $\dots$ ,  $a_{nn} = -a_{nn}$  certainly imply that  $a_{11} = 0$ ,  $a_{22} = 0$ ,  $\dots$ ,  $a_{nn} = 0$ , that is, a skew-symmetric matrix has all its diagonal entries equal to 0.

The skew-symmetric matrix

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

is interesting because it is also nonsingular since

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

Thus various combinations of the preceding concepts may occur simultaneously. By now the student must be wondering how to tell when a matrix has an inverse, and how to calculate it when it does. This is a topic we will take up in Chapter 13, for the moment we will content ourselves with the  $2 \times 2$  case.

**Proposition 11.3.** A  $2 \times 2$  matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is nonsingular iff  $ad - bc \neq 0$ . If  $ad - bc \neq 0$  then

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

**PROOF.** Suppose that  $ad - bc \neq 0$ . Let

$$\mathbf{B} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

then

$$\begin{aligned}
 \mathbf{BA} &= \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
 &= \frac{1}{ad - bc} \begin{pmatrix} da - bc & bd - bd \\ -ca + ac & -cb + ad \end{pmatrix} \\
 &= \frac{1}{ad - bc} \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I},
 \end{aligned}$$

and therefore  $\mathbf{A}$  is nonsingular with

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

as claimed.

Suppose conversely that  $\mathbf{A}$  is nonsingular, but that  $ad - bc = 0$ . We will deduce a contradiction. Let

$$\mathbf{C} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Then computing as above

$$\mathbf{CA} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (ad - bc)\mathbf{I} = \mathbf{0}$$

This gives the equation

$$\mathbf{C} = \mathbf{CI} = \mathbf{C}(\mathbf{AA}^{-1}) = (\mathbf{CA})\mathbf{A}^{-1} = \mathbf{0A}^{-1} = \mathbf{0}.$$

Therefore

$$\mathbf{C} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

so that

$$a = 0, \quad b = 0, \quad c = 0, \quad d = 0.$$

But then  $\mathbf{A} = \mathbf{0}$  also, so

$$\mathbf{I} = \mathbf{AA}^{-1} = \mathbf{0A}^{-1} = \mathbf{0}$$

so

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

and hence  $1 = 0$  which is impossible. □

The student may wonder how the above result was originally discovered. Did somebody just make a lucky guess? Perhaps, but a more logical development may be found in Chapter 13.

## EXERCISES

1. Perform the following matrix computations:

$$(a) \quad 2 \begin{pmatrix} 3 & 1 & 4 & -1 \\ 2 & 0 & -1 & 2 \end{pmatrix} - 3 \begin{pmatrix} 1 & 0 & -1 & 7 \\ -1 & -2 & 0 & -4 \end{pmatrix}$$

$$(b) \quad 2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - 4 \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + 8 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

2. Perform the following matrix multiplications:

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

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

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

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

3. Which of the following matrices are nonsingular, involutory, idempotent, nilpotent, symmetric, or skew-symmetric?

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

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \mathbf{G} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \mathbf{J} = \begin{pmatrix} 4 & 1 \\ 0 & 2 \end{pmatrix}$$

$$\mathbf{E} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \mathbf{K} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Find the inverse for those that are invertible.



4. Show that a diagonal matrix

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

is nonsingular iff  $a_{11}a_{22}a_{33} \neq 0$ . That is  $a_{11} \neq 0, a_{22} \neq 0, a_{33} \neq 0$ . If  $A$  has an inverse what is it?

5. Show that a diagonal matrix

$$A = \begin{pmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots & \\ & & & a_{nn} \end{pmatrix}$$

is nonsingular iff all its diagonal entries are nonzero.

6. If
- $A = (a_{ij})$
- is a matrix we define the
- transpose*
- of
- $A$
- to be the matrix
- $A' = (b_{ij})$
- where
- $b_{ij} = a_{ji}$
- . Find the transpose of each of the following matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

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

Show for any matrix  $A$  that  $(A')' = A$ .

7. Let  $A$  be a square matrix, show that  $A$  is symmetric iff  $A = A'$ ,  $A$  is skew-symmetric iff  $A = -A'$ .
8. Show that the product of two lower triangular matrices is again lower triangular. If you cannot work the general case do the  $2 \times 2$  and  $3 \times 3$  cases.
9. For any square matrix  $A$  show that  $A + A'$  is symmetric and  $A - A'$  is skew-symmetric.
10. Let  $A$  and  $B$  be  $3 \times 3$  matrices. Show that  $A'B' = (BA)'$ .
11. Let  $A$  be an idempotent matrix. Show that  $I - A$  is also idempotent.
12. Show that a matrix  $A$  is involutory iff  $(I - A)(I + A) = 0$ .
13. Let  $A = (a_{ij})$  be a  $3 \times 3$  matrix. Compute  $AE_{rs}$  and  $E_{rs}A$ .
14. A square matrix  $A$  is said to commute with a matrix  $B$  iff  $AB = BA$ . When does a  $3 \times 3$  matrix  $A$  commute with the matrix  $E_{rs}$ ?
15. Show that if a  $3 \times 3$  matrix  $A$  commutes with every  $3 \times 3$  matrix  $B$  then  $A$  is a scalar matrix. (Hint: If  $A$  commutes with every matrix  $B$  it commutes with the 9 matrices  $E_{rs}$ ,  $r, s = 1, 2, 3$ . Now use (14).)
16. Find all  $2 \times 2$  matrices that commute with

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

17. Construct a  $3 \times 3$  matrix  $A$  such that  $A^3 = I$ . (Try to think of a simple linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  with  $T^3 = I$  and use its matrix relative to the standard basis.)

18. Let  $A$  be a  $3 \times 3$  matrix,  $D$  be the diagonal matrix

$$D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

(a) Compute  $D \cdot A$ .

(b) Compute  $A \cdot D$ .

19. Let  $A$  be a  $3 \times 3$  matrix, Compute  $E_{12} \cdot A$ ,  $A \cdot E_{12}$ ,  $E_{21} \cdot A$ ,  $A \cdot E_{21}$ . What conclusion can you obtain in general for  $E_{rs} \cdot A$  and  $A \cdot E_{rs}$ ?

20. If  $A$  is an idempotent square matrix show  $I - 2A$  is invertible. (Hint: Idempotents correspond to projections. Interpret  $I - 2A$  as a reflection. Try the  $2 \times 2$  case first, then try to generalize.)

21. Let  $A, B$  and  $A + B$  be invertible  $n \times n$  matrices. Show that  $A^{-1} + B^{-1}$  is invertible. (Hint: If you cannot think of anything else try to show

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

22. Find all the  $2 \times 2$  matrices  $A$  satisfying the equation  $A^2 + I = 0$ .

23. Find the inverse of the matrix

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

24. Show that

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

is invertible. Find its inverse. Solve the equation

$$\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

for  $x$  and  $y$ .

25. Let  $A \in \mathbb{R}^n$ . Show that  $A = 0$  iff the matrix product  $AA'$  is the zero matrix.

# 12 Representing linear transformations by matrices

Let us return now to the ideas we developed in Chapter 10 for representing a linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  by a  $3 \times 3$  matrix. There is of course nothing sacred about  $\mathbb{R}^3$  and its standard basis as the following discussion will show.

Let us suppose given a linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{W}$$

between the finite-dimensional vector spaces  $\mathcal{V}$  and  $\mathcal{W}$ . Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  be an *ordered* basis (that is, a basis whose vectors are placed in a specified order) for  $\mathcal{V}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  an ordered basis for  $\mathcal{W}$ . It is then possible to find unique numbers  $a_{ij}$ ,  $i = 1, \dots, n$ ;  $j = 1, \dots, m$  such that

$$\begin{aligned} T(\mathbf{A}_1) &= a_{11}\mathbf{B}_1 + a_{21}\mathbf{B}_2 + \cdots + a_{m1}\mathbf{B}_m, \\ T(\mathbf{A}_2) &= a_{12}\mathbf{B}_1 + a_{22}\mathbf{B}_2 + \cdots + a_{m2}\mathbf{B}_m, \\ &\vdots \\ T(\mathbf{A}_n) &= a_{1n}\mathbf{B}_1 + a_{2n}\mathbf{B}_2 + \cdots + a_{mn}\mathbf{B}_m, \end{aligned}$$

which we have seen may be written more compactly as

$$T(\mathbf{A}_j) = \sum_{i=1}^m a_{ij}\mathbf{B}_i \quad j = 1, 2, \dots, n.$$

The  $m \times n$  matrix  $\mathbf{A} = (a_{ij})$  is called the *matrix of  $T$  relative to the ordered bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$* . Note that just as in the case of  $\mathcal{V} = \mathbb{R}^3 = \mathcal{W}$  that the *columns* of  $\mathbf{A}$  are the coordinates of the vectors  $T(\mathbf{A}_j)$  relative to the basis  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . Note also that in saying “ $\mathbf{A}$  is the matrix of  $T$ ” you **must** also specify relative to which pairs of ordered bases.

EXAMPLE 1. Let

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

be the linear transformation given by

$$T(x, y) = (y, x).$$

Calculate the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^2$ .

*Solution.* By the “standard basis of  $\mathbb{R}^2$ ” we mean the basis  $\mathbf{E}_1 = (1, 0)$ ,  $\mathbf{E}_2 = (0, 1)$ , and that we are to use this ordered basis in both the domain and range of  $T$ . We have

$$T(1, 0) = (0, 1) \quad T(0, 1) = (1, 0)$$

so the matrix we seek is

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

EXAMPLE 2. With  $T$  as in Example 1 find the matrix of  $T$  relative to the pair of ordered bases  $\{\mathbf{E}_1, \mathbf{E}_2\}$  and  $\{\mathbf{F}_1, \mathbf{F}_2\}$  where  $\mathbf{F}_1 = (0, 1)$ ,  $\mathbf{F}_2 = (1, 0)$ . (Here  $\{\mathbf{E}_1, \mathbf{E}_2\}$  is the basis in the domain of  $T$  and  $\{\mathbf{F}_1, \mathbf{F}_2\}$  the basis for the range of  $T$ .)

*Solution.* We still have

$$T(1, 0) = (0, 1) \quad T(0, 1) = (1, 0)$$

but now we must write these equations as

$$T(\mathbf{E}_1) = 1\mathbf{F}_1 + 0\mathbf{F}_2 \quad T(\mathbf{E}_2) = 0\mathbf{F}_1 + 1\mathbf{F}_2$$

so that

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

is the matrix that we now seek.

The lesson to be learned from the preceding example is that appearance can be deceiving! It also suggests that we might profitably inquire into when two matrices represent the same linear transformation relative to different ordered bases. Before doing so let us see a few more examples.

EXAMPLE 3. Calculate the matrix of the differentiation operator

$$D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$$

relative to the usual basis  $\{1, x, x^2, \dots, x^n\}$  for  $\mathcal{P}_n(\mathbb{R})$ .

*Solution.* We have for  $m = 1, 2, \dots$

$$Dx^m = mx^{m-1} = 0 + 0x + \dots + mx^{m-1} + 0x^m + \dots + 0x^n$$

and

$$D(1) = 0.$$

Thus the matrix that we seek is

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & n \\ 0 & 0 & 0 & & 0 \end{pmatrix}.$$

For example

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{when } n = 1$$

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{when } n = 2$$

$$N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{when } n = 3$$

etc. Notice that all the nonzero entries are along the *superdiagonal*. (The superdiagonal of a square matrix  $A = (a_{ij})$  are the entries  $a_{1,2}, a_{2,3}, \dots, a_{n-1,n}$  of  $A$ .)

There is of course no reason to restrict ourselves to square matrices. For example we have the linear transformation

$$D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_{n-1}(\mathbb{R})$$

and we can ask for its matrix relative to the standard bases of  $\mathcal{P}_n(\mathbb{R})$  and  $\mathcal{P}_{n-1}(\mathbb{R})$ . (What size is this matrix? Answer:  $n \times (n+1)$ .) Computing as before we see the required matrix is

$$N' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & n \end{pmatrix}.$$

For example

$$\mathbf{N}' = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad \text{when } n = 1$$

$$\mathbf{N}' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{when } n = 2$$

$$\mathbf{N}' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \quad \text{when } n = 3$$

etc.

For a related example, try to calculate the matrix of

$$\mathbf{D} : \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_{n+1}(\mathbb{R})$$

(( $n + 1$ ) not  $n - 1$ ). Before you write anything down decide what size it should be.

EXAMPLE 4. Calculate the matrix of the linear transformation:

$$\mathbf{T} : \mathbb{R}^4 \rightarrow \mathcal{P}_1(\mathbb{R})$$

given by

$$\mathbf{T}(a_1, a_2, a_3, a_4) = a_1 + a_3 + (a_2 + a_4)x$$

relative to the ordered bases

$$\mathbf{A}_1 = (1, 1, 1, 1)$$

$$\mathbf{A}_2 = (1, 1, 1, 0)$$

$$\mathbf{A}_3 = (1, 1, 0, 0)$$

$$\mathbf{A}_4 = (1, 0, 0, 0)$$

for  $\mathbb{R}^4$  and

$$\mathbf{B}_1 = 1 + x$$

$$\mathbf{B}_2 = 1 - x$$

for  $\mathcal{P}_1(\mathbb{R})$ .

*Solution.* We have (note  $1 = \frac{1}{2}(\mathbf{B}_1 + \mathbf{B}_2)$ ,  $x = \frac{1}{2}(\mathbf{B}_1 - \mathbf{B}_2)$ )

$$\mathbf{T}(\mathbf{A}_1) = \mathbf{T}(1, 1, 1, 1) = 2 + 2x = 2\mathbf{B}_1 + 0\mathbf{B}_2$$

$$\mathbf{T}(\mathbf{A}_2) = \mathbf{T}(1, 1, 1, 0) = 2 + x = \frac{3}{2}\mathbf{B}_1 + \frac{1}{2}\mathbf{B}_2$$

$$\mathbf{T}(\mathbf{A}_3) = \mathbf{T}(1, 1, 0, 0) = 1 + x = \mathbf{B}_1 + 0\mathbf{B}_2$$

$$\mathbf{T}(\mathbf{A}_4) = \mathbf{T}(1, 0, 0, 0) = 1 = \frac{1}{2}\mathbf{B}_1 + \frac{1}{2}\mathbf{B}_2$$

so the matrix that we seek is

$$\begin{pmatrix} 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

## 12: Representing linear transformations by matrices

EXAMPLE 5. Let

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

be given by

$$T(x, y, z) = (x + y, y + z).$$

Calculate the matrix of  $T$  relative to (a) the standard bases of  $\mathbb{R}^3$  and  $\mathbb{R}^2$ ,  
(b) the bases

$$\begin{aligned} \mathbf{A}_1 &= (1, 0, 0) & \mathbf{B}_1 &= (1, 0) \\ \mathbf{A}_2 &= (0, 0, 1) & \mathbf{B}_2 &= (0, 1) \\ \mathbf{A}_3 &= (1, -1, 1) \end{aligned}$$

*Solution.* We have

$$\begin{aligned} T(1, 0, 0) &= (1, 0) \\ T(0, 1, 0) &= (1, 1) \\ T(0, 0, 1) &= (0, 1) \end{aligned}$$

so that

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

is the matrix for Part (a). On the other hand

$$T(1, -1, 1) = (0, 0)$$

so that

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

is the matrix for Part (b).

EXAMPLE 6. Let  $T: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_5(\mathbb{R})$  be the linear transformation given by the formula

$$T(p(x)) = (1 + x - x^2)p(x).$$

Find the matrix of  $T$  relative to the standard bases of  $\mathcal{P}_3(\mathbb{R})$  and  $\mathcal{P}_5(\mathbb{R})$ .

*Solution.* We compute

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

and recalling that the columns of  $T$  are the vectors  $T(1)$ ,  $T(x)$ ,  $T(x^2)$ ,  $T(x^3)$  expressed in terms of the basis  $\{1, x, x^2, x^3, x^4, x^5\}$  we obtain

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

as the required matrix.

As a last example consider:

**EXAMPLE 7.** Let  $S = \{u, v, w\}$  and define

$$E: \mathcal{F}(S) \rightarrow \mathbb{R}$$

by

$$E(f) = f(u) + f(v) + f(w), \quad f \in \mathcal{F}(S).$$

Determine the matrix of  $E$  relative to the ordered bases  $\{\chi_u, \chi_v, \chi_w\}$ ,  $\{1\}$  for  $\mathcal{F}(S)$  and  $\mathbb{R}$  respectively.

*Solution.* We compute

$$E(\chi_u) = \chi_u(u) + \chi_u(v) + \chi_u(w) = 1 + 0 + 0 = 1$$

$$E(\chi_v) = \chi_v(u) + \chi_v(v) + \chi_v(w) = 0 + 1 + 0 = 1$$

$$E(\chi_w) = \chi_w(u) + \chi_w(v) + \chi_w(w) = 0 + 0 + 1 = 1$$

so the required matrix is

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

for  $E$ .

The preceding examples illustrate (I hope!) that the matrix of a linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  can be exceedingly complex if some care is not exercised in the choice of the basis relative to which the matrix is computed. Indeed finding a basis relative to which the matrix is as simple as possible should clearly be one's goal if matrices are to simplify our numerical computations with linear transformations.

With these numerical examples behind us we turn now to a more careful investigation of the relation between linear transformations and matrices.

**Theorem 12.1.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional vector spaces. Suppose that  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  are ordered bases for  $\mathcal{V}$  and  $\mathcal{W}$  respectively. Then assigning to each linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  its matrix relative to these ordered bases defines an isomorphism

$$M: \mathcal{L}(\mathcal{V}, \mathcal{W}) \rightarrow \mathcal{M}_{mn}$$



## 12: Representing linear transformations by matrices

of the vector space of linear transformations from  $\mathcal{V}$  to  $\mathcal{W}$  with the vector space of  $m \times n$  matrices.

**PROOF.** It is clear that  $M$  is a linear transformation. Indeed our definition of matrix addition and scalar multiplication was rigged up so that this would be so. To show that  $M$  is an isomorphism we will actually construct a linear map

$$L: \mathcal{M}_{mn} \rightarrow \mathcal{L}(\mathcal{V}, \mathcal{W})$$

such that

$$L \cdot M(T) = T \quad \text{for all } T \text{ in } \mathcal{L}(\mathcal{V}, \mathcal{W})$$

$$M \cdot L(A) = A \quad \text{for all } A \text{ in } \mathcal{M}_{mn}.$$

So suppose that  $A = (a_{ij})$  is an  $m \times n$  matrix. Define  $T: \mathcal{V} \rightarrow \mathcal{W}$  to be the linear extension of

$$T(A_1) = a_{11}B_1 + a_{21}B_2 + \cdots + a_{m1}B_m,$$

$$T(A_2) = a_{12}B_1 + a_{22}B_2 + \cdots + a_{m2}B_m,$$

$$\vdots$$

$$T(A_n) = a_{1n}B_1 + a_{2n}B_2 + \cdots + a_{mn}B_m.$$

Thus if

$$C = c_1A_1 + c_2A_2 + \cdots + c_nA_n$$

is an arbitrary vector of  $\mathcal{V}$  we have the horrendous formula:

$$\begin{aligned} T(C) &= (a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_n)B_1 + \cdots \\ &\quad + (a_{m1}c_1 + a_{m2}c_2 + \cdots + a_{mn}c_n)B_m \\ &= \sum_{i=1}^m \sum_{j=1}^n a_{ij}c_j B_i. \end{aligned}$$

It is immediately clear that the assignment of the linear transformation  $T$  to the matrix  $A$  defines a linear transformation

$$L: \mathcal{M}_{mn} \rightarrow \mathcal{L}(\mathcal{V}, \mathcal{W}).$$

The matrix of  $L(A)$  is seen by definition to be  $A$ . That is

$$ML(A) = A \quad \text{for all } A \text{ in } \mathcal{M}_{mn}.$$

On the other hand suppose that  $T: \mathcal{V} \rightarrow \mathcal{W}$  is in  $\mathcal{L}(\mathcal{V}, \mathcal{W})$ . Then  $T$  and  $LM(T)$  are given on the basis  $A_1, \dots, A_n$  by

$$T(A_j) = \sum_{i=1}^m a_{ij}B_i = LM(T) \quad j = 1, \dots, n.$$

Since  $\mathbf{A}_1, \dots, \mathbf{A}_n$  is a basis for  $\mathcal{V}$  one has for any vector a unique expression

$$\mathbf{C} = \sum_{j=1}^n c_j \mathbf{A}_j,$$

so we have

$$\mathbf{T}(\mathbf{C}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} c_j \mathbf{B}_i = (\mathbf{LM})(\mathbf{T})(\mathbf{C}),$$

and hence  $\mathbf{T} = \mathbf{LM}(\mathbf{T})$ , that is

$$\mathbf{LM}(\mathbf{T}) = \mathbf{T} \quad \text{for all } \mathbf{T} \text{ in } \mathcal{L}(\mathcal{V}, \mathcal{W}).$$

Therefore  $\mathbf{M}$  is an isomorphism. □

Before examining several interesting consequences of (12.1) let it be noted that the proof of (12.1) is of interest in its own right. Namely, given the following data:

- (1) a finite-dimensional vector space  $\mathcal{V}$  with basis  $\mathbf{A}_1, \dots, \mathbf{A}_n$
- (2) a finite-dimensional vector space  $\mathcal{W}$  with basis  $\mathbf{B}_1, \dots, \mathbf{B}_m$
- (3) an  $m \times n$  matrix  $\mathbf{A} = (a_{ij})$

we may manufacture a linear transformation

$$\mathbf{L}(\mathbf{A}) = \mathbf{T}: \mathcal{V} \rightarrow \mathcal{W}$$

by the formula

$$(*) \quad \mathbf{T}(\mathbf{C}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} c_j \mathbf{B}_i$$

where  $\mathbf{C} = \sum c_j \mathbf{A}_j$ , or what is the same thing, by requiring that the matrix of  $\mathbf{T}$  relative to the ordered bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  be  $\mathbf{A}$ .

**EXAMPLE 8.** Find the value of the linear transformation  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathcal{P}_2(\mathbb{R})$  whose matrix relative to the bases  $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3\}$ ,  $\{1, x, x^2\}$  is

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

on the vector  $\mathbf{C} = (1, 1, -1)$  of  $\mathbb{R}^3$ .

*Solution.* We have

$$\begin{aligned} \mathbf{T}(\mathbf{C}) &= \mathbf{T}(1\mathbf{E}_1 + 1\mathbf{E}_2 - 1\mathbf{E}_3) \\ &= \mathbf{T}(\mathbf{E}_1) + \mathbf{T}(\mathbf{E}_2) - \mathbf{T}(\mathbf{E}_3) \\ &= (1 + 2x + 3x^2) + (4x) - (-1 - 3x + 2x^2) \\ &= 1 + 1 + 2x + 4x + 3x + 3x^2 - 2x^2 \\ &= 2 + 9x + x^2. \end{aligned}$$

A moment's (an hour?) reflection on the above example and Equation (\*) will show that the coordinates of  $T(C)$  relative to the ordered basis  $\{B_1, \dots, B_n\}$  appear as the entries in the *column* vector

$$A \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

Thus we may solve problems such as Example 8 by matrix multiplications.

EXAMPLE 9. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^4$  be the linear transformation whose matrix relative to the standard bases of  $\mathbb{R}^3$  and  $\mathbb{R}^4$  is

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

Calculate the value of  $T$  on  $(1, 2, -4)$ .

*Solution.* We have

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

Therefore  $T(1, 2, 4) = (2, 7, -7, 6)$ .

While it may appear strange that we calculate the value of  $T(C)$  by using the coordinates of  $C$  to make a column matrix and that our answer appears as a column of numbers instead of a row, let us just remark that we could very well have agreed to write vectors in  $\mathbb{R}^n$  as columns of numbers instead of rows. The reason for not doing so is that English reads naturally from left to right. The interchange of rows and columns in the mathematical formalism is an unfortunate, though not accidental occurrence, and is tied up with the difference between covariance and contravariance in physics. In Chinese, column vectors would be perfectly natural.

EXAMPLE 10. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}$  be the linear transformation whose matrix is

$$(1, -2, 1)$$

relative to the standard bases. Find  $T(6, -4, 9)$ .

*Solution.* We have

$$T(6, -4, 9) = (1, -2, 1) \begin{pmatrix} 6 \\ -4 \\ 9 \end{pmatrix} = (6 + 8 + 9) = (23).$$

EXAMPLE 11. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^4$  be the linear transformation with matrix

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

relative to the standard bases of  $\mathbb{R}^3$  and  $\mathbb{R}^4$ . Find bases for the kernel and image of  $T$ .

*Solution.* If  $(x, y, z) \in \mathbb{R}^3$  then by matrix multiplication

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y + z \\ x + y \\ z \end{pmatrix}$$

so

$$T(x, y, z) = (x, y + z, x + y, z),$$

and

$$(x, y, z) \in \ker T \quad \text{iff} \quad \mathbf{0} = (x, y + z, x + y, z)$$

so

$$\begin{aligned} \left. \begin{aligned} 0 &= x \\ 0 &= x + y \end{aligned} \right\} & \quad x = 0 \quad \text{and} \quad y = 0 \\ \left. \begin{aligned} 0 &= y + z \\ 0 &= z \end{aligned} \right\} & \quad z = 0 \quad \text{and} \quad y = 0. \end{aligned}$$

Therefore  $\ker T = \{\mathbf{0}\}$ . To find a basis for the image of  $T$ , note that since  $\ker T = \{\mathbf{0}\}$  the image has dimension 3. Therefore the vectors

$$T(1, 0, 0), \quad T(0, 1, 0), \quad T(0, 0, 1)$$

are a basis for the image. By the definition of the matrix of a linear transformation the components of these vectors are the columns of the matrix, so

$$(1, 0, 1, 0), \quad (0, 1, 1, 0), \quad (0, 1, 0, 1)$$

is a basis for  $\text{Im } T$ .

Let us return now to the consequences of (12.1) we hinted at previously. First we have:

**Corollary 12.2.** *Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional vector spaces. Then  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  is also finite dimensional and moreover*

$$\dim \mathcal{L}(\mathcal{V}, \mathcal{W}) = \dim \mathcal{V} \dim \mathcal{W}.$$

## 12: Representing linear transformations by matrices

**PROOF.** This is immediate from the fact that  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  and  $\mathcal{M}_{mn}$  are isomorphic where  $m = \dim \mathcal{V}$  and  $n = \dim \mathcal{W}$ .  $\square$

To state our next corollary we will need the following very important result.

**Proposition 12.3.** Let  $\mathcal{U}$ ,  $\mathcal{V}$ , and  $\mathcal{W}$  be finite-dimensional vector spaces with bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_p\}$ ,  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ , and  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . Suppose that

$$\mathbf{S}: \mathcal{U} \rightarrow \mathcal{V}, \quad \mathbf{T}: \mathcal{V} \rightarrow \mathcal{W}$$

are linear transformations. Let  $\mathbf{B} = (b_{ij})$  be the matrix of  $\mathbf{S}$  relative to the bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_p\}$ ,  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  and  $\mathbf{A} = (a_{jk})$  the matrix of  $\mathbf{T}$  relative to the bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . Then the matrix of  $\mathbf{T} \cdot \mathbf{S}$  relative to the bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_p\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  is the matrix product  $\mathbf{AB}$ .

**PROOF.** This is an immediate consequence of the definition of the product of two matrices.  $\square$

**EXAMPLE 12.** Let

$$\mathbf{S}: \mathbb{R}^3 \rightarrow \mathbb{R}^4, \quad \mathbf{T}: \mathbb{R}^4 \rightarrow \mathbb{R}^2$$

be the linear transformations whose matrices relative to the standard bases are

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

Find the matrix of the transformation  $\mathbf{T} \cdot \mathbf{S}: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  relative to the standard bases.

**Solution.** According to (12.3) we need only calculate the matrix product  $\mathbf{AB}$  which is

$$\begin{pmatrix} -1 & 0 & 3 & -1 \\ 2 & 4 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ -2 & -1 & -2 \\ 0 & 2 & 5 \\ 4 & 3 & 0 \end{pmatrix} = \begin{pmatrix} -5 & 3 & 16 \\ 2 & 4 & -5 \end{pmatrix}$$

and so the matrix that we seek is

$$\begin{pmatrix} -5 & 3 & 16 \\ 2 & 4 & -5 \end{pmatrix}.$$

**Corollary 12.4.** A linear transformation  $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{V}$  is an isomorphism iff its matrix  $\mathbf{A}$  is invertible. (The matrix  $\mathbf{A}$  is to be computed relative to any pair  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  of ordered bases for  $\mathcal{V}$ .)

PROOF. Suppose that  $T$  is an isomorphism. Let

$$S: \mathcal{V} \rightarrow \mathcal{V}$$

be the linear transformation inverse to  $T$ . Let  $B$  be the matrix of  $S$  relative to the basis pair  $\{B_1, \dots, B_n\}, \{A_1, \dots, A_n\}$ . (N.B. We have interchanged the role of the bases  $\{B_1, \dots, B_n\}$ , and  $\{A_1, \dots, A_n\}$ . Thus if  $B = (b_{ij})$  then  $S(B_j) = \sum b_{ij} A_i$ .) According to (12.3) the matrix product  $AB$  is the matrix of the linear transformation  $T \cdot S: \mathcal{V} \rightarrow \mathcal{V}$  relative to the basis pair  $\{B_1, \dots, B_n\}, \{B_1, \dots, B_n\}$ . But  $T \cdot S(C) = C$  for all  $C \in \mathcal{V}$  since  $T$  and  $S$  are inverse isomorphisms. In particular

$$T \cdot S(B_j) = 0B_1 + 0B_2 + \dots + 0B_{j-1} + 1B_j + 0B_{j+1} + \dots + 0B_n$$

and hence the matrix of  $T \cdot S$  relative to the bases  $\{B_1, \dots, B_n\}, \{B_1, \dots, B_n\}$  is

$$I = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}.$$

Therefore  $AB = I$ . Likewise, according to (12.3) the matrix product  $BA$  is the matrix of the linear transformation  $ST: \mathcal{V} \rightarrow \mathcal{V}$  relative to the basis pair  $\{A_1, \dots, A_n\}, \{A_1, \dots, A_n\}$ . But  $ST(C) = C$  for all  $C$  in  $\mathcal{V}$  because  $S$  and  $T$  are inverse isomorphisms, and hence as before we find  $BA = I$ . This shows that if

$$T: \mathcal{V} \rightarrow \mathcal{V}$$

is an isomorphism then a matrix  $A$  for  $T$  is always invertible.

To prove the converse, we suppose that the matrix  $A$  of  $T$  is invertible. Let  $B$  be a matrix such that

$$AB = I = BA.$$

Let  $S: \mathcal{V} \rightarrow \mathcal{V}$  be the linear transformation whose matrix relative to the ordered bases  $\{B_1, \dots, B_n\}, \{A_1, \dots, A_n\}$  is  $B$ . (Note: We have again interchanged the roles of  $\{B_1, \dots, B_n\}$  and  $\{A_1, \dots, A_n\}$ .) Then the matrix of  $ST: \mathcal{V} \rightarrow \mathcal{V}$  relative to the ordered bases  $\{A_1, \dots, A_n\}, \{A_1, \dots, A_n\}$  is

$$I = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

Therefore  $S \cdot T$  and  $I$  have the same matrix relative to the bases  $\{A_1, \dots, A_n\}, \{A_1, \dots, A_n\}$  so that by (12.1)  $S \cdot T = I$ , that is

$$S \cdot T(C) = C \quad \text{for all } C \text{ in } \mathcal{V}.$$

## 12: Representing linear transformations by matrices

Likewise we see that

$$T \cdot S(C) = C \quad \text{for all } C \text{ in } \mathcal{V}$$

so that  $S$  and  $T$  are inverse isomorphisms.  $\square$

Note that in the proof of (12.4) we used the fact that the matrix of the transformation  $I: \mathcal{V} \rightarrow \mathcal{V}$  defined by  $I(C) = C$  for all  $C$  relative to the bases  $\{A_1, \dots, A_n\}$ ,  $\{A_1, \dots, A_n\}$  is the identity matrix. This is not the case if we have two *different* bases (or even different orderings on the same basis) in  $\mathcal{V}$ .

**EXAMPLE 13.** Find the matrix of the identity linear transformation  $I: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to the ordered bases

$$\begin{aligned} E_1 &= (1, 0, 0), & E_2 &= (0, 1, 0), & E_3 &= (0, 0, 1) \\ F_1 &= (1, 1, 1), & F_2 &= (1, 1, 0), & F_3 &= (1, 0, 0) \end{aligned}$$

of  $\mathbb{R}^3$ .

*Solution.* We have

$$\begin{aligned} I(E_1) &= I(1, 0, 0) = (1, 0, 0) = 0F_1 + 0F_2 + 1F_3 \\ I(E_2) &= I(0, 1, 0) = (0, 1, 0) = 0F_1 + 1F_2 + (-1)F_3 \\ I(E_3) &= I(0, 0, 1) = (0, 0, 1) = 1F_1 + (-1)F_2 + 0F_3. \end{aligned}$$

So the matrix we seek is

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

The moral of the example is that appearances are **really** deceiving.

In view of Example 13 it is reasonable to expect that when we calculate with matrices of transformations  $T: \mathcal{V} \rightarrow \mathcal{V}$  we insist upon using the same ordered basis  $\{A_1, \dots, A_n\}$  twice to do the calculation, rather than work with distinct ordered bases  $\{A_1, \dots, A_n\}$ ,  $\{B_1, \dots, B_n\}$ . Other reasons for insisting on using only one ordered basis  $\{A_1, \dots, A_n\}$  when we study a transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  from the same space to itself will appear in Chapter 14.

We turn now to another consequence of (12.1) which substantiates a remark we made concerning inverses of matrices in the last chapter.

**Corollary 12.5.** Suppose that  $A$  and  $B$  are square matrices of size  $n$  and  $AB = I$ . Then  $BA = I$ . (Thus to check that a square matrix  $B$  is the inverse of a square matrix  $A$  we need only check one of the two conditions  $AB = I$  and  $BA = I$ .)

PROOF. Let  $T, S: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the linear transformations whose matrices with respect to the standard basis of  $\mathbb{R}^n$  are  $A$  and  $B$  respectively. That is

$$L(A) = T, \quad L(B) = S$$

or what is the same thing

$$M(T) = A, \quad M(S) = B.$$

Then by (12.3)

$$M(T \cdot S) = M(T)M(S) = AB = I = M(I).$$

Therefore since  $M$  is an isomorphism (see 8.12)

$$T \cdot S = I.$$

From this we may conclude that  $S: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is an isomorphism as follows. Suppose  $C \in \ker S$ . Then

$$S(C) = 0$$

so

$$0 = T(0) = T(S(C)) = (T \cdot S)(C) = I(C) = C.$$

Therefore  $S$  has kernel  $0$  so by (8.10)

$$n = \dim \mathbb{R}^n = \dim \operatorname{Im} S + \dim \ker S = \dim \operatorname{Im} S.$$

Therefore  $\operatorname{Im} S = \mathbb{R}^n$  by (6.10) and hence  $S$  is an isomorphism by (8.11). Therefore there is a transformation

$$\bar{T}: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

such that

$$S\bar{T} = I.$$

Then

$$\begin{aligned} T(S\bar{T}) &= T(I) = T \\ (TS)\bar{T} &= I\bar{T} = \bar{T} \end{aligned}$$

so that  $T = \bar{T}$ , that is

$$ST = I.$$

Hence

$$I = M(I) = M(ST) = M(S)M(T) = B \cdot A$$

as required. □

While the proof of (12.5) may seem complicated we invite the reader to try and prove the result using only matrices. Even in the  $3 \times 3$  case such a proof will be most painful! The corollaries (12.2) and (12.5) illustrate an



important point, namely some results concerning linear transformations are best proved using matrices, and conversely some results concerning matrices are best proved using linear transformations.

In the study of linear transformations we will very often make use of matrix representations to prove theorems and make computations. The matrix representation for a linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  depends on a choice of bases for  $\mathcal{V}$  and  $\mathcal{W}$ . Our initial choice of these bases may be bad and unsuited to the problem at hand in that we obtain a matrix which does not convey the information we seek. It is therefore quite natural to change the bases to obtain a "better" representation for  $T$ , because to change the bases does nothing to the linear transformation, but it does change the matrix representation. Several natural and important questions therefore arise. First of all, what is the relation (numerically, that is) between two matrix representations of  $T$  computed with respect to different pairs of bases in  $\mathcal{V}$  and  $\mathcal{W}$ ? Secondly, if we are given two  $m \times n$  matrices  $A$  and  $B$ , when (that is, what numerical relations must hold between them) can we conclude that they represent one and the same linear transformation  $T$ , but relative to different bases? As it turns out, both questions may be answered simultaneously, however before doing so we consider one more example of the phenomena under discussion.

**EXAMPLE 14.** Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by

$$T(x, y, z) = (y + z, x + z, y + x).$$

Calculate the matrix of  $T$  relative to

- (a) the standard basis of  $\mathbb{R}^3$ ,
- (b) the basis  $\{(1, 1, 1), (1, -1, 0), (1, 1, -2)\}$  used twice.

*Solution.* To do Part (a) we compute as follows:

$$T(1, 0, 0) = (0, 1, 1)$$

$$T(0, 1, 0) = (1, 0, 1)$$

$$T(0, 0, 1) = (1, 1, 0)$$

Thus the desired matrix is

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

To do the computations of Part (b) let us set  $F_1 = (1, 1, 1)$ ,  $F_2 = (1, -1, 0)$ ,  $F_3 = (1, 1, -2)$ . Then we find

$$T(F_1) = T(1, 1, 1) = (2, 2, 2) = 2F_1 + 0F_2 + 0F_3$$

$$T(F_2) = T(1, -1, 0) = (-1, 1, 0) = 0F_1 - F_2 + 0F_3$$

$$T(F_3) = T(1, 1, -2) = (-1, -1, 2) = 0F_1 + 0F_2 - F_3.$$

So the matrix for Part (b) is

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

This should serve to illustrate that the matrix of a transformation depends heavily on the bases we use to compute it.

**Theorem 12.6.** *Let  $\mathbf{A}$  and  $\mathbf{B}$  be  $m \times n$  matrices,  $\mathcal{V}$  an  $n$ -dimensional vector space and  $\mathcal{W}$  an  $m$ -dimensional vector space. Then  $\mathbf{A}$  and  $\mathbf{B}$  represent the same linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  relative to (perhaps) different pairs of ordered bases iff there exist nonsingular matrices  $\mathbf{P}$  and  $\mathbf{Q}$  such that*

$$\mathbf{A} = \mathbf{PBQ}^{-1}.$$

(Note that  $\mathbf{P}$  is  $m \times m$  and  $\mathbf{Q}$  is  $n \times n$ .)

**PROOF.** There are two things we must prove. First, if  $\mathbf{A}$  and  $\mathbf{B}$  represent the same linear transformation relative to different bases of  $\mathcal{V}$  and  $\mathcal{W}$  we must construct invertible matrices  $\mathbf{P}$  and  $\mathbf{Q}$  such that

$$\mathbf{A} = \mathbf{PBQ}^{-1}.$$

Secondly if we are given invertible matrices  $\mathbf{P}$  and  $\mathbf{Q}$  such that

$$\mathbf{A} = \mathbf{PBQ}^{-1}$$

we must construct a linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{W}$$

and pairs of ordered bases for  $\mathcal{V}$  and  $\mathcal{W}$  so that  $\mathbf{A}$  represents  $T$  relative to one pair and  $\mathbf{B}$  relative to the other. Consider the first of these. We suppose given bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$  such that the matrix of  $T$  relative to these bases is  $\mathbf{A}$ , and bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_n\}$ ,  $\{\mathbf{D}_1, \dots, \mathbf{D}_m\}$  such that the matrix of  $T$  relative to these bases is  $\mathbf{B}$ . Let  $\mathbf{P}$  be the matrix of (remember Example 13)

$$I: \mathcal{W} \rightarrow \mathcal{W}$$

relative to the bases  $\{\mathbf{D}_1, \dots, \mathbf{D}_m\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . Then by (12.4)  $\mathbf{P}$  is invertible. Let  $\mathbf{Q}$  be the matrix of

$$I: \mathcal{V} \rightarrow \mathcal{V}$$

relative to the bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_n\}$ ,  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ . Then by (12.4)  $\mathbf{Q}$  is invertible and by (12.3)  $\mathbf{Q}^{-1}$  represents the matrix of

$$I: \mathcal{V} \rightarrow \mathcal{V}$$

relative to the bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ ,  $\{\mathbf{C}_1, \dots, \mathbf{C}_n\}$ .

Therefore by (12.4)  $\mathbf{PB}$  is the matrix of  $T: \mathcal{V} \rightarrow \mathcal{W}$  relative to the bases  $\{\mathbf{C}_1, \dots, \mathbf{C}_n\}$ ,  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . If we apply (12.4) again we see that  $\mathbf{PBQ}^{-1}$  is the matrix of  $T$  relative to the bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_m\}$ . But  $\mathbf{A}$

is also the matrix of  $T$  relative to the bases  $\{A_1, \dots, A_n\}$  and  $\{B_1, \dots, B_m\}$  so that by (12.1)

$$A = PBQ^{-1}$$

as required.

To prove the converse suppose given invertible matrices  $P$  and  $Q$  such that  $A = PBQ^{-1}$ . Choose bases  $\{A_1, \dots, A_n\}$ ,  $\{B_1, \dots, B_m\}$  for  $\mathcal{V}$  and  $\mathcal{W}$  respectively. Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be the linear transformation whose matrix is  $A$  relative to these bases. Let

$$\begin{aligned} C_1 &= P(A_1), \dots, C_n = P(A_n) \\ D_1 &= Q(B_1), \dots, D_m = Q(B_m). \end{aligned}$$

Since  $P$  and  $Q$  are isomorphisms, the collections  $\{C_1, \dots, C_n\}$ ,  $\{D_1, \dots, D_m\}$  are bases for  $\mathcal{V}$  and  $\mathcal{W}$  respectively. (See for example (8.15) and (8.16).) A brute force computation now shows that  $B$  is the matrix of  $T$  relative to the bases  $\{C_1, \dots, C_n\}$ ,  $\{D_1, \dots, D_m\}$ .  $\square$

We will return to (12.6) in chapter 15 where we will discuss in more detail when  $A$  and  $B$  represent the same linear transformation relative to distinct bases.

**EXAMPLE 15.** Let us see how (12.6) applies to Example 14. Recall that we are given the linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  defined by

$$T(x, y, z) = (y + z, x + z, y + x)$$

and

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

is the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$ , while

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

is the matrix of  $T$  relative to the ordered basis  $\{(1, 1, 1), (1, -1, 0), (1, 1, -2)\}$  of  $\mathbb{R}^3$ .

According to (12.6) there are invertible matrices  $P$  and  $Q^{-1}$  such that  $A = PBQ^{-1}$ . Our task is to compute  $P$  and  $Q^{-1}$ . The proof of (12.6) tells us how. Namely:

- (1)  $P$  is the matrix of  $I: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to the basis pair  $\{(1, 1, 1), (1, -1, 0), (1, 1, -2)\}$  and  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ .
- (2)  $Q^{-1}$  is the matrix of  $I: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  relative to the basis pair  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$  and  $\{(1, 1, 1), (1, -1, 0), (1, 1, -2)\}$ .

The computation of  $\mathbf{P}$  is easy and gives us

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

The computation of  $\mathbf{Q}^{-1}$  is not hard and depends on the following equations

$$(1, 0, 0) = \frac{1}{3}(1, 1, 1) + \frac{1}{2}(1, -1, 0) + \frac{1}{6}(1, 1, -2)$$

$$(0, 1, 0) = \frac{1}{3}(1, 1, 1) - \frac{1}{2}(1, -1, 0) + \frac{1}{6}(1, 1, -2)$$

$$(0, 0, 1) = \frac{1}{3}(1, 1, 1) + 0(1, -1, 0) - \frac{1}{3}(1, 1, -2)$$

so that

$$\mathbf{Q}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{3} & -\frac{1}{2} & \frac{1}{6} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{pmatrix}.$$

A tedious computation shows

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{Q}^{-1},$$

that is,

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ \frac{1}{3} & -\frac{1}{2} & \frac{1}{6} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{pmatrix}.$$

## EXERCISES

- Find the matrix of the following linear transformations relative to the standard bases for  $\mathbb{R}^n$

- (a)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^5$  given by

$$T(a_1, a_2, a_3) = (a_1, a_1 - a_3, a_2 + a_3, a_3, a_1 + a_2)$$

- (b)  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$  given by

$$T(a_1, a_2, a_3, a_4) = (a_1, a_1 + a_2, a_1 + a_2 + a_3, a_1 + a_2 + a_3 + a_4)$$

- (c)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}$  given by

$$T(a_1, a_2, a_3) = a_1 + a_2 + a_3$$

- (d)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^4$  given by

$$T(a_1, a_2) = (a_1 + a_2, 2a_1, 3a_2 - a_1, 2a_1 - a_2).$$

- Find the matrix of the following linear transformations relative to the usual bases of  $\mathcal{P}_n(\mathbb{R})$ .

- (a)  $D: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R})$  (Hint: think about size!)

- (b)  $D: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_2(\mathbb{R})$

- (c)  $T: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathcal{P}_2(\mathbb{R})$  given by  $T(p(x)) = p(x + 1)$

- (d)  $T: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R})$  given by  $T(p(x)) = xp(x)$ .

## 12: Representing linear transformations by matrices

3. Let  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^7$  be the linear transformation whose matrix relative to the standard bases is

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

Find  $T(1, 2, 3, 4)$ .

4. Let  $S: \mathbb{R}^7 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix relative to the standard bases is

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

and let  $T$  be the linear transformation of Problem 3. Find the matrix of  $ST: \mathbb{R}^4 \rightarrow \mathbb{R}^3$  relative to the standard bases. What is  $S \cdot T(1, -1, 1, -1)$ ?

5. Let  $S, T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  be the linear transformations with matrices

$$\begin{pmatrix} 6 & -1 & 2 \\ 2 & 4 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -5 & 0 & -7 \\ 2 & -1 & 9 \end{pmatrix}$$

respectively. What is the matrix of the linear transformation

$$3S - 7T: \mathbb{R}^3 \rightarrow \mathbb{R}^2?$$

Find  $(3S - 7T)(1, 2, 3)$ .

6. Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional vector spaces and  $T: \mathcal{V} \rightarrow \mathcal{W}$  a linear transformation. Show that there are bases  $\{A_1, \dots, A_n\}$ ,  $\{B_1, \dots, B_m\}$  for  $\mathcal{V}$  and  $\mathcal{W}$  such that the matrix of  $T$  is

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

(Hint: Choose a basis  $A'_1, \dots, A'_r$  for  $\ker T$ . Extend this to a basis  $A''_1, \dots, A''_s$ ,  $A'_1, \dots, A'_r$  for  $\mathcal{V}$ . Show that the vectors

$$B_1 = T(A''_1), \quad \dots, \quad B_s = T(A''_s)$$

are linearly independent in  $\mathcal{W}$ . Extend them to a basis  $B_1, \dots, B_s, B_{s+1}, \dots, B_m$  for  $\mathcal{W}$ . Let  $A_1 = A''_1, \dots, A_s = A''_s$ ,  $A_{s+1} = A'_1, \dots, A_n = A'_r$ . Calculate the matrix of  $T$  relative to these bases.)

7. Let

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ b & d & 0 & 0 \\ c & e & f & 0 \end{pmatrix}$$

Show that  $A^4 = 0$ . (Hint: Recall how we did an analogous problem in the  $3 \times 3$  case using the linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  whose matrix relative to the standard basis was the analog of  $A$ .) Can you generalize your result to  $n \times n$ ?

8. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation

$$T(x, y) = (x + 3y, 3x + y).$$

Find the matrix of  $T$  relative to

- (a) the standard basis of  $\mathbb{R}^2$
- (b) the basis  $\{F_1 = (1, 1), F_2 = (1, -1)\}$  used twice.
- (c)  $\{F_1, F_2\}$  in the domain and the standard basis in the range.

9. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation

$$T(x, y, z) = (3y + 4z, 3x, 4x).$$

Calculate the matrix of  $T$  relative to

- (a) the standard basis of  $\mathbb{R}^3$
- (b) the basis  $\{(0, 4, -3), (5, 3, 4), (5, -3, -4)\}$  used twice.

10. Find the matrix of the shift  $S: \mathbb{R}^n \rightarrow \mathbb{R}^n$  relative to the standard basis used twice.

11. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix relative to the standard basis is

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

Find the matrix of  $T$  relative to the basis  $\{F_1, F_2, F_3\}$  used twice, where  $F_1 = (1, 1, 1)$ ,  $F_2 = (1, 1, 0)$ ,  $F_3 = (1, 0, 0)$ .

12. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix relative to the basis of Problem 11,  $\{F_1, F_2, F_3\}$  used twice is

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

Find the matrix of  $T$  relative to the standard basis.

## 12: Representing linear transformations by matrices

13. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation  $T(x, y, z) = (x + y, y + z, z + x)$ .

- Find the matrix of  $T$  relative to the standard basis.
- Find  $A_1, A_2, A_3$  satisfying  $T(A_i) = E_i, i = 1, 2, 3$ .
- Show that  $A_1, A_2, A_3$  form a set of linearly independent vectors.
- Show that  $T(E_1), T(E_2), T(E_3)$  form a set of linearly independent vectors.
- Consider  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  the linear extension of  $S(E_1) = A_1, S(E_2) = A_2, S(E_3) = A_3$ , where  $A_i$  are that of (b).
- Find the matrix of  $S$  relative to the standard basis.
- Show that the matrices obtained in (a) and (f) are inverse matrices of each other.

14. Let  $T: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathbb{R}^2$  be the linear transformation defined by

$$T(p(x)) = \left( p(0), \int_0^1 p(x) dx \right).$$

Find the matrix of  $T$  relative to the ordered basis pair  $\{1, x, x^2, x^3\}$  and  $\{(1, 0), (0, 1)\}$ .

15. Let  $\sigma: \{a, b, c\} \rightarrow \{a, b, c\}$  be given by

$$\sigma(a) = b, \quad \sigma(b) = c, \quad \sigma(c) = a.$$

Find the matrix of

$$\sigma^* = \mathcal{F}(S) \rightarrow \mathbf{F}(S): S = \{a, b, c\}$$

with respect to the basis of characteristic functions.

16. Determine if

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

can represent the same linear transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  with respect to different pairs of ordered bases.

17. Find a pair of  $2 \times 2$  matrices  $A, B$  such that

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

18. Show that a  $3 \times 3$  matrix

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & c \\ 0 & d & e \end{pmatrix}$$

is invertible iff

$$a(be - cd) \neq 0.$$

19. Show that the  $3 \times 3$  matrix

$$\begin{pmatrix} a & 0 & c \\ 0 & b & 0 \\ c & 0 & a \end{pmatrix}$$

is invertible iff  $abc \neq 0$ .

20. Determine if the matrices

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

and

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

represent the same linear transformation with respect to different ordered parts of bases of  $\mathbb{R}^3$ .

21. Let
- $M: \mathbb{C}^2 \rightarrow \mathbb{C}$
- be the linear transformation defined by

$$M(u, v) = u \cdot v,$$

the dot denoting multiplication of complex numbers. Regard  $\mathbb{C}$  and  $\mathbb{C}^2$  as *real* vector spaces using  $\{1, i\}$  as a basis for  $\mathbb{C}$  and  $\{(1, 0), (i, 0), (0, 1), (0, i)\}$  as a basis for  $\mathbb{C}^2$ . Find the matrix of  $M$  with respect to these ordered bases.

22. Show that any non-zero linear combination of the matrices

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is invertible.

23. A vector
- $A \in \mathcal{U}$
- is called a
- fixed point*
- of a linear transformation
- $T: \mathcal{U} \rightarrow \mathcal{U}$
- iff
- $T(A) = A$
- . It is called an
- antifixed point*
- if
- $T(A) = -A$
- . Let
- $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$
- be represented by each of the following matrices in turn and find the fixed points and antifixed points if any

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

24. Let
- $\mathcal{V}$
- and
- $\mathcal{W}$
- be finite-dimensional vector spaces with ordered bases
- $\{E_1, \dots, E_n\}$
- and
- $\{F_1, \dots, F_m\}$
- . Let
- $T: \mathcal{V} \rightarrow \mathcal{W}$
- be a linear transformation with matrix
- $M$
- with respect to this basis pair. Let
- $\mathcal{V}^* = \mathcal{L}(\mathcal{V}; \mathbb{R})$
- and
- $\mathcal{W}^* = \mathcal{L}(\mathcal{W}; \mathbb{R})$
- .

- (a) Define
- $T^*: \mathcal{W}^* \rightarrow \mathcal{V}^*$
- by

$$T^*(f)(A) = f(T(A)), \quad A \in \mathcal{V}, f \in \mathcal{W}^*.$$

Show  $T$  is a linear transformation. (It is called the *dual transformation*.)

- (b) Define
- $E_1^*, \dots, E_n^* \in \mathcal{V}^*$
- to be the linear extension of the functions

$$E_i^*(E_j) = \begin{cases} 1, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Show that  $\{E_1^*, \dots, E_n^*\}$  is a basis for  $\mathcal{V}^*$ , called the *dual basis* to  $\{E_1, \dots, E_n\}$ .

- (c) Show that the matrix of
- $T^*$
- with respect to the dual basis
- $\{F_1^*, \dots, F_m^*\}$
- ,
- $\{E_1^*, \dots, E_n^*\}$
- is the transpose of
- $M$
- .



## 12: Representing linear transformations by matrices

25. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix with respect to the standard basis of  $\mathbb{R}^3$  is

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

Find bases for the kernel and image of  $T$ . Use this to show  $T$  is invertible. Find the matrix of  $T^{-1}$  relative to the standard basis.

Our purpose in this chapter is to develop further the theory and technique of representing linear transformations by matrices. We will touch on several scattered topics and techniques. It is to be emphasized that we are only scratching the surface of an iceberg!

## EXAMPLE 1. *Projections.*

One of the simplest type of linear transformation is a projection. For example in  $\mathbb{R}^3$  consider the plane  $\mathcal{V} = \{(x, y, z) | x + y + z = 0\}$  (Note  $\mathcal{V} \subset \mathbb{R}^3$  is a 2-dimensional subspace.) and the linear transformation  $P: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  that sends a vector  $\mathbf{A}$  into its projection on the plane  $\mathcal{V}$ . See Figure 12<sup>bis</sup>.1. A formula for  $P: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is not hard to find. Notice that for any vector  $\mathbf{A}$  that lies in  $\mathcal{V}$  we must have  $P(\mathbf{A}) = \mathbf{A}$ . For example

$$P(1, -1, 0) = (1, -1, 0)$$

$$P(0, 1, -1) = (0, 1, -1),$$

and notice that  $\{(1, -1, 0), (0, 1, -1)\}$  is a basis for  $\mathcal{V}$ . Notice next that for any vector  $\mathbf{B}$  lying on the line through the origin perpendicular to  $\mathcal{V}$  we must have  $P(\mathbf{B}) = \mathbf{0}$ . For example

$$P(1, 1, 1) = \mathbf{0}.$$

If we now notice that  $\{(1, -1, 0), (0, 1, -1), (1, 1, 1)\}$  is a basis for  $\mathbb{R}^3$  we can easily crank out several formulas for  $P$ . For example, the matrix of  $P$  relative to the basis  $\{(1, -1, 0), (0, 1, -1), (1, 1, 1)\}$  is just

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

Of course you may object that this is cheating somewhat, and what you really want is the matrix of  $P$  relative to the standard basis for  $\mathbb{R}^3$ . This isn't hard either. Notice that

$$\begin{aligned}(1, 0, 0) &= \frac{2}{3}(1, -1, 0) + \frac{1}{3}(0, 1, -1) + \frac{1}{3}(1, 1, 1) \\(0, 1, 0) &= -\frac{1}{3}(1, -1, 0) + \frac{1}{3}(0, 1, -1) + \frac{1}{3}(1, 1, 1) \\(0, 0, 1) &= -\frac{1}{3}(1, -1, 0) - \frac{2}{3}(0, 1, -1) + \frac{1}{3}(1, 1, 1).\end{aligned}$$

Therefore we find that

$$\begin{aligned}P(1, 0, 0) &= \frac{2}{3}P(1, -1, 0) + \frac{1}{3}P(0, 1, -1) + \frac{1}{3}P(1, 1, 1) \\&= \frac{2}{3}(1, -1, 0) + \frac{1}{3}(0, 1, -1) = (\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}) \\P(0, 1, 0) &= -\frac{1}{3}P(1, -1, 0) + \frac{1}{3}P(0, 1, -1) + \frac{1}{3}P(1, 1, 1) \\&= -\frac{1}{3}(1, -1, 0) + \frac{1}{3}(0, 1, -1) = (-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}) \\P(0, 0, 1) &= -\frac{1}{3}P(1, -1, 0) - \frac{2}{3}P(0, 1, -1) + \frac{1}{3}P(1, 1, 1) \\&= -\frac{1}{3}(1, -1, 0) - \frac{2}{3}(0, 1, -1) = (-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3})\end{aligned}$$

so that the matrix of  $P$  relative to the standard basis is

$$N = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

or if one enjoys such things

$$P(x, y, z) = (\frac{2}{3}x - \frac{1}{3}y - \frac{1}{3}z, -\frac{1}{3}x + \frac{2}{3}y - \frac{1}{3}z, -\frac{1}{3}x - \frac{1}{3}y + \frac{2}{3}z).$$

Now the first question to ask is: Which representation of  $P$  tells us the most about it? I am sure you will agree that the matrix representation  $M$  **together with the basis**  $\{(1, -1, 0), (0, 1, -1), (1, 1, 1)\}$  relative to which we computed  $M$  tells us everything we could possibly want to know about  $P$ . For it clearly tells us that  $P$  is projection onto the plane spanned by the vectors  $(1, -1, 0)$  and  $(0, 1, -1)$  which is, of course, the plane  $\mathcal{V} = \{(x, y, z): x + y + z = 0\}$ . On the other hand the matrix representation  $N$

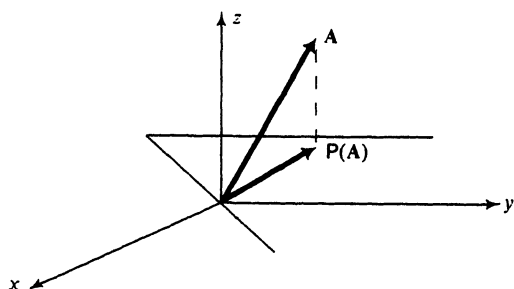


Figure 12<sup>bis</sup>.1

of  $P$  relative to the standard basis conveys relatively little about the simple geometric nature of  $P$ . This brings us to another question: How does one recognize when a linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a projection? Well to deal with this question we had better begin by defining what we mean by a projection. (Remember a definition is either useful or useless, not true or false!) A little experimentation shows that the following is a useful choice.

**Definition.** A linear transformation

$$P: \mathcal{W} \rightarrow \mathcal{W}$$

is called a *projection* iff  $P^2 = P$ , that is iff  $P(P(A)) = P(A)$  for every vector  $A$  in  $\mathcal{W}$ .

Now it pays to be warned that there are linear transformations

$$S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

that are projections in the sense of this definition that are not projections onto a plane (or line) in the sense we have spoken of so far. Let us illustrate by an example.

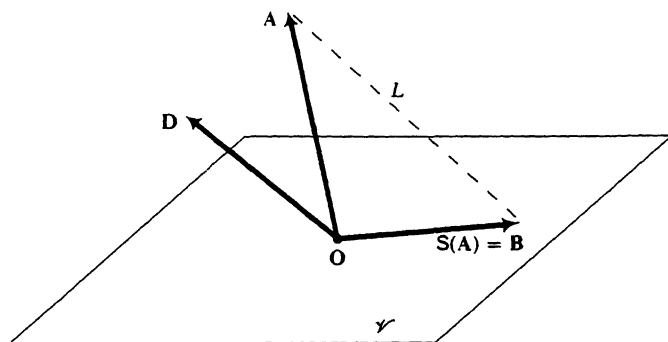


Figure 12<sup>bis</sup>.2

Let  $\mathcal{W} = \{(x, y, z) | x + y + z = 0\}$  be the plane in  $\mathbb{R}^3$  we have considered already. Let  $\mathbf{D} = (1, 1, 0)$ . Note that  $\mathbf{D}$  is not a vector in  $\mathcal{W}$ . We are going to describe a sort of *skew* projection onto  $\mathcal{W}$ . Let  $A$  be a vector in  $\mathbb{R}^3$ . Draw a line  $L$  parallel to  $\mathbf{D}$  passing through the tip of the vector  $A$ . This line will meet the plane  $\mathcal{W}$  in a unique point (why?) which determines a unique vector lying in the plane  $\mathcal{W}$  which we call  $\mathbf{B}$ . Set  $S(A) = \mathbf{B}$  as in Figure 12<sup>bis</sup>.2. It is rather easily checked from the definition that  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is indeed a linear transformation. Note that

$$\begin{aligned} S(A) &= A && \text{if } A \in \mathcal{W} \\ S(A) &= \mathbf{0} && \text{if } A \in \mathcal{L}(\mathbf{D}). \end{aligned}$$

Since  $\mathbf{D}$  is not in the plane  $\mathcal{V}$ , if we choose a basis  $\{\mathbf{F}, \mathbf{E}\}$  for  $\mathcal{V}$  then  $\{\mathbf{F}, \mathbf{E}, \mathbf{D}\}$  is a basis for  $\mathbb{R}^3$ . If  $\mathbf{A}$  is any vector in  $\mathbb{R}^3$  then

$$\mathbf{A} = a_1 \mathbf{F} + a_2 \mathbf{E} + a_3 \mathbf{D}$$

so

$$\begin{aligned} \mathbf{S}(\mathbf{A}) &= a_1 \mathbf{S}(\mathbf{F}) + a_2 \mathbf{S}(\mathbf{E}) + a_3 \mathbf{S}(\mathbf{D}) \\ &= a_1 \mathbf{F} + a_2 \mathbf{E} \end{aligned}$$

since  $\mathbf{E}$  and  $\mathbf{F}$  belong to  $\mathcal{V}$  and  $\mathbf{D}$  belongs to  $\mathcal{L}(\mathbf{D})$ . Therefore

$$\begin{aligned} \mathbf{S}^2(\mathbf{A}) &= \mathbf{S}(a_1 \mathbf{F} + a_2 \mathbf{E}) = a_1 \mathbf{S}(\mathbf{F}) + a_2 \mathbf{S}(\mathbf{E}) \\ &= a_1 \mathbf{F} + a_2 \mathbf{E}. \end{aligned}$$

Thus  $\mathbf{S}^2 = \mathbf{S}$  so  $\mathbf{S}$  is a projection.

It is a fact we are going to prove that every projection

$$\mathbf{S}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

such that  $\dim(\text{Im } \mathbf{S}) = 2$  is a skew projection of the preceding type. The more usual projections we have dealt with up until now are what are called *orthogonal* projections ( $\mathbf{D}$  is orthogonal to  $\mathcal{V}$ ) or *self-adjoint* projections (see, for example Chapter 15).

Let us return now to the question that started all this, namely how does one recognize a projection? The definition we have proposed is going to be useful if and only if we can somehow show that every projection  $\mathbf{S}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a skew projection in the preceding sense. In fact more is true.

**Theorem 12<sup>bis</sup>.1.** *Let  $\mathcal{W}$  be a finite-dimensional vector space of dimension  $n$  and  $\mathbf{S}: \mathcal{W} \rightarrow \mathcal{W}$  a projection. Then there is a basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  for  $\mathcal{W}$  such that*

$$\mathbf{S}(\mathbf{A}_i) = \begin{cases} \mathbf{A}_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } r+1 \leq i \leq n \end{cases}$$

where  $r = \dim(\text{Im } \mathbf{S})$ , and hence the matrix of  $\mathbf{S}$  relative to the basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is

$$\begin{pmatrix} \overbrace{1 \dots 1}^r & & 0 \\ & \underbrace{0 \dots 0}_{n-r} & \\ 0 & & \end{pmatrix}$$

**PROOF.** Let  $\{\mathbf{B}_1, \dots, \mathbf{B}_r\}$  be a basis for  $\text{Im } \mathbf{S}$  and  $\{\mathbf{C}_1, \dots, \mathbf{C}_s\}$  be a basis for  $\ker \mathbf{S}$ . Notice that by (8.10) we have

$$n = r + s.$$

Therefore the collection  $\{\mathbf{B}_1, \dots, \mathbf{B}_r, \mathbf{C}_1, \dots, \mathbf{C}_s\}$  contains  $n$  vectors, the correct number of vectors to be a basis for  $\mathscr{W}$ . Let us show that indeed it is a basis for  $\mathscr{W}$ . To this end notice the following two facts:

$$\begin{aligned} \mathbf{S}(\mathbf{C}_i) &= \mathbf{0} & \text{for } i = 1, \dots, s \\ \mathbf{S}(\mathbf{B}_i) &= \mathbf{B}_i & \text{for } i = 1, \dots, r. \end{aligned}$$

The first of these is clear from the definition of  $\{\mathbf{C}_1, \dots, \mathbf{C}_s\}$  while the second requires some proof. (Be sensible, we still have not used the fact that  $\mathbf{S}$  is a projection!) We note that since  $\mathbf{B}_i$  belongs to  $\text{Im } \mathbf{S}$  we may find a vector  $\mathbf{D}_i$  in  $\mathscr{W}$  such that  $\mathbf{S}(\mathbf{D}_i) = \mathbf{B}_i$ . Therefore since  $\mathbf{S}$  is a projection

$$\mathbf{S}(\mathbf{B}_i) = \mathbf{S}(\mathbf{S}(\mathbf{D}_i)) = \mathbf{S}^2(\mathbf{D}_i) = \mathbf{S}(\mathbf{D}_i) = \mathbf{B}_i,$$

which verifies the second set of equations preceding.

Now since there are  $n$  vectors in the collection  $\{\mathbf{B}_1, \dots, \mathbf{B}_r, \mathbf{C}_1, \dots, \mathbf{C}_s\}$  we may apply (6.9) to conclude that we need only show  $\{\mathbf{B}_1, \dots, \mathbf{C}_s\}$  is linearly independent for them to be a basis, that is, we must verify only one-half of the basis condition.

Suppose the vectors  $\{\mathbf{B}_1, \dots, \mathbf{B}_r, \mathbf{C}_1, \dots, \mathbf{C}_s\}$  are linearly dependent. Then we may find numbers  $b_1, \dots, b_r, c_1, \dots, c_s$ , not all zero, so that

$$(*) \quad b_1 \mathbf{B}_1 + \dots + b_r \mathbf{B}_r + c_1 \mathbf{C}_1 + \dots + c_s \mathbf{C}_s = \mathbf{0}.$$

Apply  $\mathbf{S}$  to both sides of this equation to get

$$\begin{aligned} \mathbf{0} &= \mathbf{S}(b_1 \mathbf{B}_1 + \dots + b_r \mathbf{B}_r + c_1 \mathbf{C}_1 + \dots + c_s \mathbf{C}_s) \\ &= b_1 \mathbf{S}(\mathbf{B}_1) + \dots + b_r \mathbf{S}(\mathbf{B}_r) + c_1 \mathbf{S}(\mathbf{C}_1) + \dots + c_s \mathbf{S}(\mathbf{C}_s) \\ &= b_1 \mathbf{B}_1 + \dots + b_r \mathbf{B}_r. \end{aligned}$$

But since  $\mathbf{B}_1, \dots, \mathbf{B}_r$  are a basis for  $\text{Im } \mathbf{S}$  they are linearly independent and hence  $b_1 = b_2 = \dots = b_r = 0$ . Thus our original relation  $(*)$  reads

$$c_1 \mathbf{C}_1 + \dots + c_s \mathbf{C}_s = \mathbf{0}.$$

But since  $\mathbf{C}_1, \dots, \mathbf{C}_s$  are a basis for  $\ker \mathbf{S}$  they are linearly independent also, so  $c_1 = c_2 = \dots = c_s = 0$ . Therefore the collection  $\{\mathbf{B}_1, \dots, \mathbf{B}_r, \mathbf{C}_1, \dots, \mathbf{C}_s\}$  is not linearly dependent, so must be linearly independent. Thus as we remarked earlier  $\{\mathbf{B}_1, \dots, \mathbf{B}_r, \mathbf{C}_1, \dots, \mathbf{C}_s\}$  being linearly independent is a basis for  $\mathscr{W}$ . Set

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{B}_1 \\ \mathbf{A}_2 &= \mathbf{B}_2 \\ &\vdots \\ \mathbf{A}_r &= \mathbf{B}_r & (\text{remember } r + s = n.) \\ \mathbf{A}_{r+1} &= \mathbf{C}_1 \\ &\vdots \\ \mathbf{A}_n &= \mathbf{C}_s. \end{aligned}$$

Then we will have

$$\begin{aligned} P(A_i) &= P(B_i) = B_i = A_i & \text{for } 1 \leq i \leq r \\ &= P(C_{i-r}) = 0 & \text{for } r+1 \leq i \leq n \end{aligned}$$

as required.  $\square$

It is now a rather easy consequence that every linear transformation  $S: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  such that

$$S^2 = S \quad \text{and} \quad \dim(\text{Im } S) = 2$$

is a skew projection as described preceding (12<sup>bis</sup>.1).

The reader should turn ahead to (14.1) and examine the fundamental difference between it and (12<sup>bis</sup>.1).

For a given projection  $S: \mathcal{W} \rightarrow \mathcal{W}$  the actual computation of a (not *the*!) basis as in (12<sup>bis</sup>.1) may be a numerical horror, but still the proof gives you a method for doing so in terms of the type of problem we have successfully handled before.

EXAMPLE 2. *Nilpotent transformations.*

A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is said to be *nilpotent of index  $k$*  iff  $T^k = 0$  and  $T^{k-1} \neq 0$ , that is,  $T^k(A) = 0$  for all vectors  $A$  in  $\mathcal{V}$  but there is at least one vector  $B$  in  $\mathcal{V}$  for which  $T^{k-1}(B) \neq 0$ . There is a sort of canonical example of a nilpotent transformation namely the differentiation operator

$$D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$$

is nilpotent of index  $n+1$ . Another family of simple examples is provided by the *shift operator*

$$S: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

defined by

$$S(a_1, a_2, \dots, a_n) = (0, a_1, a_2, \dots, a_{n-1}).$$

Thus if  $\{E_1, \dots, E_n\}$  denotes the usual basis for  $\mathbb{R}^n$  then

$$\begin{aligned} S(E_1) &= E_2 \\ S(E_2) &= E_3 \\ &\vdots \\ S(E_n) &= 0 \end{aligned}$$

so that the matrix of  $S$  relative to the usual basis is

$$\begin{pmatrix} 0 & 0 & & & 0 \\ 1 & 0 & & & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

In this case  $\mathbf{S}$  is nilpotent of index  $n$ . Notice in both cases the index of nilpotence is the dimension of the vector space on which the transformation is defined. This is a very special situation. In fact we can determine all such nilpotent transformations. We will need the following very important fact.

**Proposition 12<sup>bis</sup>.2.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation that is nilpotent of index  $k$ . Let  $\mathbf{B} \in \mathcal{V}$  be a vector such that  $T^{k-1}(\mathbf{B}) \neq \mathbf{0}$ . Then the vectors  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$  are linearly independent.*

PROOF. Let us suppose that

$$(*) \quad b_0 \mathbf{B} + b_1 T\mathbf{B} + \dots + b_{k-1} T^{k-1} \mathbf{B} = \mathbf{0}$$

is a linear relation among  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$ . What we must show is that  $b_0 = b_1 = \dots = b_{k-1} = 0$ . Now notice that since  $T^k(\mathbf{B}) = \mathbf{0}$  we must have

$$\mathbf{0} = T^k(\mathbf{B}) = T^{k+1}(\mathbf{B}) = \dots$$

by (8.1). Apply  $T^{k-1}$  to both sides of (\*) to get

$$\begin{aligned} \mathbf{0} &= b_0 T^{k-1}(\mathbf{B}) + b_1 T^k(\mathbf{B}) + \dots + b_{k-1} T^{2k-2}(\mathbf{B}) \\ &= b_0 T^{k-1}(\mathbf{B}) + b_1 \mathbf{0} + \dots + b_{k-1} \mathbf{0} \\ &= b_0 T^{k-1}(\mathbf{B}). \end{aligned}$$

Therefore since  $T^{k-1}(\mathbf{B}) \neq \mathbf{0}$  we must have  $b_0 = 0$ , and our linear relation (\*) has become

$$(**) \quad b_1 T(\mathbf{B}) + \dots + b_{k-1} T^{k-1}(\mathbf{B}) = \mathbf{0}.$$

Now apply  $T^{k-2}$  to both sides getting

$$\begin{aligned} \mathbf{0} &= b_1 T^{k-1}(\mathbf{B}) + b_2 T^k(\mathbf{B}) + \dots + b_{k-1} T^{2k-3}(\mathbf{B}) \\ &= b_1 T^{k-1}(\mathbf{B}) + b_2 \mathbf{0} + \dots + b_{k-1} \mathbf{0} \\ &= b_1 T^{k-1}(\mathbf{B}) \end{aligned}$$

and so  $b_1 = 0$ . Continuing in this way we may show that  $0 = b_2 = \dots = b_{k-1}$ . Therefore  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$  is linearly independent.  $\square$

**Corollary 12<sup>bis</sup>.3.** *Suppose that  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a nilpotent linear transformation and the index of  $T$  equals the dimension of  $\mathcal{V}$ . Call this common number  $k$ . Then there is a vector  $\mathbf{B} \in \mathcal{V}$  such that  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$  is a basis for  $\mathcal{V}$  and  $T^k(\mathbf{B}) = \mathbf{0}$ .*

PROOF. According to (12<sup>bis</sup>.2) the vectors  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$  are linearly independent. There are  $k$  of them and since  $k = \dim \mathcal{V}$  it follows from (6.9) that they are basis for  $\mathcal{V}$ . Since  $k$  is also the index of  $T$ ,  $T^k(\mathbf{B}) = \mathbf{0}$ .  $\square$

Therefore if  $T: \mathcal{V} \rightarrow \mathcal{V}$  is nilpotent **and** index  $T = k = \dim \mathcal{V}$  then



we may choose a vector  $\mathbf{B}$  in  $\mathcal{V}$  so that  $\{\mathbf{B}, T(\mathbf{B}), \dots, T^{k-1}(\mathbf{B})\}$  is a basis for  $\mathcal{V}$  and  $T^k \mathbf{B} = \mathbf{0}$ . The matrix of  $T$  relative to this basis is therefore

$$\begin{pmatrix} 0 & 0 & & 0 & 0 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

and if we use this basis to identify  $\mathcal{V}$  with  $\mathbb{R}^n$  as in (8.13–8.15) we find that  $T$  is identified with the shift operator. Thus the structure of nilpotent transformations  $T: \mathcal{V} \rightarrow \mathcal{V}$  of index equal to dimension  $\mathcal{V}$  is completely determined. The structure of more general nilpotent transformations is more complicated and a subject for a more advanced course in linear algebra.

### EXAMPLE 3. Cyclic transformations:

A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is called *cyclic* iff there exists a vector  $\mathbf{A}$  in  $\mathcal{V}$  such that the collection  $\{\mathbf{A}, T(\mathbf{A}), T^2(\mathbf{A}), \dots\}$  spans  $\mathcal{V}$ . The vector  $\mathbf{A}$  is called a *cyclic vector* for  $T$ . Examples of cyclic transformations are plentiful. It follows from (12<sup>bis</sup>.3) that any nilpotent transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  whose index is equal to the dimension of  $\mathcal{V}$  is cyclic. In dimension 2 essentially all nontrivial linear transformations are cyclic. We begin with a preliminary result of independent interest.

**Proposition 12<sup>bis</sup>.4.** Suppose that  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a cyclic linear transformation.

Let the dimension of  $\mathcal{V}$  be  $n$  and let  $\mathbf{A} \in \mathcal{V}$  be a cyclic vector for  $T$ . Then the vectors  $\mathbf{A}, T(\mathbf{A}), \dots, T^{n-1}(\mathbf{A})$  are a basis for  $\mathcal{V}$ .

**PROOF.** There are  $n$  vectors in the set  $\{\mathbf{A}, T(\mathbf{A}), \dots, T^{n-1}(\mathbf{A})\}$  so by (6.9) to show they are a basis in the  $n$ -dimensional space we need only show that they are linearly independent. Suppose to the contrary that they are linearly dependent. By (6.2) there will be a positive integer  $m < n$  such that

$$T^m(\mathbf{A}) \in \mathcal{L}(\mathbf{A}, T(\mathbf{A}), \dots, T^{m-1}(\mathbf{A})).$$

Therefore by applying  $T$

$$T^{m+1}(\mathbf{A}) \in \mathcal{L}(T(\mathbf{A}), T^2(\mathbf{A}), \dots, T^m(\mathbf{A})) \subset \mathcal{L}(\mathbf{A}, T(\mathbf{A}), \dots, T^{m-1}(\mathbf{A}))$$

since all of the vectors  $T(\mathbf{A}), T^2(\mathbf{A}), \dots, T^m(\mathbf{A})$  belong to  $\mathcal{L}(\mathbf{A}, T(\mathbf{A}), \dots, T^{m-1}(\mathbf{A}))$ . Applying  $T$  again we find

$$T^{m+2}(\mathbf{A}) \in \mathcal{L}(T(\mathbf{A}), T^2(\mathbf{A}), \dots, T^m(\mathbf{A})) \subset \mathcal{L}(\mathbf{A}, T(\mathbf{A}), \dots, T^{m-1}(\mathbf{A}))$$

and continuing in this way we show that

$$T^m(\mathbf{A}), T^{m+1}(\mathbf{A}), T^{m+2}(\mathbf{A}), \dots, \in \mathcal{L}(\mathbf{A}, T(\mathbf{A}), \dots, T^{m-1}(\mathbf{A}))$$

and therefore that

$$\mathcal{V} = \mathcal{L}(A, T(A), \dots, T^m(A), T^{m+1}(A), \dots) \subset \mathcal{L}(A, T(A), \dots, T^{m-1}(A))$$

so that  $\dim \mathcal{V} \leq m < n$  which is impossible. Therefore the vectors  $A, T(A), \dots, T^{n-1}(A)$  must be linearly independent.  $\square$

**Proposition 12<sup>bis</sup>.5.** *A linear transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is cyclic iff  $T \neq eI$ , where  $e$  is a number and  $I: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  the identity transformation.*

**PROOF.** Suppose that  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . A vector  $A \in \mathbb{R}^2$  is a cyclic vector for  $T$  iff  $T(A) \notin \mathcal{L}(A)$ , that is, iff  $T(A)$  does not belong to the line through the origin spanned by  $A$ . For then  $\{A, T(A)\}$  will be a basis for  $\mathbb{R}^2$  so that  $\mathbb{R}^2 = \mathcal{L}(A, T(A), \dots)$  as required. So suppose that  $T$  does not have a cyclic vector. Then  $T(A) \in \mathcal{L}(A)$  for every  $A \in \mathbb{R}^2$ . In particular

$$T(1, 0) \in \mathcal{L}(1, 0) \Rightarrow T(1, 0) = e_1(1, 0)$$

$$T(0, 1) \in \mathcal{L}(0, 1) \Rightarrow T(0, 1) = e_2(0, 1).$$

Now we claim  $e_1 = e_2$  for we also have

$$T(1, 1) \in \mathcal{L}(1, 1) \Rightarrow T(1, 1) = e(1, 1)$$

so

$$\begin{aligned} e(1, 1) &= T(1, 1) = T((1, 0) + (0, 1)) \\ &= T(1, 0) + T(0, 1) \\ &= e_1(1, 0) + e_2(0, 1) \end{aligned}$$

so

$$(e, e) = (e_1, e_2)$$

and hence  $e_1 = e = e_2$ . Therefore

$$T(1, 0) = (e, 0)$$

$$T(0, 1) = (0, e)$$

so the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^2$  is

$$\begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix} = eI$$

and hence  $T = eI$ .  $\square$

Suppose that  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a cyclic transformation with cyclic vector  $A$ . Then the matrix of  $T$  relative to the basis  $\{A, T(A), \dots, T^{n-1}(A)\}$  is

$$\begin{pmatrix} 0 & & & 0 & a_0 \\ 1 & & & & a_1 \\ \vdots & \ddots & & & \vdots \\ 0 & \cdots & 0 & 1 & a_{n-1} \end{pmatrix}$$

where  $T^n(\mathbf{A}) = a_0 \mathbf{A} + a_1 T(\mathbf{A}) + \cdots + a_{n-1} T^{n-1}(\mathbf{A})$ . This matrix is particularly simple and is one reason for looking for a cyclic vector. In general however a linear transformation need not have any cyclic vectors at all! Again this is a subject for study in a more advanced course on linear algebra.

## EXERCISES

- Find the matrix of the projection  $P: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , relative to the standard basis, where  $P$  is the projection onto the plane  $\mathcal{V} = \{(x, y, z) | x + y = 0\}$ .
- Repeat Exercise 1 for  $\mathcal{V} = \{(x, y, z) | z = 0\}$ .
- Find the basis  $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$  for  $\mathbb{R}^3$  such that the projection  $P$  of Exercise 1 relative to the basis is of the form

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

- Repeat Exercise 3 for the projection of Exercise 2.
- Show that the differentiation operator  $D: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R})$  is nilpotent of index 4. Find a basis of  $\mathcal{P}_3(\mathbb{R})$  of the form

$$\mathcal{B} = \{\mathbf{B}, D(\mathbf{B}), D^2(\mathbf{B}), D^3(\mathbf{B})\}$$

where  $\mathbf{B} \in \mathcal{P}_3(\mathbb{R})$ . Calculate the matrix of  $D$  relative to the basis  $\mathcal{B}$ .

- Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation

$$T(x, y, z) = (0, x, 2y).$$

Show that  $T$  is nilpotent of index 3. Find a vector  $\mathbf{B}$  so that

$$\{\mathbf{B}, T(\mathbf{B}), T^2(\mathbf{B})\}$$

is a basis for  $\mathbb{R}^3$ . Calculate the matrix of  $T$  relative to this basis.

Calculate the matrix of  $T$  relative to the standard basis.

- Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation

$$T(x, y, z) = (x + y, y + z, x).$$

Show that  $T$  is a cyclic linear transformation. (*Hint*: find  $\mathbf{B}$  such that  $\{\mathbf{B}, T(\mathbf{B}), T^2(\mathbf{B})\}$  is a set of independent vectors.)

- Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation defined by

$$T(x, y, z) = (y, x, z).$$

Show that  $T$  is not cyclic.

9. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation defined by

$$T(x, y, z) = (y, z, x).$$

Show that  $T$  is cyclic. Find the basis  $\{B, T(B), T^2(B)\}$  and calculate the matrix of  $T$  relative to  $\{B, T(B), T^2(B)\}$ .

10. Is the product of two projections  $P$  and  $Q$  in  $\mathcal{V}$  again a projection? What if  $P$  and  $Q$  commute with each other?
11. Is the product of two nilpotent transformations again nilpotent?
12. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is called an involution iff  $T^2 = I$ . Which of the following matrices represent involutions in  $\mathbb{R}^3$  with respect to the standard bases:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

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

13. Is the product of two involutions again an involution?
14. Two transformations  $S, T: \mathcal{V} \rightarrow \mathcal{V}$  are said to be mutually annihilating if  $ST = 0$  and  $TS = 0$ . Show that the sum of two mutually annihilating projections is again a projection.
15. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is the  $n$ -dimensional vector space,  $\mathcal{V}$  is called a *reflection* if there is an  $(n-1)$ -dimensional subspace  $\mathcal{H}$  in  $\mathcal{V}$ , the reflecting hyperplane, and a non-zero vector  $v$  in  $\mathcal{V}$  such that  $v$  does not belong to  $\mathcal{H}$  and:

$$T(h) = h \quad \text{for all } h \in \mathcal{H},$$

$$T(v) = -v.$$

Show that with respect to a suitable basis the matrix of a reflection takes the form

$$\begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & & & \ddots & \\ \vdots & 0 & & & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

16. Let  $S: \mathcal{P}_n \rightarrow \mathcal{P}_n$  be the linear transformation defined by the equation

$$S(p(x)) = p(x+1)$$

Find the matrix of  $S$  with respect to the standard basis  $\{1, x, \dots, x^n\}$  of  $\mathcal{P}_n$ .

17. Let  $A$  be the  $2 \times 2$  matrix

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

and let  $L_A: \mathcal{M}_2 \rightarrow \mathcal{M}$  be the linear transformation given by

$$L_A(M) = A \cdot M$$

for  $M$  in  $\mathcal{M}$ . Find the matrix of  $L_A$  with respect to the ordered basis

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

for  $\mathcal{M}_i$ .

18. Let  $T: \mathcal{M}_i \rightarrow \mathcal{M}_i$  be the linear transformation given by

$$T(M) = M'.$$

Find the matrix of  $T$  with respect to the same ordered basis of  $\mathcal{M}_i$  as in Exercise 17.

19. If  $P$  is a projection show that  $T = 2P - I$  is an involution.  
 20. If  $T$  is an involution show that  $P = \frac{1}{2}(T + I)$  is a projection.  
 21. Use Exercises 19 and 20 to show that with respect to a suitable basis the matrix of an involution  $T: \mathcal{V} \rightarrow \mathcal{V}$  is of the form

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

22. Find all possible  $2 \times 2$  matrices that represent a nilpotent transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  of index 2 with respect to some basis of  $\mathbb{R}^2$ .  
 23. (a) If  $A$  is a matrix show  $A + A^t$  is symmetric and  $A - A^t$  is skew-symmetric.  
 (b) Show every matrix is the sum of a symmetric and a skew-symmetric matrix.

In the historical development of linear algebra the geometry of linear transformations and the algebra of systems of linear equations played significant and important rolls. A system of linear equations has the form

$$\begin{aligned}
 (L) \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
 & \vdots \\
 & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_n.
 \end{aligned}$$

Here  $x_1, x_2, \dots, x_n$ , denote the unknowns, which are to be determined. The  $mn$  numbers  $a_{ij}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  are called the *coefficients* of the linear system (L) and are, of course, fixed. A *solution* of the system (L) is an ordered  $n$ -tuple of numbers  $(s_1, \dots, s_n)$  such that the  $m$  equations

$$\begin{aligned}
 & a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n = b_1 \\
 & a_{21}s_1 + a_{22}s_2 + \cdots + a_{2n}s_n = b_2 \\
 & \vdots \\
 & a_{m1}s_1 + a_{m2}s_2 + \cdots + a_{mn}s_n = b_n
 \end{aligned}$$

are all true. To solve the system (L) means to find all solutions of (L).

**EXAMPLE 1.** Solve the linear system

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

**Solution.** From previous experience we recall that this system may be solved as follows:

$$\left. \begin{array}{l} (1) \quad x_1 + x_2 + x_3 = 3 \\ (2) \quad 2x_2 + x_3 = 4 \\ (3) \quad x_1 - x_2 = -1 \end{array} \right\}$$

subtract (1) from (3) to get

$$\left. \begin{array}{l} x_1 + x_2 + x_3 = 3 \\ 2x_2 + x_3 = 4 \\ -2x_2 - x_3 = -4 \end{array} \right\}.$$

Erase the third equation, it follows from the second, to get

$$\begin{array}{lcl} x_1 + x_2 + x_3 = 3 & & x_1 = 3 - x_2 - x_3 \\ 2x_2 + x_3 = 4 & \text{or} & x_3 = 4 - 2x_2, \end{array}$$

or

$$\begin{array}{l} x_1 = x_2 - 1 \\ x_3 = 4 - 2x_2. \end{array}$$

Therefore the solutions are all triples of numbers

$$(s - 1, s, 4 - 2s)$$

where  $s$  is arbitrary. For example

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

are solutions but

$$(1, 1, 2)$$

is not.

On the other hand past experience should have taught us that not every system of linear equations has a solution. For example the linear system

$$\begin{array}{l} x_1 + x_2 = 1 \\ x_1 + x_2 = -1 \end{array}$$

clearly can have no solutions. Thus our first order of business in the study of linear equations should be to determine when a linear system has solutions, and only afterwards take up the discussion of actual techniques of solution.

Our study of matrices in the preceding chapters will come in handy here. In fact it is through the study of linear systems that matrices most frequently appear in modern scientific investigations. First let us introduce the matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

The matrix  $\mathbf{A}$  is called the *coefficient matrix* of the linear system and  $\mathbf{B}$  the matrix of *constants* of the linear system. We may then write the *system* of linear equations (L) in a more compact form as the single *matrix* equation

$$(M) \quad \mathbf{A}\mathbf{X} = \mathbf{B}.$$

Solving this matrix equation may be given a very simple interpretation in terms of linear transformations. We let

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

be the linear transformation whose matrix relative to the standard bases is  $\mathbf{A}$ . Thus

$$T(x_1, \dots, x_n) = \left( \sum_{j=1}^n a_{1j}x_j, \sum_{j=1}^n a_{2j}x_j, \dots, \sum_{j=1}^n a_{mj}x_j \right).$$

Let  $\mathbf{B} = (b_1, \dots, b_m)$ , which is a vector in  $\mathbb{R}^m$ . A *solution* of (M) is then a vector  $(s_1, s_2, \dots, s_n)$  in  $\mathbb{R}^n$  such that

$$T(s_1, s_2, \dots, s_n) = (b_1, b_2, \dots, b_m),$$

and to find all solutions of (M) or equivalently (L), means to find all vectors  $(s_1, \dots, s_n)$  in  $\mathbb{R}^n$  such that

$$T(s_1, s_2, \dots, s_n) = (b_1, \dots, b_m).$$

The following result is therefore clear.

**Proposition 13.1.** *With the preceding notations, the linear system (L) has a solution iff the vector  $(b_1, b_2, \dots, b_m)$  lies in the image of the transformation  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , that is, iff  $\mathbf{B} \in \text{Im } T$ .*

Now (13.1) is all well and good from a theoretical point of view, but how does one tell if  $\mathbf{B} \in \text{Im } T$  from the coefficient matrix  $\mathbf{A}$ ? The answer is really quite simple. Let us introduce the vectors

$$\begin{aligned} \mathbf{A}_{(1)} &= (a_{11}, a_{21}, \dots, a_{m1}) \\ &\vdots \\ \mathbf{A}_{(n)} &= (a_{1n}, a_{2n}, \dots, a_{mn}) \end{aligned}$$

in  $\mathbb{R}^m$ . Notice that

$$\begin{aligned} \mathbf{A}_{(1)} &= T(\mathbf{E}_1) \\ &\vdots \\ \mathbf{A}_{(n)} &= T(\mathbf{E}_n). \end{aligned}$$

Therefore by (8.4)

$$\text{Im } T = \mathcal{L}(T(\mathbf{E}_1), \dots, T(\mathbf{E}_n)) = \mathcal{L}(\mathbf{A}_{(1)}, \dots, \mathbf{A}_{(n)}),$$

that is to say, the image of  $T$  is the linear span of the vectors  $\mathbf{A}_{(1)}, \dots, \mathbf{A}_{(n)}$ . This may all be summed up in a definition and a theorem.



### 13: Systems of linear equations

**Definition.** If  $A = (a_{ij})$  is an  $m \times n$  matrix, the *column vectors* of  $A$  are the  $n$  vectors

$$\begin{aligned} A_{(1)} &= (a_{11}, a_{21}, \dots, a_{m1}) \\ A_{(2)} &= (a_{12}, a_{22}, \dots, a_{m2}) \\ &\vdots \\ A_{(n)} &= (a_{1n}, a_{2n}, \dots, a_{mn}) \end{aligned}$$

in  $\mathbb{R}^m$ . The *column space* of  $A$  is the linear span of its column vectors in  $\mathbb{R}^m$ .

**EXAMPLE 2.** What are the column vectors and what is the column space of the matrix

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

**Solution.** The column vectors are the two vectors

$$(1, -1, 0), \quad (0, -2, 2)$$

in  $\mathbb{R}^3$ . The column space of  $A$  is the linear span in  $\mathbb{R}^3$  of these two vectors. A moment's computation shows this to be the plane

$$x + y + z = 0$$

in  $\mathbb{R}^3$ .

From our preceding discussion we now obtain:

**Theorem 13.2.** Let

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

be a system of linear equations. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Then the system has a solution iff the vector  $B = (b_1, \dots, b_m)$  in  $\mathbb{R}^m$  is in the column space of  $A$ , that is, the vector  $(b_1, \dots, b_m)$  can be written as a linear combination of the column vectors of  $A$ .

**PROOF.** Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be the linear transformation whose matrix is  $A$ . Then the image of  $T$  is the column space of  $A$ . Now apply (13.1).  $\square$

**EXAMPLE 3.** Does the linear system

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

have any solutions?

*Solution.* The coefficient matrix of this linear system is

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

and its matrix of constants is

$$\mathbf{B} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Thus the column vectors of  $\mathbf{A}$  are

$$(1, 1, 2), \quad (1, 0, 1), \quad (1, 1, 2).$$

Thus the column space of  $\mathbf{A}$  is spanned by

$$(1, 1, 2) \quad \text{and} \quad (1, 0, 1)$$

while the vector  $\mathbf{B}$  is

$$(1, 1, 0).$$

Now suppose

$$(1, 1, 0) = a(1, 1, 2) + b(1, 0, 1).$$

Then

$$\begin{aligned}1 &= a + b \\1 &= a \\0 &= 2a + b\end{aligned}$$

which implies

$$\begin{aligned}b &= 0 \\a &= 1 \\b &= -2\end{aligned}$$

which is impossible (compare the first and third equations). Therefore  $\mathbf{B}$  does not belong to the column space of  $\mathbf{A}$  and the system has no solutions at all.

The simplest type of a linear system is one whose matrix of constants is the zero matrix. That is, where

$$\mathbf{B} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}.$$

This case merits a special name.

**Definition.** A system of linear equations (in matrix notation)

$$\mathbf{A}\mathbf{X} = \mathbf{B}$$

is called *homogeneous* iff

$$\mathbf{B} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}.$$

**Theorem 13.3.** Let (in matrix notation)

$$\mathbf{A}\mathbf{X} = \mathbf{0}$$

be a homogeneous system of linear equations. Then the set  $\mathcal{S} = \{(s_1, \dots, s_n)\}$  of all solutions to this linear system is a linear subspace of  $\mathbb{R}^n$ . In fact, if

$$\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

is the linear transformation whose matrix relative to the standard bases is  $\mathbf{A}$ , then  $\mathcal{S} = \ker \mathbf{T}$ .

**PROOF.** This theorem takes longer to state than to prove. Simply note that  $(s_1, \dots, s_n)$  is a solution of  $\mathbf{A}\mathbf{X} = \mathbf{0}$  iff  $\mathbf{T}(s_1, \dots, s_n) = (0, \dots, 0)$ .  $\square$

*Remark.* You might go back and look at Example 6 of Chapter 4 again.

Thus in order to solve a homogeneous system

$$\mathbf{A}\mathbf{X} = \mathbf{0}$$

we have merely to find a basis for the kernel of the linear transformation  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  whose matrix relative to the standard basis is  $\mathbf{A}$ . For by (13.3) any solution will be a linear combination of these basis vectors.

Before we turn to the case of a general linear system we pause to collect some preliminary facts.

**Definition.** Let  $\mathcal{V}$  be a vector space, let  $\mathcal{U}$  be a linear subspace of  $\mathcal{V}$ , and let  $\mathbf{A} \in \mathcal{V}$ . Denote by  $\mathcal{A} = \mathbf{A} + \mathcal{U}$  the set of all vectors of the form  $\mathbf{A} + \mathbf{X}$  where  $\mathbf{X} \in \mathcal{U}$ . The set  $\mathbf{A} + \mathcal{U}$  is called a *parallel* of  $\mathcal{U}$  in  $\mathcal{V}$ . It is said to result from *parallel translation* of  $\mathcal{U}$  by the vector  $\mathbf{A}$ . A parallel of some linear subspace of  $\mathcal{V}$  is called an *affine subspace* of  $\mathcal{V}$ .

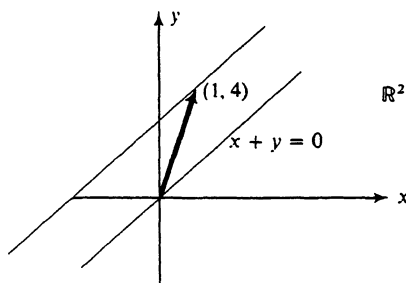


Figure 13.1

**EXAMPLE 4.** We let  $\mathcal{V} = \mathbb{R}^2$  and  $\mathcal{U}$  be the linear subspace  $x + y = 0$ . If  $\mathbf{A} = (1, 4)$  then  $\mathbf{A} + \mathcal{U}$  is the line through the point  $(1, 4)$  that is parallel to  $x + y = 0$ . (See Figure 13.1).

**Proposition 13.4.** *If  $\mathcal{U}$  is a linear subspace of a vector space  $\mathcal{V}$  then*

- (1)  $\mathbf{A} \in \mathbf{A} + \mathcal{U}$  for any  $\mathbf{A} \in \mathcal{V}$ .
- (2) If  $\mathbf{B} \in \mathbf{A} + \mathcal{U}$  then  $\mathbf{B} + \mathcal{U} = \mathbf{A} + \mathcal{U}$ .
- (3) Two parallels of  $\mathcal{U}$  either coincide or have no vector in common.
- (4) If  $\mathbf{B}, \mathbf{C} \in \mathbf{A} + \mathcal{U}$  then  $\mathbf{B} - \mathbf{C} \in \mathcal{U}$ .

**PROOF.** Since  $\mathbf{A} = \mathbf{A} + \mathbf{0}$ , and  $\mathbf{0} \in \mathcal{U}$  since  $\mathcal{U}$  is a subspace, it follows that  $\mathbf{A} \in \mathbf{A} + \mathcal{U}$  which proves (1). To prove (2) suppose that  $\mathbf{B} \in \mathbf{A} + \mathcal{U}$ . Then

$$\mathbf{B} = \mathbf{A} + \mathbf{X}$$

for some vector  $\mathbf{X}$  in  $\mathcal{U}$ . Thus if  $\mathbf{C} \in \mathbf{B} + \mathcal{U}$  we may find  $\mathbf{Y}$  in  $\mathcal{U}$  such that

$$\mathbf{C} = \mathbf{B} + \mathbf{Y} = (\mathbf{A} + \mathbf{X}) + \mathbf{Y} = \mathbf{A} + (\mathbf{X} + \mathbf{Y}) = \mathbf{A} + \mathbf{W}.$$

Now note that  $\mathbf{W} = \mathbf{X} + \mathbf{Y} \in \mathcal{U}$  because  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$ . Therefore  $\mathbf{C} \in \mathbf{A} + \mathcal{U}$ , and hence  $\mathbf{B} + \mathcal{U} \subset \mathbf{A} + \mathcal{U}$ . If we now write the equation  $\mathbf{B} = \mathbf{A} + \mathbf{X}$  in the form

$$\mathbf{A} = \mathbf{B} + (-\mathbf{X})$$

we may apply exactly the same argument to show that  $\mathbf{A} + \mathcal{U} \subset \mathbf{B} + \mathcal{U}$  and hence we may conclude that  $\mathbf{A} + \mathcal{U} = \mathbf{B} + \mathcal{U}$  as desired. To prove (3) we suppose that  $\mathbf{A} + \mathcal{U}$  and  $\mathbf{B} + \mathcal{U}$  are two parallels of  $\mathcal{U}$  in  $\mathcal{V}$ . Suppose that they have a vector  $\mathbf{C}$  in common, that is,  $\mathbf{C} \in \mathbf{A} + \mathcal{U}$  and  $\mathbf{C} \in \mathbf{B} + \mathcal{U}$ . Then by (2) we find

$$\mathbf{C} + \mathcal{U} = \mathbf{A} + \mathcal{U}$$

$$\mathbf{C} + \mathcal{U} = \mathbf{B} + \mathcal{U}$$

so that  $\mathbf{A} + \mathcal{U} = \mathbf{B} + \mathcal{U}$  as claimed. Finally to prove Part (4) we suppose that  $\mathbf{B}, \mathbf{C} \in \mathbf{A} + \mathcal{U}$ . Then we may find vectors  $\mathbf{X}$  and  $\mathbf{Y}$  in  $\mathcal{U}$  such that

$$\mathbf{B} = \mathbf{A} + \mathbf{X}$$

$$\mathbf{C} = \mathbf{A} + \mathbf{Y}$$

from which we obtain

$$\mathbf{B} - \mathbf{C} = \mathbf{X} - \mathbf{Y}.$$

But  $\mathbf{X} - \mathbf{Y}$  belongs to  $\mathcal{U}$  since  $\mathcal{U}$  is a linear subspace of  $\mathcal{V}$ .  $\square$

It follows from (13.4) that an affine subspace  $\mathcal{A}$  of  $\mathcal{V}$  is completely determined by a basis for  $\mathcal{U}$  and a *single* vector  $\mathbf{A}$  in the affine subspace  $\mathcal{A}$ .

Affine subspaces may be tied in neatly with linear equations with the aid of (13.1) and the following:

**Proposition 13.5.** *Let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be a linear transformation and  $\mathbf{C} \in \text{Im } T$ . Let  $\mathcal{A} = \{\mathbf{V} \in \mathcal{V} \mid T(\mathbf{V}) = \mathbf{C}\}$ , that is,  $\mathcal{A}$  is the set of all vectors  $\mathbf{V}$  in  $\mathcal{V}$  such that  $T(\mathbf{V}) = \mathbf{C}$ . Then  $\mathcal{A}$  is an affine subspace of  $\mathcal{V}$ . In fact if  $\mathbf{A}$  is any vector in  $\mathcal{V}$  such that  $T(\mathbf{A}) = \mathbf{C}$  then  $\mathcal{A} = \mathbf{A} + \ker T$ .*

**PROOF.** Let  $\mathbf{A}$  be any vector in  $\mathcal{V}$  such that  $T(\mathbf{A}) = \mathbf{C}$ . If  $\mathbf{B} \in \mathbf{A} + \ker T$  then  $\mathbf{B} = \mathbf{A} + \mathbf{X}$  for some  $\mathbf{X} \in \ker T$ . Hence

$$T(\mathbf{B}) = T(\mathbf{A} + \mathbf{X}) = T(\mathbf{A}) + T(\mathbf{X}) = T(\mathbf{A}) + \mathbf{0} = T(\mathbf{A}) = \mathbf{C}$$

showing that  $\mathbf{B} \in \mathcal{A}$ . Therefore  $\mathbf{A} + \ker T \subset \mathcal{A}$ . Conversely, if  $\mathbf{B} \in \mathcal{A}$  then

$$T(\mathbf{B} - \mathbf{A}) = T(\mathbf{B}) - T(\mathbf{A}) = \mathbf{C} - \mathbf{C} = \mathbf{0}$$

so that  $\mathbf{B} - \mathbf{A} = \mathbf{X}$  belongs to  $\ker T$ . Then

$$\mathbf{B} = \mathbf{A} + \mathbf{X}$$

with  $\mathbf{X} \in \ker T$ , showing  $\mathbf{B} \in \mathbf{A} + \ker T$ . Therefore  $\mathcal{A} \subset \mathbf{A} + \ker T$ . Combining this with the preceding inclusion yields  $\mathcal{A} = \mathbf{A} + \ker T$  as desired.  $\square$

**Theorem 13.6.** *Suppose (in matrix notation) that*

$$\mathbf{A}\mathbf{X} = \mathbf{B}$$

*is a system of linear equations. Then the sets of all solutions  $\mathcal{S} = \{(s_1, \dots, s_n)\}$  to this linear system is either empty or is an affine subspace of  $\mathbb{R}^n$ . In fact, if  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the linear transformation whose matrix relative to the standard bases is  $\mathbf{A}$ , then  $\mathcal{S}$  is a parallel translate of  $\ker T$  by any vector  $\mathbf{S}$  in  $\mathcal{S}$ , or  $\mathcal{S} = \emptyset$ .*

Thus we see that in order to specify all solutions of the linear system

$$\mathbf{A}\mathbf{X} = \mathbf{B}$$

we must specify a certain affine subspace of  $\mathbb{R}^n$ . This we may do (according to (13.6)) by finding a basis for the solution space of the system

$$\mathbf{A}\mathbf{X} = \mathbf{0}$$

(called the *associated homogeneous system*) and finding a single particular solution to the equation (if there is one)

$$\mathbf{AX} = \mathbf{B}.$$

EXAMPLE 5. Solve the system of linear equations

$$\begin{aligned} 2x_1 + x_2 - 2x_3 + 3x_4 &= 4 \\ 3x_1 + 2x_2 - x_3 + 2x_4 &= 6 \\ 3x_1 + 3x_2 + 3x_3 - 3x_4 &= 6. \end{aligned}$$

*Solution.* The associated homogeneous system is

$$\begin{aligned} (L_1) \quad & 2x_1 + x_2 - 2x_3 + 3x_4 = 0 \\ (L_2) \quad & 3x_1 + 2x_2 - x_3 + 2x_4 = 0 \\ (L_3) \quad & 3x_1 + 3x_2 + 3x_3 - 3x_4 = 0. \end{aligned}$$

And this may be solved as follows: Multiply the first equation by 3 and the second by 2 to get

$$\begin{aligned} (L_4) \quad & 6x_1 + 3x_2 - 6x_3 + 9x_4 = 0 \\ (L_5) \quad & 6x_1 + 4x_2 - 2x_3 + 4x_4 = 0. \end{aligned}$$

Subtract  $(L_4)$  from  $(L_5)$  to get

$$(L_6) \quad x_2 + 4x_3 - 5x_4 = 0.$$

Subtract  $(L_6)$  from  $(L_1)$  to get

$$(L_7) \quad 2x_1 - 6x_3 + 8x_4 = 0.$$

Equations  $(L_7)$  and  $(L_6)$  yield

$$\begin{aligned} x_1 &= 3x_3 - 4x_4 \\ x_2 &= -4x_3 + 5x_4. \end{aligned}$$

A basis for the solution space to the associated homogeneous system is provided by the two vectors

$$(3, -4, 1, 0) \quad (-4, 5, 0, 1).$$

Similar manipulations show that a solution to the original system of equations is

$$(1, 1, 1, 1).$$

Thus the solution space is the affine subspace

$$(1, 1, 1, 1) + \mathcal{L}((3, -4, 1, 0), (-4, 5, 0, 1))$$

of  $\mathbb{R}^4$ , or what is the same, all solutions to the original system are of the form

$$(1, 1, 1, 1) + a(3, -4, 1, 0) + b(-4, 5, 0, 1)$$

where  $a$  and  $b$  are arbitrary numbers.

It is time now to take up in more detail explicit methods for solving systems of linear equations.

## Reduction to echelon form

The first method we will describe for solving systems of linear equations is by far the simplest of the many methods available. It will always work if one perseveres enough. However it is usually far from the shortest method.

**Definition.** A matrix  $A = (a_{ij})$  is said to be an *echelon matrix* iff the first nonzero entry in any row is a 1 and it appears to the right of the first nonzero entry of the preceding row. If in addition the first nonzero entry in a given row is the only nonzero entry in its column we say that the matrix  $A$  is in *reduced echelon form*.

EXAMPLE 6. The matrices

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

are echelon matrices, but not reduced echelon matrices. The following two matrices are reduced echelon matrices

$$\begin{pmatrix} 0 & 1 & 0 & 3 & 4 \\ 0 & 0 & 1 & 5 & 6 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Notice that an echelon matrix or reduced echelon matrix is a very special type of upper triangular matrix. If the coefficient matrix of a linear system

$$AX = B$$

is in *reduced echelon form* then it is possible by inspection to read off all the solutions.

EXAMPLE 7. Solve the linear system

$$\begin{aligned} x_1 + 0x_2 + x_3 + 0x_4 + x_5 + 0x_6 &= 1 \\ 0x_1 + x_2 + x_3 + 0x_4 + 2x_5 + x_6 &= 2 \\ 0x_1 + 0x_2 + 0x_3 + x_4 + 3x_5 + 0x_6 &= 3. \end{aligned}$$

*Solution.* The coefficient matrix is in reduced echelon form. The associated homogeneous system is

$$\begin{aligned}x_1 + 0x_2 + x_3 + 0x_4 + x_5 + 0x_6 &= 0 \\0x_1 + x_2 + x_3 + 0x_4 + 2x_5 + x_6 &= 0 \\0x_1 + 0x_2 + 0x_3 + x_4 + 3x_5 + 0x_6 &= 0.\end{aligned}$$

Thus working our way up from the bottom we find

$$\begin{aligned}x_4 &= -3x_5 \\x_2 &= -x_3 - 2x_5 - x_6 \\x_1 &= -x_3 - x_5.\end{aligned}$$

A basis for the solution space of the associated homogeneous system is therefore seen to be

$$\begin{aligned}(-1, -1, 1, 0, 0, 0) \\(-1, -2, 0, -3, 1, 0) \\(0, -1, 0, 0, 0, 1)\end{aligned}$$

and a solution to the original system is gotten from

$$\begin{aligned}x_4 &= 3 - 3x_5 \\x_2 &= 2 - x_3 - 2x_5 - x_6 \\x_1 &= 1 - x_3 - x_5\end{aligned}$$

by setting  $x_3 = x_5 = x_6 = 0$  and is seen to be

$$(1, 2, 0, 3, 0, 0).$$

Thus the general solution to the original system is seen to be

$$\begin{aligned}(s_1, s_2, s_3, s_4, s_5, s_6) &= (1, 2, 0, 3, 0, 0) + a(-1, -1, 1, 0, 0, 0) \\&\quad + b(-1, -2, 0, -3, 1, 0) + c(0, -1, 0, 0, 0, 1).\end{aligned}$$

(By the *general solution* we mean that any (and all) solution to the original system is obtained for a suitable choice of the three numbers (parameters)  $a, b, c$ .)

In *solving* a system of equations

$$\mathbf{AX} = \mathbf{B}$$

how do we proceed? Well the idea is to exchange the given system for a new system

$$\overline{\mathbf{A}}\mathbf{X} = \overline{\mathbf{B}}$$



which on the one hand has the same solution set as the original system but where the coefficient matrix has a more useful distribution of zeros. In fact as the preceding example suggests we would like for the new coefficient matrix to be in reduced echelon form. What kinds of operations can we perform on the original system to accomplish this end? Clearly the following three operations and any combination of them will be allowed, because they will not change the solution space.

- (1) Interchange any two equations.
- (2) Multiply any equation by a *nonzero* number.
- (3) Add one equation to another.

Thus the following theorem says that by careful and perhaps lengthy application of these operations we may solve any linear system.

**Theorem 13.7.** *Let (in matrix notation)*

$$\mathbf{AX} = \mathbf{B}$$

*be a system of linear equations. Then by a combination of the operations of interchanging two equations, multiplying any equation by a nonzero number, and adding one equation to another we may obtain a system of equations*

$$\overline{\mathbf{A}}\mathbf{X} = \overline{\mathbf{B}}$$

*whose coefficient matrix is in reduced echelon form.*

The proof of (13.7) is an orgy of manipulation with indices and is omitted. Rather we will illustrate the technique of the proof by working some numerical examples.

**Definition.** If  $\mathbf{AX} = \mathbf{B}$  is a system of linear equations then the matrix

$$\left( \begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right)$$

is called the *augmented matrix* of the linear system.

Operations (1), (2), (3) above to be performed on the linear equations can instead be performed on the augmented matrix in order to save space. We illustrate this in the following examples also.

EXAMPLE 8. Solve the linear system

$$\begin{aligned}x + 2y - 3z &= 4 \\x + 3y + z &= 11 \\2x + 5y - 4z &= 13 \\2x + 6y + 2z &= 22.\end{aligned}$$

*Solution.* The augmented matrix for this system is

$$\left(\begin{array}{cccc}1 & 2 & -3 & 4 \\1 & 3 & 1 & 11 \\2 & 5 & -4 & 13 \\2 & 6 & 2 & 22\end{array}\right).$$

We proceed to apply the three basic operations to reduce this matrix to row echelon form

$$\begin{aligned}L_2 - L_1 \quad L_3 - 2L_1 \quad L_4 - 2L_1 & \left(\begin{array}{cccc}1 & 2 & -3 & 4 \\0 & 1 & 4 & 7 \\0 & 1 & 2 & 5 \\0 & 2 & 8 & 14\end{array}\right) \rightsquigarrow \begin{array}{l} L_1 - 2L_2 \\ L_3 - L_2 \\ L_4 - 2L_2 \end{array} \left(\begin{array}{cccc}1 & 0 & -11 & -10 \\0 & 1 & 4 & 7 \\0 & 0 & -2 & -2 \\0 & 0 & 0 & 0\end{array}\right) \\ \rightsquigarrow -\frac{1}{2}L_3 & \left(\begin{array}{cccc}1 & 0 & -11 & -10 \\0 & 1 & 4 & 7 \\0 & 0 & 1 & 1 \\0 & 0 & 0 & 0\end{array}\right) \rightsquigarrow \begin{array}{l} L_1 + 11L_3 \\ L_2 - 4L_3 \end{array} \left(\begin{array}{cccc}1 & 0 & 0 & 1 \\0 & 1 & 0 & 3 \\0 & 0 & 1 & 1 \\0 & 0 & 0 & 0\end{array}\right)\end{aligned}$$

which is the augmented matrix of the system

$$\begin{aligned}x_1 &= 1 \\x_2 &= 3 \\x_3 &= 1\end{aligned}$$

having the same solutions as the original system. Thus the original system has the unique solution  $(1, 3, 1)$ .

EXAMPLE 9. Solve the linear system

$$\begin{aligned}2x + y - 2z + 3w &= 1 \\3x + 2y - z + 2w &= 4 \\3x + 3y + 3z - 3w &= 5.\end{aligned}$$

*Solution.* The augmented matrix for this system is

$$\left(\begin{array}{cccccc}2 & 1 & -2 & 3 & 1 \\3 & 2 & -1 & 2 & 4 \\3 & 3 & 3 & -3 & 5\end{array}\right)$$

which reduces to echelon form as follows:

$$\begin{aligned}
 & \begin{matrix} 3L_1 \\ 2L_2 \\ 2L_3 \end{matrix} \begin{pmatrix} 6 & 3 & -6 & 9 & 3 \\ 6 & 4 & -2 & 4 & 8 \\ 6 & 6 & 6 & -6 & 10 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 6 & 3 & -6 & 9 & 3 \\ 0 & 1 & 4 & -5 & 5 \\ 0 & 3 & 12 & -15 & 7 \end{pmatrix} \begin{matrix} \\ L_2 - L_1 \\ L_3 - L_1 \end{matrix} \\
 & \rightsquigarrow 3L_2 \begin{pmatrix} 6 & 3 & -6 & 9 & 3 \\ 0 & 3 & 12 & -15 & 15 \\ 0 & 3 & 12 & -15 & 7 \end{pmatrix} \\
 & \rightsquigarrow \begin{matrix} L_1 - L_2 \\ L_3 - L_2 \end{matrix} \begin{pmatrix} 6 & 0 & -18 & 24 & -12 \\ 0 & 3 & 12 & -15 & 15 \\ 0 & 0 & 0 & 0 & -8 \end{pmatrix} \\
 & \rightsquigarrow \begin{matrix} \frac{1}{6}L_1 \\ \frac{1}{3}L_2 \\ -\frac{1}{8}L_3 \end{matrix} \begin{pmatrix} 1 & 0 & -3 & 4 & -2 \\ 0 & 1 & 4 & -5 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
 \end{aligned}$$

which is the augmented matrix of the linear system

$$\begin{aligned}
 x & - 3z + 4w = -2 \\
 y + 4z - 5w & = 5 \\
 0 & = 1.
 \end{aligned}$$

The impossibility of the last equation shows that this linear system has *no* solutions, hence the original system has no solutions. Note that the column space of the coefficient matrix for the new system is spanned by

$$(1, 0, 0), \quad (0, 1, 0), \quad (-3, 4, 0), \quad (4, -5, 0)$$

which is the  $xy$ -plane in  $\mathbb{R}^3$ . While the vector  $(-2, 5, 1)$  does not lie in this plane.

As a last example of this method we work:

**EXAMPLE 10.** Solve the linear system

$$\begin{aligned}
 x_2 + 3x_3 + x_4 - x_5 & = 2 \\
 x_1 - x_2 + 3x_3 - 4x_4 + 2x_5 & = 6 \\
 x_1 + x_2 - x_3 + 2x_4 + x_5 & = 1 \\
 x_1 - x_3 & = 1.
 \end{aligned}$$

**Solution.** The augmented matrix is

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

Reducing to echelon form interchange  $L_1$  and  $L_2$

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

$$\begin{array}{l} L_3 - L_1 \\ \rightarrow L_4 - L_1 \end{array} \begin{pmatrix} 1 & -1 & 3 & -4 & 2 & 6 \\ 0 & 1 & 3 & 1 & -1 & 2 \\ 0 & 2 & -4 & 6 & -1 & -5 \\ 0 & 1 & -4 & 4 & -1 & -5 \end{pmatrix}$$

$$\begin{array}{l} L_1 + L_2 \\ \rightarrow L_3 - 2L_2 \\ L_4 - L_2 \end{array} \begin{pmatrix} 1 & 0 & 6 & -3 & 1 & 8 \\ 0 & 1 & 3 & 1 & -1 & 2 \\ 0 & 0 & -10 & 4 & 1 & -9 \\ 0 & 0 & -7 & 3 & 0 & -7 \end{pmatrix}$$

$$\begin{array}{l} \rightarrow L_3 \times (-7) \\ L_4 \times 10 \end{array} \begin{pmatrix} 1 & 0 & 6 & -3 & 1 & 8 \\ 0 & 1 & 3 & 1 & -1 & 2 \\ 0 & 0 & 70 & -28 & -7 & 63 \\ 0 & 0 & -70 & 30 & 0 & -70 \end{pmatrix}$$

$$\begin{array}{l} \rightarrow L_4 + L_3 \end{array} \begin{pmatrix} 1 & 0 & 6 & -3 & 1 & 8 \\ 0 & 1 & 3 & 1 & -1 & 2 \\ 0 & 0 & 70 & -28 & -7 & 63 \\ 0 & 0 & 0 & 2 & -7 & -7 \end{pmatrix}$$

$$\begin{array}{l} \rightarrow L_3/70 \\ L_4/2 \end{array} \begin{pmatrix} 1 & 0 & 6 & -3 & 1 & 8 \\ 0 & 1 & 3 & 1 & -1 & 2 \\ 0 & 0 & 1 & -\frac{28}{70} & -\frac{1}{10} & \frac{63}{70} \\ 0 & 0 & 0 & 1 & -\frac{7}{2} & -\frac{7}{2} \end{pmatrix}$$

which is the augmented matrix of the system

$$\begin{array}{rclclcl} x_1 & + & 6x_3 & - & 3x_4 & + & x_5 & = & 8 \\ x_2 & + & 3x_3 & + & x_4 & - & x_5 & = & 2 \\ x_3 & - & \frac{2}{5}x_4 & - & \frac{1}{10}x_5 & = & \frac{9}{10} \\ x_4 & - & \frac{7}{2}x_5 & = & -\frac{7}{2}. \end{array}$$

The associated homogeneous system is

$$\begin{array}{rclclcl} x_1 & + & 6x_3 & - & 3x_4 & + & x_5 & = & 0 \\ x_2 & + & 3x_3 & + & x_4 & - & x_5 & = & 0 \\ x_3 & - & \frac{2}{5}x_4 & - & \frac{1}{10}x_5 & = & 0 \\ x_4 & - & \frac{7}{2}x_5 & = & 0 \end{array}$$

### 13: Systems of linear equations

or

$$x_4 = \frac{7}{2} x_5$$

$$x_3 = \frac{2x_4}{5} + \frac{x_5}{10}$$

$$x_2 = -3x_3 - x_4 + x_5$$

$$x_1 = -6x_3 + 3x_4 - x_5$$

and so a basis for the solution space consists of the single vector (set  $x_5 = 1$ )

$$\left( \frac{1}{2}, -7, \frac{3}{2}, \frac{7}{2}, 1 \right)$$

or if you don't like fractions

$$(1, -14, 3, 7, 2).$$

To find a particular solution to the nonhomogeneous system we use the equations

$$x_4 = -\frac{7}{2} + \frac{7}{2} x_5$$

$$x_3 = \frac{9}{10} + \frac{2}{5} x_4 + \frac{x_5}{10}$$

$$x_2 = 2 - 3x_3 - x_4 + x_5$$

$$x_1 = 8 - 6x_3 + 3x_4 - x_5.$$

Setting  $x_5 = 0$  gives the particular solution

$$x_5 = 0$$

$$x_4 = -\frac{7}{2}$$

$$x_3 = \frac{9}{10} + \frac{2}{5} \left( -\frac{7}{2} \right) = \frac{9}{10} - \frac{14}{10} = \frac{-5}{10} = -\frac{1}{2}$$

$$x_2 = 2 - 3 \left( -\frac{1}{2} \right) + \frac{7}{2} = 2 + \frac{3}{2} + \frac{7}{2} = 7$$

$$x_1 = 8 - 6 \left( -\frac{1}{2} \right) + 3 \left( -\frac{7}{2} \right) = 8 + 3 - \frac{21}{2} = \frac{1}{2}$$

that is,  $(\frac{1}{2}, 7, -\frac{1}{2}, -\frac{7}{2}, 0)$  solves the nonhomogeneous system. So the general solution to the original system is

$$(s_1, s_2, s_3, s_4, s_5) = (\frac{1}{2}, 7, -\frac{1}{2}, -\frac{7}{2}, 0) + a(1, -14, 3, 7, 2).$$

Setting  $x_5 = 1$  gives another particular solution, namely  $(1, 0, 1, 0, 1)$ , so the general solution could equally well be written

$$(s_1, s_2, s_3, s_4, s_5) = (1, 0, 1, 0, 1) + a(1, -14, 3, 7, 2)$$

where  $a$  is an arbitrary number.

## The simplex method

The method of solving a linear system

$$\mathbf{AX} = \mathbf{B}$$

by reducing the coefficient matrix  $\mathbf{A}$  to reduced echelon form is a system of trading off the given linear system for a new and simpler linear system, whose solution sets are the same,

$$\overline{\mathbf{A}}\mathbf{X} = \overline{\mathbf{B}}$$

and from which the solutions may be read off by a purely mechanical process. There is another general approach to the solution of linear systems which involves a more direct kind of horse trading of “knowns and unknowns.” To be more specific let us examine our linear system in longhand again. It looks like

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

Now we may regard this system as expressing the known quantities  $b_1, \dots, b_m$  as linear combinations of the unknown quantities  $x_1, \dots, x_n$ . To solve the system means to reverse this process, that is to express the unknowns  $x_1, \dots, x_n$  as linear combinations of the knowns  $b_1, \dots, b_m$ . For example suppose that  $a_{ij} \neq 0$ . Then we may solve the equation

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

for  $x_j$  obtaining

$$\begin{aligned} x_j = & -\frac{1}{a_{ij}} [a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ij-1}x_{j-1} \\ & - b_i + a_{ij+1}x_{j+1} + \cdots + a_{in}x_n]. \end{aligned}$$

(Note that  $x_j$  does *not* appear on the right in this formula.) We may now take this formula for  $x_j$  and put it into the other equations of the system obtaining the new system

$$\begin{aligned} &\left(a_{11} - \frac{a_{1j}a_{j1}}{a_{jj}}\right)x_1 + \cdots + \frac{a_{1j}}{a_{jj}}b_j + \cdots + \left(a_{1n} - \frac{a_{1j}a_{jn}}{a_{jj}}\right)x_n = b_1 \\ &\vdots \\ &-\frac{a_{i1}}{a_{jj}}x_1 - \cdots + \frac{b_i}{a_{jj}} - \cdots - \frac{a_{in}}{a_{jj}}x_n = x_j \\ &\vdots \\ &\left(a_{m1} - \frac{a_{mj}a_{j1}}{a_{jj}}\right)x_1 + \cdots + \frac{a_{mj}}{a_{jj}}b_j + \cdots + \left(a_{mn} - \frac{a_{mj}a_{jn}}{a_{jj}}\right)x_n = b_m. \end{aligned}$$

This process of obtaining a new system of equations is called a *pivot operation* on  $a_{ij}$ . Note that in this new system the unknown  $x_j$  does not appear in any of the equations on the left. It has been traded for the known quantity  $b_i$ . By repeated application of this process we may be able to obtain a linear system in which each right-hand side is an  $x_i$  and each left-hand side is a linear combination of the  $b_j$  and those  $x_k$  which do not appear on the left. In this way we will have solved the system. (Note that each step is reversible.) In order to apply this method (known as the *simplex method*) to solve systems of equations we must have available a simple way to write down the coefficients of the system obtained by pivoting at  $a_{ij}$ . To do this we require a simple definition.<sup>1</sup>

**Definition.** Suppose that

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

is a  $2 \times 2$  matrix. The *determinant* of  $C$ , denoted by  $\det C$ , is the number

$$\det C = c_{11}c_{22} - c_{21}c_{12}.$$

**EXAMPLE 11.** Find the determinant of the matrix

$$\begin{pmatrix} 1 & 4 \\ 3 & 4 \end{pmatrix}.$$

**Solution.** We have

$$\det \begin{pmatrix} 1 & 4 \\ 3 & 4 \end{pmatrix} = 4 - 4(3) = -12.$$

<sup>1</sup> Recall (11.3) our discussion of inverting  $2 \times 2$  matrices.

EXAMPLE 12. Find the determinant of the matrix

$$\begin{pmatrix} 1 & -3 \\ 1 & -3 \end{pmatrix}.$$

*Solution.* We have

$$\det \begin{pmatrix} 1 & -3 \\ 1 & -3 \end{pmatrix} = 1(-3) - (-3)1 = -3 + 3 = 0.$$

Notice that a matrix with all positive entries can have a negative determinant and a matrix with all nonzero entries a zero determinant.

Return now to our linear system

$$\mathbf{AX} = \mathbf{B}.$$

If we suppose that  $a_{ij} \neq 0$  then we may apply a pivot operation at  $a_{ij}$  to obtain a *new* linear system

$$\bar{\mathbf{A}}\bar{\mathbf{X}} = \bar{\mathbf{B}}$$

where  $(\bar{\mathbf{A}} = (\bar{a}_{rs}))$

$$\bar{a}_{rs} = \begin{cases} \frac{1}{a_{ij}} \det \begin{pmatrix} a_{rs} & a_{rj} \\ a_{is} & a_{ij} \end{pmatrix}, & \text{if } r \neq i \text{ and } s \neq j \\ \frac{a_{rj}}{a_{ij}} & \text{if } r \neq i \text{ and } s = j \\ -\frac{a_{is}}{a_{ij}} & \text{if } r = i \text{ and } s \neq j \\ \frac{1}{a_{ij}} & \text{if } r = i \text{ and } s = j \end{cases}$$

and

$$\bar{\mathbf{B}} = \begin{pmatrix} b_1 \\ \vdots \\ b_{i-1} \\ x_j \\ b_{i+1} \\ \vdots \\ b_m \end{pmatrix} \quad \bar{\mathbf{X}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{j-1} \\ b_j \\ x_{j+1} \\ \vdots \\ x_n \end{pmatrix}.$$

With these formulas in mind let us run through a few examples by this method.



### 13: Systems of linear equations

Note that we change not only the coefficient matrix **A** and constant matrix **B** but also the matrix of unknowns. Our notations should keep track of this fact.

**EXAMPLE 13.** Solve the linear system

$$\begin{aligned} 2x - y + z &= 2 \\ x + 2y - z &= 3 \\ 3x + y + 2z &= -1. \end{aligned}$$

*Solution:* We represent our linear system by the diagram

$$\begin{array}{ccc|c} x & y & z & \\ \hline 2 & -1 & 1 & 2 \\ 1 & 2 & -1 & 3 \\ 3 & 1 & 2 & -1 \end{array}$$

It is always easiest to pivot on a 1 so let us pivot on the  $a_{13}$  position where there is a 1. Our new system is then represented by the diagram

$$\begin{array}{ccc|c} x & y & z & \\ \hline -2 & 1 & 1 & z \\ 3 & 1 & -1 & 3 \\ -1 & 3 & 2 & -1 \end{array}$$

(Since we are pivoting on a 1 in the (1, 3) position the first row changes sign and the third column changes not at all.) For example ( $i = 1, j = 3$ )

$$\bar{a}_{21} = \det \begin{pmatrix} a_{21} & a_{23} \\ a_{11} & a_{13} \end{pmatrix} = \det \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = 1 + 2 = 3$$

$$\bar{a}_{22} = \det \begin{pmatrix} a_{22} & a_{23} \\ a_{12} & a_{13} \end{pmatrix} = \det \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = 2 - 1 = 1$$

$$\bar{a}_{31} = \det \begin{pmatrix} a_{31} & a_{33} \\ a_{11} & a_{13} \end{pmatrix} = \det \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} = 3 - 4 = -1$$

$$\bar{a}_{32} = \det \begin{pmatrix} a_{32} & a_{33} \\ a_{12} & a_{13} \end{pmatrix} = \det \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = 1 + 2 = 3.$$

This new system has a handy 1 in the (2, 2) spot, so let's perform a pivot there. The resulting system is represented by the diagram

$$\begin{array}{ccc|c} x & y & z & \\ \hline -5 & 1 & 2 & z \\ -3 & 1 & 1 & y \\ -10 & 3 & 5 & -1 \end{array}$$

Then  $i = 2$ ,  $j = 2$  and for example

$$\bar{a}_{11} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \det \begin{pmatrix} -2 & 1 \\ 3 & 1 \end{pmatrix} = -2 - 3 = -5$$

$$\bar{a}_{13} = \det \begin{pmatrix} a_{13} & a_{12} \\ a_{23} & a_{22} \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = 1 + 1 = 2$$

$$\bar{a}_{31} = \det \begin{pmatrix} a_{31} & a_{32} \\ a_{21} & a_{22} \end{pmatrix} = \det \begin{pmatrix} -1 & 3 \\ 3 & 1 \end{pmatrix} = -1 - 9 = -10$$

$$\bar{a}_{33} = \det \begin{pmatrix} a_{33} & a_{32} \\ a_{23} & a_{22} \end{pmatrix} = \det \begin{pmatrix} 2 & 3 \\ -1 & 1 \end{pmatrix} = 2 + 3 = 5.$$

Finally we pivot on the (1, 3) position to obtain

$$\begin{array}{ccc|c} -1 & 3 & 2 & \\ \hline \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & z \\ \frac{3}{10} & \frac{1}{10} & -\frac{1}{2} & y \\ -\frac{1}{10} & \frac{3}{10} & \frac{5}{10} & x \end{array}$$

$$\bar{\bar{a}}_{12} = \frac{-1}{10} \det \begin{pmatrix} 1 & -5 \\ 3 & -10 \end{pmatrix} = \frac{-1}{10} (-10 + 15) = \frac{-5}{10} = -\frac{1}{2}$$

$$\bar{\bar{a}}_{13} = \frac{-1}{10} \det \begin{pmatrix} 2 & -5 \\ 5 & -10 \end{pmatrix} = \frac{-1}{10} (-20 + 25) = -\frac{1}{2}$$

$$\bar{\bar{a}}_{22} = \frac{-1}{10} \det \begin{pmatrix} 1 & -3 \\ 3 & -10 \end{pmatrix} = \frac{-1}{10} (-10 + 9) = \frac{1}{10}$$

$$\bar{\bar{a}}_{23} = \frac{-1}{10} \det \begin{pmatrix} 1 & -3 \\ 5 & -10 \end{pmatrix} = \frac{-1}{10} (-10 + 15) = \frac{-5}{10} = -\frac{1}{2}$$

and thus we find

$$z = \frac{1}{2}(-1) - \frac{1}{2}(3) - \frac{1}{2}(2) = -\frac{1}{2} - \frac{3}{2} - 1 = -3$$

$$y = \frac{3}{10}(-1) + \frac{1}{10}(3) - \frac{1}{2}(2) = -1$$

$$x = \frac{-1}{10}(-1) + \frac{3}{10}(3) + \frac{5}{10}(2) = \frac{20}{10} = 2.$$

So there is only one solution and it is

$$x = 2, \quad y = -1, \quad z = -3.$$

As a second example consider the following:

EXAMPLE 14. Solve the linear system

$$\begin{aligned}x - y + z &= 3 \\2x + y - z &= 6 \\-x + 2y + 2z &= 1 \\3x - 2y - 2z &= -1.\end{aligned}$$

*Solution.* We represent our system by the diagram

$$\begin{array}{ccc|c}x & y & z & \\ \hline 1 & -1 & 1 & 3 \\ 2 & 1 & -1 & 6 \\ -1 & 2 & 2 & 1 \\ 3 & -2 & -2 & -1\end{array}$$

We will pivot on the  $a_{11}$  position. Our new system is then represented by the diagram

$$\begin{array}{ccc|c}3 & y & z & \\ \hline 1 & 1 & -1 & x \\ 2 & 3 & -3 & 6 \\ -1 & 1 & 3 & 1 \\ 3 & 1 & -5 & -1\end{array}$$

(since we are pivoting on a 1 in position  $a_{11}$  the first row changes sign and the first column is unchanged).

For example ( $i = 1, j = 1$ )

$$\bar{a}_{22} = \det \begin{pmatrix} a_{22} & a_{21} \\ a_{12} & a_{11} \end{pmatrix} = \det \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = 1 - (-1)(2) = 3$$

$$\bar{a}_{23} = \det \begin{pmatrix} a_{23} & a_{21} \\ a_{13} & a_{11} \end{pmatrix} = \det \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} = -1 - 2 = -3$$

$$\bar{a}_{32} = \det \begin{pmatrix} a_{32} & a_{31} \\ a_{12} & a_{11} \end{pmatrix} = \det \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = 2 - (-1)(-1) = 1$$

$$\bar{a}_{33} = \det \begin{pmatrix} a_{33} & a_{31} \\ a_{13} & a_{11} \end{pmatrix} = \det \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} = 2 + 1 = 3$$

$$\bar{a}_{42} = \det \begin{pmatrix} a_{42} & a_{41} \\ a_{12} & a_{11} \end{pmatrix} = \det \begin{pmatrix} -2 & 3 \\ -1 & 1 \end{pmatrix} = -2 - (-1)(3) = 1$$

$$\bar{a}_{43} = \det \begin{pmatrix} a_{43} & a_{41} \\ a_{13} & a_{11} \end{pmatrix} = \det \begin{pmatrix} -2 & 3 \\ 1 & 1 \end{pmatrix} = -2 - 3 = -5.$$

Let us now pivot on the  $a_{32}$  position where there is a 1 in the new matrix. This will give us a system represented by the diagram

$$\begin{array}{ccc|c}
 3 & 1 & z & \\
 \hline
 2 & 1 & -4 & x \\
 5 & 3 & -12 & 6 \\
 1 & 1 & -3 & y \\
 4 & 1 & -8 & -6
 \end{array}$$

We now have  $i = 3, j = 2$  and

$$\bar{a}_{11} = \det \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} \\ \bar{a}_{31} & \bar{a}_{32} \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = 1 - (-1) = 2$$

$$\bar{a}_{21} = \det \begin{pmatrix} 2 & 3 \\ -1 & 1 \end{pmatrix} = 2 - (-3) = 5$$

$$\bar{a}_{41} = \det \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix} = 3 + 1 = 4$$

$$\bar{a}_{13} = \det \begin{pmatrix} -1 & 1 \\ 3 & 1 \end{pmatrix} = -1 - 3 = -4$$

$$\bar{a}_{23} = \det \begin{pmatrix} -3 & 3 \\ 3 & 1 \end{pmatrix} = -3 - 9 = -12$$

$$\bar{a}_{43} = \det \begin{pmatrix} -5 & 1 \\ 3 & 1 \end{pmatrix} = -5 - 3 = -8.$$

If we pivot once more we can get rid of the  $z$  and thereby solve the system. The most convenient place to pivot is on the  $-8$  in the  $(4, 4)$  position. This will give us

$$\begin{array}{ccc|c}
 3 & 1 & -6 & \\
 \hline
 & & \frac{1}{2} & x \\
 8 & -12 & \frac{3}{2} & 6 \\
 & & \frac{1}{8} & y \\
 \frac{1}{2} & \frac{1}{8} & -\frac{1}{8} & z
 \end{array}$$

There is no need to calculate further as we have found the contradiction

$$24 - 12 - 9 = 6$$

which shows the original system has no solutions.

As our final example we consider:

### 13: Systems of linear equations

**EXAMPLE 15.** Solve the linear system

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

*Solution.* Our system is represented by

$x$	$y$	$z$	
1	2	-3	6
2	-1	4	2
4	3	-2	14

Pivoting on the 1 in the (1, 1) position gives the system

6	$y$	$z$	
1	-2	3	$x$
2	-5	10	2
4	-5	10	14

Let us pivot now on the 10 in the (3, 3) position to obtain

6	$y$	14	
$-\frac{1}{5}$	$-\frac{1}{2}$	$\frac{3}{10}$	$x$
-2	0	1	2
$-\frac{2}{5}$	$\frac{1}{2}$	$\frac{1}{10}$	$z$

from which we find

$$\begin{aligned}x &= -\frac{6}{5} - \frac{y}{2} + \frac{3(14)}{10} = \frac{42 - 12}{10} - \frac{y}{2} \\&= 3 - \frac{y}{2}\end{aligned}$$

$$-12 + 14 = 2$$

$$z = -\frac{12}{5} + \frac{1}{2}y + \frac{14}{10} = \frac{14 - 24}{10} + \frac{1}{2}y = \frac{1}{2}y - 1$$

or

$$x = 3 - \frac{y}{2}$$

$$z = \frac{1}{2}y - 1$$

so the solution space is

$$(3, 0, -1) + \mathcal{L}(-1, 2, 1).$$

## EXERCISES

1. Solve the following systems of linear equations.

$$(a) \begin{cases} x - 3y + z = -2 \\ 2x + y - z = 6 \\ x + 2y + 2z = 2 \end{cases}$$

$$(b) \begin{cases} x - y + z = 2 \\ x + y = 1 \\ x + y + z = 8 \end{cases}$$

$$(c) \begin{cases} x + 2y - 3z = 4 \\ -x + y + z = 0 \\ 4x - 2y + z = 9 \end{cases}$$

$$(d) \begin{cases} 3x - 6y + z = 7 \\ x + 2y + z = 5 \\ -2x + 5y - 2z = -1 \end{cases}$$

2. Find a system of linear equations in three unknowns whose solution space is the line through  $(1, 1, 1)$  and the origin.

3. Find the solution space of following systems of linear homogeneous equations.

$$(a) \begin{cases} x - y + z - w = 0 \\ 2x + y - z + 2w = 0 \\ 2y + 3z + w = 0 \end{cases}$$

$$(b) \begin{cases} 3x_1 + 2x_2 + x_3 - x_4 - x_5 = 0 \\ x_1 - x_2 - x_3 - x_4 + 2x_5 = 0 \\ -x_1 + 2x_2 + 3x_3 + x_4 - x_5 = 0 \end{cases}$$

$$(c) \begin{cases} 2x + 3y - z = 0 \\ x - y + z = 0 \\ x + 9y - 5z = 0 \end{cases}$$

$$(d) \begin{cases} 11x - 8y + 5z = 0 \\ 3x - y - 2z = 0 \\ -x - 12y + 13z = 0 \end{cases}$$

4. Find the solution space of the following systems.

$$(a) \begin{cases} 2x + 6y - z + w = -3 \\ x - y + z - w = 2 \\ -x - 3y + 3z + 2w = 9 \end{cases}$$

### 13: Systems of linear equations

$$(b) \begin{cases} x_2 + 2x_3 - x_4 + x_5 = 5 \\ x_1 - x_2 + x_3 + 2x_4 - x_5 = 6 \\ x_1 + x_2 - x_3 + x_5 = -2 \\ x_1 + x_3 - 3x_4 - 2x_5 = -3 \end{cases}$$

$$(c) \begin{cases} x - 2y + z = 5 \\ 2x + y - 2z = 7 \\ x - 7y + 5z = 8 \end{cases}$$

$$(d) \begin{cases} 3x + y - z = 10 \\ x - 2y - z = -2 \\ -x + y + z = 0 \\ 2x - y - 3z = 7 \end{cases}$$

5. Solve the linear systems:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

6. Write a flow chart for a computer program that reduces an  $m \times n$  matrix to a reduced echelon form. If you have access to a machine write, debug, and run the program.

7. For which value of  $a$  does the following system of equations have a solution

$$\begin{aligned} y + z &= 2, \\ x + y + z &= a, \\ x + y &= 2? \end{aligned}$$

8. For which values of  $a$  does the following system of equations have a solution

$$\begin{aligned} y + z &= 2, \\ x + ay + z &= 2, \\ x + y &= 2? \end{aligned}$$

Suppose that

$$T: \mathcal{V} \rightarrow \mathcal{V}$$

is a linear transformation of the vector space  $\mathcal{V}$  to itself. Such linear transformations have a special name (because their domain and range space are the same), they are called *endomorphisms of  $\mathcal{V}$* .

**Proposition 14.1.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional vector space  $\mathcal{V}$ . Then there exist bases  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  for  $\mathcal{V}$  such that the matrix of  $T$  is*

$$\left( \begin{array}{ccc|ccc} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & 0 & \\ \hline & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{array} \right) \left. \begin{array}{l} \left. \begin{array}{l} \text{---} \end{array} \right\} k \\ \left. \begin{array}{l} \text{---} \end{array} \right\} n - k \end{array} \right\}$$

for some integer  $k$ . (The integer  $k$  is usually called the *rank* of the linear transformation  $T$ . In fact  $k = \dim \operatorname{Im} T$ , as our proof will show.) (Compare to (12<sup>bis</sup>.1).)

**PROOF.** Let  $\mathbf{C}_1, \dots, \mathbf{C}_m$  be a basis for  $\ker T$ . By (6.7) we may find vectors  $\mathbf{C}_{m+1}, \dots, \mathbf{C}_n$  so that  $\mathbf{C}_1, \dots, \mathbf{C}_m, \mathbf{C}_{m+1}, \dots, \mathbf{C}_n$  is a basis for  $\mathcal{V}$ . Thus  $n = \dim \mathcal{V}$ . Now we will require:

**Lemma.** *The vectors  $T(\mathbf{C}_{m+1}), T(\mathbf{C}_{m+2}), \dots, T(\mathbf{C}_n)$  are linearly independent.*



PROOF. Suppose to the contrary that the vectors  $T(C_{m+1}), \dots, T(C_n)$  are linearly dependent. Then we may find numbers  $c_{m+1}, \dots, c_n$ , not all zero, so that

$$c_{m+1}T(C_{m+1}) + \dots + c_nT(C_n) = 0.$$

Therefore

$$T(c_{m+1}C_{m+1} + \dots + c_nC_n) = c_{m+1}T(C_{m+1}) + \dots + c_nT(C_n) = 0.$$

Hence the vector  $c_{m+1}C_{m+1} + \dots + c_nC_n$  belongs to the kernel of  $T$ . Remember that we chose the vectors  $C_1, \dots, C_m$  to be a basis for  $\ker T$  and therefore

$$c_{m+1}C_{m+1} + \dots + c_nC_n = c_1C_1 + \dots + c_mC_m$$

for suitable numbers  $c_1, \dots, c_n$ . But then

$$(-c_1)C_1 + (-c_2)C_2 + \dots + (-c_m)C_m + c_{m+1}C_{m+1} + \dots + c_nC_n = 0$$

which (since not all of  $c_1, \dots, c_n$  are zero) means that  $C_1, \dots, C_n$  are linearly dependent contrary to the fact that they are a basis for  $\mathcal{V}$ . Therefore the vectors  $T(C_{m+1}), \dots, T(C_n)$  are linearly independent.  $\square$

PROOF OF 14.1, CONTINUED. Note that

$$\begin{aligned} \operatorname{Im} T &= \mathcal{L}(T(C_1), \dots, T(C_n)) = \mathcal{L}(0, \dots, 0, T(C_{m+1}), \dots, T(C_n)) \\ &= \mathcal{L}(T(C_{m+1}), \dots, T(C_n)) \end{aligned}$$

and hence  $T(C_{m+1}), \dots, T(C_n)$  is a basis for  $\operatorname{Im} T$ .

Apply (6.7) again to choose vectors  $D_1, \dots, D_s$  so that  $\{D_1, \dots, D_s, T(C_{m+1}), \dots, T(C_n)\}$  is a basis for  $\mathcal{V}$ . Now note  $s = m$  since  $s + n - m = \dim \mathcal{V} = n$ . Set

$$\begin{array}{ll} A_1 = C_{m+1} & B_1 = T(C_{m+1}) \\ \vdots & \vdots \\ A_{n-m} = C_n & B_{n-m} = T(C_n) \\ A_{n-m+1} = C_1 & B_{n-m+1} = D_1 \\ \vdots & \vdots \\ A_n = C_m & B_n = D_s. \end{array}$$

Then we have

$$\begin{aligned} T(A_1) &= B_1 \\ T(A_2) &= B_2 \\ &\vdots \\ T(A_{n-m}) &= B_{n-m} \\ T(A_{n-m+1}) &= 0 \\ &\vdots \\ T(A_n) &= 0. \end{aligned}$$

If we put  $k = n - m$  then the matrix of  $T$  relative to the basis pair  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}, \{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  is

$$\begin{matrix} k \\ m = n - k \end{matrix} \left\{ \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & \ddots \\ & & & & & 0 \end{pmatrix} \right.$$

as required. □

It follows from (14.1) that if when we study an endomorphism

$$T: \mathcal{V} \rightarrow \mathcal{V}$$

on a finite-dimensional vector space through its matrix representatives we will perhaps learn nothing more useful about  $T$  than its rank if we are free to choose different bases in the domain and range of  $T$ . Since the domain and range of  $T$  are the same it is reasonable, in seeking to force the matrix to reveal more of the structure of  $T$ , to demand that we use the *same* basis in both the domain and range of  $T$ . If in this way we were to obtain a *diagonal* matrix then the structure of  $T$  will be completely revealed. Now suppose that  $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$  is a basis for  $\mathcal{V}$  such that the matrix of  $T$  relative to this basis (*used in both domain and range*) is diagonal, say

$$\begin{pmatrix} e_1 & & & 0 \\ & e_2 & & \\ & & \ddots & \\ 0 & & & e_n \end{pmatrix}.$$

What does this mean? It means that

$$T(\mathbf{E}_i) = e_i \mathbf{E}_i \quad i = 1, \dots, n.$$

That is,  $T$  is represented by a diagonal matrix iff there is a basis  $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$  for  $\mathcal{V}$  and numbers  $e_1, \dots, e_n$  such that

$$T(\mathbf{E}_i) = e_i \mathbf{E}_i \quad i = 1, \dots, n.$$

This discussion suggests that we introduce the following definition:

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of  $\mathcal{V}$ . A number  $e$  is called an *eigenvalue* of  $T$  iff there exists a **nonzero** vector  $\mathbf{E}$  such that

$$T(\mathbf{E}) = e\mathbf{E}.$$

Such a vector  $\mathbf{E}$  is then called an *eigenvector* of  $T$  associated to the eigenvalue  $e$ .

Throughout our study of endomorphisms we will insist that our matrix representatives be constructed by using the same basis for both the domain and range. If  $T: \mathcal{V} \rightarrow \mathcal{V}$  is an endomorphism and  $\{A_1, \dots, A_n\}$  a basis for  $\mathcal{V}$  we will use the phrase “the matrix of  $T$  relative to  $\{A_1, \dots, A_n\}$ ” (and similar such expressions) to mean the matrix  $A = (a_{ij})$  representing  $T$  obtained by using the basis  $\{A_1, \dots, A_n\}$  in both the domain and range of  $T$ , that is,

$$T(A_j) = \sum_{i=1}^n a_{ij} A_i$$

for  $j = 1, \dots, n$ . In this way if  $A_i$  happens to be an eigenvector of  $T$  corresponding to the eigenvalue  $e_i$  then

$$T(A_i) = 0A_1 + \dots + 0A_{i-1} + e_i A_i + 0A_{i+1} + \dots + 0A_n$$

so that in the  $i$ th column of  $A$  we will find

$$\begin{array}{c} 0 \\ \vdots \\ 0 \\ e_i \quad \text{ith row.} \\ 0 \\ \vdots \\ 0 \end{array}$$

In fact, we have the following important result,

**Proposition 14.2.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional vector space  $\mathcal{V}$ . Then  $T$  is represented by a diagonal matrix using the same basis in domain and range iff  $\mathcal{V}$  has a basis composed of eigenvectors of  $T$ .*

**Definition.** An endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$  of the finite-dimensional vector space  $\mathcal{V}$  is said to be *diagonalizable* iff there exists a basis of  $\mathcal{V}$  such that  $T$  is represented by a diagonal matrix relative to this basis.

Note that in view of (14.2) an endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$  is diagonalizable iff  $\mathcal{V}$  has a basis composed of eigenvectors of  $T$ . Thus if  $\mathcal{V}$  is  $n$ -dimensional  $T$  is diagonalizable iff there exist  $n$  linearly independent eigenvectors of  $T$ . To *diagonalize*  $T$  means to find a basis of  $\mathcal{V}$  composed of eigenvectors of  $T$  and the matrix of  $T$  with respect to this basis.

**EXAMPLE 1.** Let  $P: \mathcal{V} \rightarrow \mathcal{V}$  be a projection, that is,  $P^2 = P$  (see Chapter 12<sup>bis</sup>). Then the only eigenvalues of  $P$  are 0 and 1.

**PROOF.** Suppose that  $e$  is an eigenvalue of  $P$  and  $E \in \mathcal{V}$  a corresponding eigenvector so that

$$P(E) = eE.$$

Then computing  $P^2(\mathbf{E})$  in two different ways gives

$$P^2(\mathbf{E}) = P(\mathbf{E}) = e\mathbf{E}$$

$$P^2(\mathbf{E}) = P(P(\mathbf{E})) = P(e\mathbf{E}) = eP(\mathbf{E}) = e(e\mathbf{E}) = e^2\mathbf{E}.$$

So equating yields

$$e\mathbf{E} = e^2\mathbf{E}$$

or

$$(e^2 - e)\mathbf{E} = \mathbf{0}.$$

Since the vector  $\mathbf{E} \neq \mathbf{0}$  this implies

$$0 = e^2 - e = e(e - 1).$$

So  $e = 0$  or  $e = 1$ .

In fact in (12<sup>bis</sup>.1) we even showed that a projection has the diagonal form

$$\begin{pmatrix} \overset{1}{\text{---}} & & & & \\ & \underset{1}{\text{---}} & & & \\ & & \ddots & & \\ & & & \underset{1}{\text{---}} & \\ & 0 & & & 0 \\ & & & & \underset{0}{\text{---}} & \\ & & & & & \ddots \\ & & & & & & \underset{0}{\text{---}} \end{pmatrix}$$

where  $r = \dim(\ker T)$  and  $s = \dim(\text{Im } P)$ .

EXAMPLE 2. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation given by

$$T(x, y) = (x + 2y, 3x + 2y).$$

Find all the eigenvalues and eigenvectors of  $T$ .

*Solution.* Note that relative to the standard basis of  $\mathbb{R}^2$   $T$  is represented by the matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$$

which is hardly diagonal.

To discover the eigenvalues and eigenvectors of  $T$  we must look for numbers  $e$  and nonzero vectors  $\mathbf{E}$  such that  $T(\mathbf{E}) = e\mathbf{E}$ . That is, if  $\mathbf{E} = (x, y)$ ,

$$T(x, y) = (ex, ey).$$

On the other hand

$$T(x, y) = (x + 2y, 3x + 2y)$$

by definition. So we must look at when the system of equations

$$\begin{aligned}x + 2y &= ex \\ 3x + 2y &= ey\end{aligned}$$

has a nonzero solution. That is, for what values of  $e$  does the homogeneous linear system

$$\begin{aligned}(1 - e)x + 2y &= 0 \\ 3x + (2 - e)y &= 0\end{aligned}$$

have a nontrivial solution? According to (13.3) this system has a nontrivial solution iff the matrix

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

is *not* invertible. (That is, the linear transformation represented by this matrix is by (12.4) *not* an isomorphism.) But (11.3) says that a  $2 \times 2$  matrix  $A$  is not an isomorphism iff  $\det A = 0$ . Putting all this together we find that  $e$  is an eigenvalue of  $T$  iff

$$\det \begin{pmatrix} 1 - e & 2 \\ 3 & 2 - e \end{pmatrix} = 0.$$

But

$$\begin{aligned}\det \begin{pmatrix} 1 - e & 2 \\ 3 & 2 - e \end{pmatrix} &= (1 - e)(2 - e) - 6 \\ &= 2 - 3e + e^2 - 6 \\ &= e^2 - 3e - 4 \\ &= (e - 4)(e + 1).\end{aligned}$$

Thus the eigenvalues of  $T$  are those numbers  $e$  such that

$$0 = (e - 4)(e + 1),$$

that is

$$e = 4, -1$$

are the eigenvalues of  $T$ . Now we must find the eigenvectors of  $T$  associated to 4 and  $-1$  respectively. That means we must solve the linear equations

$$\begin{aligned}e = 4: \quad & -3x + 2y = 0 \\ & 3x - 2y = 0, \\ e = -1: \quad & 2x + 2y = 0 \\ & 3x + 3y = 0.\end{aligned}$$

The first set has as solution space the linear span of the vector  $E = (2, 3)$  and the second set has as solution space the linear span of  $(1, -1) = F$ .

Thus we see the eigenvalues of  $T$  are 4 and  $-1$  and all eigenvectors associated to 4 are multiples of  $(2, 3)$  and all eigenvectors associated to  $-1$  are multiples of  $(1, -1)$ .

While the preceding example is long and tedious it is at least partly so because we really do not have enough tools yet to work such examples comfortably. In fact the example suggests a few useful results.

**Proposition 14.3.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional vector space  $\mathcal{V}$ . A number  $e$  is an eigenvalue of  $T$  iff the endomorphism*

$$T - eI: \mathcal{V} \rightarrow \mathcal{V}$$

*is not an isomorphism.*

**PROOF.** Suppose that  $e$  is an eigenvalue of  $T$ . Let  $E$  be an eigenvector of  $T$  associated to  $e$ . Then

$$T(E) = eE.$$

Therefore

$$(T - eI)(E) = T(E) - eE = eE - eE = 0$$

so  $0 \neq E \in \ker(T - eI)$ . Hence by (8.11)  $T - eI$  is not an isomorphism. Conversely, suppose that  $T - eI$  is not an isomorphism. By (8.11) it follows that either

$$\ker(T - eI) \neq \{0\}$$

or

$$\operatorname{Im}(T - eI) \neq \mathcal{V}.$$

From (8.10) we have the equation

$$n = \dim \ker(T - eI) + \dim \operatorname{Im}(T - eI).$$

Hence in either case  $\dim \ker(T - eI) > 0$ , that is, we must have

$$\ker(T - eI) \neq \{0\}$$

and

$$\operatorname{Im}(T - eI) \neq \mathcal{V}.$$

At any rate,  $\ker(T - eI) \neq \{0\}$  and so we may select  $E \in \ker(T - eI)$ ,  $E \neq 0$ . Then

$$0 = (T - eI)(E) = T(E) - e(E)$$

so

$$T(E) = eE$$

and hence  $E$  is an eigenvector of  $T$  associated to the eigenvalue  $e$ . □

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the vector space  $\mathcal{V}$ . Suppose that  $e$  is an eigenvalue of  $T$  and set

$$\mathcal{V}_e = \{E \mid T(E) = eE\}.$$

Thus  $\mathcal{V}_e$  is the set of all eigenvectors of  $T$  associated with the eigenvalue  $e$ , **together with the zero vector**.  $\mathcal{V}_e$  is called the *eigenspace* of  $T$  associated to the eigenvalue  $e$ .

Note that by definition an eigenspace  $\mathcal{V}_e$  always contains a nonzero vector, namely an eigenvector (at least one), and therefore  $\dim \mathcal{V}_e$  is always positive. Recall that  $\ker T = \{E \in \mathcal{V} \mid T(E) = 0\}$ , and so the number 0 is an eigenvalue of  $T: \mathcal{V} \rightarrow \mathcal{V}$  iff  $\ker T \neq \{0\}$  in which case  $\mathcal{V}_0 = \ker T$ . The following example shows that there are many nonzero endomorphisms for which 0 is the only eigenvalue.

**EXAMPLE 3.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be nilpotent. Then the only eigenvalue of  $T$  is 0.

**PROOF.** Since  $T$  is nilpotent there exists an integer  $k$  such that  $T^{k-1} \neq 0$ . Let  $k$  be the smallest such integer. Then  $T^{k-1} \neq 0$ , so there is a vector  $E \in \mathcal{V}$  with  $T^{k-1}(E) \neq 0$ . Since

$$T(T^{k-1}(E)) = T^k(E) = 0 = 0E$$

it follows that  $0 \neq T^{k-1}(E) \in \ker T$ , so 0 is an eigenvalue of  $T$ ,  $\mathcal{V}_0 = \ker T$ . By definition, if  $e$  is any eigenvalue of  $T$  then there is a nonzero vector  $F \in \mathcal{V}$  such that

$$T(F) = eF.$$

Then

$$T^2(F) = T(T(F)) = T(eF) = eT(F) = eeF = e^2F,$$

$$T^3(F) = T(T^2(F)) = e^2T(F) = e^3F,$$

and continuing in this way we see

$$T^k(F) = e^kF.$$

But  $T^k = 0$  so

$$T^k(F) = 0$$

and therefore

$$e^kF = 0.$$

Since  $F$  is not the zero vector this implies  $e^k = 0$  so  $e = 0$  as required.

The only diagonalizable transformation with 0 as its only eigenvalue is the zero transformation. Therefore the only nilpotent transformation that is diagonalizable is the zero transformation! In a very strong sense this is the

reason why not all transformations can be diagonalized. In fact if complex scalars are used then any transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  can be written as a sum  $D + N$  where  $D$  is diagonalizable with the same eigenvalues (and multiplicities) as  $T$  and  $N$  is nilpotent. (This is the Jordan canonical form and the subject of Chapter 17.)

**Proposition 14.4.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the vector space  $\mathcal{V}$ . If  $e$  is an eigenvalue of  $T$  then  $\mathcal{V}_e = \ker(T - eI)$ . Hence  $\mathcal{V}_e$  is a subspace of  $\mathcal{V}$ .*

**PROOF.** Immediate from the definitions. □

In view of these results it will be convenient to have a name for endomorphisms  $T: \mathcal{V} \rightarrow \mathcal{V}$  that are *not* isomorphisms.

**Definition.** An endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$  is said to be *singular* iff  $T$  is *not* an isomorphism.

Note that in the course of proving (14.3) we also proved (look at 8.10 again):

**Proposition 14.5.** *Let  $\mathcal{V}$  be a finite-dimensional vector space and  $T: \mathcal{V} \rightarrow \mathcal{V}$  an endomorphism. Then  $T$  is singular iff  $\ker T \neq \{0\}$ .*

**EXAMPLE 4.** Let  $S: \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  be the linear transformation given by

$$S(p(x)) = xp(x).$$

Then  $S$  is an endomorphism of  $\mathcal{P}(\mathbb{R})$ . Clearly

$$\ker S = \{0\}.$$

On the other hand  $S$  is singular since the vector 1 is not in  $\text{Im } S$ . Thus (14.5) becomes false if we do not assume  $\mathcal{V}$  to be finite dimensional.

**EXAMPLE 5.** Find the eigenvalues and associated eigenspaces for the linear transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  with matrix

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

relative to the standard basis.

**Solution.** The eigenvalues occur when the matrix

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



is singular. (*Important*: note that since we have insisted on using the same basis in domain and range the matrix of  $I$  is

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

You should go back and look at Examples 1 and 2 of Chapter 12 again.) Now

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

is singular iff its determinant vanishes. So we must have

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

so

$$\begin{aligned} 0 &= (1-e)(3-e) - 8 = 3 - 4e + e^2 - 8 \\ &= e^2 - 4e - 5 = (e-5)(e+1). \end{aligned}$$

So the eigenvalues of  $T$  must be  $e = 5$ ,  $e = -1$ . To find the eigenspace of  $T$  we must solve the equations

$$\begin{aligned} e = 5: & \quad \begin{cases} -4x + 2y = 0 \\ 4x - 2y = 0 \end{cases} \\ e = -1: & \quad \begin{cases} 2x + 2y = 0 \\ 4x + 4y = 0 \end{cases} \end{aligned}$$

so

$$\begin{aligned} \mathcal{V}_5 &= \mathcal{L}(1, 2) = \{(x, y) | 2x + y = 0\} \\ \mathcal{V}_{-1} &= \mathcal{L}(1, -1) = \{(x, x) | x + y = 0\}. \end{aligned}$$

**EXAMPLE 6.** Find the eigenvalues and eigenvectors of the linear transformation with matrix

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

relative to the standard basis of  $\mathbb{R}^2$ .

*Solution.* We must find when

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

But

$$\det \begin{pmatrix} -e & 1 \\ -1 & -e \end{pmatrix} = e^2 + 1$$

and hence is **never** zero. Therefore this linear transformation

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

has *no* eigenvalues.

It should be clear from the preceding examples that the problem of calculating the eigenvalues and eigenvectors of a linear transformation

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

for  $n > 2$  depends on having a simple method to test when  $T - eI$  is singular. Such a method exists, although for large  $n$  it is not really simple, as it is enormously tedious. It involves extending the determinant to square matrices of size greater than 2. We will do this by induction.

**Definition.** Let  $A$  be an  $n \times n$  matrix. The square matrix obtained from  $A$  by deleting the  $i$ th row and  $j$ th column of  $A$  is called the *minor* of the element  $a_{ij}$  of  $A$ , and is denoted by  $M_{ij}$ .

EXAMPLE 7. Let

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

find the minors  $M_{12}$ ,  $M_{23}$ ,  $M_{33}$  of  $A$ .

*Solution.* We have

$$M_{12} = \begin{pmatrix} 2 & 2 \\ 0 & 6 \end{pmatrix}$$

$$M_{23} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

$$M_{33} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

**Definition.** Let  $A$  be a square matrix of size  $n$ . Then the *determinant* of  $A$ , denoted by  $\det A$  is defined by the inductive formula

$$\det(a) = a$$

$$\det A = a_{11} \det M_{11} - a_{12} \det M_{12} + \cdots + (-1)^{n+1} a_{1n} \det M_{1n}$$

$$= \sum_{j=1}^n (-1)^{1+j} a_{1j} \det M_{1j}.$$

EXAMPLE 8. Calculate the determinant of

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

*Solution.* We have

$$\begin{aligned} \det A &= 1 \det \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} - 0 \det \begin{pmatrix} 2 & 2 \\ 0 & 6 \end{pmatrix} + 1 \det \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix} \\ &= 6 - 8 + 0 + 8 = 6. \end{aligned}$$

EXAMPLE 9. Calculate the determinant of

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

*Solution.* We have

$$\begin{aligned} \det A &= 1 \det \begin{pmatrix} 1 & 4 & 2 \\ 1 & 0 & 4 \\ 0 & 2 & 1 \end{pmatrix} - 2 \det \begin{pmatrix} 0 & 4 & 2 \\ 0 & 0 & 4 \\ 1 & 2 & 1 \end{pmatrix} \\ &\quad - 1 \det \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \\ 1 & 0 & 1 \end{pmatrix} - 3 \det \begin{pmatrix} 0 & 1 & 4 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \\ &= 1 \left( \det \begin{pmatrix} 0 & 4 \\ 2 & 1 \end{pmatrix} - 4 \det \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} + 2 \det \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right) \\ &\quad - 2 \left( 0 \det \begin{pmatrix} 0 & 4 \\ 2 & 1 \end{pmatrix} - 4 \det \begin{pmatrix} 0 & 4 \\ 1 & 1 \end{pmatrix} + 2 \det \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix} \right) \\ &\quad - \left( 0 \det \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} - 1 \det \begin{pmatrix} 0 & 4 \\ 1 & 1 \end{pmatrix} + 2 \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \\ &\quad - 3 \left( 0 \det \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} - \det \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix} + 4 \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \\ &= (-8 - 4 + 4) - 2(0 - 4(-4) + 0) - (0 + 4 - 2) - 3(0 - 0 - 4) \\ &= -30. \end{aligned}$$

Hopefully the preceding example will convince you that the evaluation of determinants of large matrices is a most painful process. However for small values of  $n$  the utility of determinants cannot be denied.

There are a number of properties of determinants that make easier their computation. The verification of these properties requires an excursion into multilinear algebra which is undertaken in Chapter 14<sup>bis</sup>. The basic properties of determinants are summarized in properties (1)–(8) below, of which we make free use in the sequel.

**Definition.** The *transpose* of a matrix  $A = (a_{ij})$  is the matrix  $A^t = (a_{ij}^t)$  where  $a_{ij}^t = a_{ji}$ .

For example when  $n = 4$

$$\begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix}^t = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix}.$$

The transpose of  $A$  is the matrix  $A^t$  whose rows are the columns of  $A$  (see Chapter 11 Exercises 6–10).

### Properties of determinants

Let  $A$  be a matrix.

- (1) If  $A$  has a row of zeros then  $\det A = 0$ .
- (2) If  $A$  has two rows equal then  $\det A = 0$ .
- (3) If  $B$  is obtained from  $A$  by interchanging two rows then  $\det B = -\det A$ .
- (4) If  $B$  is obtained from  $A$  by adding a multiple of one row to a *different* row then  $\det B = \det A$ .
- (5) If  $B$  is obtained from  $A$  by multiplying a row of  $A$  by a number  $k$  then  $\det B = k \det A$ .
- (6)  $\det A^t = \det A$ , hence everything we said about rows in (1)–(5) works also for columns.
- (7) If  $A$  is upper triangular  $\det A$  is the product of the diagonal entries.
- (8)  $\det(AB) = \det A \det B$ .

Using these properties we may do Example 9 a little less painfully.

**EXAMPLE 9 REVISITED.** We have

$$\begin{aligned} & \det \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & 4 & 2 \\ 0 & 1 & 0 & 4 \\ 1 & 0 & 2 & 1 \end{pmatrix} \\ &= \det \begin{pmatrix} 0 & 2 & -3 & 2 \\ 0 & 1 & 4 & 2 \\ 0 & 1 & 0 & 4 \\ 1 & 0 & 2 & 1 \end{pmatrix} = \det \begin{pmatrix} 0 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ -3 & 4 & 0 & 2 \\ 2 & 2 & 4 & 1 \end{pmatrix} \\ &= (-1)^{1+4} \det \begin{pmatrix} 2 & 1 & 1 \\ -3 & 4 & 0 \\ 2 & 2 & 4 \end{pmatrix} = -\det \begin{pmatrix} 0 & 0 & 1 \\ -3 & 4 & 0 \\ -6 & -2 & 4 \end{pmatrix} \\ &= -(-1)^{1+3} \det \begin{pmatrix} -3 & 4 \\ -6 & -2 \end{pmatrix} = -(6 + 24) = -30. \end{aligned}$$

**Proposition 14.6.** Let  $A$  be an  $n \times n$  matrix. Then  $A$  is invertible iff  $\det A \neq 0$ .

**PROOF.** Suppose that  $A$  is invertible. Then

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

Taking determinants of both sides gives

$$\det(A(A^{-1})) = \det I.$$

Since  $I$  is upper triangular it follows from (7) that

$$\det I = 1.$$

From (8) we find

$$\det(A(A^{-1})) = \det A \det(A^{-1}).$$

Equating gives

$$\det A \det(A^{-1}) = 1.$$

Hence  $\det A \neq 0$  and in fact  $\det(A^{-1}) = (\det A)^{-1}$ .

To prove the converse we will actually give a method for constructing  $A^{-1}$ . Let us introduce the numbers

$$A_{ij} = (-1)^{i+j} \det \mathbf{M}_{ij}.$$

The number  $A_{ij}$  is called the *cofactor* of  $a_{ij}$ . In this notation we have

**Lemma 14.7.**  $\det A = \sum_j a_{ij} A_{ij}$  for any  $i = 1, 2, \dots$

**PROOF.** Let  $B$  be the matrix obtained from  $A$  by interchanging the first and  $i$ th rows. Then

$$\det A = -\det B = -\sum_j b_{1j} B_{1j}.$$

Next note that the minors  $\mathbf{M}_{ij}(B)$  of the  $i$ th row of  $B$  differ from the minors  $\mathbf{M}_{1j}(A)$  of the first row of  $A$  in that the first row of  $\mathbf{M}_{ij}(B)$  is the  $(i-1)$ th row of  $\mathbf{M}_{1j}(A)$ . By interchanging this row with the one below it in  $\mathbf{M}_{1j}(B)$   $i-2$  times we can put it in the  $(i-1)$ th row. Therefore  $B_{1j} = -A_{ij}$  and hence  $\det A = -\det B = -\sum_j b_{1j} B_{1j} = \sum_j a_{ij} A_{ij}$ .  $\square$

*Remark.* Note in view of the fact that  $\det A = \det A^t$  it follows from this lemma that  $\det A$  may be computed from the minors of any column also.

**Lemma 14.8.**  $\sum a_{ij} A_{kj} = 0$  if  $k \neq i$ .

**PROOF.** Let  $B$  be the matrix obtained from  $A$  by replacing the  $k$ th row of  $A$  by the  $i$ th row. Then  $B$  has two rows equal, so  $\det B = 0$ . On the other hand

$$B_{kj} = A_{kj}$$

and using (14.6) to evaluate  $\det B$  by minors of the  $k$ th row we find

$$0 = \det B = \sum b_{kj} B_{kj} = \sum a_{ij} A_{kj}$$

as required.  $\square$

Let us define now a matrix  $A^{\text{cof}} = (a_{ij}^*)$  called the *cofactor matrix* of  $A$  by  $a_{ij}^* = A_{ji}$ . (Note carefully the switch in index.) We then have

**Lemma 14.9.**  $AA^{\text{cof}} = (\det A)I$ .

**PROOF.** Let  $AA^{\text{cof}} = C = (c_{ij})$ . Then

$$c_{ik} = \sum_j a_{ij}a_{jk}^* = \sum_j a_{ij}A_{kj}$$

so by (14.7)

$$c_{ik} = \begin{cases} 0 & i \neq k \\ \det A & i = k \end{cases}$$

so that  $C = (\det A)I$  as required.  $\square$

Returning now to the proof of (14.6) we suppose that  $\det A \neq 0$ . Let

$$B = \frac{1}{\det A} A^{\text{cof}}.$$

Then

$$AB = \frac{1}{\det A} AA^{\text{cof}} = \frac{1}{\det A} (\det A)I = I$$

as required.  $\square$

**EXAMPLE 10.** Calculate the inverse of the matrix of Example 8.

*Solution.* We must calculate the cofactors of  $A$ . We find

$$A_{11} = (-1)^{1+1} \det \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} = 6 - 8 = -2$$

$$A_{12} = (-1)^{1+2} \det \begin{pmatrix} 2 & 2 \\ 0 & 6 \end{pmatrix} = -12$$

$$A_{13} = (-1)^{1+3} \det \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix} = 8$$

$$A_{21} = (-1)^{2+1} \det \begin{pmatrix} 0 & 1 \\ 4 & 6 \end{pmatrix} = -(0 - 4) = 4$$

$$A_{22} = (-1)^{2+2} \det \begin{pmatrix} 1 & 1 \\ 0 & 6 \end{pmatrix} = 6$$

$$A_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} = -4$$

$$A_{31} = (-1)^{3+1} \det \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} = -1$$

$$A_{32} = (-1)^{3+2} \det \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = 0$$

$$A_{33} = (-1)^{3+3} \det \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} = 1.$$

Hence

$$A^{\text{cof}} = \begin{pmatrix} 2 & 4 & -1 \\ -12 & 6 & 0 \\ 8 & -4 & 1 \end{pmatrix}$$

and therefore

$$A^{-1} = \frac{1}{6} \begin{pmatrix} -2 & 4 & -1 \\ -12 & 6 & 0 \\ 8 & -4 & 1 \end{pmatrix}.$$

**EXAMPLE 11.** (Compare Chapter 6 Example 17.) Determine when the vectors  $(r, 1, 1)$ ,  $(1, r, 1)$ ,  $(1, 1, r)$  are a basis for  $\mathbb{R}^3$ .

*Solution.* Consider the linear transformation

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

that is the linear extension of

$$T(1, 0, 0) = (r, 1, 1)$$

$$T(0, 1, 0) = (1, r, 1)$$

$$T(0, 0, 1) = (1, 1, r).$$

Then  $T$  is an isomorphism iff  $\{(r, 1, 1), (1, r, 1), (1, 1, r)\}$  is a basis for  $\mathbb{R}^3$ . The matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$  is

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

and so  $T$  is invertible iff

$$\begin{aligned} 0 \neq \det \begin{pmatrix} r & 1 & 1 \\ 1 & r & 1 \\ 1 & 1 & r \end{pmatrix} &= r \det \begin{pmatrix} r & 1 \\ 1 & r \end{pmatrix} + (-1) \det \begin{pmatrix} 1 & 1 \\ 1 & r \end{pmatrix} + 1 \det \begin{pmatrix} 1 & r \\ 1 & 1 \end{pmatrix} \\ &= r(r^2 - 1) - (r - 1) + (1 - r) \\ &= (r - 1)[r(r + 1) - 2] = (r - 1)(r^2 + r - 2) \\ &= (r - 1)^2(r + 2). \end{aligned}$$

Therefore  $T$  is invertible iff  $r \neq 1, -2$ , so  $\{(r, 1, 1), (1, r, 1), (1, 1, r)\}$  is a basis for  $\mathbb{R}^3$  iff  $r \neq 1, -2$ .

As a method for computing the above procedure is quite systematic but fraught with the perils of arithmetic error.

Let us return now to the problem of calculating the eigenvalues of an endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$  of the finite-dimensional vector space  $\mathcal{V}$ . Choose a basis  $\{A_1, \dots, A_n\}$  for  $\mathcal{V}$  and let  $T$  be represented by the matrix  $A = (a_{ij})$  in this basis.

**Definition.** The *characteristic polynomial* of  $T$  is the polynomial of degree  $n$ ,  $\Delta(t) = \det(A - tI)$ , where  $t$  is a variable.

*Remark.* It may seem as though the characteristic polynomial of  $T$  depends on the choice of basis  $\{A_1, \dots, A_n\}$ . To see that this is not so suppose that  $\{B_1, \dots, B_n\}$  is another basis for  $\mathcal{V}$  and that the matrix of  $T$  is  $B$  relative to this basis. Then by (12.6)

$$B = PA(P^{-1})$$

(since we are using the same bases in domain and range). Thus

$$\begin{aligned} B - tI &= PAP^{-1} - tI = PAP^{-1} - tPIP^{-1} \\ &= P(A - tI)P^{-1}. \end{aligned}$$

Therefore

$$\begin{aligned} \det(B - tI) &= \det(P(A - tI)P^{-1}) \\ &= (\det P)(\det(A - tI))(\det P^{-1}) \\ &= (\det P)(\det P^{-1})(\det(A - tI)) \\ &= (\det P)(\det P)^{-1}(\det(A - tI)) \\ &= \det(A - tI). \end{aligned}$$

Therefore it does not matter what basis we use to represent  $T$  by a matrix to compute its characteristic polynomial.

By combining (14.3) and (14.6) we obtain the important:

**Theorem 14.10** (Criteria for Eigenvalues.) *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism with characteristic polynomial  $\Delta(t)$ . Then a number  $e$  is an eigenvalue of  $T$  iff  $e$  is a root of  $\Delta(t)$ .*

**PROOF.** Choose a basis for  $\mathcal{V}$ ,  $\{A_1, \dots, A_n\}$  and let  $A = (a_{ij})$  be the matrix of  $T$  relative to this basis.

Suppose that  $e$  is an eigenvalue of  $T$ . Then by (14.3) the transformation

$$T - eI: \mathcal{V} \rightarrow \mathcal{V}$$

is singular. Its matrix relative to  $\{A_1, \dots, A_n\}$  is the matrix  $A - eI$ . Since  $T - eI$  is not an isomorphism the matrix  $A - eI$  has no inverse, (12.4). So,

$$\det(A - eI) = 0$$

which says that the number  $e$  is a root of the polynomial in  $t$

$$\Delta(t) = \det(A - tI).$$



To prove the converse suppose that  $e$  is a root of  $\Delta(t)$ . Then

$$0 = \Delta(e) = \det(\mathbf{A} - e\mathbf{I}).$$

So by (14.5)  $\mathbf{A} - e\mathbf{I}$  is not invertible and hence

$$\mathbf{T} - e\mathbf{I}: \mathcal{V} \rightarrow \mathcal{V}$$

is singular, so again by (14.3)  $e$  is an eigenvalue of  $\mathbf{T}$ . □

**EXAMPLE 12.** Let  $\mathbf{T}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by the formula

$$\mathbf{T}(x, y, z) = (0, x, y).$$

Find the characteristic polynomial of  $\mathbf{T}$ ,  $\mathbf{T}^2$ , and  $\mathbf{T}^3$ .

*Solution.* The matrix of  $\mathbf{T}$  relative to the standard basis is

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

so those of  $\mathbf{T}^2$  and  $\mathbf{T}^3$  are

$$\mathbf{A}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{A}^3 = \mathbf{0}.$$

Therefore, (since we are dealing with more than one endomorphism we will employ subscripts to indicate the transformation's characteristic polynomial),

$$\begin{aligned} \Delta_{\mathbf{T}}(t) &= \det(\mathbf{A} - t\mathbf{I}) \\ &= \det \begin{pmatrix} -t & 0 & 0 \\ 1 & -t & 0 \\ 0 & 1 & -t \end{pmatrix} = -t^3, \end{aligned}$$

$$\begin{aligned} \Delta_{\mathbf{T}^2}(t) &= \det(\mathbf{A}^2 - t\mathbf{I}) \\ &= \det \begin{pmatrix} -t & 0 & 0 \\ 0 & -t & 0 \\ 1 & 0 & -t \end{pmatrix} = -t^3, \end{aligned}$$

$$\begin{aligned} \Delta_{\mathbf{T}^3}(t) &= \det(\mathbf{A}^3 - t\mathbf{I}) \\ &= \det \begin{pmatrix} -t & 0 & 0 \\ 0 & -t & 0 \\ 0 & 0 & -t \end{pmatrix} = -t^3. \end{aligned}$$

Thus  $\mathbf{T}$ ,  $\mathbf{T}^2$  and  $\mathbf{T}^3 = \mathbf{0}$  have the same characteristic polynomial.

EXAMPLE 13. Find the characteristic polynomial and the eigenvalues of the linear transformation  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  given by

$$T(x, y, z) = (x - 2z, 0, -2x + 4z).$$

*Solution.* The matrix of  $T$  relative to the standard basis is

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

and thus the characteristic polynomial of  $T$  is

$$\begin{aligned} \det(A - tI) &= \det \begin{pmatrix} 1-t & 0 & -2 \\ 0 & -t & 0 \\ -2 & 0 & 4-t \end{pmatrix} \\ &= (1-t)(-t)(4-t) - (-2)(-t)(-2) = -t^3 + 5t^2. \end{aligned}$$

So the characteristic polynomial is  $\Delta(t) = -t^3 + 5t^2$ . The eigenvalues are  $t = 0, t = 0, t = 5$ . The corresponding eigenspaces are obtained by solving the systems of equations

$$\begin{aligned} t = 0 & \quad \begin{cases} x - 2z = 0 \\ -2x + 4z = 0 \end{cases} \\ t = 5 & \quad \begin{cases} -4x - 2z = 0 \\ -5y = 0 \\ -2x - z = 0. \end{cases} \end{aligned}$$

We find

$$\begin{aligned} \dim \mathcal{V}_0 &= 2 \quad \text{spanned by } (0, 1, 0), (2, 0, 1) \\ \dim \mathcal{V}_5 &= 1 \quad \text{spanned by } (1, 0, -2). \end{aligned}$$

Thus  $\mathbb{R}^3$  has the basis

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

composed of eigenvectors of  $T$ . Since

$$\begin{aligned} T(\mathbf{E}_1) &= 5\mathbf{E}_1 \\ T(\mathbf{E}_2) &= \mathbf{0} \\ T(\mathbf{E}_3) &= \mathbf{0} \end{aligned}$$

the matrix of  $T$  relative to this basis is

$$B = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that if we were to use the matrix **B** to compute the characteristic polynomial of **T** we would get

$$\begin{aligned}\Delta(t) &= \det(\mathbf{B} - t\mathbf{I}) = \det \begin{pmatrix} 5-t & 0 & 0 \\ 0 & -t & 0 \\ 0 & 0 & -t \end{pmatrix} \\ &= (5-t)(-t)(-t) = -t^3 + 5t^2\end{aligned}$$

as before.

The procedure we have gone through is called **diagonalizing** the linear transformation **T**. *Warning:* Not every linear transformation can be diagonalized! Go back and stare at Example 6.

EXAMPLE 14. Diagonalize, if possible, the linear transformation:

$$\mathbf{T}: \mathbb{R}^4 \rightarrow \mathbb{R}^4$$

given by

$$\mathbf{T}(x, y, z, w) = (x, 2x + 5y + 6z + 7w, 3x + 8z + 9w, 4x + 10w)$$

*Solution.* The matrix of **T** is computed first. We have

$$\begin{aligned}\mathbf{T}(1, 0, 0, 0) &= (1, 2, 3, 4) \\ \mathbf{T}(0, 1, 0, 0) &= (0, 5, 0, 0) \\ \mathbf{T}(0, 0, 1, 0) &= (0, 6, 8, 0) \\ \mathbf{T}(0, 0, 0, 1) &= (0, 7, 9, 10).\end{aligned}$$

So the matrix of **T** relative to the standard base is

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 5 & 6 & 7 \\ 3 & 0 & 8 & 9 \\ 4 & 0 & 0 & 10 \end{pmatrix}.$$

The characteristic polynomial of **T** is thus

$$\begin{aligned}\Delta(t) &= \det(\mathbf{A} - t\mathbf{I}) = \det \begin{pmatrix} 1-t & 0 & 0 & 0 \\ 2 & 5-t & 6 & 7 \\ 3 & 0 & 8-t & 9 \\ 4 & 0 & 0 & 10-t \end{pmatrix} \\ &= (1-t) \det \begin{pmatrix} 5-t & 6 & 7 \\ 0 & 8-t & 9 \\ 0 & 0 & 10-t \end{pmatrix}\end{aligned}$$

(by property 7 of determinants)

$$= (1-t)(5-t)(8-t)(10-t).$$

Thus the eigenvalues of  $T$  are

$$t = 1, 5, 8, 10.$$

The eigenspaces are computed by solving the systems

$$t = 1 \quad \begin{cases} 2x + 4y + 6z + 7w = 0 \\ 3x \quad \quad + 7z + 9w = 0 \\ 4x \quad \quad \quad + 9w = 0 \end{cases}$$

$$-4x = 9w, \quad 3x + 7z - 4x = 0, \quad 7z = x$$

$$2x + 4y + \frac{6}{7}x - \frac{28}{9}x = 0 \quad 4y = (\frac{28}{9} - \frac{6}{7} - 2)x$$

$$w = -\frac{4}{9}x \quad 4y = \frac{16}{63}x$$

$$z = \frac{1}{7}x \quad y = \frac{4}{63}x.$$

So

$$\mathcal{V}_1 = \mathcal{L}(63, 4, 9, -28).$$

$$t = 5 \quad \begin{cases} -4x \quad \quad \quad = 0 \\ 2x + 6z + 7w = 0 \\ 3x + 3z + 9w = 0 \\ 4x \quad \quad \quad 5w = 0 \end{cases}$$

$$x = 0, \quad z = 0, \quad w = 0, \quad y = \text{anything}.$$

So

$$\mathcal{V}_5 = \mathcal{L}(0, 1, 0, 0).$$

Next we have

$$t = 8 \quad \begin{cases} -7x \quad \quad \quad = 0 \\ 2x - 3y + 6z + 7w = 0 \\ 3x \quad \quad \quad + 9w = 0 \\ 4x \quad \quad \quad + 2w = 0 \end{cases}$$

$$x = 0, \quad w = 0, \quad y = 2z.$$

Therefore

$$\mathcal{V}_8 = \mathcal{L}(0, 2, 1, 0).$$

Finally

$$t = 10 \quad \begin{cases} -9x \quad \quad \quad = 0 \\ 2x - 5y + 6z + 7w = 0 \\ 3x \quad \quad - 2z + 9w = 0 \\ 4x \quad \quad \quad = 0 \end{cases}$$

so

$$x = 0$$

$$z = \frac{9}{2}w$$

$$5y = 7w + 6z = 7w + \frac{6(9)}{2}w = 34w, \quad y = \frac{34}{5}.$$

Thus

$$\mathcal{V}_{10} = \mathcal{L}(0, 68, 45, 10).$$

The vectors

$$\mathbf{E}_1 = (63, 4, 9, -28)$$

$$\mathbf{E}_2 = (0, 1, 0, 0)$$

$$\mathbf{E}_3 = (0, 2, 1, 0)$$

$$\mathbf{E}_4 = (0, 68, 45, 10)$$

may be shown to be linearly independent. (See, for example, Proposition 14.11 below.) Thus  $\mathbb{R}^4$  has as a basis the eigenvectors  $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4\}$  of  $T$ . The matrix of  $T$  relative to these vectors is

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 10 \end{pmatrix}.$$

**Proposition 14.11.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of  $\mathcal{V}$ . Suppose  $e_1, \dots, e_m$  are distinct eigenvalues of  $T$  and  $\mathbf{E}_1, \dots, \mathbf{E}_m$  are corresponding eigenvectors. Then  $\mathbf{E}_1, \dots, \mathbf{E}_m$  are linearly independent.*

**PROOF.** Suppose to the contrary that  $\mathbf{E}_1, \dots, \mathbf{E}_m$  are linearly dependent. Applying Proposition 6.2 we may find an integer  $k$  such that  $\mathbf{E}_k$  is linearly dependent on  $\mathbf{E}_1, \dots, \mathbf{E}_{k-1}$ . We may also suppose that  $k$  is the smallest integer with this property and so by (6.2) again we may suppose  $\mathbf{E}_1, \dots, \mathbf{E}_{k-1}$  are linearly independent. Now let

$$(A) \quad \mathbf{E}_k = a_1 \mathbf{E}_1 + a_2 \mathbf{E}_2 + \cdots + a_{k-1} \mathbf{E}_{k-1}.$$

Apply  $T$  to this equation and get

$$(B) \quad T(\mathbf{E}_k) = a_1 T(\mathbf{E}_1) + a_2 T(\mathbf{E}_2) + \cdots + a_{k-1} T(\mathbf{E}_{k-1}).$$

Since the  $\mathbf{E}_1, \dots, \mathbf{E}_k$  are eigenvectors of  $T$  associated to the eigenvalues  $e_1, \dots, e_k$  this equation yields

$$(C) \quad e_k \mathbf{E}_k = e_1 a_1 \mathbf{E}_1 + e_2 a_2 \mathbf{E}_2 + \cdots + e_{k-1} a_{k-1} \mathbf{E}_{k-1}.$$

Now multiply Equation (A) by  $e_k$  to obtain

$$(D) \quad e_k \mathbf{E}_k = e_k a_1 \mathbf{E}_1 + e_k a_2 \mathbf{E}_2 + \cdots + e_k a_{k-1} \mathbf{E}_{k-1}.$$

and subtract Equation (D) from Equation (C) giving

$$(E) \quad \mathbf{0} = a_1(e_1 - e_k)\mathbf{E}_1 + a_2(e_2 - e_k)\mathbf{E}_2 + \cdots + a_{k-1}(e_{k-1} - e_k)\mathbf{E}_{k-1}.$$

Since the vectors  $\mathbf{E}_1, \dots, \mathbf{E}_{k-1}$  are linearly independent it follows that

$$(F) \quad \begin{aligned} a_1(e_1 - e_k) &= 0 \\ a_2(e_2 - e_k) &= 0 \\ &\vdots \\ a_{k-1}(e_{k-1} - e_k) &= 0. \end{aligned}$$

Next we recall that the eigenvalues  $e_1, \dots, e_k$  are distinct, so that

$$(G) \quad \begin{aligned} (e_1 - e_k) &\neq 0 \\ &\vdots \\ (e_{k-1} - e_k) &\neq 0 \end{aligned}$$

so combining (F) and (G) we find

$$(H) \quad \begin{aligned} a_1 &= 0 \\ &\vdots \\ a_{k-1} &= 0. \end{aligned}$$

Going back to Equation (A) this says

$$(I) \quad \mathbf{E}_k = \mathbf{0}.$$

But an eigenvector is by definition **nonzero**, and hence our assumption that  $\{\mathbf{E}_1, \dots, \mathbf{E}_m\}$  is linearly dependent has led to a contradiction. Therefore  $\{\mathbf{E}_1, \dots, \mathbf{E}_m\}$  is linearly independent.  $\square$

If you think that the proof of (14.11) is hard compare it with the pain of proving directly that the vectors  $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$  of Example 14 are independent.

Since the characteristic polynomial of an endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$  of a vector space  $\mathcal{V}$  of dimension  $n$  is of degree  $n$  we find the following useful diagonalization result.

**Corollary 14.12.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the  $n$ -dimensional vector space  $\mathcal{V}$ . Suppose that the characteristic polynomial of  $T$  has  $n$  distinct real roots. Then there is a basis of  $\mathcal{V}$  composed of eigenvectors of  $T$ . Relative to this basis the matrix of  $T$  is diagonal.*

**PROOF.** Let  $e_1, \dots, e_n$  be the eigenvalues of  $T$ . There are  $n$  of them because the characteristic polynomial has degree  $n$ . By assumption they are distinct. Choose eigenvectors  $\mathbf{E}_1, \dots, \mathbf{E}_n$  corresponding to  $e_1, \dots, e_n$ . By (14.11)

the vectors  $\mathbf{E}_1, \dots, \mathbf{E}_n$  are linearly independent. Therefore by (6.9)  $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$  is a basis for  $\mathcal{V}$ .  $\square$

It is quite natural to inquire what the situation is when the characteristic polynomial does not factor into distinct linear factors. In general this can mean a quite complicated structure for the endomorphism  $T: \mathcal{V} \rightarrow \mathcal{V}$ . There are however a few general results we can deduce. We begin with a definition.

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional vector space  $\mathcal{V}$  and  $e$  an eigenvalue of  $T$ . The *geometric multiplicity* of  $e$  is the dimension of the eigenspace  $\mathcal{V}_e$ . The *algebraic multiplicity* of  $e$  is the multiplicity of the root  $e$  of the characteristic polynomial  $\Delta(t)$  of  $T$ .

**Proposition 14.13.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional vector space  $\mathcal{V}$  and  $e$  an eigenvalue of  $T$ . Then the geometric multiplicity of  $e$  is less than or equal to the algebraic multiplicity of  $e$ .

**PROOF.** Recall from the remark preceding Theorem 14.10 that we may calculate the characteristic polynomial of  $T$  by representing  $T$  as a matrix relative to any basis of  $\mathcal{V}$ . Let us now cleverly choose such a basis for  $\mathcal{V}$ . Let us suppose  $\{\mathbf{A}_1, \dots, \mathbf{A}_r\}$  a basis for  $\mathcal{V}_e$ , so that the geometric multiplicity of  $e$  is equal to  $r$ . Let  $n$  be the dimension of  $\mathcal{V}$  and extend  $\{\mathbf{A}_1, \dots, \mathbf{A}_r\}$  to a basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_r, \mathbf{A}_{r+1}, \dots, \mathbf{A}_n\}$  of  $\mathcal{V}$ . The matrix of  $T$  relative to this basis is easily seen to be

$$A = \begin{pmatrix} e & & & \mathbf{B} \\ & \ddots & & \\ & & e & \\ \mathbf{0} & & & \mathbf{C} \end{pmatrix}$$

where  $\mathbf{B}$  is an  $r \times n - r$  matrix,  $\mathbf{C}$  is an  $n - r \times n - r$  matrix and  $\mathbf{0}$  an  $n - r \times r$  matrix of zeros.

The characteristic polynomial of  $T$  is therefore

$$\Delta(t) = \det(A - tI) = \det \begin{pmatrix} e - t & & & \mathbf{B} \\ & e - t & & \\ & & \ddots & \\ & & & e - t \\ \mathbf{0} & & & \mathbf{C} - tI \end{pmatrix}$$

and since  $e - t$  is a factor of each of the first  $r$  columns we find

$$\Delta(t) = (e - t)^r \det \begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} - tI \end{pmatrix} = (e - t)^r q(t)$$

and therefore  $e$  is a root of  $\Delta(t) = 0$  of multiplicity at least  $r$  which shows that the algebraic multiplicity of  $e$  is at least  $r$  and the result follows.  $\square$

**Proposition 14.14.** *If  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a diagonalizable linear transformation then*

$$\Delta(t) = (t - e_1)^{m_1} \dots (t - e_k)^{m_k}$$

and

$$\dim \mathcal{V}_{e_i} = m_i \quad i = 1, 2, \dots, k$$

where  $e_1, \dots, e_k$  are the distinct eigenvalues of  $T$  and  $m_1, \dots, m_k$  their multiplicities.

**PROOF.** Since  $T$  is diagonalizable it has a matrix representation

$$A = \begin{pmatrix} \begin{matrix} e_1 & & & \\ & \ddots & & \\ & & e_1 & \\ & & & \ddots & \\ & & & & e_2 & \\ & & & & & \ddots \\ & & & & & & e_2 \\ & & & & & & & \ddots \\ & & & & & & & & e_k & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & e_k \end{matrix} & & 0 \\ 0 & & \end{pmatrix}$$

where  $n_i = \dim \mathcal{V}_{e_i}$  for  $i = 1, \dots, k$ . Using this matrix representation to compute the characteristic polynomial of  $T$  we get

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} \begin{matrix} e_1 - t & & & \\ & \ddots & & \\ & & e_1 - t & \\ & & & \ddots & \\ & & & & e_k - t & \\ & & & & & \ddots \\ & & & & & & e_k - t \end{matrix} & & 0 \\ 0 & & \end{pmatrix} \\ &= (e_1 - t)^{n_1} \dots (e_k - t)^{n_k} \\ &= (-1)^n [t - e_1]^{n_1} \dots [t - e_k]^{n_k} \end{aligned}$$

and therefore  $n_1 = m_1, \dots, n_k = m_k$  as required.  $\square$



We may summarize all this in the following:

**Theorem 14.15.** *A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  of the finite-dimensional vector space  $\mathcal{V}$  is diagonalizable iff the following two conditions hold:*

- (1) *The characteristic polynomial  $\Delta(t)$  is a product of linear factors.*
- (2) *For each eigenvalues of  $T$  the geometric and algebraic multiplicities are the same.*

At this point it is worthwhile to pause to reflect on the difference between the case of real or complex scalars. If we look at  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  given by

$$T(x, y) = (y, -x)$$

which is a rotation through  $90^\circ$  we find that the matrix of  $T$  relative to the usual basis of  $\mathbb{R}^2$  is

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

so that

$$\Delta(t) = \det \begin{pmatrix} t & 1 \\ -1 & t \end{pmatrix} = t^2 + 1,$$

and of course as we all know

$$t^2 + 1 = 0$$

has no real roots. So this simple transformation of  $\mathbb{R}^2$  has no real eigenvalues. On the other hand it does have two distinct complex roots, namely  $\pm i = \pm \sqrt{-1} \in \mathbb{C}$ . This means that the transformation

$$S: \mathbb{C}^2 \rightarrow \mathbb{C}^2$$

given by

$$S(u, v) = (v, -u)$$

which has characteristic polynomial

$$\Delta(t) = t^2 + 1$$

can be diagonalized, while the transformation  $T$  with the same (matrix and) characteristic polynomial does not have even a single eigenvalue. In fact an endomorphism  $S: \mathcal{V} \rightarrow \mathcal{V}$  of a (finite-dimensional) vector space over the complex number field  $\mathbb{C}$  must *always* have an eigenvalue. Thus it has at least one eigenvector. To see this recall the following:

**Fundamental Theorem of Algebra:** *Suppose that  $p(t) = a_n t^n + \dots + a_1 t + a_0$  is a polynomial of degree  $n$  (that is  $a_n \neq 0$ ) with complex coefficients. Then there are complex numbers  $r_1, \dots, r_n$  (unique up to change of order) such that  $p(t) = a_n(t - r_1)(t - r_2) \cdots (t - r_n)$ .*

Therefore the first condition of (14.15) for a linear transformation to be diagonalizable is always satisfied in the complex case.

**EXAMPLE 15.** Find the characteristic polynomial, eigenvalues, and eigenvectors of the endomorphism

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$$

given by

$$T(x, y, z) = (2x + y, y - z, 2y + 4z).$$

*Solution.* We begin by calculating the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$ . We find

$$T(1, 0, 0) = (2, 0, 0)$$

$$T(0, 1, 0) = (1, 1, 2)$$

$$T(0, 0, 1) = (0, -1, 4)$$

and so the matrix we seek is

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

The characteristic polynomial of  $T$  is therefore

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} 2-t & 1 & 0 \\ 0 & 1-t & -1 \\ 0 & 2 & 4-t \end{pmatrix} \\ &= (2-t) \det \begin{pmatrix} 1-t & -1 \\ 2 & 4-t \end{pmatrix} \\ &= (2-t)[(1-t)(4-t) + 2] = (2-t)[4 - 5t + t^2 + 2] \\ &= (2-t)[t^2 - 5t + 6] = (2-t)(t-2)(t-3) \\ &= -(t-2)^2(t-3). \end{aligned}$$

So the eigenvalues of  $T$  are 2, 3 with 2 having algebraic multiplicity 2.

To calculate the eigenspaces we must solve the linear systems below.

$$e = 2 \quad \begin{cases} 0x + y + 0z = 0 \\ 0 - y - z = 0 \\ 0 + 2y + 2z = 0 \end{cases}$$

which yields

$$y = 0 \quad z = 0$$

and therefore  $\mathcal{V}_2 = \mathcal{L}(1, 0, 0)$ . That is the eigenspace associated to the eigenvalue 2 is 1-dimensional and is spanned by the vector  $(1, 0, 0)$ .

$$e = 3 \quad \begin{cases} -x + y + 0z = 0 \\ 0x - 2y - z = 0 \\ 0x + 2y + z = 0 \end{cases}$$

which yields

$$x - y = 0 \quad 2y + z = 0$$

so

$$x = y \quad z = -2y$$

and therefore the eigenspace corresponding to 3 is spanned by the vector  $(1, 1, -2)$ , that is  $\mathcal{V}_3 = \mathcal{L}(1, 1, -2)$ . Notice that while the eigenvalue 2 has algebraic multiplicity 2 it has geometric multiplicity only 1. It is therefore *not* possible to find a basis for  $\mathbb{R}^3$  composed of eigenvectors of  $T$  and hence  $T$  is not diagonalizable.

EXAMPLE 16. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by

$$T(x, y, z) = (0, x, y).$$

Determine if  $T$  can be diagonalized.

*Solution.* The matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$  is

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

so the characteristic polynomial of  $T$  is

$$\Delta(t) = \det \begin{pmatrix} -t & 0 & 0 \\ 1 & -t & 0 \\ 0 & 1 & -t \end{pmatrix} = -t^3$$

and the only eigenvalue of  $T$  is 0 with algebraic multiplicity 3. The geometric multiplicity cannot also be 3 because  $\mathcal{V}_0 = \ker T$  and  $3 = \dim \mathcal{V}_0 = \dim \ker T$  implies  $T = 0$ . But certainly  $T \neq 0$ . In fact the kernel of  $T$  is the subspace spanned by  $(0, 0, 1)$  so the geometric multiplicity of the eigenvalue 0 is 1.

EXAMPLE 17. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by

$$T(x, y, z) = (-x, -z, y)$$

and let  $S: \mathbb{C}^3 \rightarrow \mathbb{C}^3$  be defined by (the complex formula)

$$S(u, v, w) = (-u, -w, v).$$

Determine if  $T$  or  $S$  can be diagonalized, and if so find their diagonal forms.

*Solution.* At least part of the problem can be worked simultaneously for both **T** and **S** because relative to the standard bases of  $\mathbb{R}^3$  and  $\mathbb{C}^3$  respectively both **T** and **S** have the same matrix:

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

so their characteristic polynomials are the same, namely,

$$\Delta(t) = \det \begin{pmatrix} -1-t & 0 & 0 \\ 0 & -t & -1 \\ 0 & 1 & -t \end{pmatrix} = -(t+1)(t^2+1),$$

and now the difference between the real and complex case appears!

$$\Delta(t) = -(t+1)(t^2+1)$$

has only one real root so **T** does not diagonalize (though geometrically it is quite simple: reflection across the  $yz$ -plane followed by a  $90^\circ$  rotation around the  $x$ -axis) and **S** does since

$$\Delta(t) = -(t+1)(t-i)(t+i)$$

has three distinct eigenvalues,  $-1, i, -i$ . The diagonal form of **S** is

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

The corresponding basis may be computed as follows.

For  $e_1 = -1$ :

$$\mathbf{S}(u, v, w) = -(u, v, w) = (-u, -v, -w)$$

so

$$(-u, -w, v) = (-u, -v, -w)$$

and passing to components gives

$$-w = -v \quad \text{and} \quad v = -w$$

so

$$w = v = 0$$

and therefore a basis for  $\mathcal{V}_{-1}$  is  $(1, 0, 0)$ .

For  $e_2 = i$ :

$$\mathbf{S}(u, v, w) = i(u, v, w)$$

so

$$(-u, -w, v) = (iu, iv, iw),$$

## 14: The elements of eigenvalue and eigenvector theory

taking components gives

$$-u = iu \quad -w = iv \quad v = iw$$

so

$$u = 0, \quad v = i, \quad w = 1.$$

Thus a basis for  $\mathcal{V}_i$  is  $(0, i, 1)$ .

For  $e_3 = -i$ : From  $\mathbf{S}(u, v, w) = -i(u, v, w)$ , as before passing to components gives

$$-u = -iu \quad -w = -iv \quad v = -iw$$

so

$$u = 0, \quad v = i, \quad w = -1.$$

Thus a basis for  $\mathcal{V}_{-i}$  is  $(0, i, -1)$ .

### EXERCISES

1. Find the rank of each of the following linear transformations

$$\mathbf{T}: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad \mathbf{T}(x, y, z) = (x + y, y, z + x)$$

$$\mathbf{T}: \mathbb{R}^4 \rightarrow \mathbb{R}^4, \quad \mathbf{T}(x, y, z, w) = (x - y, z - w, w + x, x - w)$$

$$\mathbf{T}: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R}), \quad \mathbf{T}(p(x)) = x \frac{d}{dx} p(x)$$

$$\mathbf{T}: \mathcal{P}_3(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R}), \quad \mathbf{T}(p(x)) = \frac{d}{dx} (xp(x)).$$

2. Find the cofactor of  $a_{24}$  and  $a_{42}$  in the matrix

$$\begin{pmatrix} 1 & 7 & 7 & 1 & 7 & 7 \\ 2 & 8 & 6 & 2 & 8 & 6 \\ 3 & 9 & 5 & 3 & 9 & 5 \\ 4 & 10 & 4 & 4 & 10 & 4 \\ 5 & 9 & 3 & 5 & 9 & 3 \\ 6 & 8 & 2 & 6 & 8 & 2 \end{pmatrix}$$

3. Find the determinant of each of the following matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

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

4. Determine which of the following matrices are invertible. For those that are, compute the inverses.

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

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

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

5. Can you find a  $2 \times 2$  matrix  $A$  such that  $A = A^{\text{cof}}$ ?
6. Find the characteristic polynomial, eigenvalues, and eigenvectors and diagonalize if possible each of the following linear transformations:

$$\begin{array}{ll} T: \mathbb{R}^3 \rightarrow \mathbb{R}^3 & T(x, y, z) = (x + 2y, 2y + x, z) \\ T: \mathbb{R}^3 \rightarrow \mathbb{R}^3 & T(x, y, z) = (3x - z, y - x + 2z, 4z) \\ T: \mathbb{R}^4 \rightarrow \mathbb{R}^4 & T(x, y, z, w) = (x + y, x, z + 2w, 2z + w) \\ T: \mathbb{R}^4 \rightarrow \mathbb{R}^4 & T(x, y, z, w) = (x + y, y, 2z, y + 2z + 2w) \\ T: \mathbb{R}^2 \rightarrow \mathbb{R}^2 & T(x, y) = (y, -x) \\ T: \mathbb{R}^2 \rightarrow \mathbb{R}^3 & T(x, y) = (y, -x, z). \end{array}$$

7. Prove that if a transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is nilpotent then all its eigenvalues are zero.
8. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . Prove that  $T$  has at least one eigenvalue.
9. Suppose that  $T: \mathcal{V} \rightarrow \mathcal{V}$  is nonsingular. What relation is there between the eigenvalues of  $T$  and  $T^{-1}$ ?
10. Give an example of a linear transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  without any eigenvalues. Repeat the problem for  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$ .
11. Suppose that  $e$  is an eigenvalue of  $T: \mathcal{V} \rightarrow \mathcal{V}$  with corresponding eigenvector  $E$ . Prove that  $E$  is an eigenvector of  $T^k$  corresponding to the eigenvalue  $e^k$ .
12. Suppose that  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a linear transformation with distinct eigenvalues  $e_1, \dots, e_n$ . Show that  $\det T = e_1 \cdots e_n$ .
13. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a diagonalizable endomorphism. Show that  $T^2$  is also diagonalizable. If the diagonal form of  $T$  is

$$\begin{pmatrix} e_1 & & & 0 \\ & \ddots & & \\ & & e_2 & \\ 0 & & & \ddots & \\ & & & & e_n \end{pmatrix}$$

what is the diagonal form of  $T^2$ ?

# 14: The elements of eigenvalue and eigenvector theory

14. Let  $A = (a_{ij})$ ,  $i, j = 1, 2, 3$ . Then  $\det A = a_{11} \det M_{11} - a_{12} \det M_{12} + a_{13} \det M_{13}$ , where  $M_j$ , where  $M_{1j}$ ,  $j = 1, 2, 3$ , are minors of the element  $a_{1j}$  of  $A$ . Show that

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

15. Using the above expression of  $\det A$  for  $3 \times 3$  matrix, show properties (1)–(8) of  $\det A$  hold for  $3 \times 3$  matrices.

16. Let

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}.$$

Show that  $\det A \neq 0$  iff column vectors form a set of linearly independent vectors. (Hint: consider  $A$  to be the matrix of a linear transformation with respect to the standard basis.)

17. Let  $(a, b, c)$ ,  $(a', b', c')$  be two independent vectors. Suppose  $(x, y, z)$  is another vector which lies in the linear span of  $(a, b, c)$  and  $(a', b', c')$ . Then

$$\det \begin{pmatrix} x & y & z \\ a & b & c \\ a' & b' & c' \end{pmatrix} = 0,$$

i.e.,  $(bc' - cb')x + (ca' - ac')y + (ab' - ba')z = 0$  (Compare to (\*\*) preceding Example 5 of Chapter 1.)

18. Given vectors  $(a, b, c)$ ,  $(a', b', c') \in \mathbb{R}^3$  the vector

$$(bc' - b'c, ca' - ac', ab' - ba')$$

is called the outer-product (or cross-product) of vectors  $(a, b, c)$  and  $(a', b', c')$ , and denoted by  $(a, b, c) \times (a', b', c')$ . Show

- (a)  $(a, b, c) \times (a', b', c') = -(a', b', c') \times (a, b, c)$ .  
 (b)  $(a, b, c) \times (a, b, c) = 0$ .

19. (Cramer's rule). Let

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \quad i = 1, 2, \dots, n,$$

and suppose that the matrix  $A = (a_{ij})$ , gotten from coefficients of the above system of linear equations, has the inverse  $A^{-1}$ . Then

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = A^{-1}B, \quad \text{where } B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Also,  $x_i = \det A_i / \det A$ ,  $i = 1, 2, \dots, n$ , where  $A_i$  is the matrix  $A$  whose  $i$ th column is replaced by  $B$ ,  $i = 1, 2, \dots, n$ . Verify the above statement for  $n = 3$ . The statement holds for any  $n$ , (finite!). But for all practical purpose it is useless for a large  $n$ , in

computing  $x_i$ , since you have to compute  $n + 1$  determinants of size  $n$  matrices (such a tedious task!).

20. Solve the following system of linear equations:

$$\begin{aligned}x_1 + x_2 + x_3 + x_4 &= 5 \\x_2 + 2x_3 - x_4 &= -3 \\x_1 + 3x_2 + 4x_3 - x_4 &= 0 \\2x_1 - x_2 - x_3 + x_4 &= 4\end{aligned}$$

using Cramer's rule.

21. Using Cramer's rule solve the following systems of linear equations:

$$\begin{aligned}\text{(a)} \quad x + y + z &= 0 \\2x - y + z &= 6 \\4x + 5y - z &= 2 \\ \text{(b)} \quad 3x + y - z &= 1 \\-x - y + 4z &= 7 \\2x + y - 5z &= -8\end{aligned}$$

22. Compute  $\det \mathbf{H}$ , where

$$\mathbf{H} = \begin{pmatrix} 1 & x & x^2 & x^3 \\ 1 & y & y^2 & y^3 \\ 1 & z & z^2 & z^3 \\ 1 & w & w^2 & w^3 \end{pmatrix}.$$

23. Find the inverse of each of the following matrices

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

Can you find a general pattern?

24. Diagonalize each of the matrices in Exercise 23.

25. Let  $T: \mathcal{M}_i \rightarrow \mathcal{M}_i$  be the linear transformation given by

$$T(\mathbf{M}) = \mathbf{M}^t.$$

- Find the characteristic polynomial of  $T$ .
  - Determine the eigenvalues and eigenspaces of  $T$ .
  - Is  $T$  diagonalizable? If so, diagonalize it; if not, prove it is not.
26. Recall (Chapter 12, Exercise 15) that a linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is an  $n$ -dimensional vector space  $\mathcal{V}$  is called a reflection if there is an  $(n - 1)$ -dimensional subspace  $\mathcal{H}$  in  $\mathcal{V}$  and a nonzero vector  $\mathbf{v}$  in  $\mathcal{V}$ , not in  $\mathcal{H}$  so that

$$T(\mathbf{h}) = \mathbf{h} \quad \text{for all } \mathbf{h} \in \mathcal{H},$$

$$T(\mathbf{v}) = -\mathbf{v}.$$



14: The elements of eigenvalue and eigenvector theory

- (a) Show that  $T$  is diagonalizable with diagonal form

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

- (b) Show  $\det T = -1$ .

- (c) Let  $T$  and  $S$  be reflections. Is  $TS$  also a reflection?

27. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a cyclic transformation in the  $n$ -dimensional vector space  $\mathcal{V}$  with cyclic vector  $A$ . Then there exist unique numbers  $a_0, \dots, a_{n-1}$  such that

$$T(A^n) = a_0 A + a_1 T(A) + \dots + a_{n-1} T^{n-1}(A).$$

Show that up to the  $\pm$  sign the characteristic polynomial of  $T$  is:

$$a_0 t^n + a_1 t^{n-1} + \dots + a_{n-2} t + a_{n-1}.$$

28. Show that the only eigenvalues of an involution are  $\pm 1$ .

29. Is  $\det: \mathcal{M}_n \rightarrow \mathbb{R}$  a linear transformation?

30. Is the trace  $tr: \mathcal{M}_n \rightarrow \mathbb{R}$  a linear transformation? (See p. 248 for the definition of the trace of a square matrix.)

Let  $n$  be a positive integer, and let  $\mathcal{V}_1, \dots, \mathcal{V}_n, \mathcal{W}$  be vector spaces. A function

$$f: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

is called a *multilinear form* iff for each integer  $i$ ,  $1 \leq i \leq n$ , and each  $(n-1)$ -tuple  $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$  the function

$$F: \mathcal{V}_i \rightarrow \mathcal{W}$$

defined by

$$F(v) = f(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_n)$$

is a linear transformation. For  $n = 1$  a multilinear form is simply a linear transformation. For  $n = 2$  we speak of a bilinear form. Often we just say *form* for a multilinear form.

EXAMPLE 1. Vector addition defines a form

$$\mathcal{V} \times \dots \times \mathcal{V} \rightarrow \mathcal{V}.$$

EXAMPLE 2. Let  $\mathcal{V}_1, \dots, \mathcal{V}_n$  be subspaces of  $\mathcal{W}$  then the map

$$f: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

defined by

$$f(v_1, \dots, v_n) = v_1 + v_2 + \dots + v_n$$

is a multilinear form.

**EXAMPLE 3.** Let  $a, b$  be positive integers. Then multiplication of polynomials defines a bilinear form

$$\mu: \mathcal{P}_a(\mathbb{R}) \times \mathcal{P}_b(\mathbb{R}) \rightarrow \mathcal{P}_{a+b}(\mathbb{R}).$$

**EXAMPLE 4.** Let  $a, b$  be positive integers and define

$$f: \mathbb{R}^{a+1} \times \mathbb{R}^{b+1} \rightarrow \mathbb{R}^{a+b+1}$$

by (note the indexing starts with a zero)

$$f(x, y) = (z_0, \dots, z_{a+b}),$$

where

$$x = (x_0, \dots, x_a), \quad y = (y_0, \dots, y_b),$$

and

$$z_c = \sum_{i+j=c} x_i y_j.$$

Then  $f$  is a multilinear form.

**Definition.** Two multilinear forms

$$f': \mathcal{V}'_1 \times \dots \times \mathcal{V}'_n \rightarrow \mathcal{W}',$$

$$f'': \mathcal{V}''_1 \times \dots \times \mathcal{V}''_n \rightarrow \mathcal{W}'',$$

are said to be *equivalent* iff there exist vector space isomorphisms

$$\varphi_i: \mathcal{V}'_i \rightarrow \mathcal{V}''_{\sigma(i)}, \quad i = 1, \dots, n,$$

$$\psi: \mathcal{W}' \rightarrow \mathcal{W}'',$$

where  $(\sigma(1), \dots, \sigma(n))$  is a rearrangement (i.e., *permutation*) of the integers  $(1, \dots, n)$  such that

$$f'(v'_1, \dots, v'_n) = f''(v''_1, \dots, v''_n) \quad \text{for all } v'_i \in \mathcal{V}'_i, \quad i = 1, \dots, n,$$

where

$$v''_j = \varphi_j(v'_{\tau(j)}), \quad j = 1, \dots, n,$$

and  $\tau$  is the inverse rearrangement of  $\sigma$ , that is,

$$\tau(\sigma(j)) = j, \quad j = 1, \dots, n.$$

**Theorem 14<sup>bis</sup>.1** (Linear Extension Construction for Forms). *Let  $n$  be a positive integer and let  $\mathcal{V}_1, \dots, \mathcal{V}_n, \mathcal{W}$  be vector spaces. Assume*

$$\dim \mathcal{V}_i = d_i$$

*is finite, and let  $\{e_{i,1}, \dots, e_{i,d_i}\}$  be a basis for  $\mathcal{V}_i$ ,  $i = 1, \dots, n$ . For each index  $(j_1, \dots, j_n)$  with  $1 \leq j_s \leq d_i$  let  $w_{(j_1 \dots j_n)} \in \mathcal{W}$ . Then there is one and only one form*

$$f: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

such that

$$f(e_{1,j_1}, \dots, e_{n,j_n}) = w_{(j_1, \dots, j_n)}.$$

PROOF. This is analogous to (8.13), namely, the linear extension construction. To construct  $f$  we suppose  $v_i \in \mathcal{V}_i$ ,  $i = 1, \dots, n$ . We can then write

$$v_i = a_{i,1}e_{i,1} + \dots + a_{i,d_i}e_{i,d_i}$$

for unique scalars  $a_{i,1}, \dots, a_{i,d_i}$ . We now set (this definition is forced by the desired multilinearity)

$$f(v_1, \dots, v_n) = \sum_{\substack{(j_1, \dots, j_n) \\ 1 \leq j_i \leq d_i}} a_{(1,j_1)} a_{(2,j_2)} \dots a_{(n,j_n)} w_{(j_1, \dots, j_n)}.$$

A simple check of the definitions shows that  $f$  is indeed a multilinear form. Furthermore, if

$$g: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

is any multilinear form such that

$$g(e_{1,j_1}, \dots, e_{n,j_n}) = w_{(j_1, \dots, j_n)}, \quad 1 \leq j_i \leq d_i,$$

then by multilinearity

$$g(v_1, \dots, v_n) = \sum a_{(1,j_1)} \dots a_{(n,j_n)} w_{(j_1, \dots, j_n)},$$

that is,  $f = g$ . □

If  $a$  is a scalar and

$$f, g: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

are multilinear forms, then so are

$$\begin{aligned} f + g: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}, \quad & \rightarrow (f + g)(v_1, \dots, v_n) \\ & = f(v_1, \dots, v_n) + g(v_1, \dots, v_n), \end{aligned}$$

$$af: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}, \quad \rightarrow (af)(v_1, \dots, v_n) = af(v_1, \dots, v_n).$$

Here is a partial verification of the first for  $n = 2$ .

$$\begin{aligned} (f + g)(x' + x'', y) &= f(x' + x'', y) + g(x' + x'', y) \quad \text{Def}^n. \\ &= f(x', y) + f(x'', y) + g(x', y) + g(x'', y) \quad \text{bilinearity} \\ &= f(x', y) + g(x', y) + f(x'', y) + g(x'', y) \\ &= (f + g)(x', y) + (f + g)(x'', y). \end{aligned}$$

The verification in general is completely analogous. We have thus proved:

**Proposition 14<sup>bis</sup>.2.** *The set of all multilinear forms*

$$f: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

is a vector space with respect to the addition of forms and multiplication of a form by a scalar as defined above.

**Proposition 14<sup>bis</sup>.3.** The dimension of the vector space of multilinear forms

$$f: \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

is  $d \cdot d_1 \cdots d_n$  where  $d_i = \dim \mathcal{V}_i$ ,  $i = 1, \dots, n$  and  $d = \dim \mathcal{W}$ .

**PROOF.** Follow from (14<sup>bis</sup>.1). □

The study of multilinear forms (multilinear algebra) leads to many interesting, and in some cases, totally unsolved problems. As an illustration of what is possible we first introduce one new concept.

**Definition.** A form

$$f: \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

is called *nonsingular* iff whenever

$$f(v_1, v_2, \dots, v_n) = 0$$

at least one of  $v_1, \dots, v_n$  is the zero vector.

**EXAMPLE 5.**  $\mu: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by

$$\mu(x, y) = (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1)$$

is a nonsingular bilinear form.

It is possible to construct nonsingular bilinear forms

$$\mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}^4,$$

$$\mathbb{R}^8 \times \mathbb{R}^8 \rightarrow \mathbb{R}^8.$$

Just as  $\mu$  devolves from multiplication of complex numbers these forms derive from multiplication of quaternions and Cayley numbers. These facts were known in the nineteenth century. Only in 1960 was it proved: if there is a nonsingular form

$$\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n,$$

then  $n = 1, 2, 4$  or  $8$ .

**Open Problem.** Let  $n, k$  be positive integers. Find the largest integer  $b(k, n)$  such that there exists a nonsingular bilinear form

$$\mathbb{R}^n \times \mathbb{R}^{b(k, n)} \rightarrow \mathbb{R}^k.$$

Our purpose for introducing multilinear forms is to define and establish the basic properties of the determinant of  $n \times n$  matrices. We turn to this now. We will be concerned with multilinear forms

$$f: \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \rightarrow \mathcal{W},$$

where  $\mathcal{V}_1 = \cdots = \mathcal{V}_n$ , that is, multilinear forms of the type

$$f: \mathcal{V} \times \cdots \times \mathcal{V} \rightarrow \mathcal{W}.$$

**Definition.** A form

$$f: \mathcal{V} \times \cdots \times \mathcal{V} \rightarrow \mathcal{W}$$

is called *sign-symmetric* iff for every pair of integers  $i, j$ ,  $1 \leq i \neq j \leq n$  and  $v_1, \dots, v_n \in \mathcal{V}$

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -f(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

A sign-symmetric form is also referred to as **alternating**. Note that whenever  $v_i = v_j$  for some  $1 \leq i \neq j \leq n$ , then interchanging  $v_i$  and  $v_j$  does not change the value of  $f(v_1, \dots, v_n)$ . If  $f$  is sign-symmetric on the other hand the sign of the value must change. This is only possible if

$$f(v_1, \dots, v_n) = 0,$$

whenever there is a pair of integers  $i, j$  such that  $1 \leq i \neq j \leq n$  and  $v_i = v_j$ .

**EXAMPLE 6.** Let

$$\Delta: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$

be defined by

$$\Delta(a, b), (c, d)) = ad - bc.$$

Then  $\Delta$  is bilinear and sign-symmetric. The verification, while tedious, is straightforward and left to the reader.

**Theorem 14<sup>bis</sup>.4.** Let  $\mathcal{V}$  be an  $n$ -dimensional vector space. Then the sign-symmetric forms (Note:  $n = \dim \mathcal{V}$ )

$$f: \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_n \rightarrow \mathbb{R}$$

form a one-dimensional subspace in the space of all forms.

**PROOF.** Let  $\{e_1, \dots, e_n\}$  be a basis for  $\mathcal{V}$ . By (14<sup>bis</sup>.1) a form

$$f: \mathcal{V} \times \cdots \times \mathcal{V} \rightarrow \mathbb{R}$$

is completely determined when we know the values  $f(e_{i_1}, \dots, e_{i_n}) \in \mathbb{R}$ ,  $1 \leq i_j \leq n$ . If the form is alternating then

$$f(e_{i_1}, \dots, e_{i_n}) = 0,$$

whenever  $e_{i_a} = e_{i_b}$  for a pair indexes  $i_a \neq i_b$ . This means that an alternating form

$$f: \underbrace{\mathcal{V} \times \dots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

is determined by the values

$$f(e_{i_1}, \dots, e_{i_n}),$$

where the indices are all distinct. If the form is sign-symmetric and  $(e_{i_1}, \dots, e_{i_n})$  has distinct indices, then

$$f(e_{i_1}, \dots, e_{i_n}) = (-1)^s f(e_1, e_2, \dots, e_n),$$

where  $s$  is the number of pairs of indices that have to be switched to rearrange the  $n$ -tuple  $(i_1, \dots, i_n)$  to give  $(1, \dots, n)$ . Thus a sign-symmetric/alternating form

$$f: \underbrace{\mathcal{V} \times \dots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

is completely determined by

sign-symmetry

and

$$f(e_1, e_2, \dots, e_n) \in \mathbb{R}.$$

Let

$$g: \underbrace{\mathcal{V} \times \dots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

be the sign-symmetric/alternating form defined by

$$g(e_1, \dots, e_n) = 1.$$

Let

$$f: \mathcal{V} \times \dots \times \mathcal{V} \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

be a sign-symmetric/alternating form. Then

$$h := f - f(e_1, \dots, e_n) = g, \quad \mathcal{V} \times \dots \times \mathcal{V} \rightarrow \mathbb{R}$$

is also sign-symmetric and alternating. Moreover

$$h(e_1, \dots, e_n) = 0.$$

Therefore  $h$  is identically 0. Thus  $f = f(e_1, \dots, e_n) \cdot g$  and  $g$  spans the space of sign-symmetric forms

$$f: \mathcal{V} \times \dots \times \mathcal{V} \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

so this space is one-dimensional. □

Suppose  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a linear transformation and

$$g: \underbrace{\mathcal{V} \times \dots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

is a nonzero sign-symmetric/alternating form. Then

$$g_T: \mathcal{V} \times \dots \times \mathcal{V} \rightarrow \mathbb{R}$$

defined by

$$g_T(v_1, \dots, v_n) = g(Tv_1, \dots, Tv_n) \in \mathbb{R}$$

is also a sign-symmetric/alternating form. Since the space of such forms is one-dimensional, and  $g \neq 0$ , it follows that

$$g_T = d_g(T) \cdot g$$

for a unique scalar  $d_g(T) \in \mathbb{R}$ . Thus we are led to:

**Definition.** Let  $M \in \mathcal{M}_{nn}$  be an  $n \times n$  matrix. Let

$$g: \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$$

be the sign-symmetric/alternating form defined by

$$g(E_1, \dots, E_n) = 1,$$

where  $\{E_1, \dots, E_n\}$  is the standard basis for  $\mathbb{R}^n$ . Let

$$M: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

be the linear transformation represented by the matrix  $M$  with respect to the standard basis. The *determinant* of  $M$  is the unique real number  $\det(M)$  such that  $g_M = \det(M) \cdot g$ .

**Theorem 14<sup>bis</sup>.5.** *The function*

$$\text{Det}: \mathcal{M}_{nn} \rightarrow \mathbb{R}$$

*has the following properties:*

- (1) *If  $A$  has a column of zero then  $\det A = 0$ .*
- (2) *If  $A$  has two columns equal then  $\det A = 0$ .*
- (3) *If  $B$  is obtained from  $A$  by interchanging two columns then  $\det B = -\det A$ .*



- (4) If  $\mathbf{B}$  is obtained from  $\mathbf{A}$  by adding a multiple of one column to a different column then  $\det \mathbf{B} = \det \mathbf{A}$ .  
 (5) If  $\mathbf{B}$  is obtained by multiplying a column of  $\mathbf{A}$  by a number  $k$  then  $\det \mathbf{B} = k \det \mathbf{A}$ .  
 (6)  $\det \mathbf{A}^t = \det \mathbf{A}$ ; thus everything said about columns in (1)–(5) holds also for rows.  
 (7) If  $\mathbf{A}$  is upper triangular then  $\det \mathbf{A}$  is the product of the diagonal entries.  
 (8)  $\det(\mathbf{A} \cdot \mathbf{B}) = \det \mathbf{A} \det \mathbf{B}$ .  
 (9)  $\mathbf{A}$  is invertible  $\Leftrightarrow \det \mathbf{A} \neq 0$ .

Before taking up the proof proper we require some preliminary manouvering. The proof of (14<sup>bis</sup>.4) shows that a sign-symmetric form

$$f: \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V},$$

is completely determined by the value

$$f(e_1, \dots, e_n) \in \mathbb{R},$$

where  $\{e_1, \dots, e_n\}$  is a basis for  $\mathcal{V}$ . In fact, if

$$g: \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V},$$

is the unique sign-symmetric/alternating form such that

$$g(e_1, \dots, e_n) = 1,$$

then  $f = f(e_1, \dots, e_n) \cdot g$  as forms. Thus we have shown:

**Corollary 14<sup>bis</sup>.6.** Two sign-symmetric/alternating forms

$$f', f'': \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_n \rightarrow \mathbb{R}, \quad n = \dim \mathcal{V}$$

are isomorphic  $\Leftrightarrow$  for some basis  $\{e_1, \dots, e_n\}$  of  $\mathcal{V}$

$$f'(e_1, e_2, \dots, e_n) = f''(e_1, \dots, e_n).$$

Let  $\mathbf{A} \in \mathcal{M}_n$  and write

$$\mathbf{L}(\mathbf{A}): \mathbb{R}^n \rightarrow \mathbb{R}^n$$

for the linear transformation corresponding to the matrix  $\mathbf{A}$  with respect to the standard basis  $\{\mathbf{E}_1, \dots, \mathbf{E}_n\}$  for  $\mathbb{R}^n$  (see (12.1)). Let

$$\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_n),$$

where

$$\mathbf{A}_i = \begin{pmatrix} a_{i1} \\ \vdots \\ a_{in} \end{pmatrix}, \quad i = 1, \dots, n,$$

is the  $i$ th column of  $\mathbf{A}$ . Then

$$\mathbf{L}(\mathbf{A})(\mathbf{E}_i) = \mathbf{A}_i^t, \quad i = 1, \dots, n,$$

where  $\mathbf{A}_i^t$  denotes the transpose of  $\mathbf{A}_i$ ; that is, the column  $\mathbf{A}_i$  rewritten as a row.

Let

$$g: \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_n \rightarrow \mathbb{R}$$

be the unique symmetric form defined by

$$g(\mathbf{E}_1, \dots, \mathbf{E}_n) = 1.$$

To compute the determinant of  $\mathbf{A}$  we introduce the form

$$g_{\mathbf{L}(\mathbf{A})}: \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

defined by

$$g_{\mathbf{L}(\mathbf{A})}(v_1, \dots, v_n) = g(\mathbf{A}v_1, \dots, \mathbf{A}v_n).$$

Then  $\det \mathbf{A}$  is the unique number such that

$$g_{\mathbf{A}} = (\det \mathbf{A}) \cdot g: \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}.$$

The remarks preceding (14<sup>bis</sup>.6) thus show

$$\begin{aligned} g_{\mathbf{A}} &= g_{\mathbf{A}}(\mathbf{E}_1, \dots, \mathbf{E}_n) \cdot g = g(\mathbf{L}(\mathbf{A})\mathbf{E}_1, \dots, \mathbf{L}(\mathbf{A})\mathbf{E}_n) \\ &= g(\mathbf{A}_1^t, \dots, \mathbf{A}_n^t), \end{aligned}$$

and therefore we conclude:

**Lemma 14<sup>bis</sup>.7.** *In the preceding notations we have*

$$\det(\mathbf{A}) = g(\mathbf{A}_1^t, \dots, \mathbf{A}_n^t)$$

for  $\mathbf{A} \in \mathcal{M}_n$ . □

**PROOF OF 14<sup>bis</sup>.5.** In permuted order here are the proofs.

- (2) Suppose  $\mathbf{A}$  has two columns equal, say  $\mathbf{A}_i = \mathbf{A}_j$ ,  $i \neq j$ . Then  $\mathbf{A}_i^t = \mathbf{A}_j^t$ , so  $0 = g(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_j^t, \dots, \mathbf{A}_n^t) = \det(\mathbf{A})$  because  $g$  is alternating.
- (1) Suppose  $\mathbf{A}$  has a column of zeros, say  $\mathbf{A}_i = (0)$ . Then  $\mathbf{A}_i^t = (0, \dots, 0)$  so  $0 = g(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_n^t) = \det \mathbf{A}$  by multilinearity.
- (3) If  $\mathbf{B}$  is obtained from  $\mathbf{A}$  by interchanging the two columns  $\mathbf{A}_i$  and  $\mathbf{A}_j$ , then if we arrange the notations so that  $i < j$  we see

$$\begin{aligned} \det(\mathbf{B}) &= g(\mathbf{A}_1, \dots, \mathbf{A}_j, \dots, \mathbf{A}_i, \dots, \mathbf{A}_n) \\ &= -g(\mathbf{A}_1, \dots, \mathbf{A}_i, \dots, \mathbf{A}_j, \dots, \mathbf{A}_n) = -\det(\mathbf{A}), \end{aligned}$$

because  $g$  is sign-symmetric.

(4) Suppose  $\mathbf{B}$  is obtained from  $\mathbf{A}$  by adding  $\mathbf{A}_i$  to  $\mathbf{A}_j$ ,  $i \neq j$ . Then

$$\begin{aligned}\det(\mathbf{B}) &= g(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_j^t + a\mathbf{A}_i^t, \dots, \mathbf{A}_n^t) \\ &= g(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_j^t, \dots, \mathbf{A}_n^t) \\ &\quad + g(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, a\mathbf{A}_i^t, \dots, \mathbf{A}_n^t) \\ &= \det \mathbf{A} + ag(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_n^t) \\ &= \det \mathbf{A} + 0 = \det \mathbf{A},\end{aligned}$$

by property (2).

(5) Suppose  $\mathbf{B}$  is obtained from  $\mathbf{A}$  by multiplying  $\mathbf{A}_i$  by  $a$ . Then

$$\begin{aligned}\det \mathbf{B} &= g(\mathbf{A}_1^t, \dots, a\mathbf{A}_i^t, \dots, \mathbf{A}_n^t) \\ &= ag(\mathbf{A}_1^t, \dots, \mathbf{A}_i^t, \dots, \mathbf{A}_n^t) = a \det(\mathbf{A}),\end{aligned}$$

by multilinearity.

(7) Suppose that  $\mathbf{A}$  is an upper triangular matrix, that is

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & & & \\ & a_{22} & & & \\ & & \ddots & & \\ 0 & & & \ddots & \\ & & & & a_{nn} \end{bmatrix}$$

Then

$$\begin{aligned}\mathbf{A}_1^t &= a_{11}\mathbf{E}_1, \\ \mathbf{A}_2^t &= a_{12}\mathbf{E}_1 + a_{22}\mathbf{E}_2, \\ &\vdots \\ \mathbf{A}_n^t &= a_{1n}\mathbf{E}_1 + \dots + a_{nn}\mathbf{E}_n,\end{aligned}$$

as vectors in  $\mathbb{R}^n$ . Therefore

$$\begin{aligned}\det \mathbf{A} &= g(\mathbf{A}_1^t, \dots, \mathbf{A}_n^t) \\ &= g(a_{11}\mathbf{E}_1, \mathbf{A}_2^t, \dots, \mathbf{A}_n^t),\end{aligned}$$

(by linearity in the first

$$\text{variable}) = a_{11}g(\mathbf{E}_1, a_{12}\mathbf{E}_1 + a_{22}\mathbf{E}_2, \mathbf{A}_3^t, \dots, \mathbf{A}_n^t),$$

(by linearity in the

$$\begin{aligned}\text{second variable}) &= a_{11}a_{12}g(\mathbf{E}_1, \mathbf{E}_1, \mathbf{A}_3^t, \dots, \mathbf{A}_n^t) \\ &\quad + a_{11}a_{22}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{A}_3^t, \dots, \mathbf{A}_n^t),\end{aligned}$$

$$\text{(since } g \text{ is alternating)} = a_{11}a_{12} \cdot 0 + a_{11}a_{22}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{A}_3^t, \dots, \mathbf{A}_n^t),$$

(by linearity in the

$$\begin{aligned}\text{third variable}) &= a_{11}a_{22}a_{13}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_1, \dots, \mathbf{A}_n^t) \\ &\quad + a_{11}a_{22}a_{23}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_2, \mathbf{A}_4^t, \dots, \mathbf{A}_n^t) \\ &\quad + a_{11}a_{22}a_{33}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{A}_4^t, \dots, \mathbf{A}_n^t),\end{aligned}$$

$$\begin{aligned}
(\text{since } g \text{ is alternating}) &= a_{11}a_{22}a_{13} \cdot 0 + a_{11}a_{22}a_{23} \cdot 0 \\
&\quad + a_{11}a_{22}a_{23}g(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{A}_4^t, \dots, \mathbf{A}_n^t) \\
&\quad \vdots \\
&= a_{11}a_{22}, \dots, a_{1n} \cdot 0 + \dots + a_{11}a_{22}, \dots, a_{1, n-i} \cdot 0 \\
&\quad + a_{11}, \dots, a_{nn}g(\mathbf{E}_1, \dots, \mathbf{E}_n) \\
&= a_{11}, \dots, a_{nn},
\end{aligned}$$

as required.

It is convenient, before proving (8) to prove (9).

- (9) Let  $\mathbf{A} \in \mathcal{M}_{nn}$ . Suppose that  $\mathbf{A}$  is invertible. Then  $\{\mathbf{A}\mathbf{E}_1, \dots, \mathbf{A}\mathbf{E}_n\}$  is a basis for  $\mathbb{R}^n$ . Suppose

$$g(\mathbf{A}\mathbf{E}_1, \dots, \mathbf{A}\mathbf{E}_n) = 0.$$

By (14<sup>bis</sup>.4) this implies that

$$g = 0: \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_n \rightarrow \mathbb{R}$$

as multilinear form, since, of course

$$0(\mathbf{A}\mathbf{E}_1, \dots, \mathbf{A}\mathbf{E}_n) = 0$$

and a symmetric/alternating form is completely determined by its value on *any* ordered basis, in particular, on the ordered basis  $\{\mathbf{A}\mathbf{E}_1, \dots, \mathbf{A}\mathbf{E}_n\}$ . Therefore

$$0 \neq g_{\mathbf{A}} = \det \mathbf{A} \cdot g = \det \mathbf{A} \neq 0,$$

as required.

- (8) Suppose  $\mathbf{B}$  is singular. Then so is  $\mathbf{A} \cdot \mathbf{B}$  so by (9)

$$\det(\mathbf{A}\mathbf{B}) = 0 = \det \mathbf{A} \cdot 0 = \det \mathbf{A} \cdot \det \mathbf{B}.$$

If  $\mathbf{B}$  is nonsingular, then the form

$$g_{\mathbf{B}}: \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$$

is nonzero, since by (9)

$$g_{\mathbf{B}} = \det \mathbf{B} \cdot g \neq 0.$$

Therefore  $g_{\mathbf{B}}$  is a basis for the symmetric/alternating forms since the space of such forms is one-dimensional. Thus as forms there exists  $d \in \mathbb{R}$  such that

$$g_{\mathbf{A}\mathbf{B}} = dg_{\mathbf{B}}: \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_n \rightarrow \mathbb{R}.$$

Therefore we obtain from the definitions

$$g_{\mathbf{AB}} = dg_{\mathbf{B}} = d \det(\mathbf{B})g,$$

and

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

So it remains to show that  $d = \det \mathbf{A}$ . To this end note that since  $\mathbf{B}$  is nonsingular can consider the point

$$(\mathbf{B}^{-1}\mathbf{E}_1, \dots, \mathbf{B}^{-1}\mathbf{E}_n) \in \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_n.$$

So evaluating the two sides of the equation  $g_{\mathbf{AB}} = dg_{\mathbf{B}}$  on

$$(\mathbf{B}^{-1}\mathbf{E}_1, \dots, \mathbf{B}^{-1}\mathbf{E}_n) \in \mathbb{R}^n \times \dots \times \mathbb{R}^n$$

gives, on the one hand,

$$\begin{aligned} g_{\mathbf{AB}}(\mathbf{B}^{-1}\mathbf{E}_1, \dots, \mathbf{B}^{-1}\mathbf{E}_n) &= g(\mathbf{ABB}^{-1}\mathbf{E}_1, \dots, \mathbf{ABB}^{-1}\mathbf{E}_n) \\ &= g(\mathbf{AE}_1, \dots, \mathbf{AE}_n) \\ &= \det \mathbf{A}g(\mathbf{E}_1, \dots, \mathbf{E}_n) = \det \mathbf{A}, \end{aligned}$$

and, on the other hand,

$$\begin{aligned} dg_{\mathbf{B}}(\mathbf{B}^{-1}\mathbf{E}_1, \dots, \mathbf{B}^{-1}\mathbf{E}_n) &= dg(\mathbf{BB}^{-1}\mathbf{E}_1, \dots, \mathbf{BB}^{-1}\mathbf{E}_n) \\ &= dg(\mathbf{E}_1, \dots, \mathbf{E}_n) = d, \end{aligned}$$

and so equating yields  $d = \det \mathbf{A}$  as required.

(6) We require a lemma.

**Lemma.** *Let*

$$\mathbf{H}_i: \mathbb{R}^{n-1} \rightarrow \mathbb{R}^n | H_i(x_1, \dots, x_{n-1}) = (x_1, \dots, x_{i-1}, 0, x_1, \dots, x_n),$$

and let

$$g_n: \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_n \rightarrow \mathbb{R},$$

and

$$g_{n-1}: \underbrace{\mathbb{R}^{n-1} \times \dots \times \mathbb{R}^{n-1}}_{n-1} \rightarrow \mathbb{R}$$

be the fundamental sign-symmetric/alternating forms. Define

$$g_{n-1}^i: \underbrace{\mathbb{R}^{n-1} \times \dots \times \mathbb{R}^{n-1}}_{n-1} \rightarrow \mathbb{R}$$

by

$$g_{n-1}^i(\mathbf{A}_1, \dots, \mathbf{A}_{n-1}) = g_n(\mathbf{H}_i \mathbf{A}_1, \dots, \mathbf{H}_i \mathbf{A}_{n-1}, \mathbf{E}_i).$$

Then

$$g_{n-1}^i = (-1)^{n-i} g_{n-1}.$$

PROOF. It is enough to show

$$g_{n-1}^i(\mathbf{E}_1, \dots, \mathbf{E}_{n-1}) = (-1)^{i-1}.$$

Note that

$$H_i(\mathbf{E}_j) = \begin{cases} \mathbf{E}_j, & j \leq i-1, \\ \mathbf{E}_{j+1}, & j \geq i, \end{cases}$$

so by definition

$$\begin{aligned} g_{n-1}^i(\mathbf{E}_1, \dots, \mathbf{E}_{n-1}) &= g_n(\mathbf{E}_1, \dots, \mathbf{E}_{i-1}, \mathbf{E}_{i+1}, \dots, \mathbf{E}_n, \mathbf{E}_i) \\ n - \text{interchanges} \left\{ \begin{aligned} &= -g_n(\mathbf{E}_1, \dots, \mathbf{E}_{i-1}, \mathbf{E}_{i+1}, \dots, \mathbf{E}_i, \mathbf{E}_n) \\ &= (-1)^{i-1} g_n(\mathbf{E}_1, \dots, \mathbf{E}_n) = (-1)^{n-i}, \end{aligned} \right. \end{aligned}$$

as required. □

Let  $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_n) \in \mathcal{M}_{nn}$  where

$$\mathbf{A}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}, \quad i = 1, \dots, n,$$

are the column of  $\mathbf{A}$ . Then

$$\mathbf{A}_i^t = \sum_{k=1}^n a_{ki} \mathbf{E}_k = \mathbf{B}_i + a_{ni} \mathbf{E}_n$$

which defines  $\mathbf{B}_1, \dots, \mathbf{B}_n \in \text{Span}\{\mathbf{E}_1, \dots, \mathbf{E}_{n-1}\}$ . By definition we have

$$\begin{aligned} \det \mathbf{A} &= g(\mathbf{A}_1^t, \dots, \mathbf{A}_n^t) \\ &= g(\mathbf{B}_1 + a_{n1} \mathbf{E}_n, \dots, \mathbf{B}_n + a_{nn} \mathbf{E}_n). \end{aligned}$$

We are going to expand this last expression using the multilinearity of  $g(\dots)$ . Among the terms of this expansion will be

$$g(\mathbf{B}_1, \dots, \mathbf{B}_n).$$

This term is however zero. To see this notice that

$$g(\mathbf{B}_1, \dots, \mathbf{B}_n) = \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n-1,1} & \dots & a_{n-1,n} \\ 0 & \dots & 0 \end{pmatrix}.$$

The matrix on the right is obviously singular because it has a row of zeros. By property (8) we conclude that the determinant of this matrix is zero, so

$$g(\mathbf{B}_1, \dots, \mathbf{B}_n) = 0.$$

Another typical term of the expansion is

$$g(\mathbf{B}_1, \dots, \mathbf{B}_{i-1}, a_{ni} \mathbf{E}_n, \mathbf{B}_{i+1}, \dots, a_{nj} \mathbf{E}_n, \mathbf{B}_{j+1}, \dots, \mathbf{B}_n).$$

This term also vanishes because by multilinearity and sign-symmetry it is equal to

$$a_{ni} a_{nj} g(\mathbf{B}_1, \dots, \mathbf{B}_{i-1}, \mathbf{E}_n, \mathbf{B}_{i+1}, \dots, \mathbf{B}_{j-1}, \mathbf{E}_n, \mathbf{B}_{j+1}, \dots, \mathbf{B}_n) = a_{ni} a_{nj} \cdot 0 = 0.$$

These considerations then lead to the following formula

$$\det \mathbf{A} = g(a_{n1} \mathbf{E}_n, \mathbf{B}_2, \dots, \mathbf{B}_n) + g(\mathbf{B}_1, a_{n2} \mathbf{E}_n, \mathbf{B}_3, \dots, \mathbf{B}_n) + \dots + g(\mathbf{B}_1, \dots, \mathbf{B}_{n-1} + a_{nn} \mathbf{E}_n).$$

So by linearity and the preceding lemma we get

$$\begin{aligned} \det \mathbf{A} &= a_{n1} g(\mathbf{E}_n, \mathbf{B}_2, \dots, \mathbf{B}_n) + \dots + a_{nn} g(\mathbf{B}_1, \dots, \mathbf{B}_{n-1}, \mathbf{E}_n) \\ &= \sum_{i=1}^n (-1)^{n-1} a_{ni} g(\mathbf{B}_1, \dots, \mathbf{B}_{i-1}, \mathbf{B}_{i+1}, \dots, \mathbf{B}_n) \\ &= \sum_{i=1}^n (-1)^{n-i} a_{ni} \det \mathbf{M}_{ni}, \end{aligned}$$

by the definition of the  $(n, i)$  minor of  $\mathbf{A}$  and the sign-symmetry of  $g$ . This formula is the *Lagrange expansion formula with respect to the  $n$ th row*. There is a similar formula for  $\det \mathbf{A}$  that employs the  $n$ th column of  $\mathbf{A}$ . To obtain this we write

$$\begin{aligned} \det \mathbf{A} &= g(\mathbf{A}'_1, \dots, \mathbf{A}'_n) = g\left(\mathbf{A}'_1, \dots, \mathbf{A}'_{n-1}, \sum_{k=1}^n a_{kn} \mathbf{E}_k\right) \\ &= \sum_{k=1}^n a_{kn} g(\mathbf{A}'_1, \dots, \mathbf{A}'_{n-1}, \mathbf{E}_k). \end{aligned}$$

Let us now analyze the individual terms of this sum. For  $1 \leq k \leq n$  we write

$$\mathbf{A}'_i = \mathbf{C}_{k,i} + a_{ki} \mathbf{E}_k.$$

Then reasoning as above, and using the fact that

$$g(\mathbf{C}_{k,1}, \dots, \mathbf{C}_{k,n}) = 0,$$

since it may be interpreted as the determinant of a matrix with a column of zeros, and

$$g(\mathbf{C}_{k,1}, \dots, a_{ki} \mathbf{E}_k, \mathbf{C}_{k,i+1}, \dots, a_{kj} \mathbf{E}_k, \mathbf{C}_{k,j+1}, \dots, \mathbf{C}_{k,n})$$

vanishes by sign-symmetry, we get the formula

$$\begin{aligned} \det \mathbf{A} &= \sum_{k=1}^n a_{kn} g(\mathbf{C}_{k,1}, \dots, \mathbf{C}_{k,n-1}, \mathbf{E}_k) \\ &= \sum_{k=1}^n (-1)^{n-k} a_{kn} g(\mathbf{C}_{k,1}, \dots, \mathbf{C}_{k,n-1}) \\ &= \sum_{k=1}^n (-1)^{n-k} a_{kn} \det(\mathbf{M}_{kn}), \end{aligned}$$

which is the *Lagrange expansion formula with respect to the  $n$ th column*.

We may now prove (6) by observing that the Lagrange expansion with respect to the  $n$ th column of  $A'$  is the same as the Lagrange expansion of  $A$  with respect to the  $n$ th row. But the Lagrange expansions compute  $\det$ , so

$$\begin{aligned}\det A' &= \text{Lagrange expansion of } A' \text{ with respect to the } n\text{th column} \\ &= \text{Lagrange expansion of } A \text{ with respect to the } n\text{th row} \\ &= \det A,\end{aligned}$$

as required.  $\square$

Let us derive an explicit form for the determinant of a  $2 \times 2$  matrix. Let

$$A = (A_1, A_2),$$

where

$$A_i = \begin{pmatrix} a_{1i} \\ a_{2i} \end{pmatrix}, \quad i = 1, 2,$$

are the columns of  $A$ . As in the discussion preceding the proof of (14<sup>bis</sup>.5)(7) we see

$$\begin{aligned}\det A &= g(A'_1, A'_2) \\ &= g(a_{11}E_1 + a_{21}E_2, a_{12}E_1 + a_{22}E_2) \\ &= a_{11}g(E_1, a_{12}E_1 + a_{22}E_2) + a_{21}g(E_2, a_{12}E_1 + a_{22}E_2) \\ &= a_{11}(a_{12}g(E_1, E_1) + a_{22}g(E_1, E_2)) \\ &\quad + a_{21}(a_{12}g(E_2, E_1) + a_{22}g(E_2, E_2)) \\ &= a_{11}(0 + a_{22}) + a_{21}(-a_{12} + 0) \\ &= a_{11}a_{22} - a_{21}a_{12},\end{aligned}$$

because

$$\begin{aligned}g(E_1, E_1) &= 0 = g(E_2, E_2), \\ g(E_1, E_2) &= 1 = -g(E_2, E_1).\end{aligned}$$

Thus

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

exactly as we had already defined the determinant in Chapter 13 (immediately preceding Example 11) for  $2 \times 2$  matrices.

It should be apparent that one can employ the preceding procedure to arrive at a formula for  $\det A$ , where  $A$  is an  $n \times n$  matrix. Such a formula would look like

$$(*) \quad \det A = \sum (-1)^{\sigma(j_1, \dots, j_n)} a_{j_1 1} \cdots a_{j_n n},$$

where each term in the sum

$$a_{j_1 1} \cdots a_{j_n n}$$



which is a product of matrix elements, one element chosen out of each column, so that no two elements lie in the same row, and  $\sigma(j_1, \dots, j_n)$  is the number of pairs of indices that must be interchanged to bring  $(\mathbf{E}_{j_1}, \dots, \mathbf{E}_{j_n})$  to the form  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$ .

*Note carefully.* There are as many terms in the sum (\*) as there are possible rearrangements of  $(\mathbf{E}_1, \dots, \mathbf{E}_n)$ ; or, what is the same thing,  $(1, \dots, n)$ . An elementary counting argument shows that the number of such arrangements is

$$n! = n(n-1)(\dots)(2)(1).$$

So for a  $5 \times 5$  matrix the preceding sum contains

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

terms. Thus the formula (\*) is not particularly efficacious for computations.

We close our discussion of determinants with the Lagrange expansion formula.

**Proposition 14<sup>bis</sup>.8.** *If  $\mathbf{A}$  is a square matrix of size  $n$ , then*

$$\det \mathbf{A} = \sum_{j=1}^n (-1)^{i+j} a_{ji} \det \mathbf{M}_{ji}$$

(see the definition preceding Example 7, Chapter 14 for the definition of the minor  $M_{ij}$ ).

**PROOF.** This follows from the Lagrange expansion formula with respect to the  $n$ th column, which we proved in (14<sup>bis</sup>.5)(6), by interchanging columns.  $\square$

## EXERCISES

1. Find a nonsingular bilinear form

$$f: \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}^4.$$

2. Let  $\mathcal{V}_1, \dots, \mathcal{V}_n$  be subspaces of  $\mathcal{W}$  and

$$\sigma: \mathcal{V}_1 \times \dots \times \mathcal{V}_n \rightarrow \mathcal{W}$$

the multilinear form

$$\sigma(v_1, \dots, v_n) = v_1 + \dots + v_n.$$

Show that  $\sigma$  is nonsingular iff

$$\mathcal{V}_i \cap \mathcal{V}_j = \{0\}, \quad i = j.$$

3. Show that the bilinear forms of Examples 3 and 4 are isomorphic.
4. (a) Show that  $b(k, n) \geq 1$ .  
(b) If  $k$  and  $n$  are even show  $b(k, n) \geq 2$ .

5. (a) Compute the determinant of each of the following matrices

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ 4 & -1 & -1 \end{pmatrix},$$

(b) which of the preceding matrices are invertible?

6. (a) Suppose  $\mathbf{A}, \mathbf{D} \in \mathcal{M}_{rr}$ ,  $\mathbf{B}, \mathbf{C} \in \mathcal{M}_{ss}$  and let  $\mathbf{M}$  be the matrix

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}.$$

Is it true in general that

$$(*) \quad \det \mathbf{M} = \det \mathbf{A} \det \mathbf{D} - \det \mathbf{B} \det \mathbf{C}?$$

(Review some of the examples in Exercise 5(a).)

(b) Suppose  $\mathbf{A}, \mathbf{D}$  and  $\mathbf{B}, \mathbf{C}$  commute with each other. Show that  $(*)$  holds

7. (a) Let

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

The *diagonals* of  $\mathbf{A}$  are

$$\begin{pmatrix} a_{11} & & & & \\ & a_{22} & & & \\ & & a_{33} & & \\ & & & a_{31} & \\ & & & & a_{32} \end{pmatrix},$$

and the *antidiagonals* are

$$\begin{pmatrix} & & & & \\ & & & a_{12} & \\ & a_{22} & & & \\ & & a_{21} & & \\ a_{31} & & & a_{32} & \end{pmatrix}.$$

Show that the determinant of  $\mathbf{A}$  is the sum of the products of the diagonal elements minus the sum of the products of the antidiagonal elements.

(b) Show by an example that the analogous rule is false for  $4 \times 4$  matrices.

8. Let  $\mathbf{A} = (a_1, a_2)$ ,  $\mathbf{B} = (a_1, b_2) \in \mathbb{R}^2$ ,  $\mathbf{A} \neq \mathbf{B}$ . Show that the equation of the line through  $\mathbf{A}$  and  $\mathbf{B}$  is given by

$$\det \begin{pmatrix} x & y & 1 \\ a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \end{pmatrix} = 0.$$

9. Show that the area of the triangle in  $\mathbb{R}^2$  with vertices  $(a_1, b_1)$ ,  $(a_2, b_2)$ ,  $(a_3, b_3)$  is given by

$$\det \begin{pmatrix} a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \\ a_3 & b_3 & 1 \end{pmatrix}.$$

10. If  $\mathbf{A} \in \mathcal{M}_n$  is skew-symmetric and  $n$  is odd show  $\det \mathbf{A} = 0$ .

11. Let  $\sigma$  be a *permutation* of  $(1, \dots, n)$ , that is a rearrangement  $(\sigma(1), \dots, \sigma(n))$  of  $(1, \dots, n)$  in perhaps some other order. Define  $\text{sgn}(\sigma)$  by

$$\prod_{i < j} (x_{\sigma(i)} - x_{\sigma(j)}) = \text{sgn}(\sigma) \Delta(x),$$

where

$$\Delta(x) = \prod_{i < j} (x_i - x_j).$$

- (a) Show that  $\text{sgn}(\sigma) = -1$  if  $\sigma$  is the permutation where

$$\sigma(k) = \begin{cases} k, & k \neq i, j, \\ j, & k = i, \\ i, & k = j, \end{cases}$$

such a permutation is simply the interchange of  $i$  and  $j$ .

- (b) Show any permutation can be accomplished by a succession of interchanges.  
 (c) Show  $\text{sgn}(\sigma) = (-1)^s$  where  $s$  is the number of interchanges required to accomplish  $\sigma$ .  
 (d) Let  $\sigma$  be a permutation and  $L(\sigma): \mathbb{R}^n \rightarrow \mathbb{R}^n$  by taking the linear extension of

$$L(\sigma)(\mathbf{E}_i) = \mathbf{E}_{\sigma(i)}, \quad i = 1, \dots, n.$$

The matrix of  $L(\sigma)$  with respect to the standard basis of  $\mathbb{R}^n$  is called a *permutation matrix*. Write down all the permutation matrices of size 3 and compute their determinants.

- (e) Show  $\det L(\sigma) = \text{sgn}(\sigma)$ .

12. Let  $\mathbf{A} \in \mathcal{M}_n$ . Show

$$\det \mathbf{A} = \sum (-1)^{\text{sgn}(\sigma)} a_{\sigma(1)} \cdots a_{\sigma(n)},$$

where the sum extends over all permutations of  $(1, \dots, n)$ .

13. (a) Evaluate

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}.$$

(b) Show

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & & x_n^{n-1} \end{pmatrix} = \prod_{i < j} (x_i - x_j).$$

14. Let  $\mathbf{A} \in \mathcal{M}_{mn}$ ,  $\mathbf{B} \in \mathcal{M}_{nm}$  where  $m > n$ . Show  $\det(\mathbf{AB}) = 0$ .  
(Note:  $\mathbf{A} \cdot \mathbf{B}$  is square of size  $m \times m$ .)

# 15 Inner product spaces

So far in our study of vector spaces and linear transformations we have made no use of the notions of length and angle, although these concepts play an important role in our intuition for the vector algebra of  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . In fact the length of a vector and the angle between two vectors play very important parts in the further development of linear algebra and it is now time to introduce these ingredients into our study. There are many ways to do this and in the approach that we will follow both length and angle will be derived from a more fundamental concept called a *scalar* or *inner* product of two vectors. No doubt the student has encountered the scalar product in the guise of the *dot* product of two vectors in  $\mathbb{R}^3$  which is usually defined by the equation

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos \theta$$

where  $|\mathbf{A}|$  is the length of the vector  $\mathbf{A}$ , similarly for  $\mathbf{B}$ , and  $\theta$  is the *angle* between  $\mathbf{A}$  and  $\mathbf{B}$ . In the study of vectors in  $\mathbb{R}^3$  this is a reasonable way to introduce the scalar product, because lengths and angles are already defined and well-studied concepts of geometry. In a more abstract study of linear algebra, such as we are undertaking, such an approach is not possible, for what is the length of a polynomial (vector) in  $\mathcal{P}_4(\mathbb{R})$ ? This problem can be overcome by the use of the axiomatic method. Having introduced vector spaces by axioms it is not at all unreasonable to employ additional axioms to impose further structure on them.

**Definition.** A *scalar* (or *inner*) *product* on a vector space  $\mathcal{V}$  is a function which assigns to each pair of vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$  a real number, denoted by  $\langle \mathbf{A}, \mathbf{B} \rangle$  having the following properties:

- (1)  $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{B}, \mathbf{A} \rangle$  for all  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ .
- (2) For all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$  and numbers  $r$  in  $\mathbb{R}$  we have  $\langle r\mathbf{A}, \mathbf{B} \rangle = r\langle \mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{A}, r\mathbf{B} \rangle$ .
- (3) For all vectors  $\mathbf{A}, \mathbf{B}$ , and  $\mathbf{C}$  in  $\mathcal{V}$  we have
  - (a)  $\langle \mathbf{A} + \mathbf{B}, \mathbf{C} \rangle = \langle \mathbf{A}, \mathbf{C} \rangle + \langle \mathbf{B}, \mathbf{C} \rangle$
  - (b)  $\langle \mathbf{A}, \mathbf{B} + \mathbf{C} \rangle = \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{A}, \mathbf{C} \rangle$ .
- (4) For all  $\mathbf{A}$  in  $\mathcal{V}$ ,  $\langle \mathbf{A}, \mathbf{A} \rangle \geq 0$  and  $\langle \mathbf{A}, \mathbf{A} \rangle = 0$  iff  $\mathbf{A} = \mathbf{0}$ .

Notice in reading the above axioms that we have used the notations  $+$ ,  $0$ , etc. for two different ideas, namely  $+$  for vectors and  $+$  for numbers, the vector “zero” and the number “zero,” etc. This should be taken into account in reading Axioms (2), (3), and (4).

It is not difficult to verify that for the dot product

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos \theta$$

defined in  $\mathbb{R}^3$  the above axioms are satisfied. The only condition required of that is not immediately clear is (3). To verify (3, a) consider Figure 15.1 where  $\theta$  is the angle between  $\mathbf{A}$  and  $\mathbf{B}$  and the angles with subscripts denote the angle between the vector subscript and  $\mathbf{C}$ . Now use the law of cosines. Thus the scalar product extends to general vector spaces the dot product of everyday experience.

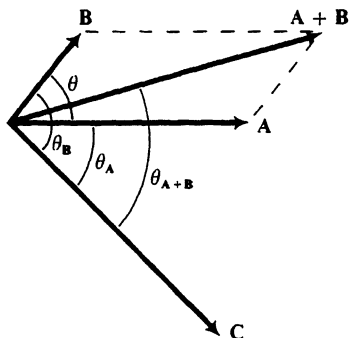


Figure 15.1

Why *scalar* product? That is, why is the concept introduced in the definition above called a *scalar* product? The answer should be apparent: because it associates to a pair of vectors a scalar (= real number). (Compare to Exercise 18 of Chapter 14.)

The notion of inner product can also be introduced into a complex vector space (see Exercise 17 at the end of this chapter where *a priori* we have not a single example where length and angle have any intuitive content.

EXAMPLE 1. *The standard scalar product on  $\mathbb{R}^n$ .*

For two vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathbb{R}^n$  we define

$$\langle \mathbf{A}, \mathbf{B} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

where  $\mathbf{A} = (a_1, \dots, a_n)$ ,  $\mathbf{B} = (b_1, \dots, b_n)$ . We assert that  $\langle \mathbf{A}, \mathbf{B} \rangle$  is a scalar product. To verify this assertion we must check that Axioms (1)–(4) of the definition of a scalar product  $\langle \cdot, \cdot \rangle$  are satisfied for the function  $\langle \mathbf{A}, \mathbf{B} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$ .

Let's see, we have

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle &= a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = b_1 a_1 + b_2 a_2 + \cdots + b_n a_n \\ &= \langle \mathbf{B}, \mathbf{A} \rangle \end{aligned}$$

so that (1) holds. Next note that

$$\begin{aligned} \langle r\mathbf{A}, \mathbf{B} \rangle &= r a_1 b_1 + r a_2 b_2 + \cdots + r a_n b_n \\ &= r(a_1 b_1 + a_2 b_2 + \cdots + a_n b_n) = r \langle \mathbf{A}, \mathbf{B} \rangle \\ &= a_1 r b_1 + a_2 r b_2 + \cdots + a_n r b_n = \langle \mathbf{A}, r\mathbf{B} \rangle \end{aligned}$$

so that Axiom (2) also holds. To verify (3a) we compute as follows

$$\begin{aligned} \langle \mathbf{A} + \mathbf{B}, \mathbf{C} \rangle &= (a_1 + b_1)c_1 + (a_2 + b_2)c_2 + \cdots + (a_n + b_n)c_n \\ &= a_1 c_1 + b_1 c_1 + a_2 c_2 + b_2 c_2 + \cdots + a_n c_n + b_n c_n \\ &= a_1 c_1 + a_2 c_2 + \cdots + a_n c_n + b_1 c_1 + b_2 c_2 + \cdots + b_n c_n \\ &= \langle \mathbf{A}, \mathbf{C} \rangle + \langle \mathbf{B}, \mathbf{C} \rangle. \end{aligned}$$

The verification of (3b) is similar. Finally note that

$$\langle \mathbf{A}, \mathbf{A} \rangle = a_1 a_1 + a_2 a_2 + \cdots + a_n a_n = a_1^2 + a_2^2 + \cdots + a_n^2$$

and since a sum of squares is always nonnegative, and is zero iff all the terms are zero we see that (4) is also satisfied.

Therefore

$$\langle \mathbf{A}, \mathbf{B} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

defines a scalar product on  $\mathbb{R}^n$  called the standard scalar product of  $\mathbb{R}^n$ .

EXAMPLE 2. *The standard scalar product in  $\mathcal{V}$  with basis.*

The preceding example may be generalized to any finite-dimensional vector space  $\mathcal{V}$  with a **fixed chosen** basis  $\{\mathbf{F}_1, \dots, \mathbf{F}_n\}$ . Namely if  $\mathbf{A}$  and  $\mathbf{B}$  belong to  $\mathcal{V}$  then we know that

$$\begin{aligned} \mathbf{A} &= a_1 \mathbf{F}_1 + \cdots + a_n \mathbf{F}_n \\ \mathbf{B} &= b_1 \mathbf{F}_1 + \cdots + b_n \mathbf{F}_n \end{aligned}$$

for unique  $n$ -tuples  $(a_1, \dots, a_n)$ ,  $(b_1, \dots, b_n)$ ; the components of  $\mathbf{A}$  and  $\mathbf{B}$  relative to the basis  $\{\mathbf{F}_1, \dots, \mathbf{F}_n\}$  (see (6.6)). Now define

$$\langle \mathbf{A}, \mathbf{B} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

The same computations made in Example 1 show that  $\langle \mathbf{A}, \mathbf{B} \rangle$  is a scalar product on  $\mathcal{V}$ . Notice that this scalar product depends on the choice of basis  $\{\mathbf{F}_1, \dots, \mathbf{F}_n\}$ . A different choice of basis will give a different scalar product. For example, if in  $\mathbb{R}^2$  we choose the standard basis and use the above definition then we obtain the standard scalar product of  $\mathbb{R}^2$ . On the other hand we could use the basis  $\mathbf{F}_1 = (1, 0)$ ,  $\mathbf{F}_2 = (1, 1)$  in which case we would obtain an inner product  $\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{F}}$  where

$$\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{F}} = a_1 b_1 - a_1 b_2 - b_2 a_1 + 2a_2 b_2$$

where  $\mathbf{A} = (a_1, a_2) = a_1 \mathbf{E}_1 + a_2 \mathbf{E}_2$  and  $\mathbf{B} = (b_1, b_2) = b_1 \mathbf{E}_1 + b_2 \mathbf{E}_2$ . Thus many “different” scalar products seem possible on  $\mathcal{V}$ , one for each choice of bases. This difference is more apparent than real as we shall presently see.

EXAMPLE 3. A scalar product on  $\mathcal{P}_k(\mathbb{R})$ .

For the space  $\mathcal{P}_k(\mathbb{R})$  of polynomials of degree at most  $k$  we may define the scalar product

$$\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx.$$

Certainly this definition is not at all like those of Example 1 or 2. In fact it is much more natural than those because it does not refer to a fixed basis. To verify the scalar product properties we freely use facts about integration from the calculus:

$$(1) \quad \langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx = \int_0^1 q(x)p(x)dx = \langle q(x), p(x) \rangle$$

$$(2) \quad \begin{aligned} \langle rp(x), q(x) \rangle &= \int_0^1 rp(x)q(x)dx = r \int_0^1 p(x)q(x)dx = r \langle p(x), q(x) \rangle \\ &= \int_0^1 p(x)r q(x)dx = \langle p(x), r q(x) \rangle. \end{aligned}$$

$$(3a) \quad \begin{aligned} \langle p(x) + q(x), r(x) \rangle &= \int_0^1 (p(x) + q(x))r(x)dx \\ &= \int_0^1 (p(x)r(x) + q(x)r(x))dx \\ &= \int_0^1 p(x)r(x)dx + \int_0^1 q(x)r(x)dx \\ &= \langle p(x), r(x) \rangle + \langle q(x), r(x) \rangle, \end{aligned}$$

and similarly for (3b).

(4) Recall that for a nonnegative continuous function  $f(x)$

$$\int_0^1 f(x)dx \geq 0$$



and

$$\int_0^1 f(x)dx = 0 \quad \text{iff } f(x) = 0 \text{ for all } x \in [0, 1].$$

Note

$$\langle p(x), p(x) \rangle = \int_0^1 (p(x))^2 dx$$

and that  $p(x)^2$  is nonnegative. Hence

$$\langle p(x), p(x) \rangle = \int_0^1 (p(x))^2 dx \geq 0.$$

Finally suppose that  $\langle p(x), p(x) \rangle = 0$ . Then

$$\int_0^1 (p(x))^2 dx = 0 \quad \text{implies} \quad p(x) = 0 \quad x \in [0, 1].$$

A nonzero polynomial of degree  $k$  has at most  $k$  roots, but  $p(x) = 0$  for all  $x \in [0, 1]$  so has lots more than  $k$  roots. Therefore  $p(x) = 0$ .

Thus the axioms for a scalar product on  $\mathcal{P}_k(\mathbb{R})$  are satisfied by the formula

$$\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx.$$

**EXAMPLE 4.** *The trace scalar product on  $\mathcal{M}_{nn}$ .*

Let  $\mathcal{M}_{nn}$  denote the vector space of  $n \times n$  matrices. Recall that  $\mathbf{E}_{rs}$  is the  $n \times n$  matrix defined by

$$\mathbf{E}_{rs} = (e_{ij}), \quad e_{ij} = \begin{cases} 1 & \text{if } i = r, j = s, \\ 0 & \text{otherwise.} \end{cases}$$

The vectors  $\{\mathbf{E}_{rs}\}$  form a basis for  $\mathcal{M}_{nn}$  (see (11.2)). Therefore

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} a_{ij}b_{ij} \quad \mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij})$$

defines a scalar product on  $\mathcal{M}_{nn}$  corresponding to the basis  $\{\mathbf{E}_{rs}\}$ . There is another way to write this scalar product that is sort of interesting.

**Definition.** If  $\mathbf{A} = (a_{ij})$  is an  $n \times n$  matrix, the *trace* of  $\mathbf{A}$  is the number  $\text{tr}(\mathbf{A})$  defined by

$$\text{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Thus  $\text{tr}(\mathbf{A})$  is the sum of the diagonal entries of  $\mathbf{A}$ . For example ( $n = 3$ )

$$\text{tr} \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} = 15,$$

and

$$\text{tr} \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = n = \text{tr}(\mathbf{I})$$

where  $\mathbf{I} \in \mathcal{M}_{nn}$  is the identity matrix.

One way in which the trace and the inner product of Example 4 are connected is by the following formula

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{AB}')$$

To verify this we need only compute the right-hand side and compare it to the definition of  $\langle \mathbf{A}, \mathbf{B} \rangle$ . We have

$$\text{tr}(\mathbf{AB}') = \sum_i \sum_j a_{ij} b'_{ji}$$

because the entry in the  $(i, i)$  position of the product matrix  $\mathbf{AB}'$  is  $\sum_j a_{ij} b'_{ji}$

$$\begin{aligned} &= \sum_i \sum_j a_{ij} b_{ij} \\ &= \sum_{i,j} a_{ij} b_{ij} = \langle \mathbf{A}, \mathbf{B} \rangle \end{aligned}$$

as we claimed.

The trace of a square matrix is quite a handy concept and although we will not make use of it, it does play a rather important role in further developments of linear algebra.

**Definition.** A vector space  $\mathcal{V}$  equipped with a fixed scalar product  $\langle \cdot, \cdot \rangle$  is called an *inner product space*. (Don't ask me why it is called an inner product space and not a scalar product space, I don't know. Because of this strange turn of the terminology, we will often call a scalar product an *inner product*.)

Let us turn now to the concepts of length and angle in an inner product space.

**Definition.** Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space and  $\mathbf{A}$  a vector of  $\mathcal{V}$ . The *length* of  $\mathbf{A}$ , denoted by  $|\mathbf{A}|$ , is the nonnegative number

$$|\mathbf{A}| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}.$$

(Notice by Axiom (4) of the definition of scalar product that  $\langle \mathbf{A}, \mathbf{A} \rangle \geq 0$  for all vectors  $\mathbf{A}$  in  $\mathcal{V}$  and hence  $\langle \mathbf{A}, \mathbf{A} \rangle$  has a unique nonnegative square root  $\sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$  which we have defined to be  $|\mathbf{A}|$ .)

**Proposition 15.1.** Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space. Then the length function  $|\cdot|$  has the following properties:

- (1) For each  $\mathbf{A}$  in  $\mathcal{V}$ ,  $|\mathbf{A}| \geq 0$  and  $|\mathbf{A}| = 0$  iff  $\mathbf{A} = \mathbf{0}$ .
- (2) For each  $\mathbf{A}$  in  $\mathcal{V}$  and number  $r$ ,  $|r\mathbf{A}| = |r||\mathbf{A}|$ .

(As is becoming usual in our use of notations we have written  $|\cdot|$  for two different concepts above. In (2)  $|r\mathbf{A}|$  is the length of the vector  $r\mathbf{A}$  while  $|r|$  is the absolute value of the number  $r$ . Actually no confusion should arise because if we recall that  $\mathbb{R} = \mathbb{R}^1$  is a vector space, then the absolute value  $|r|$  is just the length of  $r$  when  $\mathbb{R}^1$  is given the standard scalar product.)

**PROOF.** Condition (1) is just a translation of Axiom (4) for scalar products while the computation

$$\begin{aligned} |r\mathbf{A}| &= \sqrt{\langle r\mathbf{A}, r\mathbf{A} \rangle} = \sqrt{r\langle \mathbf{A}, r\mathbf{A} \rangle} \\ &= \sqrt{r^2 \langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{r^2} \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} \\ &= |r||\mathbf{A}| \end{aligned}$$

verifies condition (2). □

**Proposition 15.2** (Schwarz inequality). Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space. Then for any pair of vectors  $\mathbf{A}$  and  $\mathbf{B}$  in  $\mathcal{V}$ ,

$$|\langle \mathbf{A}, \mathbf{B} \rangle| \leq |\mathbf{A}||\mathbf{B}|.$$

**PROOF.** We begin by observing that for  $\mathbf{A} = \mathbf{0}$  we have

$$|\langle \mathbf{A}, \mathbf{B} \rangle| = 0 \quad \text{and} \quad |\mathbf{A}||\mathbf{B}| = 0$$

so that the inequality is true in this case. We may therefore assume that  $\mathbf{A} \neq \mathbf{0}$  and hence that  $|\mathbf{A}| \neq 0$ . Let

$$\mathbf{C} = \mathbf{B} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A}.$$

Then

$$\begin{aligned} 0 \leq \langle \mathbf{C}, \mathbf{C} \rangle &= \left\langle \mathbf{B} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A}, \mathbf{B} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A} \right\rangle \\ &= \left\langle \mathbf{B} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A}, \mathbf{B} \right\rangle - \left\langle \mathbf{B} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A}, \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \mathbf{A} \right\rangle \\ &= \langle \mathbf{B}, \mathbf{B} \rangle - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \langle \mathbf{A}, \mathbf{B} \rangle - \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} \langle \mathbf{B}, \mathbf{A} \rangle \\ &\quad + \frac{\langle \mathbf{A}, \mathbf{B} \rangle^2}{\langle \mathbf{A}, \mathbf{A} \rangle^2} \langle \mathbf{A}, \mathbf{A} \rangle \\ &= \langle \mathbf{B}, \mathbf{B} \rangle - \frac{\langle \mathbf{A}, \mathbf{B} \rangle^2}{\langle \mathbf{A}, \mathbf{A} \rangle} - \frac{\langle \mathbf{A}, \mathbf{B} \rangle \langle \mathbf{A}, \mathbf{B} \rangle}{\langle \mathbf{A}, \mathbf{A} \rangle} + \frac{\langle \mathbf{A}, \mathbf{B} \rangle^2}{\langle \mathbf{A}, \mathbf{A} \rangle} \\ &= \langle \mathbf{B}, \mathbf{B} \rangle - \frac{\langle \mathbf{A}, \mathbf{B} \rangle^2}{\langle \mathbf{A}, \mathbf{A} \rangle}. \end{aligned}$$

That is

$$\langle \mathbf{B}, \mathbf{B} \rangle - \frac{\langle \mathbf{A}, \mathbf{B} \rangle^2}{\langle \mathbf{A}, \mathbf{A} \rangle} \geq 0$$

Multiply both sides by  $\langle \mathbf{A}, \mathbf{A} \rangle$ , which is nonnegative and hence does not change the sense of the inequality, so we get

$$\langle \mathbf{A}, \mathbf{A} \rangle \langle \mathbf{B}, \mathbf{B} \rangle - \langle \mathbf{A}, \mathbf{B} \rangle^2 \geq 0$$

which may be rewritten as

$$|\mathbf{A}|^2 |\mathbf{B}|^2 \geq \langle \mathbf{A}, \mathbf{B} \rangle^2$$

taking positive square roots then gives

$$|\mathbf{A}| |\mathbf{B}| \geq |\langle \mathbf{A}, \mathbf{B} \rangle|$$

as required.  $\square$

**Proposition 15.3** (The triangle inequality). *Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space. Then for any vectors  $\mathbf{A}$  and  $\mathbf{B}$  in  $\mathcal{V}$  we have*

$$|\mathbf{A} + \mathbf{B}| \leq |\mathbf{A}| + |\mathbf{B}|.$$

**PROOF.** We compute as follows

$$\begin{aligned} |\mathbf{A} + \mathbf{B}|^2 &= \langle \mathbf{A} + \mathbf{B}, \mathbf{A} + \mathbf{B} \rangle \\ &= \langle \mathbf{A}, \mathbf{A} + \mathbf{B} \rangle + \langle \mathbf{B}, \mathbf{A} + \mathbf{B} \rangle \\ &= \langle \mathbf{A}, \mathbf{A} \rangle + \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{B}, \mathbf{A} \rangle + \langle \mathbf{B}, \mathbf{B} \rangle \\ &= |\mathbf{A}|^2 + \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{A}, \mathbf{B} \rangle + |\mathbf{B}|^2 \\ &= |\mathbf{A}|^2 + 2\langle \mathbf{A}, \mathbf{B} \rangle + |\mathbf{B}|^2 \\ &\leq |\mathbf{A}|^2 + 2|\langle \mathbf{A}, \mathbf{B} \rangle| + |\mathbf{B}|^2 \\ &\leq |\mathbf{A}|^2 + 2|\mathbf{A}| |\mathbf{B}| + |\mathbf{B}|^2 = (|\mathbf{A}| + |\mathbf{B}|)^2. \end{aligned}$$

That is,

$$|\mathbf{A} + \mathbf{B}|^2 \leq (|\mathbf{A}| + |\mathbf{B}|)^2$$

and taking positive square roots gives the desired conclusion.  $\square$

To see why (15.3) is called the *triangle inequality* let us define the *distance between* two vectors  $\mathbf{A}$  and  $\mathbf{B}$  in an inner product space  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  by the formula

$$d(\mathbf{A}, \mathbf{B}) = |\mathbf{A} - \mathbf{B}|.$$

The distance function  $d$  is easily seen to satisfy:

$$d(\mathbf{A}, \mathbf{B}) \geq 0$$

and

$$d(\mathbf{A}, \mathbf{B}) = 0 \quad \text{iff } \mathbf{A} = \mathbf{B}.$$

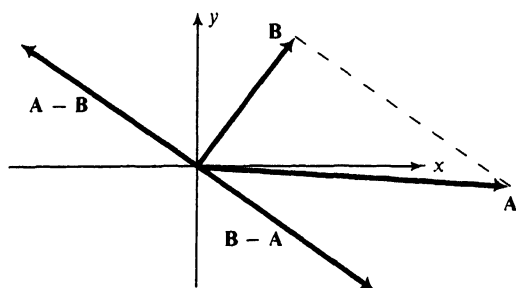


Figure 15.2

In  $\mathbb{R}^2$  one has the picture shown in Figure 15.2. Here the length of  $\mathbf{A} - \mathbf{B}$  is equal to the length of  $\mathbf{B} - \mathbf{A}$  which is the length of the dotted segment since they are opposite sides of a parallelogram. This shows that  $d(\mathbf{A}, \mathbf{B})$  is what you would think it was in this case.

The inequality of (15.3) now reads

$$d(\mathbf{A}, -\mathbf{B}) \leq |\mathbf{A}| + |-\mathbf{B}|,$$

namely the sum of the lengths of two sides of a triangle is greater than or equal to the length of the third side.

We have already used the inner product to define the length of a vector, we now use it to define angles.

**Definition.** Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space. If  $\mathbf{A}$  and  $\mathbf{B}$  are non-zero vectors of  $\mathcal{V}$  the angle between  $\mathbf{A}$  and  $\mathbf{B}$  is defined by

$$\cos \theta = \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{|\mathbf{A}||\mathbf{B}|}$$

and  $0 \leq \theta \leq \pi$ .

(Note that by the Schwarz inequality

$$-1 \leq \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{|\mathbf{A}||\mathbf{B}|} \leq 1$$

so that  $\theta$  is well defined by the above formula.)

This definition gives the formula

$$\langle \mathbf{A}, \mathbf{B} \rangle = |\mathbf{A}||\mathbf{B}| \cos \theta$$

that we are familiar with in  $\mathbb{R}^3$  so we may safely assume that the definition of  $\theta$  is a reasonable extension of the concept of angle to an arbitrary inner product space.

**Proposition 15.4.** Two vectors  $\mathbf{A}$  and  $\mathbf{B}$  in an inner product space  $\mathcal{V}$  are perpendicular (or orthogonal) iff  $\langle \mathbf{A}, \mathbf{B} \rangle = 0$ .

**PROOF.** By definition **A** and **B** are perpendicular iff the angle  $\theta$  between them is  $\pi/2$  radians  $= 90^\circ$ . For such  $\theta$  we have  $\cos \theta = 0$ . Therefore

$$\frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\|\mathbf{A}\| \|\mathbf{B}\|} = \cos \theta = 0$$

so  $\langle \mathbf{A}, \mathbf{B} \rangle = 0$ . The reverse implication, namely  $\langle \mathbf{A}, \mathbf{B} \rangle = 0$  implies  $\theta = \pi/2$  is immediate from the definition of  $\theta$ .  $\square$

Thus the inner product  $\langle \cdot, \cdot \rangle$  provides a very handy check for perpendicularity of vectors.

**EXAMPLE 5.** Are the vectors  $(1, -1, 2)$ ,  $(2, 4, 1)$  in  $\mathbb{R}^3$  perpendicular? (We assume that  $\mathbb{R}^3$  has its standard inner product).

*Solution.* One checks

$$\langle (1, -1, 2), (2, 4, 1) \rangle = 2 - 4 + 2 = 0$$

so they are perpendicular.  $\square$

**Definition.** A set of vectors  $S$  in an inner product space  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  is said to be an *orthogonal set of vectors* iff for any two distinct vectors **A**, **B** in  $S$ ,  $\langle \mathbf{A}, \mathbf{B} \rangle = 0$ . An orthogonal set of vectors is said to be *orthonormal* iff  $\|\mathbf{A}\| = 1$  for every **A** in  $S$ .

**Proposition 15.5.** An orthogonal set of nonzero vectors  $S$  in an inner product space  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  is linearly independent.

**PROOF.** Suppose that

$$a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n = \mathbf{0}$$

is a linear relation between vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  of  $\mathcal{V}$ . Then we have

$$\begin{aligned} 0 &= \langle \mathbf{A}_i, \mathbf{0} \rangle = \langle \mathbf{A}_i, a_1 \mathbf{A}_1 + \cdots + a_n \mathbf{A}_n \rangle \\ &= a_1 \langle \mathbf{A}_i, \mathbf{A}_1 \rangle + a_2 \langle \mathbf{A}_i, \mathbf{A}_2 \rangle + \cdots + a_n \langle \mathbf{A}_i, \mathbf{A}_n \rangle \\ &= a_i \langle \mathbf{A}_i, \mathbf{A}_i \rangle \end{aligned}$$

since  $\langle \mathbf{A}_i, \mathbf{A}_j \rangle = 0$  for  $i \neq j$  because the set  $S$  is orthogonal. But  $\mathbf{A}_i$  is not the zero vector and therefore  $\langle \mathbf{A}_i, \mathbf{A}_i \rangle \neq 0$ , so the equation

$$0 = a_i \langle \mathbf{A}_i, \mathbf{A}_i \rangle$$

implies  $a_i = 0$ . Since this is true for  $i = 1, 2, \dots, n$  there can be no nontrivial linear relation between the vectors of  $S$ , so  $S$  is a linearly independent set.  $\square$

EXAMPLE 6. Which of the following sets of vectors in  $\mathbb{R}^3$  are orthogonal (in the standard inner product)

- (a)  $\{(0, 0, 0), (1, 1, 1)\}$ ?  
 (b)  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ ?  
 (c)  $\{(1, 1, 1), (1, 1, 0), (1, 0, 0)\}$ ?  
 (d)  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}$ ?

*Solution.* The set (a) is orthogonal because it consists of exactly two vectors, one of which is  $\mathbf{0}$ . The set (b) is orthogonal but the set (c) is not, because

$$\langle (1, 1, 1), (1, 0, 0) \rangle = 1 \neq 0.$$

The set (d) is not orthogonal because it contains 4 *nonzero* vectors which if it were orthogonal would have to be linearly independent by (15.5) contrary to the fact that  $\mathbb{R}^3$  is only 3-dimensional.

EXAMPLE 7. Is the set of vectors

$$\left\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\right\}$$

orthogonal in  $\mathcal{P}_2(\mathbb{R})$ ?

*Solution.* The answer is yes, and you might well ask how do I know. The “real answer” lies in (15.7) and Example 10. For the moment all we can do is compute. We have

$$\begin{aligned} \left\langle 1, x - \frac{1}{2} \right\rangle &= \int_0^1 \left( x - \frac{1}{2} \right) dx = \frac{x^2}{2} - \frac{1}{2}x \Big|_0^1 = 0 \\ \left\langle 1, x^2 - x + \frac{1}{6} \right\rangle &= \int_0^1 \left( x^2 - x + \frac{1}{6} \right) dx = \frac{x^3}{3} - \frac{x^2}{2} + \frac{1}{6}x \Big|_0^1 \\ &= \frac{1}{3} - \frac{1}{2} + \frac{1}{6} = 0 \\ \left\langle x - \frac{1}{2}, x^2 - x + \frac{1}{6} \right\rangle &= \int_0^1 \left( x - \frac{1}{2} \right) \left( x^2 - x + \frac{1}{6} \right) dx \\ &= \int_0^1 \left( x^3 - \frac{3}{2}x^2 + \frac{2}{3}x - \frac{1}{12} \right) dx \\ &= \frac{x^4}{4} - \frac{1}{2}x^3 + \frac{1}{3}x^2 - \frac{1}{12}x \Big|_0^1 \\ &= \frac{1}{4} - \frac{1}{2} + \frac{1}{3} - \frac{1}{12} = \frac{3 - 6 + 4 - 1}{12} \\ &= 0. \end{aligned}$$

The remaining scalar products are seen to be zero by symmetry of  $\langle \cdot, \cdot \rangle$  in the two variables.

Note that by (15.4) the vectors  $\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\}$  are linearly independent. Since  $\dim \mathcal{P}_2(\mathbb{R}) = 3$  it follows that these vectors are actually an orthogonal basis for  $\mathcal{P}_2(\mathbb{R})$ .

The construction of orthonormal sets in  $\mathcal{P}_k(\mathbb{R})$  for large  $k$  is a fascinating business involving many areas of mathematics including analytic number theory and differential equations. The polynomials that most often occur are named after the mathematical giants of preceding centuries.

**Proposition 15.6.** Suppose that  $\{A_1, \dots, A_n\}$  is an orthogonal set of nonzero vectors in the inner product space  $\mathcal{V}$ . If

$$A \in \mathcal{L}(A_1, \dots, A_n)$$

then<sup>1</sup>

$$A = \frac{\langle A, A_1 \rangle}{\langle A_1, A_1 \rangle} A_1 + \dots + \frac{\langle A, A_n \rangle}{\langle A_n, A_n \rangle} A_n.$$

**PROOF.** The proof is very similar to that of (15.5). Suppose that  $A$  belongs to the linear span of  $A_1, \dots, A_k$ . Then there are numbers  $a_1, \dots, a_n$  such that

$$A = a_1 A_1 + \dots + a_n A_n.$$

What we are trying to prove is that the number  $a_i$  must be the particular number  $\langle A, A_i \rangle / \langle A_i, A_i \rangle$ . This is easy to prove. Take the inner product of both sides of the preceding equation with  $A_i$  giving

$$\begin{aligned} \langle A_i, A \rangle &= \langle A_i, a_1 A_1 + \dots + a_n A_n \rangle \\ &= a_1 \langle A_i, A_1 \rangle + \dots + a_n \langle A_i, A_n \rangle \\ &= a_i \langle A_i, A_i \rangle \end{aligned}$$

and since  $A_i$  is nonzero so is  $\langle A_i, A_i \rangle$  so

$$a_i = \frac{\langle A_i, A \rangle}{\langle A_i, A_i \rangle} = \frac{\langle A, A_i \rangle}{\langle A_i, A_i \rangle}$$

as required. □

**Remarks.** The numbers  $\langle A, A_i \rangle / \langle A_i, A_i \rangle$  are usually called the *Fourier coefficients* of  $A$  with respect to  $\{A_1, \dots, A_n\}$ . The vector

$$\frac{\langle A, A_i \rangle}{\langle A_i, A_i \rangle} A_i$$

is called the *component of  $A$  along  $A_i$* .

Since the vector

$$\frac{\langle A, A_1 \rangle}{\langle A_1, A_1 \rangle} A_1 + \dots + \frac{\langle A, A_n \rangle}{\langle A_n, A_n \rangle} A_n$$

<sup>1</sup> Note that since  $A_i \neq 0$ ,  $\langle A_i, A_i \rangle \neq 0$  also. Thus the right-hand side is well defined.



belongs to  $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n)$  for any vector  $\mathbf{A}$  in  $\mathcal{V}$  we have actually proven more than is claimed in (15.6), namely

$$\mathbf{A} \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n)$$

iff

$$\mathbf{A} = \frac{\langle \mathbf{A}, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1 + \dots + \frac{\langle \mathbf{A}, \mathbf{A}_n \rangle}{\langle \mathbf{A}_n, \mathbf{A}_n \rangle} \mathbf{A}_n.$$

Here are some examples to illustrate this point:

**EXAMPLE 8.** Find the Fourier coefficients of the vector

$$x^2 + x + 1$$

with respect to the orthogonal set of vectors

$$\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\}$$

in  $\mathcal{P}_2(\mathbb{R})$ .

*Solution.* We must compute

$$\langle 1, x^2 + x + 1 \rangle = \int_0^1 (x^2 + x + 1) dx = \frac{x^3}{3} + \frac{x^2}{2} + x \Big|_0^1 = \frac{11}{6}$$

$$\begin{aligned} \left\langle x - \frac{1}{2}, x^2 + x + 1 \right\rangle &= \int_0^1 \left( x - \frac{1}{2} \right) (x^2 + x + 1) dx \\ &= \int_0^1 \left( x^3 + \frac{1}{2} x^2 + \frac{1}{2} x - \frac{1}{2} \right) dx \\ &= \frac{x^4}{4} + \frac{x^3}{6} + \frac{x^2}{4} - \frac{1}{2} x \Big|_0^1 = \frac{1}{4} + \frac{1}{6} + \frac{1}{4} - \frac{1}{2} \\ &= \frac{1}{6} \end{aligned}$$

$$\begin{aligned} \left\langle x^2 - x + \frac{1}{6}, x^2 + x + 1 \right\rangle &= \int_0^1 \left( x^2 - x + \frac{1}{6} \right) (x^2 + x + 1) dx \\ &= \int_0^1 \left( x^4 + \frac{1}{6} x^2 - \frac{5}{6} x + \frac{1}{6} \right) dx \\ &= \frac{x^5}{5} + \frac{x^3}{18} - \frac{5x^2}{12} + \frac{x}{6} \Big|_0^1 = \frac{1}{5} + \frac{1}{18} - \frac{5}{12} + \frac{1}{6} \\ &= \frac{36 + 10 - 75 + 30}{180} = \frac{1}{180} \end{aligned}$$

$$\langle 1, 1 \rangle = \int_0^1 1 dx = 1$$

$$\begin{aligned}
\left\langle x - \frac{1}{2}, x - \frac{1}{2} \right\rangle &= \int_0^1 \left( x - \frac{1}{2} \right)^2 dx = \frac{1}{3} \left( x - \frac{1}{2} \right)^3 \Big|_0^1 = \frac{1}{12}, \\
\left\langle x^2 - x + \frac{1}{6}, x^2 - x + \frac{1}{6} \right\rangle &= \int_0^1 \left( x^2 - x + \frac{1}{6} \right)^2 dx \\
&= \int_0^1 \left( x^4 - 2x^3 + \frac{4}{3}x^2 - \frac{x}{3} + \frac{1}{36} \right) dx \\
&= \frac{x^5}{5} - \frac{x^4}{2} + \frac{4}{9}x^3 - \frac{x^2}{6} + \frac{1}{36}x \Big|_0^1 \\
&= \frac{1}{5} - \frac{1}{2} + \frac{4}{9} - \frac{1}{6} + \frac{1}{36} \\
&= \frac{36 - 90 + 80 - 30 + 5}{180} \\
&= \frac{1}{180}.
\end{aligned}$$

Thus the Fourier coefficients of  $1 + x + x^2$  with respect to  $\{1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}\}$  are  $\{\frac{11}{6}, 2, 1\}$ . (Note order counts!).

**EXAMPLE 9.** Write the vector  $\mathbf{A} = (1, -1, 1)$  as a linear combination of the vectors

$$\begin{aligned}
\mathbf{V}_1 &= (1, 1, 1) \\
\mathbf{V}_2 &= (0, 1, -1) \\
\mathbf{V}_3 &= (-2, 1, 1).
\end{aligned}$$

*Solution.* Before plunging ahead and solving lists of linear equations it pays to note that  $\{\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3\}$  are orthogonal and hence independent by (15.5) which means, since  $\dim \mathbb{R}^3 = 3$ , that they are a basis for  $\mathbb{R}^3$ . Accordingly by (15.6) we must have

$$\mathbf{A} = \frac{\langle \mathbf{A}, \mathbf{V}_1 \rangle}{\langle \mathbf{V}_1, \mathbf{V}_1 \rangle} \mathbf{V}_1 + \frac{\langle \mathbf{A}, \mathbf{V}_2 \rangle}{\langle \mathbf{V}_2, \mathbf{V}_2 \rangle} \mathbf{V}_2 + \frac{\langle \mathbf{A}, \mathbf{V}_3 \rangle}{\langle \mathbf{V}_3, \mathbf{V}_3 \rangle} \mathbf{V}_3$$

and rather than solve systems of equations we need only compute inner products. We find

$$\begin{aligned}
\langle \mathbf{A}, \mathbf{V}_1 \rangle &= 1 & \langle \mathbf{V}_1, \mathbf{V}_1 \rangle &= 3 \\
\langle \mathbf{A}, \mathbf{V}_2 \rangle &= -2 & \langle \mathbf{V}_2, \mathbf{V}_2 \rangle &= 2 \\
\langle \mathbf{A}, \mathbf{V}_3 \rangle &= -2 & \langle \mathbf{V}_3, \mathbf{V}_3 \rangle &= 6
\end{aligned}$$

and therefore

$$(1, -1, 1) = \frac{1}{3}(1, 1, 1) - 1(0, 1, -1) - \frac{1}{3}(-2, 1, 1),$$

which is all rather easy. □

**Proposition 15.7.** *Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space and  $\mathcal{W}$  a finite-dimensional subspace of  $\mathcal{V}$ . Then there exists in  $\mathcal{V}$  an orthonormal set of vectors  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  that is a basis for  $\mathcal{W}$ .*

**PROOF.** The method of proof that we will employ is called the *Gram-Schmidt orthonormalization* process. It is a manufacturing process that starts with any old basis for  $\mathcal{W}$  as raw material and manufactures one of the required type. We therefore begin by choosing a basis  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  for  $\mathcal{W}$ . This we know we can do by (6.1). We now define in succession vectors  $\mathbf{A}_1, \dots, \mathbf{A}_n$  by the equations

$$\mathbf{A}_1 = \mathbf{B}_1$$

$$\mathbf{A}_2 = \mathbf{B}_2 - \frac{\langle \mathbf{B}_2, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1$$

$$\mathbf{A}_n = \mathbf{B}_n - \frac{\langle \mathbf{B}_n, \mathbf{A}_{n-1} \rangle}{\langle \mathbf{A}_{n-1}, \mathbf{A}_{n-1} \rangle} \mathbf{A}_{n-1} - \dots - \frac{\langle \mathbf{B}_n, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1.$$

Before proceeding further we must check that we have not committed the cardinal sin of dividing by 0 in one of the above formulas. To do so note that  $\mathbf{A}_i$  is actually a linear combination of  $\mathbf{B}_1, \dots, \mathbf{B}_i$  in disguise, that is  $\mathbf{A}_i \in \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_i)$ . If for some  $j$ ,  $\langle \mathbf{A}_j, \mathbf{A}_j \rangle = 0$ , then  $\mathbf{A}_j = \mathbf{0}$  and therefore we get from the equation defining  $\mathbf{A}_j$

$$\mathbf{B}_j = \frac{\langle \mathbf{B}_j, \mathbf{A}_{j-1} \rangle}{\langle \mathbf{A}_{j-1}, \mathbf{A}_{j-1} \rangle} \mathbf{A}_{j-1} + \dots + \frac{\langle \mathbf{B}_j, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1 \in \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_{j-1})$$

and therefore  $\mathbf{B}_j \in \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_{j-1})$  contrary to the linear independence of  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$ . Therefore  $\langle \mathbf{A}_j, \mathbf{A}_j \rangle$  is never zero.

Next we claim that  $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_i) = \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_i)$ . We have already seen that  $\mathbf{A}_1, \dots, \mathbf{A}_i \in \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_i)$  so that  $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_i) \subset \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_i)$ . It will therefore suffice to show  $\mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_i) \subset \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_i)$  which is easy.

Our next task is to show that the vectors  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  are orthogonal. We will do this by showing successively that the sets  $\{\mathbf{A}_1\}$ ,  $\{\mathbf{A}_1, \mathbf{A}_2\}$ ,  $\dots$ ,  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  are orthogonal. The first set is orthogonal because  $\mathbf{A}_1 = \mathbf{B}_1 \neq \mathbf{0}$ . Suppose as an induction hypothesis that we have already shown the set  $\{\mathbf{A}_1, \dots, \mathbf{A}_i\}$  to be orthogonal. In order to prove that  $\{\mathbf{A}_1, \dots, \mathbf{A}_{i+1}\}$  is

orthogonal it is only necessary to show that  $\mathbf{A}_{i+1}$  is orthogonal to  $\mathbf{A}_1, \dots, \mathbf{A}_i$ . Now we compute for  $1 \leq j \leq i$

$$\begin{aligned}
 \langle \mathbf{A}_{i+1}, \mathbf{A}_j \rangle &= \left\langle \mathbf{B}_{i+1} - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_i \rangle}{\langle \mathbf{A}_i, \mathbf{A}_i \rangle} \mathbf{A}_i - \dots - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1, \mathbf{A}_j \right\rangle \\
 &= \langle \mathbf{B}_{i+1}, \mathbf{A}_j \rangle - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_i \rangle}{\langle \mathbf{A}_i, \mathbf{A}_i \rangle} \langle \mathbf{A}_i, \mathbf{A}_j \rangle - \dots \\
 &\quad - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \langle \mathbf{A}_1, \mathbf{A}_j \rangle \\
 &= \langle \mathbf{B}_{i+1}, \mathbf{A}_j \rangle - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_i \rangle}{\langle \mathbf{A}_i, \mathbf{A}_i \rangle} 0 - \dots - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_j \rangle}{\langle \mathbf{A}_j, \mathbf{A}_j \rangle} \langle \mathbf{A}_j, \mathbf{A}_j \rangle - \dots \\
 &\quad - \frac{\langle \mathbf{B}_{i+1}, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} 0 \\
 &= \langle \mathbf{B}_{i+1}, \mathbf{A}_j \rangle - \langle \mathbf{B}_{i+1}, \mathbf{A}_j \rangle = 0,
 \end{aligned}$$

where the key step was to use the fact  $\langle \mathbf{A}_k, \mathbf{A}_j \rangle = 0$  if  $1 \leq k \leq i$  and  $k \neq j$ . This computation shows that  $\{\mathbf{A}_1, \dots, \mathbf{A}_{i+1}\}$  is an orthogonal set of vectors. By induction we may conclude that  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is an orthogonal set of nonzero vectors. By (15.5) it follows that the set  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is linearly independent. Since  $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_n) \subset \mathcal{W}$  and  $n = \dim \mathcal{W}$  it follows from (6.9) that  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is a basis for  $\mathcal{W}$ .

But we are not yet done! We have only constructed an orthogonal set of vectors  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  that is a basis for  $\mathcal{W}$ . What we want is an *orthonormal* set. However this is easy. We just **normalize** the set  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  by defining  $\bar{\mathbf{A}}_i = (1/|\mathbf{A}_i|)\mathbf{A}_i$ , which we may do since  $|\mathbf{A}_i| \neq 0$ . We then have

$$|\bar{\mathbf{A}}_i| = \left| \frac{1}{|\mathbf{A}_i|} \mathbf{A}_i \right| = \frac{1}{|\mathbf{A}_i|} |\mathbf{A}_i| = 1$$

and

$$\langle \bar{\mathbf{A}}_r, \bar{\mathbf{A}}_s \rangle = \frac{1}{|\mathbf{A}_r|} \frac{1}{|\mathbf{A}_s|} \langle \mathbf{A}_r, \mathbf{A}_s \rangle = 0, \quad r \neq s$$

so that  $\{\bar{\mathbf{A}}_1, \dots, \bar{\mathbf{A}}_n\}$  is both a basis and an orthonormal set.  $\square$

The kind of basis that appears in the above proposition is very important and therefore comes in for a special name.

**Definition.** Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be a finite-dimensional inner product space.

A basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  for  $\mathcal{V}$  is called an *orthonormal basis* for  $\mathcal{V}$  iff  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is a basis for  $\mathcal{V}$  and an orthonormal set, that is

$$\langle \mathbf{A}_i, \mathbf{A}_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

or what is the same thing, the vectors are of unit length and pairwise orthogonal.

While the Gram-Schmidt process may seem quite formidable it really isn't. For once one has established that it works, one can apply it without worrying about the inductive check that the vectors it manufactures are orthogonal. However one should not forget to normalize at the end if required to do so.

**EXAMPLE 10.** Apply the Gram-Schmidt orthogonalization process to find an orthogonal basis for  $\mathcal{P}_2(\mathbb{R})$ .

*Solution.* We may start with any old basis for  $\mathcal{P}_2(\mathbb{R})$  to apply the process to, so let us choose the obvious basis  $\{1, x, x^2\}$ . We then define vectors

$$\mathbf{A}_1 = 1$$

$$\mathbf{A}_2 = x - \frac{\langle x, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1$$

$$\mathbf{A}_3 = x^2 - \frac{\langle x^2, \mathbf{A}_1 \rangle}{\langle \mathbf{A}_1, \mathbf{A}_1 \rangle} \mathbf{A}_1 - \frac{\langle x^2, \mathbf{A}_2 \rangle}{\langle \mathbf{A}_2, \mathbf{A}_2 \rangle} \mathbf{A}_2$$

and so we have some integrations to carry out.

Since  $\mathbf{A}_1 = 1$  we get

$$\langle x, \mathbf{A}_1 \rangle = \langle x, 1 \rangle = \int_0^1 x \, dx = \frac{1}{2}$$

$$\langle \mathbf{A}_1, \mathbf{A}_1 \rangle = \langle 1, 1 \rangle = \int_0^1 1 \, dx = 1$$

so that  $\mathbf{A}_2 = x - \frac{1}{2}$  and we get

$$\langle x^2, \mathbf{A}_1 \rangle = \langle x^2, 1 \rangle = \int_0^1 x^2 \, dx = \frac{1}{3}$$

$$\begin{aligned} \langle x^2, \mathbf{A}_2 \rangle &= \int_0^1 x^2 \left( x - \frac{1}{2} \right) dx = \int_0^1 \left( x^3 - \frac{1}{2} x^2 \right) dx \\ &= \frac{x^4}{4} - \frac{x^3}{6} \Big|_0^1 = \frac{1}{4} - \frac{1}{6} = \frac{1}{12} \end{aligned}$$

$$\begin{aligned} \langle \mathbf{A}_2, \mathbf{A}_2 \rangle &= \int_0^1 \left( x - \frac{1}{2} \right)^2 dx = \frac{1}{3} \left( x - \frac{1}{2} \right)^3 \Big|_0^1 \\ &= \frac{1}{3} \left[ \left( 1 - \frac{1}{2} \right)^3 - \left( 0 - \frac{1}{2} \right)^3 \right] \\ &= \frac{1}{3} \left[ \frac{1}{8} - \left( -\frac{1}{8} \right) \right] = \frac{1}{3} \left( \frac{1}{4} \right) = \frac{1}{12} \end{aligned}$$

so that we find

$$\begin{aligned} \mathbf{A}_1 &= 1 \\ \mathbf{A}_2 &= x - \frac{1}{2} \\ \mathbf{A}_3 &= x^2 - \frac{1}{3} \cdot 1 - (x - \frac{1}{2}) = x^2 - x + \frac{1}{6} \end{aligned}$$

is an orthogonal basis for  $\mathcal{P}_2(\mathbb{R})$ . (Go back and look at Example 7 again.)

**EXAMPLE 11.** Find an orthonormal basis for the subspace  $\mathcal{W}$  given by

$$x + 2y + 3z = 0$$

of  $\mathbb{R}^3$ .

*Solution.* A basis for  $\mathcal{W}$  is easily seen to be

$$\mathbf{B}_1 = (-3, 0, 1), \quad \mathbf{B}_2 = (-2, 1, 0).$$

Applying the Gram-Schmidt process we get the orthogonal basis

$$\begin{aligned} \mathbf{A}_1 &= (-3, 0, 1) \\ \mathbf{A}_2 &= (-2, 1, 0) - \frac{\langle (-3, 0, 1), (-2, 1, 0) \rangle}{\langle (-3, 0, 1), (-3, 0, 1) \rangle} (-3, 0, 1) \\ &= (-2, 1, 0) - \frac{6}{10}(-3, 0, 1) \\ &= (-\frac{1}{5}, 1, -\frac{3}{5}). \end{aligned}$$

*But we are not done yet!* The vectors  $\{\mathbf{A}_1, \mathbf{A}_2\}$  are only orthogonal, not orthonormal. We must therefore normalize them getting

$$\begin{aligned} \mathbf{C}_1 &= \frac{\mathbf{A}_1}{|\mathbf{A}_1|} = \frac{1}{\sqrt{10}}(-3, 0, 1) = \left(-\frac{3}{\sqrt{10}}, 0, \frac{1}{\sqrt{10}}\right) \\ \mathbf{C}_2 &= \frac{\mathbf{A}_2}{|\mathbf{A}_2|} = \frac{\sqrt{5}}{\sqrt{7}}\left(-\frac{1}{5}, 1, -\frac{3}{5}\right) = \left(-\frac{1}{\sqrt{35}}, \frac{\sqrt{5}}{\sqrt{7}}, -\frac{3}{\sqrt{35}}\right) \end{aligned}$$

which is an orthonormal basis for  $\mathcal{W}$ .

There is a very important consequence of (15.7), namely that a finite-dimensional inner product space has an orthonormal basis. We record this for future reference.

**Corollary 15.8.** *Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be a finite-dimensional inner product space. Then there is an orthonormal basis for  $\mathcal{V}$ .* □

One reason why orthonormal bases are so useful in an inner product space is the following

**Theorem 15.9.** Let  $\{A_1, \dots, A_n\}$  be an orthonormal basis for the inner product space  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ . If  $A$  belongs to  $\mathcal{V}$  then

$$A = \langle A, A_1 \rangle A_1 + \langle A, A_2 \rangle A_2 + \cdots + \langle A, A_n \rangle A_n.$$

If

$$B = b_1 A_1 + b_2 A_2 + \cdots + b_n A_n$$

$$C = c_1 A_1 + c_2 A_2 + \cdots + c_n A_n$$

then

$$\langle B, C \rangle = b_1 c_1 + b_2 c_2 + \cdots + b_n c_n.$$

**PROOF.** The first assertion follows directly from (15.6) because  $\langle A_1, A_1 \rangle = \langle A_2, A_2 \rangle = \cdots = \langle A_n, A_n \rangle = 1$ . To prove the second assertion we just compute:

$$\begin{aligned} \langle B, C \rangle &= \langle b_1 A_1 + \cdots + b_n A_n, c_1 A_1 + \cdots + c_n A_n \rangle \\ &= \sum_{i,j} b_i c_j \langle A_i, A_j \rangle \\ &= b_1 a_1 + b_2 a_2 + \cdots + b_n a_n \end{aligned}$$

because

$$\langle A_i, A_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

since  $\{A_1, \dots, A_n\}$  is orthonormal. □

Thus by choosing an orthonormal basis  $\{A_1, \dots, A_n\}$  in a finite-dimensional inner product space and expressing vectors in terms of their components with respect to  $\{A_1, \dots, A_n\}$  we obtain formulas for  $\langle A, B \rangle$  that look exactly like the formulas for  $\mathbb{R}^n$  with its standard inner product. There is a good reason for this as we will now explain.

In studying vector spaces we found that it was important to also consider linear transformations, that is, functions which preserved the basic vector operations. For vector spaces equipped with an inner product it seems reasonable to demand that the inner product also be preserved.

**Definition.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be inner product spaces. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  is called an *isometric embedding* iff

$$\langle T(A), T(B) \rangle = \langle A, B \rangle$$

for all vectors  $A, B$  in  $\mathcal{V}$ .

Isometric embeddings are a very restricted class of linear transformations. We have:

**Proposition 15.10.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be inner product spaces and  $T: \mathcal{V} \rightarrow \mathcal{W}$  an isometric embedding. Then  $\ker T = \{\mathbf{0}\}$ .

**PROOF.** Suppose that  $T(\mathbf{A}) = \mathbf{0}$ . Then since  $T$  is an isometric embedding

$$\langle \mathbf{A}, \mathbf{A} \rangle = \langle T(\mathbf{A}), T(\mathbf{A}) \rangle = \langle \mathbf{0}, \mathbf{0} \rangle = 0$$

and therefore  $\mathbf{A} = \mathbf{0}$ . □

**Corollary 15.11.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional inner product spaces of the same dimension and  $T: \mathcal{V} \rightarrow \mathcal{W}$  an isometric embedding. Then  $T$  is an isomorphism.

**PROOF.** By (8.11) it suffices to show that  $\text{Im } T = \mathcal{W}$ . By (8.10) we obtain  $\dim \mathcal{V} = \dim \text{Im } T$ . Since  $\dim \mathcal{V} = \dim \mathcal{W}$  by hypothesis,  $\dim \mathcal{W} = \dim \text{Im } T$ . Therefore since  $\text{Im } T$  is a linear subspace of  $\mathcal{W}$  (8.9) it follows from (6.9) that  $\text{Im } T = \mathcal{W}$ . □

**EXAMPLE 12.** An example of an isometric embedding of  $\mathbb{R}^2$  in  $\mathbb{R}^4$  is given by

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^4$$

such that

$$T(x, y) = \frac{1}{\sqrt{2}}(x, y, y, x).$$

To see this we note that

$$\begin{aligned} \langle T(\mathbf{A}), T(\mathbf{B}) \rangle &= \left\langle \frac{1}{\sqrt{2}}(a_1, a_2, a_2, a_1), \frac{1}{\sqrt{2}}(b_1, b_2, b_2, b_1) \right\rangle \\ &= \frac{1}{2}(a_1 b_1 + a_2 b_2 + a_2 b_2 + a_1 b_1) \\ &= \frac{1}{2}(2a_1 b_1 + 2a_2 b_2) \\ &= a_1 b_1 + a_2 b_2 = \langle \mathbf{A}, \mathbf{B} \rangle \end{aligned}$$

as required. □

**Definition.** Let  $\mathcal{V}$  and  $\mathcal{W}$  be inner product spaces. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{W}$  is called an *isometric isomorphism*, an *isometry* for short, iff  $T$  is an isometric embedding, and there exists an isometric embedding  $S: \mathcal{W} \rightarrow \mathcal{V}$  such that

$$S \cdot T = I: \mathcal{V} \rightarrow \mathcal{V}$$

and

$$T \cdot S = I: \mathcal{W} \rightarrow \mathcal{W}.$$

If there is an isometry  $T: \mathcal{V} \rightarrow \mathcal{W}$  we say that  $\mathcal{V}$  and  $\mathcal{W}$  are *isometric*.



**Proposition 15.12.** *Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional inner product spaces of the same dimension. If  $T: \mathcal{V} \rightarrow \mathcal{W}$  is an isometric embedding then  $T$  is an isometric isomorphism.*

**PROOF.** By (15.11)  $T$  is an isomorphism. Let  $S = T^{-1}$ . We need only show that  $S: \mathcal{W} \rightarrow \mathcal{V}$  is an isometric embedding. Let  $A, B \in \mathcal{W}$ . Write  $A = T(C)$ ,  $B = T(D)$ . Then

$$\langle S(A), S(B) \rangle = \langle C, D \rangle = \langle T(C), T(D) \rangle = \langle A, B \rangle$$

as required.  $\square$

**Proposition 15.13.** *Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional inner product spaces with  $\dim \mathcal{V} \leq \dim \mathcal{W}$ . Then there exists an isometric embedding  $T: \mathcal{V} \rightarrow \mathcal{W}$ .*

**PROOF.** Choose orthonormal bases  $\{V_1, \dots, V_m\}$  and  $\{W_1, \dots, W_n\}$  for  $\mathcal{V}$  and  $\mathcal{W}$  respectively and let  $T: \mathcal{V} \rightarrow \mathcal{W}$  be the linear extension of

$$T(V_i) = W_i \quad i = 1, \dots, m.$$

Thus if  $A, B \in \mathcal{V}$ , say

$$A = \sum a_i V_i, \quad B = \sum b_i V_i$$

we obtain from (15.9) the formula

$$\langle A, B \rangle = \sum a_i b_i,$$

while since

$$T(A) = \sum a_i W_i, \quad T(B) = \sum b_i W_i$$

we get

$$\langle T(A), T(B) \rangle = \sum a_i b_i$$

so  $T$  is an isometric embedding.  $\square$

Summarizing we obtain:

**Theorem 15.14.** *Let  $\mathcal{V}$  and  $\mathcal{W}$  be finite-dimensional inner product spaces. Then  $\mathcal{V}$  and  $\mathcal{W}$  are isometric iff  $\dim \mathcal{V} = \dim \mathcal{W}$ .*

**PROOF.** If  $\mathcal{V}$  and  $\mathcal{W}$  are isometric then they are certainly isomorphic, so  $\dim \mathcal{V} = \dim \mathcal{W}$ . On the other hand if  $\dim \mathcal{V} = \dim \mathcal{W}$  then we may apply (15.13) to construct an isometric embedding  $T: \mathcal{V} \rightarrow \mathcal{W}$  which by (15.12) will be an isometric isomorphism.  $\square$

**Corollary 15.15.** *If  $\mathcal{V}$  is an inner product space of finite dimension  $n$  then  $\mathcal{V}$  is isometric to  $\mathbb{R}^n$ .*

EXAMPLE 13. According to (15.15) the spaces  $\mathbb{R}^3$  and  $\mathcal{P}_2(\mathbb{R})$  are isometric. An explicit isometry is given by

$$T(u, v, w) = 6\sqrt{5}wx^2 + (2\sqrt{3}v - 6\sqrt{5}w)x + (u - \sqrt{3}v + \sqrt{5}w).$$

This isometry is constructed as follows. Note that  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$  is an orthonormal basis for  $\mathbb{R}^3$  while  $\{1, \sqrt{12}(x - \frac{1}{2}), \sqrt{180}(x^2 - x + \frac{1}{6})\}$  is an orthonormal basis for  $\mathcal{P}_2(\mathbb{R})$ . Therefore the linear extension of

$$T(1, 0, 0) = 1$$

$$T(0, 1, 0) = \sqrt{12}(x - \frac{1}{2})$$

$$T(0, 0, 1) = \sqrt{180}(x^2 - x + \frac{1}{6})$$

will be an isometry  $T: \mathbb{R}^3 \rightarrow \mathcal{P}_2(\mathbb{R})$ . A little algebraic manipulation brings forth the above formula.

Thus, exactly as with (8.15), we see that the added abstraction of the axiomatics has added nothing new, that is, up to isometry there is only one inner product space of finite dimension  $n$ , namely  $\mathbb{R}^n$ . Still, the axiomatic approach is quite useful in that it provides a natural setting for rather exotic inner products on  $\mathbb{R}^n$  such as that of Example 3. These inner products occur in nature and should not be denied.

For the last topic in this chapter we are going to show how a linear transformation

$$T: \mathcal{V} \rightarrow \mathbb{R}$$

of an inner product space  $\mathcal{V}$  to the 1-dimensional space  $\mathbb{R}$  may be expressed in terms of the inner product  $\langle \cdot, \cdot \rangle$ . This will prove quite useful in the next chapter.

**Definition.** Let  $\mathcal{V}$  be an inner product space and  $\mathcal{W}$  a linear subspace. The *orthogonal complement*  $\mathcal{W}^\perp$  of  $\mathcal{W}$  in  $\mathcal{V}$  is defined by

$$\mathcal{W}^\perp = \{A \in \mathcal{V} \mid \langle A, B \rangle = 0 \text{ for every } B \in \mathcal{W}\}.$$

**Proposition 15.16.** Let  $\mathcal{V}$  be an inner product space and  $\mathcal{W}$  a linear subspace. Then  $\mathcal{W}^\perp$  is a linear subspace and  $\mathcal{W} \cap \mathcal{W}^\perp = \{0\}$ .

**PROOF.** Let  $B, C$  belong to  $\mathcal{W}^\perp$  then

$$\langle A, B + C \rangle = \langle A, B \rangle + \langle A, C \rangle = 0 + 0 = 0$$

so  $B + C$  belongs to  $\mathcal{W}^\perp$ . If  $r$  is a number

$$\langle A, rB \rangle = r\langle A, B \rangle = r \cdot 0 = 0$$

so  $rB \in \mathcal{W}^\perp$ . Therefore  $\mathcal{W}^\perp$  is closed under scalar products and vector addition so is a linear subspace of  $\mathcal{V}$ .

If  $A$  belongs to  $\mathcal{W} \cap \mathcal{W}^\perp$  then  $\langle A, A \rangle = 0$  since  $A \in \mathcal{W}$  and  $A \in \mathcal{W}^\perp$ . Therefore  $A = 0$ .  $\square$

EXAMPLE 14. What is the orthogonal complement of the subspace

$$\mathcal{W} = \{(x, y, z) | x + 2y + 3z = 0\}$$

in  $\mathbb{R}^3$ ?

*Solution.* The orthogonal complement of the plane  $\mathcal{W}$  will be a line  $\mathcal{S}$ . In this case the line is the subspace spanned by  $(1, 2, 3)$ . (These coordinates are the coefficients of the equation used to define  $\mathcal{W}$ . This always works) For if  $(x, y, z) \in \mathcal{W}$  then

$$\langle (1, 2, 3), (x, y, z) \rangle = x + 2y + 3z = 0$$

because  $(x, y, z) \in \mathcal{W}$ .

EXAMPLE 15. What is the orthogonal complement of the subspace

$$\mathcal{W} = \{(x, y, z) | x + y + z = 0 \text{ and } x - y + z = 0\}?$$

*Solution.* The subspace  $\mathcal{W}$  is a line. Its orthogonal complement will therefore be a plane. A basis for  $\mathcal{W}$  is the vector  $(1, 0, -1)$ , since it satisfies the equations of both the planes

$$x + y + z = 0 \text{ and } x - y + z = 0$$

and must therefore be on their line of intersection.

The orthogonal complement of  $\mathcal{W}$  must therefore be the plane  $\mathcal{S}$  whose normal vector at the origin is  $(1, 0, -1)$ . Therefore

$$\mathcal{S} = \{(x, y, z) | x - z = 0\}.$$

as is required.

**Proposition 15.17.** Let  $\mathcal{V}$  be a finite-dimensional inner product space and  $\mathcal{W}$  a linear subspace of  $\mathcal{V}$ . Then (see (13.4))

$$\mathcal{W} + \mathcal{W}^\perp = \mathcal{V}.$$

**PROOF.** What we must do is show that any vector  $\mathbf{A}$  in  $\mathcal{V}$  may be written as a sum  $\mathbf{B} + \mathbf{C}$  with  $\mathbf{B}$  in  $\mathcal{W}$  and  $\mathbf{C}$  in  $\mathcal{W}^\perp$ . To this end apply (15.7) to choose an orthonormal basis  $\mathbf{W}_1, \dots, \mathbf{W}_k$  for  $\mathcal{W}$ . Let  $\mathbf{A} \in \mathcal{V}$  and define

$$\begin{aligned} \mathbf{B} &= \langle \mathbf{W}_1, \mathbf{A} \rangle \mathbf{W}_1 + \dots + \langle \mathbf{W}_k, \mathbf{A} \rangle \mathbf{W}_k \\ \mathbf{C} &= \mathbf{A} - \mathbf{B}. \end{aligned}$$

Clearly  $\mathbf{A} = \mathbf{B} + \mathbf{C}$  and  $\mathbf{B}$  belongs to  $\mathcal{W}$ . So what it remains to show is that  $\mathbf{C}$  belongs to  $\mathcal{W}^\perp$ . First of all note

$$\mathbf{C} = \mathbf{A} - \langle \mathbf{W}_1, \mathbf{A} \rangle \mathbf{W}_1 - \dots - \langle \mathbf{W}_k, \mathbf{A} \rangle \mathbf{W}_k.$$

Hence

$$\begin{aligned} \langle \mathbf{C}, \mathbf{W}_i \rangle &= \langle \mathbf{A}, \mathbf{W}_i \rangle - \langle \mathbf{W}_1, \mathbf{A} \rangle \langle \mathbf{W}_1, \mathbf{W}_i \rangle - \dots - \langle \mathbf{W}_k, \mathbf{A} \rangle \langle \mathbf{W}_k, \mathbf{W}_i \rangle \\ &= \langle \mathbf{A}, \mathbf{W}_i \rangle - \langle \mathbf{W}_i, \mathbf{A} \rangle = 0 \end{aligned}$$

since

$$\langle \mathbf{W}_j, \mathbf{W}_i \rangle = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Next suppose that  $\mathbf{W}$  belongs to  $\mathcal{W}$ . Then

$$\mathbf{W} = w_1 \mathbf{W}_1 + \cdots + w_k \mathbf{W}_k$$

so

$$\langle \mathbf{C}, \mathbf{W} \rangle = \sum w_i \langle \mathbf{C}, \mathbf{W}_i \rangle = \sum w_i 0 = 0$$

and hence  $\mathbf{C}$  belongs to  $\mathcal{W}^\perp$ . □

**Proposition 15.18.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space and  $\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  an orthonormal set of vectors in  $\mathcal{V}$ . Then there are vectors  $\{\mathbf{A}_{k+1}, \dots, \mathbf{A}_n\}$  such that  $\{\mathbf{A}_1, \dots, \mathbf{A}_k, \mathbf{A}_{k+1}, \dots, \mathbf{A}_n\}$  is an orthonormal basis for  $\mathcal{V}$ .*

PROOF. Let  $\mathcal{W} = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k)$ . By (14.6)  $\mathcal{W}^\perp$  is a linear subspace of  $\mathcal{V}$  and since  $\mathcal{V}$  is finite dimensional so is  $\mathcal{W}^\perp$ . Let  $\{\mathbf{B}_1, \dots, \mathbf{B}_l\}$  be an orthonormal basis for  $\mathcal{W}^\perp$ . One readily checks that the set  $\{\mathbf{A}_1, \dots, \mathbf{A}_k, \mathbf{B}_1, \dots, \mathbf{B}_l\}$  is an orthonormal set, and since none of the vectors in the set is  $\mathbf{0}$ , this set is linearly independent by (15.5). By (15.17)  $\mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k, \mathbf{B}_1, \dots, \mathbf{B}_l) = \mathcal{V}$  and so setting

$$\mathbf{A}_{k+1} = \mathbf{B}_1, \dots, \mathbf{A}_n = \mathbf{B}_l$$

we have the required orthonormal basis. □

Notice the following important consequence of the proof of (15.18).

**Corollary 15.19.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space of dimension  $n$  and  $\mathcal{W}$  a subspace of  $\mathcal{V}$  of dimension  $m$ . Then  $\dim \mathcal{W}^\perp = n - m$ .*

**Theorem 15.20.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space and  $T: \mathcal{V} \rightarrow \mathbb{R}$  a linear transformation. Then there exists a unique vector  $\mathbf{A}$  in  $\mathcal{V}$  (depending on  $T$ ) such that*

$$T(\mathbf{B}) = \langle \mathbf{B}, \mathbf{A} \rangle$$

for all  $\mathbf{B}$  in  $\mathcal{V}$ .

PROOF. Let the dimension of  $\mathcal{V}$  be  $n$ . If  $\text{Im } T = \{0\}$  set  $\mathbf{A} = \mathbf{0}$ . Then

$$0 = T(\mathbf{B})$$

$$0 = \langle \mathbf{B}, \mathbf{0} \rangle$$

so the theorem is true in this case. If  $\text{Im } T \neq \{0\}$  then we must have  $\text{Im } T = \mathbb{R}$  since  $\{0\}$  and  $\mathbb{R}$  are the only subspaces of  $\mathbb{R}$ . Therefore by (8.10) the kernel of  $T$  has dimension  $n - 1$ . Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_{n-1}\}$  be an orthonormal basis for  $\ker T$ .

By (15.18) there is a vector  $\mathbf{A}_n$  such that  $\{\mathbf{A}_1, \dots, \mathbf{A}_{n-1}, \mathbf{A}_n\}$  is a basis for  $\mathcal{V}$ . Let  $\mathbf{B} \in \mathcal{V}$ . By (15.6)

$$\mathbf{B} = \langle \mathbf{B}, \mathbf{A}_1 \rangle \mathbf{A}_1 + \dots + \langle \mathbf{B}, \mathbf{A}_n \rangle \mathbf{A}_n$$

so

$$\begin{aligned} T(\mathbf{B}) &= \langle \mathbf{B}, \mathbf{A}_1 \rangle T(\mathbf{A}_1) + \dots + \langle \mathbf{B}, \mathbf{A}_n \rangle T(\mathbf{A}_n) \\ &= \langle \mathbf{B}, \mathbf{A}_n \rangle T(\mathbf{A}_n) \\ &= \langle \mathbf{B}, T(\mathbf{A}_n) \mathbf{A}_n \rangle \end{aligned}$$

(remember  $T(\mathbf{A}_n)$  is only a number). Thus  $\mathbf{A} = T(\mathbf{A}_n) \mathbf{A}_n$  is the required vector.

To see that the vector  $\mathbf{A}$  is unique suppose that  $\mathbf{C}$  is another vector such that  $T(\mathbf{B}) = \langle \mathbf{B}, \mathbf{C} \rangle$  for all  $\mathbf{B}$  in  $\mathcal{V}$ . Therefore

$$\begin{aligned} \langle \mathbf{B}, \mathbf{A} - \mathbf{C} \rangle &= \langle \mathbf{B}, \mathbf{A} \rangle - \langle \mathbf{B}, \mathbf{C} \rangle \\ &= T(\mathbf{B}) - T(\mathbf{B}) \\ &= 0 \end{aligned}$$

for all  $\mathbf{B}$  in  $\mathcal{V}$ . Put  $\mathbf{B} = \mathbf{A} - \mathbf{C}$  to get

$$\langle \mathbf{A} - \mathbf{C}, \mathbf{A} - \mathbf{C} \rangle = 0$$

but this implies by Axiom (4) of an inner product that  $\mathbf{A} - \mathbf{C} = \mathbf{0}$  and hence  $\mathbf{A} = \mathbf{C}$ .  $\square$

EXAMPLE 16. Consider the linear transformation

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}$$

given by

$$T(x, y, z) = x + y + z.$$

The kernel of  $T$  is thus the plane

$$\ker T = \{(x, y, z) | x + y + z = 0\}$$

and a vector orthogonal to  $\ker T$  is  $(1, 1, 1)$ , which normalizes to  $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ . Since

$$T\left(\frac{1}{\sqrt{3}}(1, 1, 1)\right) = \frac{3}{\sqrt{3}} = \sqrt{3}$$

we find that

$$T(x, y, z) = \langle (x, y, z), (1, 1, 1) \rangle$$

which readily checks.

In fact more generally if

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}$$

is given by

$$T(x, y, z) = ax + by + cz$$

then the vector  $V = (a, b, c)$  satisfies

$$T(x, y, z) = \langle (x, y, z), (a, b, c) \rangle$$

as one easily checks.

EXAMPLE 17. Consider the linear transformation

$$T: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathbb{R}$$

given by

$$T(p(x)) = p(1).$$

The kernel of  $T$  is the subspace

$$\begin{aligned} \ker T &= \{p(x) | p(1) = 0\} \\ &= \{a_0 + a_1x + a_2x^2 | a_1 + a_2 + a_3 = 0\}. \end{aligned}$$

A basis for the kernel of  $T$  consists of  $1 - x$ ,  $x - x^2$ . The three vectors  $\{1 - x, x - x^2, 1\}$  are a basis for  $\mathcal{P}_2(\mathbb{R})$ . Apply the Gram-Schmidt process to orthogonalize them. (Note the order!) We get

$$\begin{aligned} A_1 &= 1 - x \\ A_2 &= x - x^2 - \frac{\langle 1, 1 - x \rangle}{\langle 1 - x, 1 - x \rangle} (1 - x) = \frac{5}{2}x - x^2 - \frac{3}{2} \\ A_3 &= 1 - \frac{\langle 1, 1 - x \rangle}{\langle 1 - x, 1 - x \rangle} (1 - x) - \frac{\langle 1, x - x^2 \rangle}{\langle x - x^2, x - x^2 \rangle} (x - x^2) \\ &= 1 - \frac{\frac{1}{2}}{\frac{1}{3}} (1 - x) - \frac{\frac{1}{6}}{\frac{1}{30}} (x - x^2) \\ &= 5x^2 - \frac{7}{2}x - \frac{1}{2} \end{aligned}$$

so that a unit vector orthogonal to  $\ker T$  is

$$\frac{\sqrt{3}}{\sqrt{2}} A_3 = \frac{\sqrt{3}}{\sqrt{2}} \left( 5x^2 - \frac{7}{2}x - \frac{1}{2} \right)$$

finally since

$$T\left(\frac{\sqrt{3}}{\sqrt{2}} \left( 5x^2 - \frac{7}{2}x - \frac{1}{2} \right)\right) = \frac{\sqrt{3}}{\sqrt{2}} \left( 5 - \frac{7}{2} - \frac{1}{2} \right) = \frac{\sqrt{3}}{\sqrt{2}}$$

we find that

$$T(p(x)) = \frac{3}{2} \int_0^1 p(x) \left( 5x^2 - \frac{7}{2}x - \frac{1}{2} \right) dx.$$

## EXERCISES

1. Compute the following inner products in  $\mathbb{R}^3$  relative to the standard inner product. Find the cosine of the angle between the indicated pairs of vectors:

$$\begin{aligned}\langle (1, 2, 3), (4, 5, 6) \rangle & \quad \langle (1, 0, 1), (0, 4, 0) \rangle \\ \langle (1, 2, 3), (3, 2, 1) \rangle & \quad \langle (1, 1, 1), (1, -2, 1) \rangle\end{aligned}$$

2. Compute the following inner products in  $\mathcal{P}_2(\mathbb{R})$ . Find the cosine of the angle between the indicated pair of vectors:

$$\begin{aligned}\langle 1 + 2x + 3x^2, 4 + 5x + 6x^2 \rangle \\ \langle 1 + 2x + 3x^2, 3 + 2x + x^2 \rangle \\ \langle 1 + x^2, 4x \rangle \\ \langle 1 + x + x^2, 1 - 2x + x^2 \rangle\end{aligned}$$

3. Compute the following inner products in  $\mathcal{M}_{33}$  using the trace inner product:

$$\begin{aligned}\left\langle \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 1 \\ 2 & 1 & 0 \\ 1 & 1 & 4 \end{pmatrix} \right\rangle \\ \left\langle \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \right\rangle\end{aligned}$$

4. Let  $(\mathcal{V}, \langle \cdot, \cdot \rangle)$  be an inner product space and  $\{A_1, \dots, A_n\}$  an orthogonal set of vectors in  $\mathcal{V}$ . Establish the following:

$$\text{Bessel's inequality: } \sum_{i=1}^n \frac{|\langle B, A_i \rangle|^2}{|A_i|^2} \leq |B|^2$$

$$\text{Parseval's identity: } \langle B, C \rangle = \sum \langle B, A_i \rangle \langle A_i, C \rangle$$

5. Let  $\mathcal{V}$  be an inner product space. Prove the following: (Draw pictures in  $\mathbb{R}^2$  to see what is going on!)

$$\text{Law of cosines: } |A - B|^2 = |A|^2 + |B|^2 - 2|A||B|\cos\theta$$

where  $\theta$  is the angle between  $A$  and  $B$ .

6. Which of the following sets of vectors in  $\mathbb{R}^3$  are orthogonal?

$$\begin{aligned}\{(0, 0, 0), (1, 1, 1), (0, 1, -1)\} \\ \{(1, 1, 1), (1, -2, 1), (-2, 1, -2)\} \\ \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\} \\ \{(1, -1, 2), (2, 0, -1), (0, 1, 1), (0, 0, 1)\}\end{aligned}$$

7. Which of the following sets of vectors in  $\mathcal{P}_3(\mathbb{R})$  are orthogonal?

$$\begin{aligned}\{1, x - \tfrac{1}{2}\} \\ \{1, x - 1\} \\ \{x, x^2 - 1\} \\ \{x, 3x^2 - 2, 2x^3\}\end{aligned}$$

8. Which of the following sets of vectors in  $\mathcal{M}_{33}$  are orthogonal?

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

$$\left\{ \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{pmatrix}, \begin{pmatrix} -3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \right\}$$

9. Find an orthonormal basis for each of the indicated subspaces:

$$\{(x, y, z) \in \mathbb{R}^3 \mid x + y - z = 0\}$$

$$\left\{ p(x) \in \mathcal{P}_3(\mathbb{R}) \mid x \frac{d}{dx} p(x) = p(x) \right\}$$

$$\{A \in \mathcal{M}_{33} \mid \text{tr}(A) = 0\}$$

10. Which of the following linear transformations are isometric embeddings?

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^4 \quad T(x, y, z) = (x + y, y + z, z + x, x + y + z)$$

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \quad T(x, y) = (x, 0, y)$$

$$T: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathcal{P}_3(\mathbb{R}) \quad T(p(x)) = xp(x)$$

$$T: \mathbb{R}^3 \rightarrow \mathcal{M}_{33} \quad T(x, y, z) = \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix}$$

$$T: \mathbb{R}^3 \rightarrow \mathcal{M}_{33} \quad T(x, y, z) = \begin{pmatrix} x & 0 & 0 \\ y & 0 & 0 \\ z & 0 & 0 \end{pmatrix}.$$

11. Find the orthogonal complement of each of the subspaces in Exercise 9.

12. A linear transformation  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is called a *rotation* iff  $T$  is an isometry and  $\det T > 0$ . Show that the matrix of  $T$  relative to the standard basis must be

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

for a unique angle  $\theta$ . In fact,  $\cos \theta = \langle \mathbf{E}_1, T(\mathbf{E}_1) \rangle$ . Draw a picture.

13. If  $T_1, T_2: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  are rotations show that  $T_1 T_2$  is also a rotation. Show  $T_1 T_2 = T_2 T_1$ .

14. Suppose  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  is an orthonormal basis for  $\mathcal{V}$ . If  $\mathbf{A}$  is a vector in  $\mathcal{V}$  such that

$$\langle \mathbf{A}, \mathbf{A}_i \rangle = 0, \quad i = 1, 2, \dots, n$$

show that  $\mathbf{A} = \mathbf{0}$ .



15. Verify that the trace of a square matrix has the following properties

- (a)  $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^t)$
- (b)  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$
- (c)  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$ .
- (d)  $\text{tr}(a\mathbf{A}) = a \text{tr} \mathbf{A}$
- (e) If  $\mathbf{A}$  is invertible then  $\text{tr}(\mathbf{ABA}^{-1}) = \text{tr}(\mathbf{B})$  for any two square matrices  $\mathbf{A}$  and  $\mathbf{B}$  and numbers  $a$ .

16. Show that there do not exist square matrices  $\mathbf{A}$  and  $\mathbf{B}$  such that  $\mathbf{AB} - \mathbf{BA} = \mathbf{I}$ .  
(Hint: see Exercise 15).

The notion of inner product can be introduced in complex vector spaces also, by a slight modification (remember: definitions are useful or useless *not* true or false) of the axioms we have used in the real case. To this end recall that for a complex number  $z = a + bi$  ( $a, b$  real numbers) the *complex conjugate* of  $z$  is the complex number

$$\bar{z} = a - bi.$$

It is very important to note that

$$z\bar{z} = (a + bi)(a - bi) = a^2 - b^2i^2 = a^2 + b^2$$

is a positive real number.

17. Let  $T: \mathbb{C} \rightarrow \mathbb{R}^2$  be given by  $T(a + bi) = (a, b)$ . Show that  $T$  is an isomorphism of real vector spaces and  $|T(z)|^2 = z\bar{z}$ .

**Definition.** If  $\mathcal{V}$  is a complex vector space, an inner-product on  $\mathcal{V}$  is a function that assigns to each pair  $\mathbf{A}, \mathbf{B}$  of vectors in  $\mathcal{V}$  a complex number  $\langle \mathbf{A}, \mathbf{B} \rangle$  such that

- (1)  $\langle \mathbf{A}, \mathbf{B} \rangle = \overline{\langle \mathbf{B}, \mathbf{A} \rangle}$  for all  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$ .
- (2) For all  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$  and  $z \in \mathbb{C}$  we have

$$\begin{aligned}\langle z\mathbf{A}, \mathbf{B} \rangle &= z\langle \mathbf{A}, \mathbf{B} \rangle, \\ \langle \mathbf{A}, z\mathbf{B} \rangle &= \bar{z}\langle \mathbf{A}, \mathbf{B} \rangle.\end{aligned}$$

- (3) For all vectors  $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{V}$  we have:

$$\begin{aligned}\langle \mathbf{A}, \mathbf{B} + \mathbf{C} \rangle &= \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{A}, \mathbf{C} \rangle, \\ \langle \mathbf{A} + \mathbf{B}, \mathbf{C} \rangle &= \langle \mathbf{A}, \mathbf{C} \rangle + \langle \mathbf{B}, \mathbf{C} \rangle.\end{aligned}$$

- (4) For all  $\mathbf{A}$  in  $\mathcal{V}$ ,  $\langle \mathbf{A}, \mathbf{A} \rangle \geq 0$  and  $\langle \mathbf{A}, \mathbf{A} \rangle = 0$  iff  $\mathbf{A} = \mathbf{0}$ .

18. On  $\mathbb{C}^n$  define

$$\langle \mathbf{Z}, \mathbf{W} \rangle = \sum_{i=1}^n z_i \bar{w}_i$$

where  $\mathbf{Z} = (z_1, z_2, \dots, z_n)$ ,  $\mathbf{W} = (w_1, \dots, w_n)$ . Verify that  $\langle \cdot, \cdot \rangle$  is an inner product on  $\mathbb{C}^n$ .

19. Compute the following inner products in  $\mathbb{C}^3$ :

$$\langle (3, i, 2 - i), (5, -i, 1) \rangle$$

$$\langle (i, i, i), (1, 1, 1) \rangle$$

$$\langle (i, 1, i), (1, i, 1) \rangle$$

20. The *length* of a vector  $\mathbf{Z}$  in a complex inner product space is by definition

$$|\mathbf{Z}| = \langle \mathbf{Z}, \mathbf{Z} \rangle^{1/2}.$$

Verify the triangle inequality in the complex case.

21. A basis for a complex inner product space is orthonormal if it consists of orthogonal vectors of length 1. Which of the following are orthonormal bases for  $\mathbb{C}^3$ ?

(a)  $(i, 0, 0), (0, i, 0), (0, 0, i)$

(b)  $(1, 0, 0), (0, i, 0), (0, 0, 1)$

(c)  $\sqrt{\frac{1}{2}}(1 + i, 1 - i, 0), \sqrt{\frac{1}{2}}(0, 1, 1), (0, 0, 1)$

(d)  $(1, i, 0), (-1, i, 0), (0, 0, 1)$

22. What changes, if any, are needed to prove the theorems of this section in the complex case. In particular show that two complex inner product spaces of finite dimensions are isometric iff they have the same dimension. In particular a complex inner product space of dimension  $n$  is isometric to  $\mathbb{C}^n$ .

23. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in an  $n$ -dimensional inner product space  $\mathcal{V}$ .  $T$  is called an *orthogonal reflection* iff there exists an  $(n - 1)$ -dimensional subspace  $\mathcal{H}$  in  $\mathcal{V}$  and a nonzero vector  $\mathbf{v}$  perpendicular to  $\mathcal{H}$  such that

$$T(\mathbf{h}) = \mathbf{h} \quad \text{for all } \mathbf{h} \text{ in } \mathcal{H},$$

$$T(\mathbf{v}) = -\mathbf{v}.$$

show

$$T(\mathbf{w}) = \mathbf{w} - 2\langle \mathbf{w}, \mathbf{v} \rangle \mathbf{v} \quad \text{for all } \mathbf{w}.$$

24. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a reflection. Show that with respect to a properly chosen inner product  $T$  is orthogonal.

25. Let  $S$  be a finite set,  $\mathcal{F}(S)$  the vector space of real valued functions on  $S$ , and  $\varphi: S \rightarrow S$  a function. Define

$$\varphi^*: \mathcal{F}(S) \rightarrow \mathcal{F}(S)$$

by

$$\varphi^*(\mathbf{f})(s) = \mathbf{f}(\varphi(s)).$$

By Chapter 8, Exercise 22,  $\varphi^*$  is a linear transformation. Show that the trace of  $\varphi^*$  is the number of fixed points of  $\varphi$ . (A point  $x \in S$  is called a fixed point of  $\varphi$  iff  $\varphi(x) = x$ .)

# 16

## The spectral theorem and quadratic forms

So far in our study of linear transformations we have concentrated on trying to find conditions that assure the matrix of the transformation has a particular form. We have not asked the related question of studying properties of those transformations whose matrix is *assumed* to have a particularly simple form. There is in fact a good reason for this and it is tied up with our work of the last chapter. For example we might propose to study those linear transformations whose matrix is symmetric. We would therefore like to introduce the following:

**Proposed Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation. We say that  $T$  is *symmetric* iff there is a basis  $\{A_1, \dots, A_n\}$  for  $\mathcal{V}$  such that the matrix of  $T$  relative to this basis is symmetric.

Now the reason we have indicated that this is only a proposed definition, and not one we intend to work with, is because it is not really a useful concept. Remember, a definition is not right or wrong, only useful or useless. The preceding one is pretty much useless. The following example will help illustrate why.

**EXAMPLE 1.** Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation given by

$$T(x, y) = (x + 2y, 2x + y).$$

Let us calculate the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^2$ . We find

$$T(1, 0) = (1, 2), \quad T(0, 1) = (2, 1)$$

so the matrix we seek is:

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

which is a symmetric matrix. On the other hand we may compute the matrix of  $T$  relative to the basis  $\{\mathbf{F}_1 = (1, 1), \mathbf{F}_2 = (1, 0)\}$ . We get

$$T(1, 1) = (3, 3) = 3\mathbf{F}_1 + 0\mathbf{F}_2$$

$$T(1, 0) = (1, 2) = 2\mathbf{F}_1 - \mathbf{F}_2$$

so the required matrix is

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

which is not symmetric.

The preceding example illustrates a crucial drawback to using the proposed definition of symmetric transformation. The concept is not a property of the linear transformation that is invariant under change of basis. This has several consequences. For one, as defined, symmetry is more a property of the coordinate system in which the matrix of the transformation is computed. Thus we will have to specify this basis and work only with it if we wish to take advantage of the symmetry property. Secondly, and more importantly, a linear transformation  $T$  might be symmetric in terms of the proposed definition and we would never know it because we are unable to discover a basis in which  $T$  has a symmetric matrix!

The aforementioned difficulties can be overcome by introducing a slightly different concept. One which depends only on the linear transformation and not on the choice of basis. *However* we must assume that the ambient vector space is equipped with a fixed inner product. The particular definition we are going to give was arrived at through a process of trial, error, and experimentation; the basis of all modern science.

We begin with an important construction.

**Lemma 16.1.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space and  $T: \mathcal{V} \rightarrow \mathcal{V}$  a linear endomorphism. For each vector  $\mathbf{A}$  in  $\mathcal{V}$  define the function*

$$T_{\mathbf{A}}: \mathcal{V} \rightarrow \mathbb{R} \quad \text{by} \quad T_{\mathbf{A}}(\mathbf{B}) = \langle T(\mathbf{B}), \mathbf{A} \rangle$$

*where  $\langle \cdot, \cdot \rangle$  is the inner product in  $\mathcal{V}$ . Then the function  $T_{\mathbf{A}}: \mathcal{V} \rightarrow \mathbb{R}$  is a linear transformation.*

**PROOF.** Suppose  $\mathbf{B}_1, \mathbf{B}_2$  belong to  $\mathcal{V}$ . Then

$$\begin{aligned} T_{\mathbf{A}}(\mathbf{B}_1 + \mathbf{B}_2) &= \langle T(\mathbf{B}_1 + \mathbf{B}_2), \mathbf{A} \rangle \\ &= \langle T(\mathbf{B}_1) + T(\mathbf{B}_2), \mathbf{A} \rangle \\ &= \langle T(\mathbf{B}_1), \mathbf{A} \rangle + \langle T(\mathbf{B}_2), \mathbf{A} \rangle \\ &= T_{\mathbf{A}}(\mathbf{B}_1) + T_{\mathbf{A}}(\mathbf{B}_2). \end{aligned}$$

A similar computation shows that  $T_A$  preserves scalar multiples, and hence  $T_A$  is a linear transformation.  $\square$

Suppose again that  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a linear transformation in the finite-dimensional inner product space  $\mathcal{V}$ . For each  $A$  in  $\mathcal{V}$  there is also the linear transformation

$$T_A: \mathcal{V} \rightarrow \mathbb{R} \quad \text{given by} \quad T_A(\mathbf{B}) = \langle T(\mathbf{B}), \mathbf{A} \rangle.$$

According to (15.20) there exists a unique vector  $A^*$  in  $\mathcal{V}$  such that

$$T_A(\mathbf{B}) = \langle \mathbf{B}, A^* \rangle$$

for all  $\mathbf{B}$  in  $\mathcal{V}$ . Define a function

$$T^*: \mathcal{V} \rightarrow \mathcal{V} \quad \text{by} \quad T^*(A) = A^*$$

The function  $T^*$  is called the *adjoint* of  $T$ .

**Proposition 16.2.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the finite-dimensional inner product space  $\mathcal{V}$ . Then the adjoint of  $T$ ,*

$$T^*: \mathcal{V} \rightarrow \mathcal{V}$$

*is also a linear transformation.*

**PROOF.** We are going to beat to death the word unique that occurs in the statement of (15.20).

The adjoint of  $T$  is characterized by the equations

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = T_A(\mathbf{B}) = \langle \mathbf{B}, T^*(A) \rangle$$

for all vectors  $A, \mathbf{B}$  in  $\mathcal{V}$ . So suppose that  $A_1, A_2$  belong to  $\mathcal{V}$ . Then for any  $\mathbf{B}$  in  $\mathcal{V}$

$$\begin{aligned} \langle \mathbf{B}, T^*(A_1 + A_2) \rangle &= \langle T(\mathbf{B}), A_1 + A_2 \rangle \\ &= \langle T(\mathbf{B}), A_1 \rangle + \langle T(\mathbf{B}), A_2 \rangle \\ &= \langle \mathbf{B}, T^*(A_1) \rangle + \langle \mathbf{B}, T^*(A_2) \rangle \\ &= \langle \mathbf{B}, T^*(A_1) + T^*(A_2) \rangle. \end{aligned}$$

Now set  $A = A_1 + A_2$ . Then the above equations show

$$\langle \mathbf{B}, T^*(A_1 + A_2) \rangle = T_A(\mathbf{B}) = \langle \mathbf{B}, T^*(A_1) + T^*(A_2) \rangle$$

for all  $\mathbf{B}$  in  $\mathcal{V}$ . By the **uniqueness** part of (15.20) we therefore find

$$T^*(A_1 + A_2) = T^*(A_1) + T^*(A_2)$$

so that  $T^*$  preserves vector sums. To show that  $T^*$  preserves scalar products we suppose  $A$  in  $\mathcal{V}$  and  $r$  a number. Then for all  $\mathbf{B}$  in  $\mathcal{V}$

$$\begin{aligned} \langle \mathbf{B}, T^*(rA) \rangle &= \langle T(\mathbf{B}), rA \rangle \\ &= r \langle T(\mathbf{B}), A \rangle \\ &= r \langle \mathbf{B}, T^*(A) \rangle \\ &= \langle \mathbf{B}, rT^*(A) \rangle \end{aligned}$$

and therefore

$$\langle \mathbf{B}, T^*(r\mathbf{A}) \rangle = T_{r\mathbf{A}}(\mathbf{B}) = \langle \mathbf{B}, rT^*(\mathbf{A}) \rangle$$

and the uniqueness part of (15.20) yields

$$T^*(r\mathbf{A}) = rT^*(\mathbf{A})$$

so that  $T^*$  preserves scalar products.  $\square$

**Proposition 16.3.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the finite-dimensional inner product space  $\mathcal{V}$ . Then*

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, T^*(\mathbf{A}) \rangle$$

for all  $\mathbf{A}, \mathbf{B} \in \mathcal{V}$ .

PROOF. This follows instantly from the formulas

$$T_{\mathbf{A}}(\mathbf{B}) = \langle T(\mathbf{B}), \mathbf{A} \rangle$$

$$T_{\mathbf{A}}(\mathbf{B}) = \langle \mathbf{B}, T^*(\mathbf{A}) \rangle$$

which have already found application in (16.2).  $\square$

Let us work a few numerical examples.

EXAMPLE 2. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation given by

$$T(x, y) = (x + 2y, 2x + y).$$

Calculate  $T^*: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .

Solution. To compute  $T^*(u, v)$  we will use the equation

$$\langle (x, y), T^*(u, v) \rangle = \langle T(x, y), (u, v) \rangle$$

for all  $(x, y), (u, v)$  in  $\mathbb{R}^2$ . We find

$$\begin{aligned} \langle (x, y), T^*(u, v) \rangle &= \langle T(x, y), (u, v) \rangle \\ &= \langle (x + 2y, y + 2x), (u, v) \rangle \\ &= xu + 2yu + yv + 2xv \\ &= xu + 2xv + yv + 2yu \\ &= x(u + 2v) + y(v + 2u) \\ &= \langle (x, y), (u + 2v, v + 2u) \rangle \end{aligned}$$

and therefore

$$T^*(u, v) = (u + 2v, v + 2u)$$

or if you prefer  $x$  and  $y$

$$T^*(x, y) = (x + 2y, y + 2x).$$

Therefore  $T^* = T$ .

EXAMPLE 3. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation given by

$$T(x, y) = (x + 3y, 2x + y).$$

Calculate  $T^*: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .

*Solution.* We have

$$\begin{aligned} \langle (x, y), T^*(u, v) \rangle &= \langle T(x, y), (u, v) \rangle \\ &= \langle (x + 3y, 2x + y), (u, v) \rangle \\ &= xu + 3yu + 2xv + yv \\ &= xu + 2xv + 3yu + yv \\ &= x(u + 2v) + y(3u + v) \\ &= \langle (x, y), (u + 2v, 3u + v) \rangle \end{aligned}$$

and therefore

$$T^*(u, v) = (u + 2v, 3u + v)$$

and  $T \neq T^*$ .

**Definition.** Let  $\mathcal{V}$  be a finite-dimensional inner product space. A linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  is called *self-adjoint* iff  $T = T^*$ .

Notice that if  $T: \mathcal{V} \rightarrow \mathcal{V}$  is self-adjoint then since  $T = T^*$

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, T(\mathbf{A}) \rangle$$

for all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . On the other hand if  $T: \mathcal{V} \rightarrow \mathcal{V}$  is a linear transformation such that

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, T(\mathbf{A}) \rangle$$

for all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$  then  $T$  is self-adjoint. To see this recall we always have the equation

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, T^*(\mathbf{A}) \rangle$$

for all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . Hence for our hypothesized  $T$  we will have

$$\langle \mathbf{B}, T(\mathbf{A}) \rangle = \langle \mathbf{B}, T^*(\mathbf{A}) \rangle$$

for all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . This may be rewritten as

$$\langle \mathbf{B}, T(\mathbf{A}) - T^*(\mathbf{A}) \rangle = 0$$

for all vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . In particular putting  $\mathbf{B} = T(\mathbf{A}) - T^*(\mathbf{A})$  we find

$$\langle T(\mathbf{A}) - T^*(\mathbf{A}), T(\mathbf{A}) - T^*(\mathbf{A}) \rangle = 0$$

for all vectors  $\mathbf{A}$  in  $\mathcal{V}$ . Therefore  $T(\mathbf{A}) - T^*(\mathbf{A}) = \mathbf{0}$  for all vectors  $\mathbf{A}$  in  $\mathcal{V}$  or what is the same thing

$$T(\mathbf{A}) = T^*(\mathbf{A})$$

for all vectors  $\mathbf{A}$  in  $\mathcal{V}$ . Hence  $T = T^*$  so that  $T$  is self-adjoint as we claimed.

As an example of how we may apply the above fact to check if a transformation is self-adjoint consider the linear transformation

$$T: \mathcal{M}_{nn} \rightarrow \mathcal{M}_{nn}$$

given by

$$T(A) = A'$$

where  $A'$  is the transpose of the  $n \times n$  matrix  $A$ . We have

$$\begin{aligned}\langle T(B'), A \rangle &= \text{tr}(T(B')A') \\ &= \text{tr}(B''A') \\ &= \text{tr}(BA')\end{aligned}$$

and

$$\begin{aligned}\langle B', T(A) \rangle &= \text{tr}(B'(T(A))') \\ &= \text{tr}(B'A'') \\ &= \text{tr}(B'A).\end{aligned}$$

Using the definition of the trace it is very easy to verify (see Chapter 15, Exercise 15)

$$\begin{aligned}\text{tr}(C') &= \text{tr}(C) \\ \text{tr}(CD) &= \text{tr}(DC).\end{aligned}$$

Recall also that

$$(CD)' = D'C'.$$

We therefore may compute as follows:

$$\text{tr}(B'A) = \text{tr}(AB') = \text{tr}((AB')') = \text{tr}(B''A') = \text{tr}(BA').$$

Therefore

$$\langle T(B'), A \rangle = \text{tr}(BA') = \text{tr}(B'A) = \langle B', T(A) \rangle$$

and hence  $T: \mathcal{M}_{nn} \rightarrow \mathcal{M}_{nn}$  must be self-adjoint.

Of course there are other methods to check if a linear transformation is self-adjoint. One such method is to examine the matrix relative to an orthonormal basis.

**Theorem 16.4.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space,  $T: \mathcal{V} \rightarrow \mathcal{V}$  a linear transformation and  $\{A_1, \dots, A_n\}$  an orthonormal basis for  $\mathcal{V}$ . Let  $M$  be the matrix of  $T$  relative to the basis  $\{A_1, \dots, A_n\}$  and  $N$  the matrix of  $T^*$  relative to the basis  $\{A_1, \dots, A_n\}$ . Then  $N = M'$ .*

**PROOF.** What we need is (15.9) and the definitions. Let  $(M) = (m_{ij})$  and  $N = (n_{ij})$ . Then

$$\begin{aligned}T(A_j) &= \sum m_{ij} A_i \\ T^*(A_s) &= \sum n_{rs} A_r.\end{aligned}$$



According to (15.9)

$$m_{ij} = \langle T(A_j), A_i \rangle$$

and

$$n_{rs} = \langle T^*(A_s), A_r \rangle.$$

But

$$\langle T(A_j), A_i \rangle = \langle A_j, T^*(A_i) \rangle = \langle T^*(A_i), A_j \rangle$$

so that

$$m_{ij} = n_{ji}.$$

But this just says  $N = M^t$  as required.  $\square$

**Theorem 16.5.** *Let  $\mathcal{V}$  be a finite-dimensional inner product space and  $T: \mathcal{V} \rightarrow \mathcal{V}$  a self-adjoint linear transformation. If  $\{A_1, \dots, A_n\}$  is an orthonormal basis for  $\mathcal{V}$  then the matrix of  $T$  relative to this basis is symmetric.*

For this reason self-adjoint transformations are often called symmetric.

**PROOF.** Let  $M$  be the matrix of  $T$  and  $N$  the matrix of  $T^*$ , both relative to the basis  $\{A_1, \dots, A_n\}$ . Then  $N = M^t$  by (16.4). But since  $T$  is self-adjoint  $T = T^*$  so  $N = M$ . Thus  $M = N = M^t$ . That is,  $M = M^t$  so  $M$  is symmetric.  $\square$

**EXAMPLE 4.** Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation given by

$$T(x, y, z) = (x - y + z, x + y, z - x).$$

Decide if  $T$  is self-adjoint.

**Solution.** The matrix of  $T$  relative to the standard orthonormal base may be easily computed. We get

$$T(1, 0, 0) = (1, 1, -1)$$

$$T(0, 1, 0) = (-1, 1, 0)$$

$$T(0, 0, 1) = (1, 0, 1),$$

so the required matrix is

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

Then

$$M^t = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

so that  $\mathbf{M} \neq \mathbf{M}^t$ , that is  $\mathbf{M}$  is not symmetric. By (16.5) it follows that  $T$  is not self-adjoint.

Actually we have enough information to compute the adjoint  $T^*: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . In view of (16.4) we need only calculate

$$\mathbf{M}^t \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + y - z \\ y - x \\ x + z \end{pmatrix}$$

from which we get the formula

$$T^*(x, y, z) = (x + y - z, y - x, x + z).$$

Of course we could also calculate  $T^*$  by using the method of Examples 2 and 3. You ought to try this.

**Theorem 16.6.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be an endomorphism of the finite-dimensional inner product space  $\mathcal{V}$ . Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  be an orthonormal basis for  $\mathcal{V}$  and suppose that the matrix of  $T$  is symmetric with respect to this basis. Then  $T$  is self-adjoint.*

**PROOF.** Let  $\mathbf{M}$  be the matrix of  $T$  relative to the basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ . By (16.4) the matrix of  $T^*$  relative to this basis is also  $\mathbf{M}$  since  $\mathbf{M}$  is symmetric. Therefore by (12.1)  $T = T^*$ .  $\square$

**ALTERNATE PROOF.** Let  $\mathbf{M}$  be the matrix of  $T$  relative to the basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$ . Let  $\mathbf{A}$  and  $\mathbf{B}$  be vectors in  $\mathcal{V}$  with

$$\begin{aligned} \mathbf{A} &= a_1 \mathbf{A}_1 + a_2 \mathbf{A}_2 + \cdots + a_n \mathbf{A}_n \\ \mathbf{B} &= b_1 \mathbf{A}_1 + b_2 \mathbf{A}_2 + \cdots + b_n \mathbf{A}_n. \end{aligned}$$

Then if  $\mathbf{M} = (m_{ij})$

$$\begin{aligned} T(\mathbf{A}) &= \sum_{i,j} m_{ij} a_j \mathbf{A}_i \\ T(\mathbf{B}) &= \sum_{i,j} m_{ij} b_j \mathbf{A}_i. \end{aligned}$$

Applying (15.9) we get

$$\begin{aligned} \langle T(\mathbf{A}), \mathbf{B} \rangle &= \sum_{i,j} m_{ij} a_j b_i = \sum_{r,s} a_r b_s m_{sr} \\ \langle \mathbf{A}, T(\mathbf{B}) \rangle &= \sum_{i,j} a_i m_{ij} b_j = \sum_{r,s} a_r b_s m_{rs}. \end{aligned}$$

Since  $\mathbf{M}$  is symmetric  $m_{rs} = m_{sr}$  so we get

$$\langle T(\mathbf{A}), \mathbf{B} \rangle = \sum_{r,s} a_r b_s m_{sr} = \sum_{r,s} a_r b_s m_{rs} = \langle \mathbf{A}, T(\mathbf{B}) \rangle$$

for all  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . But according to the definition of  $T^*$ ,

$$\langle T(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, T^*(\mathbf{A}) \rangle,$$

for all  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . Therefore we get

$$\langle \mathbf{B}, \mathbf{T}(\mathbf{A}) \rangle = \langle \mathbf{T}(\mathbf{A}), \mathbf{B} \rangle = \langle \mathbf{A}, \mathbf{T}(\mathbf{B}) \rangle = \langle \mathbf{T}(\mathbf{B}), \mathbf{A} \rangle = \langle \mathbf{B}, \mathbf{T}^*(\mathbf{A}) \rangle$$

so that

$$\langle \mathbf{B}, \mathbf{T}(\mathbf{A}) \rangle = \langle \mathbf{B}, \mathbf{T}^*(\mathbf{A}) \rangle$$

for all  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$ . The definition of the adjoint now gives  $\mathbf{T}(\mathbf{A}) = \mathbf{T}^*(\mathbf{A})$  for all  $\mathbf{A}$  in  $\mathcal{V}$ . Thus  $\mathbf{T} = \mathbf{T}^*$ .  $\square$

Notice that (16.5) provides us with a painless method for constructing loads of self-adjoint linear transformations. All we have to do is take a symmetric matrix  $\mathbf{S}$  of size  $n$ , an  $n$ -dimensional inner product space  $\mathcal{V}$  and an orthonormal basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  for  $\mathcal{V}$ . The linear transformation  $\mathbf{T} : \mathcal{V} \rightarrow \mathcal{V}$  whose matrix is  $\mathbf{S}$  in the basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  will then be self-adjoint.

Notice also that (16.5) suggests ways to modify the proposed definition we made of a symmetric linear transformation so that we get a useful concept. We will not pursue this idea here, but relegate it to a flock of exercises.

We are ready now to state the main result of this chapter (and indeed of this book so far).

**Theorem 16.7** (Spectral theorem). *Let  $\mathbf{T} : \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation in the finite-dimensional vector space  $\mathcal{V}$ . Then there exists an orthonormal basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_n\}$  and numbers  $\lambda_1, \dots, \lambda_n$  such that*

$$\mathbf{T}(\mathbf{A}_i) = \lambda_i \mathbf{A}_i \quad i = 1, 2, \dots, n.$$

Therefore the matrix of  $\mathbf{T}$  relative to this basis is

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{pmatrix}.$$

Note that it is not asserted that the numbers  $\lambda_1, \dots, \lambda_n$  are distinct; they need not be. What Theorem (16.7) says is that the characteristic polynomial of  $\mathbf{T}$  factors completely and the geometric multiplicity of any eigenvalue coincides with its algebraic multiplicity.

The proof of (16.7) is an induction argument. We will give the proof for  $n = 2$  and 3. The general case is sketched out in the exercises. We need two preliminary results.

**Proposition 16.8.** *Let  $\mathbf{T} : \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation in the finite-dimensional inner product space  $\mathcal{V}$ . Suppose that  $\mathcal{W}$  is a linear subspace of  $\mathcal{V}$  with the property that  $\mathbf{T}(\mathcal{W}) \subset \mathcal{W}$ . Then  $\mathcal{W}^\perp$  has the same property, namely,  $\mathbf{T}(\mathcal{W}^\perp) \subset \mathcal{W}^\perp$ .*

PROOF. Let  $A \in \mathcal{W}^\perp$ . We must show that

$$\langle T(A), B \rangle = 0$$

for all  $B$  in  $\mathcal{W}$ . We have

$$\begin{aligned}\langle T(A), B \rangle &= \langle A, T^*(B) \rangle \\ &= \langle A, T(B) \rangle \quad \text{since } T = T^* \\ &= 0 \quad \text{since } T(B) \in \mathcal{W} \text{ and } A \in \mathcal{W}^\perp\end{aligned}$$

as required. □

**Proposition 16.9.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation in the finite-dimensional inner product space  $\mathcal{V}$ . Suppose that  $e'$  and  $e''$  are distinct eigenvalues of  $T$  with corresponding eigenvectors  $E'$  and  $E''$ . Then  $\langle E', E'' \rangle = 0$ , that is,  $E'$  and  $E''$  are perpendicular.*

PROOF. By hypothesis we have

$$T(E') = e'E', \quad T(E'') = e''E''$$

so since  $T$  is self-adjoint we get

$$\begin{aligned}e'\langle E', E'' \rangle &= \langle e'E', E'' \rangle = \langle T(E'), E'' \rangle \\ &= \langle E', T^*(E'') \rangle = \langle E', T(E'') \rangle \\ &= \langle E', e''E'' \rangle = e''\langle E', E'' \rangle\end{aligned}$$

which means

$$e'\langle E', E'' \rangle = e''\langle E', E'' \rangle.$$

Since  $e' \neq e''$  by hypothesis we must have  $\langle E', E'' \rangle = 0$  as required. □

PROOF OF (16.7) FOR  $n = 2$ . Let  $\{E_1, E_2\}$  be an orthonormal base for  $\mathcal{V}$ . By (16.4) the matrix  $M$  of  $T$  relative to this basis is symmetric, say

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

If  $b = 0$  there is nothing to prove, so we will assume that  $b \neq 0$ . The characteristic polynomial of  $T$  is given by

$$\begin{aligned}\Delta(t) &= \det(M - tI) = \det \begin{pmatrix} a - t & b \\ b & c - t \end{pmatrix} \\ &= (a - t)(c - t) - b^2 \\ &= t^2 - (a + c)t + ac - b^2.\end{aligned}$$

By the quadratic formula the roots of  $\Delta(t)$  are given by

$$\begin{aligned} r_{\pm} &= \frac{a + c \pm \sqrt{(a + c)^2 - 4(ac - b^2)}}{2} \\ &= \frac{a + c \pm \sqrt{a^2 + 2ac + c^2 - 4ac + 4ab^2}}{2} \\ &= \frac{a + c \pm \sqrt{4b^2 + (a - c)^2}}{2}. \end{aligned}$$

Notice that since  $b \neq 0$

$$4b^2 + (a - c)^2 > 0$$

so that the characteristic roots are distinct. Let  $\{\mathbf{F}_+, \mathbf{F}_-\}$  be the corresponding eigenvectors. By (16.8)  $\mathbf{F}_+$  and  $\mathbf{F}_-$  are orthogonal. Since we may normalize  $\mathbf{F}_+$  and  $\mathbf{F}_-$  they constitute an orthonormal basis of eigenvectors completing the proof.  $\square$

**PROOF OF (16.7) FOR  $n = 3$ .** Let  $\Delta(t)$  be the characteristic polynomial of  $T$ . Then  $\Delta(t)$  is of degree 3 and hence must have at least one *real* root, say  $r$ . Then  $r$  is an eigenvalue of  $T$ . Let  $\mathbf{E}$  be a corresponding eigenvector of unit length and  $\mathcal{W} = \mathcal{L}(\mathbf{E}) \subset \mathcal{V}$ . Of course  $T(\mathcal{W}) \subset \mathcal{W}$  and therefore by (16.7)  $T(\mathcal{W}^\perp) \subset \mathcal{W}^\perp$ . Let  $\mathbb{F}_1, \mathbb{F}_2$  be an orthonormal basis for  $\mathcal{W}^\perp$ . Then  $\{\mathbf{E}, \mathbb{F}_1, \mathbb{F}_2\}$  is an orthonormal basis for  $\mathcal{V}$  (see the proof of (15.18)). By (16.4) the matrix  $\mathbf{M}$  of  $T$  relative to the basis  $\{\mathbf{E}, \mathbb{F}_1, \mathbb{F}_2\}$  will be symmetric. Since  $e$  is an eigenvalue it must look like

$$\mathbf{M} = \begin{pmatrix} e & 0 & 0 \\ 0 & a & b \\ 0 & b & c \end{pmatrix}$$

The characteristic polynomial of  $T$  is therefore

$$\Delta(t) = (e - t)(t^2 - (a + c)t + ac - b^2).$$

If  $b = 0$  there is nothing to prove. If  $b \neq 0$  then as in the case  $n = 2$  there are two distinct eigenvalues,  $e_1, e_2$  the zeros of the quadratic

$$t^2 - (a + c)t + ac - b^2.$$

If  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are corresponding eigenvectors then  $\mathbf{E}_1, \mathbf{E}_2$  are orthogonal by (16.8) while they both are orthogonal to  $\mathbf{E}$  since they belong to  $\mathcal{W}^\perp$ . Therefore  $\{\mathbf{E}, \mathbf{E}_1, \mathbf{E}_2\}$  is an orthonormal basis for  $\mathcal{V}$  and the matrix of  $T$  relative to this basis is

$$\mathbf{M} = \begin{pmatrix} e & 0 & 0 \\ 0 & e_1 & 0 \\ 0 & 0 & e_2 \end{pmatrix}.$$

(Note that it is possible that  $e = e_1$  or  $e = e_2$  even if  $b \neq 0$ . If  $b = 0$  it is even possible for  $e_1 = e = e_2$ .)  $\square$

EXAMPLE 5. Let  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation given by

$$T(x, y) = (x + 2y, 2y + x).$$

Find an orthonormal basis of eigenvectors of  $T$  for  $\mathbb{R}^2$ .

*Solution.* We already have seen that  $T$  is self-adjoint. Therefore by applying (16.7) we know that it is possible to find an orthonormal basis for  $\mathbb{R}^2$  composed of eigenvectors of  $T$ . As a matter of fact the orthogonality will take care of itself because of (16.1). This will even provide a numerical check on our work.

The matrix of  $T$  relative to the standard orthonormal basis of  $\mathbb{R}^2$  was computed to be

$$M = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

in Example 1. The characteristic polynomial of  $T$  is therefore

$$\begin{aligned} \Delta(t) &= \det(M - tI) = \det \begin{pmatrix} 1-t & 2 \\ 2 & 1-t \end{pmatrix} \\ &= (1-t)^2 - 4 = t^2 - 2t + 1 - 4 \\ &= (t^2 - 2t - 3) = (t-3)(t+1). \end{aligned}$$

The eigenvalues of  $T$  are therefore 3 and  $-1$ . Corresponding eigenvectors are easily computed

$$e = 3 \quad \begin{cases} -2x + 2y = 0 \\ -2x + 2y = 0 \end{cases}$$

$x = y$  so  $(1, 1)$  is an eigenvector and  $(\sqrt{2}/2, \sqrt{2}/2)$  an eigenvector of unit length.

$$e = -1 \quad \begin{cases} 2x + 2y = 0 \\ 2x + 2y = 0 \end{cases}$$

$x = -y$  so  $(1, -1)$  is an eigenvector and  $(\sqrt{2}/2, -\sqrt{2}/2)$  an eigenvector of unit length. Therefore  $F_1 = (\sqrt{2}/2, \sqrt{2}/2)$ ,  $F_2 = (\sqrt{2}/2, -\sqrt{2}/2)$  is, by (16.6), an orthonormal basis of eigenvectors of  $T$  for  $\mathbb{R}^2$ . We can check the orthogonality easily enough

$$\langle F_1, F_2 \rangle = \frac{2}{4} - \frac{2}{4} = 0$$

as required.

EXAMPLE 6. Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be given by

$$T(x, y, z) = (3x - z, 2y, -x + 3z).$$

Verify that  $T$  is self-adjoint and put it into diagonal form.

*Solution.* By direct calculation

$$T(1, 0, 0) = (3, 0, -1)$$

$$T(0, 1, 0) = (0, 2, 0)$$

$$T(0, 0, 1) = (-1, 0, 3).$$

Therefore the matrix of  $T$  relative to the standard basis of  $\mathbb{R}^3$  is

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

which is symmetric, so  $T$  is self-adjoint. The characteristic polynomial of  $T$  is

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

Therefore the eigenvalues of  $T$  are  $t = 2$ ,  $t = 4$  with multiplicities 2 and 1 respectively. A basis for the eigenspace corresponding to the eigenvalue 2 is obtained by solving the linear system

$$T(x, y, z) = 2(x, y, z),$$

that is,

$$\begin{aligned} 3x - z &= 2x \\ 2y &= 2y \\ -x + 3z &= 2z \end{aligned}$$

or

$$\begin{aligned} x - z &= 0 \\ -x + z &= 0 \end{aligned}$$

so a basis for the eigenspace corresponding to eigenvalue 2 consists of the vectors  $(1, 0, 1)$ ,  $(0, 1, 0)$ .

To find an eigenvector corresponding to eigenvalue 4 we solve

$$T(x, y, z) = 4(x, y, z)$$

that is,

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

or

$$\begin{aligned} -z &= x \\ -2y &= 0 \\ -x &= z \end{aligned}$$

which has  $(-1, 0, 1)$  as a solution. Therefore the matrix of  $T$  relative to the ordered basis  $\mathbf{f}_1 = (1, 0, 1)$ ,  $\mathbf{f}_2 = (0, 1, 0)$ ,  $\mathbf{f}_3 = (-1, 0, 1)$  is

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

Notice that  $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$  is an orthogonal basis.

EXAMPLE 7. Put  $T: \mathcal{M}_{2,2} \rightarrow \mathcal{M}_{2,2}$  given by  $T(A) = A^t$  into diagonal form.

*Solution.* We showed that  $T$  is self adjoint, so by the spectral theorem  $T$  is diagonalizable. As an orthonormal basis for  $\mathcal{M}_{2,2}$  we choose:

$$\begin{aligned} \mathbf{M}_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, & \mathbf{M}_2 &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ \mathbf{M}_3 &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, & \mathbf{M}_4 &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

So

$$T(\mathbf{M}_1) = \mathbf{M}_1, \quad T(\mathbf{M}_2) = \mathbf{M}_3, \quad T(\mathbf{M}_3) = \mathbf{M}_2, \quad T(\mathbf{M}_4) = \mathbf{M}_4,$$

and the matrix of  $T$  relative to this basis is

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

*Note.* From  $T(\mathbf{M}_1) = \mathbf{M}_1$ , and  $T(\mathbf{M}_4) = \mathbf{M}_4$ , we see immediately that  $\mathbf{M}_1$  and  $\mathbf{M}_4$  are eigenvectors of  $T$  corresponding to eigenvalues 1.

The characteristic polynomial of  $T$  is

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} 1-t & 0 & 0 & 0 \\ 0 & -t & 1 & 0 \\ 0 & 1 & -t & 0 \\ 0 & 0 & 0 & 1-t \end{pmatrix} \\ &= (1-t) \det \begin{pmatrix} -t & 1 & 0 \\ 1 & -t & 0 \\ 0 & 0 & 1-t \end{pmatrix} \\ &= (1-t)(-1)^{3+3}(1-t) \det \begin{pmatrix} -t & 1 \\ 1 & -t \end{pmatrix} \\ &= (1-t)^2(t^2-1) = (t-1)^3(t+1), \end{aligned}$$



so the eigenvalues of  $T$  are  $+1$ , and  $-1$  with multiplicity 3 and 1, respectively. By the spectral theorem  $\dim \mathcal{V}_1 = 3$ ,  $\dim \mathcal{V}_{-1} = 1$ . We can find three orthonormal eigenvectors of  $T$  corresponding to eigenvalue  $+1$  by solving the system

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} x \\ z \\ y \\ w \end{pmatrix}$$

which is

$$x = x, \quad y = z, \quad z = y, \quad w = w.$$

So

$$(1, 0, 0, 0), \quad \sqrt{\frac{1}{2}}(0, 1, 1, 0), \quad (0, 0, 0, 1)$$

are the components relative to  $\{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4\}$  of an orthonormal basis for  $\mathcal{V}_1$ . The corresponding vectors of  $\mathcal{M}_{2,2}$  are

$$\mathbf{N}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{N}_2 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{N}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

To find a basis for the eigenspace  $\mathcal{V}_{-1}$  requires we solve the system

$$-\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} x \\ z \\ y \\ w \end{pmatrix},$$

which is

$$-x = x, \quad -y = z, \quad -z = y, \quad -w = w.$$

Thus  $x = 0$ ,  $w = 0$ , and  $y = 1$ ,  $z = -1$ . Normalizing this we have

$$\sqrt{\frac{1}{2}}(0, 1, -1, 0)$$

as a basis for the solution space relative to the basis  $\{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4\}$  so that the vector

$$\mathbf{N}_4 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is an orthonormal base for  $\mathcal{V}_{-1}$ . The matrix of  $T$  relative to the basis  $\{\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \mathbf{N}_4\}$  is

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

The final topic we will take up is the so-called *principal axis theorem* for quadratic forms. In the calculus one spends considerable time in studying the graphs of quadratic equations; these graphs being the conic sections, ellipse, hyperbola, parabola, and various degenerate cases. There is a foolproof procedure for drawing the graph of say

$$2x^2 + 3y^2 = 7,$$

but for

$$x^2 + 4xy + y^2 = 7$$

there is a difficulty caused by the cross-product term “ $4xy$ .” Of course there is a formula for rotating the coordinate system so as to eliminate the cross-product term, and it is just this formula that we wish to discuss in the light of our work with self-adjoint linear transformations! Rather than introduce the final, sophisticated version of the theory let us consider an example first.

EXAMPLE 8. Sketch the graph of

$$x^2 + 4xy + y^2 = 7.$$

*Solution.* Make the following observation

$$x^2 + 4xy + y^2 = (x, y) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

What we wish to discover is a new coordinate system for  $\mathbb{R}^2$  in which the quadratic form has no cross-product term. Let

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

be the linear transformation whose matrix in the standard coordinate system is

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

That is,

$$T(x, y) = (x + 2y, y + 2x).$$

Then we see that

$$(x, y) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \langle (x, y), T(x, y) \rangle.$$

That is, for a vector  $\mathbf{A}$  in  $\mathbb{R}^2$  with coordinates  $(x, y)$  we have

$$x^2 + 4xy + y^2 = \langle \mathbf{A}, T(\mathbf{A}) \rangle.$$

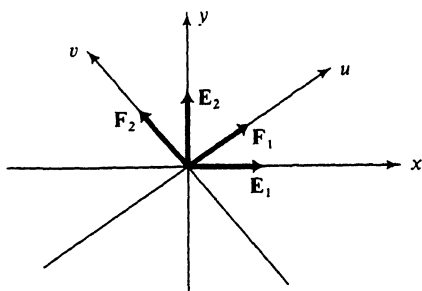


Figure 16.1

The number  $\langle A, T(A) \rangle$  depends, not on the coordinate system in which we express  $A$ , but only on  $T$ . Thus we could equally well use the coordinate system determined by the unit eigenvectors of  $T$ , namely

$$F_1 = \left( \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right), \quad F_2 = \left( -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right).$$

See Figure 16.1. Let us write  $(u, v)$  for the coordinates of a vector in the coordinate system determined by the vectors  $F_1, F_2$ . Then

$$\langle A, T(A) \rangle = (u, v) \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 3u^2 - v^2$$

since the matrix of  $T$  relative to the  $\{F_1, F_2\}$  base is

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

Thus our original quadratic form

$$7 = x^2 + 4xy + y^2$$

may be written (changing coordinate systems has not changed the value of  $\langle A, T(A) \rangle$ )

$$7 = 3u^2 - v^2$$

in the  $(u, v)$  coordinate system. One readily checks that this is a hyperbola and the graph is easily sketched as in Figure 16.2. The vertices of the hyperbola are at the points  $(u, v) = (\pm\sqrt{7/3}, 0)$ . The  $(x, y)$  coordinates of these points are easily computed

$$(x, y) = \pm\sqrt{\frac{7}{3}} F_1 + 0F_2 = \pm\sqrt{\frac{7}{3}} \left( \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right).$$

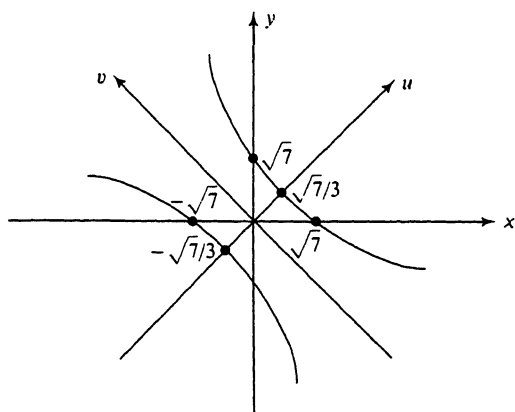


Figure 16.2

The relation between the  $(x, y)$  and  $(u, v)$  coordinates of a point  $A$  in  $\mathbb{R}^2$  is computed thusly:

$$\begin{aligned} x\mathbf{E}_1 + y\mathbf{E}_2 = \mathbf{A} &= u\mathbf{F}_1 + v\mathbf{F}_2 \\ &= u\left(\frac{\sqrt{2}}{2}\mathbf{E}_1 + \frac{\sqrt{2}}{2}\mathbf{E}_2\right) + v\left(-\frac{\sqrt{2}}{2}\mathbf{E}_1 + \frac{\sqrt{2}}{2}\mathbf{E}_2\right) \\ &= \frac{\sqrt{2}}{2}(u - v)\mathbf{E}_1 + \frac{\sqrt{2}}{2}(u + v)\mathbf{E}_2 \end{aligned}$$

whence,

$$\begin{aligned} x &= \frac{\sqrt{2}}{2}u - \frac{\sqrt{2}}{2}v \\ y &= \frac{\sqrt{2}}{2}u + \frac{\sqrt{2}}{2}v \end{aligned}$$

or equivalently

$$\begin{aligned} u\mathbf{F}_1 + v\mathbf{F}_2 = \mathbf{A} &= x\mathbf{E}_1 + y\mathbf{E}_2 \\ &= x\left(\frac{\sqrt{2}}{2}\mathbf{F}_1 - \frac{\sqrt{2}}{2}\mathbf{F}_2\right) + y\left(\frac{\sqrt{2}}{2}\mathbf{F}_1 + \frac{\sqrt{2}}{2}\mathbf{F}_2\right) \\ &= \frac{\sqrt{2}}{2}(x + y)\mathbf{F}_1 + \frac{\sqrt{2}}{2}(y - x)\mathbf{F}_2 \end{aligned}$$

whence

$$\begin{aligned} u &= \frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y \\ v &= -\frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y. \end{aligned}$$

The equations

$$\sqrt{3}u \pm v = \sqrt{7}$$

for the asymptotes of the hyperbola may thus be rewritten

$$\sqrt{3} \left( \frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y \right) \pm \left( -\frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y \right) = \sqrt{7}$$

or

$$\frac{\sqrt{6} \mp \sqrt{2}}{2}x + \frac{\sqrt{6} \pm \sqrt{2}}{2}y = \sqrt{7}.$$

Notice that the angle  $\theta$  between the  $(u, v)$  and  $(x, y)$  coordinate systems (that is, the angle between  $\mathbf{E}_1$  and  $\mathbf{F}_1$ ) may be computed from the formula

$$\cos \theta = \langle \mathbf{E}_1, \mathbf{F}_1 \rangle = \frac{\sqrt{2}}{2}.$$

It is now time to reveal the basic theory behind the preceding example. This will require a definition.

**Definition.** Let  $\mathcal{V}$  be a finite-dimensional inner product space. A function  $\varphi: \mathcal{V} \rightarrow \mathbb{R}$  is called *quadratic* iff there exists a self-adjoint linear transformation  $T: \mathcal{V} \rightarrow \mathcal{V}$  such that

$$\varphi(\mathbf{A}) = \langle \mathbf{A}, T(\mathbf{A}) \rangle$$

for all vectors  $\mathbf{A}$  in  $\mathcal{V}$ .

EXAMPLE 9. The function

$$\varphi(x, y) = x^2 + 4xy - y^2$$

is quadratic. For if

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

is the linear transformation with matrix

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

in the standard orthonormal coordinate system then  $T$  is self-adjoint by (16.5). Moreover

$$\langle (x, y), T(x, y) \rangle = (x, y) \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 4xy - y^2 = \varphi(x, y)$$

and so  $\varphi$  is quadratic. More generally

$$\varphi(x, y) = ax^2 + 2bxy + cy^2$$

is quadratic. We let

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

have matrix

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

in the standard orthonormal basis, so that  $T$  is self adjoint and one checks

$$\varphi(x, y) = \langle (x, y), T(x, y) \rangle.$$

Thus  $\varphi$  is quadratic.

Strictly speaking we should call a function  $\varphi: \mathcal{V} \rightarrow \mathbb{R}$  given by

$$\varphi(A) = \langle A, T(A) \rangle$$

a *homogeneous* quadratic function, because it has no linear terms.

**Theorem 16.10** (Principal Axis Theorem). *Suppose that*

$$\varphi: \mathcal{V} \rightarrow \mathbb{R}$$

*is a homogeneous quadratic function on the finite-dimensional inner product space  $\mathcal{V}$ . Then there exists a basis  $\{A_1, \dots, A_n\}$  for  $\mathcal{V}$  and numbers  $\lambda_1, \dots, \lambda_n$  such that for a vector  $X = x_1 A_1 + \dots + x_n A_n$  in  $\mathcal{V}$  the value of  $\varphi$  is given by the formula*

$$\varphi(X) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2.$$

**PROOF.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint transformation such that

$$\varphi(X) = \langle X, T(X) \rangle$$

for all  $X$  in  $\mathcal{V}$ . By the spectral theorem we may find a basis  $A_1, \dots, A_n$  for  $\mathcal{V}$  composed of eigenvectors of  $T$ . Let  $\lambda_1, \dots, \lambda_n$  be the corresponding eigenvalues. If  $X = x_1 A_1 + \dots + x_n A_n$  then by (15.9) we get

$$\begin{aligned} \varphi(X) &= \langle X, T(X) \rangle \\ &= \langle x_1 A_1 + \dots + x_n A_n, T(x_1 A_1 + \dots + x_n A_n) \rangle \\ &= \langle x_1 A_1 + \dots + x_n A_n, x_1 T(A_1) + \dots + x_n T(A_n) \rangle \\ &= \langle x_1 A_1 + \dots + x_n A_n, x_1 \lambda_1 A_1 + \dots + x_n \lambda_n A_n \rangle \\ &= x_1 \lambda_1 x_1 + x_2 \lambda_2 x_2 + \dots + x_n \lambda_n x_n \\ &= \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_n x_n^2, \end{aligned}$$

as required. □

Note that the proof of (16.10) is itself important as it provides a method for finding the basis  $A_1, \dots, A_n$  whose elements are called the *principal axes* of the quadratic form  $\varphi$ .

EXAMPLE 10. Find the principal axes of the quadratic function

$$\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$$

given by

$$\varphi(x, y, z) = 3x^2 + 2y^2 + 3z^2 - 2xy - 2yz.$$

*Solution.* A little experimentation will show that

$$\varphi(x, y, z) = \langle (x, y, z), T(x, y, z) \rangle$$

where  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the self-adjoint linear transformation whose matrix is (do you see the gimmick?)

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

relative to the standard orthonormal basis of  $\mathbb{R}^3$ . The characteristic polynomial of  $T$  is

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} 3-t & -1 & 0 \\ -1 & 2-t & -1 \\ 0 & -1 & 3-t \end{pmatrix} \\ &= (3-t) \det \begin{pmatrix} 2-t & -1 \\ -1 & 3-t \end{pmatrix} + \det \begin{pmatrix} -1 & -1 \\ 0 & 3-t \end{pmatrix} \\ &= (3-t)[(2-t)(3-t) - 1] - (3-t) \\ &= (3-t)[6 - 5t + t^2 - 1 - 1] \\ &= (3-t)(t^2 - 5t + 4) \\ &= -(t-3)(t-4)(t-1), \end{aligned}$$

so the eigenvalues of  $T$  are 1, 3, and 4. Corresponding eigenvectors may be readily found. We get:

$$t = 1 \quad \begin{cases} 2x - y = 0 \\ -x + y - z = 0 \\ -y + 2z = 0 \end{cases}$$

which yields

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

so

$$\begin{aligned} y &= 2x \\ z &= x \end{aligned}$$

which has a nontrivial solution

$$x = 1, y = 2, z = 1$$

and a unit eigenvector is thus  $\mathbf{A}_1 = (1/\sqrt{6}, 2/\sqrt{6}, 1/\sqrt{6})$ .

Likewise we find  $\mathbf{A}_2 = (1/\sqrt{2}, 0, -1/\sqrt{2})$  is a unit eigenvector corresponding to  $t = 3$  and  $\mathbf{A}_3 = (1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})$  is an eigenvector corresponding to  $t = 4$ .

Thus the principal axes of  $\varphi$  are

$$\mathbf{A}_1 = \left( \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6} \right)$$

$$\mathbf{A}_2 = \left( \frac{\sqrt{2}}{2}, 0, -\frac{\sqrt{2}}{2} \right)$$

$$\mathbf{A}_3 = \left( \frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3} \right)$$

and if  $(u, v, w)$  denotes the coordinates of a vector  $\mathbf{X}$  with respect to the principal axes

$$\varphi(\mathbf{X}) = u^2 + 3v^2 + 4w^2.$$

**EXAMPLE 11.** Sketch the graph of the quadratic equation

$$2x^2 + 3xy - 2y^2 = 10.$$

*Solution.* Consider the function

$$\varphi(x, y) = 2x^2 + 3xy - 2y^2$$

Of course  $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$  is a quadratic function and

$$\varphi(x, y) = \langle (x, y), T(x, y) \rangle$$

where  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is the self-adjoint transformation whose matrix in the standard orthonormal basis of  $\mathbb{R}^2$  is

$$\begin{pmatrix} 2 & \frac{3}{2} \\ \frac{3}{2} & -2 \end{pmatrix}.$$

The characteristic polynomial of  $T$  is

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} 2-t & \frac{3}{2} \\ \frac{3}{2} & -2-t \end{pmatrix} = (2-t)(-2-t) - \frac{9}{4} \\ &= t^2 - 4 - \frac{9}{4} \\ &= t^2 - \frac{25}{4}. \end{aligned}$$



## 16: The spectral theorem and quadratic forms

The eigenvalues of  $T$  are therefore  $\pm 5/2$ . Corresponding unit eigenvectors are easily found to be

$$\left(\frac{3\sqrt{10}}{10}, \frac{\sqrt{10}}{10}\right) = \mathbf{F}_+ \quad \text{corresponding to } \frac{5}{2}$$

$$\left(-\frac{\sqrt{10}}{10}, \frac{3\sqrt{10}}{10}\right) = \mathbf{F}_- \quad \text{corresponding to } -\frac{5}{2}.$$

If we write  $(u, v)$  for the coordinates of a vector  $\mathbf{A}$  in  $\mathbb{R}^2$  with respect to the orthonormal basis  $\{\mathbf{F}_+, \mathbf{F}_-\}$  we find

$$\varphi(\mathbf{A}) = \frac{5u^2}{2} - \frac{5v^2}{2}.$$

We seek the graph of the equation

$$\varphi(\mathbf{A}) = 10.$$

But if we employ the  $(u, v)$  coordinates determined by the principal axes of  $T$  then this is just the graph of the equation

$$\frac{5u^2}{2} - \frac{5v^2}{2} = 10$$

or

$$u^2 - v^2 = 4.$$

One recognizes this straight off as a hyperbola (see Figure 16.3). The angle  $\theta$  between the principal axes of  $\theta$  and the usual axes is given by

$$\cos \theta = \langle \mathbf{E}_1, \mathbf{F}_+ \rangle = \frac{3\sqrt{10}}{10}$$

which is about  $18.5^\circ$ .

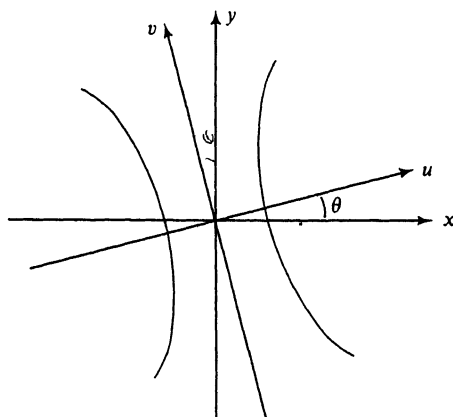


Figure 16.3

The coordinates of the vertices of the hyperbola are

$$(u, v) = (\pm 2, 0)$$

or

$$(x, y) = \left( \frac{3\sqrt{10}}{5}, \frac{\sqrt{10}}{5} \right).$$

Etc.

As a final example we wish to indicate that the application of the principal axis theorem to the study of conic sections is not limited to the homogeneous case.

**EXAMPLE 12.** Sketch the graph of the equation

$$16x^2 - 24xy + 9y^2 - 30x - 40y = 0.$$

*Solution.* Let us first find the principal axes of the homogeneous quadratic function  $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  given by

$$\varphi(x, y) = 16x^2 - 24xy + 9y^2.$$

We begin by noting that

$$\varphi(x, y) = \langle (x, y), T(x, y) \rangle$$

where  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is the self-adjoint transformation with matrix

$$\begin{pmatrix} 16 & -12 \\ -12 & 9 \end{pmatrix}$$

relative to the standard orthonormal basis of  $\mathbb{R}^2$ . The characteristic polynomial of  $T$  is

$$\begin{aligned} \Delta(t) &= \det \begin{pmatrix} 16-t & -12 \\ -12 & 9-t \end{pmatrix} = (9-t)(16-t) - 144 \\ &= t^2 - 25t + 144 - 144 \\ &= t^2 - 25t \end{aligned}$$

The eigenvalues of  $T$  are therefore  $t = 0, t = 25$ . Corresponding eigenvectors are  $F_1 = (\frac{3}{5}, \frac{4}{5})$  and  $F_2 = (-\frac{4}{5}, \frac{3}{5})$ . If we let  $(u, v)$  denote the coordinates of a vector in the  $\{F_1, F_2\}$  basis then

$$\begin{aligned} xE_1 + yE_2 = A &= uF_1 + vF_2 = u\left(\frac{3}{5}E_1 + \frac{4}{5}E_2\right) + v\left(\frac{-4}{5}E_1 + \frac{3}{5}E_2\right) \\ &= \left(\frac{3u}{5} - \frac{4v}{5}\right)E_1 + \left(\frac{4u}{5} + \frac{3v}{5}\right)E_2. \end{aligned}$$

That is, the relation between the  $(x, y)$  and  $(u, v)$  coordinate systems is

$$x = \frac{3u}{5} - \frac{4v}{5}$$

$$y = \frac{4u}{5} + \frac{3v}{5}.$$

To graph the equation

$$16x - 24xy - 9y - 30x - 40y = 0$$

is therefore the same as to graph the equation

$$\varphi(x, y) - 30x - 40y = 0$$

which is the same as

$$\varphi(u, v) - 30\left(\frac{3u}{5} - \frac{4v}{5}\right) - 40\left(\frac{4u}{5} + \frac{3v}{5}\right) = 0$$

which is the same as

$$0u^2 + 25v^2 - 18u + 24v - 32u - 24v = 0$$

or

$$25v^2 = 50u$$

or

$$v^2 = 2u$$

which is instantly seen to be a parabola (see Figure 16.4). The angle  $\theta$  between the  $(u, v)$  and  $(x, y)$  coordinate systems is given by

$$\cos \theta = \langle \mathbf{E}_1, \mathbf{F}_1 \rangle = \frac{3}{5}$$

which is about  $53.1^\circ$ .

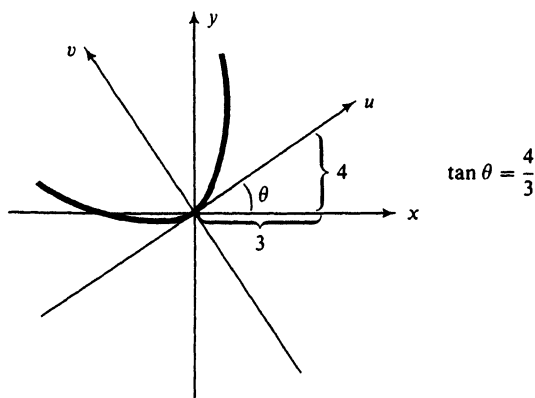


Figure 16.4

## EXERCISES

1. Compute the adjoint of each of the following

(a)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$   $T(x, y, z) = (x + y + z, x + 2y + 2z, x + 2y + 3z)$

(b)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$   $T(x, y, z) = (x + y - z, -x + 2y + 2z, x + 2y + 3z)$

(c)  $T: \mathcal{P}_2(\mathbb{R}) \rightarrow \mathcal{P}_2(\mathbb{R})$   $T(p(x)) = x \frac{d}{dx} p(x) - \frac{d}{dx} (xp(x)).$

(d)  $T: \mathbb{R}^4 \rightarrow \mathbb{R}^4$   $T(x, y, z, w) = (x + y, y + z, z + w, w + x).$

(e)  $T: \mathcal{M}_{33} \rightarrow \mathcal{M}_{33}$   $T(A) = A' + A$

(f)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$   $T(x, y, z) = (-y + z, -x + 2z, x + 2y)$

2. Let  $S, T: \mathcal{V} \rightarrow \mathcal{V}$ . Prove:

(a)  $(S + T)^* = S^* + T^*.$

(b)  $(aS)^* = aS^*.$

(c)  $(ST)^* = T^*S^*.$

(d)  $T + T^*$  is self-adjoint.

3. Diagonalize each of the following self-adjoint transformations

(a)  $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$   $T(x, y) = (2x + y, 2y + x)$

(b)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$   $T(x, y, z) = (2x + y + z, x + 2y + z, x + y + 2z)$

(c)  $T: \mathbb{R}^5 \rightarrow \mathbb{R}^5$   $T(x, y, z, u, v) = (2x + y, 2y + x, 2z + u + v, z + 2u + v, z + u + 2v)$

(d)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$   $T(x, y, z) = (y + z, x + z, x + y)$

(e)  $T: \mathcal{M}_{33} \rightarrow \mathcal{M}_{33}$   $T(A) = A'$

(f)  $T: \mathcal{M}_{33} \rightarrow \mathcal{M}_{33}$   $T(A) = A - A'$

4. Sketch the Graph of each of the following

(a)  $8x^2 - 12xy + 17y^2 = 80$

(b)  $3x^2 + 2xy + 3y^2 = 4$

(c)  $5x^2 - 8xy + 5y^2 = 9$

(d)  $11x^2 - 24xy + 4y^2 + 6x + 8y = -15$

(e)  $16x^2 - 24xy + 9y^2 - 30x + 40y = 5$

5. Suppose that  $\varphi: \mathcal{V} \rightarrow \mathbb{R}$  is a quadratic form. Show that for any two vectors  $\mathbf{A}, \mathbf{B}$  in  $\mathcal{V}$  that

$$\langle \mathbf{A}, \mathbf{B} \rangle = \frac{1}{2}(\varphi(\mathbf{A} + \mathbf{B}) - \varphi(\mathbf{A}) - \varphi(\mathbf{B}))$$

that is, the inner product in  $\mathcal{V}$  may be computed from *any* quadratic form.6. Suppose that  $\varphi: \mathcal{V} \rightarrow \mathbb{R}$  is a quadratic form and

$$\varphi(\mathbf{A}) = \langle \mathbf{A}, \mathbf{T}(\mathbf{A}) \rangle$$

$$\varphi(\mathbf{A}) = \langle \mathbf{A}, \mathbf{S}(\mathbf{A}) \rangle$$

## 16: The spectral theorem and quadratic forms

for two self-adjoint transformations  $S, T: \mathcal{V} \rightarrow \mathcal{V}$ . Show that  $S = T$  (*Hint*: use (4) to show that  $\langle B, T(A) \rangle = \langle B, S(A) \rangle$  for all  $B$  and  $A$  in  $\mathcal{V}$ . Then recall that  $\langle S(B), A \rangle = \langle B, S(A) \rangle$  for all  $A, B$  in  $\mathcal{V}$ . Use this to show that  $T = S^*$ . Appeal to self-adjointness.)

7. Let  $*$ :  $\mathcal{L}(\mathcal{V}) \rightarrow \mathcal{L}(\mathcal{V})$ . Show that  $*$  is a linear transformation and is self-adjoint.

The exercises that follow are intended to lead to a proof of the spectral theorem for arbitrary dimensions  $n$ . There are two basic steps in all the proofs that I know of. The first of these is to show that the transformation always has at least one eigenvector. Two different ways to obtain this are offered in Exercises 8 and 9. The second step is to show that the subspace orthogonal to an eigenvector is invariant. The proof can then be completed by induction. This is the purpose of Exercise 10.

8. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation on the finite-dimensional inner product space  $\mathcal{V}$ . Let  $S(\mathcal{V}) = \{x \in \mathcal{V} \mid \|x\| = 1\}$  be the unit sphere of  $\mathcal{V}$ . We assume from the multivariable calculus the fundamental theorem that a real-valued function on  $S(\mathcal{V})$  always has a minimum. Define

$$f: S(\mathcal{V}) \rightarrow \mathbb{R}$$

by the formula

$$f(x) = \langle T(x), x \rangle$$

and let  $v \in S(\mathcal{V})$  be a vector at which  $f$  assumes its minimum value. Let  $y \in S(\mathcal{V})$  be a fixed vector orthogonal to  $v$ .

- (a) Show that the vector  $(v + ty)/\sqrt{1 + t^2}$  has unit length for all real numbers  $t$ . Define the function  $g: \mathbb{R} \rightarrow \mathbb{R}$  by

$$g(t) = f\left(\frac{v + ty}{\sqrt{1 + t^2}}\right)$$

- (b) Show that  $g(0) = f(v)$  is the minimum value of  $g(t)$ .  
(c) Show that

$$\frac{dg}{dt} = \frac{2t}{(1 + t^2)^2} \langle T(v + ty), v + ty \rangle.$$

- (*Hint*: Write  $g(t) = \langle T((v + ty)/\sqrt{1 + t^2}), (v + ty)/\sqrt{1 + t^2} \rangle$  and expand the inner product remembering  $v$  and  $y$  are fixed, have unit length, and are perpendicular.)  
(d) Evaluate

$$\left. \frac{dg}{dt} \right|_{t=0}.$$

(Answer:  $2\langle T(y), v \rangle$ ).

- (e) Show that  $T(v)$  is perpendicular to  $y$ . (*Hint*: Use (d), self-adjointness and the fact that  $t = 0$  is a critical point of  $g$ .)

- (f) Show  $T(v) = \lambda v$  for some number  $\lambda$ , that is,  $v$  is an eigenvector of  $T$ . (*Hint*: Use (e) to show  $T(v)$  is perpendicular to the orthogonal complement of  $\mathcal{L}(v)$ , and so  $T(v)$  belongs to  $\mathcal{L}(v)$ .)

9. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation on the finite-dimensional inner product space  $\mathcal{V}$ . Let  $\{e_1, \dots, e_n\}$  be an orthonormal basis for  $\mathcal{V}$  and  $A$  the matrix of  $T$  relative to  $\{e_1, \dots, e_n\}$ . Introduce the linear transformation  $T_C: \mathbb{C}^n \rightarrow \mathbb{C}^n$  whose matrix with respect to the standard basis is  $A$ .

(a) Show the characteristic polynomial of  $T_C$  is equal to the characteristic polynomial of  $T$ .

The fundamental theorem of algebra tells us that  $\Delta_{T_C}(x)$  must have a root, which may, however, be complex. We want to show that it is real so that  $\Delta_T(x)$  has a real root, and hence  $T$  an eigenvalue. To do this let  $\lambda$  be a root of  $\Delta_{T_C}(x)$ ,  $w \in \mathbb{C}^n$  a corresponding eigenvector, and  $\bar{w}$  the vector whose coordinates are the complex conjugates of those of  $w$ .

(b) Show that  $T_C(\bar{w}) = \bar{\lambda}\bar{w}$  where  $\bar{\lambda}$  is the complex conjugate of  $\lambda$ . (Hint: Compute the complex conjugate of the matrix product  $Aw$  and use the fact that the entries of  $A$  are real.)

(c) Show that the number  $\bar{w} \cdot w^t$  is nonzero ( $\cdot$  denotes the matrix product of the row vector  $\bar{w}$  with the column vector  $w^t$ ).

(d) Show  $\bar{w}Aw^t = \lambda(\bar{w} \cdot w^t)$ ,  $wAw^t = \bar{\lambda}(w \cdot \bar{w}^t)$ .

(e) Show that  $\lambda$  is real. (Hint: Use (d) and take transposes to show  $\bar{w}Aw^t = wAw^t$  because  $A$  is symmetric.)

(f) Show that the characteristic polynomial of  $T$  has a real root.

10. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a self-adjoint linear transformation on the finite-dimensional inner product space  $\mathcal{V}$ . Using either (8) or (9) we know that  $T$  has at least one eigenvector. Let  $v$  be such an eigenvector with corresponding eigenvalue  $\lambda$ .

(a) If  $y \in \mathcal{V}$  is orthogonal to  $v$  then  $T(y)$  is also orthogonal to  $v$ . (Hint: Use  $\langle T(y), v \rangle = \langle y, T(v) \rangle$ .)

(b) Let  $\mathcal{W}$  be the subspace of  $\mathcal{V}$  orthogonal to  $\mathcal{L}(v)$ . Show  $T(\mathcal{W}) \subset \mathcal{W}$ .

(c) Let  $T|_{\mathcal{W}}: \mathcal{W} \rightarrow \mathcal{W}$  be the restriction of  $T$  to  $\mathcal{W}$ . Show  $T|_{\mathcal{W}}$  is self-adjoint.

(d) Show that there is an orthonormal basis  $\{v, w_1, \dots, w_m\}$  for  $\mathcal{V}$  with respect to which the matrix of  $T$  is

$$\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}$$

where  $\mathbf{B}$  is symmetric.

(e) Use Exercise (8f) or (9f), (10d) and induction to prove the Spectral theorem (Theorem 16.7).

11. (a) Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be the linear transformation whose matrix with respect to the standard orthonormal basis of  $\mathbb{R}^3$  is

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

Diagonalize  $T$ .

16: The spectral theorem and quadratic forms

(b) Same problem for each of the following matrices

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}.$$

12. Let  $T: \mathbb{C}^2 \rightarrow \mathbb{C}^2$  be the linear transformation given by

$$T(a, b) = (2ia + b, a),$$

where  $a, b$  are complex numbers.

- Find the matrix of  $T$  with respect to the standard basis of  $\mathbb{C}^2$ .
  - Show the matrix of  $T$  is symmetric.
  - Find the eigenvalues of  $T$ .
  - Show that  $T$  is *not* diagonalizable.
  - What do you conclude about the validity of the following statement? Symmetric matrices are diagonalizable. (*Hint*: Show this is true if the entries of the matrix are real, and can be false if the entries are allowed to be complex.)
13. Suppose  $A \in \mathcal{M}_n$  and that the columns of  $A$  regarded as vectors in  $\mathbb{R}^n$  are orthogonal. Show that regarded as vectors in  $\mathbb{R}^n$  the rows are also orthogonal. (*Hint*: Think about the relation between matrices and linear transformations.)

This chapter is of a distinctly more difficult nature than the preceding ones. In it we will treat the problem of finding a particular canonical, simple, matrix, representative of a linear transformation. In the preceding chapter we treated this problem for *self-adjoint* linear transformations in a finite-dimensional inner product space. For a self-adjoint linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{V}$$

in the finite-dimensional inner product space  $\mathcal{V}$  we saw that we could always find an orthonormal basis for  $\mathcal{V}$  so that the corresponding matrix of  $T$  was diagonal. The required basis would be composed of the normed eigenvectors of  $T$  and the diagonal entries of the corresponding matrix are the eigenvalues of  $T$ . The problem we now consider is: Can we achieve a similar *normal form* in general? An answer to this question leads to the Jordan normal form. Before we arrive at this goal we will require quite a few preparatory steps.

To begin with, let us consider where we might encounter difficulties in case we were to proceed by analogy with the case of self-adjoint linear transformations. What we achieved in the self-adjoint case is a decomposition

$$\mathcal{V} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_q$$

of the vector space  $\mathcal{V}$  into a direct sum of eigenspaces  $\mathcal{U}_i$ ,  $1 \leq i \leq q$ , of  $T$ . For a general linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{V},$$

there are two problems that arise to prevent an analogous result.

- (1)  $T$  may not have any eigenvalues at all.
- (2)  $T$  has eigenvalues, but the sum of the eigenspaces  $\mathcal{U}_1 + \cdots + \mathcal{U}_q$  is not all of  $\mathcal{V}$ .



The first problem is not so serious. We can avoid it by working with vector spaces over the *complex numbers*. (The advantage so gained has already been indicated in the exercises to Chapter 16.) On the other hand the second problem is more serious. It reflects more of a defect in the linear transformation  $T$  rather than in that of the scalars. Here are some examples to illustrate this point.

*Note.* A direct sum decomposition of a vector space with respect to subspaces will be defined, explained, and illustrated in what follows.

EXAMPLE 1. We will consider three linear transformations from  $\mathbb{R}^3$  to itself, viz.

$$T_i: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad i = 1, 2, 3;$$

each one of which has the real number  $e$  as eigenvalue with multiplicity 3, and furthermore, the corresponding eigenspace  $\mathcal{V}_e$  will have dimensions 1, 2, and 3, respectively.

(1)  $T_1: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear transformation represented by the matrix

$$A_1 = \begin{pmatrix} e & 0 & 0 \\ 0 & e & 0 \\ 0 & 0 & e \end{pmatrix},$$

with respect to the standard basis of  $\mathbb{R}^3$ . Then  $e$  is an eigenvalue of  $T_1$  with the geometric multiplicity 3 and the standard basis vectors are eigenvectors of  $T_1$ .

(2)  $T_2: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear transformation represented by the matrix

$$A_2 = \begin{pmatrix} e & 1 & 0 \\ 0 & e & 0 \\ 0 & 0 & e \end{pmatrix},$$

with respect to the standard basis of  $\mathbb{R}^3$ . The characteristic polynomial of  $T_2$  is

$$p(t) = (t - e)^3,$$

so  $e$  is an eigenvalue from algebraic multiplicity 3.

But there are only two linearly independent eigenvectors belonging to this eigenvalue. To see this remember the eigenvectors  $\mathbf{E}$  are the nonzero solutions of

$$(*) \quad (T_2 - eI)(\mathbf{E}) = 0.$$

The vectors  $\mathbf{E} = (x, y, z) \in \mathbb{R}^3$  that satisfy  $(*)$  are the solutions of the linear system

$$y = 0.$$

So  $(1, 0, 0), (0, 0, 1)$  is a basis for  $\mathcal{V}_e$  and  $e$  is of geometric multiplicity 2.

(3)  $T_3: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear transformation represented by the matrix

$$A_3 = \begin{pmatrix} e & 1 & 0 \\ 0 & e & 1 \\ 0 & 0 & e \end{pmatrix},$$

with respect to the standard basis of  $\mathbb{R}^3$ . The characteristic polynomial is again

$$p(t) = (t - e)^3,$$

so  $e$  is again an eigenvalue with algebraic multiplicity 3. To find the geometric dimension we must find a basis for

$$\text{Ker}(T_3 - eT).$$

This leads to the system of linear equations

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

namely

$$y = 0,$$

$$z = 0,$$

so that  $\mathcal{V}_e = \mathcal{L}((1, 0, 0))$  and  $e$  has only geometric dimension 1.

From this example we learn that there are linear transformations that simply do not have enough eigenvectors to be representable by a diagonal matrix. In order to represent such a linear transformation by as simple a matrix as possible, we must attempt to cleverly enlarge the eigenspaces whose dimension is smaller than the algebraic multiplicity of the corresponding eigenvalue. The enlargement should be made in such a way that the enlarged space admits a basis with respect to which the matrix of  $T$  is “practically” diagonal.

To achieve this program we are going to need a number of new ideas.

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation. A vector subspace  $\mathcal{U}$  in  $\mathcal{V}$  is called *T-invariant* iff  $T(\mathcal{U}) \subseteq \mathcal{U}$ .

**EXAMPLE 2.** If  $T: \mathcal{V} \rightarrow \mathcal{V}$  has an eigenvalue  $e$  then the eigenspace  $\mathcal{V}_e$  is  $T$ -invariant. To see this note that  $\mathbf{E} \in \mathcal{V}_e$ , then  $T(\mathbf{E}) = e\mathbf{E}$  is a multiple of  $\mathbf{E}$  and so belongs to  $\mathcal{V}_e$  since  $\mathcal{V}_e$  is a subspace.

**EXAMPLE 3.** Let  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by the matrix (Example 1(2))

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

Then  $\mathcal{U} = \mathcal{L}((1, 0, 0), (0, 1, 0))$  is an invariant subspace. To see this note that

$$T(1, 0, 0) = (e, 0, 0) = e(1, 0, 0) + 0(0, 1, 0),$$

$$T(0, 1, 0) = (1, e, 0) = 1(1, 0, 0) + e(0, 1, 0),$$

so  $T(1, 0, 0), T(0, 1, 0) \in \mathcal{U}$ . By linearity of  $T$  this implies  $T(\mathcal{U}) \subseteq \mathcal{U}$ , as  $\mathcal{U}$  is invariant.

When there is an invariant subspace we can find a simple matrix representation.

**Proposition 17.1.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the  $n$ -dimensional vector space  $\mathcal{V}$ . Suppose  $\mathcal{U}$  is an  $m$ -dimensional  $T$ -invariant subspace of  $\mathcal{V}$ . Then with respect to a suitable basis for  $\mathcal{V}$ ,  $T$  can be represented by a matrix in block form*

$$\left( \begin{array}{c|c} \mathbf{A} & \mathbf{C} \\ \hline 0 & \mathbf{B} \end{array} \right) \begin{matrix} \} m \\ \} n - m \end{matrix}$$

$\underbrace{\hspace{1.5cm}}_m \quad \quad \underbrace{\hspace{1.5cm}}_{n-m}$

PROOF. Choose a basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_m\}$  for  $\mathcal{U}$  and extend this to a basis

$$\{\mathbf{A}_1, \dots, \mathbf{A}_m, \mathbf{B}_1, \dots, \mathbf{B}_{n-m}\}$$

for  $\mathcal{V}$ . Then because  $T\mathbf{A}_i \in \mathcal{U} = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_m)$  we have

$$T(\mathbf{A}_i) = \sum_{j=1}^m a_{ji} \mathbf{A}_j,$$

so the first  $m$  columns of the matrix of  $T$  with respect to this basis have no entries in rows  $m+1$  through  $n$ . This says the matrix is of the above block form.  $\square$

**Definition.** Let  $\mathcal{V}$  be a vector space and let  $\mathcal{S}, \mathcal{T}$  be subspaces of  $\mathcal{V}$ . We say  $\mathcal{V}$  is the *direct sum* of  $\mathcal{S}$  and  $\mathcal{T}$  iff

$$\mathcal{V} = \mathcal{S} + \mathcal{T} \quad \text{and} \quad \mathcal{S} \cap \mathcal{T} = \{0\}.$$

In this case we write  $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$ .

**Proposition 17.2.** *Let  $\mathcal{V}$  be a vector space and let  $\mathcal{S}, \mathcal{T}$  be subspaces of  $\mathcal{V}$ . Then  $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$  iff every vector in  $\mathcal{V}$  can be written uniquely in the form*

$$\mathbf{A} = \mathbf{S} + \mathbf{T}, \quad \mathbf{S} \in \mathcal{S}, \quad \mathbf{T} \in \mathcal{T}.$$

PROOF. Suppose  $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$ . Let  $\mathbf{A} \in \mathcal{V}$ . Since  $\mathcal{V} = \mathcal{S} + \mathcal{T}$ . There exist vectors  $\mathbf{S} \in \mathcal{S}, \mathbf{T} \in \mathcal{T}$  such that

$$(*) \quad \mathbf{A} = \mathbf{S} + \mathbf{T}.$$

If there were vectors  $\bar{\mathbf{S}} \in \mathcal{S}$ ,  $\bar{\mathbf{T}} \in \mathcal{T}$  so that

$$\mathbf{A} = \bar{\mathbf{S}} + \bar{\mathbf{T}}$$

then

$$\mathbf{S} + \mathbf{T} = \mathbf{A} = \bar{\mathbf{S}} + \bar{\mathbf{T}}$$

implies

$$\mathbf{S} - \bar{\mathbf{S}} = \bar{\mathbf{T}} - \mathbf{T}.$$

So  $\mathbf{S} - \bar{\mathbf{S}} \in \mathcal{T}$ . But  $\mathbf{S} - \bar{\mathbf{S}} \in \mathcal{S}$  since  $\mathcal{S}$  is a subspace. Therefore

$$\mathbf{S} - \bar{\mathbf{S}} \in \mathcal{S} \cap \mathcal{T} = \{0\},$$

so  $\mathbf{S} - \bar{\mathbf{S}} = 0$  or  $\mathbf{S} = \bar{\mathbf{S}}$ . A similar argument shows  $\mathbf{T} - \bar{\mathbf{T}} = 0$  or  $\mathbf{T} = \bar{\mathbf{T}}$ , so the representation of  $\mathbf{A}$  in (\*) is unique.

The converse is immediate.  $\square$

**Proposition 17.3.** Let  $\mathcal{V}$  be a finite-dimensional vector space and let  $\mathcal{S}, \mathcal{T}$  be subspaces of  $\mathcal{V}$ . If  $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$  then

$$\dim \mathcal{V} = \dim \mathcal{S} + \dim \mathcal{T}.$$

Moreover if  $\{\mathbf{A}_1, \dots, \mathbf{A}_s\}$  is a basis for  $\mathcal{S}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_t\}$  a basis for  $\mathcal{T}$ , then  $\{\mathbf{A}_1, \dots, \mathbf{A}_s, \mathbf{B}_1, \dots, \mathbf{B}_t\}$  is a basis for  $\mathcal{V}$ .

PROOF. The proof is similar to (8.10).  $\square$

**Proposition 17.4.** Let  $\mathcal{V}$  be a finite-dimensional vector space, and let  $\mathcal{S}, \mathcal{T}$  be subspaces of  $\mathcal{V}$  such that  $\mathcal{V} = \mathcal{S} + \mathcal{T}$ . Then  $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$  iff

$$\dim \mathcal{V} = \dim \mathcal{S} + \dim \mathcal{T}$$

PROOF. Do Exercises 8 and 9 of Chapter 6.  $\square$

**Corollary 17.5.** Let  $\mathcal{V} \langle \cdot, \cdot \rangle$  be an inner product space and let  $\mathcal{W}$  be a subspace of  $\mathcal{V}$ . Then

$$\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp.$$

PROOF. Follow from (15.16).  $\square$

EXAMPLE 4. In  $\mathbb{R}^3$  consider the subspaces

$$\mathcal{S} = \mathcal{L}((1, -1, 0), (0, 1, -1)),$$

$$\mathcal{T} = \mathcal{L}((1, 1, 1)).$$

Then

$$\mathbb{R}^3 = \mathcal{S} \oplus \mathcal{T}.$$

If  $\mathcal{R} = \mathcal{L}((1, 1, 1), (1, -2, 1))$ . Then

$$\mathbb{R}^3 = \mathcal{R} + \mathcal{S}.$$

But  $\mathbb{R}^3 \neq \mathcal{R} \oplus \mathcal{S}$  because

$$\mathcal{R} \cap \mathcal{S} \neq \{0\}.$$

In fact

$$1 \cdot (1, -1, 0) + (-1)(0, 1, -1) = (1, -2, 1)$$

shows

$$(1, -2, 1) \in \mathcal{R} \cap \mathcal{S}.$$

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation. A  $T$ -invariant subspace  $\mathcal{U}$  in  $\mathcal{V}$  is called *completely invariant* or *fully invariant* iff there exists a  $T$ -invariant subspace in  $\mathcal{W}$  such that  $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$ . The subspace  $\mathcal{W}$  is called a  $T$ -invariant complement for  $\mathcal{U}$ .

**Corollary 17.6.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the  $n$ -dimensional vector space  $\mathcal{V}$ . Suppose  $\mathcal{U}$  is a  $T$ -completely invariant subspace of dimension  $m$  in  $\mathcal{V}$ . Then with respect to a suitable basis for  $\mathcal{V}$ ,  $T$  can be represented in block matrix form

$$\begin{array}{c} m\{ \begin{array}{c|c} \mathbf{A} & 0 \\ \hline 0 & \mathbf{B} \end{array} \} \\ n\{ \underbrace{\hspace{1.5cm}}_m \quad \underbrace{\hspace{1.5cm}}_n \end{array}$$

PROOF. Let  $\mathcal{W}$  be a  $T$ -invariant complement for  $\mathcal{U}$ . Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_m\}$  be a basis for  $\mathcal{U}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_n\}$  a basis for  $\mathcal{W}$ . Then  $\{\mathbf{A}_1, \dots, \mathbf{A}_m, \mathbf{B}_1, \dots, \mathbf{B}_n\}$  is a basis for  $\mathcal{V}$ . The matrix of  $T$  with respect to this basis is of the required form.  $\square$

This corollary shows that one way to simplify a matrix representative for  $T$  is to find a complementary pair of  $T$ -invariant subspaces. But, how does one find such pairs? The following results explain how.

**Proposition 17.7.** Suppose given linear transformations

$$S, T: \mathcal{V} \rightarrow \mathcal{V}$$

such that

$$S \circ T = T \circ S$$

(that is,  $S$  and  $T$  commute with each other). Then  $\text{Ker } S$  and  $\text{Im } S$  are  $T$ -invariant subspaces.

PROOF. If  $\mathbf{A} \in \text{Ker } \mathbf{S}$ , then

$$\mathbf{S} \circ \mathbf{T}(\mathbf{A}) = \mathbf{T} \circ \mathbf{S}(\mathbf{A}) = \mathbf{T}(\mathbf{0}) = \mathbf{0}$$

shows that  $\mathbf{T}\mathbf{A} \in \text{Ker } \mathbf{S}$ , so  $\text{Ker } \mathbf{S}$  is a  $\mathbf{T}$ -invariant subspace. Similarly, if  $\mathbf{B} \in \text{Im } \mathbf{S}$ , then  $\mathbf{B} = \mathbf{S}(\mathbf{C})$  for some  $\mathbf{C} \in \mathcal{V}$  and, therefore

$$\mathbf{T}(\mathbf{B}) = \mathbf{T} \circ \mathbf{S}(\mathbf{C}) = \mathbf{S} \circ (\mathbf{T}(\mathbf{C}))$$

shows that  $\mathbf{T}(\mathbf{B}) \in \text{Im } \mathbf{S}$ , so  $\text{Im } \mathbf{S}$  is a  $\mathbf{T}$ -invariant subspace.  $\square$

**Corollary 17.8.** *Let  $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation and*

$$\mathbf{S} = a_0 \mathbf{I} + a_1 \mathbf{T} + \cdots + a_m \mathbf{T}^m: \mathcal{V} \rightarrow \mathcal{V},$$

*where  $a_0, \dots, a_m \in \mathbb{C}$ . Then  $\text{Ker } \mathbf{S}$  and  $\text{Im } \mathbf{S}$  are  $\mathbf{T}$ -invariant.*

PROOF. Clearly  $\mathbf{S}$  and  $\mathbf{T}$  commute, so we may apply (17.7).

The next step in our program, the key that unlocks the door to the Jordan canonical form, is the generalization of (12<sup>bis</sup>.3) on the structure of nilpotent transformations with a cyclic vector to completely general nilpotent transformations.

**Theorem 17.9.** *Let  $\mathcal{V}$  be a finite-dimensional vector space and suppose that  $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{V}$  is a nilpotent linear transformation. Then there exists a direct sum decomposition*

$$\mathcal{V} = \mathcal{U}_1 \oplus \cdots \oplus \mathcal{U}_t$$

*such that  $\mathcal{U}_i$  is  $\mathbf{T}$ -invariant and the induced map  $\mathbf{T}: \mathcal{U}_i \rightarrow \mathcal{U}_i$  has a cyclic vector for  $i = 1, \dots, t$ .*

PROOF. Let  $p = \text{Index}(\mathbf{T})$ , that is,  $p$  is the smallest integer such that  $\mathbf{T}^p = \mathbf{0}$ . Define

$$H_i = \text{Ker } \mathbf{T}^i, \quad i = 0, 1, \dots, p.$$

Then the  $H_i$  are vector subspaces of  $\mathcal{V}$ . Notice that  $H_0 = \{0\}$  and  $H_p = \mathcal{V}$ . Moreover, if  $0 \leq i \leq p$  then we have

$$\mathbf{A} \in H_{i+1} \Rightarrow \mathbf{T}^i \mathbf{A} = \mathbf{0} \Rightarrow \mathbf{T}^{i+1} = \mathbf{T}(\mathbf{T}^i(\mathbf{A})) = \mathbf{T}(\mathbf{0}) = \mathbf{0},$$

so  $\mathbf{A} \in H_{i+1}$ . Thus we have a chain of subspaces

$$\{0\} = H_0 \subset H_1 \subset \cdots \subset H_p = \mathcal{V},$$

and  $\mathbf{T}(H_i) \subset H_{i-1}$ .

*Remark.* In the case that  $\mathbf{T}$  has a cyclic vector  $\mathbf{A}$  then

$$H_i = \mathcal{L}(\mathbf{T}^{p-i}(\mathbf{A}), \dots, \mathbf{T}^{p-1}(\mathbf{A}))$$

for  $i = 1, \dots, p$  and  $H_0 = \{0\}$ .

We proceed now to construct the invariant subspaces  $\mathcal{U}_1, \dots, \mathcal{U}_t$  of  $\mathcal{V}$ . To do this we need a new idea.

**Definition.** Let  $\mathcal{V}$  be a vector space and  $\mathcal{W}$  a subspace. A set of vectors  $\mathbf{E}$  is said to be *linearly dependent over  $\mathcal{W}$*  or *modulo  $\mathcal{W}$*  iff there are distinct vectors  $\mathbf{A}_1, \dots, \mathbf{A}_k$  in  $\mathbf{E}$  and numbers  $a_1, \dots, a_k$ , not all zero such that

$$(*) \quad a_1 \mathbf{A}_1 + \dots + a_k \mathbf{A}_k \in \mathcal{W}.$$

A set of vectors  $\mathbf{E}$  that is not linearly dependent over  $\mathcal{W}$  is said to be *linearly independent over  $\mathcal{W}$* .

**Proposition A.** Let  $\mathcal{V}$  be a vector space and  $\mathcal{W}$  a finite-dimensional subspace of  $\mathcal{V}$ . A set of vectors  $\mathbf{E}$  is linearly independent over  $\mathcal{W}$  iff for every basis  $\{\mathbf{B}_1, \dots, \mathbf{B}_s\}$  of  $\mathcal{W}$ , the set of vectors  $\mathbf{E} \cup \{\mathbf{B}_1, \dots, \mathbf{B}_s\}$  is linearly independent in  $\mathcal{V}$ .

PROOF. Suppose that  $\mathbf{E}$  is linearly independent over  $\mathcal{W}$  and  $\{\mathbf{B}_1, \dots, \mathbf{B}_s\}$  is a basis for  $\mathcal{W}$ . Let  $\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  be vectors of  $\mathbf{E}$ ,  $a_1, \dots, a_k$ , and  $b_1, \dots, b_s$  numbers, such that

$$a_1 \mathbf{A}_1 + \dots + a_k \mathbf{A}_k + b_1 \mathbf{B}_1 + \dots + b_s \mathbf{B}_s = 0.$$

Then, of course

$$a_1 \mathbf{A}_1 + \dots + a_k \mathbf{A}_k = -(b_1 \mathbf{B}_1 + \dots + b_s \mathbf{B}_s) \in \mathcal{W},$$

so  $a_1 = \dots = a_k = 0$ . But then

$$b_1 \mathbf{B}_1 + \dots + b_s \mathbf{B}_s = 0,$$

so  $b_1 = \dots = b_s = 0$  also. The converse is just as easy.  $\square$

**Definition.** Let  $\mathcal{V}$  be a vector space and  $\mathcal{W}$  a subspace of  $\mathcal{V}$ . A set of vectors  $\mathbf{E}$  is said to *span  $\mathcal{V}$  over  $\mathcal{W}$*  or *modulo  $\mathcal{W}$*  iff  $\mathcal{V} = \mathcal{L}(\mathbf{E}) + \mathcal{W}$ . If, in addition,  $\mathbf{E}$  is linearly independent over  $\mathcal{W}$ , then we say that  $\mathbf{E}$  is a *basis for  $\mathcal{V}$  over  $\mathcal{W}$* .

**Proposition B.** If  $\mathcal{V}$  is a finite-dimensional vector space and  $\mathcal{W}$  is a subspace then there exists a basis for  $\mathcal{V}$  over  $\mathcal{W}$ . The number of vectors in a basis for  $\mathcal{V}$  over  $\mathcal{W}$  is  $\dim \mathcal{V} - \dim \mathcal{W}$ .

PROOF. To prove the existence of a basis for  $\mathcal{V}$  over  $\mathcal{W}$  is easy. Choose a basis  $\{\mathbf{B}_1, \dots, \mathbf{B}_s\}$  for  $\mathcal{W}$ . Extend this to a basis  $\{\mathbf{A}_1, \dots, \mathbf{A}_k, \mathbf{B}_1, \dots, \mathbf{B}_s\}$  for  $\mathcal{V}$ . Then by Proposition A the set  $\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  is linearly independent over  $\mathcal{W}$ . Since

$$\begin{aligned} \mathcal{V} &= \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k, \mathbf{B}_1, \dots, \mathbf{B}_s) = \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k) + \mathcal{L}(\mathbf{B}_1, \dots, \mathbf{B}_s) \\ &= \mathcal{L}(\mathbf{A}_1, \dots, \mathbf{A}_k) + \mathcal{W} \end{aligned}$$

it follows that  $\{\mathbf{A}_1, \dots, \mathbf{A}_k\}$  is a basis for  $\mathcal{V}$  over  $\mathcal{W}$ . Since  $s + k = \dim \mathcal{V}$  it follows that  $k = \dim \mathcal{V} - \dim \mathcal{W}$ .  $\square$





so  $T$  is nilpotent of index 3. The matrix of  $T$  with respect to the standard basis of  $\mathbb{R}^5$  is

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

which despite the large number of zero entries is not very informative. The procedure of (17.9) provides a more suitable basis for  $\mathbb{R}^5$ . We first need the chain of subspaces

$$\{0\} = H_0 \subset H_1 \subset H_2 \subset H_3 = \mathbb{R}^5.$$

A basis for  $H_2 = \text{Ker } T^2$  is given by

$$\mathbf{B}_1 = (1, 0, 0, 0, 0),$$

$$\mathbf{B}_2 = (0, 1, 0, 0, 0),$$

$$\mathbf{B}_3 = (0, 0, 0, 1, 0),$$

$$\mathbf{B}_4 = (0, 0, 0, 0, 1).$$

This can be extended to a basis for  $\mathbb{R}^5$  by adjoining

$$\mathbf{A}_1 = (0, 0, 1, 0, 0).$$

Then  $T\mathbf{A}_1 \in H_2$ , which together with the basis for  $H_1$

$$\mathbf{C}_1 = (0, 1, 0, 0, 0),$$

$$\mathbf{C}_2 = (0, 0, 0, 0, 1),$$

fails by one vector to be a basis for  $H_2$ . So we add the vector

$$\mathbf{A}_2 = (1, 0, 0, 0, 0)$$

to obtain the basis

$$\mathbf{A}_1, T\mathbf{A}_1, \mathbf{A}_2, \mathbf{C}_1, \mathbf{C}_2$$

for  $H_2$ . The vectors

$$T^2\mathbf{A}_1, T\mathbf{A}_2 \in H_1,$$

and are a basis for  $H_1$ . Here the process stops with the basis scheme

$$\begin{aligned} &\mathbf{A}_1, \\ &T\mathbf{A}_1, \quad \mathbf{A}_2, \\ &T^2\mathbf{A}_1, \quad T\mathbf{A}_2, \end{aligned}$$

for  $\mathbb{R}^5$ . Set

$$\mathcal{U}_1 = \mathcal{L}(\mathbf{A}_1, \mathbf{T}\mathbf{A}_1, \mathbf{T}^2\mathbf{A}_1),$$

$$\mathcal{U}_2 = \mathcal{L}(\mathbf{A}_2, \mathbf{T}\mathbf{A}_2).$$

Then  $\mathcal{U}_1, \mathcal{U}_2$  are complementary  $\mathbf{T}$ -invariant subspaces of  $\mathbb{R}^5$ , so

$$\mathbb{R}^5 = \mathcal{U}_1 \oplus \mathcal{U}_2.$$

$\mathbf{A}_1 \in \mathcal{U}_1$  is a cyclic vector of index 3 and  $\mathbf{A}_2 \in \mathcal{U}_2$  is a cyclic vector of index 2. The matrix of  $\mathbf{T}$  with respect to the basis

$$\mathbf{A}_1, \mathbf{T}\mathbf{A}_1, \mathbf{T}^2\mathbf{A}_1, \mathbf{A}_2, \mathbf{T}^2\mathbf{A}_2$$

for  $\mathbb{R}^5$  is

$$\left( \begin{array}{ccc|cc} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array} \right)$$

which is the block sum of two shift operators; one of index 3 and one of index 2.

The matrix interpretation of (17.9) is contained in the following corollary.

**Corollary 17.10.** *Let  $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{V}$  be a nilpotent linear transformation in the finite-dimensional vector space  $\mathcal{V}$ . Let  $p$  be the index of  $\mathbf{T}$ . Then with respect to an appropriate basis for  $\mathcal{V}$  the matrix of  $\mathbf{T}$  is of the form*

$$\left( \begin{array}{cccc} \mathbf{J}_p \cdots \mathbf{J}_p & & & \\ & \mathbf{J}_{p-1} \cdots \mathbf{J}_{p-1} & & 0 \\ & & \ddots & \\ 0 & & & \mathbf{J}_1 \cdots \mathbf{J}_1 \end{array} \right),$$

where  $\mathbf{J}_i$  occurs  $n_i$  times

$$\mathbf{J}_i = \underbrace{\begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & \cdots & 1 & 0 \end{pmatrix}}_i$$

and  $n_i \geq 0$ ,  $i = 1, \dots, p$ . (Note:  $n_p \neq 0$ , but it can happen that  $n_{p-1} = 0$ , for example.)

PROOF. By (17.9) we can find a direct sum decomposition

$$\mathcal{V} = \mathcal{U}_1 \oplus \dots \oplus \mathcal{U}_h$$

of  $\mathcal{V}$  where each  $\mathcal{U}_i$  is  $T$ -invariant and the induced linear transformation

$$T: \mathcal{U}_i \rightarrow \mathcal{U}_i$$

has a cyclic vector. If we choose a basis for  $\mathcal{U}_i$  composed of  $\mathbf{A}_i, T\mathbf{A}_i, \dots, T^{p_i-1}\mathbf{A}_i$  where  $T^{p_i}(\mathbf{A}_i) = 0$ , then the matrix of

$$T: \mathcal{U}_i \rightarrow \mathcal{U}_i$$

with respect to this basis is

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

Assembling these blocks we obtain a matrix of the stated form representing  $T$ .  $\square$

*Remark.* In a finite-dimensional vector space  $\mathcal{V}$  an ascending chain of subspaces

$$\{0\} = \mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \dots \subseteq \mathcal{U}_n \subseteq \dots \subseteq \mathcal{V}$$

must eventually stabilize; that is, there must be an integer  $m$  such that

$$\mathcal{U}_i = \mathcal{U}_m \quad \text{if } i \geq m.$$

This is because

$$0 = \dim \mathcal{U}_0 \leq \dim \mathcal{U}_1 \leq \dots \leq \dim \mathcal{U}_n \leq \dots \leq \dim \mathcal{V},$$

and since  $\mathcal{U}_j \subseteq \mathcal{U}_{j+1}$   $\dim \mathcal{U}_j = \dim \mathcal{U}_{j+1}$  iff  $\mathcal{U}_j = \mathcal{U}_{j+1}$ .

Similarly, a descending chain of subspaces

$$\mathcal{V} = \mathcal{W}_0 \supseteq \mathcal{W}_1 \supseteq \dots \supseteq \mathcal{W}_n \supseteq \dots \supseteq \{0\}$$

must also stabilize at a subspace  $\mathcal{W}_l$  such that

$$\mathcal{W}_j = \mathcal{W}_l \quad \text{if } j \geq l$$

as one easily sees from the inequality

$$\dim \mathcal{V} = \dim \mathcal{W}_0 \geq \dim \mathcal{W}_1 \geq \dots \geq \dim \mathcal{W}_n \geq \dots \geq 0.$$

This simple property of *finite-dimensional* vector spaces will play a central role in the construction of the Jordan normal form.

We want to obtain a direct sum decomposition of  $\mathcal{V}$  into  $T$ -invariant subspaces for an arbitrary linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{V},$$

where  $\mathcal{V}$  is finite dimensional. The idea is to *enlarge* the eigenspaces of  $T$ . But how? Here is the idea. Suppose that  $e$  is an eigenvalue of  $T$ . Then

$$\mathcal{V}_e = \text{Ker}(T - eT).$$

Let us now look at

$$\mathcal{V}_{e,i} = \text{Ker}(T - eI)^i, \quad i = 1, 2, \dots$$

Notice that these subspaces are nested one within the other, that is,

$$\mathcal{V}_e = \mathcal{V}_{e,1} \subseteq \mathcal{V}_{e,2} \subseteq \mathcal{V}_{e,3} \subseteq \dots \subseteq \mathcal{V}.$$

If  $\mathcal{V}$  is finite dimensional then this chain of subspaces cannot go on getting bigger forever, for that would lead to an infinite set of linearly independent vectors in  $\mathcal{V}$ , contradicting the finite dimensionality of  $\mathcal{V}$ . Therefore there must be an integer  $n$  (depending on  $T, e$ , etc.) such that

$$\mathcal{V}_{e,j} = \mathcal{V}_{e,n} \quad \text{for all } j \geq n.$$

The vector space  $\mathcal{V}_{e,n} = \text{Ker}(T - eI)^n$  at which the chain

$$\mathcal{V}_e = \mathcal{V}_{e,1} \subseteq \mathcal{V}_{e,2} \subseteq \dots \subseteq \mathcal{V}_{e,n} \subseteq \mathcal{V}$$

stops growing is  $T$ -invariant, since

$$(T - eI)^n = T^n - eT^{n-1} + \dots + (-1)^n e^n I$$

is a polynomial in  $T$  and we may therefore apply (17.6). Let us introduce a name for  $\mathcal{V}_{e,n}$ .

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation with eigenvalue  $e$ . Let

$$\mathcal{V}_{e,i} = \text{Ker}(T - eI)^i, \quad i = 1, 2, \dots$$

Let  $n$  be an integer such that

$$\mathcal{V}_{e,n} = \mathcal{V}_{e,n+1} = \dots,$$

that is

$$\mathcal{V}_{e,j} = \mathcal{V}_{e,n} \quad \text{if } j \geq n.$$

The subspace  $\mathcal{V}_{e,n}$  is called the *closure* of the eigenspace  $\mathcal{V}_e$ . The closure of  $\mathcal{V}_e$  will be denoted by  $\overline{\mathcal{V}}_e$ . Note

$$\mathcal{V}_e = \mathcal{V}_{e,1} \subset \mathcal{V}_{e,2} \subset \dots \subset \mathcal{V}_{e,n} = \overline{\mathcal{V}}_e$$

is a chain of subspaces of  $\mathcal{V}$ .

*Remark.* Suppose that somewhere in the chain

$$\mathcal{V}_e = \mathcal{V}_{e,1} \subseteq \mathcal{V}_{e,2} \subseteq \cdots \subseteq \overline{\mathcal{V}}_e,$$

we find  $\mathcal{V}_{e,i} = \mathcal{V}_{e,i+1}$ . Then, in fact

$$\mathcal{V}_{e,i} = \mathcal{V}_{e,i+1} = \mathcal{V}_{e,i+2} = \cdots = \overline{\mathcal{V}}_e,$$

that is, the first time two of the subspaces  $\mathcal{V}_{e,i}$  agree, the chain stabilize; that is, nothing “new” happens after that point.

To see this suppose  $\mathcal{V}_{e,i} = \mathcal{V}_{e,i+1}$ . By definition  $\mathcal{V}_{e,i+2} = \text{Ker}(T - eI)^{i+2}$ . Suppose  $A \in \mathcal{V}_{e,i+2}$ . Then

$$(T - eI)(A) \in \mathcal{V}_{e,i+1} = \mathcal{V}_{e,i},$$

so

$$(T - eI)^{i+1}(A) = (T - eI)^i((T - eI)(A)) = 0,$$

since  $\mathcal{V}_{e,i} = \text{Ker}(T - eI)^i$ . But this says  $A \in \mathcal{V}_{e,i+1}$ . Therefore  $\mathcal{V}_{e,i+2} = \mathcal{V}_{e,i+1}$ . Since the reverse inclusion also holds we must have equality.

The closure of each eigenspace  $\mathcal{V}_e$  of a linear transformation  $T$  is  $T$ -invariant. Therefore  $T(\overline{\mathcal{V}}_e) \subset \overline{\mathcal{V}}_e$ . The following proposition tells us about the behavior of the linear transformation  $T: \overline{\mathcal{V}}_e \rightarrow \overline{\mathcal{V}}_e$ .

**Proposition 17.11.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation with eigenvalue  $e$ . Let  $\overline{\mathcal{V}}_e$  be the closure of the corresponding eigenspace  $\mathcal{V}_e$ . Then the induced linear transformation*

$$T - eI: \overline{\mathcal{V}}_e \rightarrow \overline{\mathcal{V}}_e$$

*is nilpotent of index  $k$ . Furthermore*

$$\text{Ker}(T - eI)^i = \overline{\mathcal{V}}_e, \quad i \geq k.$$

**PROOF.** Recall the chain of subspaces

$$\mathcal{V}_{e,i} = \text{Ker}(T - eI)^i, \quad i = 1, 2, \dots$$

We have  $\overline{\mathcal{V}}_e = \mathcal{V}_{e,n}$  where  $n$  is, say, the smallest integer such that

$$\mathcal{V}_{e,j} = \mathcal{V}_{e,n} \quad \text{if } j \geq n.$$

Therefore by the definitions

$$(T - eI)^n = 0: \overline{\mathcal{V}}_e \rightarrow \overline{\mathcal{V}}_e,$$

so

$$(T - eI): \overline{\mathcal{V}}_e \rightarrow \overline{\mathcal{V}}_e$$

is nilpotent of index  $k \leq n$ . To see that the index is exactly  $n$  we reason as follows:

$$(T - eI)^k = 0: \overline{\mathcal{V}}_e \rightarrow \overline{\mathcal{V}}_e$$

implies

$$\text{Ker}(T - eI)^k = \mathcal{V}_{e,k} \quad \overline{\mathcal{V}}_e = \mathcal{V}_{e,n}.$$

But  $n$  is the smallest integer such that the chain

$$\mathcal{V}_e = \mathcal{V}_{e,1} \subseteq \cdots \subseteq \mathcal{V}_{e,n} = \overline{\mathcal{V}}_e$$

stabilizes, so  $\mathcal{V}_{e,k} \subseteq \mathcal{V}_{e,n} = \overline{\mathcal{V}}_e$ . Therefore  $\mathcal{V}_{e,k} \subseteq \overline{\mathcal{V}}_e \subseteq \mathcal{V}_{e,k}$  and  $\overline{\mathcal{V}}_e = \mathcal{V}_{e,k}$ , which says  $k = n$ .  $\square$

We are now in a position to state the main result of this chapter

**Theorem 17.12.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the finite-dimensional complex vector space  $\mathcal{V}$ . Let*

$$p(t) = (t - e_1)^{d_1} \cdots (t - e_q)^{d_q}$$

*be the characteristic polynomial of  $T$ . Then:*

- (1) *The closure of each eigenspace  $\overline{\mathcal{V}}_{e_i}$  is fully  $T$ -invariant.*
- (2)  *$\dim \overline{\mathcal{V}}_{e_i} = d_i$ .*
- (3) *The characteristic polynomial of the induced linear transformation*

$$T: \overline{\mathcal{V}}_{e_i} \rightarrow \overline{\mathcal{V}}_{e_i}$$

*is*

$$(t - e_i)^{d_i}.$$

- (4)  *$\mathcal{V} = \overline{\mathcal{V}}_{e_1} \oplus \cdots \oplus \overline{\mathcal{V}}_{e_q}$ .*

We will need a number of preliminary steps before undertaking a proof of this result. But before embarking on this undertaking notice that combining (17.12) with the canonical matrix representation for nilpotent transformations of (17.10) will give us a matrix canonical form

$$\begin{pmatrix} \mathbf{M}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{M}_q \end{pmatrix},$$

where

$$\mathbf{M}_j = \begin{pmatrix} \mathbf{J}_{1,j} & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_{r_j,j} \end{pmatrix},$$

and

$$\mathbf{J}_{i,j} = \begin{pmatrix} e_j & & 0 \\ 1 & & \\ & \ddots & \\ 0 & & 1 & e_j \end{pmatrix}, \quad 1 \leq i \leq r_j; \quad 1 \leq j \leq q.$$

(See (17.14) and the ensuing discussion.) We separate out the crucial technical step in the proof of (17.12) in the following lemma:

**Lemma 17.13.** *Suppose  $S: \mathcal{V} \rightarrow \mathcal{V}$  a linear transformation in the finite-dimensional vector space  $\mathcal{V}$ . Then there exists an integer  $k$  such that:*

$$\left. \begin{array}{l} \text{Im } S^j = \text{Im } S^k, \\ \text{Ker } S^j = \text{Ker } S^k, \end{array} \right\} \text{ whenever } j \geq k,$$

*and, moreover,  $\text{Ker } S^k$  and  $\text{Im } S^k$  are a complementary pair of  $S$ -invariant subspaces that is, both are  $S$ -invariant and*

$$\mathcal{V} = \text{Im } S^k \oplus \text{Ker } S^k.$$

**PROOF.** The nested chains of subspaces

$$\{0\} = \text{Ker } S^0 \subseteq \text{Ker } S^1 \subseteq \cdots \subseteq \text{Ker } S^n \subseteq \cdots \subseteq \mathcal{V},$$

$$\mathcal{V} = \text{Im } S^0 \supseteq \text{Im } S^1 \supseteq \cdots \supseteq \text{Im } S^n \supseteq \cdots \supseteq \{0\},$$

must eventually stabilize because  $\mathcal{V}$  is finite dimensional. Say

$$\text{Ker } S^i = \text{Ker } S^{n'} \quad \text{if } i \geq n',$$

$$\text{Im } S^i = \text{Im } S^{n''} \quad \text{if } i \geq n''.$$

If  $k = \max\{n', n''\}$ , then from the  $k$ th term, both sequences are certainly stable. We claim

$$\text{Im } S^k \cap \text{Ker } S^k = \{0\}.$$

For if  $A \in \text{Im } S^k \cap \text{Ker } S^k$ , then there is a vector  $\bar{A} \in \mathcal{V}$  such that

$$A = S^k \bar{A}.$$

Then, of course

$$0 = S^k A = S^{2k} \bar{A},$$

so  $A \in \text{Ker } S^{2k}$ . But by choice of  $k$

$$\text{Ker } S^k = \text{Ker } S^{k+1} = \cdots = \text{Ker } S^{2k},$$

so  $A \in \text{Ker } S^k$ . Then

$$A = S^k A = 0$$

as claimed. For the linear transformation

$$S^k: \mathcal{V} \rightarrow \mathcal{V},$$

we have

$$\dim \mathcal{V} = \dim \text{Im } S^k + \dim \text{Ker } S^k.$$

On the other hand, since

$$\text{Im } S^k \cap \text{Ker } S^k = \{0\},$$

we have

$$\dim(\text{Im } S^k + \text{Ker } S^k) = \dim \text{Im } S^k + \dim \text{Ker } S^k$$

also, therefore

$$\dim \mathcal{V} = \dim(\text{Im } S^k + \text{Ker } S^k),$$

so  $\text{Im } S^k + \text{Ker } S^k$  is a subspace of  $\mathcal{V}$  with the same dimension as  $\mathcal{V}$ , and so must equal  $\mathcal{V}$ .  $\square$

PROOF OF 17.12. We begin by applying (17.11) to the linear transformation

$$S_1 = T - e_1 I: \mathcal{V} \rightarrow \mathcal{V}.$$

We conclude that for a suitable integer  $k$

$$\mathcal{V} = \text{Ker}(T - e_1 I)^k \oplus \text{Im}(T - e_1 I)^k,$$

where by the definition of  $\overline{\mathcal{V}}_{e_1}$  we have

$$\overline{\mathcal{V}}_{e_1} = \text{Ker}(T - e_1 I)^k.$$

By (17.7) the subspaces

$$\text{Im}(T - e_1 I)^k, \text{Ker}(T - e_1 I)^k$$

are  $T$ -invariant, so by (17.6)  $T$  has a matrix representation

$$M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

where  $A$  is a matrix representation for

$$T: \overline{\mathcal{V}}_{e_1} \rightarrow \overline{\mathcal{V}}_{e_1}$$

and  $B$  is a matrix representation for

$$T: \text{Im}(T - e_1 I)^k \rightarrow \text{Im}(T - e_1 I)^k.$$

From the matrix  $M$  we can compute the characteristic polynomial of  $T$ . We get

$$p(t) = p_1(t) \cdot p_2(t),$$

where

$$p_1(t) = \det(A - tI),$$

$$p_2(t) = \det(B - tI),$$

are the characteristic polynomials of  $T$  on  $\overline{\mathcal{V}}_{e_1}$  and  $\hat{\mathcal{V}}_{e_1} = \text{Im}(T - e_1 I)^k$ , respectively. Since  $\overline{\mathcal{V}}_{e_1}$  contains the eigenspace  $\mathcal{V}_{e_1}$  it follows that

$$p_1(t) = (t - e_1)^{d_1} p(t)$$



and  $p_2(t)$  cannot contain  $(t - e_1)$  as a factor. We claim  $p(t) = +1$ . To this end recall that

$$T - e_1 I: \overline{\mathcal{V}}_{e_1} \rightarrow \overline{\mathcal{V}}_{e_1}$$

is nilpotent of order  $k$ , so has a matrix representation in the form

$$\begin{pmatrix} 0 & & & 0 \\ * & & & \\ & \ddots & & \\ 0 & & * & 0 \end{pmatrix}, \quad * = 0 \text{ or } 1.$$

Therefore  $T: \overline{\mathcal{V}}_{e_1} \rightarrow \overline{\mathcal{V}}_{e_1}$  has a matrix representation

$$\begin{pmatrix} e_1 & & & 0 \\ * & & & \\ & \ddots & & \\ 0 & & * & e_1 \end{pmatrix}$$

and therefore

$$p_1(t) = \det \begin{pmatrix} e_1 - t & & & 0 \\ * & & & \\ & \ddots & & \\ 0 & & * & e_1 - t \end{pmatrix} = (e_1 - t)^{\dim \overline{\mathcal{V}}_{e_1}}.$$

Therefore  $\dim \overline{\mathcal{V}}_{e_1} = d_1$  and  $p_1(t) = +1$ , because the factorization of a polynomial into linear factors is unique. Therefore

$$\begin{cases} p_1(t) = (t - e_1)^{d_1}, \\ p_2(t) = (t - e_2)^{d_2} \cdots (t - e_q)^{d_q} \end{cases}$$

If we now repeat the preceding argument for the eigenvalue  $e_2$  of the linear transformation

$$T: \widehat{\mathcal{V}}_{e_1} \rightarrow \widehat{\mathcal{V}}_{e_1}$$

and so on, we arrive at the desired direct sum decomposition.  $\square$

Let us put this all together and see what it says about matrix representations of linear transformations. We start with a linear transformation

$$T: \mathcal{V} \rightarrow \mathcal{V}$$

in a finite-dimensional complex vector space  $\mathcal{V}$ . Let

$$p(t) = (t - e_1)^{d_1} \cdots (t - e_q)^{d_q}$$

be the characteristic polynomial of  $T$ , where  $e_1, \dots, e_q$  are the distinct eigenvalues of  $T$ . We apply (17.12) to obtain the direct sum decomposition

$$\mathcal{V} = \overline{\mathcal{V}}_{e_1} \oplus \cdots \oplus \overline{\mathcal{V}}_{e_q}$$

of  $\mathcal{V}$  into the sum of the closures of the eigenspaces. If we look at (17.6) again we see that we can certainly obtain a block matrix representation

$$\begin{pmatrix} \mathbf{M}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{M}_q \end{pmatrix}$$

for  $\mathbf{T}$ , where  $\mathbf{M}_j$  represents

$$\mathbf{T}: \overline{\mathcal{V}}_{e_j} \rightarrow \overline{\mathcal{V}}_{e_j}$$

for  $j = 1, \dots, q$ . We can use (17.12) to obtain more information about

$$\mathbf{T}: \overline{\mathcal{V}}_{e_j} \rightarrow \overline{\mathcal{V}}_{e_j}, \quad j = 1, \dots, q,$$

because

$$(\mathbf{T} - e_j \mathbf{I}) = \mathbf{N}_j: \overline{\mathcal{V}}_{e_j} \rightarrow \overline{\mathcal{V}}_{e_j}$$

is a nilpotent transformation of index  $d_j$ . We apply the knowledge we gained in (17.10) to  $\mathbf{N}_1, \dots, \mathbf{N}_q$ , to find a suitable basis for  $\overline{\mathcal{V}}_{e_j}$  with respect to which the nilpotent transformation  $\mathbf{N}_j$  has a matrix representative

$$\underbrace{\begin{pmatrix} \mathbf{N}_{n_j,j} & & 0 \\ & \ddots & \\ 0 & & \mathbf{N}_{s_j,j} \end{pmatrix}}_{d_j} \Bigg\} d_j, \quad n_j \geq \dots \geq s_j \geq 1,$$

and each block  $\mathbf{N}_{r_j,j}$  that appears has the form

$$\mathbf{N}_{r_i,j} = \underbrace{\begin{pmatrix} 0 & & 0 \\ 1 & \ddots & \\ & \ddots & 0 \\ 0 & & 1 & 0 \end{pmatrix}}_{r_j} \Bigg\} r_j$$

Since  $\mathbf{N}_j = \mathbf{T} - e_j \mathbf{I}$  this means that with respect to the same basis

$$\mathbf{T}: \overline{\mathcal{V}}_{e_j} \rightarrow \overline{\mathcal{V}}_{e_j}$$

is represented by the matrix

$$\mathbf{J}_j = \begin{pmatrix} \mathbf{J}_{n_j,j} & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_{s_j,j} \end{pmatrix},$$

where

$$\mathbf{J}_{r_i,j} = \begin{pmatrix} e_j & & 0 \\ 1 & \ddots & \\ & \ddots & e_j \\ 0 & & 1 & e_j \end{pmatrix}.$$

Altogether this says that  $T$  has the matrix representation:

$$\begin{pmatrix} \mathbf{J}_{e_1,1} & & & & & \\ & \mathbf{J}_{r_1,1} & & & & \\ & & \mathbf{J}_{e_2,2} & & & \\ & & & \mathbf{J}_{r_2,2} & & \\ & 0 & & & \ddots & \\ & & & & & \mathbf{J}_{r_q,q} \end{pmatrix},$$

where

$$\mathbf{J}_{sij} = \begin{pmatrix} e_j & 0 \\ 1 & \\ 0 & \underbrace{e_j \quad 1 \quad e_j}_{s_i} \end{pmatrix},$$

or

The diagram shows a curved boundary enclosing a region. Inside, a sequence of points and segments is shown. The points are labeled  $e_1, d_1, e_1, e_2, d_2, e_2, \dots, d_q, e_q$ . The points  $e_i$  are marked with an asterisk (\*). Segments connect  $e_1$  to  $d_1$ ,  $d_1$  to  $e_1$ ,  $e_1$  to  $e_2$ ,  $e_2$  to  $d_2$ , and so on. The regions between the curve and the segments are labeled 0.

where  $\ast = 0$  or  $1$ . The pattern of  $0$ 's and  $1$ 's being determined as in (17.10) by the nilpotent transformations

$$T - e_i | : \overline{\mathcal{V}}_i \rightarrow \overline{\mathcal{V}}_i, \quad i = 1, \dots, q.$$

Let us put this together as a theorem.

**Theorem 17.14.** *Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the finite-dimensional complex vector space  $\mathcal{V}$ . Let*

$$p(t) = (t - e_1)^{d_1} \cdots (t - e_q)^{d_q}$$

be the characteristic polynomial of  $T$ . Then there exists a basis for  $\mathcal{V}$  with respect to which the matrix representation of  $T$  has the block matrix form

$$\begin{pmatrix} \mathbf{M}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{M}_q \end{pmatrix}$$

and

$$\mathbf{M}_j = \underbrace{\begin{pmatrix} \mathbf{J}_{n_j, j} & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_{s_j, j} \end{pmatrix}}_{d_j} \Bigg\} d_j, \quad n_j \geq \cdots \geq s_j \geq 1,$$

and

$$\mathbf{J}_{r_j j} = \begin{pmatrix} e_j & & 0 \\ 1 & \ddots & \\ & \ddots & e_j \\ 0 & & 1 & e_j \end{pmatrix}, \quad j = 1, \dots, q. \quad \square$$

This matrix representation is called the *Jordan canonical form*. The matrices

$$\mathbf{J}_{r_j j}: j = 1, \dots, q, n_j \geq r_j \geq s_j$$

are called the *Jordan blocks*.

Let us look at some examples!

EXAMPLE 6. Let  $T: \mathbb{C}^3 \rightarrow \mathbb{C}^3$  be given by

$$T(x, y, z) = (5x + 4y + 3z, -x - 3z, x - 2y + z).$$

With respect to the standard basis for  $\mathbb{C}^3$  the matrix representation of  $T$  is

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

The characteristic polynomial of  $T$  is

$$\begin{aligned} p(t) &= \det(\mathbf{A} - t\mathbf{I}) = \det \begin{pmatrix} 5-t & 4 & 3 \\ -1 & -t & -3 \\ 1 & -2 & 1-t \end{pmatrix} \\ &= (5-t)[-t(1-t) - 6] - 4[-1(1-t) + 3] + 3[2 - (-t)] \\ &= -t^3 + 6t^2 - 32 \\ &= -(t-4)^2(t+2), \end{aligned}$$

so there are two eigenvalues

$$\lambda_1 = 4 \text{ of algebraic multiplicity } 2,$$

$$\lambda_2 = -2 \text{ of algebraic multiplicity } 1.$$

Let us find the eigenspace corresponding to the eigenvalue 4. To do so we must solve the linear system

$$\begin{pmatrix} 1 & 4 & 3 \\ -1 & -4 & -3 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This leads to the equivalent systems

$$\begin{pmatrix} 1 & 4 & 3 \\ 0 & 0 & 0 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 2 & 2 & 0 \\ 0 & 0 & 0 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

so

$$x = -y = z$$

and the eigenspace is one-dimensional with basis

$$(1 \quad -1 \quad 1) = E_1.$$

To find a basis for the eigenspace corresponding to the eigenvalue  $-2$  we need to solve the linear system

$$\begin{pmatrix} 7 & 4 & 3 \\ -1 & 2 & -3 \\ 1 & -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which leads to the solution

$$x = -y = -z.$$

Therefore the eigenspace corresponding to the eigenvalue  $-2$  has dimension 1 with basis  $E_2 = (1, -1, -1)$ .

The linear transformation  $T$  *cannot* be put into diagonal form because the eigenvalue 4 has algebraic multiplicity 2 but the geometric multiplicity is only 1.

The Jordan normal form of  $T$  is

$$\begin{pmatrix} 4 & 0 & 0 \\ * & 4 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

where  $\ast = 0, 1$ ,  $\ast$  cannot be 0 because  $T$  is not diagonalizable. Thus we already know

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

is the Jordan canonical form for  $T$  *although we have not yet found the basis with respect to which  $T$  is represented by this matrix.*

To find the corresponding basis for  $\mathcal{V} = \mathbb{C}^3$  we must find a basis for  $\overline{\mathcal{V}}_4$ . Since we know  $\dim \overline{\mathcal{V}}_4 = 2$  and  $\mathcal{V}_4 \subset \overline{\mathcal{V}}_4$  we must find a vector  $F_4$  such that

$$(T - 4I)^2(F_4) = 0,$$

but

$$(T - 4I)(F_4) \neq 0.$$

To find  $F$  notice

$$(T - 4I)^2 = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -4 & -3 \\ 1 & -2 & -3 \end{pmatrix}^2 = \begin{pmatrix} 0 & -18 & -18 \\ 0 & 18 & 18 \\ 0 & 18 & 18 \end{pmatrix}.$$

So a basis for  $\text{Ker}(T - 4I)^2$  is the solution space of the linear system

$$y = z = 0.$$

One solution is the eigenvector  $E_1 = (1, -1, 1)$  and a *linearly independent* solution is  $F = (0, 1, -1)$ . The matrix of  $T$  with respect to the ordered basis  $(F_1, F_2, F_3)$  is the Jordan normal form

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

where  $F_1 = F$ ,  $F_2 = (T - 4I)(F_1)$  and  $F_3 = E_2$ . Another way to obtain the missing vector  $F$  is to reason as follows.

To find  $\text{Ker}(T - 4I)^2$  we can use the fact that  $E_1 \in \text{Ker}(T - 4I)$  and look for a solution  $F$  to the vector equation

$$(T - 4I)(F) = E_1.$$

We will then have

$$(T - 4I)^2(F) = (T - 4I)(E_1) = 0.$$

To solve the vector equation

$$(T - 4I)(F) = E_1,$$

we solve the linear system

$$\begin{pmatrix} 1 & 4 & 3 \\ -1 & -4 & -3 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},$$

or what is the same

$$\begin{pmatrix} 1 & 4 & 3 \\ 0 & 0 & 0 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

so

$$\begin{pmatrix} 2 & 2 & 0 \\ 0 & 0 & 0 \\ 1 & -2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix},$$

so

$$x + y = 1,$$

$$x - y - 3z = 1.$$

A solution is

$$x = 0,$$

$$y = 1,$$

$$z = -1,$$

so  $\mathbf{F} = (0, 1, -1)$  is a vector of the type we are looking for.

This second method is the more systematic of the two but does not always apply as the following example serves to illustrate.

**EXAMPLE 7.** Let  $T: \mathbb{C}^3 \rightarrow \mathbb{C}^3$  be the linear transformation represented by the matrix

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

with respect to the standard basis of  $\mathbb{C}^3$ .

The characteristic polynomial of  $T$  is

$$p(t) = \det \begin{pmatrix} 2-t & 2 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & 2 \end{pmatrix} = \dots = -(t-1)^3.$$

Therefore  $t = 1$  is an eigenvalue of algebraic multiplicity 3. To determine the corresponding eigenspace we solve the linear system

$$\begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \\ -1 & -2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This leads to the equation

$$x + 2y - z = 0,$$

which defines a plane in  $\mathbb{C}^3$  with a basis  $\mathbf{E}_1 = (1, 0, 1)$ ,  $\mathbf{E}_2 = (0, 1, 2)$ . Since the eigenspace  $\mathcal{V}_1$  is two-dimensional and 1 is the only eigenvalue we have  $\overline{\mathcal{V}}_1 = \mathbb{C}^3$ , and so  $\overline{\mathcal{V}}_1$  has dimension 1 over  $\mathcal{V}_1$ . To find a basis for  $\mathcal{V}_1$  over  $\mathcal{V}_1$  we may choose any vector in  $\mathbb{C}^3$  not in  $\mathcal{V}_1$ , since it will automatically be in  $\text{Ker}(\mathbf{T} - \mathbf{I})^2$ . (Note: Of course  $(\mathbf{T} - \mathbf{I})^2 = 0$  in this case.)

One such vector is  $\mathbf{F}_1 = (1, 0, 0)$ . We now follow the prescription of (17.14)

$$(\mathbf{T} - \mathbf{I})(\mathbf{F}_1) = \begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \\ -1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$$

and  $\mathbf{F}_2 = (\mathbf{T} - \mathbf{I})(\mathbf{F}_1) = (1, -1, -1)$  belongs to  $\mathcal{V}_1$ . If we extend  $\mathbf{F}_2$  to a basis for  $\mathcal{V}_1$  by adjoining say  $\mathbf{F}_3 = (0, 1, -2)$  we find that the matrix of  $\mathbf{T}$  with respect to the basis  $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\}$  is in the Jordan normal form

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

Note that in this example we could not simply proceed as in Example 6, and once we found  $\mathcal{V}_1$  look for a vector  $\mathbf{F}$  that satisfies

$$(*) \quad (\mathbf{T} - \mathbf{I})(\mathbf{F}) = \mathbf{E}_1 \quad (\text{or } \mathbf{E}_2).$$

This is because  $\mathcal{V}_1 = \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2)$  is two-dimensional. In fact  $(*)$  has no solutions whatsoever in this case. What we could do instead is look for a vector  $\mathbf{F} \neq 0$  such that

$$(\mathbf{T} - \mathbf{I})(\mathbf{F}) \in \mathcal{L}(\mathbf{E}_1, \mathbf{E}_2).$$

This approach will be illustrated in the next example.

**EXAMPLE 8.** As a last example we consider the linear transformation represented by the  $4 \times 4$  matrix

$$\mathbf{A} = \begin{pmatrix} 7 & 1 & 2 & 2 \\ 1 & 4 & -1 & -1 \\ -2 & 1 & 5 & -1 \\ 1 & 1 & 2 & 8 \end{pmatrix}.$$



We proceed step-by-step as in the two preceding examples. First we compute the characteristic polynomial

$$\begin{aligned} p(t) &= \det \begin{pmatrix} 7-t & 1 & 2 & 2 \\ 1 & 4-t & -1 & -1 \\ -2 & 1 & 5-t & -1 \\ 1 & 1 & 2 & 8-t \end{pmatrix} \\ &= (7-t)[\cdots] \\ &= (t-6)^4. \end{aligned}$$

So there is exactly one eigenvalue, namely  $t = 6$  with algebraic multiplicity 4. To find a basis for the eigenspace  $\mathcal{V}_4$  we must solve the linear system

$$0 = (A - 6I) \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}.$$

Since the first and last row are identical and the sum of the second and third is the first row we can work with the equivalent system

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

which easily leads to

$$\begin{aligned} y + z + w &= 0, \\ x + w + z &= 0, \end{aligned}$$

and the basis

$$\mathbf{A}_1 = (-1, -1, 1, 0), \quad \mathbf{A}_2 = (-1, -1, 0, 1)$$

for the solution space. Therefore

$$2 = \dim \mathcal{V}_6 < \dim \overline{\mathcal{V}}_6 = 4.$$

In the next step we need to extend  $\mathbf{A}_1, \mathbf{A}_2$  to a basis for

$$\mathcal{V}_{6,2} = \text{Ker}(A - 6I)^2.$$

So we look for vectors  $\mathbf{F}$  such that

$$(A - 6I)(\mathbf{F}) \in \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2) = \mathcal{V}_{6,1}$$

That means we are looking for solutions to

$$\begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = a\mathbf{A}_1 + b\mathbf{A}_2 = \begin{pmatrix} -a & -b \\ -a & -b \\ a & b \\ b & \end{pmatrix},$$

where  $a$  and  $b$  must first be chosen so that solutions exist. From the first and fourth equation we see  $b = -a - b$ , so  $2b = -a$ . If we set  $b = 1$  and  $a = -2$  we obtain the *consistent* system

$$\begin{aligned}x - 2y - z - w &= 1, \\ -2x + y - z - w &= -2.\end{aligned}$$

(Note: Equations 1 and 4 are consequences of these.) From these we obtain

$$y = x - 1 = -(z + w)$$

so a solution is

$$\mathbf{A}_3 = (1, 0, 0, 0).$$

There are in fact no further linear independent solutions to the system

$$(\mathbf{A} - 6\mathbf{I})^2 \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

because any such solution must be of the form

$$c\mathbf{A}_3 + \mathbf{B}, \quad \mathbf{B} \in \mathcal{L}(\mathbf{A}_1, \mathbf{A}_2).$$

Finally, we must extend  $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$  to a basis for

$$\overline{\mathcal{V}}_6 = \mathcal{V}_{6,3} = \text{Ker}(\mathbf{A} - 6\mathbf{I})^3.$$

This is equivalent to finding a solution to the system

$$(\mathbf{A} - 6\mathbf{I}) \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = a_1\mathbf{A}_1 + a_2\mathbf{A}_2 + a_3\mathbf{A}_3.$$

Thus we must solve the system

$$\begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & 1 & 2 & 2 \end{pmatrix} = \begin{pmatrix} -a_1 & -a_2 & +a_3 \\ -a_1 & -a_2 & \\ & a_1 & \\ & a_2 & \end{pmatrix}.$$

From the first and fourth row we get

$$-a_1 - a_2 + a_3 = a_2$$

and since the sum of the two middle rows is the negative of the first we get

$$-a_1 - a_2 + a_1 = -a_1 - a_2 + a_3.$$

From which we see

$$a_1 = 0, \quad a_2 = 1, \quad a_3 = -3$$

leads to a consistent system with solution

$$\mathbf{A}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

As basis for  $\mathbb{C}^4$  we have

$$\mathbf{B}_1 = \mathbf{A}_4,$$

$$\mathbf{B}_2 = (\mathbf{A} - 6\mathbf{I})(\mathbf{B}_1) = \begin{pmatrix} 2 \\ -1 \\ -1 \\ 2 \end{pmatrix},$$

$$\mathbf{B}_3 = (\mathbf{A} - 6\mathbf{I})(\mathbf{B}_2) = \begin{pmatrix} 3 \\ 3 \\ -6 \\ 3 \end{pmatrix},$$

$$\mathbf{B}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -2 \end{pmatrix},$$

where  $\mathbf{B}_4$  is an eigenvector linearly independent from the eigenvector  $\mathbf{B}_1$ . With respect to this basis the matrix of  $T$  is

$$\begin{pmatrix} 6 & 0 & 0 & 0 \\ 1 & 6 & 0 & 0 \\ 0 & 1 & 6 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}.$$

The main results of this chapter may be summarized in the following two theorems.

**Theorem 17.15.** *Let  $A, B \in \mathcal{M}_n$ . Then  $A$  and  $B$  represent the same linear transformation  $T: \mathbb{C}^n \rightarrow \mathbb{C}^n$  with respect to different bases iff  $A$  and  $B$  have the same Jordan normal form.*  $\square$

**Theorem 17.16.** Let  $S, T: \mathcal{V} \rightarrow \mathcal{V}$  be linear endomorphisms of the finite-dimensional complex vector space  $\mathcal{V}$ . Then there exists an isomorphism  $P: \mathcal{V} \rightarrow \mathcal{V}$  such that  $T = PSP^{-1}$  iff  $S$  and  $T$  have the same Jordan normal form.  $\square$

## EXERCISES

1. Give an example of a linear transformation in a finite-dimensional *real* vector space with no nontrivial invariant subspace.
2. Let  $T: \mathbb{C}^2 \rightarrow \mathbb{C}^2$  be given the linear transformation given by the matrix

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

with respect to some basis. What are the invariant subspaces of  $T$ .

3. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation with invariant subspaces  $\mathcal{U}, \mathcal{W}$ . Show  $\mathcal{U} + \mathcal{W}$  and  $\mathcal{U} \cap \mathcal{W}$  are also invariant subspaces of  $T$ .
4. Let  $D: \mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$  be the differentiation operator. Show  $\mathcal{P}_i(\mathbb{R})$ ,  $i = 1, \dots, n$  are all invariant subspaces of  $D$ . Does any of them have an invariant complement?
5. Which of the following subspaces of  $\mathbb{C}^3$  are complementary to

$$\mathcal{U} = \{(x, y, z) | x + y + z = 0\}.$$

- (a)  $\{(x, y, z) | x + 2y + z = 0\}$ ;
- (b)  $\{(x, y, z) | x - y + z = 0\}$ ;
- (c)  $\{(x, y, z) | x + 2y + z = 0 \text{ and } x - y + z = 0\}$ ;
- (d)  $\mathcal{L}((1, -1, 1))$ ;
- (e)  $\mathcal{L}((1, -2, 1))$ .

6. Let us define two subspaces of  $\mathcal{P}_n(\mathbb{R})$  by

$$\mathcal{S} = \{p \in \mathcal{P}_n(\mathbb{R}) | p(x) = p(-x)\},$$

$$\mathcal{A} = \{p \in \mathcal{P}_n(\mathbb{R}) | p(x) = -p(-x)\}.$$

Show that  $\mathcal{S}$  and  $\mathcal{A}$  are complementary subspaces. Are they left invariant by the differentiation operator  $D$ ?

7. Let  $\mathcal{V}$  be an inner product space and let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation with invariant subspace  $\mathcal{W}$ . Show that  $\mathcal{W}^\perp$  is invariant for the adjoint  $T^*: \mathcal{V} \rightarrow \mathcal{V}$ .
8. For each of the subspaces  $\mathcal{W}$  of  $\mathbb{C}^3$  in Exercise 5 find a basis for  $\mathbb{C}^3$  over  $\mathcal{W}$ .

9. Verify that each of the following transformations is nilpotent and find their Jordan normal form

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

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \\ -1 & -2 & 1 \end{pmatrix},$$

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

10. Let  $\mathbf{T}: \mathbb{C}^2 \rightarrow \mathbb{C}^2$  with just one eigenvalue  $\lambda$ . Show that  $\mathbf{T} - \lambda \mathbf{I}$  is a nilpotent linear transformation.
11.  $\mathbf{N}: \mathcal{V} \rightarrow \mathcal{V}$  be a nilpotent transformation. Show that for any linear transformation  $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{V}$  that  $\mathbf{NT} - \mathbf{TN} = \mathbf{S}$  is a nilpotent transformation.
12. Let  $\mathbf{N}: \mathcal{V} \rightarrow \mathcal{V}$  be a nilpotent transformation in an  $n$ -dimensional vector space  $\mathcal{V}$ . Show that the characteristic polynomial  $\mathbf{P}_{\mathbf{N}}(t)$  is  $(-1)^n t^n$ .
13. Show that the trace of a nilpotent transformation is zero. Is the converse true?
14. Let  $\mathbf{T}: \mathbb{C}^3 \rightarrow \mathbb{C}^3$  be represented with respect to the standard basis by each of the following matrices in turn. Find the Jordan normal form of  $\mathbf{T}$

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

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 1 \end{pmatrix}.$$

15. Let  $T: \mathbb{C}^5 \rightarrow \mathbb{C}^5$  be represented by the matrix

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

with respect to the standard basis of  $\mathbb{C}^5$ . Find the Jordan normal form of  $T$ .

16. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in a finite-dimensional complex vector space  $\mathcal{V}$ . Show that  $T = A + N$ , where  $A: \mathcal{V} \rightarrow \mathcal{V}$  is an isomorphism and  $N: \mathcal{V} \rightarrow \mathcal{V}$  is nilpotent.

**Definition.** Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a linear transformation in the finite-dimensional complex vector space  $\mathcal{V}$ . For each eigenvalue  $e_i$  of  $T$  let the integer  $n_i$  be defined by  $n_i = \dim \overline{\mathcal{V}}_{e_i}^{n_i}$ . Let  $m(t) = (t - e_1)^{n_1} \cdots (t - e_q)^{n_q}$  where  $e_1, \dots, e_q$  are the distinct eigenvalues of  $T$ . The polynomial  $m_T(t)$  is called the *minimal polynomial* of  $T$ .

17. Show that  $n_i$  is the smallest integer such that  $\text{Ker}(T - e_i I)^{n_i} = \overline{\mathcal{V}}_{e_i}$ .
18. Show that  $(T - e_i I): \mathcal{V}_{e_i} \rightarrow \mathcal{V}_{e_i}$  is nilpotent of index  $n_i$ .
19. Show that  $n_i$  is less than or equal to the algebraic multiplicity of the eigenvalue  $e_i$ .
20. Show that the linear transformation  $m(T): \mathcal{V} \rightarrow \mathcal{V}$  is the zero transformation.
21. (Hamilton–Cayley Theorem). Show that the linear transformation  $p(T): \mathcal{V} \rightarrow \mathcal{V}$  is the zero transformation where  $p(t)$  is the characteristic polynomial of  $T$ . (*Hint:* Use Exercise 19 to show  $m(t)$  divides  $p(t)$ ; that is, that there exists a third polynomial  $q(t)$  such that  $p(t) = q(t) \cdot m(t)$ . Then use Exercise 20.)
22. Let  $T: \mathcal{V} \rightarrow \mathcal{V}$  be a nonsingular linear transformation in a finite-dimensional complex vector space  $\mathcal{V}$ . Show that there exists a linear transformation  $S: \mathcal{V} \rightarrow \mathcal{V}$  such that  $S^2 = T$ . (*Hint:* Use Jordan normal form.) We call  $S$  the square root of  $T$ .
23. Find a square root of each of the following

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

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

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

# 18 Applications to linear differential equations

Differential equations have played an important part in the development of mathematics since the time of Newton. They have also been central to many applications of mathematics to the physical sciences and technology. Often the equations relevant to practical applications are so difficult to solve explicitly that they can only be handled with approximation techniques on large computer systems. In this chapter we will be concerned with a simple form of differential equation, and systems thereof, namely, linear differential equations with constant coefficients. These systems have a great deal in common with systems of linear equations, and we are in a position to apply the hard-won knowledge about Jordan canonical forms to solve such systems. We begin with a short excursion into the subject of differential equations.

**EXAMPLE 1.** We are looking for all differentiable functions  $y: \mathbb{R} \rightarrow \mathbb{R}$  with the property that the derivative  $dy/dt$  of  $y$  satisfies the equation

$$\frac{dy}{dt}(x) = r \cdot y(x), \quad x \in \mathbb{R},$$

where  $r$  is some fixed number.

*Solution.*  $y = ce^{rt}$  where  $c$  is a constant.

*Reason.* Fundamental theorem of the calculus.

**Definition.** Let  $n$  be a positive integer,  $B \subset \mathbb{R}^{n+2}$  a nonempty subset, and

$$\mathbf{F}: B \rightarrow \mathbb{R}$$

a function. A function

$$y: J \rightarrow \mathbb{R}$$

$J$  an interval (i.e.,  $J = \{t \in \mathbb{R}, a \leq t \leq b\}$ ) is called a *solution to the differential equation*

$$F\left(t, y, \frac{dy}{dt}, \dots, \frac{d^n y}{dt^n}\right) = 0$$

iff the following conditions are fulfilled:

- (a)  $y$  is an  $n$ -times differentiable functions;
- (b)  $\forall x \in J$  the  $(n + 2)$ -tuple

$$\left(x, y(x), \frac{dy}{dt}(x), \dots, \frac{d^n y}{dt^n}(x)\right)$$

belongs to  $B$ ; and

$$(c) \quad \forall x \in J: F\left(x, y(x), \frac{dy}{dt}(x), \dots, \frac{d^n y}{dt^n}(x)\right) = 0.$$

A class of differential equations of particular interest are those of the form

$$\frac{d^n y}{dt^n} - f\left(t, y, \dots, \frac{d^{n-1} y}{dt^{n-1}}\right) = 0,$$

or, equivalently

$$\frac{d^n y}{dt^n} = f\left(t, y, \dots, \frac{d^{n-1} y}{dt^{n-1}}\right).$$

Such an equation is of the  $n$ th order.

We can also consider *systems* of differential equations. They are analogous to simultaneous equations.

**Definition.** A solution to the first-order system

$$\frac{dy_1}{dt} = f_1(x, y, \dots, y_n),$$

$$\frac{dy_2}{dt} = f_2(x, y_1, \dots, y_n),$$

$$\frac{dy_n}{dt} = f_n(x, y_1, \dots, y_n),$$

where  $f_i: B \rightarrow \mathbb{R}$ ,  $i = 1, \dots, n$ , are functions defined on  $\emptyset \neq B \subset \mathbb{R}^{n+1}$ , consists of an interval  $J \subset \mathbb{R}$  and functions

$$y_i: J \rightarrow \mathbb{R}, \quad i = 1, \dots, n,$$



such that:

- (a)  $y_1, \dots, y_n$  are differentiable on  $J$ ;
- (b)  $\forall x \in J$  the  $(n+1)$ -tuple  $(x, y_1(x), \dots, y_n(x))$  belongs to  $B$ ; and
- (c)  $\forall x \in J$  and each  $1 \leq i \leq n$

$$\frac{dy_i}{dt}(x) = f_i(x, y_1(x), \dots, y_n(x)).$$

If each  $f_i$ ,  $1 \leq i \leq n$  is a linear function, that is

$$f_i(x_1, \dots, x_{n+1}) = \sum_{j=1}^{n+1} a_{ij}x_j, \quad 1 \leq i \leq n,$$

for  $a_{ij} \in \mathbb{R}$ , then the above system is called a *linear differential system of the first order*. If, furthermore,  $a_{i1} = 0$ ,  $i = 1, \dots, n$  (i.e., “ $x$  does not appear on the right-hand side”) then the system is called *homogeneous*.

*Remarks.* 1. Do not let these definitions scare you! We will only be concerned with linear differential systems of the first order.

2. Note that for linear systems that the functions  $f_i$  are defined on all of  $\mathbb{R}^{n+1}$ .

3. We can write a linear differential system in the matrix form

$$\begin{pmatrix} \frac{dy_1}{dt}(x) \\ \vdots \\ \frac{dy_n}{dt}(x) \end{pmatrix} = bx + A \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix}$$

where

$$b = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad A = \begin{pmatrix} a_{12} & \cdots & a_{1,n+1} \\ & \ddots & \\ a_{n2} & \cdots & a_{n,n+1} \end{pmatrix}.$$

The system is homogeneous when  $b = 0 \in \mathbb{R}^n$ .

Suppose that  $f: B \rightarrow \mathbb{R}$ ,  $B \subset \mathbb{R}^{n+1}$ , and that we are looking for a solution  $y: J \rightarrow \mathbb{R}$  to the  $n$ th-order equation

$$(*) \quad \frac{d^n y}{dt^n}(x) = f\left(x, y, \frac{dy}{dt}, \dots, \frac{d^{n-1}y}{dt^{n-1}}\right).$$

If we set

$$\begin{aligned} y_1 &= y, \\ y_2 &= \frac{dy}{dt} = \frac{dy_1}{dt}, \\ &\vdots \\ y_n &= \frac{d^{n-1}y}{dt^{n-1}} = \frac{dy_{n-1}}{dt} \end{aligned}$$

then  $y$  is a solution to (\*) iff  $(y_1, \dots, y_n)$  is a solution to the first-order system

$$\begin{aligned}\frac{dy_1}{dt} &= y_2, \\ \frac{dy_{n-2}}{dt} &= y_{n-1}, \\ \frac{dy_{n-1}}{dt} &= y_n, \\ \frac{dy_n}{dt} &= f(x, y_1, \dots, y_n).\end{aligned}$$

In this way solving a single differential equation of order  $n$  is equivalent to solving a first-order system.

Before we take up some examples let us consider the structure of the set of all solutions to a linear differential system (compare (13.6)).

**Proposition 18.1.** *Suppose*

$$(*) \quad \begin{pmatrix} \frac{dy_1}{dt} \\ \vdots \\ \frac{dy_n}{dt} \end{pmatrix} = bx + A \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$A \in \mathcal{M}_n$ ,  $b \in \mathbb{R}^n$ , is a linear differential system of the first order. Let  $\mathcal{V}$  be the set of all solutions on some interval  $J = [a, b]$  to the associated homogeneous system

$$(H*) \quad \frac{dy}{dt} = Ay, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then

- (a)  $\mathcal{V}$  is a vector subspace of  $\mathcal{C}(a, b) \oplus \dots \oplus \mathcal{C}(a, b)$  (Chapter 4, Example 5).
- (b) If  $z = (z_1, \dots, z_n)$  is a solution to (\*) then every solution to (\*) is of the form  $z + y$  for some  $y \in \mathcal{V}$ .

**PROOF.** (a) First of all  $\mathbf{0} \in \mathcal{V}$  because if

$$\theta = (0, \dots, 0),$$

then

$$\frac{d\theta}{dt} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = A \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = A\theta.$$

Next, if  $\varphi, \psi \in \mathcal{V}$  and  $a, b \in \mathbb{R}$  then using the linearity of differentiation we find

$$\begin{aligned}\frac{d}{dt}(a\varphi + b\psi) &= a \frac{d\varphi}{dt} + b \frac{d\psi}{dt} \\ &= a\mathbf{A}\varphi + b\mathbf{A}\psi \\ &= \mathbf{A}(a\varphi + b\psi),\end{aligned}$$

which says  $a\varphi + b\psi$  solves the equation (H\*) and so  $a\varphi + b\psi \in \mathcal{V}$

(b) If  $z$  is a fixed solution to (\*) and  $w$  any other solution, then

$$\begin{aligned}\frac{d}{dt}(z - w) &= \frac{dz}{dt} - \frac{dw}{dt} \\ &= bx + \mathbf{A}z - (bx + \mathbf{A}w) \\ &= \mathbf{A}(z - w),\end{aligned}$$

so  $z - w$  solves (H\*) and hence  $z - w = y \in \mathcal{V}$ . □

**Definition.** A linear homogeneous differential system of the first order

$$(*) \quad \frac{dy}{dt} = \mathbf{A}y, \quad \mathbf{A} \in \mathcal{M}_n,$$

is said to be *diagonalizable* if the matrix  $\mathbf{A}$  is diagonalizable.

Suppose that

$$(*) \quad \frac{dy}{dt} = \mathbf{A}y$$

is a diagonalizable system. Then there exists a diagonal matrix  $\mathbf{D}$  and an invertible matrix  $\mathbf{P} \in \mathcal{M}_n$  such that

$$\mathbf{A} = \mathbf{PDP}^{-1}.$$

Thus we can write (\*) in the form

$$\frac{dy}{dt} = \mathbf{PDP}^{-1}y.$$

If we multiply both sides of this equation by  $\mathbf{P}^{-1}$  we obtain the equivalent equation

$$\mathbf{P}^{-1} \frac{dy}{dt} = \mathbf{DP}^{-1}y.$$

Since  $\mathbf{P}^{-1}$  is a matrix of constants, i.e., does not depend on the variable  $t$ , we can write further

$$\mathbf{P}^{-1} \frac{dy}{dt} = \frac{d}{dt}(\mathbf{P}^{-1}y).$$

Now comes the crucial step. Set

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and combine the two preceding equations in

$$\frac{dw}{dt} = \mathbf{D}w.$$

Now  $\mathbf{D}$  being a diagonal matrix, this latter system is just

$$\begin{aligned} \frac{dw_1}{dt} &= d_1 w_1, \\ &\vdots \\ \frac{dw_n}{dt} &= d_n w_n, \end{aligned}$$

and so has the solution

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} c_1 e^{d_1 t} \\ \vdots \\ c_n e^{d_n t} \end{pmatrix}.$$

But since

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

we have

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{P} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \mathbf{P} \begin{pmatrix} c_1 e^{d_1 t} \\ \vdots \\ c_n e^{d_n t} \end{pmatrix}.$$

and thus we have shown:

**Theorem 18.2.** *Let*

$$(*) \quad \frac{dy}{dt} = \mathbf{A}y, \quad \mathbf{A} \in \mathcal{M}_{nn},$$

*be a diagonalizable linear system, with*

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1}$$

for a diagonal matrix  $\mathbf{D}$  and invertible matrix  $\mathbf{P}$ . Then the solutions of (\*) are of the form

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{P} \begin{pmatrix} c_1 e^{d_1 t} \\ \vdots \\ c_n e^{d_n t} \end{pmatrix},$$

where  $\mathbf{D} = \text{diag}\{d_1, \dots, d_n\}$  and  $c_1, \dots, c_n$  are constants. □

EXAMPLE 2. Solve the linear differential system

$$\frac{dy_1}{dt} = y_2,$$

$$\frac{dy_2}{dt} = y_1.$$

In matrix form this is the linear differential system

$$\frac{dy}{dt} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The coefficient matrix

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

is symmetric, and therefore diagonalizable. To diagonalize  $\mathbf{A}$  we first compute the characteristic polynomial

$$p(t) = \det \begin{pmatrix} -t & 1 \\ 1 & -t \end{pmatrix} = (t^2 - 1) = (t - 1)(t + 1).$$

There are two eigenvalues  $t = \pm 1$  and corresponding eigenvectors can be easily found, viz.,

$$t = 1: \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad \text{Sol.: } \mathbf{E}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$t = -1: \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad \text{Sol.: } \mathbf{E}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We can therefore write

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

where

$$\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = (\mathbf{E}_1 \quad \mathbf{E}_2)$$

$\begin{matrix} \uparrow & \uparrow \\ \mathbf{E}_1 & \mathbf{E}_2 \end{matrix}$

Now the associated diagonal system is

$$\begin{aligned}\frac{dw_1}{dt} &= w_1, \\ \frac{dw_2}{dt} &= -w_2,\end{aligned}$$

and has as solutions the functions

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} c_1 e^t \\ c_2 e^{-t} \end{pmatrix}.$$

Thus the solutions to the equation (\*) are

$$\begin{aligned}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \mathbf{P} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} c_1 e^t \\ c_2 e^{-t} \end{pmatrix} \\ &= \begin{pmatrix} c_1 e^t + c_2 e^{-t} \\ c_1 e^t - c_2 e^{-t} \end{pmatrix} \quad **.\end{aligned}$$

EXAMPLE 3. Let us solve the differential equation

$$(*) \quad \frac{d^3 z}{dt^3} + 4 \frac{d^2 z}{dt^2} + \frac{dz}{dt} - 6z = 0.$$

First of all we obtain an equivalent first-order system by setting

$$\begin{aligned}y_1 &= z, \\ y_2 &= \frac{dz}{dt}, \\ y_3 &= \frac{d^2 z}{dt^2}.\end{aligned}$$

We then have

$$\begin{aligned}\frac{dy_1}{dt} &= y_2, \\ \frac{dy_2}{dt} &= y_3,\end{aligned}$$

and the original equation now reads

$$\frac{dy_3}{dt} = -4y_3 - y_2 + 6y_1,$$

so we have the system

$$\frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -1 & -4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

The characteristic polynomial of the coefficient matrix is

$$\begin{aligned} p(\lambda) &= \det \begin{pmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 6 & -1 & -4 - \lambda \end{pmatrix} \\ &= -\lambda(\lambda(\lambda + 4) + 1) + 6(1) \\ &= -\lambda^3 - 4\lambda^2 - \lambda + 6 \\ &= -(\lambda - 1)(\lambda + 2)(\lambda + 3). \end{aligned}$$

Therefore the coefficient matrix has three distinct eigenvalues and can be diagonalized. The computation of eigenvectors proceeds as usual. We find

$$\lambda_1 = 1: \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 6 & -1 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \mathbf{E}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ is an eigenvector.}$$

$$\lambda_2 = -2: \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 6 & -1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow y = -2x, z = -2y \text{ so } \mathbf{E}_2 = \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix} \text{ is an eigenvector.}$$

$$\lambda_3 = -3: \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 6 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow y = -3x, z = -3y \text{ so } \mathbf{E}_3 = \begin{pmatrix} 1 \\ -3 \\ 9 \end{pmatrix} \text{ is an eigenvector.}$$

The matrix of the basis change is

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & -3 \\ 1 & 4 & 9 \end{pmatrix},$$

that is,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -1 & -4 \end{pmatrix} = \mathbf{P} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \mathbf{P}^{-1}.$$

Set

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

The diagonal system

$$\frac{d}{dt} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

has the solutions

$$\begin{aligned} w_1 &= c_1 e^t, \\ w_2 &= c_2 e^{-2t}, \\ w_3 &= c_3 e^{-3t}. \end{aligned}$$

If

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

then

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{P} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix},$$

and thus

$$\begin{aligned} y_1 &= c_1 e^t + c_2 e^{-2t} + c_3 e^{-3t}, \\ y_2 &= c_1 e t - 2c_2 e^{-2t} - 3c_3 e^{-3t}, \\ y_3 &= c_1 e t + 4c_2 t + 9c_3 e^{-3t}. \end{aligned}$$

Since  $z = y_1$  we obtain that

$$z = c_1 e^t + c_2 e^{-2t} + c_3 e^{-3t}$$

is the general solution to (\*). \*\*



This example illustrates that the differential equation

$$\frac{d^n z}{dt^n} + a_{n-1} \frac{d^{n-1} z}{dt} + \cdots + a_1 \frac{dz}{dt} + a_0 z = 0$$

is equivalent to the linear system

$$\frac{d}{dt} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & & & & 1 \\ -a_0 & \cdots & a_{n-2} & & -a_{n-1} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

To apply the preceding solution method we need to know when a matrix of the form

$$\begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & 1 \\ -a_0 & \cdots & & -a_{n-1} \end{pmatrix} = \mathbf{A}$$

is diagonalizable. Matrices of this form are called *companion matrices*.

**Proposition 18.3.** *The characteristic polynomial of the matrix*

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & 0 & 1 \\ b_0 & \cdots & b_{n-2} & b_{n-1} \end{pmatrix}$$

is

$$p(t) = (-1)^n (t^n - b_{n-1}t^{n-1} - \cdots - b_1t - b_0).$$

**PROOF.** For  $n = 2$  we have

$$\begin{aligned} p(t) &= \det \begin{pmatrix} -t & 1 \\ b_0 & b_1 - t \end{pmatrix} = -t(b_1 - t) - b_0 \\ &= t^2 - b_1t - b_0, \end{aligned}$$

so the proposition is true for  $n = 2$ . Suppose that we knew that the proposition were true for all companion matrices of size  $n - 1$  and that  $\mathbf{B}$  is a companion

matrix of size  $n$ . Then using Lagrange expansion with respect to the first column gives

$$\begin{aligned} p(t) &= \det(\mathbf{B} - t\mathbf{I}) = \det \begin{pmatrix} -t & 1 & 0 & \cdots & 0 \\ 0 & -t & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -t & 1 \\ b_0 & \cdots & \cdots & \cdots & b_{n-1} - t \end{pmatrix} \\ &= -t \det \begin{pmatrix} -t & 1 & 0 & \cdots & 0 \\ 0 & -t & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1 & b_2 & \cdots & b_{n-1} - t \end{pmatrix} + (-1)^{n+1} b_0 \det \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}. \end{aligned}$$

Now notice that

$$\det \begin{pmatrix} -t & 1 & 0 & \cdots & 0 \\ 0 & -t & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -t & \cdots & 1 \\ b_1 & \cdots & b_{n-2} & b_{n-1} & -t \end{pmatrix}$$

is the characteristic polynomial of the companion matrix

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ b_1 & \cdots & \cdots & b_{n-1} & 0 \end{pmatrix}$$

of size  $n - 1$ . By our inductive assumption we know its characteristic polynomial, and substituting this and also that the determinant of the identity matrix is 1 in the preceding equation gives

$$\begin{aligned} p(t) &= -t((-1)^{n-1}(t^{n-1} - b_{n-1}t^{n-1} - \cdots - b_1)) + (-1)^{n+1}b_0 \\ &= (-1)^n[t^n - b_{n-1}t^{n-1} - \cdots - b_1t - b_0], \end{aligned}$$

establishing the proposition by induction.  $\square$

**Theorem 18.4.** Consider the differential equation

$$(*) \quad \frac{d^n z}{dt^n} + a_{n-1} \frac{d^{n-1} z}{dt^{n-1}} + \cdots + a_1 \frac{dz}{dt} + a_0 z = 0.$$

If the polynomial

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

has  $n$  distinct real roots,  $\lambda_1, \dots, \lambda_n$ , then the general solution of (\*) is

$$z = c_1 e^{\lambda_1 t} + \dots + c_n e^{\lambda_n t}.$$

PROOF. Set

$$\begin{aligned} y_1 &= z, \\ y_2 &= \frac{dz}{dt}, \\ &\vdots \\ y_n &= \frac{d^n z}{dt^n}. \end{aligned}$$

Then (\*) is equivalent to the linear differential system

$$\frac{d}{dt} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & & \\ 0 & \dots & 0 & 1 \\ -a_0 & \dots & -a_{n-1} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{A} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

The characteristic polynomial of the coefficient matrix is (by (18.3))

$$p(\lambda) = \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_0.$$

By hypothesis this has  $n$  distinct roots  $\lambda_1, \dots, \lambda_n$ , so the coefficient matrix  $\mathbf{A}$  is diagonalizable, say

$$\mathbf{A} = \mathbf{P} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \mathbf{P}^{-1}.$$

The corresponding diagonal system

$$\frac{d}{dt} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

has as solution

$$\begin{aligned} w_1 &= c_1 e^{\lambda_1 t}, \\ &\vdots \\ w_n &= c_n e^{\lambda_n t}, \end{aligned}$$

where  $c_1, \dots, c_n$  are arbitrary constants, and

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

So, in particular

$$z(t) = y_1(t) = p_{11}c_1e^{\lambda_1 t} + \cdots + p_{1n}c_ne^{\lambda_n t}.$$

It only remains to show that none of the entries  $p_{1j}$  is zero, so that none of the constants  $p_{1j}c_j$  vanish, and thus remain arbitrary. So suppose  $p_{1j} = 0$ . Then from the equation

$$y_2 = \frac{dy_1}{dt}$$

we get by substitution

$$\begin{aligned} p_{21}c_1e^{\lambda_1 t} + \cdots + p_{2j}c_je^{\lambda_j t} + \cdots + p_{2n}c_ne^{\lambda_n t} \\ = p_{11}c_1e^{\lambda_1 t} + \cdots + pc_je^{\lambda_j t} + \cdots + p_{1n}c_ne^{\lambda_n t}, \end{aligned}$$

for all values of the constants  $c_1, \dots, c_n$ . In particular, for  $c_1 = \cdots = c_{j-1} = c_{j+1} = \cdots = c_n = 0$  and  $c_j = 1$ . This gives

$$p_{2j}e^{\lambda_j t} = 0$$

so  $p_{2j} = 0$ . Continuing in this way we obtain in succession  $p_{3j} = 0, \dots, p_{nj} = 0$ . But this says that  $\mathbf{P}$  has a column of zeros. So the determinant of  $\mathbf{P}$  is zero, contrary to the fact that  $\mathbf{P}$  is invertible.  $\square$

Let us try another example.

**EXAMPLE 4.** Solve the differential equation

$$\frac{d^3 z}{dt^3} - \frac{dz}{dt} = 0.$$

*Solution.* The associated polynomial is

$$p(\lambda) = \lambda^3 - \lambda = \lambda(\lambda^2 - 1) = \lambda(\lambda - 1)(\lambda + 1).$$

So by (18.4) the general solution we seek is (remember  $e^0 = 1$ )

$$z = c_1 e^{0t} + c_2 e^t + c_3 e^{-t} = c_1 + c_2 e^t + c_3 e^{-t}. \quad **$$

Up until now we have only considered linear differential systems

$$\frac{dy}{dt} = \mathbf{A}y,$$

where the coefficient matrix  $\mathbf{A}$  is diagonalizable. We know, of course, that not all matrices are diagonalizable. What happens then? Here is where the Jordan canonical form comes into play! Let us begin by looking at a simple example.

EXAMPLE 5. Consider the differential system

$$\begin{pmatrix} \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 1 & a \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

where  $a$  is a fixed real number. The characteristic polynomial of the coefficient matrix is

$$p(\lambda) = (a - \lambda)^2$$

so  $a$  is a double root, and the coefficient matrix cannot be diagonalized. (Think this through!) On the other hand, we can easily solve this differential system. We have

$$\frac{dy_1}{dt} = ay_1 \Rightarrow y_1 = c_1 e^{at}.$$

If we put this into the second equation we find

$$\frac{dy_2}{dt} = c_1 e^{at} + y_2.$$

Consider  $y_2 e^{-at}$ , and compute its derivative by the product rule. We find

$$\begin{aligned} \frac{d}{dt}(y_2 e^{-at}) &= -y_2 e^{-at} + (c_1 e^{at} + y_2) e^{-at} \\ &= -y_2 e^{-at} + c_1 + y_2 e^{-at} \\ &= c_1. \end{aligned}$$

That is,  $y_2 e^{-at}$  has a constant derivative. Therefore by the fundamental theorem of the calculus

$$y_2 e^{-at} = c_1 t + c_2,$$

or

$$y_2 = c_1 t e^{-at} + c_2 e^{-at},$$

where  $c_2$  is a second arbitrary constant. So the general solution is

$$\begin{aligned} y_1 &= c_1 e^{at}, \\ y_2 &= c_1 t e^{at} + c_2 e^{at}. \quad ** \end{aligned}$$

Here is another example.

EXAMPLE 6. We want to solve the equation

$$\frac{d^2 z}{dt^2} + \frac{2 dz}{dt} + z = 0.$$

The equivalent linear differential system is

$$\begin{pmatrix} \frac{dy_1}{dt} \\ \frac{dy_2}{dt} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The characteristic polynomial of the coefficient matrix  $\mathbf{A}$  is

$$p(\lambda) = \lambda^2 + 2\lambda + 1 = (\lambda + 1),$$

which has  $\lambda = -1$  as a double root. The matrix  $\mathbf{A}$  is therefore not diagonalizable. (Because, if it were, the diagonal form would be scalar, and since scalar matrices commute with every matrix the equation  $\mathbf{A} = \mathbf{PDP}^{-1}$  would yield  $\mathbf{A} = \mathbf{D}$ . But  $\mathbf{A}$  is not itself diagonal.) On the other hand  $\mathbf{A}$  has Jordan canonical form

$$\mathbf{J}(\mathbf{A}) = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix},$$

and we saw how to solve

$$\frac{dw}{dt} = \mathbf{J}(\mathbf{A})w$$

In Example 5. To make use of this information we need the transformation matrix  $\mathbf{Q}^{-1}$  such that

$$\mathbf{A} = \mathbf{QJ}(\mathbf{A})\mathbf{Q}^{-1}.$$

That means we must find the Jordan basis. This proceeds as in Chapter 17. Let  $\mathbf{T}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation represented by the matrix  $\mathbf{A}$  with respect to the standard basis of  $\mathbb{R}^2$ . Let  $\mathbf{F}_1 = (u, v) \in \mathbb{R}^2$  be a vector such that

$$\mathbf{F}_2 = \mathbf{A} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Such a vector exists because  $\mathbf{A}$  is not diagonalizable. Furthermore, since  $-1$  is the only eigenvalue of  $\mathbf{A}$  it follows  $\mathbf{F}_2 \in \mathcal{L}(\mathbf{F}_1)$ . Therefore,  $\{\mathbf{F}_1, \mathbf{F}_2\}$  is a basis for  $\mathbb{R}^2$ . Moreover

$$\begin{aligned} \mathbf{T}(\mathbf{F}_1) &= \mathbf{F}_1 + \mathbf{F}_2, \\ \mathbf{T}(\mathbf{F}_2) &= (\mathbf{T} + \mathbf{I} - \mathbf{I})(\mathbf{T} + \mathbf{I})(\mathbf{F}_1) \\ &= (\mathbf{T} + \mathbf{I})^2(\mathbf{F}_1) - (\mathbf{T} + \mathbf{I})(\mathbf{F}_1) \\ &= \mathbf{0} - \mathbf{F}_2, \end{aligned}$$

because  $\mathbf{T} + \mathbf{I}$  is nilpotent of index 2. Thus  $\mathbf{J}(\mathbf{A})$  is the matrix of  $\mathbf{T}$  with respect to  $\{\mathbf{F}_1, \mathbf{F}_2\}$ . A possible choice for  $\mathbf{F}_1$  is  $(0, 1)$  whence  $\mathbf{F}_2 = (1, -1)$ . Therefore setting

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

we have

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

We now proceed as in the diagonalizable case and set

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{Q}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

so

$$\frac{d}{dt} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{J}(\mathbf{A}) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

Example 5 gives

$$w_1 = c_1 e^{-t},$$

$$w_2 = c_1 t e^{-t} + c_2 e^{-t},$$

and so,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{Q} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} c_1 e^{-t} \\ c_1 t e^{-t} + c_2 e^{-t} \end{pmatrix} \\ = \begin{pmatrix} c_1 t e^{-t} + c_2 e^{-t} \\ c_1 e^{-t} - c_1 t e^{-t} - c_2 e^{-t} \end{pmatrix},$$

whence

$$z = c_1 t e^{-t} + c_2 e^{-t},$$

where  $c_1, c_2$  are arbitrary constants. \*\*

This example leads naturally to the differential system

$$\frac{dy}{dt} = \mathbf{J}_n(a) \mathbf{Y},$$

where

$$\mathbf{J}_n(a) = \begin{pmatrix} a & & & 0 \\ 1 & & \ddots & \\ & \ddots & & \\ 0 & & 1 & a \end{pmatrix}$$

is a typical Jordan block. What does the solution to such a system look like? That is easy. We proceed as in Example 5

$$\frac{dy_1}{dt} = ay_1$$

so

$$y_1 = ce^{at},$$

$$\frac{dy_2}{dt} = y_1 + ay_2,$$

so by the product rule (see Example 5 again)

$$\frac{d}{dt}(e^{-at}y_2) = c_1.$$

This says the function  $y_2 e^{-at}$  is a constant. Therefore

$$y_2 e^{-at} = c_1 t + c_2,$$

and

$$y_2 = c_1 t e^{at} + c_2 e^{at} = (c_1 t + c_2) e^{at}.$$

Next

$$\frac{dy_3}{dt} = y_2 + ay_3$$

so by the product rule

$$\begin{aligned}\frac{d}{dt}(y_3 e^{-at}) &= y_2 e^{-at} \\ &= c_1 t + c_2.\end{aligned}$$

So by the fundamental theorem of the calculus

$$y_3 e^{-at} = \frac{c_1}{2} t^2 + c_2 t + c_3,$$

where  $c_3$  is yet another constant. This says

$$y_3 = \left( \frac{c_1}{2} t^2 + c_2 t + c_3 \right) e^{at}.$$

Continuing in this way gives

$$y_k = \frac{c_1}{(k-1)!} t^{k-1} + \frac{c_1}{(k-2)!} t^{k-2} + \cdots + c_{k-1} t + c_k$$

for  $k = 1, 2, \dots, n$ .

We may summarize this in a theorem.



**Theorem 18.5.** *The general solution to the linear system*

$$\frac{dy}{dt} = \mathbf{J}_n(a)y$$

is given by

$$y_k = \left( \frac{c_1}{(k-1)!} t^{k-1} + \cdots + c_{k-1}t + c_k \right) e^{at}, \quad k = 1, \dots, n,$$

where  $c_1, \dots, c_n$  are arbitrary constants.  $\square$

This result can now be applied to nondiagonalizable systems as the following example illustrates.

**EXAMPLE 7.** Solve the differential system

$$\frac{dy}{dt} = \begin{pmatrix} 2 & 2 & -1 \\ -1 & -1 & 1 \\ -1 & -2 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

*Solution.* The characteristic polynomial of the coefficient matrix  $\mathbf{A}$  is

$$p(\lambda) = \det \begin{pmatrix} 2 - \lambda & 2 & -1 \\ -1 & -1 - \lambda & 1 \\ -1 & -2 & 2 - \lambda \end{pmatrix} = \cdots = (\lambda - 1)^3.$$

There is therefore just one eigenvalue 1, of multiplicity 3. The Jordan normal form of  $\mathbf{A}$  is (see Example 7 in Chapter 17)

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

and a corresponding Jordan basis for  $\mathbb{R}^3$  is

$$\mathbf{F}_1 = (1, 0, 0), \quad \mathbf{F}_2 = (1, -1, 1), \quad \mathbf{F}_3 = (0, 1, 2).$$

So

$$\mathbf{A} = \mathbf{Q}\mathbf{J}(\mathbf{A})\mathbf{Q}^{-1},$$

where

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

If we set

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \mathbf{Q}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

then

$$\frac{dw}{dt} = \mathbf{J}(\mathbf{A})w = \begin{pmatrix} \mathbf{J}_2(1) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

and working block by block we find

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} c_1 e^t \\ \left(\frac{c_1}{2}t + c_2\right)e^t \end{pmatrix}$$

and

$$w_3 = c_3 e^t,$$

where  $c_1, c_2, c_3$  are arbitrary constants. Thus

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = Q \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \cdots = \begin{pmatrix} c_1 e^t + \frac{c_1}{2} t e^t + c_2 e^t \\ -\frac{c_1}{2} e^t - c_2 e^t + c_3 e^t \\ \frac{c_1}{2} t e^t + c_2 e^t + 2c_3 e^t \end{pmatrix}. \quad **$$

As a last example we consider

**EXAMPLE 8.** Solve the differential equation

$$\frac{d^3 y}{dt^3} - \frac{d^2 y}{dt^2} - \frac{dy}{dt} + y = 0.$$

*Solution.* The coefficient matrix of the associated linear system is

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

The characteristic polynomial is (by (18.3))

$$\begin{aligned} p(t) &= (-1)^3(t^3 - t^2 - t + 1) \\ &= -(t-1)^2(t+1), \end{aligned}$$

so  $\mathbf{A}$  has two distinct eigenvalues  $\lambda_1 = 1$  and  $\lambda_2 = -1$  with multiplicities 2 and 1. The Jordan form of  $\mathbf{A}$  is therefore

$$\mathbf{J}(\mathbf{A}) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

Following the procedure of Chapter 17 we find the corresponding Jordan basis for  $\mathbb{R}^3$  is

$$\mathbf{F}_1 = (0, 1, 2), \quad \mathbf{F}_2 = (1, 1, 1), \quad \mathbf{F}_3 = (1, -1, 1).$$

So if we set

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

then

$$\mathbf{A} = \mathbf{P}\mathbf{J}(\mathbf{A})\mathbf{P}^{-1}.$$

Set

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \mathbf{P}^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

where as usual

$$\begin{aligned} y_1 &= y, \\ y_2 &= \frac{dy}{dt}, \\ y_3 &= \frac{d^2y}{dt^2}. \end{aligned}$$

Our original differential equation is equivalent to the system

$$\frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{A} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

which in turn is equivalent to

$$\frac{d}{dt} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \mathbf{J}(\mathbf{A}) \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

By (18.5) the general solution of the linear system

$$\frac{d}{dt} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

is given by (work block by block)

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} c_1 e^t \\ (c_1 t + c_2) e^t \\ c_3 e^t \end{pmatrix}.$$

From the definition of  $(w_1, w_2, w_3)$  we thus find

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= \mathbf{P} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 e^t \\ (c_1 t + c_2) e^t \\ c_3 e^t \end{pmatrix} \begin{pmatrix} (c_1 t + c_2) e^t + c_3 e^{-t} \\ c_1 e^t + (c_1 t + c_2) e^t - c_3 e^{-t} \\ 2c_1 + (c_1 t + c_2) e^t + c_3 e^{-t} \end{pmatrix}, \end{aligned}$$

so we conclude

$$y = y_1 = (c_1 t + c_2) e^t + c_3 e^{-t}$$

is the general solution to the original equation (\*). \*\*

## EXERCISES

1. Among all the solutions of the differential system

$$\frac{dy_1}{dt} = -y_2,$$

$$\frac{dy_2}{dt} = -y_1,$$

find a solution that satisfies the *initial conditions*

$$y_1(0) = 2,$$

$$y_2(0) = 0.$$

(Hint: First find all solutions. Then determine for which values of the constants the initial conditions are satisfied.)

2. Solve the differential system

$$\frac{dy_1}{dt} = y_2 + y_3,$$

$$\frac{dy_2}{dt} = y_1 + y_3,$$

$$\frac{dy_3}{dt} = y_1 + y_2.$$

## 18: Applications to linear differential equations

Find a solution satisfying

$$y_1(0) = 1,$$

$$y_2(0) = 0,$$

$$y_3(0) = -1.$$

3. Solve each of the following differential equations

$$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = 0,$$

$$\frac{d^2y}{dt^2} - 3\frac{dy}{dt} + 2y = 0,$$

$$\frac{d^2y}{dt^2} - y = 0,$$

$$\frac{d^2y}{dt^2} + y = 0.$$

4. Solve each of the following linear systems

$$(a) \frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

$$(b) \frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

$$(c) \frac{d}{dt} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

5. Solve each of the following differential equations

$$(a) \frac{d^3y}{dt^3} + \frac{d^2y}{dt^2} - \frac{dy}{dt} - y = 0.$$

$$(b) \frac{d^3y}{dt^3} + 3\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + y = 0.$$

$$(c) \frac{d^3y}{dt^3} - 6\frac{d^2y}{dt^2} - 7\frac{dy}{dt} - 6y = 0.$$

In each case determine a solution satisfying the initial conditions

$$y(0) = 1,$$

$$\frac{dy}{dt} = 0,$$

$$\frac{d^2y}{dt^2} = 1.$$

- abuse of notation 20
- addition
  - of linear transformations 70
  - of matrices 113
  - of position vectors 2
  - of vectors 13
- adjoint of  $T$  276
- affine subspace 170
- algebraic multiplicity of eigenvalue 214
- angle between  $\mathbf{A}$  and  $\mathbf{B}$  244, 252
- associated homogeneous system 172
- associative law 13
  - general 18
- axiomatic method 13
- axioms of a vector space 13–14
  
- basic rules
  - of matrix operations 117
  - of vector algebra 2–3
- basis 43
  - ordered 49, 105, 130
  - orthonormal 259
  - over a subspace 310
  - standard 85, 106
  - usual 85, 106, 131
- Bessel's inequality 270
  
- Cartesian  $k$ -space 16
  - complex 19
  - (real) 16
  
- characteristic function of  $s$  37
- characteristic polynomial of  $T$  207
- closure of eigenspace 315
- cofactor 204
- cofactor matrix 204
- column of a matrix 105, 112
- column space of  $\mathbf{A}$  168
- column vectors of  $\mathbf{A}$  168
- commutative law of vector addition 13
  - generalized 18
- completely invariant 308
- component
  - of  $\mathbf{A}$  along  $\mathbf{A}_i$  255
  - of a position vector 5
  - of a vector 16
    - relative to the ordered basis 49
- coordinate transformation 13
- coordinates
  - of a point 4
  - of a vector (relative to the ordered basis) 49
- Cramer's rule 222
- criteria for Eigenvalues 207
- cyclic transformation 160
- cyclic vector for  $T$  160
  
- determinant 182, 201
  - properties of 203
- diagonal entries 119
- diagonalizable 194
- diagonalize or diagonalizing  $T$  194

## Index

- differentiation operator 74
- dimension of a vector space 48
- direct sum 306
- displacement 1
- distance 251
- dot product 244
  
- echelon form 174
- eigenspace of  $T$  198
  - closure 315
- eigenvalue 193
  - algebraic multiplicity 214
  - geometric multiplicity 214
  - of  $T$  193
- eigenvector 193
  - of  $T$  associated to  $e$  193
- endomorphisms 191
- equation of the line through  $P, Q$  5
  - of the plane 9
- Euclidean space 16
  
- finite-dimensional vector space 42
- force 1
- Fourier coefficients 255
- fully invariant 308
- fundamental theorem of algebra 216
  
- general solution 175
- geometric multiplicity 214
- Gram-Schmidt orthonormalization process 258
  
- homogeneous equation, linear 29, 170
- homogeneous quadratic function 293
  
- identity matrix 118
  - transformation 81
- index of nilpotence 120, 158
- injective 88
- infinite-dimensional vector space 47
- initial point 1
- injective 88
- inner product 249
  - trace 248, 249
- inner product space 249
  - complex 272
  - (real) 249
- intersection (of sets) 19
- invariant complement 308
- invariant subspace 305
  - fully 308
- involution 123
- involutory 123
- isometric 263
  - embedding 262
  - isomorphism 263
- isometry 263
- isomorphic 76
- isomorphism 76
  - natural 85
  
- Jordan blocks 323
- Jordan canonical form 323
  
- $k$ -tuple 16
  
- law of cosines 270
- length function 250
- length of  $A$  250
- linear
  - combination 22
  - dependence 35
  - extension 82
  - homogeneous equation 29
  - independence 35
  - over a subspace 310
  - relation 35
  - span 22
  - subspace 21
  - transformation 66
- linear equations
  - coefficients of 165
  - homogeneous system of 170
  - solution of 165
  - system of 165
- linear transformation
  - composition of 70
  - definition of 66
  - homogeneity of 78
  - image of a 72
  - injective 88
  - kernel of a 72
  - rank of a 191
  - scalar multiplication of 71
  - self-adjoint 278
  - sum of two 70
  - symmetric 280
- linear transformations
  - sum of 70
  - vector space of 71

- linearly dependent 35
  - on a set of vectors 38
  - set of vectors 35
- linearly independent 35
  - set of vectors 35
- matrix
  - augmented 176
  - coefficient 167
  - cofactor 204
  - definition of 105, 112
  - diagonal 119
  - echelon 174
  - equality of 113
  - idempotent 120
  - inverse 122
  - invertible 122
  - involutory 123
  - $i$ th row of 112
  - $j$ th column of 112
  - lower triangular 120
  - minor 201
  - $m \times n$ - 112
  - multiplication of 107-108, 114
  - nilpotent 120
  - nonsingular 122
  - scalar 118
  - skew symmetric 124
  - square (of size  $n$ ) 116
  - strictly triangular 120
  - sum of 113
  - symmetric 124
  - the identity 118
  - transpose of a 128, 203
  - triangular 120
  - zero 120
- matrix
  - equation of a linear system 167
  - of a linear transformation 130
  - of constants of a linear system 167
  - of ST 107-108, 140
  - product 114
- model 15
- multiplicity of eigenvalue 214
  - algebraic 214
  - geometric 214
- nilpotent 75, 158
  - matrix 120
  - of index  $k$  158
  - transformation 75, 158
- normalize 259
- orthogonal
  - complement 265
  - projection 156
  - reflection 273
  - set of vectors 253
- orthonormal
  - basis 259
  - set 253
- parallel (translation) of  $\mathcal{U}$  170
- Parseval's inequality 270
- pivot operation 182
- point
  - initial 1
  - terminal 1
- principal axes 293
- principal axis theorem 293
- product
  - dot 245
  - inner 245
  - scalar 245
  - scalar on  $\mathcal{P}_k(\mathbb{R})$  247
  - standard scalar on  $\mathbb{R}^n$  246
  - standard scalar on  $\mathcal{V}$  with basis 246
  - trace scalar 248, 249
- projection 68, 129
  - definition of 155
  - orthogonal 156
  - self-adjoint 156
  - skew 156
- quadratic form 293
- quadratic function 292
  - homogeneous 293
- rank of the linear transformation 191
- real vector space
  - of degrees less than or equal to  $n$  27
  - of polynomials 28
- reduced echelon form 174
- reduction to the echelon form 174-176
- reflection 129, 273
- representing a linear transformation by
  - matrix 130
- restriction of  $T$  to  $\mathcal{W}$  301
- rotation 67, 271
- row of a matrix 105, 112



## Index

- scalar 13–14
- scalar multiplication 13–14
  - of position vectors 2, 3
- scalar product 245
  - on  $\mathcal{P}_k(\mathbb{R})$  247
  - standard scalar product on  $\mathbb{R}^n$  246
  - standard scalar product on  $\mathcal{V}$  246
  - trace 248–249
- Schwarz inequality 250
- self-adjoint 278
  - projection 156
- shift 87
- simplex method 180–183
- singular endomorphism 199
- size of a matrix 112
- solution space 29–30, 172
- space
  - column 168
  - inner product 249
  - linear 13
  - vector 13
- spectral theorem 282
- standard basis 85, 106
- subspace
  - affine 170
  - linear 21
- sum
  - of linear spaces 24
  - of matrices 113
  - of position vectors 2, 5
  - of vectors 13
- superdiagonal (of a matrix) 132
- surjective = onto 111
- symmetric 124, 280
- system of linear equations 170
- terminal point 1
- trace 248
- transformation
  - cyclic 160
  - identity 81
  - injective 88
  - nilpotent 75, 158
  - zero 71
- transpose 128, 203
- triangle inequality 251
- triangular matrix 120
- union (of sets) 19
- unit vector 260
- vector
  - addition 13–14
  - algebra 2
  - bound 1
  - column (of a matrix) 168
  - equality of 1
  - free 1
  - from  $\mathbf{P}$  to  $\mathbf{Q}$  1
  - in the plane and space 1–10
  - position 2
  - quantities 1
  - same 1
  - unit 260
  - zero 13
- vector equation
  - on the line  $L$  4
  - of the plane  $\pi$  8
- vector space 13
  - complex 19, 272
  - real 19
- velocity 1
- weighted shift 122
- zero
  - matrix 120
  - transformation 71
  - vector 13

# List of Notations

$a, b, c, \dots$	as numbers	15
$A, B, \dots, P, Q, \dots, O,$	as points	1
	as vectors	15
$O$	the origin	1
$\vec{P}, \vec{E}_i$	position vector from the origin to $P, E_i$	1, 10
$\vec{PQ}$	vector from $P$ to $Q$	1
$\in$	"is an element of"	19
$\subset$	"is contained" or "is a subset of"	19
$(a, b)$	open interval	29
$[a, b]$	closed interval	29
$S \cap T$	set theoretic intersection	19
$S \cup T$	set theoretic union	19
$\Rightarrow$	"implies"	17
$ S $	number of elements in the set $A$	48
$ \mathbf{A} $	length of vector $\mathbf{A}$	250
$\langle \ , \ \rangle$	"inner product"	245
$d(\mathbf{A}, \mathbf{B})$	distance	251
$\mathbf{A} \cdot \mathbf{B}$	dot/scalar product	245
$\mathbf{A}^{-1}$	inverse matrix of $\mathbf{A}$	122
$\mathbf{A}^t$	transpose $d$ matrix of $\mathbf{A}$	128
$\mathbf{A}^{\text{cof}}$	cofactor matrix of $\mathbf{A}$	205
$\mathbf{A}^*$	adjoint of $\mathbf{A}$	276
$\mathbf{A} + \mathcal{U}$	parallel (translation) of $\mathcal{U}$	170
$A_{ij}$	cofactor	204
$a_{ij}$	the entry in the $i$ th row, $j$ th column of a matrix $A$	65
$(a_{ij})$	the matrix with entries $a_{ij}$	105
$\mathbb{C}$	the set of all complex numbers	19, 28
$\mathbb{C}^k$	the complex Cartesian space	19
$\mathcal{C}(a, b)$	continuous functions on the interval $[a, b] \subset \mathbb{R}$	29
		361

## List of Notations

$\det \mathbf{C}$	the determinant of the matrix $\mathbf{C}$	182
$D$	differentiation operator	74
$\Delta(t)$	characteristic polynomial	207
$\dim$	dimension of	48
$\mathbf{E}_i$	$(0, 0, \dots, 1, 0 \dots 0) \in \mathbb{R}^n$ ; standard basis vector	85
$\mathbf{E}_{rs}$	the matrix with a 1 in the $(r, s)$ position and zero otherwise	113
$\mathcal{F}(S)$	the set of all real valued functions defined on $S$	31
$\mathcal{F}(S, T)$	the elements of $\mathcal{F}(S)$ that vanish on $T \subseteq S$	31
$\mathcal{F}_{\mathbb{C}}(S)$	complex valued functions defined on $S$	31
$g _T$		80
$I$	the identity transformation	81
$\mathbf{I}$	the identity matrix	118
$\text{Im } T$	the image of $T$	72
$\ker T$	the kernel of $T$	72
$\mathcal{L}(\quad)$	the linear span of	22, 23
$L_{k(x)}$		97
$L$	integration operator	74
$M_{k(x)}$	multiplication operator by $k(x)$	100
$\mathcal{M}_{mn}$	the set of all $m \times n$ matrices	113
$\mathcal{P}(\mathbb{R})$	the set of all polynomials	28
$\mathcal{P}_n(\mathbb{R})$	the set of all polynomials of degree $\leq n$	27
$\mathbb{R}$	the real numbers	16
$\mathbb{R}^k$	the real Cartesian space	16
$\mathbf{R}_T$		81
$\gamma$	the solution space	170
$S$	the shift operator	87
$T_A$		275
$\text{tr}(\mathbf{A})$	trace of $\mathbf{A}$	248
$T(\mathcal{U})$		69
$T_\phi$		86
$T _{\mathcal{W}}$	restriction of $T$ to the subspace $\mathcal{W}$	301
$\mathcal{V}_e$	the eigenspace corresponding to eigenvalue $e$	198
$\mathcal{V}_{xy}$	the $xy$ plane	72
$\mathcal{W}^\perp$	the orthogonal complement of	265
$\chi_s$	the characteristic function of $s$	37
$\emptyset$	the empty set	94
$\mathcal{V}, \mathcal{W}, \mathcal{U}, \mathcal{L}, \mathcal{T}$	vector spaces	21
$T, S, \dots$	linear transformations	66
iff	if and only if	