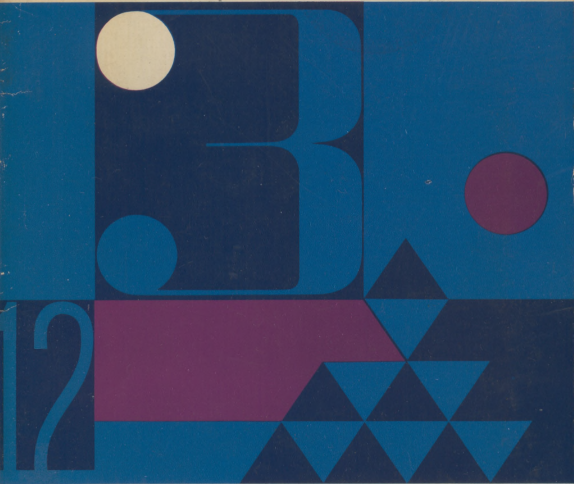
# elementary linear algebra

J. R. MUNKRES




ADDISON-WESLEY PUBLISHING COMPANY, INC.

# elementary linear algebra

JAMES R. MUNKRES

*Massachusetts Institute of Technology*

ADDISON-WESLEY PUBLISHING COMPANY, INC.

Reading, Massachusetts • Palo Alto • London

This book is in the
ADDISON-WESLEY SERIES IN MATHEMATICS

*Consulting Editor*
LYNN H. LOOMIS



*Printed in the United States of America*

# PREFACE

Our purpose in this book is to study, in some detail, the general system of $k$ linear equations in $n$ unknowns. For this purpose, we begin by studying vector spaces, subspaces, and linear independence of vectors. With these tools of linear algebra at our disposal, we then obtain general theorems about the consistency of a system of linear equations and about the dimension of its set of solutions. We conclude by studying determinants and their relation to systems of linear equations. A supplementary section deals with linear independence of functions.

The background we expect of the reader is not extensive. He will need to be familiar with vectors, having studied them either in calculus or in physics, if he is to understand the geometric ideas behind some of our definitions; and in the last section, we assume some familiarity with elementary calculus.

It is our contention that the most appropriate time for a student to begin his study of linear algebra is during the final stages of a course in calculus. For one thing, if he waits until then, he will not only have acquired some familiarity with vectors, but also, and more important, he should be prepared to read and to understand exact definitions and careful arguments, at least when they do not involve too heavy a dose of abstraction.

On the other hand, if his study of linear algebra precedes a course in differential equations, he will have at hand the various existence theorems for systems of linear equations which he will need in any careful treatment of the subject.* He will also have acquired some familiarity with the fundamental notion of linear independence, which arises again there.

This book is clearly not intended as a text in linear algebra. Only the most elementary topics are considered—those which are relevant to the study of systems of linear equations. It might be more appropriately considered as an introduction to a more thorough and abstract treatment of the subject, such as appears in Halmos' *Finite-dimensional Vector Spaces* or Hoffman and Kunze's *Linear Algebra.*

---

* See, for example, Chapter 4 of Kaplan's *Ordinary Differential Equations*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1961.

Our selection of material has been guided by these ideas as to the use to which this book might be put; likewise, the manner of presentation is what we consider to be appropriate for students at this level. We are grateful to Arthur Mattuck, Edward M. Brown, and others of the staff of the Massachusetts Institute of Technology for their comments and suggestions on both of these matters.

I dedicate this little book with affection to two of my teachers, Irene Kucera and Clint Gass.

J. R. M.

*Cambridge, Mass.*
*January 1964*

# CONTENTS

## §1. VECTOR SPACES

**Definition.** Let $V^n$ denote the set of all $n$-tuples $[a_1, a_2, \ldots, a_n]$ of real numbers; $V^n$ is called a *vector space*, and its elements are called *vectors*. We often denote a vector $[a_1, \ldots, a_n]$ by the symbol $\mathbf{a}$; the numbers $a_1, \ldots, a_n$ are called the *components* of $\mathbf{a}$.

If $\mathbf{a} = [a_1, \ldots, a_n]$ and $\mathbf{b} = [b_1, \ldots, b_n]$ are vectors, and $c$ is a scalar (a real number), we define $\mathbf{a} + \mathbf{b}$, $c\mathbf{a}$, and $\mathbf{a} \cdot \mathbf{b}$ by the equations

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n],$$
$$c\mathbf{a} = [ca_1, ca_2, \ldots, ca_n],$$
$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

These operations are called *addition of vectors*, *multiplication of a vector by a scalar*, and *dot product* (or *scalar product*) *of vectors*, respectively. They have the following properties (1) through (12):

(1) $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$ (associativity).
(2) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ (commutativity).
(3) There is a unique vector $\mathbf{0}$ such that $\mathbf{a} + \mathbf{0} = \mathbf{a}$ for every $\mathbf{a}$.
(4) $\mathbf{a} + (-1)\mathbf{a} = \mathbf{0}$.
(5) $c(d\mathbf{a}) = (cd)\mathbf{a}$ (associativity).
(6) $(c + d)\mathbf{a} = c\mathbf{a} + d\mathbf{a}$ (distributivity).
(7) $c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}$ (distributivity).
(8) $1\mathbf{a} = \mathbf{a}$.

These eight properties, which involve only the first two operations, are called the *vector space properties; any* set of objects having two operations which possess these properties is called a *vector space*. (More precisely, it is called a vector space over the real numbers, since we take the scalars to be real numbers. One could instead take the scalars to be complex numbers, but this we will not do.)

The dot product operation has the following properties:

(9) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$.
(10) $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \cdot \mathbf{b}) + (\mathbf{a} \cdot \mathbf{c})$.
(11) $(c\mathbf{a}) \cdot \mathbf{b} = c(\mathbf{a} \cdot \mathbf{b})$.
(12) $\mathbf{a} \cdot \mathbf{a} > 0$ unless $\mathbf{a} = \mathbf{0}$.

Verification of the properties is routine. For instance, to check (1), one first applies the definition of vector addition to compute

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = [a_1 + (b_1 + c_1), \ldots, a_n + (b_n + c_n)],$$
$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = [(a_1 + b_1) + c_1, \ldots, (a_n + b_n) + c_n].$$

Then one notes that $a_i + (b_i + c_i) = (a_i + b_i) + c_i$ for $i = 1, \ldots, n$; this is one of the familiar properties of the addition operation for real numbers. It shows that the above two $n$-tuples have all their components equal, and hence are equal.

The vector spaces $V^1$, $V^2$, and $V^3$ have simple geometric interpretations which will help us to understand them. To picture $V^3$, for instance, we represent the vector $[a_1, a_2, a_3]$ by the directed line segment in space having its initial point at the origin and its end point at the point with coordinates $(a_1, a_2, a_3)$. Hence our geometric interpretation of the 3-tuple $[a_1, a_2, a_3]$ is the directed line segment which is commonly denoted by $a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ in physics and in calculus. Our definitions for the addition of vectors, and for multiplication of a vector by a scalar, correspond to the familiar ways of adding directed line segments, and multiplying a directed line segment by a scalar. Similarly, the definition of the dot product becomes a familiar formula in this geometric interpretation.

Let us note that it is usual in physics and in calculus to allow directed line segments to lie anywhere in space. This we shall not do; instead we shall consider only those directed line segments whose initial points lie at the origin. This will be more convenient for our purpose, which is to use directed line segments as a means of picturing $V^3$.

Finally, we should remark that we make a mental distinction between the triple $[a_1, a_2, a_3]$ which we call a *vector*, and the triple $(a_1, a_2, a_3)$ which represents a *point*. The reason is that we *do* different things with them. We add two triples of the first kind, for instance, but we certainly do not add two points together. On the other hand, we speak of the distance between two points, but not of the distance between two vectors.

## EXERCISES

1. Evaluate:

(a) $[1, -1, 1, 1] + 2[4, 1, 1, -1]$ (b) $a[1, 0, 0, 0] + b[1, 1, 1, 1]$
(c) $[1, 1] \cdot ([-1, 2] + [1, 0])$ (d) $[a, b, 0] \cdot [1, 0, 7]$

2. Give geometric interpretations for $V^1$ and $V^2$ similar to the one for $V^3$.

3. Check properties (2) through (12) by reducing each to a familiar property of the real numbers.

4. Properties (1) through (12) involve *five* operations: the three vector operations which we defined above, as well as the operations of addition and multiplication for real numbers. In stating these properties, we have used the symbol $+$ ambiguously for both addition of vectors and addition of real numbers. Examine properties (1) through (12) above, and decide for each $+$ symbol, which type of addition is meant. Then do the same for each indicated multiplication. Does any one of the properties involve all five operations? four of the operations? three of them?

5. We define $\|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$ and call it the *norm* of $\mathbf{a}$. If $\mathbf{a}$ is in $V^3$, what is the geometric interpretation of $\|\mathbf{a}\|$? If $\|\mathbf{a}\| = 1$, $\mathbf{a}$ is called a *unit vector*.

6. Consider the following statements and decide for each whether it is true, false, or meaningless. If it is true, prove it, using only properties (1) through (12).

(a) $0\mathbf{a} = \mathbf{0}$ (b) $\mathbf{a} + c = c + \mathbf{a}$ (c) $c\mathbf{0} = \mathbf{0}$
(d) Given $\mathbf{a}$, there is a unique $\mathbf{b}$ such that $\mathbf{a} + \mathbf{b} = \mathbf{0}$.
(e) Given $\mathbf{a}$, there is a unique $c$ such that $c\mathbf{a} = \mathbf{0}$.
(f) Given $\mathbf{a}$, there is a unique $c$ such that $c\mathbf{a} = 1$.
(g) Given $\mathbf{a}$, there is a unique $\mathbf{b}$ such that $\mathbf{a} \cdot \mathbf{b} = 1$.
(h) $(c\mathbf{a}) \cdot (d\mathbf{b}) = (cd)(\mathbf{a} \cdot \mathbf{b})$ (i) $\mathbf{0} \cdot \mathbf{a} = 0$
(j) $(\mathbf{a} \cdot \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \cdot \mathbf{c})$ (k) $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = (\mathbf{a} \cdot \mathbf{c}) + (\mathbf{b} \cdot \mathbf{c})$
(l) $(\mathbf{a} \cdot \mathbf{b})\mathbf{c} = (\mathbf{b} \cdot \mathbf{c})\mathbf{a}$ (m) $\|c\mathbf{a}\| = c\|\mathbf{a}\|$
(n) $\|c\mathbf{a}\| = |c|\,\|\mathbf{a}\|$ (o) If $\mathbf{a} \neq \mathbf{0}$, $\|\mathbf{a}\| > 0$.
(p) $\|\mathbf{0}\| = 0$
(q) $(1/\|\mathbf{a}\|)\mathbf{a}$ has norm 1 if $\mathbf{a} \neq \mathbf{0}$.

7. Let $\mathbf{a}$ and $\mathbf{b}$ be nonzero vectors. Prove that

$$-1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|\,\|\mathbf{b}\|} \leq 1,$$

using only properties (1) through (12). [*Hint*: Let $c = 1/\|\mathbf{a}\|$ and $d = 1/\|\mathbf{b}\|$ and use the fact that $\|c\mathbf{a} + d\mathbf{b}\|^2 \geq 0$ and $\|c\mathbf{a} - d\mathbf{b}\|^2 \geq 0$.]

8. Under the hypotheses of Exercise 7, show that $|\mathbf{a} \cdot \mathbf{b}| = \|\mathbf{a}\|\,\|\mathbf{b}\|$ if and only if $\mathbf{a}$ is a scalar multiple of $\mathbf{b}$.

9. The inequality $|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\|\,\|\mathbf{b}\|$, proved in Exercise 7, is called the *Cauchy-Schwarz inequality*. It allows us to define the *angle* between the nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ to be the number $\cos^{-1}(\mathbf{a} \cdot \mathbf{b}/\|\mathbf{a}\|\,\|\mathbf{b}\|)$. Why do we need the Cauchy-Schwarz inequality for this? Does the inequality hold if $\mathbf{a}$ or $\mathbf{b}$ is zero?

10. Show that $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$. [*Hint*: Use the Cauchy-Schwarz inequality to show that $(\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) \leq (\|\mathbf{a}\| + \|\mathbf{b}\|)^2$.]

11. If $\mathbf{a}$ and $\mathbf{b}$ are vectors in $V^3$, interpret the inequality of Exercise 10 geometrically.

## §2. SUBSPACES

**Definition.** Let $W$ be a nonempty set of vectors of $V^n$. Suppose $W$ has the property that

(a) whenever $\mathbf{a}$ and $\mathbf{b}$ are in $W$, so is the vector $\mathbf{a} + \mathbf{b}$, and
(b) whenever $\mathbf{a}$ is in $W$, so is the vector $c\mathbf{a}$, for any scalar $c$.

Then $W$ is called a *subspace* of $V^n$.

Now the three vector operations are defined for the vectors of $W$, and it is trivial to check that they still possess properties (1) through (12). (For instance, $\mathbf{0}$ is always in $W$ because $\mathbf{0} = 0\mathbf{a}$, where $\mathbf{a}$ is any vector of $W$.) Hence $W$ may be looked on as a vector space in its own right.

Note that $W$ might possibly be all of $V^n$, or it might possibly consist of the zero vector alone; these are the two extremes of size for $W$.

EXAMPLE. Consider the set $W$ of all vectors $\mathbf{a} = [a_1, a_2, a_3]$ in $V^3$ such that $a_1 = a_3$; this is a subspace of $V^3$. To check this, suppose $\mathbf{b} = [b_1, b_2, b_3]$ is another vector in $W$. Then $\mathbf{a} + \mathbf{b}$ is in $W$, since $a_1 + b_1 = a_3 + b_3$; and $c\mathbf{a}$ is in $W$, since $ca_1 = ca_3$.

EXAMPLE. Consider the set $W$ of all vectors $[a_1, a_2, a_3, a_4]$ in $V^4$ such that $a_3 \geq 0$; this is not a subspace of $V^4$. It is true that the sum of any two vectors in $W$ is in $W$, but not every scalar multiple of an element of $W$ is in $W$. For instance, let $\mathbf{a} = [1, 0, 2, -1]$. Then $\mathbf{a}$ is in $W$, but $(-2)\mathbf{a}$ is not in $W$.

**Definition.** If $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is a set of vectors in $V^n$, a *linear combination* of these vectors is an expression of the form

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_p\mathbf{a}_p,$$

where some or all of the scalars $c_i$ may be zero, of course.

**Definition.** Let $W$ be a subspace of $V^n$. If $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is a set of vectors in $W$ such that every vector in $W$ may be written as a linear combination of them, then the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$ are said to *span* $W$, or to form a *spanning set* for $W$.

To illustrate, in the first example above, the vectors $[1, 0, 1]$ and $[0, 1, 0]$ span $W$, for if $\mathbf{a}$ is in $W$, then $[a_1, a_2, a_3] = a_1[1, 0, 1] + a_2[0, 1, 0]$. Another spanning set for $W$ is the set consisting of $[1, -1, 1]$ and $[0, 2, 0]$, since $[a_1, a_2, a_3] = a_1[1, -1, 1] + ((a_1 + a_2)/2)[0, 2, 0]$.

We shall prove in §4 that, given *any* subspace $W$ of $V^n$, the problem of finding a finite set of vectors which span $W$ may be solved. For the present, we content ourselves with stating the following theorem:

**2.1 Theorem.** There is a spanning set for $V^n$ which consists of $n$ vectors.

REMARK. We might consider the reverse of the problem just mentioned, and ask whether, given any finite set $\mathbf{a}_1, \ldots, \mathbf{a}_p$ of vectors in $V^n$, there is a subspace $W$ which they span. The answer is yes; in fact, there is only one such subspace. It consists of all vectors in $V^n$ which may be written as linear combinations of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$. To check that this subset is in fact a subspace, we only need to make the following computations:

$$(c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p) + (d_1\mathbf{a}_1 + \cdots + d_p\mathbf{a}_p) = (c_1 + d_1)\mathbf{a}_1 + \cdots + (c_p + d_p)\mathbf{a}_p,$$

$$d(c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p) = (dc_1)\mathbf{a}_1 + \cdots + (dc_p)\mathbf{a}_p.$$

This is one way of determining a subspace of $V^n$; another useful way of determining a subspace is provided by the following definition:

**Definition.** Consider a system of $k$ linear equations in $n$ unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1,\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2,\\ &\vdots\\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n &= b_k. \end{aligned}$$

Here $x_1, \ldots, x_n$ are the unknowns, and the symbols $a_{ij}$ and $b_i$ denote scalars. Note that $i$ denotes the number of the equation in which $a_{ij}$ appears, and $j$ tells you the unknown by which it is multiplied. If the constants $b_i$ are all equal to zero, the system is said to be *homogeneous*.

By a *solution* of the system, we mean a vector $[d_1, \ldots, d_n]$ of $V^n$ such that when $d_1$ is substituted for $x_1$, $d_2$ is substituted for $x_2$, and so on, all the equations are satisfied. If the system is homogeneous, the set of all such vectors $[d_1, \ldots, d_n]$ will form a subspace of $V^n$, as the following theorem states; it is called the *solution space* of the system of linear equations.

**2.2 Theorem.** Consider the set of all solutions of a given system of $k$ linear equations in $n$ unknowns. If the system is homogeneous, the set of solutions is a subspace of $V^n$. If the system is not homogeneous, the set of solutions is not a subspace of $V^n$.

## EXERCISES

1. Determine which of the following subsets of $V^3$ are subspaces of $V^3$. For each which is a subspace, find a spanning set.

(a) The set of all vectors of the form $[a, b, 0]$, where $a$ and $b$ are scalars.
(b) The set of all vectors of the form $[a - b, a + b, 2a]$, where $a$ and $b$ are scalars.
(c) The set of all vectors of the form $[a, a, a]$.
(d) The set of all vectors of the form $[1, 0, a]$.

*(cont.)*

(e) The set of all vectors of the form

$$a[1, 0, 1] + b[0, -1, 0] + c[1, 1, 1].$$

(f) The set of all vectors $\mathbf{a}$ such that

$$[1, -1, 2]\cdot\mathbf{a} = 0.$$

(g) The set of all vectors $\mathbf{a}$ such that

$$[1, -1, 2]\cdot\mathbf{a} = 1.$$

(h) The set of all vectors $\mathbf{a}$ such that $\|\mathbf{a}\| \geq 1$.

(i) The set of all vectors $\mathbf{a}$ such that $\|\mathbf{a}\| = 0$.

2. Interpret geometrically each of the subspaces of $V^3$ given in Exercise 1.

3. Find spanning sets for the solution spaces of the following systems. Each solution space is a subspace of $V^3$; interpret it geometrically.

(a)
$$\begin{aligned} x_1 - x_2 + x_3 &= 0 \\ x_1 \phantom{{}-x_2} + 2x_3 &= 0 \end{aligned}$$

(b)
$$\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ 3x_2 - x_3 &= 0 \\ 2x_1 - x_2 - x_3 &= 0 \end{aligned}$$

(c)
$$\begin{aligned} x_1 - x_2 - x_3 &= 0 \\ x_2 + x_3 &= 0 \\ x_1 + 2x_2 - x_3 &= 0 \end{aligned}$$

(d)
$$x_1 + x_2 + x_3 = 0$$

4. What would you conjecture are the possible kinds of subspaces which $V^3$ may have?

5. Show that if $\mathbf{b}_1, \ldots, \mathbf{b}_q$ are vectors which lie in the subspace $W$ of $V^n$, then any linear combination of them also lies in $W$.

6. Prove Theorem 2.1.

7. Prove Theorem 2.2.

8. Show that the vector $[1, 1, 1, 1]$ spans the solution space of the system

$$\begin{aligned} x_1 - x_2 \phantom{{}-x_3-x_4} &= 0 \\ x_2 - x_3 \phantom{{}-x_4} &= 0 \\ x_3 - x_4 &= 0 \end{aligned}$$

9. Consider the directed line segments in space which we use to picture the vectors $[1, -1, 2]$, $[1, 0, 3]$, and $[1, -2, 1]$ of $V^3$. Show that these line segments are coplanar. Show that these three vectors do not span all of $V^3$. Show that one of these vectors equals a linear combination of the others.

10. Use a geometrical proof to show that if $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors of $V^3$ whose corresponding directed line segments are coplanar, then one of them must equal a linear combination of the others.

## §3. A USEFUL THEOREM

For later use, we need the following fact about linear equations:

> **3.1 Theorem.** Given a homogeneous system of $k$ linear equations in $n$ unknowns. If $k$ is less than $n$, then the solution space contains some vector other than **0**.

*Proof.* We are concerned here only with proving the *existence* of some solution other than **0**, not with actually finding such a solution in practice, nor with finding all possible solutions. (We will study the practical problem in much greater detail in a later section.)

We start with a system of $k$ equations in $n$ unknowns:

$$(*)\qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0, \\ &\vdots \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n &= 0. \end{aligned}$$

Our procedure will be to reduce the size of this system step-by-step by eliminating first $x_1$, then $x_2$, and so on. After $k - 1$ steps, we will be reduced to solving just one equation and this will be easy. But a certain amount of care is needed in the description—for instance, if $a_{11} = \cdots = a_{k1} = 0$, it is nonsense to speak of "eliminating" $x_1$, since all its coefficients are zero. We have to allow for this possibility.

To begin then, if all the coefficients of $x_1$ are zero, you may verify that the vector $[1, 0, \ldots, 0]$ is a solution of the system which is different from **0**, and you are done. Otherwise, at least one of the coefficients of $x_1$ is nonzero, and we may suppose for convenience that the equations have been arranged so that this happens in the first equation, with the result that $a_{11} \neq 0$. We multiply the first equation by the scalar $a_{21}/a_{11}$ and then subtract it from the second, eliminating the $x_1$-term from the second equation. Similarly, we eliminate the $x_1$-term in each of the remaining equations. The result is a *new* system of linear equations of the form

$$(**)\qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\ \boxed{\begin{aligned} b_{22}x_2 + \cdots + b_{2n}x_n &= 0, \\ &\vdots \\ b_{k2}x_2 + \cdots + b_{kn}x_n &= 0. \end{aligned}} \end{aligned}$$

Now any solution of this new system of equations is also a solution of the old system $(*)$, because we can recover the old system from the new one:

we merely multiply the first equation of the system (**) by the same scalars we used before, and then we *add* it to the corresponding later equations of this system.

The crucial thing about what we have done is contained in the following statement: If the smaller system enclosed in the box above has a solution other than the zero vector, then the larger system (**) also has a solution other than the zero vector [so that the original system (*) we started with has a solution other than the zero vector]. We prove this as follows: Suppose $[d_2, \ldots, d_n]$ is a solution of the smaller system, different from $[0, \ldots, 0]$. We substitute into the first equation and solve for $x_1$, thereby obtaining the following vector,

$$[(-1/a_{11})(a_{12}d_2 + \cdots + a_{1n}d_n), d_2, \ldots, d_n],$$

which you may verify is a solution of the larger system (**).

In this way we have reduced the size of our problem; we now need only to prove our theorem for a system of $k - 1$ equations in $n - 1$ unknowns. If we apply this reduction a second time, we reduce the problem to proving the theorem for a system of $k - 2$ equations in $n - 2$ unknowns. Continuing in this way, after $k - 1$ elimination steps in all, we will be down to a system consisting of only one equation, in $n - k + 1$ unknowns. Now $n - k + 1 \geq 2$, because we assumed as our hypothesis that $n > k$; thus our problem reduces to proving the following statement: a "system" consisting of *one* linear homogeneous equation in *two or more* unknowns always has a solution other than $\mathbf{0}$.

## EXERCISES

1. Show that this last statement is actually a special case of the theorem we are trying to prove.

2. Prove this last statement, and thus complete the proof of the theorem. Note that you must consider the possibility that one or more of the coefficients is zero.

3. Use the method of elimination given in the proof to reduce the system (*)

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

to the system (**)

$$\begin{aligned} 2x_1 + \quad 3x_2 \quad - x_3 &= 0, \\ \boxed{\tfrac{1}{2}x_2 + \tfrac{5}{2}x_3 = 0,} \end{aligned}$$

Find a nonzero solution of the smaller "system" $\frac{1}{2}x_2 + \frac{5}{2}x_3 = 0$ and obtain from it a solution of the system (**). Show that this is also a solution of the system (*).

Prove that (*) and (**) have exactly the same solution spaces.

## §4. LINEAR INDEPENDENCE, BASES, AND DIMENSION

**Definition.** The set $\mathbf{a}_1, \ldots, \mathbf{a}_p$ of vectors in $V^n$ is said to be *linearly independent* (or simply, *independent*) if no one of them equals a linear combination of the others. The only exception is the set consisting of the zero vector alone, which we do not consider to be independent.

For example, the set $\mathbf{a}_1, \mathbf{a}_2$ is independent if neither vector is a scalar multiple of the other. A second example: A set consisting of three vectors in $V^3$ is independent if their corresponding directed line segments are not coplanar. (See Exercise 10 of §2.) A set consisting of a single vector is independent if that vector is not the zero vector.

A useful condition which is equivalent to linear independence is given by the following theorem:

**4.1 Theorem.** A set $\mathbf{a}_1, \ldots, \mathbf{a}_p$ of vectors in $V^n$ is independent *if and only if* the only linear combination

$$c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p$$

of them which equals the zero vector $\mathbf{0}$ is the trivial linear combination $0\mathbf{a}_1 + 0\mathbf{a}_2 + \cdots + 0\mathbf{a}_p$.

*Proof.* First, assume $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is an independent set. Then suppose $c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p = \mathbf{0}$. If some coefficient, say $c_i$, were different from zero, then we could solve for $\mathbf{a}_i$ and obtain

$$\mathbf{a}_i = (-1/c_i)(c_1\mathbf{a}_1 + \cdots + c_{i-1}\mathbf{a}_{i-1} + c_{i+1}\mathbf{a}_{i+1} + \cdots + c_p\mathbf{a}_p),$$

which would contradict the definition of linear independence. Hence all the coefficients $c_i$ must equal 0.

Conversely, suppose that only the trivial linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_p$ equals $\mathbf{0}$. If $\mathbf{a}_1, \ldots, \mathbf{a}_p$ were not an independent set, we would have

$$\mathbf{a}_i = d_1\mathbf{a}_1 + \cdots + d_{i-1}\mathbf{a}_{i-1} + d_{i+1}\mathbf{a}_{i+1} + \cdots + d_p\mathbf{a}_p$$

for some $i$, by definition. But then the nontrivial linear combination

$$d_1\mathbf{a}_1 + \cdots + d_{i-1}\mathbf{a}_{i-1} - \mathbf{a}_i + d_{i+1}\mathbf{a}_{i+1} + \cdots + d_p\mathbf{a}_p$$

would equal $\mathbf{0}$, contrary to hypothesis.

**4.2 Corollary.** Let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be an independent set of vectors in $V^n$; let $W$ denote the subspace they span. If $\mathbf{b}$ is a vector of $V^n$ not lying in $W$, then the set $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{b}$ is independent.

*Proof.* Suppose $c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p + d\mathbf{b} = \mathbf{0}$. Now if $d \neq 0$, we may solve for $\mathbf{b}$ and have

$$\mathbf{b} = (-1/d)(c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p),$$

which implies that $\mathbf{b}$ lies in $W$, contrary to hypothesis. Hence $d = 0$. But then $c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p = \mathbf{0}$, from which it follows that $c_1 = c_2 = \cdots = c_p = 0$, since $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is an independent set. Applying the preceding theorem, it follows that $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{b}$ must be an independent set.

**4.3 Theorem.** Let $W$ be a subspace of $V^n$ which is spanned by a set consisting of $p$ vectors $\mathbf{b}_1, \ldots, \mathbf{b}_p$. Then any *independent* set of vectors lying in $W$ can contain at most $p$ vectors.

*Proof.* The way we prove this theorem is to show that any set $\mathbf{a}_1, \ldots, \mathbf{a}_q$ consisting of $q$ vectors of $W$, where $q > p$, cannot be independent. It then follows that if a set of vectors of $W$ is to be independent, it must not contain more than $p$ vectors.

To prove that the set $\mathbf{a}_1, \ldots, \mathbf{a}_q$ is not independent, we need to find scalars $x_1, \ldots, x_q$, not all zero, such that $x_1\mathbf{a}_1 + \cdots + x_q\mathbf{a}_q = \mathbf{0}$. Now since each vector $\mathbf{a}_i$ lies in $W$, and since the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_p$ span $W$ by hypothesis, we have equations

$$\mathbf{a}_i = c_{1i}\mathbf{b}_1 + c_{2i}\mathbf{b}_2 + \cdots + c_{pi}\mathbf{b}_p$$

for $i = 1, \ldots, q$, where the $c_{ij}$ are scalars. Multiplying the $i$th equation by $x_i$ and adding equations, we have

$$\begin{aligned} x_1\mathbf{a}_1 + \cdots + x_q\mathbf{a}_q = {} & (x_1c_{11} + x_2c_{12} + \cdots + x_qc_{1q})\mathbf{b}_1 \\ & + (x_1c_{21} + x_2c_{22} + \cdots + x_qc_{2q})\mathbf{b}_2 + \cdots \\ & + (x_1c_{p1} + x_2c_{p2} + \cdots + x_qc_{pq})\mathbf{b}_p. \end{aligned}$$

Let us examine the coefficients of the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_p$ in this equation and ask the question: What values of $x_1, \ldots, x_q$ will make all these coefficients vanish? Finding such values of $x_1, \ldots, x_q$ simply involves solving a homogeneous system of $p$ linear equations in $q$ unknowns. Since $q > p$, we have more unknowns than equations, so Theorem 3.1 tells us that the system has a solution other than the zero solution $x_1 = x_2 = \cdots = x_q = 0$. But this means that there are values of $x_1, \cdots, x_q$, not all zero, such that

$$x_1\mathbf{a}_1 + \cdots + x_q\mathbf{a}_q = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_p = \mathbf{0},$$

as desired.

**Definition.** Let $W$ be a subspace of $V^n$. Suppose we consider all possible independent sets of vectors lying in $W$. By the preceding theorem, none of these sets can contain more than $n$ vectors, since $V^n$ has a spanning set consisting of $n$ vectors (see Theorem 2.1). Let us take as large an independent set in $W$ as possible; the number of vectors in this set is called the *dimension* of $W$. $V^n$ itself has dimension $n$ (see Exercise 10).

**4.4 Theorem.** Let $p$ be the dimension of $W$. Then *any* independent set of vectors of $W$ consisting of $p$ vectors necessarily spans $W$.

*Proof.* Let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be an independent set of vectors of $W$. If this set does not span $W$, then there is a vector $\mathbf{b}$ of $W$ lying outside the subspace $W_1$ which it does span. By Corollary 4.2, the set $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{b}$ is an independent set of vectors lying in $W$. But this set consists of more than $p$ vectors, contradicting the hypothesis that the dimension of $W$ is $p$.

REMARK. This theorem shows that any subspace $W$ of $V^n$ has a finite spanning set; to find one, merely choose as large an independent set in $W$ as possible. This is fine in theory, but not much help in practice. We will tackle the practical problem in the next two sections.

**Definition.** If $W$ is a subspace of $V^n$, an independent set $\mathbf{a}_1, \ldots, \mathbf{a}_p$ of vectors of $W$ which spans $W$ is called a *basis* for $W$. The number of vectors in this set necessarily equals the dimension of $W$, for since these vectors span $W$, $W$ cannot contain an independent set consisting of more than $p$ vectors, and thus this is as large an independent set as $W$ contains.

We now prove two theorems which will be useful to us later on:

**4.5 Theorem.** Let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be an independent set of vectors lying in $V^n$. There exist vectors $\mathbf{a}_{p+1}, \ldots, \mathbf{a}_n$ of $V^n$ such that $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is a basis for $V^n$.

*Proof.* If $\mathbf{a}_1, \ldots, \mathbf{a}_p$ spans $V^n$, we are through; if not, we may choose a vector $\mathbf{a}_{p+1}$ of $V^n$ lying outside the subspace of $V^n$ these vectors do span. By 4.2, the set $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{a}_{p+1}$ is independent.

Now if this set spans $V^n$, we are through; if not, we may choose a vector $\mathbf{a}_{p+2}$ lying outside the subspace these vectors do span. Similarly, we continue choosing vectors, one at a time. The process stops after we choose $\mathbf{a}_n$, because then we have an independent set consisting of $n$ vectors, and this set necessarily spans $V^n$, by Theorem 4.4.

**4.6 Theorem.** Suppose $W$ is a subspace of $V^n$ spanned by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$. There is a subset of the set $\mathbf{a}_1, \ldots, \mathbf{a}_k$ whose elements form a basis for $W$.

*Proof.* If the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are independent, they already form a basis for $W$. If not, let us consider all possible independent sets of vectors which are subsets of the set $\mathbf{a}_1, \ldots, \mathbf{a}_k$; choose one such set which is as large as possible. Suppose for convenience it consists of the first $p$ vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$. These $p$ vectors span a subspace $W_1$ of $V^n$ which is contained in $W$; in fact they form a basis for $W_1$.

Now the vector $\mathbf{a}_{p+1}$ must lie in $W_1$, for otherwise the set $\mathbf{a}_1, \ldots, \mathbf{a}_p, \mathbf{a}_{p+1}$ would be independent, by 4.2. This would give us a contradiction. Similarly, each of the vectors $\mathbf{a}_{p+2}, \ldots, \mathbf{a}_k$ must lie in $W_1$. As a result, all the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ lie in $W_1$, so that any linear combination of them also lies in $W_1$. Thus $W_1$ is necessarily all of $W$.

## EXERCISES

1. Show that the set of vectors of $V^3$ consisting of $[1, -1, 1]$ and $[1, 0, 1]$ is independent. Does this set span $V^3$?

2. Let $\mathbf{b}_1$ and $\mathbf{b}_2$ span the subspace $W$ of $V^n$. Let $\mathbf{a}_1 = \mathbf{b}_1 - \mathbf{b}_2$, $\mathbf{a}_2 = 3\mathbf{b}_1 + 3\mathbf{b}_2$, $\mathbf{a}_3 = 2\mathbf{b}_1 - 5\mathbf{b}_2$. Show directly that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ do not form an independent set.

3. Show that the vectors $[1, 0, 1, 1, 0]$, $[0, 1, 2, 3, 0]$, and $[0, 0, 0, 0, 1]$ form an independent set. Find two additional vectors which along with these will form a basis for $V^5$.

4. Show that $[1, 3]$, $[-1, 2]$, and $[7, 6]$ do not form an independent set.

5. Show that any vector $[a, b, c]$ of $V^3$ is equal to a linear combination of the vectors $[1, -1, 2]$, $[1, 1, 0]$, and $[0, 1, 0]$. Conclude that these three vectors necessarily form an independent set.

6. Prove: Any set of vectors of $V^n$ which includes $\mathbf{0}$ is not independent. (Hence, if $W$ is the subspace of $V^n$ consisting of $\mathbf{0}$ alone, then the dimension of $W$ is 0.)

7. Prove that if you delete some of the vectors from an independent set, the remaining vectors still form an independent set.

8. Prove the following: If $W$ is a subspace of $V^n$ of dimension $p$, then $0 \leq p \leq n$, and if $p = n$, then $W = V^n$.

9. Prove that if $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is a basis for $W$, then any vector $\mathbf{b}$ of $W$ can be written *uniquely* as a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_p$:

$$\mathbf{b} = c_1\mathbf{a}_1 + \cdots + c_p\mathbf{a}_p.$$

The scalars $c_1, \ldots, c_p$ are called the *components of* $\mathbf{b}$ *relative to the basis* $\mathbf{a}_1, \ldots, \mathbf{a}_p$.

10. Show that the vectors $\mathbf{e}_1 = [1, 0, 0, \ldots, 0]$, $\mathbf{e}_2 = [0, 1, 0, \ldots, 0]$, $\ldots, \mathbf{e}_n = [0, 0, 0, \ldots, 1]$ are a basis for $V^n$. This is called the *natural basis* for $V^n$. What are the components of the vector $\mathbf{a} = [a_1, \ldots, a_n]$ relative to this basis?

11. Two vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be *orthogonal* if $\mathbf{a} \cdot \mathbf{b} = 0$. (See Exercise 9 of §1.) Let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be a set of vectors which are mutually

orthogonal (that is, each is orthogonal to all the others). Show that if none of the vectors is the zero vector, then this set is independent.

12. Given the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ of $V^n$. Show that there exist scalars $c_1$, $d_1$, $d_2$ such that the vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ are orthogonal, where

$$\mathbf{b}_1 = \mathbf{a}_1,$$

$$\mathbf{b}_2 = \mathbf{a}_2 + c_1\mathbf{b}_1,$$

$$\mathbf{b}_3 = \mathbf{a}_3 + d_1\mathbf{b}_1 + d_2\mathbf{b}_2.$$

Show that if the set $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ is independent, so is the set $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$. Do these two sets span the same subspace of $V^n$?

13. Let $W$ be the subspace of $V^4$ spanned by $[1, 1, -1, 2]$, $[0, 1, 0, 2]$, and $[1, 1, 0, 1]$. Find an orthogonal basis for $W$.

14. Generalize the method of Exercise 12 to prove that given a basis $\mathbf{a}_1, \ldots, \mathbf{a}_p$ for a subspace $W$ of $V^n$, one can construct a basis for $W$ consisting of vectors which are mutually orthogonal. This construction is usually called the *Gram-Schmidt process.*

15. A basis for $W$ is called *orthonormal* if the vectors in it are mutually orthogonal and each has norm 1. Show that every subspace $W$ of $V^n$ has an orthonormal basis. Find an orthonormal basis for the subspace $W$ of Exercise 13.

16. If $\mathbf{a}_1, \ldots, \mathbf{a}_p$ is an orthonormal basis for $W$, and $\mathbf{b}$ is a vector of $W$, what are the components of $\mathbf{b}$ relative to this basis? Find a formula for $\|\mathbf{b}\|$ in terms of these components.

17. Show that Theorem 4.1 is true when the set of vectors consists of a single vector $\mathbf{a}_1$. The proof we gave assumed implicitly that $p > 1$.

## §5. ROW OPERATIONS ON MATRICES

Up to this point, we have not had a reasonable computational procedure for determining whether a given set of vectors is independent or not. Nor have we learned how to determine, in the case where the set is not independent, just how large an independent set it includes. We now give a procedure for doing this.

To say the same thing in a different way, suppose we have a subspace $W$ of $V^n$ which is specified by means of a spanning set, and we want to determine the dimension of $W$. Our procedure will tell us how to do this; in fact, it will construct for us a basis for $W$.

On the other hand, a subspace $W$ of $V^n$ might be specified as the solution space of a homogeneous system of linear equations. Here again, we would like to be able to determine the dimension of $W$ and to find a basis for $W$. Our procedure will enable us to do this as well, as we shall see in §6.

**Definition.** Suppose $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are vectors of $V^n$. Let us write each out in terms of its components:

$$\mathbf{a}_i = [a_{i1}, a_{i2}, \cdots, a_{in}],$$

and let us form a rectangular array whose rows are made up of the components of these vectors, successively:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \cdots a_{1n} \\ a_{21} & a_{22} & a_{23} \cdots a_{2n} \\ \vdots & & \vdots \\ a_{k1} & a_{k2} & a_{k3} \cdots a_{kn} \end{bmatrix}$$

Such a rectangular array of scalars is called a *matrix*; since it has $k$ rows and $n$ columns, it is said to be a $k$ *by* $n$ matrix. The scalars $a_{ij}$ are called the *entries* of the matrix.

Given such a $k$ by $n$ matrix $A$, the subspace $W$ of $V^n$ spanned by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ which form the rows of this matrix is called the *row space* of the matrix. The dimension of the row space is called the *rank* $r$ of the matrix.

We may apply Theorem 4.6 to conclude that $r$ may also be described as the largest independent set of vectors which is a subset of the set $\mathbf{a}_1, \ldots, \mathbf{a}_k$. In particular, if $r = k$, then the rows of $A$ form an independent set.

**Definition.** We now define three operations on matrices which will aid us in determining the rank of a matrix. They are called the *elementary row operations*:

(1) Interchange two rows, say the $i$th and $j$th rows.

(2) Multiply every entry in a given row, say the $i$th row, by a nonzero scalar $c$.

(3) Replace a given row, say the $i$th row, by itself plus a scalar multiple $d$ of some other row, say the $j$th row.

Each of these operations, if applied to a matrix, gives a new matrix; a useful fact for us is that each of these operations is reversible, so that we can recover the old matrix from the new one by applying one of these operations. To prove this fact, we note that the reverse of operation (1) is operation (1) itself; and the reverse of operation (2) is the operation which multiplies every entry in the $i$th row by the scalar $1/c$. The reverse of operation (3) is to replace the $i$th row of the new matrix by itself plus the scalar multiple $(-d)$ of the $j$th row of the new matrix.

The usefulness of the elementary row operations in determining the rank of a matrix comes from the following theorem.

**5.1 Theorem.** Let $B$ be a matrix obtained by applying an elementary row operation to the matrix $A$. Then the row space of $B$ is the same as the row space of $A$.

*Proof.* What we shall do is to show first that each vector in the row space of $B$ also lies in the row space of $A$. Then since $A$ may be obtained by applying an elementary row operation to $B$, it follows in the same way that every vector in the row space of $A$ also lies in the row space of $B$, and so the two row spaces must be the same.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_k$ be the vectors whose components form the rows of $A$. If $B$ is obtained by applying operation (1) to $A$, the same vectors form the rows of $B$, although they are written in a different order. If $B$ is obtained by applying operation (2) to $A$, the rows of the new matrix are $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, c\mathbf{a}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k$; and if it is obtained by applying operation (3), the rows of $B$ are $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}_i + d\mathbf{a}_j, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k$. It is now clear that any linear combination of the rows of $B$ is also equal to a linear combination of the rows of $A$. Since any vector in the row space of $B$ is equal to a linear combination of the rows of $B$, our theorem follows.

---

We now give the procedure for finding a basis for the subspace $W$ of $V^n$ spanned by a given set of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$. We form the matrix $A$ whose rows are the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$, and then we apply elementary row operations to the matrix $A$ until we reduce it to a matrix $B$ in a certain standard form, called reduced row-echelon form. For such a matrix $B$, the nonzero rows of $B$ automatically form a basis for the row space of $B$. Hence they are a basis for the row space of $A$ as well, since $A$ and $B$ have the same row space (by Theorem 5.1).

**First Part of the Reduction Process.** Examine the first column of the matrix $A$.

(a) If some entry in the first column is nonzero, bring it to the upper left-hand corner of the array by some elementary operation of type (1). Then apply operation (3) several times, replacing each of the rows after the first by itself plus an appropriate scalar multiple of the first row, where the scalar is chosen so that the resulting matrix has the form

$$\begin{bmatrix} b_{11} & b_{12} \cdots b_{1n} \\ 0 & \boxed{\begin{matrix} b_{22} \cdots b_{2n} \\ \vdots \\ b_{k2} \cdots b_{kn} \end{matrix}} \\ \vdots & \\ 0 & \end{bmatrix}.$$

Now consider the matrix enclosed in the box, and start the reduction process over again, applying it now to this smaller matrix.

(b) If each entry in the first column is zero, the matrix is in the form

$$\begin{bmatrix} 0 & \boxed{\begin{matrix} a_{12} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k2} & \cdots & a_{kn} \end{matrix}} \\ \vdots & \\ 0 & \end{bmatrix},$$

and nothing needs to be done. Consider the matrix enclosed in the box, and start the reduction process over again, applying it now to this smaller matrix.

Example. Consider the matrix

$$\begin{bmatrix} 0 & 0 & 2 & 7 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -4 & 5 \\ -2 & 2 & -5 & 4 \end{bmatrix}.$$

Applying the reduction process to it, we obtain:

$$\begin{bmatrix} 0 & 0 & 2 & 7 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -4 & 5 \\ -2 & 2 & -5 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 0 & 2 & 7 \\ -1 & 1 & -4 & 5 \\ -2 & 2 & -5 & 4 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & \boxed{\begin{matrix} 0 & 2 & 7 \\ 0 & -3 & 6 \\ 0 & -3 & 6 \end{matrix}} \\ 0 & \\ 0 & \end{bmatrix} \text{ by (a)}, \quad \to \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 0 & \boxed{\begin{matrix} 2 & 7 \\ -3 & 6 \\ -3 & 6 \end{matrix}} \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix} \text{ by (b)},$$

$$\to \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & \boxed{\begin{matrix} \frac{33}{2} \\ \frac{33}{2} \end{matrix}} \\ 0 & 0 & 0 & \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & \frac{33}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ by (a)}.$$

REMARK. The reduction process eventually stops, because the matrix enclosed by the box gets smaller at each step. At the end of the process the matrix will be in what is called *row-echelon* (stairstep) *form,* of which the following is an example:

$$\begin{bmatrix} \surd & & & & & & & \\ 0 & 0 & \surd & & & & & \\ 0 & 0 & 0 & \surd & & & & \\ 0 & 0 & 0 & 0 & 0 & \surd & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \surd & \cdots \\ & \cdots & & & & & & \end{bmatrix}.$$

All the entries lying beneath the stairsteps are zeros, and the entries at each of the points marked $\surd$ are nonzero, because if such an entry were zero, we would have drawn the stairsteps so that this entry lay beneath the stairs. We will call these entries we have checked the *corner entries* of the row-echelon form.

Note that up to now we have used only the elementary row operations (1) and (3).

**Second Part of the Reduction Process.** We start with a matrix in row-echelon form.

By several applications of operations of type (2), bring the matrix to the form where each corner entry is equal to 1. Then by multiplying each row by appropriate scalars and adding it to the rows lying *above* it, bring the matrix to the form in which each entry which lies directly above a corner entry is 0.

Note that these operations do not affect the position of the stairsteps, nor do they affect any of the zeros lying beneath the stairsteps. At the conclusion of this part of the process, the matrix will be in what is called *reduced row-echelon form.*

EXAMPLE. If we apply the second part of the reduction process to the example considered above, we obtain the matrix

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

**5.2 Theorem.** If a matrix $B$ is in reduced row-echelon form, its nonzero rows form a basis for its row space.

*Proof.* We know that the rows of $B$ span the row space of $B$, so that the nonzero rows also span this row space. Hence we need only to show that they are independent. But no one of them can equal a linear combination of the others, because each row has an entry of 1 in some position where the other rows all have entries of 0.

For example, the third row $\mathbf{a}_3$ of the matrix

$$\begin{bmatrix} 1 & 0 & -2 & 0 & 3 \\ 0 & 1 & 7 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

cannot possibly equal a linear combination of the others, because the general linear combination of the first two rows $\mathbf{a}_1$ and $\mathbf{a}_2$,

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 = [c_1,\ c_2,\ -2c_1 + 7c_2,\ 0,\ 3c_1 + 2c_2],$$

necessarily has a 0 as its fourth component, whereas $\mathbf{a}_3$ has a 1 as its fourth component.

**5.3 Corollary.** If a matrix is in reduced row-echelon form, the rank of the matrix equals the number of nonzero rows.

## EXERCISES

1. Carry out the reductions indicated:

(a) $$\begin{bmatrix} 0 & 1 & 2 & 0 & 1 \\ 0 & 2 & 4 & 1 & 1 \\ 0 & 3 & 6 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(b) $$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 \\ 0 & 1 & 3 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2. Find a basis for the subspace $W$ of $V^3$ spanned by

$$[1,\ -1,\ 0],\quad [0,\ 1,\ 1],\quad [-1,\ 2,\ 1],\quad \text{and}\quad [-1,\ 3,\ 2].$$

3. Find a basis for the subspace $W$ of $V^5$ spanned by

$$[2,\ 4,\ 5,\ 0,\ 3],\quad [1,\ 3,\ 3,\ -2,\ -1],\quad [-1,\ -1,\ -2,\ 1,\ 3],$$
$$[0,\ 2,\ 1,\ -1,\ 0],\quad \text{and}\quad [-2,\ 0,\ -3,\ 1,\ 5].$$

4. Find a basis for the subspace $W$ of $V^6$ consisting of all vectors of the form

$$[a + b, b - 2c, a - b + c, b, a - c, 2a + c],$$

where $a$, $b$, and $c$ are arbitrary scalars.

5. Consider the matrices

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

The second matrix is called the *transpose* of the first, since it is obtained by flipping the first matrix over so that all the rows become columns. Find the rank of each of these matrices. Then, for each of the matrices in Exercise 1, compare its rank with the rank of its transpose. Can you make a conjecture here?

6. Show that at the end of the first part of the reduction process, when the matrix is in row-echelon form, the nonzero rows already form a basis for the row space.

7. Given a matrix $A$, our row-reduction procedure enables us to construct a basis for the row space of $A$, but usually the basis we obtain will consist of vectors that are different from those which form the rows of $A$. Devise a procedure for choosing a basis for the row space of $A$ from among the rows of $A$.

8. (a) Find a subset of the spanning set given in Exercise 2 which forms a basis for the space $W$. Is there more than one such subset?

(b) Do the same for the spanning set in Exercise 3.

(c) Do the same for the rows of the matrix in the first example of this section.

## §6. HOMOGENEOUS SYSTEMS

We now consider the second problem mentioned at the beginning of §5. Suppose a subspace $W$ is specified as the solution space of a homogeneous system of linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0, \\ &\vdots \\ a_{k1}x_1 + \cdots + a_{kn}x_n &= 0. \end{aligned}$$

Can we determine the dimension of $W$, and can we find a basis for $W$?

We are going to use a procedure similar to that in 3.1, but now we are interested in finding *all* solutions of this system. Again we use the familiar process of "eliminating unknowns"; this process involves forming a new system of equations by multiplying some of these equations by scalars, and by adding or subtracting them from each other. What we are going to do here is to describe a procedure for carrying out this elimination process which is both efficient and systematic.

First of all, there is no need to copy down the equations each time. It saves labor if, instead, we form the *coefficient matrix*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix}$$

of the system, and work with the rows of this matrix rather than with the equations. This simply saves copying the $x_i$ over and over again.

Now our elementary row operations, if applied to this matrix, correspond respectively to (1) interchanging two of the equations of the system, (2) multiplying both sides of an equation by a nonzero scalar $c$, and (3) replacing an equation by itself plus a scalar multiple of some other equation. Then it is clear that any solution of our given system of equations will also satisfy the system which corresponds to the new matrix. *Because* the elementary row operations are reversible, the converse is also true. This fact is summed up in the following theorem:

**6.1 Theorem.** Suppose $B$ is a matrix obtained by applying an elementary row operation to the matrix $A$. If $A$ and $B$ are the coefficient matrices of two homogeneous systems of linear equations, these systems have precisely the same solutions.

**The Solution Procedure.** Given a homogeneous system of linear equations, we form the coefficient matrix of the system, and then apply elementary row operations to this matrix so as to bring it into reduced row-echelon form. At this point we claim it is easy to determine a basis for the solution space of the system. Instead of proving this fact in general, we shall illustrate it by means of an example.

Suppose the reduced row-echelon form of the coefficient matrix is:

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

The corresponding system of equations is

$$\begin{aligned}
x_2 \quad + 2x_4 \quad - 3x_6 \quad &= 0,\\
x_3 - \ x_4 \quad + 2x_6 \quad &= 0,\\
x_5 + \ x_6 \quad &= 0,\\
x_7 &= 0.
\end{aligned}$$

Now each of the unknowns $x_2$, $x_3$, $x_5$ and $x_7$ corresponding to a corner entry of the matrix appears in only *one* equation of the system. Hence we may solve for these unknowns in terms of the other unknowns. As a result, the solution space of the system necessarily consists of all vectors $[x_1, \ldots, x_7]$ in $V^7$ such that

$$\begin{aligned}
x_2 &= -2x_4 + 3x_6,\\
x_3 &= \quad x_4 - 2x_6,\\
x_5 &= \qquad\quad - x_6,\\
x_7 &= 0.
\end{aligned}$$

The most general vector which satisfies these four equations is the vector

$$[x_1,\ -2x_4 + 3x_6,\ x_4 - 2x_6,\ x_4,\ -x_6,\ x_6,\ 0].$$

Said differently, for every assignment of specific real values to $x_1$, $x_4$, and $x_6$ in this expression, we get a solution; and conversely, every solution comes from such an assignment. Hence we can say that this is the *general solution* of the system of equations.

However, this is not the best form in which to write the general solution. It is better to separate out the parts involving $x_1$, $x_4$, and $x_6$ and write it as

$$[x_1, 0, 0, 0, 0, 0, 0] + [0, -2x_4, x_4, x_4, 0, 0, 0] + [0, 3x_6, -2x_6, 0, -x_6, x_6, 0],$$

or as

$$x_1[1, 0, 0, 0, 0, 0, 0] + x_4[0, -2, 1, 1, 0, 0, 0] + x_6[0, 3, -2, 0, -1, 1, 0].$$

When the general solution is written in this form, we see that since $x_1$, $x_4$, and $x_6$ are arbitrary scalars, the general solution of the system is the general linear combination of the three vectors

$$[1, 0, 0, 0, 0, 0, 0], \quad [0, -2, 1, 1, 0, 0, 0], \quad \text{and} \quad [0, 3, -2, 0, -1, 1, 0].$$

Hence these three vectors span the solution space of the system. We claim even more—they are a *basis* for the solution space. To show this, we must show that they are independent. Now each has an entry of 1 in a position where the other two have entries of 0: The vector whose coefficient is $x_1$ has such an entry as its 1st component; the vector whose coefficient is $x_4$ has such an entry as its 4th component; and the vector whose coefficient is $x_6$ has such an entry as its 6th component. Thus no one of them equals a linear combination of the others, so they must be independent.

Because $x_1$, $x_4$, and $x_6$ are arbitrary scalars, it is customary to replace them by $c_1$, $c_2$, and $c_3$, and to write the general solution of the system in the form

$$c_1[1, 0, 0, 0, 0, 0, 0] + c_2[0, -2, 1, 1, 0, 0, 0] + c_3[0, 3, -2, 0, -1, 1, 0].$$

In this form, $c_1$, $c_2$, $c_3$ are often thought of as parameters, and one says that the general solution depends on three parameters or "arbitrary constants." They are said to be *independent parameters*, because the vectors by which they are multiplied form an independent set.

Let us look at the dimensions involved in this example. The coefficient matrix of our system has rank 4, since it is in reduced row-echelon form and has four nonzero rows. There are seven unknowns, but only three of them may be specified arbitrarily, because the four equations determine the remaining unknowns in terms of these. We have thus the fact that

$$\underset{7}{\text{number of unknowns}} - \underset{4}{\text{rank of matrix}} = \underset{3}{\text{dimension of solution space.}}$$

A similar fact holds in general; it is expressed in the following theorem:

> **6.2 Theorem.** Given a homogeneous system of $k$ linear equations in $n$ unknowns. If the rank of the coefficient matrix is $r$, then the dimension of the solution space is $n - r$.

*Proof.* Form the coefficient matrix of the system, and by means of elementary row operations bring it to reduced row-echelon form. These operations do not affect the rank $r$ of the matrix (Theorem 5.1) nor do they affect the solution space (Theorem 6.1). Hence it suffices to prove the theorem in the special case in which the coefficient matrix is in reduced row-echelon form. This proof proceeds exactly as in the example worked out above, and so we leave it as an exercise.

---

It is sometimes mistakenly thought that the general solution of a homogeneous system of $k$ linear equations in $n$ unknowns necessarily will involve $n - k$ arbitrary constants. The preceding analysis shows that it is not the *number* of equations that matters, but the number $r$ of *independent* equations. (A homogeneous system of linear equations is said to be independent if the rows of the coefficient matrix form an independent set.) If one has a homogeneous system of $r$ *independent* linear equations in $n$ unknowns, then the general solution will involve $n - r$ arbitrary constants.

## EXERCISES

1. By reducing to row-echelon form, find a basis for the solution spaces $W$ of the systems:

(a)
$$\begin{aligned} x_1 - x_2 + x_3 - x_4 &= 0, \\ x_1 \phantom{{}-x_2} - x_3 + x_4 &= 0, \\ x_1 + x_2 \phantom{{}+x_3} - 2x_4 &= 0. \end{aligned}$$

(b)
$$\begin{aligned} x_1 \phantom{{}+x_2} + 2x_3 \phantom{{}+x_4+x_5} - x_6 &= 0, \\ x_2 \phantom{{}+x_3} - 5x_4 \phantom{{}+x_5+x_6} &= 0, \\ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 &= 0. \end{aligned}$$

2. Find the general solution of the system

$$\begin{aligned} x_1 + x_2 - x_3 + 2x_4 &= 0, \\ 3x_2 - x_3 + 3x_4 &= 0, \\ 2x_1 - x_2 - x_3 + x_4 &= 0. \end{aligned}$$

3. Complete the proof of Theorem 6.2.

4. Consider the system of five equations in seven unknowns which was solved as an example in this section. We exhibited the general solution as the general linear combination of the three independent vectors $[1, 0, 0, 0, 0, 0, 0]$, $[0, -2, 1, 1, 0, 0, 0]$, and $[0, 3, -2, 0, -1, 1, 0]$. Show that this is not the only form in which the general solution may be written; specifically, show that any vector of the form

$$\begin{aligned} c_1[1, -4, 2, 2, 0, 0, 0] &+ c_2[1, -1, 1, -1, 1, -1, 0] \\ &+ c_3[1, 1, -1, 1, -1, 1, 0] \end{aligned}$$

is a solution of the system, and that every solution of the system is equal to a vector of this form.

5. Show that any vector of the form

$$\begin{aligned} c_1[1, -2, 1, 1, 0, 0, 0] &+ c_2[0, 1, -1, 1, -1, 1, 0] \\ &+ c_3[-1, 4, -3, 1, -2, 2, 0] \\ &+ c_4[1, -5, 3, 1, 1, -1, 0] \end{aligned}$$

is a solution of the system considered in Exercise 4, and that every solution is equal to a vector of this form. How do you explain the fact that there are four arbitrary constants here?

6. Find two forms which involve none of the same vectors for the general solution of Exercise 2.

7. Let $W$ be a subspace of $V^n$. Let $W_1$ be the set of all vectors $\mathbf{a}$ of $V^n$ such that $\mathbf{a}$ is orthogonal to every vector of $W$. Show that $W_1$ is a subspace of $V^n$. It is called the *orthogonal complement* of $W$.

8. Let $W$ be a subspace of $V^n$; let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be a spanning set for $W$. Show that if $\mathbf{a}$ is orthogonal to each of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$, then $\mathbf{a}$ lies in the orthogonal complement of $W$.

9. Find a basis for the orthogonal complement of the subspace of $V^3$ spanned by $[1, -1, 2]$. Interpret your result geometrically.

Find a basis for the orthogonal complement of:

10. The subspace $W$ in Exercise 2 of §5.
11. The space $W$ in Exercise 3 of §5.
12. The space $W$ in Exercise 4 of §5.
13. The spaces $W$ in Exercise 1 of this section.
14. Given a homogeneous system of equations whose coefficient matrix is $A$. Show that the solution space of the system is the orthogonal complement of the row space of $A$.
15. If $W$ is a subspace of $V^n$ and $W$ has dimension $p$, what is the dimension of the orthogonal complement of $W$?
16. Prove that if $W$ is a subspace of $V^n$, then $W$ is the solution space of some homogeneous system of linear equations.

## §7. THE GENERAL SYSTEM

We now consider the general system of $k$ linear equations in $n$ unknowns. Such a system need not have a solution; for example, the system

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

clearly has no solution, since the sum of the first two equations contradicts the third. On the other hand, such a system may have many solutions. We shall study this situation in some detail; and then we shall give a practical procedure for finding the general solution of a system of linear equations.

**Definition.** Suppose we are given a $k$ by $n$ matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix}.$$

If $\mathbf{x} = [x_1, \ldots, x_n]$ is any vector in $V^n$, let $A(\mathbf{x})$ denote the vector $\mathbf{b} = [b_1, \ldots, b_k]$ of $V^k$ for which the following equations hold:

$$(*) \qquad \begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1, \\ &\vdots \\ a_{k1}x_1 + \cdots + a_{kn}x_n &= b_k. \end{aligned}$$

In this way, we consider $A$ as a *function* which assigns to each vector $\mathbf{x}$ of $V^n$ a vector $\mathbf{b}$ of $V^k$, by the rule just given.

The function we have defined is often called a *linear transformation.* Its most important property is expressed in the following statement:

If $A(\mathbf{x}) = \mathbf{b}$ and $A(\mathbf{y}) = \mathbf{c}$, then

(a) $A(\mathbf{x} + \mathbf{y}) = \mathbf{b} + \mathbf{c}$, and
(b) $A(d\mathbf{x}) = d\mathbf{b}$, for any scalar $d$.

---

To check this property, we note that by hypothesis,

$$\begin{array}{lcl} a_{11}x_1 + \cdots + a_{1n}x_n = b_1, & & a_{11}y_1 + \cdots + a_{1n}y_n = c_1, \\ \vdots & \text{and} & \vdots \\ a_{k1}x_1 + \cdots + a_{kn}x_n = b_k & & a_{k1}y_1 + \cdots + a_{kn}y_n = c_k. \end{array}$$

If we add corresponding equations of these two systems, we get the system

$$\begin{array}{l} a_{11}(x_1 + y_1) + \cdots + a_{1n}(x_n + y_n) = b_1 + c_1, \\ \vdots \\ a_{k1}(x_1 + y_1) + \cdots + a_{kn}(x_n + y_n) = b_k + c_k, \end{array}$$

which states that $A(\mathbf{x} + \mathbf{y}) = \mathbf{b} + \mathbf{c}$, as desired. Similarly, multiplying through the first system by $d$ gives us the system

$$\begin{array}{l} a_{11}(dx_1) + \cdots + a_{1n}(dx_n) = db_1, \\ \vdots \\ a_{k1}(dx_1) + \cdots + a_{kn}(dx_n) = db_k, \end{array}$$

which states that $A(d\mathbf{x}) = d\mathbf{b}$.

---

REMARK. The solution space of the system of homogeneous equations having $A$ as coefficient matrix may be defined, in these new terms, as the set of all $\mathbf{x}$ in $V^n$ such that $A(\mathbf{x}) = \mathbf{0}$. Another space associated with $A$ is given in the following definition:

**Definition.** Given a $k$ by $n$ matrix $A$, consider the set $W$ of all vectors $\mathbf{b}$ of $V^k$ such that $A(\mathbf{x}) = \mathbf{b}$ for some $\mathbf{x}$ in $V^n$; this set is called the *range space* of the linear transformation $A$. Stated differently, the range space of $A$ consists of those vectors $[b_1, \ldots, b_k]$ for which the system $(*)$ above has a solution.

Let us check that $W$ is a subspace of $V^k$. If $\mathbf{b}$ and $\mathbf{c}$ are in $W$, then $A(\mathbf{x}) = \mathbf{b}$ for some $\mathbf{x}$ and $A(\mathbf{y}) = \mathbf{c}$ for some $\mathbf{y}$. Then $A(\mathbf{x} + \mathbf{y}) = \mathbf{b} + \mathbf{c}$, and $A(d\mathbf{x}) = d\mathbf{b}$, so that $\mathbf{b} + \mathbf{c}$ and $d\mathbf{b}$ belong to $W$, as desired.

EXAMPLE. Consider the system

$$\begin{aligned} x_1 - x_2 + \phantom{2}x_3 &= b_1, \\ x_1 \phantom{- x_2} + 2x_3 &= b_2, \\ 2x_1 - x_2 + 3x_3 &= b_3. \end{aligned}$$

It does not have any solution unless $b_1 + b_2 = b_3$. However, if this condition is satisfied, the system does have a solution. (One such solution is the vector $\mathbf{x} = [b_2, b_2 - b_1, 0]$, but there are others.) Thus, in this case, the range space of the coefficient matrix consists *precisely* of all vectors $[b_1, b_2, b_3]$ of $V^3$ such that $b_3 = b_1 + b_2$. Note that the dimension of the range space is 2, and the rank of the coefficient matrix is also 2. This is not accidental, as the following theorem shows.

**7.1 Theorem.** If $A$ is a matrix of rank $r$, the dimension of the range space of $A$ equals $r$.

*Proof.* Suppose $A$ is of size $k$ by $n$. Let $W_1$ be the set of all vectors $\mathbf{x}$ in $V^n$ such that $A(\mathbf{x}) = \mathbf{0}$. If $W_1$ has dimension $p$, then by Theorem 6.2, $p = n - r$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_p$ be a basis for $W_1$. Apply Theorem 4.5 to choose vectors $\mathbf{a}_{p+1}, \ldots, \mathbf{a}_n$ in $V^n$ in such a way that the set $\mathbf{a}_1, \ldots, \mathbf{a}_n$ forms a basis for $V^n$. Now $A(\mathbf{a}_1) = \mathbf{0}, \ldots, A(\mathbf{a}_p) = \mathbf{0}$ by hypothesis; let

$$\mathbf{d}_{p+1} = A(\mathbf{a}_{p+1}), \ldots, \mathbf{d}_n = A(\mathbf{a}_n).$$

Our claim is that the vectors $\mathbf{d}_{p+1}, \ldots, \mathbf{d}_n$ form a basis for the range space of $A$. This will prove our theorem, since there are $n - p$ vectors in this set, and we know $p = n - r$.

First we show that the vectors $\mathbf{d}_{p+1}, \ldots, \mathbf{d}_n$ are independent. This is easy, for if $c_{p+1}\mathbf{d}_{p+1} + \cdots + c_n\mathbf{d}_n = \mathbf{0}$, then

$$A(c_{p+1}\mathbf{a}_{p+1} + \cdots + c_n\mathbf{a}_n) = c_{p+1}\mathbf{d}_{p+1} + \cdots + c_n\mathbf{d}_n = \mathbf{0},$$

by the basic properties (a) and (b) of the linear transformation $A$. This means, by definition, that the vector $c_{p+1}\mathbf{a}_{p+1} + \cdots + c_n\mathbf{a}_n$ is in the space $W_1$, so that it equals a linear combination of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$. But this is impossible unless all the coefficients vanish, because the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ form an independent set.

We now show that the vectors $\mathbf{d}_{p+1}, \ldots, \mathbf{d}_n$ span the range space of $A$. If $\mathbf{b}$ is any vector in the range space of $A$, then $\mathbf{b} = A(\mathbf{x})$ for some $\mathbf{x}$ in $V^n$, by definition. The vector $\mathbf{x}$ equals some linear combination $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$, since the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ span $V^n$. If we compute $A(\mathbf{x})$, using properties (a) and (b), we see that

$$A(\mathbf{x}) = \mathbf{b} = c_1\mathbf{0} + \cdots + c_p\mathbf{0} + c_{p+1}\mathbf{d}_{p+1} + \cdots + c_n\mathbf{d}_n,$$

which gives us our desired result.

**7.2 Corollary.** Let $A$ be a $k$ by $n$ matrix of rank $r$. If $r < k$, there is a vector $\mathbf{b} = [b_1, \ldots, b_k]$ such that the system $A(\mathbf{x}) = \mathbf{b}$ has no solution. If $r = k$, the system has a solution for every choice of $\mathbf{b}$.

*Proof.* The dimension of the range space of $A$ is $r$; if $r < k$, the range space $W$ is not all of $V^k$, so we may choose a vector $\mathbf{b}$ in $V^k$ not lying in $W$.

The corresponding system has no solution. If $r = k$, then the range space is all of $V^k$, so the system has a solution for every $\mathbf{b}$ in $V^k$.

**7.3 Corollary.** The rank of a matrix $A$ equals the rank of the transpose $B$ of $A$. (See Exercise 5 of §5.)

*Proof.* Let $A$ be of size $k$ by $n$; let $W$ be the range space of $A$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the natural basis for $V^n$ (see Exercise 10 of §4). We claim that the vectors $A(\mathbf{e}_1), \ldots, A(\mathbf{e}_n)$ necessarily span $W$. To prove this, suppose $\mathbf{b} = A(\mathbf{x})$ for some $\mathbf{x}$ in $V^n$; let $x = [x_1, \ldots, x_n]$. Then

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, \qquad \text{and} \qquad \mathbf{b} = x_1A(\mathbf{e}_1) + \cdots + x_nA(\mathbf{e}_n),$$

as desired.

If we compute $A(\mathbf{e}_1)$, we find that it is just the $k$-tuple whose components are the entries in the first column of $A$. Said differently, $A(\mathbf{e}_1)$ is the first *row* of the transposed matrix $B$. Similarly, $A(\mathbf{e}_2)$ is the second row of $B$, and so on. Thus the vectors $A(\mathbf{e}_1), \ldots, A(\mathbf{e}_n)$ span the row space of $B$. But they also span the range space of $A$, which by Theorem 7.1 has dimension equal to the rank of $A$.

---

We may rephrase Corollary 7.3 as follows: If $A$ is a matrix, then the dimension of the row space of $A$ equals the dimension of the column space of $A$. That is, if $r$ of its rows form an independent set, then so do $r$ of its columns, and conversely. This fact is one of the more difficult theorems of linear algebra; you may appreciate its difficulty better if you try to prove it directly, using just the definition of linear independence!

---

Given a system of $k$ linear equations in $n$ unknowns, we know that the system need not have a solution. Nevertheless, we can still ask the following question: If the system does have a solution, does it have more than one, and if so, what form does the general solution take?

The answer to this question is easy. Suppose we are given $A$ and $\mathbf{b}$, and we know that the vector $\mathbf{d}$ is one solution of the system $A(\mathbf{x}) = \mathbf{b}$. Then for any other solution $\mathbf{x}$, we have the fact that $A(\mathbf{x} - \mathbf{d}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Hence the vector $\mathbf{y} = \mathbf{x} - \mathbf{d}$ must be a solution of the *related homogeneous system* $A(\mathbf{y}) = \mathbf{0}$. The converse also holds; if $\mathbf{y}$ is a solution of the system $A(\mathbf{y}) = \mathbf{0}$, then $\mathbf{d} + \mathbf{y}$ is a solution of the system $A(\mathbf{x}) = \mathbf{b}$. We have thus proved the following theorem:

**7.4 Theorem.** Given a system of linear equations which has a particular solution $\mathbf{d}$. Let $c_1\mathbf{y}_1 + \cdots + c_p\mathbf{y}_p$ be the general solution of the related homogeneous system, where $c_1, \ldots, c_p$ are arbitrary scalars. Then $\mathbf{x} = \mathbf{d} + c_1\mathbf{y}_1 + \cdots + c_p\mathbf{y}_p$ is the general solution of the given system of linear equations.

**Finding the General Solution of a System of Linear Equations.** The preceding discussion seems to suggest that solving a system of linear equations should involve three steps: first determining whether the system has a solution; if so, finding a particular solution of the system, and then finding the general solution of the related homogeneous system. In practice, however, we do not do this at all. Instead, we find it more efficient to carry out all three steps at once.

The idea, as before, is to apply addition and subtraction operations until enough of the unknowns are eliminated so that the general solution can be written down. Again, instead of performing these operations on the equations themselves, we perform them on the rows of an appropriate matrix. In the present situation, we take the coefficient matrix of the system and adjoin to it an extra column consisting of the constants of the system:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{k1} & \cdots & a_{kn} & b_k \end{bmatrix}.$$

This matrix is called the *augmented matrix* of the system. Applying an elementary row operation to this matrix corresponds to (1) exchanging two equations, (2) multiplying both sides of an equation by a nonzero scalar $c$, or (3) replacing an equation by itself plus a scalar multiple of another equation. For the same reasons as before, the system corresponding to the new matrix will have precisely the same solutions as the system corresponding to the old matrix.

Now we apply the elementary row operations until we bring the part of the matrix which is the coefficient matrix into reduced row-echelon form. (We do not bother to reduce the column of constants.) At this point we can determine whether the system has a solution, and if it does, we can write down the general solution immediately.

EXAMPLE. Suppose we have a system of three equations in four unknows whose augmented matrix is reduced by elementary row operations to the form

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

The part of this matrix which is the coefficient matrix is then in reduced row-echelon form. The system corresponding to this matrix clearly has no solution, for the third equation of the system is

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 2.$$

EXAMPLE. Consider a system of four equations in six unknowns whose augmented matrix reduces to the following form:

$$\begin{bmatrix} 1 & 0 & 3 & 7 & 0 & -2 & 5 \\ 0 & 1 & 0 & -4 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Although there is a final row of zeros in the coefficient matrix, there is also a corresponding zero in the column of constants, so that the final equation gives us no trouble. We may now write down the general solution at once. For, as in the homogeneous case, the unknowns $x_1$, $x_2$, and $x_5$ each appear in only *one* equation of the system, so we can solve for them in terms of the other unknowns:

$$\begin{aligned} x_1 &= 5 - 3x_3 - 7x_4 + 2x_6, \\ x_2 &= 4x_4 - x_6, \\ x_5 &= -1 + 3x_6, \end{aligned}$$

so that the general solution is the vector

$$[5 - 3x_3 - 7x_4 + 2x_6,\ 4x_4 - x_6,\ x_3,\ x_4,\ -1 + 3x_6,\ x_6].$$

The best way to express the general solution is as follows:

$$\begin{aligned} \mathbf{x} = [x_1, \cdots, x_6] &= [5, 0, 0, 0, -1, 0] + c_1[-3, 0, 1, 0, 0, 0] \\ &\quad + c_2[-7, 4, 0, 1, 0, 0] + c_3[2, -1, 0, 0, 3, 1], \end{aligned}$$

where we have replaced $x_3$, $x_4$, and $x_6$ by the parameters $c_1$, $c_2$, and $c_3$, respectively. In this form, the general solution is exhibited as the sum of a particular solution of the system and the general solution of the related homogeneous system. It involves three independent parameters.

---

It should be clear now how to use this procedure to solve any system of $k$ linear equations in $n$ unknowns: First form the augmented matrix of the system. Then apply elementary row operations until the coefficient matrix is brought into reduced row-echelon form. If there is any row in which the last entry $b_i$ is the only nonzero one, the system has no solution, for that row corresponds to the equation

$$0x_1 + 0x_2 + \cdots + 0x_n = b_i.$$

If there is no such row, the system has a solution, and one finds the general solution by solving for the unknowns corresponding to the corner entries of the matrix in terms of the remaining unknowns.

## EXERCISES

1. Find the general solution of the following systems of linear equations:

(a) $$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1, \\ x_2 - x_3 + x_4 &= 0, \\ x_1 - x_2 + 2x_3 &= 0. \end{aligned}$$

(b) $$\begin{aligned} x_1 - x_2 - 2x_3 + 3x_4 &= 2, \\ x_1 + 3x_2 + 4x_3 - x_4 &= 1, \\ x_1 + x_2 + x_3 + x_4 &= 1. \end{aligned}$$

(c) $$\begin{aligned} x_1 - x_2 + x_3 + x_4 &= 1, \\ -3x_2 + 5x_3 + 2x_4 &= 3, \\ 2x_1 + x_2 - 3x_3 &= -1. \end{aligned}$$

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

(e) $$\begin{aligned} x_1 + x_2 - x_3 &= 1, \\ x_1 - x_2 + x_3 &= 0, \\ -x_1 + 2x_2 + x_3 &= -1, \\ x_1 + 7x_2 - 3x_3 &= 2. \end{aligned}$$

2. Consider the system of four equations in six unknowns which was solved as the last example of this section. Show that

$$\mathbf{x} = [1, 4, -1, 1, -1, 0] + c_1[4, -4, 1, -1, 0, 0] + c_2[-1, -1, 1, 0, 3, 1] + c_3[0, 5, -3, 1, -3, -1]$$

is another form of the general solution of this system.

3. Show that every vector of the form

$$\mathbf{x} = [0, 3, 0, 1, 2, 1] + c_1[4, 1, -2, 0, -3, -1] + c_2[4, -4, 1, -1, 0, 0]$$

is a solution of the system considered in Exercise 2. Is every solution of the system equal to a vector of this form?

4. Show that every vector of the form

$$\mathbf{x} = [0, 3, 0, 1, 2, 1] + c_1[4, -4, 1, -1, 0, 0] + c_2[0, 5, -3, 1, -3, -1] + c_3[4, 6, -5, 1, -6, -2]$$

is a solution of the system considered in Exercise 2. Is every solution of the system equal to a vector of this form?

5. For each of the systems given in Exercise 1, find two forms for the general solution (whenever this is possible).

6. Choose constants which make the following systems have no solution, if it is possible to do so:

(a) $$\begin{aligned} x_1 - x_2 + x_3 + x_4 &= b_1, \\ x_2 - x_3 + 2x_4 &= b_2, \\ x_1 - x_2 - x_3 &= b_3. \end{aligned}$$

(b) $$\begin{aligned} x_1 + x_2 + 2x_3 - 2x_4 &= b_1, \\ x_1 + x_2 + \tfrac{3}{2}x_3 - x_4 &= b_2, \\ -x_2 - x_3 + x_4 &= b_3, \\ 2x_1 + x_3 + 3x_4 &= b_4. \end{aligned}$$

7. Prove the following theorems:

**Theorem (a).** Given a system of $n$ linear equations in $n$ unknowns. If the rank of the coefficient matrix equals $n$, then the system has one and only one solution.

**Theorem (b).** Given a homogeneous system of $n$ linear equations in $n$ unknowns. If the rank of the coefficient matrix is less than $n$, the system has infinitely many solutions.

8. Prove the following theorem:

**Theorem.** A system of linear equations has a solution if and only if the rank of the augmented matrix equals the rank of the coefficient matrix.

9. Find conditions on the constants in the system in Exercise 6(a) which are both necessary and sufficient for the existence of a solution. Do the same for Exercise 6(b).

## §8. DETERMINANTS

The determinant is a certain function which assigns, to every square matrix $A$, a real number which we denote by $\det A$. This function has a number of interesting properties, of which the following are important for us:

(1) If $A'$ is obtained from $A$ by interchanging two rows, then $\det A' = -\det A$.

(2) If $A'$ is obtained from $A$ by multiplying each entry in row $i$ by the nonzero scalar $c$, then $\det A' = c \det A$.

(3) If $A'$ is obtained from $A$ by replacing row $i$ by row $i$ plus $d$ times row $j$, then $\det A' = \det A$.

(4) If $A$ has a row of zeros, then $\det A = 0$.

(5) If $A$ is the matrix

$$\begin{bmatrix} 1 & 0 & 0 \cdots 0 \\ 0 & 1 & 0 \cdots 0 \\ 0 & 0 & 1 \cdots 0 \\ \vdots & & \vdots \\ 0 & 0 & 0 \cdots 1 \end{bmatrix},$$

then $\det A = 1$. This matrix is called the *identity matrix*.

For the time being, we assume these properties of the determinant function. In the final part of this section, the definition of the determinant is given and these properties are derived from it.

**Definition.** Given a matrix, one can obtain from it a number of matrices, of various sizes, merely by deleting some rows and/or columns. We shall call a matrix obtained in this way a *submatrix* of the original matrix.

EXAMPLE. The matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \qquad \text{has the submatrix} \qquad \begin{bmatrix} 5 & 7 \\ 9 & 11 \end{bmatrix},$$

obtained by deleting the first row and the second and fourth columns. It has in all eighteen submatrices of size 2 by 2. It has three submatrices of size 2 by 4 and twelve of size 1 by 1.

**8.1 Theorem.** Let $B$ be a matrix of size $p$ by $n$. The rows of $B$ form an independent set if and only if $B$ contains some submatrix of size $p$ by $p$ whose determinant is nonzero.

*Proof.* Given the matrix $B$, suppose we apply some elementary row operations to it, obtaining a new matrix $B'$, and then we delete some of the columns of $B'$, obtaining a $p$ by $p$ submatrix $C'$. Because of the way elementary row operations work, we would have achieved precisely the same result if we first had deleted the corresponding columns from $B$, obtaining a submatrix $C$, and then had applied those same elementary row operations to $C$. Hence det $C'$ must equal a nonzero scalar multiple of det $C$; this follows from properties (1), (2), and (3).

Now we choose elementary row operations to be applied to $B$ so as to bring it into reduced row-echelon form $B'$.

Suppose the rows of $B$ are independent. Then the rank of $B$ equals $p$, so that $B'$ has no row of zeros, and there is a "corner entry" of 1 appearing in each row of $B'$. Let us form a $p$ by $p$ submatrix $C'$ of $B'$ by deleting all the columns of $B'$ which do not contain a corner entry. Then $C'$ is necessarily the identity matrix, so det $C' = 1$. If $C$ is the corresponding $p$ by $p$ submatrix of $B$, then det $C$ is nonzero, since it equals a nonzero scalar multiple of det $C'$.

To prove the converse, suppose $B$ contains a $p$ by $p$ submatrix $C$ whose determinant is nonzero. Let $B'$ be the reduced row-echelon form of $B$; let $C'$ be the submatrix of $B'$ corresponding to $C$. If $B'$ contained a row of zeros, $C'$ would contain a row of zeros also, and we would have det $C' = 0$. This cannot happen, because det $C'$ equals a nonzero scalar multiple of det $C$. Hence the rank of $B$ equals $p$, so the rows of $B$ form an independent set.

**8.2 Theorem.** Let $A$ be a matrix of size $k$ by $n$. Let the largest square submatrix of $A$ whose determinant is nonzero be of size $p$ by $p$. Then $p$ is equal to the rank of $A$.

*Proof.* Let $r$ be the rank of $A$. We show that $A$ contains a submatrix of size $r$ by $r$ whose determinant is nonzero, and that $A$ contains no larger such submatrix.

Since $A$ has rank $r$, it is possible to choose $r$ of the rows of $A$ which form an independent set. Let $B$ be the submatrix of $A$ obtained by deleting the remaining rows of $A$. Theorem 8.1 implies that $B$ contains an $r$ by $r$ submatrix whose determinant is nonzero.

On the other hand, suppose $C$ were an $m$ by $m$ submatrix of $A$ having nonzero determinant, where $m > r$. Let $B_1$ be the $m$ by $n$ submatrix of $A$ which contains $C$. By Theorem 8.1, the rows of $B_1$ form an independent set, so that the dimension of the row space of $A$ must be at least $m$. This contradicts the fact that the rank of $A$ is $r$.

**8.3 Corollary.** An $n$ by $n$ matrix $A$ has rank $n$ if and only if $\det A \neq 0$.

**Construction of the Determinant Function.** Suppose we take the positive integers $1, 2, \ldots, k$ and write them down in some arbitrary order, say $j_1, j_2, \ldots, j_k$. This new ordering is called a *permutation* of these integers. For each integer $j_i$ in this ordering, let us count how many integers *follow* it in this ordering, but *precede* it in the natural ordering $1, 2, \ldots, k$. This number is called the *number of inversions caused by* the integer $j_i$. If we determine this number for each integer $j_i$ in the ordering and add the results together, the number we get is called the total number of inversions which occur in this ordering. If the number is odd, we say the permutation is an *odd permutation*; if the number is even, we say it is an *even permutation*.

For example, consider the following reordering of the integers between 1 and 6:

$$2, 5, 1, 3, 6, 4.$$

If we count up the inversions, we see that the integer 2 causes one inversion, 5 causes three inversions, 1 and 3 cause no inversions, 6 causes one inversion, and 4 causes none. The sum is five, so the permutation is odd.

If a permutation is odd, we say the *sign* of that permutation is $-$; if it is even, we say its sign is $+$. A useful fact about the sign of a permutation is the following:

**8.4 Theorem.** If we interchange two adjacent elements of a permutation, we change the sign of the permutation.

*Proof.* Let us suppose the elements $j_i$ and $j_{i+1}$ of the permutation $j_1, \ldots, j_i, j_{i+1}, \ldots, j_k$ are the two we interchange, obtaining the permutation

$$j_1, \ldots, j_{i+1}, j_i, \ldots, j_k.$$

The number of inversions caused by the integers $j_1, \ldots, j_{i-1}$ clearly is the same in the new permutation as in the old one, and so is the number of inversions caused by $j_{i+2}, \ldots, j_k$. It remains to compare the number of inversions caused by $j_{i+1}$ and by $j_i$ in the two permutations.

*Case* I: $j_i$ precedes $j_{i+1}$ in the natural ordering $1, \ldots, k$. In this case, the number of inversions caused by $j_i$ is the same in both permutations, but the number of inversions caused by $j_{i+1}$ is one larger in the second permutation than in the first, for $j_i$ follows $j_{i+1}$ in the second permutation, but not in the first. Hence the total number of inversions is increased by one.

*Case* II: $j_i$ follows $j_{i+1}$ in the natural ordering $1, \ldots, k$. In this case, the number of inversion caused by $j_{i+1}$ is the same in both permutations, but the number of inversions caused by $j_i$ is one less in the second permutation than in the first.

In either case the total number of inversions changes by one, so that the sign of the permutation changes.

EXAMPLE. If we interchange the second and third elements of the permutation considered in the previous example, we obtain 2, 1, 5, 3, 6, 4, in which the total number of inversions is four, so the permutation is even.

**Definition.** Consider a $k$ by $k$ matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$

Pick out one entry from each row of $A$; do this in such a way that these entries all lie in different columns of $A$. Take the product of these entries,

$$a_{1j_1}a_{2j_2}a_{3j_3} \cdots a_{kj_k},$$

and prefix a $\pm$ sign according as the permutation $j_1, \ldots, j_k$ is even or odd. (Note that we arrange the entries in the order of the rows they come from, and then we compute the sign of the resulting permutation of the column indices.)

If we write down *all possible* such expressions and add them together, the number we get is defined to be the *determinant* of $A$.

REMARK. We apply this definition to the general 2 by 2 matrix, and obtain the formula

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

If we apply it to a 3 by 3 matrix, we find that

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{matrix} + a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\ - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{matrix}$$

The formula for the determinant of a 4 by 4 matrix involves 24 terms, and for a 5 by 5 matrix it involves 120 terms; we will not write down these formulas. The reader will readily believe that the definition we have given is not very useful for computational purposes!

The definition is, however, very convenient for theoretical purposes. For example, we can verify properties (2), (4), and (5) listed above immediately. To check (4), note that if $A$ has a row consisting entirely of zeros, then each term in the expansion of $\det A$ has a factor of zero in it, so that every term vanishes and $\det A = 0$. Similarly, if $A$ is the identity matrix, then every term has a factor of zero in it except for the term $a_{11}a_{22}a_{33} \cdots a_{kk} = 1 \cdot 1 \cdot 1 \cdots 1 = 1$, so that $\det A = 1$. Hence property (5) holds. To prove (2), note that multiplying row $i$ by the scalar $c$ has the effect of replacing the entry $a_{ij}$ by $ca_{ij}$, for each $j$. The general term in the expansion of $\det A'$ is then $\pm a_{1j_1}a_{2j_2} \cdots (ca_{ij_i}) \cdots a_{kj_k}$. Since this holds for each term in the expansion of $\det A'$, we see that $\det A' = c \det A$.

The other two properties require more work to prove.

**8.5 Theorem.** If $A'$ is obtained from $A$ by interchanging two rows, then $\det A' = - \det A$.

*Proof.* We consider first the case in which the two rows are adjacent to each other. Suppose they are rows $i$ and $i + 1$. Note that each term in the expansion of $\det A'$ also appears in the expansion of $\det A$, because we make *all possible* choices of one entry from each row and column when we write down this expansion. The only thing we have to do is to compare what signs this term has when it appears in the two expansions.

Let $a_{1j_1} \cdots a_{ij_i}a_{i+1,j_{i+1}} \cdots a_{kj_k}$ be a term in the expansion of $\det A$. If we look at the corresponding term in the expansion of $\det A'$, we see that we have the same factors, *but they are arranged differently*. For to compute the sign of this term, we agreed to arrange the entries in the order of the rows they came from, and then to take the sign of the corresponding permutation of the column indices. Thus in the expansion of det $A'$, this term will appear as

$$a_{1j_1} \cdots a_{i+1,j_i+1} a_{i,j_i} \cdots a_{kj_k}.$$

The permutation of the column indices here is the same as above except that elements $j_i$ and $j_{i+1}$ have been interchanged. By Theorem 8.4, this means that this term appears in the expansion of det $A'$ with the sign opposite to its sign in the expansion of det $A$.

Since this result holds for each term in the expansion of det $A'$, we have det $A' = -$ det $A$.

Now let us consider the more general case, in which we form $A'$ by interchanging rows $i$ and $j$. Suppose $i < j$; let the rows of $A$ be denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_k$. We can just as well obtain the matrix $A'$ by several exchanges of adjacent rows, as follows: First interchange $\mathbf{a}_i$ with $\mathbf{a}_{i+1}$, then interchange $\mathbf{a}_i$ with $\mathbf{a}_{i+2}$, and so on, until $\mathbf{a}_i$ is brought to a position just above $\mathbf{a}_j$. (This involves interchanging $\mathbf{a}_i$ with $\mathbf{a}_{i+1}, \ldots, \mathbf{a}_{j-1}$, so that $j - i - 1$ interchanges are involved.) Then interchange $\mathbf{a}_i$ with $\mathbf{a}_j$. Finally, interchange $\mathbf{a}_j$ successively with $\mathbf{a}_{j-1}, \mathbf{a}_{j-2}, \ldots, \mathbf{a}_{i+1}$; this involves $j - i - 1$ additional interchanges. At the end of this process we have the matrix $A'$. Since the process involved $2(j - i - 1) + 1$ interchanges of adjacent rows, the determinant changed signs an odd number of times, so that det $A' = -$ det $A$.

**8.6 Theorem.** If $A'$ is obtained by replacing row $i$ of $A$ by row $i$ plus $d$ times row $m$, then det $A' =$ det $A$.

*Proof.* The general term in the expansion of det $A$ is

$$(*) \qquad \pm\, a_{1j_1} a_{2j_2} \cdots a_{ij_i} \cdots a_{kj_k}.$$

The corresponding term in the expansion of det $A'$ may be obtained from this by replacing the factor $a_{ij_i}$ by the expression $(a_{ij_i} + d a_{mj_i})$. If we multiply out, we get the term $(*)$ plus $d$ times a copy of term $(*)$ in which the $i$th factor $a_{ij_i}$ has been replaced by the entry $a_{mj_i}$.

From this it follows that the determinant of $A'$ equals the determinant of $A$, plus $d$ times the determinant of a matrix $B$ which is obtained from $A$ by replacing row $i$ by row $m$. Note that $B$ is not formed by *interchanging* rows $i$ and $m$, but by erasing row $i$ and copying the entries of row $m$ in the empty spaces.

If we can prove that det $B = 0$, our theorem will follow. Note that $B$ is a matrix with two identical rows. What happens if we interchange these two rows, obtaining a matrix $B'$? Since the rows are identical, $B = B'$, so that det $B =$ det $B'$. On the other hand, Theorem 8.5 tells us that det $B' = -$ det $B$. Hence we must have det $B = 0$.

**How to Compute Determinants.** Properties (1), (2), and (3) give us a relatively efficient way of computing determinants. To do this, first apply elementary row operations until the matrix is brought into row-echelon form. This is the first part of the reduction process given in §5; it involves only elementary operations of types (1) and (3), so the most it can do is change the sign of the determinant. If you keep track of the number of row exchanges you make, you will know whether or not you have changed the sign of the determinant. When the matrix is in row-echelon form, it may contain a row of zeros, in which case the determinant is zero. Otherwise, the matrix will be of the form

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1k} \\ 0 & b_{22} & b_{23} & \cdots & b_{2k} \\ 0 & 0 & b_{33} & \cdots & b_{3k} \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & b_{kk} \end{bmatrix},$$

which is called *triangular form*. We leave it to you to prove the following theorem:

**8.7 Theorem.** If a matrix $B$ is in triangular form, then

$$\det B = b_{11}b_{22}b_{33} \cdots b_{kk}.$$

EXAMPLE. Let us compute a determinant by this method.

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

$$= -\det \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & -5 & 2 & -4 \\ 0 & 1 & 0 & 3 \end{bmatrix} = -\det \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 7 & -9 \\ 0 & 0 & -1 & 4 \end{bmatrix}$$

$$= -\det \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 7 & -9 \\ 0 & 0 & 0 & \frac{19}{7} \end{bmatrix} = -19.$$

A good way to check this answer is to make an even number of row exchanges in the original matrix, and then reduce this new matrix to row-echelon form.

## EXERCISES

1. Compute the signs of the following permutations:

(a) 2, 4, 1, 3 (b) 5, 4, 3, 2, 1 (c) 7, 6, 5, 3, 4, 1, 2

2. Using Theorem 8.2, show that the following matrix has rank 2:

$$\begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & 2 & 1 \\ 1 & -1 & -7 & 0 \end{bmatrix}.$$

3. Compute the determinants of the following matrices; check your answers.

$$\text{(a)} \begin{bmatrix} 1 & -1 & 0 & 4 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & 0 & 3 \\ 1 & 6 & 5 & 0 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & -1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & 3 & 1 & 0 \\ 0 & 0 & -1 & 4 & 2 \end{bmatrix}.$$

4. Write out the formula for the determinant of a 4 by 4 matrix.

5. Show that if the numbers $a_1, \ldots, a_k$ are distinct, then the determinant of the following matrix does not vanish:

$$\begin{bmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{k-1} \\ \vdots & & & & \vdots \\ 1 & a_k & a_k^2 & \cdots & a_k^{k-1} \end{bmatrix}.$$

[*Hint*: Show that the columns form an independent set.]

6. Consider a system of linear equations whose coefficient matrix $A$ is of size $n$ by $n$.

(a) If det $A \neq 0$, can you say anything about the general solution?
(b) If det $A \neq 0$ and the system is homogeneous, can you say anything about the general solution?
(c) If det $A = 0$, can you say anything about the general solution?
(d) If det $A = 0$ and the system is homogeneous, can you say anything about the general solution?

7. Prove Theorem 8.7.

## §9. FUNCTION SPACES

Suppose we consider all real-valued functions $f(x)$ which are defined on a given interval $a \leq x \leq b$. The sum of two such functions is again such a function, as is the product of such a function by a real number. If we let $F$ denote the set of all such functions, it is easy to check that these two operations have all of the vector space properties (see §1), so that we may speak of $F$ as being a vector space, even though it is a very different one in many ways from $V^3$ or $V^n$.

The only question which might arise in checking the vector space properties is the question as to what the zero vector $\mathbf{0}$ means in this context. A little thought will show that if property (3) is to hold, $\mathbf{0}$ must stand for the *zero function*, the constant function whose value is everywhere zero.

We define what is meant by a subspace of $F$ in the same way as we did in defining subspaces of $V^n$. It turns out that one of the more important subspaces of $F$ is the set $C$ of all continuous functions $f(x)$ defined on $a \leq x \leq b$. The set $C$ *is* in fact a subspace of $F$, because the sum of two continuous functions is a continuous function, and so is the product of a continuous function and a real number.

Then we define what we mean by a *linearly independent* set of functions; it is a set $f_1, \ldots, f_p$ of functions such that the only linear combination $c_1f_1 + \cdots + c_pf_p$ of them which equals the zero function is the trivial linear combination $0f_1 + \cdots + 0f_p$.

EXAMPLE. The functions $x$ and $\sin x$ restricted to the interval $0 \leq x \leq 1$ form a linearly independent set. For if $c_1x + c_2 \sin x$ is to equal the zero function, then in particular, $c_1 \pi/6 + c_2 \sin \pi/6 = 0$ and $c_1 \pi/4 + c_2 \sin \pi/4 = 0$. Hence $c_1 = -c_2\, 3/\pi$ and $c_1 = -c_2\, 2\sqrt{2}/\pi$, so that $c_1 = c_2 = 0$.

One obvious question is whether one can define a scalar product for $F$ which satisfies properties (9) through (12) of §1. The ordinary product of functions will not do, because that gives a function, not a scalar. The following definition is one which has turned out to be extremely useful, although it seems strange at first glance. We define the scalar product $\langle f, g\rangle$ of the functions $f$ and $g$ by the equation

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

(We denote the scalar product by $\langle f, g\rangle$ rather than $f \cdot g$, because the latter usually means the ordinary product of two functions.) Now the function $f(x)g(x)$ must be integrable before this definition makes sense, but if we restrict ourselves to the space $C$ of continuous functions, then $\langle f, g\rangle$ is always defined. We leave it to you to check that properties (9) through (12) are satisfied.

## EXERCISES

1. Check that $C$ satisfies the properties (1) through (12) of §1. The only one which requires some work is the one which says that $\langle f, f\rangle > 0$ if $f$ is not the zero function.

2. Show that the functions 1, $x$, and $e^x$ when restricted to the interval $0 \leq x \leq 1$ form an independent set.

3. Show that the functions 1, $x$, $x^2$ when restricted to the interval $a \leq x \leq b$ form a linearly independent set. [*Hint*: If the function $c_0 + c_1x + c_2x^2$ equals zero for all $x$ in the interval $a \leq x \leq b$, then so do all its derivatives.]

4. Show that the functions 1, $x$, $x^2, \ldots, x^n$ when restricted to the interval $a \leq x \leq b$ form an independent set.

5. Show that the functions 1, $e^x$, $e^{2x}$ when restricted to the interval $a \leq x \leq b$ form a linearly independent set.

6. Show that the functions $\sin x$, $\cos x$ when restricted to the interval $a \leq x \leq b$ form a linearly independent set. [*Hint*: If $f(x) = c_1 \sin x + c_2 \cos x$ for $a \leq x \leq b$, then $(f'(x))^2 + (f(x))^2 = c_1^2 + c_2^2$.]

7. Determine whether the functions 1, $x - 1$, and $x^2 - 1$ form an independent set when restricted to the interval $a \leq x \leq b$.

8. Do the same for the functions 1, $x^2 - 1$, $x^2 + 1$.

9. The functions $f$ and $g$ are said to be *orthogonal* if $\langle f, g\rangle = 0$. Show that the functions 1, $\cos x$, $\sin x$ form an orthogonal set when restricted to the interval $0 \leq x \leq 2\pi$.

10. Suppose that $f$ is a function defined for $0 \leq x \leq 2\pi$ such that

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x.$$

Take the scalar product of both sides successively with 1, $\cos x$, and $\sin x$ to obtain formulas for the coefficients $a_0$, $a_1$, and $b_1$.

11. Show that the functions 1, $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x, \ldots, \cos nx$, $\sin nx$ form an orthogonal set when restricted to the interval $0 \leq x \leq 2\pi$.*

12. Suppose that $f$ is a function such that

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + \cdots + a_n \cos nx + b_n \sin nx.$$

Find formulas for the coefficients $a_0, \ldots, b_n$. The numbers given by these formulas are called the *Fourier coefficients* of $f$. They are defined even if $f$ is not equal to a linear combination of 1, $\cos x, \ldots, \sin nx$, although their meaning is more complicated in that case.

13. Suppose that $\phi_0(x), \phi_1(x), \ldots, \phi_n(x)$ form an orthogonal set when restricted to the interval $a \leq x \leq b$. If $f$ is a function such that $f(x) =$

---

* See Section 16-8 of Thomas' *Calculus and Analytic Geometry*, 3rd ed. (1960), or Chapter 7 of Kaplan's *Advanced Calculus*, (1952), Addison-Wesley Publishing Co. Inc., Reading, Mass.

$c_0\phi_0(x) + \ldots + c_n\phi_n(x)$, find formulas for the coefficients $c_0, \ldots, c_n$. The numbers given by these formulas are called the Fourier coefficients of $f$ with respect to the orthogonal set $\phi_0, \ldots, \phi_n$. The fact that they may be readily computed is one of the aspects that makes orthogonal sets of functions very useful for applications.

14. Choose scalars $c_0$, $d_0$, $d_1$ so that the functions $P_0$, $P_1$, and $P_2$ given by the formulas

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x + c_0P_0(x), \\ P_2(x) &= x^2 + d_0P_0(x) + d_1P_1(x) \end{aligned}$$

are orthogonal when restricted to the interval $-1 \leq x \leq 1$. Compare with Exercise 12 of §4.

15. Generalize the procedure of Exercise 14 to show that there exist functions $P_0(x), P_1(x), \ldots, P_n(x)$ which are orthogonal when restricted to the interval $-1 \leq x \leq 1$, where $P_i(x)$ is a polynomial of degree $i$, for $i = 0, \ldots, n$. Except for a scalar factor, the functions constructed in this way are the mathematically very well known ones called the *Legendre polynomials*. They are of considerable importance in the application of mathematics to physical problems.

16. Show that if the numbers $a_1, \cdots, a_k$ are distinct, then the functions $e^{a_1x}, \cdots, e^{a_kx}$ form a linearly independent set, when restricted to the interval $a \leq x \leq b$. Show that no two of these functions are orthogonal.

# INDEX

**THE AUTHOR**

James R. Munkres, who received the Ph.D. degree from the University of Michigan, is Associate Professor of Mathematics at the Massachusetts Institute of Technology. Prior to joining the Massachusetts Institute of Technology faculty, he taught at the University of Michigan and Princeton University. Professor Munkres is a member of The American Mathematical Society, and his fields of special interest include combinatorial and differential topology. He is the author of *Elementary Differential Topology,* published in 1963.

---

The purpose of this short text is to study the general system of linear equations in some detail. Using some ideas of modern linear algebra—vector spaces, subspaces, and linear independence of vectors—theorems are obtained about the consistency of a system of linear equations and about the dimension of its set of solutions. Determinants and linear independence of functions are also discussed.

The text will be particularly useful as a supplement to a course in calculus or differential equations. It is also appropriate for independent reading or as a prelude to a more abstract course in linear or modern algebra.

**ADDISON-WESLEY PUBLISHING COMPANY, INC.**

Reading, Massachusetts • Palo Alto • London

PRINTED IN U.S.A. 4895