Chapter 1

Matrix Operations

> Erſtlich wird alles dasjenige eine Größe genennt,
> welches einer Vermehrung oder einer Verminderung fähig iſt,
> oder wozu ſich noch etwas hinzuſetzen oder davon wegnehmen läßt.
>
> Leonhard Euler

Matrices play a central role in this book. They form an important part of the theory, and many concrete examples are based on them. Therefore it is essential to develop facility in matrix manipulation. Since matrices pervade much of mathematics, the techniques needed here are sure to be useful elsewhere.

The concepts which require practice to handle are *matrix multiplication* and *determinants*.

1. THE BASIC OPERATIONS

Let m, n be positive integers. An $m \times n$ matrix is a collection of mn numbers arranged in a rectangular array:

(1.1)
$$\begin{array}{cc} & n \text{ columns} \\ m \text{ rows} & \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

For example, $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 5 \end{bmatrix}$ is a 2×3 matrix.

The numbers in a matrix are called the *matrix entries* and are denoted by a_{ij}, where i, j are indices (integers) with $1 \le i \le m$ and $1 \le j \le n$. The index i is called the *row index*, and j is the *column index*. So a_{ij} is the entry which appears in the ith row and jth column of the matrix:

$$\begin{array}{cc} & j \\ i & \begin{bmatrix} \cdots & a_{ij} & \cdots\cdots \\ & \vdots & \\ & \vdots & \end{bmatrix} \end{array}$$

In the example above, $a_{11} = 2$, $a_{13} = 0$, and $a_{23} = 5$.

We usually introduce a symbol such as A to denote a matrix, or we may write it as (a_{ij}).

A $1 \times n$ matrix is called an n-dimensional *row vector*. We will drop the index i when $m = 1$ and write a row vector as

(1.2) $$A = [a_1 \cdots a_n], \quad \text{or as} \quad A = (a_1, \ldots, a_n).$$

The commas in this row vector are optional. Similarly, an $m \times 1$ matrix is an m-dimensional *column vector*:

(1.3) $$B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

A 1×1 matrix $[a]$ contains a single number, and we do not distinguish such a matrix from its entry.

(1.4) *Addition* of matrices is vector addition:

$$(a_{ij}) + (b_{ij}) = (s_{ij}),$$

where $s_{ij} = a_{ij} + b_{ij}$ for all i, j. Thus

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

The sum of two matrices A, B is defined only when they are both of the same shape, that is, when they are $m \times n$ matrices with the same m and n.

(1.5) *Scalar multiplication* of a matrix by a number is defined as with vectors. The result of multiplying a number c and a matrix (a_{ij}) is another matrix:

$$c(a_{ij}) = (b_{ij}),$$

where $b_{ij} = ca_{ij}$ for all i, j. Thus

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

Numbers will also be referred to as *scalars*.

The complicated notion is that of *matrix multiplication*. The first case to learn is the product AB of a row vector A (1.2) and a column vector B (1.3) which is defined when both are the same size, that is, $m = n$. Then the product AB is the 1×1 matrix or scalar

$$a_1b_1 + a_2b_2 + \cdots + a_mb_m. \tag{1.6}$$

(This product is often called the "dot product" of the two vectors.) Thus

$$\begin{bmatrix} 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix} = 3 \cdot 1 + 1 \cdot (-1) + 2 \cdot 4 = 10.$$

The usefulness of this definition becomes apparent when we regard A and B as vectors which represent indexed quantities. For example, consider a candy bar containing m ingredients. Let a_i denote the number of grams of $(ingredient)_i$ per candy bar, and let b_i denote the cost of $(ingredient)_i$ per gram. Then the matrix product $AB = c$ computes the cost per candy bar:

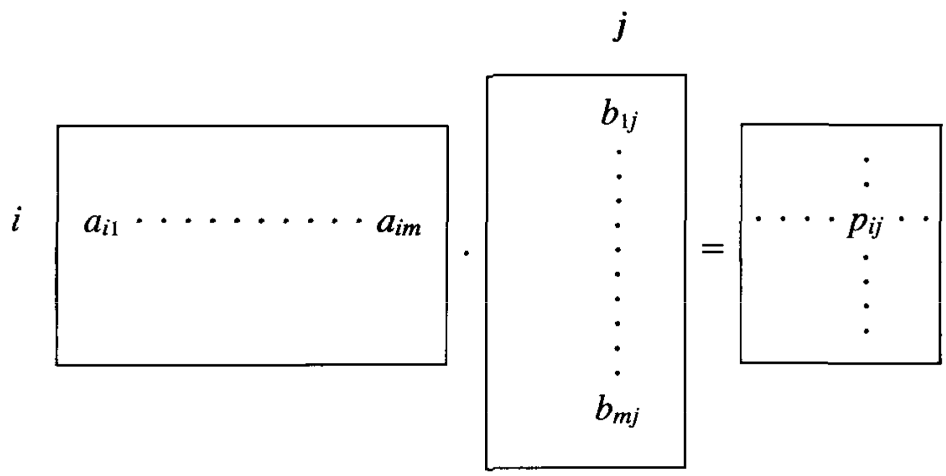
$$(\text{grams/bar}) \cdot (\text{cost/gram}) = (\text{cost/bar}).$$

On the other hand, the fact that we consider this to be the product of a row by a column is an arbitrary choice.

In general, the product of two matrices A and B is defined if the number of columns of A is equal to the number of rows of B, say if A is an $\ell \times m$ matrix and B is an $m \times n$ matrix. In this case, the product is an $\ell \times n$ matrix. Symbolically, $(\ell \times m) \cdot (m \times n) = (\ell \times n)$. The entries of the product matrix are computed by multiplying all rows of A by all columns of B, using rule (1.6) above. Thus if we denote the product AB by P, then

$$p_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}. \tag{1.7}$$

This is the product of the ith row of A and the jth column of B.



For example,

$$\begin{bmatrix} 0 & -1 & 2 \\ 3 & 4 & -6 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \tag{1.8}$$

This definition of matrix multiplication has turned out to provide a very convenient computational tool.

Going back to our candy bar example, suppose that there are ℓ candy bars. Then we may form a matrix A whose ith row measures the ingredients of $(bar)_i$. If the cost is to be computed each year for n years, we may form a matrix B whose jth column measures the cost of the ingredients in $(year)_j$. The matrix product $AB = P$ computes the cost per bar: p_{ij} = cost of $(bar)_i$ in $(year)_j$.

Matrix notation was introduced in the nineteenth century to provide a shorthand way of writing linear equations. The system of equations

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

can be written in matrix notation as

$$AX = B, \tag{1.9}$$

where A denotes the coefficient matrix (a_{ij}), X and B are column vectors, and AX is the matrix product

$$\boxed{\qquad\quad A \qquad\quad} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

Thus the matrix equation

$$\begin{bmatrix} 0 & -1 & 2 \\ 3 & 4 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

represents the following system of two equations in three unknowns:

$$\begin{aligned} -x_2 + 2x_3 &= 2 \\ 3x_1 + 4x_2 - 6x_3 &= 1. \end{aligned}$$

Equation (1.8) exhibits one solution: $x_1 = 1$, $x_2 = 4$, $x_3 = 3$.

Formula (1.7) defining the product can also be written in "sigma" notation as

$$p_{ij} = \sum_{k=1}^{m} a_{ik}b_{kj} = \sum_{k} a_{ik}b_{kj}.$$

Each of these expressions is a shorthand notation for the sum (1.7) which defines the product matrix.

Our two most important notations for handling sets of numbers are the Σ or sum notation as used above and matrix notation. The Σ notation is actually the more versatile of the two, but because matrices are much more compact we will use them whenever possible. One of our tasks in later chapters will be to translate complicated mathematical structures into matrix notation in order to be able to work with them conveniently.

Various *identities* are satisfied by the matrix operations, such as the *distributive laws*

$$A(B + B') = AB + AB', \quad \text{and} \quad (A + A')B = AB + A'B \tag{1.10}$$

and the *associative law*

$$(AB)C = A(BC). \tag{1.11}$$

These laws hold whenever the matrices involved have suitable sizes, so that the products are defined. For the associative law, for example, the sizes should be $A = \ell \times m$, $B = m \times n$ and, $C = n \times p$, for some ℓ, m, n, p. Since the two products (1.11) are equal, the parentheses are not required, and we will denote them by ABC. The triple product ABC is then an $\ell \times p$ matrix. For example, the two ways of computing the product

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

are

$$(AB)C = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} \quad \text{and} \quad A(BC) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix}.$$

Scalar multiplication is compatible with matrix multiplication in the obvious sense:

$$c(AB) = (cA)B = A(cB). \tag{1.12}$$

The proofs of these identities are straightforward and not very interesting.

In contrast, *the commutative law does not hold* for matrix multiplication; that is,

$$AB \neq BA, \text{ usually.} \tag{1.13}$$

In fact, if A is an $\ell \times m$ matrix and B is an $m \times \ell$ matrix, so that AB and BA are both defined, then AB is $\ell \times \ell$ while BA is $m \times m$. Even if both matrices are square, say $m \times m$, the two products tend to be different. For instance,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \text{ while } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Since matrix multiplication is not commutative, care must be taken when working with matrix equations. We can multiply both sides of an equation $B = C$ on the left by a matrix A, to conclude that $AB = AC$, provided that the products are defined. Similarly, if the products are defined, then we can conclude that $BA = CA$. We can not derive $AB = CA$ from $B = C$!

Any matrix all of whose entries are 0 is called a *zero matrix* and is denoted by 0, though its size is arbitrary. Maybe $0_{m \times n}$ would be better.

The entries a_{ii} of a matrix A are called its *diagonal entries,* and a matrix A is called a *diagonal matrix* if its only nonzero entries are diagonal entries.

The square $n \times n$ matrix whose only nonzero entries are 1 in each diagonal position,

(1.14)
$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix},$$

is called the $n \times n$ *identity* matrix. It behaves like 1 in multiplication: If A is an $m \times n$ matrix, then

$$I_m A = A \quad \text{and} \quad A I_n = A.$$

Here are some shorthand ways of drawing the matrix I_n:

$$I_n = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}.$$

We often indicate that a whole region in a matrix consists of zeros by leaving it blank or by putting in a single 0.

We will use $*$ to indicate an arbitrary undetermined entry of a matrix. Thus

$$\begin{bmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{bmatrix}$$

may denote a square matrix whose entries below the diagonal are 0, the other entries being undetermined. Such a matrix is called an *upper triangular* matrix.

Let A be a (square) $n \times n$ matrix. If there is a matrix B such that

(1.15)
$$AB = I_n \quad \text{and} \quad BA = I_n,$$

then B is called an *inverse* of A and is denoted by A^{-1}:

(1.16)
$$A^{-1}A = I_n = AA^{-1}.$$

When A has an inverse, it is said to be an *invertible* matrix. For example, the matrix $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$ is invertible. Its inverse is $A^{-1} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}$, as is seen by computing the products AA^{-1} and $A^{-1}A$. Two more examples are:

$$\begin{bmatrix} 1 & \\ & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \\ & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ & 1 \end{bmatrix}.$$

We will see later that A is invertible if there is a matrix B such that either one of the two relations $AB = I_n$ or $BA = I_n$ holds, and that B is then the inverse [see (2.23)]. But since multiplication of matrices is not commutative, this fact is not obvious. It fails for matrices which aren't square. For example, let $A = [1 \quad 2]$ and let $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then $AB = [1] = I_1$, but $BA = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \neq I_2$.

On the other hand, an inverse is unique if it exists at all. In other words, there can be only one inverse. Let B, B' be two matrices satisfying (1.15), for the same matrix A. We need only know that $AB = I_n$ (B is a *right inverse*) and that $B'A = I_n$ (B' is a *left inverse*). By the associative law, $B'(AB) = (B'A)B$. Thus

$$B' = B'I = B'(AB) = (B'A)B = IB = B, \tag{1.17}$$

and so $B' = B$. □

(1.18) **Proposition.** Let A, B be $n \times n$ matrices. If both are invertible, so is their product AB, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

More generally, if $A_1, \ldots, A_m$ are invertible, then so is the product $A_1 \cdots A_m$, and its inverse is $A_m^{-1} \cdots A_1^{-1}$.

Thus the inverse of $\begin{bmatrix} 1 & \\ & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & -1 \\ & 1 \end{bmatrix}\begin{bmatrix} 1 & \\ & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} \\ & \frac{1}{2} \end{bmatrix}$.

Proof. Assume that A, B are invertible. Then we check that $B^{-1}A^{-1}$ is the inverse of AB:

$$ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I,$$

and similarly

$$B^{-1}A^{-1}AB = \cdots = I.$$

The last assertion is proved by induction on m [see Appendix (2.3)]. When $m = 1$, the assertion is that if A_1 is invertible then A_1^{-1} is the inverse of A_1, which is trivial. Next we assume that the assertion is true for $m = k$, and we proceed to check it for $m = k + 1$. We suppose that $A_1, \ldots, A_{k+1}$ are invertible $n \times n$ matrices, and we denote by P the product $A_1 \cdots A_k$ of the first k matrices. By the induction hypothesis, P is invertible, and its inverse is $A_k^{-1} \cdots A_1^{-1}$. Also, A_{k+1} is invertible. So, by what has been shown for two invertible matrices, the product $PA_{k+1} = A_1 \cdots A_k A_{k+1}$ is invertible, and its inverse is $A_{k+1}^{-1}P^{-1} = A_{k+1}^{-1}A_k^{-1} \cdots A_1^{-1}$. This shows that the assertion is true for $m = k + 1$, which completes the induction proof. □

Though this isn't clear from the definition of matrix multiplication, we will see that most square matrices are invertible. But finding the inverse explicitly is not a simple problem when the matrix is large.

The set of all invertible $n \times n$ matrices is called the n-dimensional *general linear group* and is denoted by GL_n. The general linear groups will be among our most important examples when we study the basic concept of a group in the next chapter.

Various tricks simplify matrix multiplication in favorable cases. *Block multiplication* is one of them. Let M, M' be $m \times n$ and $n \times p$ matrices, and let r be an integer less than n. We may decompose the two matrices into blocks as follows:

$$M = [A|B] \quad \text{and} \quad M' = \left[\begin{array}{c} A' \\ \hline B' \end{array}\right],$$

where A has r columns and A' has r rows. Then the matrix product can be computed as follows:

$$MM' = AA' + BB'. \tag{1.19}$$

This decomposition of the product follows directly from the definition of multiplication, and it may facilitate computation. For example,

$$\left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 7 \end{array}\right]\left[\begin{array}{cc} 2 & 3 \\ 4 & 8 \\ \hline 0 & 0 \end{array}\right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 4 & 8 \end{bmatrix} + \begin{bmatrix} 5 \\ 7 \end{bmatrix}[0 \quad 0] = \begin{bmatrix} 2 & 3 \\ 4 & 8 \end{bmatrix}.$$

Note that formula (1.19) looks the same as rule (1.6) for multiplying a row vector and a column vector.

We may also multiply matrices divided into more blocks. For our purposes, a decomposition into four blocks will be the most useful. In this case the rule for block multiplication is the same as for multiplication of 2×2 matrices. Let $r + s = n$ and let $k + \ell = m$. Suppose we decompose an $m \times n$ matrix M and an $n \times p$ matrix M' into submatrices

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \quad M' = \left[\begin{array}{c|c} A' & B' \\ \hline C' & D' \end{array}\right],$$

where the number of columns of A is equal to the number of rows of A'. Then the rule for block multiplication is

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]\left[\begin{array}{c|c} A' & B' \\ \hline C' & D' \end{array}\right] = \left[\begin{array}{c|c} AA' + BC' & AB' + BD' \\ \hline CA' + DC' & CB' + DD' \end{array}\right]. \tag{1.20}$$

For example,

$$\left[\begin{array}{cc|c} 1 & 0 & 5 \\ \hline 0 & 1 & 7 \end{array}\right] \cdot \left[\begin{array}{cc|cc} 2 & 3 & 1 & 1 \\ 4 & 1 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 \end{array}\right] = \left[\begin{array}{cc|cc} 2 & 8 & 6 & 1 \\ \hline 4 & 8 & 7 & 0 \end{array}\right].$$

In this product, the upper left block is $[1\ \ 0]\begin{bmatrix}2 & 3\\ 4 & 1\end{bmatrix} + [5][0\ \ 1] = [2\ \ 8]$, etc.

Again, this rule can be verified directly from the definition of matrix multiplication. In general, block multiplication can be used whenever two matrices are decomposed into submatrices in such a way that the necessary products are defined.

Besides facilitating computations, block multiplication is a useful tool for proving facts about matrices by induction.

2. ROW REDUCTION

Let $A = (a_{ij})$ be an $m \times n$ matrix, and consider a variable $n \times p$ matrix $X = (x_{ij})$. Then the matrix equation

$$Y = AX \tag{2.1}$$

defines the $m \times p$ matrix $Y = (y_{ij})$ as a function of X. This operation is called *left multiplication by A*:

$$y_{ij} = a_{i1}x_{1j} + \cdots + a_{in}x_{nj}. \tag{2.2}$$

Notice that in formula (2.2) the entry y_{ij} depends only on $x_{1j},\dots,x_{nj}$, that is, on the jth column of X and on the ith row of the matrix A. Thus A operates separately on each column of X, and we can understand the way A operates by considering its action on column vectors:

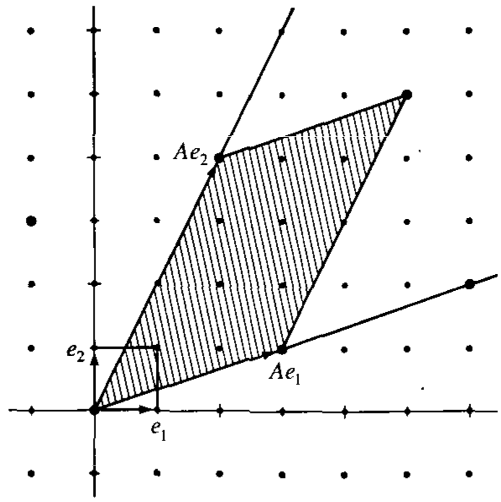
$$\boxed{\qquad A \qquad}\begin{bmatrix}x_1\\ \cdot\\ \cdot\\ \cdot\\ x_n\end{bmatrix} = \begin{bmatrix}y_1\\ \cdot\\ \cdot\\ y_m\end{bmatrix}.$$

Left multiplication by A on column vectors can be thought of as a function from the space of n-dimensional column vectors X to the space of m-dimensional column vectors Y, or a collection of m functions of n variables:

$$y_i = a_{i1}x_1 + \cdots + a_{in}x_n \qquad (i = 1,\dots,m).$$

It is called a *linear transformation,* because the functions are homogeneous and linear. (A *linear* function of a set of variables $u_1,\dots,u_k$ is one of the form $a_1u_1 + \cdots + a_ku_k + c$, where $a_1,\dots,a_k,c$ are scalars. Such a function is *homogeneous linear* if the constant term c is zero.)

A picture of the operation of the 2×2 matrix $\begin{bmatrix}3 & 2\\ 1 & 4\end{bmatrix}$ is shown below. It maps 2-space to 2-space:



(2.3) **Figure.**

Going back to the operation of A on an $n \times p$ matrix X, we can interpret the fact that A acts in the same way on each column of X as follows: Let Y_i denote the ith row of Y, which we view as a *row vector:*

$$Y = \begin{bmatrix} \text{---} & Y_1 & \text{---} \\ \text{---} & Y_2 & \text{---} \\ & \vdots & \\ \text{---} & Y_p & \text{---} \end{bmatrix}$$

We can compute Y_i in terms of the rows X_j of X, in vector notation, as

(2.4) $$Y_i = a_{i1}X_1 + \cdots + a_{in}X_n.$$

This is just a restatement of (2.2), and it is another example of block multiplication. For example, the bottom row of the product

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

can be computed as $3[1 \quad 0] + 4[4 \quad 2] - 6[3 \quad 2] = [1 \quad -4]$.

When A is a square matrix, we often speak of left multiplication by A as a *row operation*.

The simplest nonzero matrices are the matrix units, which we denote by e_{ij}:

(2.5) $$e_{ij} = \begin{array}{cc} & j \\ i & \begin{bmatrix} & \cdot & \\ \cdot\ \cdot & 1 & \cdot\ \cdot\ \cdot\ \cdot \\ & \cdot & \\ & \vdots & \end{bmatrix} \end{array}.$$

This matrix e_{ij} has a 1 in the (i, j) position as its only nonzero entry. (We usually denote matrices by capital letters, but the use of a small letter for the matrix units is traditional.) Matrix units are useful because every matrix $A = (a_{ij})$ can be written out as a sum in the following way:

$$A = a_{11}e_{11} + a_{12}e_{12} + \cdots + a_{nn}e_{nn} = \sum_{i,j} a_{ij}e_{ij}.$$

The indices i, j under the sigma mean that the sum is to be taken over all values of i and all values of j. For instance

$$\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & \\ & \end{bmatrix} + \begin{bmatrix} & 2 \\ & \end{bmatrix} + \begin{bmatrix} & \\ 1 & \end{bmatrix} + \begin{bmatrix} & \\ & 4 \end{bmatrix} = 3e_{11} + 2e_{12} + 1e_{21} + 4e_{22}.$$

Such a sum is called a linear combination of the matrices e_{ij}.

The matrix units are convenient for the study of addition and scalar multiplication of matrices. But to study matrix multiplication, some square matrices called *elementary matrices* are more useful. There are three types of elementary matrix:

$$\text{(2.6i)} \qquad \begin{bmatrix} 1 & & & \\ & \ddots & a & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & a & \ddots & \\ & & & 1 \end{bmatrix} = I + ae_{ij} \qquad (i \neq j).$$

Such a matrix has diagonal entries 1 and one nonzero off-diagonal entry.

$$\text{(2.6ii)} \qquad \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & & 1 \\ & & & \ddots & \\ & & 1 & & 0 \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix} = I + e_{ij} + e_{ji} - e_{ii} - e_{jj}.$$

Here the ith and jth diagonal entries of I are replaced by zero, and two 1's are added in the (i, j) and (j, i) positions. (The formula in terms of the matrix units is rather ugly, and we won't use it much.)

$$\text{(2.6iii)} \qquad \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix} = I + (c - 1)e_{ii}, \qquad (c \neq 0).$$

One diagonal entry of the identity matrix is replaced by a nonzero number c.

The elementary 2×2 matrices are

$$\text{(i)} \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix}, \quad \text{(ii)} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{(iii)} \begin{bmatrix} c & \\ & 1 \end{bmatrix}, \begin{bmatrix} 1 & \\ & c \end{bmatrix},$$

where, as above, a is arbitrary and c is an arbitrary nonzero number.

The elementary matrices E operate on a matrix X as described below.

(2.7) To get the matrix EX, you must:

Type (i): *Replace the* ith *row* X_i *by* $X_i + aX_j$, or *add* $a\cdot$(row j) *to* (row i);

Type (ii): *Interchange* (row i) *and* (row j);

Type (iii): *Multiply* (row i) *by a nonzero scalar* c.

These operations are called *elementary row operations*. Thus multiplication by an elementary matrix is an elementary row operation. You should verify these rules of multiplication carefully.

(2.8) **Lemma.** Elementary matrices are invertible, and their inverses are also elementary matrices.

The proof of this lemma is just a calculation. The inverse of an elementary matrix is the matrix corresponding to the inverse row operation: If $E = I + ae_{ij}$ is of Type (i), then $E^{-1} = I - ae_{ij}$; "*subtract* $a\cdot$(row j) *from* (row i)". If E is of Type (ii), then $E^{-1} = E$, and if E is of Type (iii), then E^{-1} is of the same type, with c^{-1} in the position that c has in E; "*multiply* (row i) *by* c^{-1}". □

We will now study the effect of elementary row operations (2.7) on a matrix A, with the aim of ending up with a simpler matrix A':

$$A \xrightarrow{\text{sequence of operations}} \cdots \longrightarrow A'.$$

Since each elementary row operation is obtained as the result of multiplication by an elementary matrix, we can express the result of a succession of such operations as multiplication by a sequence $E_1, \ldots, E_k$ of elementary matrices:

(2.9)
$$A' = E_k \cdots E_2 E_1 A.$$

This procedure is called *row reduction*, or *Gaussian elimination*. For example, we can simplify the matrix

(2.10)
$$M = \begin{bmatrix} 1 & 0 & 2 & 1 & 5 \\ 1 & 1 & 5 & 2 & 7 \\ 1 & 2 & 8 & 4 & 12 \end{bmatrix}$$

by using the first type of elementary operation to clear out as many entries as possible:

$$\begin{bmatrix} 1 & 0 & 2 & 1 & 5 \\ 1 & 1 & 5 & 2 & 7 \\ 1 & 2 & 8 & 4 & 12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 2 & 1 & 5 \\ 0 & 1 & 3 & 1 & 2 \\ 1 & 2 & 8 & 4 & 12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 2 & 1 & 5 \\ 0 & 1 & 3 & 1 & 2 \\ 0 & 2 & 6 & 3 & 7 \end{bmatrix} \longrightarrow$$

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

Row reduction is a useful method of solving systems of linear equations. Suppose we are given a system of m equations in n unknowns, say $AX = B$ as in (1.9), where A is an $m \times n$ matrix, X is an unknown column vector, and B is a given column vector. To solve this system, we form the $m \times (n+1)$ block matrix

$$\text{(2.11)} \qquad M = [A \mid B] = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_n \end{array}\right],$$

and we perform row operations to simplify M. Note that $EM = [EA \mid EB]$. Let

$$M' = [A' \mid B']$$

be the result of a sequence of row operations. The key observation follows:

(2.12) **Proposition.** The solutions of $A'X = B'$ are the same as those of $AX = B$.

Proof. Since M' is obtained by a sequence of elementary row operations,

$$M' = E_r \cdots E_1 M.$$

Let $P = E_r \cdots E_1$. This matrix is invertible, by Lemma (2.8) and Proposition (1.18). Also, $M' = [A' \mid B'] = [PA \mid PB]$. If X is a solution of the original system $AX = B$, then $PAX = PB$, which is to say, $A'X = B'$. So X also solves the new system. Conversely, if $A'X = B'$, then $AX = P^{-1}A'X = P^{-1}B' = B$, so X solves the system $AX = B$ too. □

For example, consider the system

$$\text{(2.13)} \qquad \begin{aligned} x_1 + \phantom{x_2 + {}} 2x_3 + x_4 &= 5 \\ x_1 + x_2 + 5x_3 + 2x_4 &= 7 \\ x_1 + 2x_2 + 8x_3 + 4x_4 &= 12. \end{aligned}$$

Its augmented matrix is the matrix M considered above (2.10), so our row reduction of this matrix shows that this system of equations is equivalent to

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

We can read off the solutions of this system immediately: We may choose x_3 arbitrarily and then solve for x_1, x_2, and x_4. The general solution of (2.13) can therefore be written in the form

$$x_3 = c_3,\ x_1 = 1 - 2c_3,\ x_2 = -1 - 3c_3,\ x_4 = 3,$$

where c_3 is arbitrary.

We now go back to row reduction of an arbitrary matrix. It is not hard to see that, by a sequence of row operations, any matrix A can be reduced to one which looks roughly like this:

$$\text{(2.14)} \qquad A = \boxed{\begin{array}{cccccccc} 1 & *..* & 0 & *..* & 0 & *...* & 0 & \\ & & 1 & *..* & 0 & *...* & 0 & \\ & & & & 1 & *...* & 0 & \cdots \\ & & & & & & 1 & \\ & & & & & & & \ddots \end{array}},$$

where $*$ denotes an arbitrary number and the large blank space consists of zeros. This is called a *row echelon matrix*. For instance,

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

is a row echelon matrix. So is the end result of our reduction of (2.10). The definition of a row echelon matrix is given in (2.15):

(2.15)

(a) The first nonzero entry in every row is 1. This entry is called a *pivot*.

(b) The first nonzero entry of row $i + 1$ is to the right of the first nonzero entry of row i.

(c) The entries above a pivot are zero.

To make a row reduction, find the first column which contains a nonzero entry. (If there is none, then $A = 0$, and 0 is a row echelon matrix.) Interchange rows using an elementary operation of Type (ii), moving a nonzero entry to the top row. Normalize this entry to 1 using an operation of Type (iii). Then clear out the other entries in its column by a sequence of operations of Type (i). The resulting matrix will have the block form

$$\left[\begin{array}{ccc|c|ccc}
0 & \cdots & 0 & 1 & * & \cdots & * \\
\hline
0 & \cdots & 0 & 0 & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & * & \cdots & *
\end{array}\right], \text{ which we may write as } \left[\begin{array}{c|c|c} & 1 & B \\ \hline & & D \end{array}\right] = A'.$$

We now continue, performing row operations on the smaller matrix D (cooking until done). Formally, this is induction on the size of the matrix. The principle of complete induction [see Appendix (2.6)] allows us to assume that every matrix with fewer rows than A can be reduced to row echelon form. Since D has fewer rows, we may assume that it can be reduced to a row echelon matrix, say D''. The row operations we perform to reduce D to D'' will not change the other blocks making up A'. Therefore A' can be reduced to the matrix

$$\left[\begin{array}{c|c|c} & 1 & B \\ \hline & & D'' \end{array}\right] = A'',$$

which satisfies requirements (2.15a and b) for a row echelon matrix. Therefore our original matrix A can be reduced to this form. The entries in B above the pivots of D'' can be cleared out at this time, to finish the reduction to row echelon form. $\square$

It can be shown that the row echelon matrix obtained from a given matrix A by row reduction is unique, that is, that it does not depend on the particular sequence of operations used. However, this is not a very important point, so we omit the proof.

The reason that row reduction is useful is that we can solve a system of equations $A'X = B'$ immediately if A' is in row echelon form. For example, suppose that

$$[A' \,|\, B'] = \left[\begin{array}{cccc|c} 1 & 6 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

There is no solution to $A'X = B'$ because the third equation is $0 = 1$. On the other hand,

$$[A' \,|\, B'] = \left[\begin{array}{cccc|c} 1 & 6 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

has solutions. Choosing x_2, x_4 arbitrarily, we can solve the first equation for x_1 and the second for x_3. This is the procedure we use to solve system (2.13).

The general rule is as follows:

(2.16) **Proposition.** Let $M' = [A' \,|\, B']$ be a row echelon matrix. Then the system of equations $A'X = B'$ has a solution if and only if there is no pivot in the last column B'. In that case, an arbitrary value can be assigned to the unknown x_i if column i does not contain a pivot. $\square$

Of course every *homogeneous* linear system $AX = 0$ has the trivial solution $X = 0$. But looking at the row echelon form again, we can conclude that if there are more unknowns than equations then the homogeneous equation $AX = 0$ has a *nontrivial* solution for X :

(2.17) **Corollary.** Every system $AX = 0$ of m homogeneous equations in n unknowns, with $m < n$, has a solution X in which some x_i is nonzero.

For, let $A'X = 0$ be the associated row echelon equation, and let r be the number of pivots of A'. Then $r \leq m$. According to the proposition, we may assign arbitrary values to $n - r$ variables x_i. □

We will now use row reduction to characterize square invertible matrices.

(2.18) **Proposition.** Let A be a square matrix. The following conditions are equivalent:

(a) A can be reduced to the identity by a sequence of elementary row operations.
(b) A is a product of elementary matrices.
(c) A is invertible.
(d) The system of homogeneous equations $AX = 0$ has only the trivial solution $X = 0$.

Proof. We will prove this proposition by proving the implications (a)⇒(b)⇒(c)⇒(d)⇒(a). To show that (a) implies (b), suppose that A can be reduced to the identity by row operations: $E_k \cdots E_1 A = I$. Multiplying both sides of this equation on the left by $E_1^{-1} \cdots E_k^{-1}$, we obtain $A = E_1^{-1} \cdots E_k^{-1}$. Since the inverse of an elementary matrix is elementary, this shows that A is a product of elementary matrices. Because a product of elementary matrices is invertible, (b) implies (c). If A is invertible we can multiply both sides of the equation $AX = 0$ by A^{-1} to derive $X = 0$. So the equation $AX = 0$ has only the trivial solution. This shows that (c) implies (d).

To prove the last implication, that (d) implies (a), we take a look at *square* row echelon matrices M. We note the following dichotomy:

(2.19) *Let M be a square row echelon matrix.*
Either M is the identity matrix, or its bottom row is zero.

This is easy to see, from (2.15).

Suppose that (a) does not hold for a given matrix A. Then A can be reduced by row operations to a matrix A' whose bottom row is zero. In this case there are at most $n-1$ nontrivial equations in the linear system $A'X = 0$, and so Corollary (2.17) tells us that this system has a nontrivial solution. Since the equation $AX = 0$ is equivalent to $A'X = 0$, it has a nontrivial solution as well. This shows that if (a) fails then (d) does too; hence (d) implies (a). This completes the proof of Proposition (2.18). □

(2.20) **Corollary.** If a row of a square matrix A is zero, then A is not invertible. □

Row reduction provides a method of computing the inverse of an invertible matrix A: We reduce A to the identity by row operations:

$$E_k \cdots E_1 A = I$$

as above. Multiplying both sides of this equation on the right by A^{-1}, we have

$$E_k \cdots E_1 I = A^{-1}.$$

(2.21) **Corollary.** Let A be an invertible matrix. To compute its inverse A^{-1}, apply elementary row operations $E_1, \ldots, E_k$ to A, reducing it to the identity matrix. The same sequence of operations, when applied to I, yields A^{-1}.

The corollary is just a restatement of the two equations. □

(2.22) **Example.** We seek the inverse of the matrix

$$A = \begin{bmatrix} 5 & 4 \\ 6 & 5 \end{bmatrix}.$$

To compute it we form the 2×4 block matrix

$$[A \mid I] = \left[\begin{array}{cc|cc} 5 & 4 & 1 & 0 \\ 6 & 5 & 0 & 1 \end{array}\right].$$

We perform row operations to reduce A to the identity, carrying the right side along, and thereby end up with A^{-1} on the right because of Corollary (2.21).

$$\begin{aligned}
[A \mid I] = \left[\begin{array}{cc|cc} 5 & 4 & 1 & 0 \\ 6 & 5 & 0 & 1 \end{array}\right] & \qquad \text{Subtract (row 1) from (row 2)} \\
\longrightarrow \left[\begin{array}{cc|cc} 5 & 4 & 1 & 0 \\ 1 & 1 & -1 & 1 \end{array}\right] & \qquad \text{Subtract } 4 \cdot \text{(row 2) from (row 1)} \\
\longrightarrow \left[\begin{array}{cc|cc} 1 & 0 & 5 & -4 \\ 1 & 1 & -1 & 1 \end{array}\right] & \qquad \text{Subtract (row 1) from (row 2)} \\
\longrightarrow \left[\begin{array}{cc|cc} 1 & 0 & 5 & -4 \\ 0 & 1 & -6 & 5 \end{array}\right] & = [I \mid A^{-1}].
\end{aligned}$$

Thus $A^{-1} = \begin{bmatrix} 5 & -4 \\ -6 & 5 \end{bmatrix}$.

(2.23) **Proposition.** Let A be a square matrix which has either a left inverse B: $BA = I$, or a right inverse: $AB = I$. Then A is invertible, and B is its inverse.

Proof. Suppose that $AB = I$. We perform row reduction on A. According to (2.19), there are elementary matrices $E_1, \ldots, E_k$ so that $A' = E_k \ldots E_1 A$ either is the identity matrix or has bottom row zero. Then $A'B = E_k \dots E_1$, which is an invertible matrix. Hence the bottom row of $A'B$ is not zero, and it follows that A' has a nonzero bottom row too. So $A' = I$. By (2.18), A is invertible, and the equations $I = E_k \dots E_1 A$ and $AB = I$ show that $A^{-1} = E_k \dots E_1 = B$ (see (1.17)). The other case is that $BA = I$. Then we can interchange A and B in the above argument and conclude that B is invertible and A is its inverse. So A is invertible too. □

For most of this discussion, we could have worked with columns rather than rows. We chose to work with rows in order to apply the results to systems of linear equations; otherwise columns would have served just as well. Rows and columns are interchanged by the matrix *transpose*. The transpose of an $m \times n$ matrix A is the $n \times m$ matrix A^t obtained by reflecting about the diagonal: $A^t = (b_{ij})$, where

$$b_{ij} = a_{ji}.$$

For instance,

$$\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}^t = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^t = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The rules for computing with the transpose are given in (2.24):

(2.24)

(a) $$(A + B)^t = A^t + B^t.$$

(b) $$(cA)^t = cA^t.$$

(c) $$(AB)^t = B^t A^t!$$

(d) $$(A^t)^t = A.$$

Using formulas (2.24c and d), we can deduce facts about *right multiplication, XP*, from the corresponding facts about left multiplication.

The elementary matrices (2.6) act by right multiplication as the following *elementary column operations:*

(2.25)

(a) *Add* $a \cdot$ (column i) *to* (column j).
(b) *Interchange* (column i) *and* (column j).
(c) *Multiply* (column i) *by* $c \neq 0$.

3. DETERMINANTS

Every square matrix A has a number associated to it called its *determinant*. In this section we will define the determinant and derive some of its properties. The determinant of a matrix A will be denoted by det A.

The determinant of a 1×1 matrix is just its unique entry

$$\det [a] = a, \tag{3.1}$$

and the determinant of a 2×2 matrix is given by the formula

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc. \tag{3.2}$$

If we think of a 2×2 matrix A as an operator on the space $\mathbb{R}^2$ of real two-dimensional vectors, as in Section 2, then $\det A$ can be interpreted geometrically. Its absolute value is the area of the parallelogram which forms the image of a unit square under the operation. For example, the area of the shaded region of Figure (2.3) is 10. The determinant is positive or negative according to whether the orientation of the square is preserved or reversed by the operation. Moreover, $\det A = 0$ if and only if the parallelogram degenerates to a line segment, and this occurs if and only if the two columns of A are proportional.

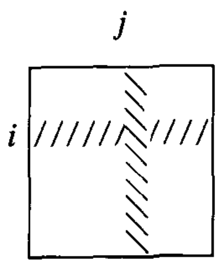
The set of all $n \times n$ matrices forms a space of dimension n^2, which we denote by $\mathbb{R}^{n \times n}$. We will regard the determinant of $n \times n$ matrices as a *function* from this space to the real numbers:

$$\det\colon \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}.$$

This just means that det is a function of the n^2 matrix entries. There is one such function for each positive integer n. Unfortunately there are many formulas for the determinant, and all of them are complicated when n is large. The determinant is important because it has very nice properties, though there is no simple formula for it. Not only are the formulas complicated, but it may not be easy to show directly that two of them define the same function. So we will use the following strategy: We choose one formula essentially at random and take it as the definition of the determinant. In that way we are talking about a particular function. We show that the function we have chosen has certain very special properties. We also show that our chosen function is the *only* one having these properties. Then, to check that some other formula defines the same determinant, we have to check only that the function which it defines has these same properties. It turns out that this is usually relatively easy.

The determinant of an $n \times n$ matrix can be computed in terms of certain $(n-1) \times (n-1)$ determinants by a process called *expansion by minors*. This expansion allows us to give a recursive definition of the determinant function. Let A be an $n \times n$ matrix and let A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained by crossing out the ith row and the jth column of A:

(3.3)



$= A_{ij}.$

For example, if

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 2 \\ 0 & 5 & 1 \end{bmatrix}, \text{ then } \quad A_{21} = \begin{bmatrix} 0 & 3 \\ 5 & 1 \end{bmatrix}.$$

Expansion by minors on the first column is the formula

(3.4) $$\det A = a_{11} \det A_{11} - a_{21} \det A_{21} +, - \cdots \pm a_{n1} \det A_{n1}.$$

The signs alternate. We take this formula, together with (3.1), as a recursive *definition of the determinant.* Notice that the formula agrees with (3.2) for 2×2 matrices.

The determinant of the matrix A shown above is

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

The three 2×2 determinants which appear here can be computed by expanding by minors again and using (3.1), or by using (3.2), to get

$$\det A = 1 \cdot (-9) - 2 \cdot (-15) + 0 \cdot (-3) = 21.$$

There are other formulas for the determinant, including expansions by minors on other columns and on rows, which we will derive presently [see (4.11, 5.1, 5.2)].

It is important, both for computation of determinants and for theoretical considerations, to know some of the many special properties satisfied by determinants. Most of them can be verified by direct computation and induction on n, using expansion by minors (3.4). We will list some without giving formal proofs. In order to be able to interpret these properties for functions other than the determinant, we will denote the determinant by the symbol d for the time being.

(3.5) $$\mathrm{d}(I) = 1.$$

(3.6) *The function* $\mathrm{d}(A)$ *is linear in the rows of the matrix.*

By this we mean the following: Let R_i denote the row vector which is the ith row of the matrix, so that A can be written symbolically as

$$A = \begin{bmatrix} \text{---} R_1 \text{---} \\ \vdots \\ \text{---} R_n \text{---} \end{bmatrix}.$$

By definition, *linearity* in the ith row means that whenever R and S are row vectors then

$$\mathrm{d}\begin{bmatrix} \vdots \\ \text{---} R+S \text{---} \\ \vdots \end{bmatrix} = \mathrm{d}\begin{bmatrix} \vdots \\ \text{---} R \text{---} \\ \vdots \end{bmatrix} + \mathrm{d}\begin{bmatrix} \vdots \\ \text{---} S \text{---} \\ \vdots \end{bmatrix},$$

and

$$\mathrm{d}\begin{bmatrix} \vdots \\ \text{——}cR\text{——} \\ \vdots \end{bmatrix} = c\,\mathrm{d}\begin{bmatrix} \vdots \\ \text{——}R\text{——} \\ \vdots \end{bmatrix},$$

where the other rows of the matrices appearing in these relations are the same throughout. For example,

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

and

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

Linearity allows us to operate on *one row at a time*, with the other rows left fixed.

Another property:

(3.7) *If two adjacent rows of a matrix A are equal, then* $\mathrm{d}(A) = 0$.

Let us prove this fact by induction on n. Suppose that rows j and $j+1$ are equal. Then the matrices A_{i1} defined by (3.3) also have two rows equal, except when $i = j$ or $i = j+1$. When A_{i1} has two equal rows, its determinant is zero by induction. Thus only two terms of (3.4) are different from zero, and

$$\mathrm{d}(A) = \pm a_{j1}\mathrm{d}(A_{j1}) \mp a_{j+1\,1}\mathrm{d}(A_{j+1\,1}).$$

Moreover, since the rows R_j and R_{j+1} are equal, it follows that $A_{j1} = A_{j+1\,1}$ and that $a_{j1} = a_{j+1\,1}$. Since the signs alternate, the two terms on the right side cancel, and the determinant is zero.

Properties (3.5–3.7) characterize determinants uniquely [see (3.14)], and we will derive further relations from them without going back to definition (3.4).

(3.8) *If a multiple of one row is added to an adjacent row, the determinant is unchanged.*

For, by (3.6) and (3.7),

$$\mathrm{d}\begin{bmatrix} \vdots \\ \text{——}R\text{——} \\ \text{——}S+cR\text{——} \\ \vdots \end{bmatrix} = \mathrm{d}\begin{bmatrix} \vdots \\ \text{——}R\text{——} \\ \text{——}S\text{——} \\ \vdots \end{bmatrix} + c\,\mathrm{d}\begin{bmatrix} \vdots \\ \text{——}R\text{——} \\ \text{——}R\text{——} \\ \vdots \end{bmatrix} = \mathrm{d}\begin{bmatrix} \vdots \\ \text{——}R\text{——} \\ \text{——}S\text{——} \\ \vdots \end{bmatrix}.$$

The same reasoning works if S is above R.

(3.9) *If two adjacent rows are interchanged, the determinant is multiplied by* -1.

We apply (3.8) repeatedly:

$$
\mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & R & \text{---} \\ \text{---} & S & \text{---} \\ & \vdots & \end{bmatrix}
= \mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & R & \text{---} \\ \text{---} & (S-R) & \text{---} \\ & \vdots & \end{bmatrix}
= \mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & R+(S-R) & \text{---} \\ \text{---} & (S-R) & \text{---} \\ & \vdots & \end{bmatrix}
$$

$$
= \mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & S & \text{---} \\ \text{---} & (S-R) & \text{---} \\ & \vdots & \end{bmatrix}
= \mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & S & \text{---} \\ \text{---} & (-R) & \text{---} \\ & \vdots & \end{bmatrix}
= -\mathrm{d}\begin{bmatrix} & \vdots & \\ \text{---} & S & \text{---} \\ \text{---} & R & \text{---} \\ & \vdots & \end{bmatrix}.
$$

(3.7′) *If two rows of a matrix A are equal, then* $\mathrm{d}(A) = 0$.

For, interchanging adjacent rows a few times results in a matrix A' with two adjacent rows equal. By (3.7) $\mathrm{d}(A') = 0$, and by (3.9) $\mathrm{d}(A) = \pm\det(A')$.

Using (3.7′), the proofs of (3.8) and (3.9) show the following:

(3.8′) *If a multiple of one row is added to another row, the determinant is not changed.*

(3.9′) *If two rows are interchanged, the determinant is multiplied by* -1.

Also, (3.6) implies the following:

(3.10) *If a row of A is zero, then* $\mathrm{d}(A) = 0$.

If a row is zero, then A doesn't change when we multiply that row by 0. But according to (3.6), $\mathrm{d}(A)$ gets multiplied by 0. Thus $\mathrm{d}(A) = 0\mathrm{d}(A) = 0$.

Rules (3.8′), (3.9′), and (3.6) describe the effect of an elementary row operation (2.7) on the determinant, so they can be rewritten in terms of the elementary matrices. They tell us that $\mathrm{d}(EA) = \mathrm{d}(A)$ if E is an elementary matrix of the first kind, that $\mathrm{d}(EA) = -d(A)$ if E is of the second kind, and (3.6) that $\mathrm{d}(EA) = c\mathrm{d}(A)$ if E is of the third kind. Let us apply these rules to compute $\mathrm{d}(E)$ when E is an elementary matrix. We substitute $A = I$. Then, since $\mathrm{d}(I) = 1$, the rules determine $\mathrm{d}(EI) = \mathrm{d}(E)$:

(3.11) The determinant of an elementary matrix is:

(i) First kind (*add a multiple of one row to another*): $\mathrm{d}(E) = 1$, by (3.8′).
(ii) Second kind (*row interchange*): $\mathrm{d}(E) = -1$, by (3.9′).
(iii) Third kind (*multiply a row by a nonzero constant*): $\mathrm{d}(E) = c$, by (3.6).

Moreover, if we use rules (3.8′), (3.9′), and (3.6) again, applying them this time to an arbitrary matrix A and using the values for $\mathrm{d}(E)$ which have just been determined, we obtain the following:

(3.12) *Let E be an elementary matrix and let A be arbitrary. Then*

$$\mathrm{d}(EA) = \mathrm{d}(E)\mathrm{d}(A).$$

Recall from (2.19) that every square matrix A can be reduced by elementary row operations to a matrix A' which is either the identity I or else has its bottom row zero:

$$A' = E_k \cdots E_1 A.$$

We know by (3.5) and (3.10) that $\mathrm{d}(A)' = 1$ or $\mathrm{d}(A') = 0$ according to the case. By (3.12) and induction,

$$\mathrm{d}(A') = \mathrm{d}(E_k) \cdots \mathrm{d}(E_1)\mathrm{d}(A). \tag{3.13}$$

We also know $\mathrm{d}(E_i)$, by (3.11), and hence we can use this formula to compute $\mathrm{d}(A)$.

(3.14) **Theorem.** *Axiomatic Characterization of the Determinant:* The determinant function (3.4) is the *only* one satisfying rules (3.5–3.7).

Proof. We used only these rules to arrive at equations (3.11) and (3.13), and they determine $\mathrm{d}(A)$. Since the expansion by minors (3.4) satisfies (3.5–3.7), it agrees with (3.13). □

We will now return to our usual notation $\det A$ for the determinant of a matrix.

(3.15) **Corollary.** A square matrix A is invertible if and only if $\det A \neq 0$.

This follows from formulas (3.11), (3.13), and (2.18). By (3.11), $\det E_i \neq 0$ for all i. Thus if A' is as in (3.13), then $\det A \neq 0$ if and only if $\det A' \neq 0$, which is the case if and only if $A' = I$. By (2.18), $A' = I$ if and only if A is invertible. □

We can now prove one of the most important properties of the determinant function: its compatibility with matrix multiplications.

(3.16) **Theorem.** Let A, B be any two $n \times n$ matrices. Then

$$\det(AB) = (\det A)(\det B).$$

Proof. We note that this is (3.12) if A is an elementary matrix.

Case 1: A is invertible. By (2.18b), A is a product of elementary matrices: $A = E_1 \cdots E_k$. By (3.12) and induction, $\det A = (\det E_1)\cdots(\det E_k)$, and $\det AB = \det(E_1 \cdots E_k B) = (\det E_1)\cdots(\det E_k)(\det B) = (\det A)(\det B)$.

Case 2: A is not invertible. Then $\det A = 0$ by (3.15), and so the theorem will follow in this case if we show that $\det(AB) = 0$ too. By (2.18), A can be reduced to a matrix $A' = E_k \cdots E_1 A$ having bottom row zero. Then the bottom row of $A'B$ is also zero; hence

$$0 = \det(A'B) = \det(E_k \cdots E_1 AB) = (\det E_k)\cdots(\det E_1)(\det AB).$$

Since $\det E_i \neq 0$, it follows that $\det AB = 0$. □

(3.17) **Corollary.** If A is invertible, $\det(A^{-1}) = \dfrac{1}{\det A}$.

Proof. $(\det A)(\det A^{-1}) = \det I = 1$. □

Note. It is a natural idea to try to define determinants using rules (3.11) and (3.16). These rules certainly determine $\det A$ for every invertible matrix A, since we can write such a matrix as a product of elementary matrices. But there is a problem. Namely, there are many ways to write a given matrix as a product of elementary matrices. Without going through some steps as we have, it is not clear that two such products would give the same answer for the determinant. It is actually not particularly easy to make this idea work.

The proof of the following proposition is a good exercise.

(3.18) **Proposition.** Let A^t denote the transpose of A. Then

$$\det A = \det A^t. \ \square$$

(3.19) **Corollary.** Properties (3.6–3.10) continue to hold if the word *row* is replaced by *column* throughout. □

4. PERMUTATION MATRICES

A bijective map p from a set S to itself is called a *permutation* of the set:

$$p: S \longrightarrow S. \tag{4.1}$$

For example,

$$\begin{aligned} 1 &\rightsquigarrow 3 \\ 2 &\rightsquigarrow 1 \\ 3 &\rightsquigarrow 2 \end{aligned} \tag{4.2}$$

is a permutation of the set $\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}$. It is called a *cyclic* permutation because it operates as

$$\begin{tikzcd} \mathbf{1} \arrow[dr] & & \mathbf{2} \arrow[ll] \\ & \mathbf{3} \arrow[ur] & \end{tikzcd} .$$

There are several notations for permutations. We will use function notation in this section, so that $p(\mathbf{x})$ denotes the value of the permutation p on the element $\mathbf{x}$. Thus if p is the permutation given in (4.2), then

$$p(\mathbf{1}) = \mathbf{3},\ p(\mathbf{2}) = \mathbf{1},\ p(\mathbf{3}) = \mathbf{2}.$$

A *permutation matrix* P is a matrix with the following property: The operation of left multiplication by P is a permutation of the rows of a matrix. The elementary matrices of the second type (2.6ii) are the simplest examples. They correspond to the permutations called *transpositions*, which interchange two rows of a matrix, leaving the others alone. Also,

(4.3)
$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

is a permutation matrix. It acts on a column vector $X = (x, y, z)^t$ as

$$PX = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix}.$$

The entry in the first position is sent to the third position, and so on, so P has permuted rows according to the cyclic permutation p given in (4.2).

There is one point which can cause confusion and which makes it important for us to establish our notation carefully. When we permute the *entries* of a vector $(x_1, \ldots, x_n)^t$ according to a permutation p, the *indices* are permuted in the opposite way. For instance, multiplying the column vector $X = (x_1, x_2, x_3)^t$ by the matrix in (4.3) gives

(4.4)
$$PX = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ x_1 \end{bmatrix}.$$

The indices in (4.4) are permuted by $\mathbf{1} \rightsquigarrow \mathbf{2} \rightsquigarrow \mathbf{3} \rightsquigarrow \mathbf{1}$, which is the inverse of the permutation p. Thus there are two ways to associate a permutation to a permutation matrix P: the permutation p which describes how P permutes the entries, and the inverse operation which describes the effect on indices. We must make a decision, so we will say that the permutation associated to P is the one which describes its action on the entries of a column vector. Then the indices are permuted in the opposite way, so

(4.5)
$$PX = \begin{bmatrix} x_{p^{-1}(1)} \\ \vdots \\ x_{p^{-1}(n)} \end{bmatrix}.$$

Multiplication by P has the corresponding effect on the rows of an $n \times r$ matrix A.

The permutation matrix P can be written conveniently in terms of the matrix units (2.5) or in terms of certain column vectors called the *standard basis* and denoted by e_i. The vector e_i has a 1 in the ith position as its single nonzero entry, so these vectors are the matrix units for an $n \times 1$ matrix.

(4.6) **Proposition.** Let P be the permutation matrix associated to a permutation p.

(a) The jth column of P is the column vector $e_{p(j)}$.

(b) P is a sum of n matrix units: $P = e_{p(1)1} + \cdots + e_{p(n)n} = \sum_j e_{p(j)j}$. □

A permutation matrix P always has a single 1 in each row and in each column, the rest of its entries being 0. Conversely, any such matrix is a permutation matrix.

(4.7) **Proposition.**

(a) Let p, q be two permutations, with associated permutation matrices P,Q. Then the matrix associated to the permutation pq is the product PQ.

(b) A permutation matrix P is invertible, and its inverse is the transpose matrix: $P^{-1} = P^t$.

Proof. By pq we mean the composition of the two permutations

$$pq(i) = p(q(i)). \tag{4.8}$$

Since P operates by permuting rows according to p and Q operates by permuting according to q, the associative law for matrix multiplication tells us that PQ permutes according to pq:

$$(PQ)X = P(QX).$$

Thus PQ is the permutation matrix associated to pq. This proves (a). We leave the proof of (b) as an exercise. □

The determinant of a permutation matrix is easily seen to be ± 1, using rule (3.9). This determinant is called the *sign of a permutation:*

$$\operatorname{sign} p = \det P = \pm 1. \tag{4.9}$$

The permutation (4.2) has sign $+1$, while any transposition has sign -1 [see (3.11ii)]. A permutation p is called *odd* or *even* according to whether its sign is -1 or $+1$.

Let us now go back to an arbitrary $n \times n$ matrix A and use linearity of the determinant (3.6) to expand $\det A$. We begin by working on the first row. Applying (3.6), we find that

$$\det A = \det\begin{bmatrix} a_{11}0 \;\ldots\ldots\; 0 \\ \text{———} R_2 \text{———} \\ \vdots \\ \text{———} R_n \text{———} \end{bmatrix} + \det\begin{bmatrix} 0a_{12}0 \;\ldots\ldots\; 0 \\ \text{———} R_2 \text{———} \\ \vdots \\ \text{———} R_n \text{———} \end{bmatrix} + \cdots + \det\begin{bmatrix} 0 \;\ldots\ldots\; 0a_{1n} \\ \text{———} R_2 \text{———} \\ \vdots \\ \text{———} R_n \text{———} \end{bmatrix}.$$

We continue expanding each of these determinants on the second row, and so on. When we are finished, $\det A$ is expressed as a sum of many terms, each of which is the determinant of a matrix M having only one entry left in each row:

$$M = \begin{bmatrix} & a_{1?} & \\ a_{2?} & & \\ & & \\ & & a_{n?} \end{bmatrix}.$$

Many of these determinants will be zero because a whole column vanishes. Thus the determinant of a 2×2 matrix is the sum of four terms:

$$\begin{aligned} \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \det\begin{bmatrix} a & 0 \\ c & d \end{bmatrix} + \det\begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \\ &= \det\begin{bmatrix} a & 0 \\ c & 0 \end{bmatrix} + \det\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} + \det\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} + \det\begin{bmatrix} 0 & b \\ 0 & d \end{bmatrix}. \end{aligned}$$

But the first and fourth terms are zero; therefore

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \det\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} + \det\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix}.$$

In fact, the matrices M having no column zero must have one entry a_{ij} left in each row and each column. They are like permutation matrices P, except that the 1's in P are replaced by the entries of A:

$$(4.10) \qquad P = \sum_j e_{p(j)j}, \quad M = \sum_j a_{p(j)j} e_{p(j)j}.$$

By linearity of the determinant (3.6),

$$\begin{aligned} \det M &= (a_{p(1)1} \cdots a_{p(n)n})(\det P) \\ &= (\operatorname{sign} p)(a_{p(1)1} \cdots a_{p(n)n}). \end{aligned}$$

There is one such term for each permutation p. This leads to the formula

$$(4.11) \qquad \det A = \sum_{\text{perm}\, p} (\operatorname{sign} p) a_{p(1)1} \cdots a_{p(n)1},$$

where the sum is over all permutations of the set $\{1,\ldots,n\}$. It seems slightly nicer to write this formula in its transposed form:

$$\text{(4.12)} \qquad \det A = \sum_{\text{perm } p} (\text{sign } p) a_{1p(1)} \cdots a_{np(n)}.$$

This is called the *complete expansion* of the determinant.

For example, the complete expansion of the determinant of a 3×3 matrix has six terms:

$$\text{(4.13)} \qquad \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$
$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

The complete expansion is more of theoretical than of practical importance, because it has too many terms to be useful for computation unless n is small. Its theoretical importance comes from the fact that determinants are exhibited as *polynomials* in the n^2 variable matrix entries a_{ij}, with coefficients ± 1. This has important consequences. Suppose, for example, that each matrix entry a_{ij} is a differentiable function of a single variable: $a_{ij} = a_{ij}(t)$. Then $\det A$ is also a differentiable function of t, because sums and products of differentiable functions are differentiable.

5. CRAMER'S RULE

The name *Cramer's Rule* is applied to a group of formulas giving solutions of systems of linear equations in terms of determinants. To derive these formulas we need to use expansion by minors on columns other than the first one, as well as on rows.

(5.1) *Expansion by minors on the jth column:*

$$\det A = (-1)^{j+1} a_{1j} \det A_{1j} + (-1)^{j+2} a_{2j} \det A_{2j} + \cdots + (-1)^{j+n} a_{nj} \det A_{nj}.$$

(5.2) *Expansion by minors on the ith row:*

$$\det A = (-1)^{i+1} a_{i1} \det A_{i1} + (-1)^{i+2} a_{i2} \det A_{i2} + \cdots + (-1)^{i+n} a_{in} \det A_{in}.$$

In these formulas A_{ij} is the matrix (3.3). The terms $(-1)^{i+j}$ provide alternating signs depending on the position (i, j) in the matrix. (I doubt that such tricky notation is really helpful, but it has become customary.) The signs can be read off of the following figure:

$$(5.3)\qquad \begin{bmatrix} + & - & + & - & \cdots \\ - & + & & & \\ + & - & \cdot & & \\ \vdots & & & \cdot & \\ & & & & \cdot \end{bmatrix}.$$

To prove (5.1), one can proceea in either of two ways:

(a) Verify properties (3.5–3.7) for (5.1) directly and apply Theorem (3.14), or
(b) Interchange (column j) with (column 1) and apply (3.9′) and (3.19).

We omit these verifications. Once (5.1) is proved, (5.2) can be derived from it by transposing the matrix and applying (3.18).

(5.4) **Definition.** Let A be an $n \times n$ matrix. The *adjoint* of A is the $n \times n$ matrix whose (i, j) entry $(\mathrm{adj})_{ij}$ is $(-1)^{i+j} \det A_{ji} = \alpha_{ji}$, where A_{ij} is the matrix obtained by crossing out the ith row and the jth column, as in (3.3):

$$(\operatorname{adj} A) = (\alpha_{ij})^{\mathrm{t}},$$

where $\alpha_{ij} = (-1)^{i+j} \det A_{ij}$. Thus

$$(5.5)\qquad \operatorname{adj}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

and

$$(5.6)\qquad \operatorname{adj}\begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 1 & -2 \\ -2 & 0 & 1 \\ -3 & -1 & 2 \end{bmatrix}^{\mathrm{t}} = \begin{bmatrix} 4 & -2 & -3 \\ 1 & 0 & -1 \\ -2 & 1 & 2 \end{bmatrix}.$$

We can now proceed to derive the formula called Cramer's Rule.

(5.7) **Theorem.** Let $\delta = \det A$. Then

$$(\operatorname{adj} A)\cdot A = \delta I, \quad \text{and} \quad A\cdot(\operatorname{adj} A) = \delta I.$$

Note that in these equations

$$\delta I = \begin{bmatrix} \delta & & \\ & \ddots & \\ & & \delta \end{bmatrix}.$$

(5.8) **Corollary.** Suppose that the determinant δ of A is not zero. Then

$$A^{-1} = \frac{1}{\delta}(\text{adj } A).$$

For example, the inverse of the 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$\frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The determinant of the 3×3 matrix whose adjoint is computed in (5.6) happens to be 1; therefore for that matrix, $A^{-1} = \text{adj } A$.

The proof of Theorem (5.7) is easy. The (i, j) entry of $(\text{adj } A) \cdot A$ is

$$(\text{adj})_{i1}a_{1j} + \cdots + (\text{adj})_{in}a_{nj} = \alpha_{1i}a_{1j} + \cdots + \alpha_{ni}a_{nj}. \tag{5.9}$$

If $i = j$, this is formula (5.1) for δ, which is the required answer. Suppose $i \neq j$. Consider the matrix B obtained by replacing (column i) by (column j) in the matrix A. So (column j) appears twice in the matrix B. Then (5.9) is expansion by minors for B on its ith column. But $\det B = 0$ by (3.7′) and (3.19). So (5.9) is zero, as required. The second equation of Theorem (5.7) is proved similarly. □

Formula (5.8) can be used to write the solution of a system of linear equations $AX = B$, where A is an $n \times n$ matrix in a compact form, provided that $\det A \neq 0$. Multiplying both sides by A^{-1}, we obtain

$$X = A^{-1}B = \frac{1}{\delta}(\text{adj } A)B, \tag{5.10}$$

where $\delta = \det A$. The product on the right can be expanded out to obtain the formula

$$x_j = \frac{1}{\delta}(b_1\alpha_{1j} + \cdots b_n\alpha_{nj}), \tag{5.11}$$

where $\alpha_{ij} = \pm\det A_{ij}$ as above.

Notice that the main term $(b_1\alpha_{1j} + \cdots + b_n\alpha_{nj})$ on the right side of (5.11) looks like the expansion of the determinant by minors on the jth column, except that b_i has replaced a_{ij}. We can incorporate this observation to get another expression for the solution of the system of equations. Let us form a new matrix M_j, replacing the jth column of A by the column vector B. Expansion by minors on the jth column shows that

$$\det M_j = (b_1\alpha_{1j} + \cdots + b_n\alpha_{nj}).$$

This gives us the tricky formula

$$x_j = \frac{\det M_j}{\det A}. \tag{5.12}$$

For some reason it is popular to write the solution of the system of equations $AX = B$ in this form, and it is often this form that is called *Cramer's Rule*. However, this expression does not simplify computation. The main thing to remember is expression (5.8) for the inverse of a matrix in terms of its adjoint; the other formulas follow from this expression.

As with the complete expansion of the determinant (4.10), formulas (5.8–5.11) have theoretical as well as practical significance, because the answers A^{-1} and X are exhibited explicitly as quotients of polynomials in the variables $\{a_{ij}, b_i\}$, with integer coefficients. If, for instance, a_{ij} and b_j are all continuous functions of t, so are the solutions x_i.

A general algebraical determinant in its developed form
may be likened to a mixture of liquids seemingly homogeneous,
but which, being of differing boiling points, admit of being separated
by the process of fractional distillation.

James Joseph Sylvester

EXERCISES

1. The Basic Operations

1. What are the entries a_{21} and a_{23} of the matrix $\begin{bmatrix} 1 & 2 & 5 \\ 2 & 7 & 8 \\ 0 & 9 & 4 \end{bmatrix}$?

2. Compute the products AB and BA for the following values of A and B.

(a) $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 3 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -8 & -4 \\ 9 & 5 \\ -3 & -2 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 4 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 6 & -4 \\ -3 & 2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $B = [1 \ \ 2 \ \ 1]$

3. Let $A = (a_1, \ldots, a_n)$ be a row vector, and let $B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ be a column vector. Compute the products AB and BA.

4. Verify the associative law for the matrix product

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

Notice that this is a self-checking problem. You have to multiply correctly, or it won't come out. If you need more practice in matrix multiplication, use this problem as a model.

5. Compute the product $\begin{bmatrix} 1 & a \\ & 1 \end{bmatrix}\begin{bmatrix} 1 & b \\ & 1 \end{bmatrix}$.

6. Compute $\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}^n$.

7. Find a formula for $\begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}^n$, and prove it by induction.

8. Compute the following matrix products by block multiplication:

$$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 5 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right]\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 3 \end{array}\right], \quad \left[\begin{array}{c|cc} 0 & 1 & 2 \\ \hline 0 & 1 & 0 \\ 3 & 0 & 1 \end{array}\right]\left[\begin{array}{c|cc} 1 & 2 & 3 \\ \hline 4 & 2 & 3 \\ 5 & 0 & 4 \end{array}\right].$$

9. Prove rule (1.20) for block multiplication.

10. Let A, B be square matrices.
(a) When is $(A + B)(A - B) = A^2 - B^2$?
(b) Expand $(A + B)^3$.

11. Let D be the diagonal matrix

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix},$$

and let $A = (a_{ij})$ be any $n \times n$ matrix.
(a) Compute the products DA and AD.
(b) Compute the product of two diagonal matrices.
(c) When is a diagonal matrix invertible?

12. An $n \times n$ matrix is called *upper triangular* if $a_{ij} = 0$ whenever $i > j$. Prove that the product of two upper triangular matrices is upper triangular.

13. In each case, find all real 2×2 matrices which commute with the given matrix.
(a) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 2 & 3 \\ 0 & 6 \end{bmatrix}$

14. Prove the properties $0 + A = A$, $0A = 0$, and $A0 = 0$ of zero matrices.

15. Prove that a matrix which has a row of zeros is not invertible.

16. A square matrix A is called *nilpotent if* $A^k = 0$ for some $k > 0$. Prove that if A is nilpotent, then $I + A$ is invertible.

17. (a) Find infinitely many matrices B such that $BA = I_2$ when

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

(b) Prove that there is no matrix C such that $AC = I_3$.

18. Write out the proof of Proposition (1.18) carefully, using the associative law to expand the product $(AB)(B^{-1}A^{-1})$.

19. The *trace* of a square matrix is the sum of its diagonal entries:

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

(a) Show that $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$, and that $\operatorname{tr} AB = \operatorname{tr} BA$.

(b) Show that if B is invertible, then $\operatorname{tr} A = \operatorname{tr} BAB^{-1}$.

20. Show that the equation $AB - BA = I$ has no solutions in $n \times n$ matrices with real entries.

2. Row Reduction

1. **(a)** For the reduction of the matrix M (2.10) given in the text, determine the elementary matrices corresponding to each operation.

(b) Compute the product P of these elementary matrices and verify that PM is indeed the end result.

2. Find all solutions of the system of equations $AX = B$ when

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

and B has the following value:

$$\textbf{(a)} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \textbf{(b)} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \textbf{(c)} \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$$

3. Find all solutions of the equation $x_1 + x_2 + 2x_3 - x_4 = 3$.

4. Determine the elementary matrices which are used in the row reduction in Example (2.22) and verify that their product is A^{-1}.

5. Find inverses of the following matrices:

$$\begin{bmatrix} 1 & \\ & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}, \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}\begin{bmatrix} & 1 \\ 1 & \end{bmatrix}\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}.$$

6. Make a sketch showing the effect of multiplication by the matrix $A = \begin{bmatrix} 2 & -1 \\ 2 & 3 \end{bmatrix}$ on the plane $\mathbb{R}^2$.

7. How much can a matrix be simplified if both row and column operations are allowed?

8. (a) Compute the matrix product $e_{ij}e_{k\ell}$.

(b) Write the identity matrix as a sum of matrix units.

(c) Let A be any $n \times n$ matrix. Compute $e_{ii}Ae_{jj}$.

(d) Compute $e_{ij}A$ and Ae_{ij}.

9. Prove rules (2.7) for the operations of elementary matrices.

10. Let A be a square matrix. Prove that there is a set of elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A$ either is the identity or has its bottom row zero.

11. Prove that every invertible 2×2 matrix is a product of at most four elementary matrices.

12. Prove that if a product AB of $n \times n$ matrices is invertible then so are the factors A, B.

13. A matrix A is called symmetric if $A = A^t$. Prove that for any matrix A, the matrix AA^t is symmetric and that if A is a square matrix then $A + A^t$ is symmetric.

14. (a) Prove that $(AB)^t = B^tA^t$ and that $A^{tt} = A$.
(b) Prove that if A is invertible then $(A^{-1})^t = (A^t)^{-1}$.

15. Prove that the inverse of an invertible symmetric matrix is also symmetric.

16. Let A and B be symmetric $n \times n$ matrices. Prove that the product AB is symmetric if and only if $AB = BA$.

17. Let A be an $n \times n$ matrix. Prove that the operator "left multiplication by A" determines A in the following sense: If $AX = BX$ for very column vector X, then $A = B$.

18. Consider an arbitrary system of linear equations $AX = B$ where A and B have real entries.
(a) Prove that if the system of equations $AX = B$ has more than one solution then it has infinitely many.
(b) Prove that if there is a solution in the complex numbers then there is also a real solution.

***19.** Prove that the reduced row echelon form obtained by row reduction of a matrix A is uniquely determined by A.

3. Determinants

1. Evaluate the following determinants:

(a) $\begin{bmatrix} 1 & i \\ 2-i & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 5 & 2 & 0 & 0 \\ 8 & 6 & 3 & 0 \\ 0 & 9 & 7 & 4 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 4 & 1 & 3 \\ 2 & 3 & 5 & 0 \\ 4 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$

2. Prove that $\det \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 1 & 7 & 7 \\ 0 & 0 & 2 & 3 \\ 4 & 2 & 1 & 5 \end{bmatrix} = -\det \begin{bmatrix} 2 & 1 & 5 & 1 \\ 1 & 3 & 7 & 0 \\ 0 & 0 & 2 & 1 \\ 2 & 4 & 1 & 4 \end{bmatrix}$.

3. Verify the rule $\det AB = (\det A)(\det B)$ for the matrices $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 5 & -2 \end{bmatrix}$. Note that this is a self-checking problem. It can be used as a model for practice in computing determinants.

4. Compute the determinant of the following $n \times n$ matrices by induction on n.

(a) $\begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & \cdot^{\cdot^{\cdot}} & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}$ (b) $\begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & \ddots & & \\ & & & & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$

5. Evaluate $\det \begin{bmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 2 & 3 & & \vdots \\ 3 & 3 & 3 & & \vdots \\ \vdots & & & \ddots & \vdots \\ n & & \cdots & & n \end{bmatrix}$

*6. Compute $\det \begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix}$.

7. Prove that the determinant is linear in the rows of a matrix, as asserted in (3.6).
8. Let A be an $n \times n$ matrix. What is $\det(-A)$?
9. Prove that $\det A^t = \det A$.
10. Derive the formula $\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$ from the properties (3.5, 3.6, 3.7, 3.9).
11. Let A and B be square matrices. Prove that $\det(AB) = \det(BA)$.
12. Prove that $\det\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = (\det A)(\det D)$, if A and D are square blocks.

*13. Let a $2n \times 2n$ matrix be given in the form $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, where each block is an $n \times n$ matrix. Suppose that A is invertible and that $AC = CA$. Prove that $\det M = \det(AD - CB)$. Give an example to show that this formula need not hold when $AC \neq CA$.

4. Permutation Matrices

1. Consider the permutation p defined by $1 \rightsquigarrow 3$, $2 \rightsquigarrow 1$, $3 \rightsquigarrow 4$, $4 \rightsquigarrow 2$.
   (a) Find the associated permutation matrix P.
   (b) Write p as a product of transpositions and evaluate the corresponding matrix product.
   (c) Compute the sign of p.
2. Prove that every permutation matrix is a product of transpositions.
3. Prove that every matrix with a single 1 in each row and a single 1 in each column, the other entries being zero, is a permutation matrix.
4. Let p be a permutation. Prove that $\operatorname{sign} p = \operatorname{sign} p^{-1}$.
5. Prove that the transpose of a permutation matrix P is its inverse.
6. What is the permutation matrix associated to the permutation $i \rightsquigarrow n-i$?
7. (a) The complete expansion for the determinant of a 3×3 matrix consists of six triple products of matrix entries, with sign. Learn which they are.
   (b) Compute the determinant of the following matrices using the complete expansion, and check your work by another method:

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 4 & -1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 1 \end{bmatrix}, \begin{bmatrix} a & b & c \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

8. Prove that the complete expansion (4.12) defines the determinant by verifying rules (3.5–3.7).
9. Prove that formulas (4.11) and (4.12) define the same number.

5. Cramer's Rule

1. Let $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a matrix with determinant 1. What is A^{-1}?
2. (self-checking) Compute the adjoints of the matrices $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix}$, $\begin{bmatrix} 4 & -1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 1 \end{bmatrix}$, and $\begin{bmatrix} a & b & c \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, and verify Theorem (5.7) for them.
3. Let A be an $n \times n$ matrix with integer entries a_{ij}. Prove that A^{-1} has integer entries if and only if $\det A = \pm 1$.
4. Prove that expansion by minors on a row of a matrix defines the determinant function.

Miscellaneous Problems

1. Write the matrix $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ as a product of elementary matrices, using as few as you can. Prove that your expression is as short as possible.
2. Find a representation of the complex numbers by real 2×2 matrices which is compatible with addition and multiplication. Begin by finding a nice solution to the matrix equation $A^2 = -I$.
3. (*Vandermonde determinant*) **(a)** Prove that $\det \begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix} = (b-a)(c-a)(c-b)$.

   ***(b)** Prove an analogous formula for $n \times n$ matrices by using row operations to clear out the first column cleverly.

*4. Consider a general system $AX = B$ of m linear equations in n unknowns. If the coefficient matrix A has a left inverse A', a matrix such that $A'A = I_n$, then we may try to solve the system as follows:

$$\begin{aligned} AX &= B \\ A'AX &= A'B \\ X &= A'B. \end{aligned}$$

But when we try to check our work by running the solution backward, we get into trouble:

$$\begin{aligned} X &= A'B \\ AX &= AA'B \\ AX &\overset{?}{=} B. \end{aligned}$$

We seem to want A' to be a right inverse: $AA' = I_m$, which isn't what was given. Explain. (Hint: Work out some examples.)

5. **(a)** Let A be a real 2×2 matrix, and let A_1, A_2 be the rows of A. Let P be the parallelogram whose vertices are $0, A_1, A_2, A_1 + A_2$. Prove that the area of P is the absolute value of the determinant $\det A$ by comparing the effect of an elementary row operation on the area and on $\det A$.

***(b)** Prove an analogous result for $n \times n$ matrices.

***6.** Most invertible matrices can be written as a product $A = LU$ of a lower triangular matrix L and an upper triangular matrix U, where in addition all diagonal entries of U are 1.

(a) Prove uniqueness, that is, prove that there is at most one way to write A as a product.

(b) Explain how to compute L and U when the matrix A is given.

(c) Show that every invertible matrix can be written as a product LPU, where L, U are as above and P is a permutation matrix.

7. Consider a system of n linear equations in n unknowns: $AX = B$, where A and B have *integer* entries. Prove or disprove the following.

(a) The system has a rational solution if $\det A \neq 0$.

(b) If the system has a rational solution, then it also has an integer solution.

***8.** Let A, B be $m \times n$ and $n \times m$ matrices. Prove that $I_m - AB$ is invertible if and only if $I_n - BA$ is invertible.