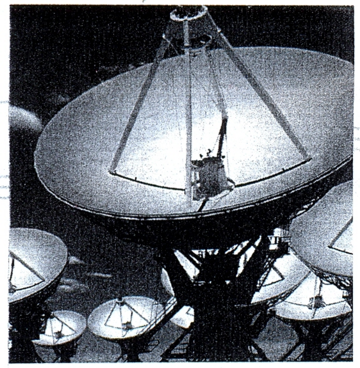


CHAPTER 7

Linear Algebra: Matrices, Vectors, Determinants. Linear Systems

This is the first of two chapters on **linear algebra,** which concerns mainly systems of linear equations and linear transformations (to be discussed in this chapter) and eigenvalue problems (to follow in Chap. 8).

Systems of linear equations, briefly called **linear systems,** arise in electrical networks, mechanical frameworks, economic models, optimization problems, numerics for differential equations, as we shall see in Chaps. 21–23, and so on.

As main tools, linear algebra uses **matrices** (rectangular arrays of numbers or functions) and **vectors.** Calculations with matrices handle matrices as single objects, denote them by single letters, and calculate with them in a very compact form, almost as with numbers, so that matrix calculations constitute a powerful "mathematical shorthand".

Calculations with matrices and vectors are defined and explained in Secs. 7.1–7.2. Sections 7.3–7.8 center around linear systems, with a thorough discussion of Gauss elimination, the role of rank, the existence and uniqueness problem for solutions (Sec. 7.5), and matrix inversion. This also includes determinants (Cramer's rule) in Sec. 7.6 (for quick reference) and Sec. 7.7. Applications are considered throughout this chapter. The last section (Sec. 7.9) on vector spaces, inner product spaces, and linear transformations is more abstract. Eigenvalue problems follow in Chap. 8.

COMMENT. ***Numeric linear algebra (Secs. 20.1–20.5) can be studied immediately after this chapter.***

Prerequisite: None.
Sections that may be omitted in a short course: 7.5, 7.9.
References and Answers to Problems: App. 1 Part B, and App. 2.

7.1 Matrices, Vectors: Addition and Scalar Multiplication

In this section and the next one we introduce the basic concepts and rules of matrix and vector algebra. The main application to **linear systems** (systems of linear equations) begins in Sec. 7.3.

A **matrix** is a rectangular array of numbers (or functions) enclosed in brackets. These numbers (or functions) are called the **entries** (or sometimes the *elements*) of the matrix. For example,

$$(1)\qquad \begin{bmatrix} 0.3 & 1 & -5 \\ 0 & -0.2 & 16 \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

$$\begin{bmatrix} e^{-x} & 2x^2 \\ e^{6x} & 4x \end{bmatrix}, \quad [a_1 \quad a_2 \quad a_3], \quad \begin{bmatrix} 4 \\ \frac{1}{2} \end{bmatrix}$$

are matrices. The first matrix has two **rows** (horizontal lines of entries) and three **columns** (vertical lines). The second and third matrices are **square matrices,** that is, each has as many rows as columns (3 and 2, respectively). The entries of the second matrix have two indices giving the location of the entry. The first index is the number of the row and the second is the number of the column in which the entry stands. Thus, a_{23} (read *a two three*) is in Row 2 and Column 3, etc. This notation is standard, regardless of whether a matrix is square or not.

Matrices having just a single row or column are called **vectors.** Thus the fourth matrix in (1) has just one row and is called a **row vector.** The last matrix in (1) has just one column and is called a **column vector.**

We shall see that matrices are practical in various applications for storing and processing data. As a first illustration let us consider two simple but typical examples.

EXAMPLE 1 Linear Systems, a Major Application of Matrices

In a system of linear equations, briefly called a **linear system,** such as

$A x = b$

$$\begin{aligned} 4x_1 + 6x_2 + 9x_3 &= 6 \\ 6x_1 \qquad - 2x_3 &= 20 \\ 5x_1 - 8x_2 + x_3 &= 10 \end{aligned}$$

the coefficients of the **unknowns** x_1, x_2, x_3 are the entries of the **coefficient matrix,** call it **A,**

$b = \begin{bmatrix} 6 \\ 20 \\ 10 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$

$$\mathbf{A} = \begin{bmatrix} 4 & 6 & 9 \\ 6 & 0 & -2 \\ 5 & -8 & 1 \end{bmatrix}. \qquad \text{The matrix} \qquad \tilde{\mathbf{A}} = \begin{bmatrix} 4 & 6 & 9 & 6 \\ 6 & 0 & -2 & 20 \\ 5 & -8 & 1 & 10 \end{bmatrix}$$

is obtained by augmenting **A** by the right sides of the linear system and is called the **augmented matrix** of the system. In **A** the coefficients of the system are displayed in the pattern of the equations. That is, their position in **A** corresponds to that in the system when written as shown. The same is true for $\tilde{\mathbf{A}}$.

We shall see that the augmented matrix $\tilde{\mathbf{A}}$ contains all the information about the solutions of a system, so that we can solve a system just by calculations on its augmented matrix. We shall discuss this in great detail, beginning in Sec. 7.3. Meanwhile you may verify by substitution that the solution is $x_1 = 3$, $x_2 = \frac{1}{2}$, $x_3 = -1$.

The notation x_1, x_2, x_3 for the unknowns is practical but not essential; we could choose x, y, z or some other letters. ■

EXAMPLE 2 Sales Figures in Matrix Form

Sales figures for three products I, II, III in a store on Monday (M), Tuesday (T), $\cdots$ may for each week be arranged in a matrix

$$\begin{array}{ccc} & \begin{array}{cccccc} \text{M} & \text{T} & \text{W} & \text{Th} & \text{F} & \text{S} \end{array} & \\ \mathbf{A} = & \left[\begin{array}{cccccc} 400 & 330 & 810 & 0 & 210 & 470 \\ 0 & 120 & 780 & 500 & 500 & 960 \\ 100 & 0 & 0 & 270 & 430 & 780 \end{array}\right] & \begin{array}{c} \text{I} \\ \text{II} \\ \text{III} \end{array} \end{array}$$

If the company has ten stores, we can set up ten such matrices, one for each store. Then by adding corresponding entries of these matrices we can get a matrix showing the total sales of each product on each day. Can you think of other data for which matrices are feasible? For instance, in transportation or storage problems? Or in recording phone calls, or in listing distances in a network of roads? ∎

General Concepts and Notations

We shall denote matrices by capital boldface letters **A, B, C,** $\cdots$, or by writing the general entry in brackets; thus $\mathbf{A} = [a_{jk}]$, and so on. By an ***m* × *n* matrix** (read *m by n matrix*) we mean a matrix with m rows and n columns—rows come always first! $m \times n$ is called the **size** of the matrix. Thus an $m \times n$ matrix is of the form

$A \in R^{m \times n}$

$$(2) \qquad \mathbf{A} = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

a_{jk} : The jth row - kth column element

The matrices in (1) are of sizes 2×3, 3×3, 2×2, 1×3, and 2×1, respectively.

Each entry in (2) has two subscripts. The first is the *row number* and the second is the *column number*. Thus a_{21} is the entry in Row 2 and Column 1.

If $m = n$, we call **A** an $n \times n$ **square matrix.** Then its diagonal containing the entries $a_{11}, a_{22}, \cdots, a_{nn}$ is called the **main diagonal** of **A.** Thus the main diagonals of the two square matrices in (1) are a_{11}, a_{22}, a_{33} and e^{-x}, $4x$, respectively.

Square matrices are particularly important, as we shall see. A matrix that is not square is called a **rectangular matrix.**

Vectors

A **vector** is a matrix with only one row or column. Its entries are called the **components** of the vector. We shall denote vectors by ***lowercase*** boldface letters **a, b,** $\cdots$ or by its general component in brackets, $\mathbf{a} = [a_j]$, and so on. Our special vectors in (1) suggest that a (general) **row vector** is of the form

$a \in R^{1 \times n}$

$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots, \quad a_n]. \qquad \text{For instance,} \qquad \mathbf{a} = [-2 \quad 5 \quad 0.8 \quad 0 \quad 1].$$

no commas!

A **column vector** is of the form

$b \in R^{n \times 1}$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \qquad \text{For instance,} \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ -7 \end{bmatrix}.$$

or $(b_1, b_2, \cdots, b_m)$

Matrix Addition and Scalar Multiplication

What makes matrices and vectors really useful and particularly suitable for computers is the fact that we can calculate with them almost as easily as with numbers. Indeed, we now introduce rules for addition and for scalar multiplication (multiplication by numbers) that were suggested by practical applications. (Multiplication of matrices by matrices follows in the next section.) We first need the concept of equality.

DEFINITION

Equality of Matrices

Two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ are **equal,** written $\mathbf{A} = \mathbf{B}$, if and only if they have the same size and the corresponding entries are equal, that is, $a_{11} = b_{11}$, $a_{12} = b_{12}$, and so on. Matrices that are not equal are called **different.** Thus, matrices of different sizes are always different.

EXAMPLE 3 **Equality of Matrices**

$A = B \Leftrightarrow a_{jk} = b_{jk}, \; j = 1, \cdots, m, \; k = 1, \cdots, n$

Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 4 & 0 \\ 3 & -1 \end{bmatrix}.$$

Then

$$\mathbf{A} = \mathbf{B} \quad \text{if and only if} \quad \begin{aligned} a_{11} &= 4, & a_{12} &= 0, \\ a_{21} &= 3, & a_{22} &= -1. \end{aligned}$$

The following matrices are all different. Explain!

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \quad \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 0 \\ 4 & 2 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 3 \\ 0 & 4 & 2 \end{bmatrix} \blacksquare$$

DEFINITION

Addition of Matrices

The **sum** of two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ ***of the same size*** is written $\mathbf{A} + \mathbf{B}$ and has the entries $a_{jk} + b_{jk}$ obtained by adding the corresponding entries of **A** and **B.** Matrices of different sizes cannot be added.

As a special case, the **sum a + b** of two row vectors or two column vectors, which must have the same number of components, is obtained by adding the corresponding components.

Let $C \triangleq A + B$

Then, $c_{jk} \triangleq a_{jk} + b_{jk}, \; j = 1, \cdots, m, \; k = 1, \cdots, n$

EXAMPLE 4 Addition of Matrices and Vectors

If $\mathbf{A} = \begin{bmatrix} -4 & 6 & 3 \\ 0 & 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & -1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$, then $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 5 & 3 \\ 3 & 2 & 2 \end{bmatrix}$.

$\mathbf{A}$ in Example 3 and our present $\mathbf{A}$ cannot be added. If $\mathbf{a} = [5 \quad 7 \quad 2]$ and $\mathbf{b} = [-6 \quad 2 \quad 0]$, then $\mathbf{a} + \mathbf{b} = [-1 \quad 9 \quad 2]$.

An application of matrix addition was suggested in Example 2. Many others will follow. ■

DEFINITION

Scalar Multiplication (Multiplication by a Number)

The **product** of any $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ and any **scalar** c (number c) is written $c\mathbf{A}$ and is the $m \times n$ matrix $c\mathbf{A} = [ca_{jk}]$ obtained by multiplying each entry of $\mathbf{A}$ by c.

Here $(-1)\mathbf{A}$ is simply written $-\mathbf{A}$ and is called the **negative** of $\mathbf{A}$. Similarly, $(-k)\mathbf{A}$ is written $-k\mathbf{A}$. Also, $\mathbf{A} + (-\mathbf{B})$ is written $\mathbf{A} - \mathbf{B}$ and is called the **difference** of $\mathbf{A}$ and $\mathbf{B}$ (which must have the same size!).

EXAMPLE 5 Scalar Multiplication

If $\mathbf{A} = \begin{bmatrix} 2.7 & -1.8 \\ 0 & 0.9 \\ 9.0 & -4.5 \end{bmatrix}$, then $-\mathbf{A} = \begin{bmatrix} -2.7 & 1.8 \\ 0 & -0.9 \\ -9.0 & 4.5 \end{bmatrix}$, $\frac{10}{9}\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 0 & 1 \\ 10 & -5 \end{bmatrix}$, $0\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

If a matrix $\mathbf{B}$ shows the distances between some cities in miles, $1.609\mathbf{B}$ gives these distances in kilometers. ■

Rules for Matrix Addition and Scalar Multiplication. From the familiar laws for the addition of numbers we obtain similar laws for the addition of matrices of the same size $m \times n$, namely,

$$
\text{(3)} \qquad
\begin{aligned}
&\text{(a)} & \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\
&\text{(b)} & (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad \text{(written } \mathbf{A} + \mathbf{B} + \mathbf{C}\text{)} \\
&\text{(c)} & \mathbf{A} + \mathbf{0} &= \mathbf{A} \\
&\text{(d)} & \mathbf{A} + (-\mathbf{A}) &= \mathbf{0}.
\end{aligned}
$$

Here $\mathbf{0}$ denotes the **zero matrix** (of size $m \times n$), that is, the $m \times n$ matrix with all entries zero. (The last matrix in Example 5 is a zero matrix.)

Hence matrix addition is *commutative* and *associative* [by (3a) and (3b)].

Similarly, for scalar multiplication we obtain the rules

$$
\text{(4)} \qquad
\begin{aligned}
&\text{(a)} & c(\mathbf{A} + \mathbf{B}) &= c\mathbf{A} + c\mathbf{B} \\
&\text{(b)} & (c + k)\mathbf{A} &= c\mathbf{A} + k\mathbf{A} \\
&\text{(c)} & c(k\mathbf{A}) &= (ck)\mathbf{A} \qquad \text{(written } ck\mathbf{A}\text{)} \\
&\text{(d)} & 1\mathbf{A} &= \mathbf{A}.
\end{aligned}
$$

§ 7.2 Matrix Multiplication

DEFINITION

Multiplication of a Matrix by a Matrix

The **product** $\mathbf{C} = \mathbf{AB}$ (in this order) of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ times an $r \times p$ matrix $\mathbf{B} = [b_{jk}]$ is defined if and only if $r = n$ and is then the $m \times p$ matrix $\mathbf{C} = [c_{jk}]$ with entries

$$c_{jk} = \sum_{l=1}^{n} a_{jl}b_{lk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \qquad \begin{array}{l} j = 1, \cdots, m \\ k = 1, \cdots, p. \end{array} \tag{1}$$

The condition $r = n$ means that the second factor, **B,** must have as many rows as the first factor has columns, namely n. As a diagram of sizes (denoted as shown):

$$\begin{array}{ccccc} \mathbf{A} & \mathbf{B} & = & \mathbf{C} \\ [m \times n] & [n \times r] & = & [m \times r]. \end{array}$$

c_{jk} in (1) is obtained by multiplying each entry in the jth row of **A** by the corresponding entry in the kth column of **B** and then adding these n products. For instance, $c_{21} = a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1}$, and so on. One calls this briefly a ***"multiplication of rows into columns."*** See the illustration in Fig. 155, where $n = 3$.

$$m = 4 \left\{ \overbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}}^{n=3} \overbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}}^{p=2} = \overbrace{\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix}}^{p=2} \right\} m = 4$$

Fig. 155. Notations in a product $\mathbf{AB} = \mathbf{C}$

EXAMPLE 1 Matrix Multiplication

$$\mathbf{AB} = \begin{bmatrix} 3 & 5 & -1 \\ 4 & 0 & 2 \\ -6 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -2 & 3 & 1 \\ 5 & 0 & 7 & 8 \\ 9 & -4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 22 & -2 & 43 & 42 \\ 26 & -16 & 14 & 6 \\ -9 & 4 & -37 & -28 \end{bmatrix}$$

Here $c_{11} = 3 \cdot 2 + 5 \cdot 5 + (-1) \cdot 9 = 22$, and so on. The entry in the box is $c_{23} = 4 \cdot 3 + 0 \cdot 7 + 2 \cdot 1 = 14$. The product **BA** is not defined. ■

EXAMPLE 2 Multiplication of a Matrix and a Vector

$$\begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \cdot 3 + 2 \cdot 5 \\ 1 \cdot 3 + 8 \cdot 5 \end{bmatrix} = \begin{bmatrix} 22 \\ 43 \end{bmatrix} \qquad \text{whereas} \qquad \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 1 & 8 \end{bmatrix} \qquad \text{is undefined.}$$ ■

EXAMPLE 3 Products of Row and Column Vectors

$$\begin{bmatrix} 3 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = [19], \qquad \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 3 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 1 \\ 6 & 12 & 2 \\ 12 & 24 & 4 \end{bmatrix}.$$ ■

EXAMPLE 4 **CAUTION! Matrix Multiplication Is Not Commutative, AB ≠ BA in General**

This is illustrated by Examples 1 and 2, where one of the two products is not even defined, and by Example 3, where the two products have different sizes. But it also holds for square matrices. For instance,

$$\begin{bmatrix} 1 & 1 \\ 100 & 100 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 100 & 100 \end{bmatrix} = \begin{bmatrix} 99 & 99 \\ -99 & -99 \end{bmatrix}.$$

It is interesting that this also shows that $\mathbf{AB} = \mathbf{0}$ does *not* necessarily imply $\mathbf{BA} = \mathbf{0}$ or $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$. We shall discuss this further in Sec. 7.8, along with reasons when this happens. ■

Our examples show that the ***order of factors*** in matrix products ***must always be observed very carefully.*** Otherwise matrix multiplication satisfies rules similar to those for numbers, namely.

$$\text{(2)}\quad\begin{aligned} &\text{(a)} && (k\mathbf{A})\mathbf{B} = k(\mathbf{AB}) = \mathbf{A}(k\mathbf{B}) && \textit{written } k\mathbf{AB} \textit{ or } \mathbf{A}k\mathbf{B} \\ &\text{(b)} && \mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} && \textit{written } \mathbf{ABC} \\ &\text{(c)} && (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \\ &\text{(d)} && \mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \end{aligned}$$

provided **A, B,** and **C** are such that the expressions on the left are defined; here, k is any scalar. (2b) is called the **associative law.** (2c) and (2d) are called the **distributive laws.**

Since matrix multiplication is a multiplication of rows into columns, we can write the defining formula (1) more compactly as

$$\text{(3)}\qquad c_{jk} = \mathbf{a}_j\mathbf{b}_k, \qquad j = 1, \cdots, m; \quad k = 1, \cdots, p,$$

where $\mathbf{a}_j$ is the jth row vector of **A** and $\mathbf{b}_k$ is the kth column vector of **B,** so that in agreement with (1),

$$\mathbf{a}_j\mathbf{b}_k = [a_{j1} \quad a_{j2} \quad \cdots \quad a_{jn}]\begin{bmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{bmatrix} = a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk}.$$

EXAMPLE 5 **Product in Terms of Row and Column Vectors**

If $\mathbf{A} = [a_{jk}]$ is of size 3 × 3 and $\mathbf{B} = [b_{jk}]$ is of size 3 × 4, then

$$AB = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}[b_1 \cdots b_n]$$

$$\text{(4)}\qquad \mathbf{AB} = \begin{bmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 & \mathbf{a}_1\mathbf{b}_3 & \mathbf{a}_1\mathbf{b}_4 \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 & \mathbf{a}_2\mathbf{b}_3 & \mathbf{a}_2\mathbf{b}_4 \\ \mathbf{a}_3\mathbf{b}_1 & \mathbf{a}_3\mathbf{b}_2 & \mathbf{a}_3\mathbf{b}_3 & \mathbf{a}_3\mathbf{b}_4 \end{bmatrix}.$$

Taking $\mathbf{a}_1 = [3 \quad 5 \quad -1]$, $\mathbf{a}_2 = [4 \quad 0 \quad 2]$, etc., verify (4) for the product in Example 1. ■

where

$a_i \in R^{1\times n}$

$b_j \in R^{m\times 1}$

Parallel processing of products on the computer is facilitated by a variant of (3) for computing $\mathbf{C} = \mathbf{AB}$, which is used by standard algorithms (such as in Lapack). In this method, **A** is used as given, **B** is taken in terms of its column vectors, and the product is computed columnwise; thus,

$$\text{(5)}\qquad \mathbf{AB} = \mathbf{A}[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p] = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \cdots \quad \mathbf{Ab}_p].$$

$$(5)' \qquad Ab = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} b = \begin{bmatrix} a_1 b \\ \vdots \\ a_m b \end{bmatrix}$$

Columns of **B** are then assigned to different processors (individually or several to each processor), which simultaneously compute the columns of the product matrix $\mathbf{Ab}_1$, $\mathbf{Ab}_2$, etc.

EXAMPLE 6 Computing Products Columnwise by (5)

To obtain

$$\mathbf{AB} = \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 & 7 \\ -1 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 11 & 4 & 34 \\ -17 & 8 & -23 \end{bmatrix}$$

from (5), calculate the columns

$$\begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 11 \\ -17 \end{bmatrix}, \quad \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}, \quad \begin{bmatrix} 4 & 1 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 34 \\ -23 \end{bmatrix}$$

of **AB** and then write them as a single matrix, as shown in the first formula on the right. ■

Motivation of Multiplication by Linear Transformations

Let us now motivate the "unnatural" matrix multiplication by its use in **linear transformations.** For $n = 2$ variables these transformations are of the form

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned} \tag{6*}$$

and suffice to explain the idea. (For general n they will be discussed in Sec. 7.9.) For instance, (6*) may relate an x_1x_2-coordinate system to a y_1y_2-coordinate system in the plane. In vectorial form we can write (6*) as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}. \tag{6}$$

Now suppose further that the x_1x_2-system is related to a w_1w_2-system by another linear transformation, say,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{Bw} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} b_{11}w_1 + b_{12}w_2 \\ b_{21}w_1 + b_{22}w_2 \end{bmatrix}. \tag{7}$$

Then the y_1y_2-system is related to the w_1w_2-system indirectly via the x_1x_2-system, and we wish to express this relation directly. Substitution will show that this direct relation is a linear transformation, too, say,

$$\mathbf{y} = \mathbf{Cw} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} c_{11}w_1 + c_{12}w_2 \\ c_{21}w_1 + c_{22}w_2 \end{bmatrix}. \tag{8}$$

Indeed, substituting (7) into (6), we obtain

$$\begin{aligned} y_1 &= a_{11}(b_{11}w_1 + b_{12}w_2) + a_{12}(b_{21}w_1 + b_{22}w_2) \\ &= (a_{11}b_{11} + a_{12}b_{21})w_1 + (a_{11}b_{12} + a_{12}b_{22})w_2 \\ y_2 &= a_{21}(b_{11}w_1 + b_{12}w_2) + a_{22}(b_{21}w_1 + b_{22}w_2) \\ &= (a_{21}b_{11} + a_{22}b_{21})w_1 + (a_{21}b_{12} + a_{22}b_{22})w_2. \end{aligned}$$

§7.3 HW

7
16. '24

Comparing this with (8), we see that

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} \qquad c_{12} = a_{11}b_{12} + a_{12}b_{22}$$
$$c_{21} = a_{21}b_{11} + a_{22}b_{21} \qquad c_{22} = a_{21}b_{12} + a_{22}b_{22}.$$

This proves that $\mathbf{C} = \mathbf{AB}$ with the product defined as in (1). For larger matrix sizes the idea and result are exactly the same. Only the number of variables changes. We then have m variables y and n variables x and p variables w. The matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C} = \mathbf{AB}$ then have sizes $m \times n$, $n \times p$, and $m \times p$, respectively. And the requirement that $\mathbf{C}$ be the product $\mathbf{AB}$ leads to formula (1) in its general form. *This motivates matrix multiplication completely.*

$$y = Ax = A(Bw) = (AB)w = Cw$$

Transposition

Transposition provides a transition from row vectors to column vectors and conversely. More generally, it gives us a choice to work either with a matrix or with its transpose, whatever will be more practical in a specific situation.

DEFINITION

Transposition of Matrices and Vectors

The transpose of an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ is the $n \times m$ matrix $\mathbf{A}^{\mathsf{T}}$ (read *A transpose*) that has the first *row* of $\mathbf{A}$ as its first *column*, the second *row* of $\mathbf{A}$ as its second *column*, and so on. Thus the transpose of $\mathbf{A}$ in (2) is $\mathbf{A}^{\mathsf{T}} = [a_{kj}]$, written out

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}$$

$$\textbf{(9)} \qquad \mathbf{A}^{\mathsf{T}} = [a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdot & \cdot & \cdots & \cdot \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

As a special case, transposition converts row vectors to column vectors and conversely.

EXAMPLE 7 Transposition of Matrices and Vectors

If
$$\mathbf{A} = \begin{bmatrix} 5 & -8 & 1 \\ 4 & 0 & 0 \end{bmatrix}, \quad \text{then} \quad \mathbf{A}^{\mathsf{T}} = \begin{bmatrix} 5 & 4 \\ -8 & 0 \\ 1 & 0 \end{bmatrix}.$$

A little more compactly, we can write

$$\begin{bmatrix} 5 & -8 & 1 \\ 4 & 0 & 0 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} 5 & 4 \\ -8 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} 3 & 8 \\ 0 & -1 \end{bmatrix}$$

$$[6 \quad 2 \quad 3]^{\mathsf{T}} = \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}, \qquad \begin{bmatrix} 6 \\ 2 \\ 3 \end{bmatrix}^{\mathsf{T}} = [6 \quad 2 \quad 3].$$

Note that for a square matrix, the transpose is obtained by interchanging entries that are symmetrically positioned with respect to the main diagonal, e.g., a_{12} and a_{21}, and so on. ■

Rules for transposition are

$$\text{(10)}\qquad \begin{aligned} &\text{(a)} && (\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A} \\ &\text{(b)} && (\mathbf{A}+\mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} + \mathbf{B}^\mathsf{T} \\ &\text{(c)} && (c\mathbf{A})^\mathsf{T} = c\mathbf{A}^\mathsf{T} \\ &\text{(d)} && (\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}. \end{aligned}$$

CAUTION! Note that in (10d) the transposed matrices are ***in reversed order***. We leave the proofs to the student. (See Prob. 22.)

Special Matrices

Certain kinds of matrices will occur quite frequently in our work, and we now list the most important ones of them.

square matrices
$A \in R^{n\times n}$

Symmetric and Skew-Symmetric Matrices. Transposition gives rise to two useful classes of matrices, as follows. *Symmetric matrices* and *skew-symmetric matrices* are ***square*** matrices whose transpose equals the matrix itself or minus the matrix, respectively:

$$\text{(11)}\qquad \mathbf{A}^\mathsf{T} = \mathbf{A} \quad (\text{thus } a_{kj} = a_{jk}), \qquad \mathbf{A}^\mathsf{T} = -\mathbf{A} \quad (\text{thus } a_{kj} = -a_{jk}, \text{ hence } a_{jj} = 0).$$

Symmetric Matrix Skew-Symmetric Matrix

EXAMPLE 8 Symmetric and Skew-Symmetric Matrices

$$\mathbf{A} = \begin{bmatrix} 20 & 120 & 200 \\ 120 & 10 & 150 \\ 200 & 150 & 30 \end{bmatrix} \quad \text{is symmetric, and} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & -2 \\ 3 & 2 & 0 \end{bmatrix} \quad \text{is skew-symmetric.}$$

For instance, if a company has three building supply centers C_1, C_2, C_3, then $\mathbf{A}$ could show costs, say, a_{jj} for handling 1000 bags of cement on center C_j, and a_{jk} $(j \neq k)$ the cost of shipping 1000 bags from C_j to C_k. Clearly, $a_{jk} = a_{kj}$ because shipping in the opposite direction will usually cost the same.

Symmetric matrices have several general properties which make them important. This will be seen as we proceed. ■

Triangular Matrices. Upper triangular matrices are square matrices that can have nonzero entries only on and ***above*** the main diagonal, whereas any entry below the diagonal must be zero. Similarly, **lower triangular matrices** can have nonzero entries only on and ***below*** the main diagonal. Any entry on the main diagonal of a triangular matrix may be zero or not.

EXAMPLE 9 Upper and Lower Triangular Matrices

$$\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & 6 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 0 & 0 \\ 8 & -1 & 0 \\ 7 & 6 & 8 \end{bmatrix}, \quad \begin{bmatrix} 3 & 0 & 0 & 0 \\ 9 & -3 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 9 & 3 & 6 \end{bmatrix}. \qquad ■$$

Upper triangular Lower triangular

Diagonal Matrices. These are square matrices that can have nonzero entries only on the main diagonal. Any entry above or below the main diagonal must be zero.

If all the diagonal entries of a diagonal matrix **S** are equal, say, c, we call **S** a **scalar matrix** because multiplication of any square matrix **A** of the same size by **S** has the same effect as the multiplication by a scalar, that is,

$$\mathbf{AS} = \mathbf{SA} = c\mathbf{A}. \tag{12}$$

In particular, a scalar matrix whose entries on the main diagonal are all 1 is called a **unit matrix** (or **identity matrix**) and is denoted by $\mathbf{I}_n$ or simply by **I.** For **I,** formula (12) becomes

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}. \tag{13}$$

EXAMPLE 10 Diagonal Matrix D. Scalar Matrix S. Unit Matrix I

$$\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ ■

Applications of Matrix Multiplication

Matrix multiplication will play a crucial role in connection with linear systems of equations, beginning in the next section. For the time being we mention some other simple applications that need no lengthy explanations.

EXAMPLE 11 Computer Production. Matrix Times Matrix

Supercomp Ltd produces two computer models PC1086 and PC1186. The matrix **A** shows the cost per computer (in thousands of dollars) and **B** the production figures for the year 2005 (in multiples of 10000 units.) Find a matrix **C** that shows the shareholders the cost per quarter (in millions of dollars) for raw material, labor, and miscellaneous.

$$\mathbf{A} = \begin{array}{c} \begin{array}{cc} \text{PC1086} & \text{PC1186} \end{array} \\ \begin{bmatrix} 1.2 & 1.6 \\ 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Raw Components} \\ \text{Labor} \\ \text{Miscellaneous} \end{array} \qquad \mathbf{B} = \begin{array}{c} \text{Quarter} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 3 & 8 & 6 & 9 \\ 6 & 2 & 4 & 3 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{PC1086} \\ \text{PC1186} \end{array}$$

Solution.

$$\mathbf{C} = \mathbf{AB} = \begin{array}{c} \text{Quarter} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 13.2 & 12.8 & 13.6 & 15.6 \\ 3.3 & 3.2 & 3.4 & 3.9 \\ 5.1 & 5.2 & 5.4 & 6.3 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{Raw Components} \\ \text{Labor} \\ \text{Miscellaneous} \end{array}$$

Since cost is given in multiples of \$1000 and production in multiples of 10 000 units, the entries of **C** are multiples of \$10 millions; thus $c_{11} = 13.2$ means \$132 million, etc. ■

HW

Prob. 7.2.19 (b)
Prob. 7.2.22
Prob. 7.2.28 (a)

Linear System, Coefficient Matrix, Augmented Matrix

A **linear system of m equations in n unknowns** $x_1, \cdots, x_n$ is a set of equations of the form

$$
\begin{aligned}
a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m
\end{aligned} \tag{1}
$$

The system is called *linear* because each variable x_j appears in the first power only, just as in the equation of a straight line. $a_{11}, \cdots, a_{mn}$ are given numbers, called the **coefficients** of the system. $b_1, \cdots, b_m$ on the right are also given numbers. If all the b_j are zero, then (1) is called a **homogeneous system.** If at least one b_j is not zero, then (1) is called a **nonhomogeneous system.**

A **solution** of (1) is a set of numbers $x_1, \cdots, x_n$ that satisfies all the m equations. A **solution vector** of (1) is a vector $\mathbf{x}$ whose components form a solution of (1). If the system (1) is homogeneous, it has at least the **trivial solution** $x_1 = 0, \cdots, x_n = 0$.

Matrix Form of the Linear System (1). From the definition of matrix multiplication we see that the m equations of (1) may be written as a single vector equation

$$\mathbf{Ax} = \mathbf{b} \tag{2}$$

where the **coefficient matrix** $\mathbf{A} = [a_{jk}]$ is the $m \times n$ matrix

$A \in R^{m\times n}$

$$
\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}
$$

are column vectors. We assume that the coefficients a_{jk} are not all zero, so that $\mathbf{A}$ is not a zero matrix. Note that $\mathbf{x}$ has n components, whereas $\mathbf{b}$ has m components. The matrix

$\tilde{A} \in R^{m\times(n+1)}$

$$
\tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]
$$

is called the **augmented matrix** of the system (1). The dashed vertical line could be omitted (as we shall do later); it is merely a reminder that the last column of $\tilde{\mathbf{A}}$ does not belong to $\mathbf{A}$.

The augmented matrix $\tilde{\mathbf{A}}$ *determines the system* (1) *completely* because it contains all the given numbers appearing in (1).

EXAMPLE 1 Geometric Interpretation. Existence and Uniqueness of Solutions

If $m = n = 2$, we have two equations in two unknowns x_1, x_2

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2.$$

If we interpret x_1, x_2 as coordinates in the x_1x_2-plane, then each of the two equations represents a straight line, and (x_1, x_2) is a solution if and only if the point P with coordinates x_1, x_2 lies on both lines. Hence there are three possible cases:

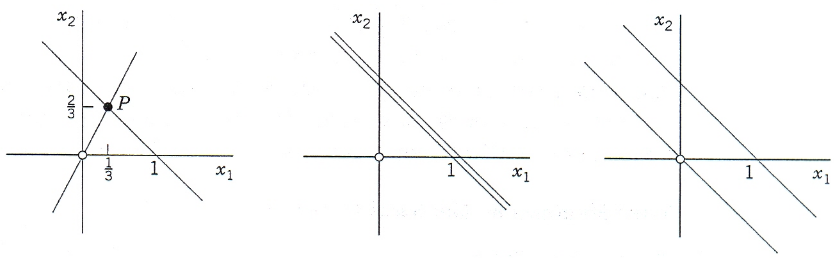
(*a*) Precisely one solution if the lines intersect.

(*b*) Infinitely many solutions if the lines coincide.

(*c*) No solution if the lines are parallel

For instance,

Case (*a*)	Case (*b*)	Case (*c*)
$x_1 + x_2 = 1$	$x_1 + x_2 = 1$	$x_1 + x_2 = 1$
$2x_1 - x_2 = 0$	$2x_1 + 2x_2 = 2$	$x_1 + x_2 = 0$



If the system is homogenous, Case (c) cannot happen, because then those two straight lines pass through the origin, whose coordinates 0, 0 constitute the trivial solution. If you wish, consider three equations in three unknowns as representations of three planes in space and discuss the various possible cases in a similar fashion. See Fig. 156. ■

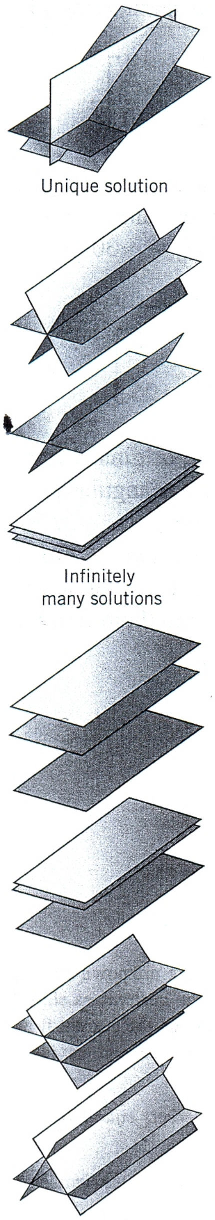


Fig. 156. Three equations in three unknowns interpreted as planes in space

Our simple example illustrates that a system (1) may perhaps have no solution. This poses the following problem. Does a given system (1) have a solution? Under what conditions does it have precisely one solution? If it has more than one solution, how can we characterize the set of all solutions? How can we actually obtain the solutions? Perhaps the last question is the most immediate one from a practical viewpoint. We shall answer it first and discuss the other questions in Sec. 7.5.

Gauss Elimination and Back Substitution

This is a standard elimination method for solving linear systems that proceeds systematically irrespective of particular features of the coefficients. It is a method of great practical importance and is reasonable with respect to computing time and storage demand (two aspects we shall consider in Sec. 20.1 in the chapter on numeric linear algebra). We begin by motivating the method. If a system is in **"triangular form,"** say,

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned}$$

we can solve it by **"back substitution,"** that is, solve the last equation for the variable, $x_2 = -26/13 = -2$, and then work backward, substituting $x_2 = -2$ into the first equation and solve it for x_1, obtaining $x_1 = \frac{1}{2}(2 - 5x_2) = \frac{1}{2}(2 - 5\cdot(-2)) = 6$. This gives us the idea of first reducing a general system to triangular form. For instance, let the given system be

The solutions of (*) and (*)' are identical

(*)
$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ -4x_1 + 3x_2 &= -30. \end{aligned} \qquad \text{Its augmented matrix is} \qquad \begin{bmatrix} 2 & 5 & 2 \\ -4 & 3 & -30 \end{bmatrix}.$$

We leave the first equation as it is. We eliminate x_1 from the second equation, to get a triangular system. For this we add twice the first equation to the second, and we do the same operation on the ***rows*** of the augmented matrix. This gives $-4x_1 + 4x_1 + 3x_2 + 10x_2 = -30 + 2\cdot 2$, that is,

(*)'
$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned} \qquad \text{Row 2 + 2 Row 1} \qquad \begin{bmatrix} 2 & 5 & 2 \\ 0 & 13 & -26 \end{bmatrix}$$

where Row 2 + 2 Row 1 means "Add twice Row 1 to Row 2" in the original matrix. This is the **Gauss elimination** (for 2 equations in 2 unknowns) giving the triangular form, from which back substitution now yields $x_2 = -2$ and $x_1 = 6$, as before.

Since a linear system is completely determined by its augmented matrix, ***Gauss elimination can be done by merely considering the matrices,*** as we have just indicated. We do this again in the next example, emphasizing the matrices by writing them first and the equations behind them, just as a help in order not to lose track.

EXAMPLE 2 Gauss Elimination. Electrical Network

Solve the linear system

$$\tilde{A} = \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right] \Leftarrow$$

Augmented Matrix

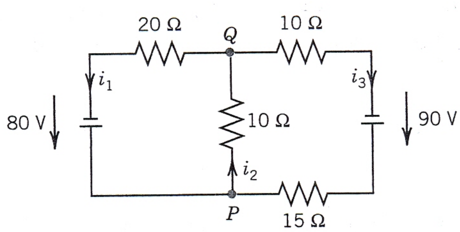
$$\begin{aligned} x_1 - x_2 + x_3 &= 0 \\ -x_1 + x_2 - x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ 20x_1 + 10x_2 \qquad &= 80. \end{aligned}$$

Derivation from the circuit in Fig. 157 (Optional). This is the system for the unknown currents $x_1 = i_1$, $x_2 = i_2$, $x_3 = i_3$ in the electrical network in Fig. 157. To obtain it, we label the currents as shown, choosing directions arbitrarily; if a current will come out negative, this will simply mean that the current flows against the direction of our arrow. The current entering each battery will be the same as the current leaving it. The equations for the currents result from Kirchhoff's laws:

Kirchhoff's current law (KCL). *At any point of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.*

Kirchhoff's voltage law (KVL). *In any closed loop, the sum of all voltage drops equals the impressed electromotive force.*

Node P gives the first equation, node Q the second, the right loop the third, and the left loop the fourth, as indicated in the figure.



Node P: $i_1 - i_2 + i_3 = 0$

Node Q: $-i_1 + i_2 - i_3 = 0$

Right loop: $10i_2 + 25i_3 = 90$

Left loop: $20i_1 + 10i_2 = 80$

Fig. 157. Network in Example 2 and equations relating the currents

Solution by Gauss Elimination. This system could be solved rather quickly by noticing its particular form. But this is not the point. The point is that the Gauss elimination is systematic and will work in general, also for large systems. We apply it to our system and then do back substitution. As indicated let us write the augmented matrix of the system first and then the system itself:

Augmented Matrix $\tilde{\mathbf{A}}$ | Equations

$$\begin{array}{r} \text{Pivot } 1 \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array} \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right] \qquad \begin{array}{r} \text{Pivot } 1 \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array} \begin{aligned} x_1 - x_2 + x_3 &= 0 \\ -x_1 + x_2 - x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ 20x_1 + 10x_2 \qquad &= 80. \end{aligned}$$

Step 1. Elimination of x_1

Call the first row of $\mathbf{A}$ the **pivot row** and the first equation the **pivot equation.** Call the coefficient 1 of its x_1-term the **pivot** in this step. Use this equation to eliminate x_1 (get rid of x_1) in the other equations. For this, do:

Add 1 times the pivot equation to the second equation.

Add -20 times the pivot equation to the fourth equation.

This corresponds to **row operations** on the augmented matrix as indicated in BLUE behind the ***new matrix*** in (3). So the operations are performed on the ***preceding matrix.*** The result is

$$\text{(3)} \qquad \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{array}\right] \qquad \begin{array}{l} \\ \text{Row 2 + Row 1} \\ \\ \text{Row 4} - 20 \text{ Row 1} \end{array} \qquad \begin{aligned} x_1 - x_2 + x_3 &= 0 \\ 0 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ 30x_2 - 20x_3 &= 80. \end{aligned}$$

Step 2. Elimination of x_2

The first equation remains as it is. We want the new second equation to serve as the next pivot equation. But since it has no x_2-term (in fact, it is $0 = 0$), we must first change the order of the equations and the corresponding rows of the new matrix. We put $0 = 0$ at the end and move the third equation and the fourth equation one place up. This is called **partial pivoting** (as opposed to the rarely used *total pivoting*, in which also the order of the unknowns is changed). It gives

$$\begin{array}{r} \\ \text{Pivot } 10 \longrightarrow \\ \text{Eliminate } 30 \longrightarrow \\ \\ \end{array} \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & \boxed{30} & -20 & 80 \\ 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{array}{r} \\ \text{Pivot } 10 \longrightarrow \\ \text{Eliminate } 30x_2 \longrightarrow \\ \\ \end{array} \begin{aligned} x_1 - x_2 + x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ \boxed{30x_2} - 20x_3 &= 80 \\ 0 &= 0 \end{aligned}$$

To eliminate x_2, do:

Add -3 times the pivot equation to the third equation.

The result is

$$\text{(4)} \qquad \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 0 & -95 & -190 \\ 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{array}{l} \\ \\ \text{Row 3} - 3 \text{ Row 2} \\ \\ \end{array} \qquad \begin{aligned} x_1 - x_2 + x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ -95x_3 &= -190 \\ 0 &= 0 \end{aligned}$$

Back Substitution. ***Determination of x_3, x_2, x_1 (in this order)***

Working backward from the last to the first equation of this "triangular" system (4), we can now readily find x_3, then x_2, and then x_1:

$$\begin{aligned} -95x_3 &= -190 \qquad & x_3 &= i_3 = 2 \text{ [A]} \\ 10x_2 + 25x_3 &= 90 \qquad & x_2 &= \tfrac{1}{10}(90 - 25x_3) = i_2 = 4 \text{ [A]} \\ x_1 - x_2 + x_3 &= 0 \qquad & x_1 &= x_2 - x_3 = i_1 = 2 \text{ [A]} \end{aligned}$$

where A stands for "amperes." This is the answer to our problem. The solution is unique. ■

Elementary Row Operations. Row-Equivalent Systems

Example 2 illustrates the operations of the Gauss elimination. These are the first two of three operations, which are called

Elementary Row Operations for Matrices:

Interchange of two rows

Addition of a constant multiple of one row to another row

Multiplication of a row by a ***nonzero*** *constant c.*

CAUTION! These operations are for rows, ***not for columns!*** They correspond to the following

Elementary Operations for Equations:

Interchange of two equations

Addition of a constant multiple of one equation to another equation

Multiplication of an equation by a ***nonzero*** *constant c.*

Clearly, the interchange of two equations does not alter the solution set. Neither does that addition because we can undo it by a corresponding subtraction. Similarly for that multiplication, which we can undo by multiplying the new equation by $1/c$ (since $c \neq 0$), producing the original equation.

We now call a linear system S_1 **row-equivalent** to a linear system S_2 if S_1 can be obtained from S_2 by (finitely many!) row operations. Thus we have proved the following result, which also justifies the Gauss elimination.

THEOREM 1

Row-Equivalent Systems

Row-equivalent linear systems have the same set of solutions.

Because of this theorem, systems having the same solution sets are often called *equivalent systems*. But note well that we are dealing with ***row operations.*** No column operations on the augmented matrix are permitted in this context because they would generally alter the solution set.

A linear system (1) is called **overdetermined** if it has more equations than unknowns, as in Example 2, **determined** if $m = n$, as in Example 1, and **underdetermined** if it has fewer equations than unknowns.

Furthermore, a system (1) is called **consistent** if it has at least one solution (thus, one solution or infinitely many solutions), but **inconsistent** if it has no solutions at all, as $x_1 + x_2 = 1$, $x_1 + x_2 = 0$ in Example 1.

Gauss Elimination: The Three Possible Cases of Systems

The Gauss elimination can take care of linear systems with a unique solution (see Example 2), with infinitely many solutions (Example 3, below), and without solutions (inconsistent systems; see Example 4).

EXAMPLE 3 **Gauss Elimination if Infinitely Many Solutions Exist**

Solve the following linear systems of three equations in four unknowns whose augmented matrix is

$$(5)\quad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0.6 & 1.5 & 1.5 & -5.4 & 2.7 \\ 1.2 & -0.3 & -0.3 & 2.4 & 2.1 \end{array}\right]. \qquad \text{Thus,} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ \boxed{0.6x_1} + 1.5x_2 + 1.5x_3 - 5.4x_4 &= 2.7 \\ \boxed{1.2x_1} - 0.3x_2 - 0.3x_3 + 2.4x_4 &= 2.1. \end{aligned}$$

Solution. As in the previous example, we circle pivots and box terms of equations and corresponding entries to be eliminated. We indicate the operations in terms of equations and operate on both equations and matrices.

Step 1. Elimination of x_1 from the second and third equations by adding

$$-\,0.6/3.0 = -0.2 \text{ times the first equation to the second equation,}$$

$$-\,1.2/3.0 = -0.4 \text{ times the first equation to the third equation.}$$

This gives the following, in which the pivot of the next step is circled.

$$(6)\quad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & -1.1 & -1.1 & 4.4 & -1.1 \end{array}\right] \quad \begin{array}{l} \\ \text{Row 2} - 0.2 \text{ Row 1} \\ \text{Row 3} - 0.4 \text{ Row 1} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ \boxed{-1.1x_2} - 1.1x_3 + 4.4x_4 &= -1.1 \end{aligned}$$

Step 2. Elimination of x_2 from the third equation of (6) by adding

$$1.1/1.1 = 1 \text{ times the second equation to the third equation.}$$

This gives

$$(7)\quad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \begin{array}{l} \\ \\ \text{Row 3} + \text{Row 2} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ 0 &= 0\ . \end{aligned}$$

Back Substitution. From the second equation, $x_2 = 1 - x_3 + 4x_4$. From this and the first equation, $x_1 = 2 - x_4$. Since x_3 and x_4 remain arbitrary, we have infinitely many solutions. If we choose a value of x_3 and a value of x_4, then the corresponding values of x_1 and x_2 are uniquely determined.

On Notation. If unknowns remain arbitrary, it is also customary to denote them by other letters $t_1, t_2, \cdots$. In this example we may thus write $x_1 = 2 - x_4 = 2 - t_2$, $x_2 = 1 - x_3 + 4x_4 = 1 - t_1 + 4t_2$, $x_3 = t_1$ (first arbitrary unknown), $x_4 = t_2$ (second arbitrary unknown). ■

EXAMPLE 4 **Gauss Elimination if no Solution Exists**

What will happen if we apply the Gauss elimination to a linear system that has no solution? The answer is that in this case the method will show this fact by producing a contradiction. For instance, consider

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 \\ 6 & 2 & 4 & 6 \end{array}\right] \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ \boxed{2x_1} + x_2 + x_3 &= 0 \\ \boxed{6x_1} + 2x_2 + 4x_3 &= 6. \end{aligned}$$

Step 1. Elimination of x_1 from the second and third equations by adding

$$-\tfrac{2}{3} \text{ times the first equation to the second equation,}$$

$$-\tfrac{6}{3} = -2 \text{ times the first equation to the third equation.}$$

This gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{array}\right] \quad \begin{array}{l} \\ \text{Row 2} - \frac{2}{3} \text{ Row 1} \\ \text{Row 3} - 2 \text{ Row 1} \end{array} \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ \boxed{-2x_2} + 2x_3 &= 0. \end{aligned}$$

Step 2. Elimination of x_2 from the third equation gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{array}\right] \Rightarrow$$

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right] \quad \text{Row 3} - 6 \text{ Row 2} \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ 0 &= 12. \end{aligned}$$

The false statement $0 = 12$ shows that the system has no solution. ■

Row Echelon Form and Information From It

echelon

[éʃəlan]

At the end of the Gauss elimination the form of the coefficient matrix, the augmented matrix, and the system itself are called the **row echelon form.** In it, rows of zeros, if present, are the last rows, and in each nonzero row the leftmost nonzero entry is farther to the right than in the previous row. For instance, in Example 4 the coefficient matrix and its augmented in row echelon form are

$$\begin{bmatrix} 3 & 2 & 1 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right].$$

Note that we do not require that the leftmost nonzero entries be 1 since this would have no theoretic or numeric advantage. (The so-called *reduced echelon form,* in which those entries *are* 1, will be discussed in Sec. 7.8.)

Assume: the first r columns of A are linearly independent

or

Also, reordering the unknown variables $x_1, \dots, x_n$

(Example)

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & X \\ 0 & 0 & 0 & X \\ 0 & 0 & 1 & X \end{array}\right]$$

At the end of the Gauss elimination (before the back substitution) the row echelon form of the augmented matrix will be

$$\text{(8)} \qquad m\left\{ \overbrace{\left[\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ & c_{22} & \cdots & \cdots & c_{2n} & \tilde{b}_2 \\ & 0 & \ddots & & \vdots & \vdots \\ & & & k_{rr} & \cdots\ k_{rn} & \tilde{b}_r \\ & & & & & \tilde{b}_{r+1} \\ & & 0 & & & \vdots \\ & & & & & \tilde{b}_m \end{array}\right]}^{n} \right. .$$

Here, $r \leqq m$ and $a_{11} \neq 0$, $c_{22} \neq 0, \cdots, k_{rr} \neq 0$, and all the entries in the blue triangle as well as in the blue rectangle are zero. From this we see that with respect to solutions of the system with augmented matrix (8) (and thus with respect to the originally given system) there are three possible cases:

(a) ***Exactly one solution*** if $r = n$ and $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$, if present, are zero. To get the solution, solve the nth equation corresponding to (8) (which is $k_{nn}x_n = \tilde{b}_n$) for x_n, then the $(n-1)$st equation for x_{n-1}, and so on up the line. See Example 2, where $r = n = 3$ and $m = 4$.

$\begin{bmatrix} \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{bmatrix} = 0$ in (8) $\Leftarrow$

(b) ***Infinitely many solutions*** if $r < n$ and $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$, if present, are zero. To obtain any of these solutions, choose values of $x_{r+1}, \cdots, x_n$ arbitrarily. Then solve the rth equation for x_r, then the $(r-1)$st equation for x_{r-1}, and so on up the line. See Example 3.

$\begin{bmatrix} \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{bmatrix} \neq 0$ in (8) $\Leftarrow$

(c) ***No solution*** if $r < m$ and one of the entries $\tilde{b}_{r+1}, \cdots, \tilde{b}_m$ is not zero. See Example 4, where $r = 2 < m = 3$ and $\tilde{b}_{r+1} = \tilde{b}_3 = 12$.

(a) $\left[\begin{array}{cccc|c} a_{11} & \cdots & \cdots & a_{1n} & b_1 \\ 0 & c_{22} & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & k_{nn} & \tilde{b}_n \\ 0 & 0 & & 0 & 0 \end{array}\right] \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ -1 \end{bmatrix} = 0$

HW

Prob. 7.3.16

Prob. 7.3.24

20. (Wheatstone bridge) Show that if $R_x/R_3 = R_1/R_2$ in the figure, then $I = 0$. (R_0 is the resistance of the instrument by which I is measured.) This bridge is a method for determining R_x. R_1, R_2, R_3 are known. R_3 is variable. To get R_x, make $I = 0$ by varing R_3. Then calculate $R_x = R_3R_1/R_2$.

21. (Traffic flow) Methods of electrical circuit analysis have applications to other fields. For instance, applying the analog of Kirchhoff's current law, find the traffic flow (cars per hour) in the net of one-way streets (in the directions indicated by the arrows) shown in the figure. Is the solution unique?

22. (Models of markets) Determine the equilibrium solution ($D_1 = S_1$, $D_2 = S_2$) of the two-commodity market with linear model (D, S, P = demand, supply, price; index 1 = first commodity, index 2 = second commodity)

$$D_1 = 60 - 2P_1 - P_2, \qquad S_1 = 4P_1 - 2P_2 + 14$$
$$D_2 = 4P_1 - P_2 + 10, \qquad S_2 = 5P_2 - 2.$$

23. (Equivalence relation) By definition, an *equivalence relation* on a set is a relation satisfying three conditions (named as indicated):

(i) Each element A of the set is equivalent to itself (*"Reflexivity"*).

(ii) If A is equivalent to B, then B is equivalent to A (*"Symmetry"*).

(iii) If A is equivalent to B and B is equivalent to C, then A is equivalent to C (*"Transitivity"*).

Show that row equivalence of matrices satisfies these three conditions. *Hint.* Show that for each of the three elementary row operations these conditions hold.

24. PROJECT. Elementary Matrices. The idea is that elementary operations can be accomplished by matrix multiplication. If $\mathbf{A}$ is an $m \times n$ matrix on which we want to do an elementary operation, then there is a matrix $\mathbf{E}$ such that $\mathbf{EA}$ is the new matrix after the operation. Such an $\mathbf{E}$ is called an **elementary matrix.** This idea can be helpful, for instance, in the design of algorithms. (*Computationally,* it is generally preferable to do row operations *directly,* rather than by multiplication by $\mathbf{E}$.)

(a) Show that the following are elementary matrices, for interchanging Rows 2 and 3, for adding -5 times the first row to the third, and for multiplying the fourth row by 8.

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

Apply $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ to a vector and to a 4×3 matrix of your choice. Find $\mathbf{B} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$, where $\mathbf{A} = [a_{jk}]$ is the general 4×2 matrix. Is $\mathbf{B}$ equal to $\mathbf{C} = \mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$?

(b) Conclude that $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ are obtained by doing the corresponding elementary operations on the 4×4 unit matrix. Prove that *if* $\mathbf{M}$ *is obtained from* $\mathbf{A}$ *by an elementary row operation, then*

$$\mathbf{M} = \mathbf{EA},$$

where $\mathbf{E}$ *is obtained from the* $n \times n$ *unit matrix* $\mathbf{I}_n$ *by the same row operation.*

25. CAS PROJECT. Gauss Elimination and Back Substitution. Write a program for Gauss elimination and back substitution (a) that does not include pivoting, (b) that does include pivoting. Apply the programs to Probs. 13–16 and to some larger systems of your choice.

7.4 Linear Independence. Rank of a Matrix. Vector Space

In the last section we explained the Gauss elimination with back substitution, the most important numeric solution method for linear systems of equations. It appeared that such a system may have a unique solution or infinitely many solutions, or it may be inconsistent, that is, have no solution at all. Hence we are confronted with the questions of **existence and uniqueness of solutions.** We shall answer these questions in the next section. As the key concept for this (and other questions) we introduce the ***rank of a matrix.*** To define rank, we first need the following concepts, which are of general importance.

Linear Independence and Dependence of Vectors

Given any set of m vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ (with the same number of components), a **linear combination** of these vectors is an expression of the form

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)}$$

where $c_1, c_2, \cdots, c_m$ are any scalars. Now consider the equation

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0}. \quad \textbf{(1)}$$

(1) ⇒ $c_1 = \cdots = c_m = 0$

Clearly, this vector equation (1) holds if we choose all c_j's zero, because then it becomes $\mathbf{0} = \mathbf{0}$. If this is the only m-tuple of scalars for which (1) holds, then our vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ are said to form a *linearly independent set* or, more briefly, we call them **linearly independent.** Otherwise, if (1) also holds with scalars not all zero, we call these vectors **linearly dependent,** because then we can express (at least) one of them as a linear combination of the others. For instance, if (1) holds with, say, $c_1 \neq 0$, we can solve (1) for $\mathbf{a}_{(1)}$:

$$\mathbf{a}_{(1)} = k_2\mathbf{a}_{(2)} + \cdots + k_m\mathbf{a}_{(m)} \qquad \text{where } k_j = -c_j/c_1.$$

(Some k_j's may be zero. Or even all of them, namely, if $\mathbf{a}_{(1)} = \mathbf{0}$.)

Why is this important? Well, in the case of linear dependence we can get rid of some of the vectors until we arrive at a linearly independent set that is optimal to work with because it is smallest possible in the sense that it consists only of the "really essential" vectors, which can no longer be expressed linearly in terms of each other. This motivates the idea of a "basis" used in various contexts, notably later in our present section.

EXAMPLE 1 Linear Independence and Dependence

The three vectors

$$\begin{aligned}\mathbf{a}_{(1)} &= [\;3 \quad 0 \quad 2 \quad 2]\\ \mathbf{a}_{(2)} &= [\;-6 \quad 42 \quad 24 \quad 54]\\ \mathbf{a}_{(3)} &= [\;21 \quad -21 \quad 0 \quad -15]\end{aligned}$$

are linearly dependent because

$$6\mathbf{a}_{(1)} - \tfrac{1}{2}\mathbf{a}_{(2)} - \mathbf{a}_{(3)} = \mathbf{0}.$$

Although this is easily checked (do it!), it is not so easy to discover. However, a systematic method for finding out about linear independence and dependence follows below.

The first two of the three vectors are linearly independent because $c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} = \mathbf{0}$ implies $c_2 = 0$ (from the second components) and then $c_1 = 0$ (from any other component of $\mathbf{a}_{(1)}$). ■

Rank of a Matrix

DEFINITION

> The **rank** of a matrix **A** is the maximum number of linearly independent row vectors of **A**. It is denoted by rank **A**.

Further important properties will result from the basic

THEOREM 3

Rank in Terms of Column Vectors

The rank r of a matrix $\mathbf{A}$ equals the maximum number of linearly independent column vectors of $\mathbf{A}$.

Hence $\mathbf{A}$ and its transpose $\mathbf{A}^{\mathsf{T}}$ have the same rank.

PROOF In this proof we write simply "rows" and "columns" for row and column vectors. Let $\mathbf{A}$ be an $m \times n$ matrix of rank $\mathbf{A} = r$. Then by definition of rank, $\mathbf{A}$ has r linearly independent rows which we denote by $\mathbf{v}_{(1)}, \cdots, \mathbf{v}_{(r)}$ (regardless of their position in $\mathbf{A}$), and all the rows $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ of $\mathbf{A}$ are linear combinations of those, say,

$$A = \begin{bmatrix} a_{(1)} \\ \vdots \\ a_{(m)} \end{bmatrix}$$

$$\begin{aligned}
\mathbf{a}_{(1)} &= c_{11}\mathbf{v}_{(1)} + c_{12}\mathbf{v}_{(2)} + \cdots + c_{1r}\mathbf{v}_{(r)} \\
\mathbf{a}_{(2)} &= c_{21}\mathbf{v}_{(1)} + c_{22}\mathbf{v}_{(2)} + \cdots + c_{2r}\mathbf{v}_{(r)} \\
&\ \ \vdots \\
\mathbf{a}_{(m)} &= c_{m1}\mathbf{v}_{(1)} + c_{m2}\mathbf{v}_{(2)} + \cdots + c_{mr}\mathbf{v}_{(r)}.
\end{aligned} \tag{3}$$

These are vector equations for rows. To switch to columns, we write (3) in terms of components as n such systems, with $k = 1, \cdots, n$,

$$\begin{aligned}
a_{1k} &= c_{11}v_{1k} + c_{12}v_{2k} + \cdots + c_{1r}v_{rk} \\
a_{2k} &= c_{21}v_{1k} + c_{22}v_{2k} + \cdots + c_{2r}v_{rk} \\
&\ \ \vdots \\
a_{mk} &= c_{m1}v_{1k} + c_{m2}v_{2k} + \cdots + c_{mr}v_{rk}
\end{aligned} \tag{4}$$

and collect components in columns. Indeed, we can write (4) as

$$\begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = v_{1k}\begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} + v_{2k}\begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix} + \cdots + v_{rk}\begin{bmatrix} c_{1r} \\ c_{2r} \\ \vdots \\ c_{mr} \end{bmatrix} \tag{5}$$

where $k = 1, \cdots, n$. Now the vector on the left is the kth column vector of $\mathbf{A}$. We see that each of these n columns is a linear combination of the same r columns on the right. Hence $\mathbf{A}$ cannot have more linearly independent columns than rows, whose number is rank $\mathbf{A} = r$. Now rows of $\mathbf{A}$ are columns of the transpose $\mathbf{A}^{\mathsf{T}}$. For $\mathbf{A}^{\mathsf{T}}$ our conclusion is that $\mathbf{A}^{\mathsf{T}}$ cannot have more linearly independent columns than rows, so that $\mathbf{A}$ cannot have more linearly independent rows than columns. Together, the number of linearly independent columns of $\mathbf{A}$ must be r, the rank of $\mathbf{A}$. This completes the proof. ■

EXAMPLE 4 **Illustration of Theorem 3**

The matrix in (2) has rank 2. From Example 3 we see that the first two row vectors are linearly independent and by "working backward" we can verify that Row 3 = 6 Row 1 $-\frac{1}{2}$ Row 2. Similarly, the first two columns are linearly independent, and by reducing the last matrix in Example 3 by columns we find that

Column 3 = $\frac{2}{3}$ Column 1 + $\frac{2}{3}$ Column 2 and Column 4 = $\frac{2}{3}$ Column 1 + $\frac{29}{21}$ Column 2. ■

Our further discussion will show that the rank of a matrix is an important key concept for understanding general properties of matrices and linear systems of equations.

EXAMPLE 2 Rank

The matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \tag{2}$$

has rank 2, because Example 1 shows that the first two row vectors are linearly independent, whereas all three row vectors are linearly dependent.

Note further that rank $\mathbf{A} = 0$ if and only if $\mathbf{A} = \mathbf{0}$. This follows directly from the definition. ■

We call a matrix $\mathbf{A}_1$ **row-equivalent** to a matrix $\mathbf{A}_2$ if $\mathbf{A}_1$ can be obtained from $\mathbf{A}_2$ by (finitely many!) elementary row operations.

Now the maximum number of linearly independent row vectors of a matrix does not change if we change the order of rows or multiply a row by an nonzero c or take a linear combination by adding a multiple of a row to another row. This proves that rank is **invariant** under elementary row operations:

THEOREM 1 Row-Equivalent Matrices

Row-equivalent matrices have the same rank.

Hence we can determine the rank of a matrix by reduction to row-echelon form (Sec. 7.3) and then see the rank directly.

EXAMPLE 3 Determination of Rank

For the matrix in Example 2 we obtain successively

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \quad \text{(given)}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & -21 & -14 & -29 \end{bmatrix} \quad \begin{matrix} \\ \text{Row 2} + 2 \text{ Row 1} \\ \text{Row 3} - 7 \text{ Row 1} \end{matrix}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \\ \\ \text{Row 3} + \tfrac{1}{2} \text{ Row 2} \end{matrix}$$ ■

Since rank is defined in terms of two vectors, we immediately have the useful

THEOREM 2 Linear Independence and Dependence of Vectors

p vectors with n components each are linearly independent if the matrix with these vectors as row vectors has rank p, but they are linearly dependent if that rank is less than p.

Remark

$$\text{rank}\begin{bmatrix} a_{(1)} \\ a_{(2)} \\ \vdots \\ a_{(m)} \end{bmatrix} = \text{rank}\begin{bmatrix} a_{(1)} \\ a_{(2)} \\ \alpha_1 a_{(1)} + \alpha_2 a_{(2)} + a_{(3)} \\ \vdots \\ a_{(m)} \end{bmatrix} = \text{rank}\begin{bmatrix} a_{(1)} \\ c\, a_{(2)} \\ \vdots \\ a_{(m)} \end{bmatrix} = \text{rank}\begin{bmatrix} a_{(2)} \\ a_{(1)} \\ \vdots \\ a_{(m)} \end{bmatrix}$$

Combining Theorems 2 and 3 we obtain

THEOREM 4 **Linear Dependence of Vectors**

p vectors with $n < p$ components are always linearly dependent.

n columns
p rows

PROOF The matrix $\mathbf{A}$ with those p vectors as row vectors has p rows and $n < p$ columns; hence by Theorem 3 it has rank $\mathbf{A} \leqq n < p$, which implies linear dependence by Theorem 2. ■

$$A = \begin{bmatrix} a_{(1)} \\ \vdots \\ a_{(p)} \end{bmatrix} \in R^{p \times n}$$

Vector Space

The following related concepts are of general interest in linear algebra. In the present context they provide a clarification of essential properties of matrices and their role in connection with linear systems.

A **vector space** is a (nonempty) set V of vectors such that with any two vectors $\mathbf{a}$ and $\mathbf{b}$ in V all their linear combinations $\alpha\mathbf{a} + \beta\mathbf{b}$ (α, β any real numbers) are elements of V, and these vectors satisfy the laws (3) and (4) in Sec. 7.1 (written in lowercase letters $\mathbf{a}$, $\mathbf{b}$, $\mathbf{u}$, $\cdots$, which is our notation for vectors). (This definition is presently sufficient. General vector spaces will be discussed in Sec. 7.9.)

$a \in V$, $b \in V$
$\Downarrow$
$\alpha a + \beta b \in V$

The maximum number of linearly independent vectors in V is called the **dimension** of V and is denoted by dim V. Here we assume the dimension to be finite; infinite dimension will be defined in Sec. 7.9.

A linearly independent set in V consisting of a maximum possible number of vectors in V is called a **basis** for V. Thus the number of vectors of a basis for V equals dim V.

The set of all linear combinations of given vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(p)}$ with the same number of components is called the **span** of these vectors. Obviously, a span is a vector space.

$W \subset V$

This holds for W instead of V

By a **subspace** of a vector space V we mean a nonempty subset of V (including V itself) that forms itself a vector space with respect to the two algebraic operations (addition and scalar multiplication) defined for the vectors of V.

EXAMPLE 5 **Vector Space, Dimension, Basis**

The span of the three vectors in Example 1 is a vector space of dimension 2, and a basis is $\mathbf{a}_{(1)}$, $\mathbf{a}_{(2)}$, for instance, or $\mathbf{a}_{(1)}$, $\mathbf{a}_{(3)}$, etc. ■

We further note the simple

THEOREM 5 **Vector Space R^n**

The vector space R^n consisting of all vectors with n components (n real numbers) has dimension n.

PROOF A basis of n vectors is $\mathbf{a}_{(1)} = [1 \quad 0 \quad \cdots \quad 0]$, $\mathbf{a}_{(2)} = [0 \quad 1 \quad 0 \quad \cdots \quad 0]$, $\cdots$, $\mathbf{a}_{(n)} = [0 \quad \cdots \quad 0 \quad 1]$. ■

In the case of a matrix $\mathbf{A}$ we call the span of the row vectors the **row space** of $\mathbf{A}$ and the span of the column vectors the **column space** of $\mathbf{A}$.

Now, Theorem 3 shows that a matrix **A** has as many linearly independent rows as columns. By the definition of dimension, their number is the dimension of the row space or the column space of **A.** This proves

$N(A) \triangleq \{ x \in R^n \mid Ax = 0 \}$

Let $u, v \in N(A)$.

Then,

THEOREM 6 **Row Space and Column Space**

The row space and the column space of a matrix **A** *have the same dimension, equal to rank* **A.**

Finally, for a given matrix **A** the solution set of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ is a vector space, called the **null space** of **A,** and its dimension is called the **nullity** of **A.** In the next section we motivate and prove the basic relation

$$\text{rank } \mathbf{A} + \text{nullity } \mathbf{A} = \text{Number of columns of } \mathbf{A}. \tag{6}$$

$A(\alpha u + \beta v) = \alpha Au + \beta Av = 0 \Rightarrow \alpha u + \beta v \in N(A) \Rightarrow N(A)$ is a subspace of R^n

PROBLEM SET 7.4

1–12 RANK, ROW SPACE, COLUMN SPACE

Find the rank and a basis for the row space and for the column space. *Hint.* Row-reduce the matrix and its transpose. (You may omit obvious factors from the vectors of these bases.)

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

2. $\begin{bmatrix} 8 & 2 & 5 \\ 16 & 6 & 29 \\ 4 & 0 & -7 \end{bmatrix}$

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

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

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

6. $\begin{bmatrix} 1 & 1 & a \\ 1 & a & 1 \\ a & 1 & 1 \end{bmatrix}$

7. $\begin{bmatrix} 8 & 0 & 4 \\ 0 & 2 & 0 \\ 4 & 0 & 2 \\ 0 & 4 & 0 \end{bmatrix}$

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

9. $\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 5 & 8 & -37 \\ 3 & 8 & 7 & 0 \\ 0 & -37 & 0 & 37 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix}$

11. $\begin{bmatrix} 2 & 4 & 8 & 16 \\ 16 & 8 & 4 & 2 \\ 4 & 8 & 16 & 2 \\ 2 & 16 & 8 & 4 \end{bmatrix}$

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

13–20 LINEAR INDEPENDENCE

Are the following sets of vectors linearly independent? (Show the details.)

13. [3 −2 0 4], [5 0 0 1], [−6 1 0 1], [2 0 0 3]

14. [1 1 0], [1 0 0], [1 1 1]

15. [6 0 3 1 4 2], [0 −1 2 7 0 5], [12 3 0 −19 8 −11]

16. [3 4 7], [2 0 3], [8 2 3], [5 5 6]

17. [0.2 1.2 5.3 2.8 1.6], [4.3 3.4 0.9 2.0 −4.3]

HW

Prob. 7.4.9

Prob. 7.4.22 ~ 26

18. [3 2 1], [0 0 0], [4 3 6]

19. $[1\ \ \frac{1}{2}\ \ \frac{1}{3}\ \ \frac{1}{4}]$, $[\frac{1}{2}\ \ \frac{1}{3}\ \ \frac{1}{4}\ \ \frac{1}{5}]$, $[\frac{1}{3}\ \ \frac{1}{4}\ \ \frac{1}{5}\ \ \frac{1}{6}]$, $[\frac{1}{4}\ \ \frac{1}{5}\ \ \frac{1}{6}\ \ \frac{1}{7}]$

20. [1 2 3 4], [2 3 4 5], [3 4 5 6], [4 5 6 7]

21. CAS Experiment. Rank. (a) Show experimentally that the $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ with $a_{jk} = j + k - 1$ has rank 2 for any n. (Problem 20 shows $n = 4$.) Try to prove it.

(b) Do the same when $a_{jk} = j + k + c$, where c is any positive integer.

(c) What is rank $\mathbf{A}$ if $a_{jk} = 2^{j+k-2}$? Try to find other large matrices of low rank independent of n.

22–26 PROPERTIES OF RANK AND CONSEQUENCES

Show the following.

22. rank $\mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$ = rank $\mathbf{AB}$. (Note the order!)

23. rank $\mathbf{A}$ = rank $\mathbf{B}$ does *not* imply rank $\mathbf{A}^2$ = rank $\mathbf{B}^2$. (Give a counterexample.)

24. If $\mathbf{A}$ is not square, either the row vectors or the column vectors of $\mathbf{A}$ are linearly dependent.

25. If the row vectors of a square matrix are linearly independent, so are the column vectors, and conversely.

26. Give examples showing that the rank of a product of matrices cannot exceed the rank of either factor.

27–36 VECTOR SPACES

Is the given set of vectors a vector space? (Give reason.) If your answer is yes, determine the dimension and find a basis. ($v_1, v_2, \cdots$ denote components.)

27. All vectors in R^3 such that $v_1 + v_2 = 0$

28. All vectors in R^4 such that $2v_2 - 3v_4 = k$

29. All vectors in R^3 with $v_1 \geqq 0$, $v_2 = -4v_3$

30. All vectors in R^2 with $v_1 \leqq v_2$

31. All vectors in R^3 with $4v_1 + v_3 = 0$, $3v_2 = v_3$

32. All vectors in R^4 with $v_1 - v_2 = 0$, $v_3 = 5v_1$, $v_4 = 0$

33. All vectors in R^n with $|v_j| \leqq 1$ for $j = 1, \cdots, n$

34. All ordered quadruples of positive real numbers

35. All vectors in R^5 with $v_1 = 2v_2 = 3v_3 = 4v_4 = 5v_5$

36. All vectors in R^4 with $3v_1 - v_3 = 0$, $2v_1 + 3v_2 - 4v_4 = 0$

7.5 Solutions of Linear Systems: Existence, Uniqueness

Rank as just defined gives complete information about existence, uniqueness, and general structure of the solution set of linear systems as follows.

A linear system of equations in n unknowns has a unique solution if the coefficient matrix and the augmented matrix have the same rank n, and infinitely many solution if that common rank is less than n. The system has no solution if those two matrices have different rank.

To state this precisely and prove it, we shall use the (generally important) concept of a **submatrix** of **A**. By this we mean any matrix obtained from **A** by omitting some rows or columns (or both). By definition this includes **A** itself (as the matrix obtained by omitting no rows or columns); this is practical.

THEOREM 1

Fundamental Theorem for Linear Systems

(a) Existence. *A linear system of m equations in n unknowns $x_1, \cdots, x_n$*

$$
(1)\qquad
\begin{array}{c}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m
\end{array}
$$

$Ax = b$ where $x = [x_1 \cdots x_n]^T$

$b = [b_1 \cdots b_n]^T$

is **consistent,** *that is, has solutions, if and only if the coefficient matrix* **A** *and the augmented matrix* $\tilde{\mathbf{A}}$ *have the same rank. Here,*

common

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

$Ax = b$

(b) Uniqueness. *The system* (1) *has precisely one solution if and only if* ~~*this*~~ *common rank r of* **A** *and* $\tilde{\mathbf{A}}$ *equals n.* ($\Rightarrow \exists$ a solution)

(c) Infinitely many solutions. *If this common rank r is less than n, the system* (1) *has infinitely many solutions. All of these solutions are obtained by determining r suitable unknowns* (*whose submatrix of coefficients must have rank r*) *in terms of the remaining n − r unknowns, to which arbitrary values can be assigned.* (See Example 3 in Sec. 7.3.)

$A = [c_{(1)} \cdots c_{(n)}]$

$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

(d) Gauss elimination (Sec. 7.3). *If solutions exist, they can all be obtained by the Gauss elimination.* (This method will automatically reveal whether or not solutions exist; see Sec. 7.3.)

PROOF **(a)** We can write the system (1) in vector form $\mathbf{Ax} = \mathbf{b}$ or in terms of column vectors $\mathbf{c}_{(1)}, \cdots, \mathbf{c}_{(n)}$ of **A**:

$\tilde{A} = [c_{(1)} \cdots c_{(n)}\ b]$

$$\mathbf{c}_{(1)}x_1 + \mathbf{c}_{(2)}x_2 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{b}. \tag{2}$$

$\tilde{\mathbf{A}}$ is obtained by augmenting **A** by a single column **b.** Hence, by Theorem 3 in Sec. 7.4, rank $\tilde{\mathbf{A}}$ equals rank **A** or rank **A** + 1. Now if (1) has a solution **x,** then (2) shows that **b** must be a linear combination of those column vectors, so that $\tilde{\mathbf{A}}$ and **A** have the same maximum number of linearly independent column vectors and thus the same rank.

$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$

Conversely, if rank $\tilde{\mathbf{A}}$ = rank **A,** then **b** must be a linear combination of the column vectors of **A,** say,

$$\mathbf{b} = \alpha_1\mathbf{c}_{(1)} + \cdots + \alpha_n\mathbf{c}_{(n)} \tag{2*}$$

since otherwise rank $\tilde{\mathbf{A}}$ = rank **A** + 1. But (2*) means that (1) has a solution, namely, $x_1 = \alpha_1, \cdots, x_n = \alpha_n$, as can be seen by comparing (2*) and (2).

(b) If rank **A** = n, the n column vectors in (2) are linearly independent by Theorem 3 in Sec. 7.4. We claim that then the representation (2) of **b** is unique because otherwise

$$\mathbf{c}_{(1)}x_1 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{c}_{(1)}\tilde{x}_1 + \cdots + \mathbf{c}_{(n)}\tilde{x}_n. \tag{*}$$

This would imply (take all terms to the left, with a minus sign)

$$(x_1 - \tilde{x}_1)\mathbf{c}_{(1)} + \cdots + (x_n - \tilde{x}_n)\mathbf{c}_{(n)} = \mathbf{0} \tag{*'}$$

and $x_1 - \tilde{x}_1 = 0, \cdots, x_n - \tilde{x}_n = 0$ by linear independence. But this means that the scalars $x_1, \cdots, x_n$ in (2) are uniquely determined, that is, the solution of (1) is unique.

If the solution of (1) is unique, then (*) implies $x_1 = \tilde{x}_1, \ldots, x_n = \tilde{x}_n$. In otherwords, (*)' implies always that all coefficients $(x_i - \tilde{x}_i)$, $i = 1, \ldots, n$ are zero. By the definition of vector independence, this implies that rank A = rank $\tilde{A} = n$.

(c) If rank $\mathbf{A}$ = rank $\widetilde{\mathbf{A}} = r < n$, then by Theorem 3 in Sec. 7.4 there is a linearly independent set K of r column vectors of $\mathbf{A}$ such that the other $n - r$ column vectors of $\mathbf{A}$ are linear combinations of those vectors. We renumber the columns and unknowns, denoting the renumbered quantities by ˆ, so that $\{\hat{\mathbf{c}}_{(1)}, \cdots, \hat{\mathbf{c}}_{(r)}\}$ is that linearly independent set K. Then (2) becomes

$$c_{(1)}x_1 + c_{(2)}x_2 + \cdots + c_{(n)}x_n = \hat{\mathbf{c}}_{(1)}\hat{x}_1 + \cdots + \hat{\mathbf{c}}_{(r)}\hat{x}_r + \hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1} + \cdots + \hat{\mathbf{c}}_{(n)}\hat{x}_n = \mathbf{b},$$

$Ax = \hat{A}\hat{x} = b$

$\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{\mathbf{c}}_{(n)}$ are linear combinations of the vectors of K, and so are the vectors $\hat{x}_{r+1}\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{x}_n\hat{\mathbf{c}}_{(n)}$. Expressing these vectors in terms of the vectors of K and collecting terms, we can thus write the system in the form

$$\hat{\mathbf{c}}_{(1)}y_1 + \cdots + \hat{\mathbf{c}}_{(r)}y_r = \mathbf{b} \tag{3}$$

with $y_j = \hat{x}_j + \beta_j$, where β_j results from the $n - r$ terms $\hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1}, \cdots, \hat{\mathbf{c}}_{(n)}\hat{x}_n$; here, $j = 1, \cdots, r$. Since the system has a solution, there are $y_1, \cdots, y_r$ satisfying (3). These scalars are unique since K is linearly independent. Choosing $\underbrace{\hat{x}_{r+1}, \cdots, \hat{x}_n}_{(n-r)개}$ fixes the β_j and corresponding $\hat{x}_j = y_j - \beta_j$, where $j = 1, \cdots, r$.

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_r \\ x \\ \vdots \\ x \end{bmatrix} \begin{matrix} \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} (n-r)개 \end{matrix}$$

(d) This was discussed in Sec. 7.3 and is restated here as a reminder. ∎

The theorem is illustrated in Sec. 7.3. In Example 2 there is a unique solution since rank $\widetilde{\mathbf{A}}$ = rank $\mathbf{A} = n = 3$ (as can be seen from the last matrix in the example). In Example 3 we have rank $\widetilde{\mathbf{A}}$ = rank $\mathbf{A} = 2 < n = 4$ and can choose x_3 and x_4 arbitrarily. In Example 4 there is no solution because rank $\mathbf{A} = 2 <$ rank $\widetilde{\mathbf{A}} = 3$.

Homogeneous Linear System

Recall from Sec. 7.3 that a linear system (1) is called **homogeneous** if all the b_j's are zero, and **nonhomogeneous** if one or several b_j's are not zero. For the homogeneous system we obtain from the Fundamental Theorem the following results.

THEOREM 2

Homogeneous Linear System

A homogeneous linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{4}$$

$Ax = 0$

always has the **trivial solution** $x_1 = 0, \cdots, x_n = 0$. *Nontrivial solutions exist if and only if rank* $\mathbf{A} < n$. *If* rank $\mathbf{A} = r < n$, *these solutions, together with* $\mathbf{x} = \mathbf{0}$, *form a vector space* (see Sec. 7.4) *of dimension* $n - r$, called the **solution space** of (4).

In particular, if $\mathbf{x}_{(1)}$ *and* $\mathbf{x}_{(2)}$ *are solution vectors of* (4), *then* $\mathbf{x} = \mathbf{c}_1\mathbf{x}_{(1)} + \mathbf{c}_2\mathbf{x}_{(2)}$ *with any scalars* c_1 *and* c_2 *is a solution vector of* (4). (This ***does not hold*** for nonhomogeneous systems. Also, the term *solution space* is used for *homogeneous* systems only.)

$$c_{(1)}x_1 + c_{(2)}x_2 + \cdots + c_{(n)}x_n = 0$$

where

$$c_{(k)} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

$\mathcal{N}(A) \triangleq [\{x \in R^n \mid Ax = 0\}$

$x_{(1)} \in \mathcal{N}(A)$

$x_{(2)} \in \mathcal{N}(A)$

$\Downarrow$

$A(c_1 x_{(1)} + c_2 x_{(2)}) = 0$

$\Downarrow$

$c_1 x_{(1)} + c_2 x_{(2)} \in \mathcal{N}(A)$

PROOF The first proposition can be seen directly from the system. It agrees with the fact that $\mathbf{b} = \mathbf{0}$ implies that rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$, so that a homogeneous system is always ***consistent.*** If rank $\mathbf{A} = n$, the trivial solution is the unique solution according to (b) in Theorem 1. If rank $\mathbf{A} < n$, there are nontrivial solutions according to (c) in Theorem 1. The solutions form a vector space because if $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are any of them, then $\mathbf{Ax}_{(1)} = \mathbf{0}$, $\mathbf{Ax}_{(2)} = \mathbf{0}$, and this implies $\mathbf{A}(\mathbf{x}_{(1)} + \mathbf{x}_{(2)}) = \mathbf{Ax}_{(1)} + \mathbf{Ax}_{(2)} = \mathbf{0}$ as well as $\mathbf{A}(c\mathbf{x}_{(1)}) = c\mathbf{Ax}_{(1)} = \mathbf{0}$, where c is arbitrary. If rank $\mathbf{A} = r < n$, Theorem 1 (c) implies that we can choose $n - r$ suitable unknowns, call them $\hat{x}_{r+1}, \cdots, \hat{x}_n$, in an arbitrary fashion, and every solution is obtained in this way. Hence a basis for the solution space, briefly called **a basis of solutions** of (4), is $\mathbf{y}_{(1)}, \cdots, \mathbf{y}_{(n-r)}$, where the basis vector $\mathbf{y}_{(j)}$ is obtained by choosing $\hat{x}_{r+j} = 1$ and the other $\hat{x}_{r+1}, \cdots, \hat{x}_n$ zero; the corresponding first r components of this solution vector are then determined. Thus the solution space of (4) has dimension $n - r$. This proves Theorem 2. ■

$\hat{A} y_{(j)} = 0$, $j = 1, \cdots, (n-r)$

$\mathcal{N}(A)$

The solution space of (4) is also called the **null space** of $\mathbf{A}$ because $\mathbf{Ax} = \mathbf{0}$ for every $\mathbf{x}$ in the solution space of (4). Its dimension is called the **nullity** of $\mathbf{A}$. Hence Theorem 2 states that

$$\underbrace{\text{rank } \mathbf{A}}_{r} + \underbrace{\text{nullity } \mathbf{A}}_{(n-r)} = n \tag{5}$$

where n is the number of unknowns (number of columns of $\mathbf{A}$).

Furthermore, by the definition of rank we have rank $\mathbf{A} \leqq m$ in (4). Hence if $m < n$, then rank $\mathbf{A} < n$. By Theorem 2 this gives the practically important

THEOREM 3 **Homogeneous Linear System with Fewer Equations Than Unknowns**

A homogeneous linear system with fewer equations than unknowns has always nontrivial solutions.

$n - r > 0$

Nonhomogeneous Linear Systems

The characterization of all solutions of the linear system (1) is now quite simple, as follows.

THEOREM 4 **Nonhomogeneous Linear System**

If a nonhomogeneous linear system (1) *is consistent, then all of its solutions are obtained as*

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_h \tag{6}$$

where $\mathbf{x}_0$ *is any* (fixed) *solution of* (1) *and* $\mathbf{x}_h$ *runs through all the solutions of the corresponding homogeneous system* (4).

PROOF The difference $\mathbf{x}_h = \mathbf{x} - \mathbf{x}_0$ of any two solutions of (1) is a solution of (4) because $\mathbf{Ax}_h = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{Ax} - \mathbf{Ax}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Since $\mathbf{x}$ is any solution of (1), we get all the solutions of (1) if in (6) we take any solution $\mathbf{x}_0$ of (1) and let $\mathbf{x}_h$ vary throughout the solution space of (4). ■

7.6 For Reference: Second- and Third-Order Determinants

We explain these determinants separately from the general theory in Sec. 7.7 because they will be sufficient for many of our examples and problems. Since this section is for reference, ***go on to the next section, consulting this material only when needed.***

A **determinant of second order** is denoted and defined by

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{1}$$

So here we have ***bars*** (whereas a matrix has ***brackets***).

Cramer's rule for solving linear systems of two equations in two unknowns

$$\text{(2)} \qquad \begin{aligned} &\text{(a)} \quad a_{11}x_1 + a_{12}x_2 = b_1 \\ &\text{(b)} \quad a_{21}x_1 + a_{22}x_2 = b_2 \end{aligned}$$

is

$$\text{(3)} \qquad \begin{aligned} x_1 &= \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{D} = \frac{b_1a_{22} - a_{12}b_2}{D}, \\ x_2 &= \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{D} = \frac{a_{11}b_2 - b_1a_{21}}{D}. \end{aligned}$$

with D as in (1), provided

$$D \neq 0.$$

The value $D = 0$ appears for inconsistent nonhomogeneous systems and for homogeneous systems with nontrivial solutions.

X **PROOF** We prove (3). To eliminate x_2, multiply (2a) by a_{22} and (2b) by $-a_{12}$ and add,

$$(a_{11}a_{22} - a_{12}a_{21})x_1 = b_1a_{22} - a_{12}b_2.$$

Similarly, to eliminate x_1, multiply (2a) by $-a_{21}$ and (2b) by a_{11} and add,

$$(a_{11}a_{22} - a_{12}a_{21})x_2 = a_{11}b_2 - b_1a_{21}.$$

Assuming that $D = a_{11}a_{22} - a_{12}a_{21} \neq 0$, dividing, and writing the right sides of these two equations as determinants, we obtain (3). ∎

EXAMPLE 1 Cramer's Rule for Two Equations

If
$$\begin{aligned} 4x_1 + 3x_2 &= 12 \\ 2x_1 + 5x_2 &= -8 \end{aligned}$$
then
$$x_1 = \frac{\begin{vmatrix} 12 & 3 \\ -8 & 5 \end{vmatrix}}{\begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix}} = \frac{84}{14} = 6, \qquad x_2 = \frac{\begin{vmatrix} 4 & 12 \\ 2 & -8 \end{vmatrix}}{\begin{vmatrix} 4 & 3 \\ 2 & 5 \end{vmatrix}} = \frac{-56}{14} = -4.$$ ■

Third-Order Determinants

A **determinant of third order** can be defined by

$$\text{(4)} \qquad D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

Note the following. The signs on the right are $+ - +$. Each of the three terms on the right is an entry in the first column of D times its **minor,** that is, the second-order determinant obtained from D by deleting the row and column of that entry; thus, for a_{11} delete the first row and first column, and so on.

If we write out the minors in (4), we obtain

$$\text{(4*)} \qquad D = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{13}a_{32} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{13}a_{22}.$$

Cramer's Rule for Linear Systems of Three Equations

$$\text{(5)} \qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$$

is

$$\text{(6)} \qquad x_1 = \frac{D_1}{D}, \qquad x_2 = \frac{D_2}{D}, \qquad x_3 = \frac{D_3}{D} \qquad (D \neq 0)$$

with the *determinant D of the system* given by (4) and

$$D_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}, \qquad D_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}, \qquad D_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}.$$

Note that D_1, D_2, D_3 are obtained by replacing Columns 1, 2, 3, respectively, by the column of the right sides of (5).

Cramer's rule (6) can be derived by eliminations similar to those for (3), but it also follows from the general case (Theorem 4) in the next section.

7.7 Determinants. Cramer's Rule

Determinants were originally introduced for solving linear systems. Although ***impractical in computations,*** they have important engineering applications in eigenvalue problems (Sec. 8.1), differential equations, vector algebra (Sec. 9.3), and so on. They can be introduced in several equivalent ways. Our definition is particularly practical in connection with linear systems.

A **determinant of order** n is a scalar associated with an $n \times n$ (hence ***square!***) matrix $\mathbf{A} = [a_{jk}]$, which is written

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{1}$$

and is defined for $n = 1$ by

$$D = a_{11} \tag{2}$$

and for $n \geqq 2$ by

$$D = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn} \qquad (j = 1, 2, \cdots, \text{ or } n) \tag{3a}$$

or

$$D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk} \qquad (k = 1, 2, \cdots, \text{ or } n) \tag{3b}$$

Here,

$$C_{jk} = (-1)^{j+k}M_{jk}$$

and M_{jk} is a determinant of order $n - 1$, namely, the determinant of the submatrix of $\mathbf{A}$ obtained from $\mathbf{A}$ by omitting the row and column of the entry a_{jk}, that is, the jth row and the kth column.

In this way, D is defined in terms of n determinants of order $n - 1$, each of which is, in turn, defined in terms of $n - 1$ determinants of order $n - 2$, and so on; we finally arrive at second-order determinants, in which those submatrices consist of single entries whose determinant is defined to be the entry itself.

From the definition it follows that *we may* ***expand*** *D by any row or column,* that is, choose in (3) the entries in any row or column, similarly when expanding the C_{jk}'s in (3), and so on.

This definition is unambiguous, that is, yields the same value for D no matter which columns or rows we choose in expanding. A proof is given in App. 4.

Terms used in connection with determinants are taken from matrices. In D we have n^2 **entries** a_{jk}, also n **rows** and n **columns,** and a **main diagonal** on which $a_{11}, a_{22}, \cdots, a_{nn}$ stand. Two terms are new:

M_{jk} is called the **minor** *of* a_{jk} *in* D, and C_{jk} the **cofactor** *of* a_{jk} *in* D.

For later use we note that (3) may also be written in terms of minors

$C_{jk} = (-1)^{j+k} M_{jk}$

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (j = 1, 2, \cdots, \text{ or } n) \tag{4a}$$

$$D = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (k = 1, 2, \cdots, \text{ or } n). \tag{4b}$$

EXAMPLE 1 Minors and Cofactors of a Third-Order Determinant

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

In (4) of the previous section the minors and cofactors of the entries in the first column can be seen directly. For the entries in the second row the minors are

$$M_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}, \qquad M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \qquad M_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

and the cofactors are $C_{21} = -M_{21}$, $C_{22} = +M_{22}$, and $C_{23} = -M_{23}$. Similarly for the third row—write these down yourself. And verify that the signs in C_{jk} form a **checkerboard pattern**

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

■

EXAMPLE 2 Expansions of a Third-Order Determinant

$$D = \begin{vmatrix} 1 & 3 & 0 \\ 2 & 6 & 4 \\ -1 & 0 & 2 \end{vmatrix} = 1\begin{vmatrix} 6 & 4 \\ 0 & 2 \end{vmatrix} - 3\begin{vmatrix} 2 & 4 \\ -1 & 2 \end{vmatrix} + 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix}$$

$$= 1(12 - 0) - 3(4 + 4) + 0(0 + 6) = -12.$$

This is the expansion by the first row. The expansion by the third column is

$$D = 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix} - 4\begin{vmatrix} 1 & 3 \\ -1 & 0 \end{vmatrix} + 2\begin{vmatrix} 1 & 3 \\ 2 & 6 \end{vmatrix} = 0 - 12 + 0 = -12,$$

Verify that the other four expansions also give the value -12. ■

EXAMPLE 3 Determinant of a Triangular Matrix

$$\begin{vmatrix} -3 & 0 & 0 \\ 6 & 4 & 0 \\ -1 & 2 & 5 \end{vmatrix} = -3\begin{vmatrix} 4 & 0 \\ 2 & 5 \end{vmatrix} = -3 \cdot 4 \cdot 5 = -60.$$

Inspired by this, can you formulate a little theorem on determinants of triangular matrices? Of diagonal matrices? ■

Remarks

$\det A = \sum (\pm) a_{1*} a_{2*} \cdots a_{n*}$

where the sequence of second subscript is a permutation of $1, 2, \cdots, n$. The $+$ sign is chosen if the number of inversions in this sequence is even. The number of inversion is the number of pairs in which a large number precedes a smaller.

General Properties of Determinants

To obtain the value of a determinant (1), we can first simplify it systematically by elementary row operations, similar to those for matrices in Sec. 7.3, as follows.

THEOREM 1 **Behavior of an nth-Order Determinant under Elementary Row Operations**

(a) *Interchange of two rows multiplies the value of the determinant by* -1.

(b) *Addition of a multiple of a row to another row does not alter the value of the determinant.*

(c) *Multiplication of a row by a nonzero constant c multiplies the value of the determinant by c.* (This holds also when $c = 0$, but gives no longer an elementary row operation.)

PROOF **(a)** By induction. The statement holds for $n = 2$ because

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \qquad \text{but} \qquad \begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad.$$

$n \geqq 3$

Now, suppose that the statement holds for $n = l \, (= n-1)$

We now make the induction hypothesis that (a) holds for determinants of order $n - 1 \geqq 2$ and show that it then holds for determinants of order n. Let D be of order n. Let E be obtained from D by the interchange of two rows. Expand D and E by a row that is **not** one of those interchanged, call it the jth row. Then by (4a),

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}, \qquad E = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} N_{jk} \tag{5}$$

where N_{jk} is obtained from the minor M_{jk} of a_{jk} in D by the interchange of those two rows which have been interchanged in D (and which N_{jk} must both contain because we expand by another row!). Now these minors are of order $n - 1$. Hence the induction hypothesis applies and gives $N_{jk} = -M_{jk}$. Thus $E = -D$ by (5).

the statement holds for $n = l+1 \, (=n)$

First, prove (c) and then (b)

(b) Add c times Row i to Row j. Let $\tilde{D}$ be the new determinant. Its entries in Row j are $a_{jk} + ca_{ik}$. If we expand $\tilde{D}$ by this Row j, we see that we can write it as $\tilde{D} = D_1 + cD_2$, where $D_1 = D$ has in Row j the a_{jk}, whereas D_2 has in that Row j the a_{ik} from the addition. Hence D_2 has a_{ik} in both Row i and Row j. Interchanging these two rows gives D_2 back, but on the other hand it gives $-D_2$ by (a). Together $D_2 = -D_2 = 0$, so that $\tilde{D} = D_1 = D$.

(c) Expand the determinant by the row that has been multiplied. ■

CAUTION! $\det(c\mathbf{A}) = c^n \det \mathbf{A}$ (not $c \det \mathbf{A}$). Explain why.

EXAMPLE 4 **Evaluation of Determinants by Reduction to Triangular Form**

Because of Theorem 1 we may evaluate determinants by reduction to triangular form, as in the Gauss elimination for a matrix. For instance (with the blue explanations always referring to the *preceding determinant*)

$$D = \begin{vmatrix} 2 & 0 & -4 & 6 \\ 4 & 5 & 1 & 0 \\ 0 & 2 & 6 & -1 \\ -3 & 8 & 9 & 1 \end{vmatrix}$$

$$D_1 + \det \begin{bmatrix} \vdots & & \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \\ ca_{i1} & \cdots & ca_{in} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

<u>Remarks</u>

$$\|$$

$$\tilde{D} = \det \begin{bmatrix} \vdots & & & \\ a_{i1} & \cdots & a_{ik} & \cdots & a_{in} \\ \vdots & & & \\ (a_{j1} + ca_{i1}) & \cdots & (a_{jk} + ca_{ik}) & \cdots & (a_{jn} + ca_{in}) \\ \vdots & & & \\ a_{m1} & \cdots & & \cdots & a_{mn} \end{bmatrix}$$

$$D_1 = \det \begin{bmatrix} \vdots & & \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 2 & 6 & -1 \\ 0 & 8 & 3 & 10 \end{vmatrix} \quad \begin{matrix} \\ \text{Row 2} - 2 \text{ Row 1} \\ \\ \text{Row 4} + 1.5 \text{ Row 1} \end{matrix}$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 0 & 2.4 & 3.8 \\ 0 & 0 & -11.4 & 29.2 \end{vmatrix} \quad \begin{matrix} \\ \\ \text{Row 3} - 0.4 \text{ Row 2} \\ \text{Row 4} - 1.6 \text{ Row 2} \end{matrix}$$

$$= \begin{vmatrix} 2 & 0 & -4 & 6 \\ 0 & 5 & 9 & -12 \\ 0 & 0 & 2.4 & 3.8 \\ 0 & 0 & -0 & 47.25 \end{vmatrix} \quad \begin{matrix} \\ \\ \\ \text{Row 4} + 4.75 \text{ Row 3} \end{matrix}$$

$$= 2 \cdot 5 \cdot 2.4 \cdot 47.25 = 1134.$$

THEOREM 2

Further Properties of *n*th-Order Determinants

(a)–(c) *in Theorem 1 hold also for columns.*

(d) ***Transposition*** *leaves the value of a determinant unaltered.*

(e) ***A zero row or column*** *renders the value of a determinant zero.*

(f) ***Proportional rows or columns*** *render the value of a determinant zero. In particular, a determinant with two identical rows or columns has the value zero.*

PROOF **(a)–(e)** follow directly from the fact that a determinant can be expanded by any row column. In **(d)**, transposition is defined as for matrices, that is, the jth row becomes the jth column of the transpose.

(f) If Row $j = c$ times Row i, then $D = cD_1$, where D_1 has Row j = Row i. Hence an interchange of these rows reproduces D_1, but it also gives $-D_1$ by Theorem 1(a). Hence $D_1 = 0$ and $D = cD_1 = 0$. Similarly for columns.

It is quite remarkable that the important concept of the rank of a matrix **A**, which is the maximum number of linearly independent row or column vectors of **A** (see Sec. 7.4), can be related to determinants. Here we may assume that rank $\mathbf{A} > 0$ because the only matrices with rank 0 are the zero matrices (see Sec. 7.4).

THEOREM 3

Rank in Terms of Determinants

*An $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ has rank $r \geqq 1$ if and only if **A** has an $r \times r$ submatrix with nonzero determinant, whereas every square submatrix with more than r rows that **A** has (or does not have!) has determinant equal to zero.*

*In particular, if **A** is square, $n \times n$, it has rank n if and only if*

$$\det \mathbf{A} \neq 0.$$

PROOF The key idea is that elementary row operations (Sec. 7.3) alter neither rank (by Theorem 1 in Sec. 7.4) nor the property of a determinant being nonzero (by Theorem 1 in this section). The echelon form $\hat{\mathbf{A}}$ of $\mathbf{A}$ (see Sec. 7.3) has r nonzero row vectors (which are the first r row vectors) if and only if rank $\mathbf{A} = r$. Let $\hat{\mathbf{R}}$ be the $r \times r$ submatrix in the left upper corner of $\hat{\mathbf{A}}$ (so that the entries of $\hat{\mathbf{R}}$ are in both the first r rows and r columns of $\hat{\mathbf{A}}$). Now $\hat{\mathbf{R}}$ is triangular, with all diagonal entries r_{jj} nonzero. Thus, det $\hat{\mathbf{R}} = r_{11} \cdots r_{rr} \neq 0$. Also det $\mathbf{R} \neq 0$ for the corresponding $r \times r$ submatrix $\mathbf{R}$ of $\mathbf{A}$ because $\hat{\mathbf{R}}$ results from $\mathbf{R}$ by elementary row operations. Similarly, det $\mathbf{S} = 0$ for any square submatrix $\mathbf{S}$ of $r + 1$ or more rows perhaps contained in $\mathbf{A}$ because the corresponding submatrix $\hat{\mathbf{S}}$ of $\hat{\mathbf{A}}$ must contain a row of zeros (otherwise we would have rank $\mathbf{A} \geqq r + 1$), so that det $\hat{\mathbf{S}} = 0$ by Theorem 2. This proves the theorem for an $m \times n$ matrix.

In particular, if $\mathbf{A}$ is square, $n \times n$, then rank $\mathbf{A} = n$ if and only if $\mathbf{A}$ contains an $n \times n$ submatrix with nonzero determinant. But the only such submatrix can be $\mathbf{A}$ itself, hence det $\mathbf{A} \neq 0$. ■

Cramer's Rule

Theorem 3 opens the way to the classical solution formula for linear systems known as **Cramer's rule**[2]**,** which gives solutions as quotients of determinants. ***Cramer's rule is not practical in computations*** (for which the methods in Secs. 7.3 and 20.1–20.3 are suitable), but is of ***theoretical interest*** in differential equations (Secs. 2.10, 3.3) and other theories that have engineering applications.

THEOREM 4 **Cramer's Theorem (Solution of Linear Systems by Determinants)**

(a) *If a linear system of n equations in the same number of unknowns* $x_1, \cdots, x_n$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{6}$$

has a nonzero coefficient determinant $D = \det \mathbf{A}$, *the system has precisely one solution. This solution is given by the formulas*

$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \quad \cdots, \quad x_n = \frac{D_n}{D} \qquad \textbf{(Cramer's rule)} \tag{7}$$

where D_k is the determinant obtained from D by replacing in D the kth column by the column with the entries $b_1, \cdots, b_n$.

(b) *Hence if the system* (6) *is* ***homogeneous*** *and* $D \neq 0$, *it has only the trivial solution* $x_1 = 0$, $x_2 = 0, \cdots, x_n = 0$. *If* $D = 0$, *the homogeneous system also has nontrivial solutions.*

---

[2] GABRIEL CRAMER (1704–1752), Swiss mathematician.

HW

Prob. 7.7. 13
Prob. 7.7. 20

21. $\begin{bmatrix} 8 & 4 \\ -2 & -1 \\ 6 & 3 \end{bmatrix}$

22. $\begin{bmatrix} 2 & 1 & 0 \\ 13 & -13 & 12 \\ -3 & 5 & -4 \end{bmatrix}$

23. $\begin{bmatrix} 0.4 & 0 & -2.4 & 3.0 \\ 1.2 & 0.6 & 0 & 0.3 \\ 0 & 1.2 & 1.2 & 0 \end{bmatrix}$

24. TEAM PROJECT. Geometrical Applications: Curves and Surfaces Through Given Points. The idea is to get an equation from the vanishing of the determinant of a homogeneous linear system as the condition for a nontrivial solution in Cramer's theorem. We explain the trick for obtaining such a system for the case of a line L through two given points P_1: (x_1, y_1) and P_2: (x_2, y_2). The unknown line is $ax + by = -c$, say. We write it as $ax + by + c \cdot 1 = 0$. To get a nontrivial solution a, b, c, the determinant of the "coefficients" x, y, 1 must be zero. The system is

$$\begin{aligned} &ax + by + c \cdot 1 = 0 \quad (\text{Line } L) \\ &ax_1 + by_1 + c \cdot 1 = 0 \quad (P_1 \text{ on } L) \\ &ax_2 + by_2 + c \cdot 1 = 0 \quad (P_2 \text{ on } L). \end{aligned} \tag{12}$$

(a) Line through two points. Derive from $D = 0$ in (12) the familiar formula

$$\frac{x - x_1}{x_1 - x_2} = \frac{y - y_1}{y_1 - y_2}.$$

(b) Plane. Find the analog of (12) for a plane through three given points. Apply it when the points are (1, 1, 1), (3, 2, 6), (5, 0, 5).

(c) Circle. Find a similar formula for a circle in the plane through three given points. Find and sketch the circle through (2, 6), (6, 4), (7, 1).

(d) Sphere. Find the analog of the formula in (c) for a sphere through four given points. Find the sphere through (0, 0, 5), (4, 0, 1), (0, 4, 1), (0, 0, −3) by this formula or by inspection.

(e) General conic section. Find a formula for a general conic section (the vanishing of a determinant of 6th order). Try it out for a quadratic parabola and for a more general conic section of your own choice.

25. WRITING PROJECT. General Properties of Determinants. Illustrate each statement in Theorems 1 and 2 with an example of your choice.

26. CAS EXPERIMENT. Determinant of Zeros and Ones. Find the value of the determinant of the $n \times n$ matrix $\mathbf{A}_n$ with main diagonal entries all 0 and all others 1. Try to find a formula for this. Try to prove it by induction. Interpret $\mathbf{A}_3$ and $\mathbf{A}_4$ as *"incidence matrices"* (as in Problem Set 7.1 but without the minuses) of a triangle and a tetrahedron, respectively; similarly for an *"n-simplex"*, having n vertices and $n(n-1)/2$ edges (and spanning R^{n-1}, $n = 5, 6, \cdots$).

7.8 Inverse of a Matrix. Gauss–Jordan Elimination

In this section we consider square matrices exclusively.

The **inverse** of an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is denoted by $\mathbf{A}^{-1}$ and is an $n \times n$ matrix such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{1}$$

where $\mathbf{I}$ is the $n \times n$ unit matrix (see Sec. 7.2).

If $\mathbf{A}$ has an inverse, then $\mathbf{A}$ is called a **nonsingular matrix.** If $\mathbf{A}$ has no inverse, then $\mathbf{A}$ is called a **singular matrix.**

If $\mathbf{A}$ *has an inverse, the inverse is unique.*

Indeed, if both $\mathbf{B}$ and $\mathbf{C}$ are inverses of $\mathbf{A}$, then $\mathbf{AB} = \mathbf{I}$ and $\mathbf{CA} = \mathbf{I}$, so that we obtain the uniqueness from

$$\mathbf{B} = \mathbf{IB} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{CI} = \mathbf{C}.$$

We prove next that **A** has an inverse (is nonsingular) if and only if it has maximum possible rank n. The proof will also show that $\mathbf{Ax} = \mathbf{b}$ implies $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ provided $\mathbf{A}^{-1}$ exists, and will thus give a motivation for the inverse as well as a relation to linear systems. (But this will ***not*** give a good method of solving $\mathbf{Ax} = \mathbf{b}$ ***numerically*** because the Gauss elimination in Sec. 7.3 requires fewer computations.)

THEOREM 1 **Existence of the Inverse**

The inverse $\mathbf{A}^{-1}$ *of an* $n \times n$ *matrix* **A** *exists if and only if* rank $\mathbf{A} = n$, *thus* (by Theorem 3, Sec. 7.7) *if and only if* $\det \mathbf{A} \neq 0$. *Hence* **A** *is nonsingular if* rank $\mathbf{A} = n$, *and is singular if* rank $\mathbf{A} < n$.

PROOF Let **A** be a given $n \times n$ matrix and consider the linear system

$$\mathbf{Ax} = \mathbf{b}. \tag{2}$$

($\Rightarrow$) If the inverse $\mathbf{A}^{-1}$ exists, then multiplication from the left on both sides and use of (1) gives

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

(Let $\bar{x}, \tilde{x}$ satisfy (2). Then, $A(\bar{x}-\tilde{x}) = 0$ $\Rightarrow \bar{x}-\tilde{x} = A^{-1}0 = 0$: unique

? This shows that (2) has a unique solution **x**. Hence **A** must have rank n by the Fundamental Theorem in Sec. 7.5.

($\Leftarrow$) Conversely, let rank $\mathbf{A} = n$. Then by the same theorem, the system (2) has a unique solution **x** for any **b**. Now the back substitution following the Gauss elimination (Sec. 7.3) shows that the components x_j of **x** are linear combinations of those of **b**. Hence we can write

$\begin{bmatrix} x & \cdots & x \\ & \ddots & \vdots \\ 0 & & x \end{bmatrix}$ $x = E_i E_j \cdots b$

$$\mathbf{x} = \mathbf{Bb} \tag{3}$$

with **B** to be determined. Substitution into (2) gives

$$\mathbf{Ax} = \mathbf{A}(\mathbf{Bb}) = (\mathbf{AB})\mathbf{b} = \mathbf{Cb} = \mathbf{b} \qquad (\mathbf{C} = \mathbf{AB})$$

for any **b**. Hence $\mathbf{C} = \mathbf{AB} = \mathbf{I}$, the unit matrix. Similarly, if we substitute (2) into (3) we get

$$\mathbf{x} = \mathbf{Bb} = \mathbf{B}(\mathbf{Ax}) = (\mathbf{BA})\mathbf{x}$$

for any **x** (and $\mathbf{b} = \mathbf{Ax}$). Hence $\mathbf{BA} = \mathbf{I}$. Together, $\mathbf{B} = \mathbf{A}^{-1}$ exists. ■

[3]WILHELM JORDAN (1842–1899), German mathematician and geodesist. [See *American Mathematical Monthly* **94** (1987), 130–142.]

We do ***not recommend*** it as a method for solving systems of linear equations, since the number of operations in addition to those of the Gauss elimination is larger than that for back substitution, which the Gauss–Jordan elimination avoids. See also Sec. 20.1.

Remark

$Ax = 0$ for any x $\Leftrightarrow$ $A = 0$
where
$A \in R^{m\times n}$

Determination of the Inverse by the Gauss–Jordan Method

For the practical determination of the inverse $\mathbf{A}^{-1}$ of a nonsingular $n \times n$ matrix $\mathbf{A}$ we can use the Gauss elimination (Sec. 7.3), actually a variant of it, called the **Gauss-Jordan elimination**[3] (footnote of p. 316). The idea of the method is as follows.

Using $\mathbf{A}$, we form n linear systems

$$\mathbf{A}\mathbf{x}_{(1)} = \mathbf{e}_{(1)}, \qquad \cdots, \qquad \mathbf{A}\mathbf{x}_{(n)} = \mathbf{e}_{(n)}$$

where $\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(n)}$ are the columns of the $n \times n$ unit matrix $\mathbf{I}$; thus, $\mathbf{e}_{(1)} = [1 \;\; 0 \;\; \cdots \;\; 0]^\mathsf{T}$, $\mathbf{e}_{(2)} = [0 \;\; 1 \;\; 0 \;\; \cdots \;\; 0]^\mathsf{T}$, etc. These are n vector equations in the unknown vectors $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. We combine them into a single matrix equation $\mathbf{AX} = \mathbf{I}$, with the unknown matrix $\mathbf{X}$ having the columns $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. Correspondingly, we combine the n augmented matrices $[\mathbf{A} \;\; \mathbf{e}_{(1)}], \cdots, [\mathbf{A} \;\; \mathbf{e}_{(n)}]$ into one $n \times 2n$ "augmented matrix" $\widetilde{\mathbf{A}} = [\mathbf{A} \;\; \mathbf{I}]$. Now multiplication of $\mathbf{AX} = \mathbf{I}$ by $\mathbf{A}^{-1}$ from the left gives $\mathbf{X} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}$. Hence, to solve $\mathbf{AX} = \mathbf{I}$ for $\mathbf{X}$, we can apply the Gauss elimination to $\widetilde{\mathbf{A}} = [\mathbf{A} \;\; \mathbf{I}]$. This gives a matrix of the form $[\mathbf{U} \;\; \mathbf{H}]$ with upper triangular $\mathbf{U}$ because the Gauss elimination triangularizes systems. The Gauss–Jordan method reduces $\mathbf{U}$ by further elementary row operations to diagonal form, in fact to the unit matrix $\mathbf{I}$. This is done by eliminating the entries of $\mathbf{U}$ above the main diagonal and making the diagonal entries all 1 by multiplication (see the example below). Of course, the method operates on the entire matrix $[\mathbf{U} \;\; \mathbf{H}]$, transforming $\mathbf{H}$ into some matrix $\mathbf{K}$, hence the entire $[\mathbf{U} \;\; \mathbf{H}]$ to $[\mathbf{I} \;\; \mathbf{K}]$. This is the "augmented matrix" of $\mathbf{IX} = \mathbf{K}$. Now $\mathbf{IX} = \mathbf{X} = \mathbf{A}^{-1}$, as shown before. By comparison, $\mathbf{K} = \mathbf{A}^{-1}$, so that we can read $\mathbf{A}^{-1}$ directly from $[\mathbf{I} \;\; \mathbf{K}]$.

The following example illustrates the practical details of the method.

Solve $\left(AX = I \text{ for } X \right)$
where $X = [x_{(1)} \cdots x_{(n)}]$
$\Downarrow$
$[Ax_{(1)} \cdots Ax_{(n)}] = I$

$[I \;\; K]$
$\Downarrow$
$IX = K$
$\Downarrow$
$X = K$

EXAMPLE 1 **Inverse of a Matrix. Gauss–Jordan Elimination**

Determine the inverse $\mathbf{A}^{-1}$ of

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

Solution. We apply the Gauss elimination (Sec. 7.3) to the following $n \times 2n = 3 \times 6$ matrix, where BLUE always refers to the previous matrix.

$$[\mathbf{A} \;\; \mathbf{I}] = \left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 3 & -1 & 1 & 0 & 1 & 0 \\ -1 & 3 & 4 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 2 & 2 & -1 & 0 & 1 \end{array}\right] \quad \begin{array}{l} \\ \text{Row 2} + 3 \text{ Row 1} \\ \text{Row 3} - \text{Row 1} \end{array}$$

$$\left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 0 & -5 & -4 & -1 & 1 \end{array}\right] \quad \begin{array}{l} \\ \\ \text{Row 3} - \text{Row 2} \end{array}$$

$E_j [A \;\; I]$
$= E_j [A \; e_{(1)} \cdots e_{(n)}]$
$= [E_j A \;\; E_j e_{(1)} \cdots E_j e_{(n)}]$
$= [I \;\; K]$

Remark

$AX = I \Rightarrow XA = I$ (why?)
where
$A \in \mathbb{R}^{n \times n}$

This is $[\mathbf{U} \quad \mathbf{H}]$ as produced by the Gauss elimination. Now follow the additional Gauss–Jordan steps, reducing $\mathbf{U}$ to $\mathbf{I}$, that is, to diagonal form with entries 1 on the main diagonal.

$$\left[\begin{array}{ccc|ccc} 1 & -1 & -2 & -1 & 0 & 0 \\ 0 & 1 & 3.5 & 1.5 & 0.5 & 0 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} -\text{Row } 1 \\ 0.5 \text{ Row } 2 \\ -0.2 \text{ Row } 3 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 0.6 & 0.4 & -0.4 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row } 1 + 2 \text{ Row } 3 \\ \text{Row } 2 - 3.5 \text{ Row } 3 \\ \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -0.7 & 0.2 & 0.3 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row } 1 + \text{Row } 2 \\ \\ \end{array}$$

The last three columns constitute $\mathbf{A}^{-1}$. Check:

$$\begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix} \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Similarly, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. ■

Useful Formulas for Inverses

The explicit formula (4) in the following theorem is often useful in theoretical studies (as opposed to *computing* inverses). In fact, the special case $n = 2$ occurs quite frequently in geometrical and other applications.

THEOREM 2 Inverse of a Matrix

The inverse of a nonsingular $n \times n$ *matrix* $\mathbf{A} = [a_{jk}]$ *is given by*

$$(4) \qquad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} [C_{jk}]^{\mathsf{T}} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix},$$

where C_{jk} *is the cofactor of* a_{jk} *in* $\det \mathbf{A}$ (see Sec. 7.7). (CAUTION! Note well that in $\mathbf{A}^{-1}$, the cofactor C_{jk} occupies the same place as a_{kj} (not a_{jk}) does in $\mathbf{A}$.)

In particular, the inverse of

$$(4^*) \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \textit{is} \quad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

PROOF We denote the right side of (4) by $\mathbf{B}$ and show that $\mathbf{BA} = \mathbf{I}$. We first write

$$\mathbf{BA} = \mathbf{G} = [g_{kl}] \tag{5}$$

and then show that $\mathbf{G} = \mathbf{I}$. Now by the definition of matrix multiplication and because of the form of $\mathbf{B}$ in (4), we obtain (CAUTION! C_{sk}, not C_{ks})

$$g_{kl} = \sum_{s=1}^{n} \frac{C_{sk}}{\det \mathbf{A}} a_{sl} = \frac{1}{\det \mathbf{A}} (a_{1l}C_{1k} + \cdots + a_{nl}C_{nk}). \tag{6}$$

Now (9) and (10) in Sec. 7.7 show that the sum $(\cdots)$ on the right is $D = \det \mathbf{A}$ when $l = k$, and is zero when $l \neq k$. Hence

$$g_{kk} = \frac{1}{\det \mathbf{A}} \det \mathbf{A} = 1,$$

$$g_{kl} = 0 \quad (l \neq k),$$

In particular, for $n = 2$ we have in (4) in the first row $C_{11} = a_{22}$, $C_{21} = -a_{12}$ and in the second row $C_{12} = -a_{21}$, $C_{22} = a_{11}$. This gives (4*). ■

EXAMPLE 2 **Inverse of a 2 × 2 Matrix**

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{10}\begin{bmatrix} 4 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix}$$ ■

EXAMPLE 3 **Further Illustration of Theorem 2**

Using (4), find the inverse of

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

Solution. We obtain $\det \mathbf{A} = -1(-7) - 1 \cdot 13 + 2 \cdot 8 = 10$, and in (4),

$$C_{11} = \begin{vmatrix} -1 & 1 \\ 3 & 4 \end{vmatrix} = -7, \qquad C_{21} = -\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 2, \qquad C_{31} = \begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = 3,$$

$$C_{12} = -\begin{vmatrix} 3 & 1 \\ -1 & 4 \end{vmatrix} = -13, \qquad C_{22} = \begin{vmatrix} -1 & 2 \\ -1 & 4 \end{vmatrix} = -2, \qquad C_{32} = -\begin{vmatrix} -1 & 2 \\ 3 & 1 \end{vmatrix} = 7,$$

$$C_{13} = \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 8, \qquad C_{23} = -\begin{vmatrix} -1 & 1 \\ -1 & 3 \end{vmatrix} = 2, \qquad C_{33} = \begin{vmatrix} -1 & 1 \\ 3 & -1 \end{vmatrix} = -2,$$

so that by (4), in agreement with Example 1,

$$\mathbf{A}^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}.$$ ■

Remarks

In particular, if we choose as new numbers the entries $a_{1l}, \cdots, a_{nl}$ of the lth column of D (where $l \neq k$), we have a new determinant $\hat{D}$ which has twice the column $[a_{1l} \quad \cdots \quad a_{nl}]^{\mathsf{T}}$, once as its lth column, and once as its kth because of the replacement. Hence $\hat{D} = 0$ by Theorem 2(f). If we now expand $\hat{D}$ by the column that has been replaced (the kth column), we thus obtain

$$a_{1l}C_{1k} + a_{2l}C_{2k} + \cdots + a_{nl}C_{nk} = 0 \qquad (l \neq k). \tag{10}$$

Diagonal matrices $\mathbf{A} = [a_{jk}]$, $a_{jk} = 0$ when $j \neq k$, have an inverse if and only if all $a_{jj} \neq 0$. Then $\mathbf{A}^{-1}$ is diagonal, too, with entries $1/a_{11}, \cdots, 1/a_{nn}$.

PROOF For a diagonal matrix we have in (4)

$$\frac{C_{11}}{D} = \frac{a_{22} \cdots a_{nn}}{a_{11}a_{22} \cdots a_{nn}} = \frac{1}{a_{11}}, \qquad \text{etc.} \qquad \blacksquare$$

EXAMPLE 4 Inverse of a Diagonal Matrix

Let

$$\mathbf{A} = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

Then the inverse is

$$\mathbf{A}^{-1} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \blacksquare$$

Products can be inverted by taking the inverse of each factor and multiplying these inverses *in reverse order,*

$$(\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}. \tag{7}$$

Hence for more than two factors,

$$(\mathbf{AC} \cdots \mathbf{PQ})^{-1} = \mathbf{Q}^{-1}\mathbf{P}^{-1} \cdots \mathbf{C}^{-1}\mathbf{A}^{-1}. \tag{8}$$

PROOF The idea is to start from (1) for $\mathbf{AC}$ instead of $\mathbf{A}$, that is, $\mathbf{AC}(\mathbf{AC})^{-1} = \mathbf{I}$, and multiply it on both sides from the left, first by $\mathbf{A}^{-1}$, which because of $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ gives

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{AC}(\mathbf{AC})^{-1} &= \mathbf{C}(\mathbf{AC})^{-1} \\ &= \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}, \end{aligned}$$

and then multiplying this on both sides from the left, this time by $\mathbf{C}^{-1}$ and by using $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$,

$$\mathbf{C}^{-1}\mathbf{C}(\mathbf{AC})^{-1} = (\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}.$$

This proves (7), and from it, (8) follows by induction. $\blacksquare$

We also note that *the inverse of the inverse is the given matrix,* as you may prove,

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}. \tag{9}$$

Unusual Properties of Matrix Multiplication. Cancellation Laws

Section 7.2 contains warnings that some properties of matrix multiplication deviate from those for numbers, and we are now able to explain the restricted validity of the so-called **cancellation laws** [2.] and [3.] below, using rank and inverse, concepts that were not yet available in Sec. 7.2. The deviations from the usual are of great practical importance and must be carefully observed. They are as follows.

[1.] Matrix multiplication is not commutative, that is, in general we have

$$\mathbf{AB} \neq \mathbf{BA}.$$

[2.] $\mathbf{AB} = \mathbf{0}$ does not generally imply $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$ (or $\mathbf{BA} = \mathbf{0}$); for example,

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

[3.] $\mathbf{AC} = \mathbf{AD}$ does not generally imply $\mathbf{C} = \mathbf{D}$ (even when $\mathbf{A} \neq \mathbf{0}$).

Complete answers to [2.] and [3.] are contained in the following theorem.

THEOREM 3

Cancellation Laws

Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *be* $n \times n$ *matrices. Then:*

(a) *If* rank $\mathbf{A} = n$ *and* $\mathbf{AB} = \mathbf{AC}$, *then* $\mathbf{B} = \mathbf{C}$.

(b) *If* rank $\mathbf{A} = n$, *then* $\mathbf{AB} = \mathbf{0}$ *implies* $\mathbf{B} = \mathbf{0}$. *Hence if* $\mathbf{AB} = \mathbf{0}$, *but* $\mathbf{A} \neq \mathbf{0}$ *as well as* $\mathbf{B} \neq \mathbf{0}$, *then* rank $\mathbf{A} < n$ *and* rank $\mathbf{B} < n$.

(c) *If* $\mathbf{A}$ *is singular, so are* $\mathbf{BA}$ *and* $\mathbf{AB}$.

PROOF **(a)** The inverse of $\mathbf{A}$ exists by Theorem 1. Multiplication by $\mathbf{A}^{-1}$ from the left gives $\mathbf{A}^{-1}\mathbf{AB} = \mathbf{A}^{-1}\mathbf{AC}$, hence $\mathbf{B} = \mathbf{C}$.

(b) Let rank $\mathbf{A} = n$. Then $\mathbf{A}^{-1}$ exists, and $\mathbf{AB} = \mathbf{0}$ implies $\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B} = \mathbf{0}$. Similarly when rank $\mathbf{B} = n$. This implies the second statement in (b).

($\mathbf{c}_1$) Rank $\mathbf{A} < n$ by Theorem 1. Hence $\mathbf{Ax} = \mathbf{0}$ has nontrivial solutions by Theorem 2 in Sec. 7.5. Multiplication by $\mathbf{B}$ shows that these solutions are also solutions of $\mathbf{BAx} = \mathbf{0}$, so that rank $(\mathbf{BA}) < n$ by Theorem 2 in Sec. 7.5 and $\mathbf{BA}$ is singular by Theorem 1.

($\mathbf{c}_2$) $\mathbf{A}^\mathsf{T}$ is singular by Theorem 2(d) in Sec. 7.7. Hence $\mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$ is singular by part (c_1), and is equal to $(\mathbf{AB})^\mathsf{T}$ by (10d) in Sec. 7.2. Hence $\mathbf{AB}$ is singular by Theorem 2(d) in Sec. 7.7. ■

Determinants of Matrix Products

The determinant of a matrix product $\mathbf{AB}$ or $\mathbf{BA}$ can be written as the product of the determinants of the factors, and it is interesting that det $\mathbf{AB}$ = det $\mathbf{BA}$, although $\mathbf{AB} \neq \mathbf{BA}$ in general. The corresponding formula (10) is needed occasionally and can be obtained by Gauss-Jordan elimination (see Example 1) and from the theorem just proved.

THEOREM 4

Determinant of a Product of Matrices

For any $n \times n$ *matrices* $\mathbf{A}$ *and* $\mathbf{B}$,

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det \mathbf{A} \det \mathbf{B}. \tag{10}$$

PROOF If $\mathbf{A}$ or $\mathbf{B}$ is singular, so are $\mathbf{AB}$ and $\mathbf{BA}$ by Theorem 3(c), and (10) reduces to $0 = 0$ by Theorem 3 in Sec. 7.7.

Now let $\mathbf{A}$ and $\mathbf{B}$ be nonsingular. Then we can reduce $\mathbf{A}$ to a diagonal matrix $\hat{\mathbf{A}} = [a_{jk}]$ by Gauss–Jordan steps. Under these operations, det $\mathbf{A}$ retains its value, by Theorem 1 in Sec. 7.7, (a) and (b) [not (c)] except perhaps for a sign reversal in row interchanging when pivoting. But the same operations reduce $\mathbf{AB}$ to $\hat{\mathbf{A}}\mathbf{B}$ with the same effect on det ($\mathbf{AB}$). Hence it remains to prove (10) for $\hat{\mathbf{A}}\mathbf{B}$; written out,

$$\hat{\mathbf{A}}\mathbf{B} = \begin{bmatrix} \hat{a}_{11} & 0 & \cdots & 0 \\ 0 & \hat{a}_{22} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \hat{a}_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ & & \vdots & \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \hat{a}_{11}b_{11} & \hat{a}_{11}b_{12} & \cdots & \hat{a}_{11}b_{1n} \\ \hat{a}_{22}b_{21} & \hat{a}_{22}b_{22} & \cdots & \hat{a}_{22}b_{2n} \\ & & \vdots & \\ \hat{a}_{nn}b_{n1} & \hat{a}_{nn}b_{n2} & \cdots & \hat{a}_{nn}b_{nn} \end{bmatrix}.$$

We now take the determinant det ($\hat{\mathbf{A}}\mathbf{B}$). On the right we can take out a factor $\hat{a}_{11}$ from the first row, $\hat{a}_{22}$ from the second, $\cdots$, $\hat{a}_{nn}$ from the nth. But this product $\hat{a}_{11}\hat{a}_{22}\cdots\hat{a}_{nn}$ equals det $\hat{\mathbf{A}}$ because $\hat{\mathbf{A}}$ is diagonal. The remaining determinant is det $\mathbf{B}$. This proves (10) for det ($\mathbf{AB}$), and the proof for det ($\mathbf{BA}$) follows by the same idea. ■

This completes our discussion of linear systems (Secs. 7.3–7.8). Section 7.9 on vector spaces and linear transformations is optional. ***Numeric methods*** are discussed in Secs. 20.1–20.4, which are independent of other sections on numerics.

PROBLEM SET 7.8

1–12 **INVERSE**

Find the inverse by Gauss–Jordan [or by (4*) if $n = 2$] or state that it does not exist. Check by using (1).

1. $\begin{bmatrix} 1.20 & 4.64 \\ 0.50 & 3.60 \end{bmatrix}$

2. $\begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}$

3. $\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{bmatrix}$

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

HW

Prob. 7.8.6

5. $\begin{bmatrix} 3 & -1 & 1 \\ -15 & 6 & -5 \\ 5 & -2 & 2 \end{bmatrix}$

6. $\begin{bmatrix} 29 & -11 & 10 \\ -160 & 61 & -55 \\ 55 & -21 & 19 \end{bmatrix}$

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

8. $\begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & 2 \\ 2 & 4 & 11 \end{bmatrix}$

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

10. $\begin{bmatrix} 0 & 8 & 0 \\ 0 & 0 & 4 \\ 2 & 0 & 0 \end{bmatrix}$

11. $\begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & 2 \\ 2 & 4 & 10 \end{bmatrix}$

12. $\begin{bmatrix} 1 & 2 & -9 \\ -2 & -4 & 19 \\ 0 & -1 & 2 \end{bmatrix}$

13. **(Triangular matrix)** Is the inverse of a triangular matrix always triangular (as in Prob. 7)? Give reason.

14. **(Rotation)** Give an application of the matrix in Prob. 3 that makes the form of its inverse obvious.

15. **(Inverse of the square)** Verify $(\mathbf{A}^2)^{-1} = (\mathbf{A}^{-1})^2$ for $\mathbf{A}$ in Prob. 5.

16. Prove the formula in Prob. 15.

17. **(Inverse of the transpose)** Verify $(\mathbf{A}^{\mathsf{T}})^{-1} = (\mathbf{A}^{-1})^{\mathsf{T}}$ for $\mathbf{A}$ in Prob. 5.

18. Prove the formula in Prob. 17.

19. **(Inverse of the inverse)** Prove that $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

20. **(Row interchange)** Same question as in Prob. 14 for the matrix in Prob. 9.

21–23 EXPLICIT FORMULA (4) FOR THE INVERSE

Formula (4) is generally not very practical. To understand its use, apply it:

21. To Prob. 9. 22. To Prob. 4. 23. To Prob. 7.

7.9 Vector Spaces, Inner Product Spaces, Linear Transformations *Optional*

In Sec. 7.4 we have seen that ***special vector spaces*** arise quite naturally in connection with matrices and linear systems, that their elements, called *vectors,* satisfy rules quite similar to those for numbers [(3) and (4) in Sec. 7.1], and that they are often obtained as *spans* (sets of linear combinations) of finitely many given vectors. Each such vector has n real numbers as its *components.* Look this up before going on.

Now if we take *all* vectors with n real numbers as components (*"real vectors"*), we obtain the very important **real n-dimensional vector space R^n**. This is a standard name and notation. Thus, each vector in R^n is an ordered n-tuple of real numbers.

Particular cases are R^2, the space of all ordered pairs (**"vectors in the plane"**) and R^3, the space of all ordered triples (**"vectors in 3-space"**). These vectors have wide applications in mechanics, geometry, and calculus that are basic to the engineer and physicist.

Similarly, if we take all ordered n-tuples of *complex numbers* as vectors and complex numbers as scalars, we obtain the **complex vector space C^n**, which we shall consider in Sec. 8.5.

This is not all. There are other sets of practical interest (sets of matrices, functions, transformations, etc.) for which addition and scalar multiplication can be defined in a natural way so that they form a "vector space". This suggests to create from the ***"concrete model"*** R^n the ***"abstract concept"*** of a ***"real vector space"*** V by taking the basic properties (3) and (4) in Sec. 7.1 as axioms. These axioms guarantee that one obtains a useful and applicable theory of those more general situations. Note that each axiom expresses a simple property of R^n or, as a matter of fact, of R^3. Selecting good axioms needs experience and is a process of trial and error that often extends over a long period of time.

DEFINITION

Real Vector Space

A nonempty set V of elements $\mathbf{a}, \mathbf{b}, \cdots$ is called a **real vector space** (or *real linear space*), and these elements are called **vectors** (regardless of their nature, which will come out from the context or will be left arbitrary) if in V there are defined two algebraic operations (called *vector addition* and *scalar multiplication*) as follows.

I. Vector addition associates with every pair of vectors $\mathbf{a}$ and $\mathbf{b}$ of V a unique vector of V, called the *sum* of $\mathbf{a}$ and $\mathbf{b}$ and denoted by $\mathbf{a} + \mathbf{b}$, such that the following axioms are satisfied.

I.1 *Commutativity.* For any two vectors $\mathbf{a}$ and $\mathbf{b}$ of V,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

I.2 *Associativity.* For any three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ of V,

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) \qquad (\text{written } \mathbf{u} + \mathbf{v} + \mathbf{w}).$$

I.3 There is a unique vector in V, called the *zero vector* and denoted by $\mathbf{0}$, such that for every $\mathbf{a}$ in V,

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

I.4 For every $\mathbf{a}$ in V there is a unique vector in V that is denoted by $-\mathbf{a}$ and is such that

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

II. Scalar multiplication. The real numbers are called **scalars.** Scalar multiplication associates with every $\mathbf{a}$ in V and every scalar c a unique vector of V, called the *product* of c and $\mathbf{a}$ and denoted by $c\mathbf{a}$ (or $\mathbf{a}c$) such that the following axioms are satisfied.

II.1 *Distributivity.* For every scalar c and vectors $\mathbf{a}$ and $\mathbf{b}$ in V,

$$c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b}.$$

II.2 *Distributivity.* For all scalars c and k and every $\mathbf{a}$ in V,

$$(c + k)\mathbf{a} = c\mathbf{a} + k\mathbf{a}.$$

II.3 *Associativity.* For all scalars c and k and every $\mathbf{a}$ in V,

$$c(k\mathbf{a}) = (ck)\mathbf{a} \qquad (\text{written } ck\mathbf{a}).$$

II.4 For every $\mathbf{a}$ in V,

$$1\mathbf{a} = \mathbf{a}.$$

A **complex vector space** is obtained if, instead of real numbers, we take complex numbers as scalars.

Basic concepts related to the concept of a vector space are defined as in Sec. 7.4.

A **linear combination** of vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ in a vector space V is an expression

$$c_1\mathbf{a}_{(1)} + \cdots + c_m\mathbf{a}_{(m)} \qquad (c_1, \cdots, c_m \text{ any scalars}).$$

These vectors form a **linearly independent set** (briefly, they are called **linearly independent**) if

$$c_1\mathbf{a}_{(1)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0} \tag{1}$$

implies that $c_1 = 0, \cdots, c_m = 0$. Otherwise, if (1) also holds with scalars not all zero, the vectors are called **linearly dependent.**

Note that (1) with $m = 1$ is $c\mathbf{a} = \mathbf{0}$ and shows that a single vector $\mathbf{a}$ is linearly independent if and only if $\mathbf{a} \neq \mathbf{0}$.

V has **dimension n,** or is **n-dimensional,** if it contains a linearly independent set of n vectors, whereas any set of more than n vectors in V is linearly dependent. That set of n linearly independent vectors is called a **basis** for V. Then every vector in V can be written as a linear combination of the basis vectors; for a given basis, this representation is unique (see Prob. 14).

EXAMPLE 1 Vector Space of Matrices

The real 2×2 matrices form a four-dimensional real vector space. A basis is

$$\mathbf{B}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{B}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

because any 2×2 matrix $\mathbf{A} = [a_{jk}]$ has a unique representation $\mathbf{A} = a_{11}\mathbf{B}_{11} + a_{12}\mathbf{B}_{12} + a_{21}\mathbf{B}_{21} + a_{22}\mathbf{B}_{22}$. Similarly, the real $m \times n$ matrices with fixed m and n form an mn-dimensional vector space. What is the dimension of the vector space of all 3×3 skew-symmetric matrices? Can you find a basis? ■

EXAMPLE 2 Vector Space of Polynomials

The set of all constant, linear, and quadratic polynomials in x together is a vector space of dimension 3 with basis $\{1, x, x^2\}$ under the usual addition and multiplication by real numbers because these two operations give polynomials not exceeding degree 2. What is the dimension of the vector space of all polynomials of degree not exceeding a given fixed n? Can you find a basis? ■

If a vector space V contains a linearly independent set of n vectors for every n, no matter how large, then V is called **infinite dimensional,** as opposed to a *finite dimensional* (n-dimensional) vector space just defined. An example of an infinite dimensional vector space is the space of all continuous functions on some interval $[a, b]$ of the x-axis, as we mention without proof.

Inner Product Spaces

If $\mathbf{a}$ and $\mathbf{b}$ are vectors in R^n, regarded as column vectors, we can form the product $\mathbf{a}^\mathsf{T}\mathbf{b}$. This is a 1×1 matrix, which we can identify with its single entry, that is, with a number. This product is called the **inner product** or **dot product of a and b.** Other notations for it are $(\mathbf{a}, \mathbf{b})$ and $\mathbf{a} \cdot \mathbf{b}$. Thus

$$\mathbf{a}^\mathsf{T}\mathbf{b} = (\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b} = [a_1 \cdots a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{l=1}^{n} a_l b_l = a_1 b_1 + \cdots + a_n b_n.$$

We now extend this concept to general real vector spaces by taking basic properties of $(\mathbf{a}, \mathbf{b})$ as axioms for an "abstract inner product" $(\mathbf{a}, \mathbf{b})$ as follows.

DEFINITION

Real Inner Product Space

A real vector space V is called a **real inner product space** (or *real pre-Hilbert*[4] *space*) if it has the following property. With every pair of vectors $\mathbf{a}$ and $\mathbf{b}$ in V there is associated a real number, which is denoted by $(\mathbf{a}, \mathbf{b})$ and is called the **inner product** of $\mathbf{a}$ and $\mathbf{b}$, such that the following axioms are satisfied.

I. For all scalars q_1 and q_2 and all vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in V,

$$(q_1\mathbf{a} + q_2\mathbf{b}, \mathbf{c}) = q_1(\mathbf{a}, \mathbf{c}) + q_2(\mathbf{b}, \mathbf{c}) \qquad \textit{(Linearity)}.$$

II. For all vectors $\mathbf{a}$ and $\mathbf{b}$ in V,

$$(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a}) \qquad \textit{(Symmetry)}.$$

III. For every $\mathbf{a}$ in V,

$$\left.\begin{aligned} &(\mathbf{a}, \mathbf{a}) \geqq 0, \\ (\mathbf{a}, \mathbf{a}) = 0 \quad &\text{if and only if} \quad \mathbf{a} = \mathbf{0} \end{aligned}\right\} \qquad \textit{(Positive-definiteness)}.$$

Vectors whose inner product is zero are called **orthogonal.**

The *length* or **norm** of a vector in V is defined by

$$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} \qquad (\geqq 0). \tag{2}$$

A vector of norm 1 is called a **unit vector.**

From these axioms and from (2) one can derive the basic inequality

$$|(\mathbf{a}, \mathbf{b})| \leqq \|\mathbf{a}\|\,\|\mathbf{b}\| \qquad \textit{(Cauchy–Schwarz}^5 \textit{ inequality)}. \tag{3}$$

From this follows

$$\|\mathbf{a} + \mathbf{b}\| \leqq \|\mathbf{a}\| + \|\mathbf{b}\| \qquad \textit{(Triangle inequality)}. \tag{4}$$

A simple direct calculation gives

$$\|\mathbf{a} + \mathbf{b}\|^2 + \|\mathbf{a} - \mathbf{b}\|^2 = 2(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2) \qquad \textit{(Parallelogram equality)}. \tag{5}$$

[4]DAVID HILBERT (1862–1943), great German mathematician, taught at Königsberg and Göttingen and was the creator of the famous Göttingen mathematical school. He is known for his basic work in algebra, the calculus of variations, integral equations, functional analysis, and mathematical logic. His "Foundations of Geometry" helped the axiomatic method to gain general recognition. His famous 23 problems (presented in 1900 at the International Congress of Mathematicians in Paris) considerably influenced the development of modern mathematics.

If V is finite dimensional, it is actually a so-called *Hilbert space;* see Ref. [GR7], p. 73, listed in App. I.

[5]HERMANN AMANDUS SCHWARZ (1843–1921). German mathematician, known by his work in complex analysis (conformal mapping) and differential geometry. For Cauchy see Sec. 2.5.

EXAMPLE 3 n-Dimensional Euclidean Space

R^n with the inner product

$$(\mathbf{a}, \mathbf{b}) = \mathbf{a}^\mathsf{T}\mathbf{b} = a_1b_1 + \cdots + a_nb_n \tag{6}$$

(where both **a** and **b** are *column* vectors) is called the **n-dimensional Euclidean space** and is denoted by E^n or again simply by R^n. Axioms I–III hold, as direct calculation shows. Equation (2) gives the **"Euclidean norm"**

$$\|\mathbf{a}\| = \sqrt{(\mathbf{a}, \mathbf{a})} = \sqrt{\mathbf{a}^\mathsf{T}\mathbf{a}} = \sqrt{a_1^{\,2} + \cdots + a_n^{\,2}}. \tag{7}$$

■

EXAMPLE 4 An Inner Product for Functions. Function Space

The set of all real-valued continuous functions $f(x)$, $g(x)$, $\cdots$ on a given interval $\alpha \leqq x \leqq \beta$ is a real vector space under the usual addition of functions and multiplication by scalars (real numbers). On this **"function space"** we can define an inner product by the integral

$$(f, g) = \int_\alpha^\beta f(x)\, g(x)\, dx. \tag{8}$$

Axioms I–III can be verified by direct calculation. Equation (2) gives the norm

$$\|f\| = \sqrt{(f, f)} = \sqrt{\int_\alpha^\beta f(x)^2\, dx}. \tag{9}$$

■

Our examples give a first impression of the great generality of the abstract concepts of vector spaces and inner product spaces. Further details belong to more advanced courses (on functional analysis, meaning abstract modern analysis; see Ref. [GR7] listed in App. 1) and cannot be discussed here. Instead we now take up a related topic where matrices play a central role.

Linear Transformations

Let X and Y be any vector spaces. To each vector $\mathbf{x}$ in X we assign a unique vector $\mathbf{y}$ in Y. Then we say that a **mapping** (or **transformation** or **operator**) of X into Y is given. Such a mapping is denoted by a capital letter, say F. The vector $\mathbf{y}$ in Y assigned to a vector $\mathbf{x}$ in X is called the **image** of $\mathbf{x}$ under F and is denoted by $F(\mathbf{x})$ [or $F\mathbf{x}$, without parentheses].

F is called a **linear mapping** or **linear transformation** if for all vectors $\mathbf{v}$ and $\mathbf{x}$ in X and scalars c,

$$\begin{aligned} F(\mathbf{v} + \mathbf{x}) &= F(\mathbf{v}) + F(\mathbf{x}) \\ F(c\mathbf{x}) &= cF(\mathbf{x}). \end{aligned} \tag{10}$$

Linear Transformation of Space R^n into Space R^m

From now on we let $X = R^n$ and $Y = R^m$. Then any real $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ gives a transformation of R^n into R^m,

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{11}$$

Since $\mathbf{A}(\mathbf{u} + \mathbf{x}) = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{x}$ and $\mathbf{A}(c\mathbf{x}) = c\mathbf{A}\mathbf{x}$, this transformation is linear.

We show that, conversely, every linear transformation F of R^n into R^m can be given in terms of an $m \times n$ matrix $\mathbf{A}$, after a basis for R^n and a basis for R^m have been chosen. This can be proved as follows.

Let $\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(n)}$ be any basis for R^n. Then every $\mathbf{x}$ in R^n has a unique representation

$$\mathbf{x} = x_1\mathbf{e}_{(1)} + \cdots + x_n\mathbf{e}_{(n)}.$$

Since F is linear, this representation implies for the image $F(\mathbf{x})$:

$$F(\mathbf{x}) = F(x_1\mathbf{e}_{(1)} + \cdots + x_n\mathbf{e}_{(n)}) = x_1F(\mathbf{e}_{(1)}) + \cdots + x_nF(\mathbf{e}_{(n)}).$$

Hence F is uniquely determined by the images of the vectors of a basis for R^n. We now choose for R^n the **"standard basis"**

$$\text{(12)} \qquad \mathbf{e}_{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \mathbf{e}_{(n)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

where $\mathbf{e}_{(j)}$ has its jth component equal to 1 and all others 0. We show that we can now determine an $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ such that for every $\mathbf{x}$ in R^n and image $\mathbf{y} = F(\mathbf{x})$ in R^m,

$$\mathbf{y} = F(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

Indeed, from the image $\mathbf{y}^{(1)} = F(\mathbf{e}_{(1)})$ of $\mathbf{e}_{(1)}$ we get the condition

$$\mathbf{y}^{(1)} = \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_m^{(1)} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

from which we can determine the first column of $\mathbf{A}$, namely $a_{11} = y_1^{(1)}$, $a_{21} = y_2^{(1)}, \cdots,$ $a_{m1} = y_m^{(1)}$. Similarly, from the image of $\mathbf{e}_{(2)}$ we get the second column of $\mathbf{A}$, and so on. This completes the proof. ■

We say that $\mathbf{A}$ **represents** F, or *is a representation of* F, with respect to the bases for R^n and R^m. Quite generally, the purpose of a **"representation"** is the replacement of one object of study by another object whose properties are more readily apparent.

In three-dimensional Euclidean space E^3 the standard basis is usually written $\mathbf{e}_{(1)} = \mathbf{i}$, $\mathbf{e}_{(2)} = \mathbf{j}$, $\mathbf{e}_{(3)} = \mathbf{k}$. Thus,

$$\text{(13)} \qquad \mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

These are the three unit vectors in the positive directions of the axes of the **Cartesian coordinate system in space,** that is, the usual coordinate system with the same scale of measurement on the three mutually perpendicular coordinate axes.

EXAMPLE 5 Linear Transformations

Interpreted as transformations of Cartesian coordinates in the plane, the matrices

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

represent a reflection in the line $x_2 = x_1$, a reflection in the x_1-axis, a reflection in the origin, and a stretch (when $a > 1$, or a contraction when $0 < a < 1$) in the x_1-direction, respectively. ■

EXAMPLE 6 Linear Transformations

Our discussion preceding Example 5 is simpler than it may look at first sight. To see this, find **A** representing the linear transformation that maps (x_1, x_2) onto $(2x_1 - 5x_2, 3x_1 + 4x_2)$.

Solution. Obviously, the transformation is

$$y_1 = 2x_1 - 5x_2$$
$$y_2 = 3x_1 + 4x_2.$$

From this we can directly see that the matrix is

$$\mathbf{A} = \begin{bmatrix} 2 & -5 \\ 3 & 4 \end{bmatrix}. \qquad \text{Check:} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & -5 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 - 5x_2 \\ 3x_1 + 4x_2 \end{bmatrix}.$$ ■

If **A** in (11) is square, $n \times n$, then (11) maps R^n into R^n. If this **A** is nonsingular, so that $\mathbf{A}^{-1}$ exists (see Sec. 7.8), then multiplication of (11) by $\mathbf{A}^{-1}$ from the left and use of $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ gives the **inverse transformation**

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}. \tag{14}$$

It maps every $\mathbf{y} = \mathbf{y}_0$ onto that **x,** which by (11) is mapped onto $\mathbf{y}_0$. *The inverse of a linear transformation is itself linear,* because it is given by a matrix, as (14) shows.

PROBLEM SET 7.9

1–12 VECTOR SPACES

(Additional problems in Problem Set 7.4.)

Is the given set (taken with the usual addition and scalar multiplication) a vector space? (Give a reason.) If your answer is yes, find the dimension and a basis.

1. All vectors in R^3 satisfying $5v_1 - 3v_2 + 2v_3 = 0$
2. All vectors in R^3 satisfying $2v_1 + 3v_2 - v_3 = 0$, $v_1 - 4v_2 + v_3 = 0$
3. All 2×3 matrices with all entries nonnegative
4. All symmetric 3×3 matrices
5. All vectors in R^5 with the first three components 0
6. All vectors in R^4 with $v_1 + v_2 = 0$, $v_3 - v_4 = 1$
7. All skew-symmetric 2×2 matrices
8. All $n \times n$ matrices **A** with fixed n and $\det \mathbf{A} = 0$
9. All polynomials with positive coefficients and degree 3 or less
10. All functions $f(x) = a \cos x + b \sin x$ with any constants a and b
11. All functions $f(x) = (ax + b)e^{-x}$ with any constants a and b
12. All 2×3 matrices with the second row any multiple of $[4 \quad 0 \quad -9]$