LINEAR ALGEBRA

Second Edition

KENNETH HOFFMAN
Professor of Mathematics
Massachusetts Institute of Technology

RAY KUNZE
Professor of Mathematics
University of California, Irvine

PRENTICE-HALL, INC., Englewood Cliffs, New Jersey

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

Current printing (last digit):

10 9 8 7 6

Library of Congress Catalog Card No. 75-142120
Printed in the United States of America

Preface

Our original purpose in writing this book was to provide a text for the undergraduate linear algebra course at the Massachusetts Institute of Technology. This course was designed for mathematics majors at the junior level, although three-fourths of the students were drawn from other scientific and technological disciplines and ranged from freshmen through graduate students. This description of the M.I.T. audience for the text remains generally accurate today. The ten years since the first edition have seen the proliferation of linear algebra courses throughout the country and have afforded one of the authors the opportunity to teach the basic material to a variety of groups at Brandeis University, Washington University (St. Louis), and the University of California (Irvine).

Our principal aim in revising *Linear Algebra* has been to increase the variety of courses which can easily be taught from it. On one hand, we have structured the chapters, especially the more difficult ones, so that there are several natural stopping points along the way, allowing the instructor in a one-quarter or one-semester course to exercise a considerable amount of choice in the subject matter. On the other hand, we have increased the amount of material in the text, so that it can be used for a rather comprehensive one-year course in linear algebra and even as a reference book for mathematicians.

The major changes have been in our treatments of canonical forms and inner product spaces. In Chapter 6 we no longer begin with the general spatial theory which underlies the theory of canonical forms. We first handle characteristic values in relation to triangulation and diagonalization theorems and then build our way up to the general theory. We have split Chapter 8 so that the basic material on inner product spaces and unitary diagonalization is followed by a Chapter 9 which treats sesqui-linear forms and the more sophisticated properties of normal operators, including normal operators on real inner product spaces.

We have also made a number of small changes and improvements from the first edition. But the basic philosophy behind the text is unchanged.

We have made no particular concession to the fact that the majority of the students may not be primarily interested in mathematics. For we believe a mathematics course should not give science, engineering, or social science students a hodgepodge of techniques, but should provide them with an understanding of basic mathematical concepts.

On the other hand, we have been keenly aware of the wide range of backgrounds which the students may possess and, in particular, of the fact that the students have had very little experience with abstract mathematical reasoning. For this reason, we have avoided the introduction of too many abstract ideas at the very beginning of the book. In addition, we have included an Appendix which presents such basic ideas as set, function, and equivalence relation. We have found it most profitable not to dwell on these ideas independently, but to advise the students to read the Appendix when these ideas arise.

Throughout the book we have included a great variety of examples of the important concepts which occur. The study of such examples is of fundamental importance and tends to minimize the number of students who can repeat definition, theorem, proof in logical order without grasping the meaning of the abstract concepts. The book also contains a wide variety of graded exercises (about six hundred), ranging from routine applications to ones which will extend the very best students. These exercises are intended to be an important part of the text.

Chapter 1 deals with systems of linear equations and their solution by means of elementary row operations on matrices. It has been our practice to spend about six lectures on this material. It provides the student with some picture of the origins of linear algebra and with the computational technique necessary to understand examples of the more abstract ideas occurring in the later chapters. Chapter 2 deals with vector spaces, subspaces, bases, and dimension. Chapter 3 treats linear transformations, their algebra, their representation by matrices, as well as isomorphism, linear functionals, and dual spaces. Chapter 4 defines the algebra of polynomials over a field, the ideals in that algebra, and the prime factorization of a polynomial. It also deals with roots, Taylor's formula, and the Lagrange interpolation formula. Chapter 5 develops determinants of square matrices, the determinant being viewed as an alternating n-linear function of the rows of a matrix, and then proceeds to multilinear functions on modules as well as the Grassman ring. The material on modules places the concept of determinant in a wider and more comprehensive setting than is usually found in elementary textbooks. Chapters 6 and 7 contain a discussion of the concepts which are basic to the analysis of a single linear transformation on a finite-dimensional vector space; the analysis of characteristic (eigen) values, triangulable and diagonalizable transformations; the concepts of the diagonalizable and nilpotent parts of a more general transformation, and the rational and Jordan canonical forms. The primary and cyclic decomposition theorems play a central role, the latter being arrived at through the study of admissible subspaces. Chapter 7 includes a discussion of matrices over a polynomial domain, the computation of invariant factors and elementary divisors of a matrix, and the development of the Smith canonical form. The chapter ends with a discussion of semi-simple operators, to round out the analysis of a single operator. Chapter 8 treats finite-dimensional inner product spaces in some detail. It covers the basic geometry, relating orthogonalization to the idea of 'best approximation to a vector' and leading to the concepts of the orthogonal projection of a vector onto a subspace and the orthogonal complement of a subspace. The chapter treats unitary operators and culminates in the diagonalization of self-adjoint and normal operators. Chapter 9 introduces sesqui-linear forms, relates them to positive and self-adjoint operators on an inner product space, moves on to the spectral theory of normal operators and then to more sophisticated results concerning normal operators on real or complex inner product spaces. Chapter 10 discusses bilinear forms, emphasizing canonical forms for symmetric and skew-symmetric forms, as well as groups preserving non-degenerate forms, especially the orthogonal, unitary, pseudo-orthogonal and Lorentz groups.

We feel that any course which uses this text should cover Chapters 1, 2, and 3 thoroughly, possibly excluding Sections 3.6 and 3.7 which deal with the double dual and the transpose of a linear transformation. Chapters 4 and 5, on polynomials and determinants, may be treated with varying degrees of thoroughness. In fact, polynomial ideals and basic properties of determinants may be covered quite sketchily without serious damage to the flow of the logic in the text; however, our inclination is to deal with these chapters carefully (except the results on modules), because the material illustrates so well the basic ideas of linear algebra. An elementary course may now be concluded nicely with the first four sections of Chapter 6, together with (the new) Chapter 8. If the rational and Jordan forms are to be included, a more extensive coverage of Chapter 6 is necessary.

Our indebtedness remains to those who contributed to the first edition, especially to Professors Harry Furstenberg, Louis Howard, Daniel Kan, Edward Thorp, to Mrs. Judith Bowers, Mrs. Betty Ann (Sargent) Rose and Miss Phyllis Ruby. In addition, we would like to thank the many students and colleagues whose perceptive comments led to this revision, and the staff of Prentice-Hall for their patience in dealing with two authors caught in the throes of academic administration. Lastly, special thanks are due to Mrs. Sophia Koulouras for both her skill and her tireless efforts in typing the revised manuscript.

K. M. H. / R. A. K.

Contents

-1. With the usual operations of addition and multiplication, the set of integers satisfies all of the conditions (1)–(9) except condition (8).

EXAMPLE 3. The set of **rational numbers,** that is, numbers of the form p/q, where p and q are integers and $q \neq 0$, is a subfield of the field of complex numbers. The division which is not possible within the set of integers is possible within the set of rational numbers. The interested reader should verify that any subfield of C must contain every rational number.

EXAMPLE 4. The set of all complex numbers of the form $x + y\sqrt{2}$, where x and y are rational, is a subfield of C. We leave it to the reader to verify this.

In the examples and exercises of this book, the reader should assume that the field involved is a subfield of the complex numbers, unless it is expressly stated that the field is more general. We do not want to dwell on this point; however, we should indicate why we adopt such a convention. If F is a field, it may be possible to add the unit 1 to itself a finite number of times and obtain 0 (see Exercise 5 following Section 1.2):

$$1 + 1 + \cdots + 1 = 0.$$

That does not happen in the complex number field (or in any subfield thereof). If it does happen in F, then the least n such that the sum of n 1's is 0 is called the **characteristic** of the field F. If it does not happen in F, then (for some strange reason) F is called a field of **characteristic zero.** Often, when we assume F is a subfield of C, what we want to guarantee is that F is a field of characteristic zero; but, in a first exposure to linear algebra, it is usually better not to worry too much about characteristics of fields.

1.2. Systems of Linear Equations

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

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

where $y_1, \ldots, y_m$ and A_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, are given elements of F. We call (1-1) a **system of m linear equations in n unknowns.** Any n-tuple $(x_1, \ldots, x_n)$ of elements of F which satisfies each of the equations in (1-1) is called a **solution** of the system. If $y_1 = y_2 = \cdots = y_m = 0$, we say that the system is **homogeneous,** or that each of the equations is homogeneous.

Perhaps the most fundamental technique for finding the solutions of a system of linear equations is the technique of elimination. We can illustrate this technique on the homogeneous system

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0. \end{aligned}$$

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

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

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

$$7x_1 + 7x_3 = 0$$

or, $x_1 = -x_3$. So we conclude that if (x_1, x_2, x_3) is a solution then $x_1 = x_2 = -x_3$. Conversely, one can readily verify that any such triple is a solution. Thus the set of solutions consists of all triples $(-a, -a, a)$.

We found the solutions to this system of equations by 'eliminating unknowns,' that is, by multiplying equations by scalars and then adding to produce equations in which some of the x_j were not present. We wish to formalize this process slightly so that we may understand why it works, and so that we may carry out the computations necessary to solve a system in an organized manner.

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

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

Such an equation we shall call a **linear combination** of the equations in (1-1). Evidently, any solution of the entire system of equations (1-1) will also be a solution of this new equation. This is the fundamental idea of the elimination process. If we have another system of linear equations

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

in which each of the k equations is a linear combination of the equations in (1-1), then every solution of (1-1) is a solution of this new system. Of course it may happen that some solutions of (1-2) are not solutions of (1-1). This clearly does not happen if each equation in the original system is a linear combination of the equations in the new system. Let us say that two systems of linear equations are **equivalent** if each equation in each system is a linear combination of the equations in the other system. We can then formally state our observations as follows.

Theorem 1. *Equivalent systems of linear equations have exactly the same solutions.*

If the elimination process is to be effective in finding the solutions of a system like (1-1), then one must see how, by forming linear combinations of the given equations, to produce an equivalent system of equations which is easier to solve. In the next section we shall discuss one method of doing this.

Exercises

1. Verify that the set of complex numbers described in Example 4 is a subfield of C.

2. Let F be the field of complex numbers. Are the following two systems of linear equations equivalent? If so, express each equation in each system as a linear combination of the equations in the other system.

$$\begin{aligned} x_1 - x_2 &= 0 \qquad & 3x_1 + x_2 &= 0 \\ 2x_1 + x_2 &= 0 \qquad & x_1 + x_2 &= 0 \end{aligned}$$

3. Test the following systems of equations as in Exercise 2.

$$\begin{aligned} -x_1 + x_2 + 4x_3 &= 0 \qquad & x_1 \qquad - x_3 &= 0 \\ x_1 + 3x_2 + 8x_3 &= 0 \qquad & x_2 + 3x_3 &= 0 \\ \tfrac{1}{2}x_1 + x_2 + \tfrac{5}{2}x_3 &= 0 \end{aligned}$$

4. Test the following systems as in Exercise 2.

$$\begin{aligned} 2x_1 + (-1 + i)x_2 \qquad + x_4 &= 0 \qquad & \left(1 + \frac{i}{2}\right)x_1 + 8x_2 - ix_3 - x_4 &= 0 \\ 3x_2 - 2ix_3 + 5x_4 &= 0 \qquad & \tfrac{2}{3}x_1 - \tfrac{1}{2}x_2 + x_3 + 7x_4 &= 0 \end{aligned}$$

5. Let F be a set which contains exactly two elements, 0 and 1. Define an addition and multiplication by the tables:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Verify that the set F, together with these two operations, is a field.

6. Prove that if two homogeneous systems of linear equations in two unknowns have the same solutions, then they are equivalent.

7. Prove that each subfield of the field of complex numbers contains every rational number.

8. Prove that each field of characteristic zero contains a copy of the rational number field.

1.3. Matrices and Elementary Row Operations

One cannot fail to notice that in forming linear combinations of linear equations there is no need to continue writing the 'unknowns' $x_1, \ldots, x_n$, since one actually computes only with the coefficients A_{ij} and the scalars y_i. We shall now abbreviate the system (1-1) by

$$AX = Y$$

where

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

We call A the **matrix of coefficients** of the system. Strictly speaking, the rectangular array displayed above is not a matrix, but is a representation of a matrix. An $m \times n$ **matrix over the field** F is a function A from the set of pairs of integers (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, into the field F. The **entries** of the matrix A are the scalars $A(i, j) = A_{ij}$, and quite often it is most convenient to describe the matrix by displaying its entries in a rectangular array having m rows and n columns, as above. Thus X (above) is, or defines, an $n \times 1$ matrix and Y is an $m \times 1$ matrix. For the time being, $AX = Y$ is nothing more than a shorthand notation for our system of linear equations. Later, when we have defined a multiplication for matrices, it will mean that Y is the product of A and X.

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

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

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

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

In defining $e(A)$, it is not really important how many columns A has, but the number of rows of A is crucial. For example, one must worry a little to decide what is meant by interchanging rows 5 and 6 of a 5×5 matrix. To avoid any such complications, we shall agree that an elementary row operation e is defined on the class of all $m \times n$ matrices over F, for some fixed m but any n. In other words, a particular e is defined on the class of all m-rowed matrices over F.

One reason that we restrict ourselves to these three simple types of row operations is that, having performed such an operation e on a matrix A, we can recapture A by performing a similar operation on $e(A)$.

Theorem 2. *To each elementary row operation* e *there corresponds an elementary row operation* e_1*, of the same type as* e*, such that* $e_1(e(A)) = e(e_1(A)) = A$ *for each* A*. In other words, the inverse operation (function) of an elementary row operation exists and is an elementary row operation of the same type.*

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

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

Using Theorem 2, the reader should find it easy to verify the following. Each matrix is row-equivalent to itself; if B is row-equivalent to A, then A is row-equivalent to B; if B is row-equivalent to A and C is row-equivalent to B, then C is row-equivalent to A. In other words, row-equivalence is an equivalence relation (see Appendix).

Theorem 3. *If* A *and* B *are row-equivalent* $m \times n$ *matrices, the homogeneous systems of linear equations* $AX = 0$ *and* $BX = 0$ *have exactly the same solutions.*

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

$$A = A_0 \rightarrow A_1 \rightarrow \cdots \rightarrow A_k = B.$$

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

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

Example 5. Suppose F is the field of rational numbers, and

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

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

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

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

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

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

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

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

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 \qquad\quad - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

and

$$\begin{aligned} x_3 - \tfrac{11}{3}x_4 &= 0 \\ x_1 \qquad\quad + \tfrac{17}{3}x_4 &= 0 \\ x_2 \qquad - \tfrac{5}{3}x_4 &= 0 \end{aligned}$$

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

EXAMPLE 6. Suppose F is the field of complex numbers and

$$A = \begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix}.$$

In performing row operations it is often convenient to combine several operations of type (2). With this in mind

$$\begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Thus the system of equations

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

has only the trivial solution $x_1 = x_2 = 0$.

In Examples 5 and 6 we were obviously not performing row operations at random. Our choice of row operations was motivated by a desire to simplify the coefficient matrix in a manner analogous to 'eliminating unknowns' in the system of linear equations. Let us now make a formal definition of the type of matrix at which we were attempting to arrive.

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

(a) *the first non-zero entry in each non-zero row of* R *is equal to* 1;

(b) *each column of* R *which contains the leading non-zero entry of some row has all its other entries* 0.

EXAMPLE 7. One example of a row-reduced matrix is the $n \times n$ (square) **identity matrix** I. This is the $n \times n$ matrix defined by

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

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

In Examples 5 and 6, the final matrices in the sequences exhibited there are row-reduced matrices. Two examples of matrices which are *not* row-reduced are:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The second matrix fails to satisfy condition (a), because the leading non-zero entry of the first row is not 1. The first matrix does satisfy condition (a), but fails to satisfy condition (b) in column 3.

We shall now prove that we can pass from any given matrix to a row-reduced matrix, by means of a finite number of elementary row operations. In combination with Theorem 3, this will provide us with an effective tool for solving systems of linear equations.

Theorem 4. *Every* m $\times$ n *matrix over the field* F *is row-equivalent to a row-reduced matrix.*

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

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

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

Exercises

1. Find all solutions to the system of equations

$$\begin{aligned}(1 - i)x_1 - ix_2 &= 0\\ 2x_1 + (1 - i)x_2 &= 0.\end{aligned}$$

2. If

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

find all solutions of $AX = 0$ by row-reducing A.

3. If

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

find all solutions of $AX = 2X$ and all solutions of $AX = 3X$. (The symbol cX denotes the matrix each entry of which is c times the corresponding entry of X.)

4. Find a row-reduced matrix which is row-equivalent to

$$A = \begin{bmatrix} i & -(1+i) & 0 \\ 1 & -2 & 1 \\ 1 & 2i & -1 \end{bmatrix}.$$

5. Prove that the following two matrices are *not* row-equivalent:

$$\begin{bmatrix} 2 & 0 & 0 \\ a & -1 & 0 \\ b & c & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & -1 \\ 1 & 3 & 5 \end{bmatrix}.$$

6. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a 2×2 matrix with complex entries. Suppose that A is row-reduced and also that $a + b + c + d = 0$. Prove that there are exactly three such matrices.

7. Prove that the interchange of two rows of a matrix can be accomplished by a finite sequence of elementary row operations of the other two types.

8. Consider the system of equations $AX = 0$ where

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is a 2×2 matrix over the field F. Prove the following.

(a) If every entry of A is 0, then every pair (x_1, x_2) is a solution of $AX = 0$.

(b) If $ad - bc \neq 0$, the system $AX = 0$ has only the trivial solution $x_1 = x_2 = 0$.

(c) If $ad - bc = 0$ and some entry of A is different from 0, then there is a solution (x_1^0, x_2^0) such that (x_1, x_2) is a solution if and only if there is a scalar y such that $x_1 = yx_1^0$, $x_2 = yx_2^0$.

1.4. Row-Reduced Echelon Matrices

Until now, our work with systems of linear equations was motivated by an attempt to find the solutions of such a system. In Section 1.3 we established a standardized technique for finding these solutions. We wish now to acquire some information which is slightly more theoretical, and for that purpose it is convenient to go a little beyond row-reduced matrices.

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

(a) R *is row-reduced;*

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

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

One can also describe an $m \times n$ row-reduced echelon matrix R as follows. Either every entry in R is 0, or there exists a positive integer r, $1 \leq r \leq m$, and r positive integers $k_1, \ldots, k_r$ with $1 \leq k_i \leq n$ and

(a) $R_{ij} = 0$ for $i > r$, and $R_{ij} = 0$ if $j < k_i$.

(b) $R_{ik_j} = \delta_{ij}$, $1 \leq i \leq r$, $1 \leq j \leq r$.

(c) $k_1 < \cdots < k_r$.

EXAMPLE 8. Two examples of row-reduced echelon matrices are the $n \times n$ identity matrix, and the $m \times n$ **zero matrix** $0^{m,n}$, in which all entries are 0. The reader should have no difficulty in making other examples, but we should like to give one non-trivial one:

$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Theorem 5. *Every* m × n *matrix* A *is row-equivalent to a row-reduced echelon matrix.*

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

In Examples 5 and 6, we saw the significance of row-reduced matrices in solving homogeneous systems of linear equations. Let us now discuss briefly the system $RX = 0$, when R is a row-reduced echelon matrix. Let rows $1, \ldots, r$ be the non-zero rows of R, and suppose that the leading non-zero entry of row i occurs in column k_i. The system $RX = 0$ then consists of r non-trivial equations. Also the unknown x_{k_i} will occur (with non-zero coefficient) only in the ith equation. If we let $u_1, \ldots, u_{n-r}$ denote the $(n - r)$ unknowns which are different from $x_{k_1}, \ldots, x_{k_r}$, then the r non-trivial equations in $RX = 0$ are of the form

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

All the solutions to the system of equations $RX = 0$ are obtained by assigning any values whatsoever to $u_1, \ldots, u_{n-r}$ and then computing the corresponding values of $x_{k_1}, \ldots, x_{k_r}$ from (1-3). For example, if R is the matrix displayed in Example 8, then $r = 2$, $k_1 = 2$, $k_2 = 4$, and the two non-trivial equations in the system $RX = 0$ are

$$\begin{aligned} x_2 - 3x_3 \quad + \tfrac{1}{2}x_5 &= 0 \quad \text{or} \quad x_2 = 3x_3 - \tfrac{1}{2}x_5 \\ x_4 + 2x_5 &= 0 \quad \text{or} \quad x_4 = -2x_5. \end{aligned}$$

So we may assign any values to x_1, x_3, and x_5, say $x_1 = a$, $x_3 = b$, $x_5 = c$, and obtain the solution $(a, 3b - \frac{1}{2}c, b, -2c, c)$.

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

Theorem 6. *If A is an m $\times$ n matrix and m $<$ n, then the homogeneous system of linear equations AX $=$ 0 has a non-trivial solution.*

Proof. Let R be a row-reduced echelon matrix which is row-equivalent to A. Then the systems $AX = 0$ and $RX = 0$ have the same solutions by Theorem 3. If r is the number of non-zero rows in R, then certainly $r \leq m$, and since $m < n$, we have $r < n$. It follows immediately from our remarks above that $AX = 0$ has a non-trivial solution. ∎

Theorem 7. *If A is an n $\times$ n (square) matrix, then A is row-equivalent to the n $\times$ n identity matrix if and only if the system of equations AX $=$ 0 has only the trivial solution.*

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

Let us now ask what elementary row operations do toward solving a system of linear equations $AX = Y$ which is not homogeneous. At the outset, one must observe one basic difference between this and the homogeneous case, namely, that while the homogeneous system always has the trivial solution $x_1 = \cdots = x_n = 0$, an inhomogeneous system need have no solution at all.

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

$$\begin{aligned} A'_{ij} &= A_{ij}, \quad \text{if} \quad j \leq n \\ A'_{i(n+1)} &= y_i. \end{aligned}$$

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

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

which results from applying the sequence of row operations to the matrix Y. It should be clear to the reader that, just as in the proof of Theorem 3, the systems $AX = Y$ and $RX = Z$ are equivalent and hence have the same solutions. It is very easy to determine whether the system $RX = Z$ has any solutions and to determine all the solutions if any exist. For, if R has r non-zero rows, with the leading non-zero entry of row i occurring in column k_i, $i = 1, \ldots, r$, then the first r equations of $RX = Z$ effectively express $x_{k_1}, \ldots, x_{k_r}$ in terms of the $(n - r)$ remaining x_j and the scalars $z_1, \ldots, z_r$. The last $(m - r)$ equations are

$$\begin{aligned} 0 &= z_{r+1} \\ \vdots &\qquad \vdots \\ 0 &= z_m \end{aligned}$$

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

EXAMPLE 9. Let F be the field of rational numbers and

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

and suppose that we wish to solve the system $AX = Y$ for some y_1, y_2, and y_3. Let us perform a sequence of row operations on the augmented matrix A' which row-reduces A:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}.$$

The condition that the system $AX = Y$ have a solution is thus

$$2y_1 - y_2 + y_3 = 0$$

and if the given scalars y_i satisfy this condition, all solutions are obtained by assigning a value c to x_3 and then computing

$$\begin{aligned} x_1 &= -\tfrac{3}{5}c + \tfrac{1}{5}(y_1 + 2y_2) \\ x_2 &= \tfrac{1}{5}c + \tfrac{1}{5}(y_2 - 2y_1). \end{aligned}$$

Let us observe one final thing about the system $AX = Y$. Suppose the entries of the matrix A and the scalars $y_1, \ldots, y_m$ happen to lie in a subfield F_1 of the field F. If the system of equations $AX = Y$ has a solution with $x_1, \ldots, x_n$ in F, it has a solution with $x_1, \ldots, x_n$ in F_1. For, over either field, the condition for the system to have a solution is that certain relations hold between $y_1, \ldots, y_m$ in F_1 (the relations $z_i = 0$ for $i > r$, above). For example, if $AX = Y$ is a system of linear equations in which the scalars y_k and A_{ij} are real numbers, and if there is a solution in which $x_1, \ldots, x_n$ are complex numbers, then there is a solution with $x_1, \ldots, x_n$ real numbers.

Exercises

1. Find all solutions to the following system of equations by row-reducing the coefficient matrix:

$$\begin{aligned} \tfrac{1}{3}x_1 + 2x_2 - 6x_3 &= 0 \\ -4x_1 + 5x_3 &= 0 \\ -3x_1 + 6x_2 - 13x_3 &= 0 \\ -\tfrac{7}{3}x_1 + 2x_2 - \tfrac{8}{3}x_3 &= 0 \end{aligned}$$

2. Find a row-reduced echelon matrix which is row-equivalent to

$$A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1 + i \end{bmatrix}.$$

What are the solutions of $AX = 0$?

3. Describe explicitly all 2×2 row-reduced echelon matrices.

4. Consider the system of equations

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

Does this system have a solution? If so, describe explicitly all solutions.

5. Give an example of a system of two linear equations in two unknowns which has no solution.

6. Show that the system

$$\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= 1 \\ x_1 + x_2 - x_3 + x_4 &= 2 \\ x_1 + 7x_2 - 5x_3 - x_4 &= 3 \end{aligned}$$

has no solution.

7. Find all solutions of

$$\begin{aligned} 2x_1 - 3x_2 - 7x_3 + 5x_4 + 2x_5 &= -2 \\ x_1 - 2x_2 - 4x_3 + 3x_4 + x_5 &= -2 \\ 2x_1 - 4x_3 + 2x_4 + x_5 &= 3 \\ x_1 - 5x_2 - 7x_3 + 6x_4 + 2x_5 &= -7. \end{aligned}$$

8. Let

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

For which triples (y_1, y_2, y_3) does the system $AX = Y$ have a solution?

9. Let

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

For which (y_1, y_2, y_3, y_4) does the system of equations $AX = Y$ have a solution?

10. Suppose R and R' are 2×3 row-reduced echelon matrices and that the systems $RX = 0$ and $R'X = 0$ have exactly the same solutions. Prove that $R = R'$.

1.5. Matrix Multiplication

It is apparent (or should be, at any rate) that the process of forming linear combinations of the rows of a matrix is a fundamental one. For this reason it is advantageous to introduce a systematic scheme for indicating just what operations are to be performed. More specifically, suppose B is an $n \times p$ matrix over a field F with rows $\beta_1, \ldots, \beta_n$ and that from B we construct a matrix C with rows $\gamma_1, \ldots, \gamma_m$ by forming certain linear combinations

$$\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \cdots + A_{in}\beta_n. \tag{1-4}$$

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

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

we see that the entries of C are given by

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

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

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

EXAMPLE 10. Here are some products of matrices with rational entries.

(a)
$$\begin{bmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_1 &= (5 \quad -1 \quad 2) = 1 \cdot (5 \quad -1 \quad 2) + 0 \cdot (15 \quad 4 \quad 8) \\ \gamma_2 &= (0 \quad 7 \quad 2) = -3(5 \quad -1 \quad 2) + 1 \cdot (15 \quad 4 \quad 8) \end{aligned}$$

(b)
$$\begin{bmatrix} 0 & 6 & 1 \\ 9 & 12 & -8 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_2 &= (9 \quad 12 \quad -8) = -2(0 \quad 6 \quad 1) + 3(3 \quad 8 \quad -2) \\ \gamma_3 &= (12 \quad 62 \quad -3) = 5(0 \quad 6 \quad 1) + 4(3 \quad 8 \quad -2) \end{aligned}$$

(c)
$$\begin{bmatrix} 8 \\ 29 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

(d)
$$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} [2 \quad 4]$$

Here

$$\gamma_2 = (6 \quad 12) = 3(2 \quad 4)$$

(e)
$$[2 \quad 4] \begin{bmatrix} -1 \\ 3 \end{bmatrix} = [10]$$

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

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

It is important to observe that the product of two matrices need not be defined; the product is defined if and only if the number of columns in the first matrix coincides with the number of rows in the second matrix. Thus it is meaningless to interchange the order of the factors in (a), (b), and (c) above. Frequently we shall write products such as AB without explicitly mentioning the sizes of the factors and in such cases it will be understood that the product is defined. From (d), (e), (f), (g) we find that even when the products AB and BA are both defined it need not be true that $AB = BA$; in other words, matrix multiplication is *not commutative.*

EXAMPLE 11.

(a) If I is the $m \times m$ identity matrix and A is an $m \times n$ matrix, $IA = A$.

(b) If I is the $n \times n$ identity matrix and A is an $m \times n$ matrix, $AI = A$.

(c) If $0^{k,m}$ is the $k \times m$ zero matrix, $0^{k,n} = 0^{k,m}A$. Similarly, $A0^{n,p} = 0^{m,p}$.

EXAMPLE 12. Let A be an $m \times n$ matrix over F. Our earlier shorthand notation, $AX = Y$, for systems of linear equations is consistent with our definition of matrix products. For if

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

with x_i in F, then AX is the $m \times 1$ matrix

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

such that $y_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n$.

The use of column matrices suggests a notation which is frequently useful. If B is an $n \times p$ matrix, the columns of B are the $1 \times n$ matrices $B_1, \ldots, B_p$ defined by

$$B_j = \begin{bmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{bmatrix}, \qquad 1 \leq j \leq p.$$

The matrix B is the succession of these columns:

$$B = [B_1, \ldots, B_p].$$

The i, j entry of the product matrix AB is formed from the ith row of A

and the jth column of B. The reader should verify that the jth column of AB is AB_j:

$$AB = [AB_1, \ldots, AB_p].$$

In spite of the fact that a product of matrices depends upon the order in which the factors are written, it is independent of the way in which they are associated, as the next theorem shows.

Theorem 8. *If* A, B, C *are matrices over the field* F *such that the products* BC *and* A(BC) *are defined, then so are the products* AB, (AB)C *and*

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

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

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

for each i, j. By definition

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

When A is an $n \times n$ (square) matrix, the product AA is defined. We shall denote this matrix by A^2. By Theorem 8, $(AA)A = A(AA)$ or $A^2A = AA^2$, so that the product AAA is unambiguously defined. This product we denote by A^3. In general, the product $AA \cdots A$ (k times) is unambiguously defined, and we shall denote this product by A^k.

Note that the relation $A(BC) = (AB)C$ implies among other things that linear combinations of linear combinations of the rows of C are again linear combinations of the rows of C.

If B is a given matrix and C is obtained from B by means of an elementary row operation, then each row of C is a linear combination of the rows of B, and hence there is a matrix A such that $AB = C$. In general there are many such matrices A, and among all such it is convenient and possible to choose one having a number of special properties. Before going into this we need to introduce a class of matrices.

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

EXAMPLE 13. A 2×2 elementary matrix is necessarily one of the following:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \quad c \neq 0, \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}, \quad c \neq 0.$$

Theorem 9. *Let* e *be an elementary row operation and let* E *be the* m $\times$ m *elementary matrix* E = e(I). *Then, for every* m $\times$ n *matrix* A,

$$\mathrm{e(A) = EA.}$$

Proof. The point of the proof is that the entry in the ith row and jth column of the product matrix EA is obtained from the ith row of E and the jth column of A. The three types of elementary row operations should be taken up separately. We shall give a detailed proof for an operation of type (ii). The other two cases are even easier to handle than this one and will be left as exercises. Suppose $r \neq s$ and e is the operation 'replacement of row r by row r plus c times row s.' Then

$$E_{ik} = \begin{cases} \delta_{ik}, & i \neq r \\ \delta_{rk} + c\delta_{sk}, & i = r. \end{cases}$$

Therefore,

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} A_{ik}, & i \neq r \\ A_{rj} + cA_{sj}, & i = r. \end{cases}$$

In other words $EA = e(A)$. ∎

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

Proof. Suppose $B = PA$ where $P = E_s \cdots E_2E_1$ and the E_i are $m \times m$ elementary matrices. Then E_1A is row-equivalent to A, and $E_2(E_1A)$ is row-equivalent to E_1A. So E_2E_1A is row-equivalent to A; and continuing in this way we see that $(E_s \cdots E_1)A$ is row-equivalent to A.

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

Exercises

1. Let

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \quad C = [1 \quad -1].$$

Compute ABC and CAB.

2. Let

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

Verify directly that $A(AB) = A^2B$.

3. Find two different 2×2 matrices A such that $A^2 = 0$ but $A \neq 0$.

4. For the matrix A of Exercise 2, find elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_k \cdots E_2E_1A = I.$$

5. Let

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

Is there a matrix C such that $CA = B$?

6. Let A be an $m \times n$ matrix and B an $n \times k$ matrix. Show that the columns of $C = AB$ are linear combinations of the columns of A. If $\alpha_1, \ldots, \alpha_n$ are the columns of A and $\gamma_1, \ldots, \gamma_k$ are the columns of C, then

$$\gamma_j = \sum_{r=1}^{n} B_{rj}\alpha_r.$$

7. Let A and B be 2×2 matrices such that $AB = I$. Prove that $BA = I$.

8. Let

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

be a 2×2 matrix. We inquire when it is possible to find 2×2 matrices A and B such that $C = AB - BA$. Prove that such matrices can be found if and only if $C_{11} + C_{22} = 0$.

1.6. Invertible Matrices

Suppose P is an $m \times m$ matrix which is a product of elementary matrices. For each $m \times n$ matrix A, the matrix $B = PA$ is row-equivalent to A; hence A is row-equivalent to B and there is a product Q of elementary matrices such that $A = QB$. In particular this is true when A is the

$m \times m$ identity matrix. In other words, there is an $m \times m$ matrix Q, which is itself a product of elementary matrices. such that $QP = I$. As we shall soon see, the existence of a Q with $QP = I$ is equivalent to the fact that P is a product of elementary matrices.

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

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

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

$$B = BI = B(AC) = (BA)C = IC = C. \quad \blacksquare$$

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

Theorem 10. *Let* A *and* B *be* $n \times n$ *matrices over* F.

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

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

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

Corollary. *A product of invertible matrices is invertible.*

Theorem 11. *An elementary matrix is invertible.*

Proof. Let E be an elementary matrix corresponding to the elementary row operation e. If e_1 is the inverse operation of e (Theorem 2) and $E_1 = e_1(I)$, then

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

and

$$E_1E = e_1(E) = e_1(e(I)) = I$$

so that E is invertible and $E_1 = E^{-1}$. $\blacksquare$

EXAMPLE 14.

(a)
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(b)
$$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -c \\ 0 & 1 \end{bmatrix}$$

(c)
$$\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}$$

(d) When $c \neq 0$,
$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} c^{-1} & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & c^{-1} \end{bmatrix}.$$

Theorem 12. *If* A *is an* n $\times$ n *matrix, the following are equivalent.*

(i) A *is invertible.*
(ii) A *is row-equivalent to the* n $\times$ n *identity matrix.*
(iii) A *is a product of elementary matrices.*

Proof. Let R be a row-reduced echelon matrix which is row-equivalent to A. By Theorem 9 (or its corollary),
$$R = E_k \cdots E_2 E_1 A$$
where $E_1, \ldots, E_k$ are elementary matrices. Each E_j is invertible, and so
$$A = E_1^{-1} \cdots E_k^{-1} R.$$
Since products of invertible matrices are invertible, we see that A is invertible if and only if R is invertible. Since R is a (square) row-reduced echelon matrix, R is invertible if and only if each row of R contains a non-zero entry, that is, if and only if $R = I$. We have now shown that A is invertible if and only if $R = I$, and if $R = I$ then $A = E_k^{-1} \cdots E_1^{-1}$. It should now be apparent that (i), (ii), and (iii) are equivalent statements about A. ∎

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

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

Theorem 13. *For an* n $\times$ n *matrix* A, *the following are equivalent.*

(i) A *is invertible.*
(ii) *The homogeneous system* AX = 0 *has only the trivial solution* X = 0.
(iii) *The system of equations* AX = Y *has a solution* X *for each* n $\times$ 1 *matrix* Y.

Proof. According to Theorem 7, condition (ii) is equivalent to the fact that A is row-equivalent to the identity matrix. By Theorem 12, (i) and (ii) are therefore equivalent. If A is invertible, the solution of $AX = Y$ is $X = A^{-1}Y$. Conversely, suppose $AX = Y$ has a solution for each given Y. Let R be a row-reduced echelon matrix which is row-equivalent to A. We wish to show that $R = I$. That amounts to showing that the last row of R is not (identically) 0. Let

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

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

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

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

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

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

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

We should like to make one final comment about the solution of linear equations. Suppose A is an $m \times n$ matrix and we wish to solve the system of equations $AX = Y$. If R is a row-reduced echelon matrix which is row-equivalent to A, then $R = PA$ where P is an $m \times m$ invertible matrix. The solutions of the system $AX = Y$ are exactly the same as the solutions of the system $RX = PY$ $(= Z)$. In practice, it is not much more difficult to find the matrix P than it is to row-reduce A to R. For, suppose we form the augmented matrix A' of the system $AX = Y$, with arbitrary scalars $y_1, \ldots, y_m$ occurring in the last column. If we then perform on A' a sequence of elementary row operations which leads from A to R, it will become evident what the matrix P is. (The reader should refer to Example 9 where we essentially carried out this process.) In particular, if A is a square matrix, this process will make it clear whether or not A is invertible and if A is invertible what the inverse P is. Since we have already given the nucleus of one example of such a computation, we shall content ourselves with a 2×2 example.

EXAMPLE 15. Suppose F is the field of rational numbers and

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

Then

$$\begin{bmatrix} 2 & -1 & y_1 \\ 1 & 3 & y_2 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 1 & 3 & y_2 \\ 2 & -1 & y_1 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 3 & y_2 \\ 0 & -7 & y_1 - 2y_2 \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 1 & 3 & y_2 \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 0 & \frac{1}{7}(y_2 + 3y_1) \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix}$$

from which it is clear that A is invertible and

$$A^{-1} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{bmatrix}.$$

It may seem cumbersome to continue writing the arbitrary scalars $y_1, y_2, \ldots$ in the computation of inverses. Some people find it less awkward to carry along two sequences of matrices, one describing the reduction of A to the identity and the other recording the effect of the same sequence of operations starting from the identity. The reader may judge for himself which is a neater form of bookkeeping.

EXAMPLE 16. Let us find the inverse of

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

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{4}{45} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{180} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{6} & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -6 & 12 & 0 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -9 & 60 & -60 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

It must have occurred to the reader that we have carried on a lengthy discussion of the rows of matrices and have said little about the columns. We focused our attention on the rows because this seemed more natural from the point of view of linear equations. Since there is obviously nothing sacred about rows, the discussion in the last sections could have been carried on using columns rather than rows. If one defines an elementary column operation and column-equivalence in a manner analogous to that of elementary row operation and row-equivalence, it is clear that each $m \times n$ matrix will be column-equivalent to a 'column-reduced echelon' matrix. Also each elementary column operation will be of the form $A \to AE$, where E is an $n \times n$ elementary matrix—and so on.

Exercises

1. Let

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

Find a row-reduced echelon matrix R which is row-equivalent to A and an invertible 3×3 matrix P such that $R = PA$.

2. Do Exercise 1, but with

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

3. For each of the two matrices

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

use elementary row operations to discover whether it is invertible, and to find the inverse in case it is.

4. Let

$$A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}.$$

For which X does there exist a scalar c such that $AX = cX$?

5. Discover whether

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

is invertible, and find A^{-1} if it exists.

6. Suppose A is a 2×1 matrix and that B is a 1×2 matrix. Prove that $C = AB$ is not invertible.

7. Let A be an $n \times n$ (square) matrix. Prove the following two statements:

(a) If A is invertible and $AB = 0$ for some $n \times n$ matrix B, then $B = 0$.

(b) If A is not invertible, then there exists an $n \times n$ matrix B such that $AB = 0$ but $B \neq 0$.

8. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Prove, using elementary row operations, that A is invertible if and only if $(ad - bc) \neq 0$.

9. An $n \times n$ matrix A is called **upper-triangular** if $A_{ij} = 0$ for $i > j$, that is, if every entry below the main diagonal is 0. Prove that an upper-triangular (square) matrix is invertible if and only if every entry on its main diagonal is different from 0.

10. Prove the following generalization of Exercise 6. If A is an $m \times n$ matrix, B is an $n \times m$ matrix and $n < m$, then AB is not invertible.

11. Let A be an $m \times n$ matrix. Show that by means of a finite number of elementary row and/or column operations one can pass from A to a matrix R which is both 'row-reduced echelon' and 'column-reduced echelon,' i.e., $R_{ij} = 0$ if $i \neq j$, $R_{ii} = 1$, $1 \leq i \leq r$, $R_{ii} = 0$ if $i > r$. Show that $R = PAQ$, where P is an invertible $m \times m$ matrix and Q is an invertible $n \times n$ matrix.

12. The result of Example 16 suggests that perhaps the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

is invertible and A^{-1} has integer entries. Can you prove that?

5. Let F be a field and let n be a positive integer ($n \geq 2$). Let V be the vector space of all $n \times n$ matrices over F. Which of the following sets of matrices A in V are subspaces of V?

(a) all invertible A;
(b) all non-invertible A;
(c) all A such that $AB = BA$, where B is some fixed matrix in V;
(d) all A such that $A^2 = A$.

6. (a) Prove that the only subspaces of R^1 are R^1 and the zero subspace.

(b) Prove that a subspace of R^2 is R^2, or the zero subspace, or consists of all scalar multiples of some fixed vector in R^2. (The last type of subspace is, intuitively, a straight line through the origin.)

(c) Can you describe the subspaces of R^3?

7. Let W_1 and W_2 be subspaces of a vector space V such that the set-theoretic union of W_1 and W_2 is also a subspace. Prove that one of the spaces W_i is contained in the other.

8. Let V be the vector space of all functions from R into R; let V_e be the subset of even functions, $f(-x) = f(x)$; let V_o be the subset of odd functions, $f(-x) = -f(x)$.

(a) Prove that V_e and V_o are subspaces of V.
(b) Prove that $V_e + V_o = V$.
(c) Prove that $V_e \cap V_o = \{0\}$.

9. Let W_1 and W_2 be subspaces of a vector space V such that $W_1 + W_2 = V$ and $W_1 \cap W_2 = \{0\}$. Prove that for each vector α in V there are *unique* vectors α_1 in W_1 and α_2 in W_2 such that $\alpha = \alpha_1 + \alpha_2$.

2.3. Bases and Dimension

We turn now to the task of assigning a dimension to certain vector spaces. Although we usually associate 'dimension' with something geometrical, we must find a suitable algebraic definition of the dimension of a vector space. This will be done through the concept of a basis for the space.

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

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

A set which is not linearly dependent is called **linearly independent.** *If the set* S *contains only finitely many vectors* $\alpha_1, \alpha_2, \ldots, \alpha_n$, *we sometimes say that* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *are dependent* (*or independent*) *instead of saying* S *is dependent* (*or independent*).

The following are easy consequences of the definition.

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

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

EXAMPLE 12. Let F be a subfield of the complex numbers. In F^3 the vectors

$$\begin{aligned}\alpha_1 &= (\ \ 3, 0, -3)\\ \alpha_2 &= (-1, 1, \ \ \ 2)\\ \alpha_3 &= (\ \ 4, 2, -2)\\ \alpha_4 &= (\ \ 2, 1, \ \ \ 1)\end{aligned}$$

are linearly dependent, since

$$2\alpha_1 + 2\alpha_2 - \alpha_3 + 0 \cdot \alpha_4 = 0.$$

The vectors

$$\begin{aligned}\epsilon_1 &= (1, 0, 0)\\ \epsilon_2 &= (0, 1, 0)\\ \epsilon_3 &= (0, 0, 1)\end{aligned}$$

are linearly independent

EXAMPLE 13. Let F be a field and in F^n let S be the subset consisting of the vectors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ defined by

$$\begin{aligned}\epsilon_1 &= (1, 0, 0, \ldots, 0)\\ \epsilon_2 &= (0, 1, 0, \ldots, 0)\\ &\cdots\cdots\\ \epsilon_n &= (0, 0, 0, \ldots, 1).\end{aligned}$$

Let $x_1, x_2, \ldots, x_n$ be scalars in F and put $\alpha = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n$. Then

$$\alpha = (x_1, x_2, \ldots, x_n). \tag{2-12}$$

This shows that $\epsilon_1, \ldots, \epsilon_n$ span F^n. Since $\alpha = 0$ if and only if $x_1 = x_2 = \cdots = x_n = 0$, the vectors $\epsilon_1, \ldots, \epsilon_n$ are linearly independent. The set $S = \{\epsilon_1, \ldots, \epsilon_n\}$ is accordingly a basis for F^n. We shall call this particular basis the **standard basis** of F^n.

EXAMPLE 14. Let P be an invertible $n \times n$ matrix with entries in the field F. Then $P_1, \ldots, P_n$, the columns of P, form a basis for the space of column matrices, $F^{n \times 1}$. We see that as follows. If X is a column matrix, then

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

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

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

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

EXAMPLE 15. Let A be an $m \times n$ matrix and let S be the solution space for the homogeneous system $AX = 0$ (Example 7). Let R be a row-reduced echelon matrix which is row-equivalent to A. Then S is also the solution space for the system $RX = 0$. If R has r non-zero rows, then the system of equations $RX = 0$ simply expresses r of the unknowns $x_1, \ldots, x_n$ in terms of the remaining $(n - r)$ unknowns x_j. Suppose that the leading non-zero entries of the non-zero rows occur in columns $k_1, \ldots, k_r$. Let J be the set consisting of the $n - r$ indices different from $k_1, \ldots, k_r$:

$$J = \{1, \ldots, n\} - \{k_1, \ldots, k_r\}.$$

The system $RX = 0$ has the form

$$\begin{array}{ccc} x_{k_1} + \sum\limits_J c_{1j}x_j & = & 0 \\ \vdots \qquad \vdots & & \vdots \\ x_{k_r} + \sum\limits_J c_{rj}x_j & = & 0 \end{array}$$

where the c_{ij} are certain scalars. All solutions are obtained by assigning (arbitrary) values to those x_j's with j in J and computing the corresponding values of $x_{k_1}, \ldots, x_{k_r}$. For each j in J, let E_j be the solution obtained by setting $x_j = 1$ and $x_i = 0$ for all other i in J. We assert that the $(n - r)$ vectors E_j, j in J, form a basis for the solution space.

Since the column matrix E_j has a 1 in row j and zeros in the rows indexed by other elements of J, the reasoning of Example 13 shows us that the set of these vectors is linearly independent. That set spans the solution space, for this reason. If the column matrix T, with entries $t_1, \ldots, t_n$, is in the solution space, the matrix

$$N = \sum_J t_jE_j$$

is also in the solution space *and* is a solution such that $x_j = t_j$ for each j in J. The solution with that property is unique; hence, $N = T$ and T is in the span of the vectors E_j.

EXAMPLE 16. We shall now give an example of an infinite basis. Let F be a subfield of the complex numbers and let V be the space of polynomial functions over F. Recall that these functions are the functions from F into F which have a rule of the form

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

Let $f_k(x) = x_k$, $k = 0, 1, 2, \ldots$. The (infinite) set $\{f_0, f_1, f_2, \ldots\}$ is a basis for V. Clearly the set spans V, because the function f (above) is

$$f = c_0f_0 + c_1f_1 + \cdots + c_nf_n.$$

The reader should see that this is virtually a repetition of the definition of polynomial function, that is, a function f from F into F is a polynomial function if and only if there exists an integer n and scalars $c_0, \ldots, c_n$ such that $f = c_0f_0 + \cdots + c_nf_n$. Why are the functions independent? To show that the set $\{f_0, f_1, f_2, \ldots\}$ is independent means to show that each finite subset of it is independent. It will suffice to show that, for each n, the set $\{f_0, \ldots, f_n\}$ is independent. Suppose that

$$c_0f_0 + \cdots + c_nf_n = 0.$$

This says that

$$c_0 + c_1x + \cdots + c_nx^n = 0$$

for every x in F; in other words, every x in F is a root of the polynomial $f(x) = c_0 + c_1x + \cdots + c_nx^n$. We assume that the reader knows that a polynomial of degree n with complex coefficients cannot have more than n distinct roots. It follows that $c_0 = c_1 = \cdots = c_n = 0$.

We have exhibited an infinite basis for V. Does that mean that V is not finite-dimensional? As a matter of fact it does; however, that is not immediate from the definition, because for all we know V might also have a finite basis. That possibility is easily eliminated. (We shall eliminate it in general in the next theorem.) Suppose that we have a finite number of polynomial functions $g_1, \ldots, g_r$. There will be a largest power of x which appears (with non-zero coefficient) in $g_1(x), \ldots, g_r(x)$. If that power is k, clearly $f_{k+1}(x) = x^{k+1}$ is not in the linear span of $g_1, \ldots, g_r$. So V is not finite-dimensional.

A final remark about this example is in order. Infinite bases have nothing to do with 'infinite linear combinations.' The reader who feels an irresistible urge to inject power series

$$\sum_{k=0}^{\infty} c_kx^k$$

into this example should study the example carefully again. If that does not effect a cure, he should consider restricting his attention to finite-dimensional spaces from now on.

Theorem 4. *Let V be a vector space which is spanned by a finite set of vectors $\beta_1, \beta_2, \ldots, \beta_m$. Then any independent set of vectors in V is finite and contains no more than m elements.*

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

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

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

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

Since $n > m$, Theorem 6 of Chapter 1 implies that there exist scalars $x_1, x_2, \ldots, x_n$ not all 0 such that

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

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

Corollary 1. *If V is a finite-dimensional vector space, then any two bases of V have the same (finite) number of elements.*

Proof. Since V is finite-dimensional, it has a finite basis

$$\{\beta_1, \beta_2, \ldots, \beta_m\}.$$

By Theorem 4 every basis of V is finite and contains no more than m elements. Thus if $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a basis, $n \leq m$. By the same argument, $m \leq n$. Hence $m = n$. ▌

This corollary allows us to define the **dimension** of a finite-dimensional vector space as the number of elements in a basis for V. We shall denote the dimension of a finite-dimensional space V by dim V. This allows us to reformulate Theorem 4 as follows.

Corollary 2. *Let V be a finite-dimensional vector space and let n = dim V. Then*

(a) *any subset of* V *which contains more than* n *vectors is linearly dependent;*

(b) *no subset of* V *which contains fewer than* n *vectors can span* V.

EXAMPLE 17. If F is a field, the dimension of F^n is n, because the standard basis for F^n contains n vectors. The matrix space $F^{m\times n}$ has dimension mn. That should be clear by analogy with the case of F^n, because the mn matrices which have a 1 in the i, j place with zeros elsewhere form a basis for $F^{m\times n}$. If A is an $m \times n$ matrix, then the solution space for A has dimension $n - r$, where r is the number of non-zero rows in a row-reduced echelon matrix which is row-equivalent to A. See Example 15.

If V is any vector space over F, the zero subspace of V is spanned by the vector 0, but $\{0\}$ is a linearly dependent set and not a basis. For this reason, we shall agree that the zero subspace has dimension 0. Alternatively, we could reach the same conclusion by arguing that the empty set is a basis for the zero subspace. The empty set spans $\{0\}$, because the intersection of all subspaces containing the empty set is $\{0\}$, and the empty set is linearly independent because it contains no vectors.

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

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

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

Then $b = 0$; for otherwise,

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

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

Theorem 5. *If* W *is a subspace of a finite-dimensional vector space* V, *every linearly independent subset of* W *is finite and is part of a (finite) basis for* W.

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

We extend S_0 to a basis for W, as follows. If S_0 spans W, then S_0 is a basis for W and we are done. If S_0 does not span W, we use the preceding lemma to find a vector β_1 in W such that the set $S_1 = S_0 \cup \{\beta_1\}$ is independent. If S_1 spans W, fine. If not, apply the lemma to obtain a vector β_2

in W such that $S_2 = S_1 \cup \{\beta_2\}$ is independent. If we continue in this way, then (in not more than dim V steps) we reach a set

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

which is a basis for W. ∎

Corollary 1. *If* W *is a proper subspace of a finite-dimensional vector space* V, *then* W *is finite-dimensional and dim* W < *dim* V.

Proof. We may suppose W contains a vector $\alpha \neq 0$. By Theorem 5 and its proof, there is a basis of W containing α which contains no more than dim V elements. Hence W is finite-dimensional, and dim $W \leq$ dim V. Since W is a proper subspace, there is a vector β in V which is not in W. Adjoining β to any basis of W, we obtain a linearly independent subset of V. Thus dim $W <$ dim V. ∎

Corollary 2. *In a finite-dimensional vector space* V *every non-empty linearly independent set of vectors is part of a basis.*

Corollary 3. *Let* A *be an* n × n *matrix over a field* F, *and suppose the row vectors of* A *form a linearly independent set of vectors in* F^n. *Then* A *is invertible.*

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the row vectors of A, and suppose W is the subspace of F^n spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. Since $\alpha_1, \alpha_2, \ldots, \alpha_n$ are linearly independent, the dimension of W is n. Corollary 1 now shows that $W = F^n$. Hence there exist scalars B_{ij} in F such that

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

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

$$BA = I. \quad ∎$$

Theorem 6. *If* W_1 *and* W_2 *are finite-dimensional subspaces of a vector space* V, *then* $W_1 + W_2$ *is finite-dimensional and*

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

Proof. By Theorem 5 and its corollaries, $W_1 \cap W_2$ has a finite basis $\{\alpha_1, \ldots, \alpha_k\}$ which is part of a basis

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

and part of a basis

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

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

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

and these vectors form an independent set. For suppose

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

Then

$$-\Sigma\, z_r\gamma_r = \Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j$$

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

$$\Sigma\, z_r\gamma_r = \Sigma\, c_i\alpha_i$$

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

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

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

$$\Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j = 0$$

and since

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

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

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

is a basis for $W_1 + W_2$. Finally

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

Let us close this section with a remark about linear independence and dependence. We defined these concepts for sets of vectors. It is useful to have them defined for finite sequences (ordered n-tuples) of vectors: $\alpha_1, \ldots, \alpha_n$. We say that the vectors $\alpha_1, \ldots, \alpha_n$ are **linearly dependent** if there exist scalars $c_1, \ldots, c_n$, not all 0, such that $c_1\alpha_1 + \cdots + c_n\alpha_n = 0$. This is all so natural that the reader may find that he has been using this terminology already. What is the difference between a finite sequence $\alpha_1, \ldots, \alpha_n$ and a set $\{\alpha_1, \ldots, \alpha_n\}$? There are two differences, identity and order.

If we discuss the set $\{\alpha_1, \ldots, \alpha_n\}$, usually it is presumed that no two of the vectors $\alpha_1, \ldots, \alpha_n$ are identical. In a sequence $\alpha_1, \ldots, \alpha_n$ all the α_i's may be the same vector. If $\alpha_i = \alpha_j$ for some $i \neq j$, then the sequence $\alpha_1, \ldots, \alpha_n$ is linearly dependent:

$$\alpha_i + (-1)\alpha_j = 0.$$

Thus, if $\alpha_1, \ldots, \alpha_n$ are linearly independent, they are distinct and we may talk about the set $\{\alpha_1, \ldots, \alpha_n\}$ and know that it has n vectors in it. So, clearly, no confusion will arise in discussing bases and dimension. The dimension of a finite-dimensional space V is the largest n such that some n-tuple of vectors in V is linearly independent—and so on. The reader who feels that this paragraph is much ado about nothing might ask himself whether the vectors

$$\alpha_1 = (e^{\pi/2}, 1)$$
$$\alpha_2 = (\sqrt[3]{110}, 1)$$

are linearly independent in R^2.

The elements of a sequence are enumerated in a specific order. A set is a collection of objects, with no specified arrangement or order. Of course, to describe the set we may list its members, and that requires choosing an order. But, the order is not part of the set. The sets $\{1, 2, 3, 4\}$ and $\{4, 3, 2, 1\}$ are identical, whereas 1, 2, 3, 4 is quite a different sequence from 4, 3, 2, 1. The order aspect of sequences has no bearing on questions of independence, dependence, etc., because dependence (as defined) is not affected by the order. The sequence $\alpha_n, \ldots, \alpha_1$ is dependent if and only if the sequence $\alpha_1, \ldots, \alpha_n$ is dependent. In the next section, order will be important.

Exercises

1. Prove that if two vectors are linearly dependent, one of them is a scalar multiple of the other.

2. Are the vectors

$$\alpha_1 = (1, 1, 2, 4), \qquad \alpha_2 = (2, -1, -5, 2)$$
$$\alpha_3 = (1, -1, -4, 0), \qquad \alpha_4 = (2, 1, 1, 6)$$

linearly independent in R^4?

3. Find a basis for the subspace of R^4 spanned by the four vectors of Exercise 2.

4. Show that the vectors

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 2, 1), \qquad \alpha_3 = (0, -3, 2)$$

form a basis for R^3. Express each of the standard basis vectors as linear combinations of α_1, α_2, and α_3.

5. Find three vectors in R^3 which are linearly dependent, and are such that any two of them are linearly independent.

6. Let V be the vector space of all 2×2 matrices over the field F. Prove that V has dimension 4 by exhibiting a basis for V which has four elements.

7. Let V be the vector space of Exercise 6. Let W_1 be the set of matrices of the form

$$\begin{bmatrix} x & -x \\ y & z \end{bmatrix}$$

and let W_2 be the set of matrices of the form

$$\begin{bmatrix} a & b \\ -a & c \end{bmatrix}.$$