CHAPTER II

Matrices and Linear Equations

You have met linear equations in elementary school. Linear equations are simply equations like

$$2x + y + z = 1,$$
$$5x - y + 7z = 0.$$

You have learned to solve such equations by the successive elimination of the variables. In this chapter, we shall review the theory of such equations, dealing with equations in n variables, and interpreting our results from the point of view of vectors. Several geometric interpretations for the solutions of the equations will be given.

The first chapter is used here very little, and can be entirely omitted if you know only the definition of the dot product between two n-tuples. The multiplication of matrices will be formulated in terms of such a product. One geometric interpretation for the solutions of homogeneous equations will however rely on the fact that the dot product between two vectors is 0 if and only if the vectors are perpendicular, so if you are interested in this interpretation, you should refer to the section in Chapter I where this is explained.

II, §1. Matrices

We consider a new kind of object, matrices.

Let n, m be two integers $\geqq 1$. An array of numbers

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

is called a **matrix**. We can abbreviate the notation for this matrix by writing it (a_{ij}), $i = 1,\ldots,m$ and $j = 1,\ldots,n$. We say that it is an m by n matrix, or an $m \times n$ matrix. The matrix has m **rows** and n **columns**. For instance, the first column is

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

and the second row is $(a_{21}, a_{22},\ldots,a_{2n})$. We call a_{ij} the ***ij*-entry** or ***ij*-component** of the matrix.

Look back at Chapter I, §1. The example of 7-space taken from economics gives rise to a 7×7 matrix (a_{ij}) $(i, j = 1,\ldots,7)$, if we define a_{ij} to be the amount spent by the i-th industry on the j-th industry. Thus keeping the notation of that example, if $a_{25} = 50$, this means that the auto industry bought 50 million dollars worth of stuff from the chemical industry during the given year.

Example 1. The following is a 2×3 matrix:

$$\begin{pmatrix} 1 & 1 & -2 \\ -1 & 4 & -5 \end{pmatrix}.$$

It has two rows and three columns.

The rows are $(1, 1, -2)$ and $(-1, 4, -5)$. The columns are

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ -5 \end{pmatrix}.$$

Thus the rows of a matrix may be viewed as n-tuples, and the columns may be viewed as vertical m-tuples. A vertical m-tuple is also called a **column vector.**

A vector $(x_1,\dots,x_n)$ is a $1 \times n$ matrix. A column vector

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

is an $n \times 1$ matrix.

When we write a matrix in the form (a_{ij}), then i denotes the row and j denotes the column. In Example 1, we have for instance

$$a_{11} = 1,\ a_{23} = -5.$$

A single number (a) may be viewed as a 1×1 matrix.

Let (a_{ij}), $i = 1,\dots,m$ and $j = 1,\dots,n$ be a matrix. If $m = n$, then we say that it is a **square** matrix. Thus

$$\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -1 & 5 \\ 2 & 1 & -1 \\ 3 & 1 & -1 \end{pmatrix}$$

are both square matrices.

We define the **zero matrix** to be the matrix such that $a_{ij} = 0$ for all i, j. It looks like this:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We shall write it O. We note that we have met so far with the zero number, zero vector, and zero matrix.

We shall now define addition of matrices and multiplication of matrices by numbers.

We define addition of matrices only when they have the same size. Thus let m, n be fixed integers $\geqq 1$. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. We define $A + B$ to be the matrix whose entry in the i-th row and j-th column is $a_{ij} + b_{ij}$. In other words, we add matrices of the same size componentwise.

Example 2. Let

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 1 & -1 \\ 2 & 1 & -1 \end{pmatrix}.$$

Then

$$A + B = \begin{pmatrix} 6 & 0 & -1 \\ 4 & 4 & 3 \end{pmatrix}.$$

If A, B are both $1 \times n$ matrices, i.e. n-tuples, then we note that our addition of matrices coincides with the addition which we defined in Chapter I for n-tuples.

If O is the zero matrix, then for any matrix A (of the same size, of course), we have $O + A = A + O = A$.

This is trivially verified. We shall now define the multiplication of a matrix by a number. Let c be a number, and $A = (a_{ij})$ be a matrix. We define cA to be the matrix whose ij-component is ca_{ij}. We write

$$cA = (ca_{ij}).$$

Thus we multiply each component of A by c.

Example 3. Let A, B be as in Example 2. Let $c = 2$. Then

$$2A = \begin{pmatrix} 2 & -2 & 0 \\ 4 & 6 & 8 \end{pmatrix} \quad \text{and} \quad 2B = \begin{pmatrix} 10 & 2 & -2 \\ 4 & 2 & -2 \end{pmatrix}.$$

We also have

$$(-1)A = -A = \begin{pmatrix} -1 & 1 & 0 \\ -2 & -3 & -4 \end{pmatrix}.$$

In general, for any matrix $A = (a_{ij})$ we let $-A$ (minus A) be the matrix $(-a_{ij})$. Since we have the relation $a_{ij} - a_{ij} = 0$ for numbers, we also get the relation

$$A + (-A) = O$$

for matrices. The matrix $-A$ is also called the **additive inverse** of A.

We define one more notion related to a matrix. Let $A = (a_{ij})$ be an $m \times n$ matrix. The $n \times m$ matrix $B = (b_{ji})$ such that $b_{ji} = a_{ij}$ is called the **transpose** of A, and is also denoted by ${}^tA$. Taking the transpose of a matrix amounts to changing rows into columns and vice versa. If A is the matrix which we wrote down at the beginning of this section, then ${}^tA$ is the matrix

$$\begin{pmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{pmatrix}.$$

To take a special case:

$$\text{If } A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 5 \end{pmatrix}, \qquad \text{then} \qquad {}^tA = \begin{pmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 5 \end{pmatrix}.$$

If $A = (2, 1, -4)$ is a *row vector*, then

$${}^tA = \begin{pmatrix} 2 \\ 1 \\ -4 \end{pmatrix}$$

is a *column vector.*

A matrix A which is equal to its transpose, that is $A = {}^tA$, is called **symmetric**. Such a matrix is necessarily a square matrix.

Remark on notation. I have written the transpose sign on the left, because in many situations one considers the inverse of a matrix written A^{-1}, and then it is easier to write ${}^tA^{-1}$ rather than $(A^{-1})^t$ or $(A^t)^{-1}$, which are in fact equal. The mathematical community has no consensus as to where the transpose sign should be placed, on the right or left.

Exercises II, §1

1. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -1 & 5 & -2 \\ 1 & 1 & -1 \end{pmatrix}.$$

Find $A + B$, $3B$, $-2B$, $A + 2B$, $2A + B$, $A - B$, $A - 2B$, $B - A$.

2. Let

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -1 & 1 \\ 0 & -3 \end{pmatrix}.$$

Find $A + B$, $3B$, $-2B$, $A + 2B$, $A - B$, $B - A$.

3. (a) Write down the row vectors and column vectors of the matrices A, B in Exercise 1.
   (b) Write down the row vectors and column vectors of the matrices A, B in Exercise 2.

4. (a) In Exercise 1, find ${}^tA$ and ${}^tB$.
   (b) In Exercise 2, find ${}^tA$ and ${}^tB$.

5. If A, B are arbitrary $m \times n$ matrices, show that

$${}^t(A + B) = {}^tA + {}^tB.$$

6. If c is a number, show that ${}^t(cA) = c\,{}^tA$.

7. If $A = (a_{ij})$ is a square matrix, then the elements a_{ii} are called the **diagonal** elements. How do the diagonal elements of A and ${}^tA$ differ?

8. Find ${}^t(A + B)$ and ${}^tA + {}^tB$ in Exercise 2.

9. Find $A + {}^tA$ and $B + {}^tB$ in Exercise 2.

10. (a) Show that for any square matrix, the matrix $A + {}^tA$ is symmetric.
    (b) Define a matrix A to be **skew-symmetric** if ${}^tA = -A$. Show that for any square matrix A, the matrix $A - {}^tA$ is skew-symmetric.
    (c) If a matrix is skew-symmetric, what can you say about its diagonal elements?

11. Let

$$E_1 = (1, 0, \dots, 0), \quad E_2 = (0, 1, 0, \dots, 0), \quad \dots, \quad E_n = (0, \dots, 0, 1)$$

be the standard unit vectors of $\mathbf{R}^n$. Let $x_1, \dots, x_n$ be numbers. What is $x_1E_1 + \cdots + x_nE_n$? Show that if

$$x_1E_1 + \cdots + x_nE_n = O$$

then $x_i = 0$ for all i.

II, §2. Multiplication of Matrices

We shall now define the product of matrices. Let $A = (a_{ij})$, $i = 1, \dots, m$ and $j = 1, \dots, n$ be an $m \times n$ matrix. Let $B = (b_{jk})$, $j = 1, \dots, n$ and let $k = 1, \dots, s$ be an $n \times s$ matrix:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{ns} \end{pmatrix}.$$

We define the **product** AB to be the $m \times s$ matrix whose ik-coordinate is

$$\sum_{j=1}^{n} a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}.$$

If $A_1, \dots, A_m$ are the row vectors of the matrix A, and if $B^1, \dots, B^s$ are the column vectors of the matrix B, then the ik-coordinate of the product AB is equal to $A_i \cdot B^k$. Thus

$$AB = \begin{pmatrix} A_1 \cdot B^1 & \cdots & A_1 \cdot B^s \\ \vdots & & \vdots \\ A_m \cdot B^1 & \cdots & A_m \cdot B^s \end{pmatrix}.$$

Multiplication of matrices is therefore a generalization of the dot product.

Example. Let

$$A = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Then AB is a 2×2 matrix, and computations show that

$$AB = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 15 & 15 \\ 4 & 12 \end{pmatrix}.$$

Example. Let

$$C = \begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix}.$$

Let A, B be as in Example 1. Then

$$BC = \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 5 \\ -3 & -5 \\ 1 & 5 \end{pmatrix}$$

and

$$A(BC) = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} -1 & 5 \\ -3 & -5 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 30 \\ -8 & 0 \end{pmatrix}.$$

Compute $(AB)C$. What do you find?

If $X = (x_1, \dots, x_m)$ is a row vector, i.e. a $1 \times m$ matrix, then we can form the product XA, which looks like this:

$$(x_1, \dots, x_m) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (y_1, \dots, y_n),$$

where

$$y_k = x_1 a_{1k} + \cdots + x_m a_{mk}.$$

In this case, XA is a $1 \times n$ matrix, i.e. a row vector.

On the other hand, if X is a column vector,

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then $AX = Y$ where Y is also a column vector, whose coordinates are given by

$$y_i = \sum_{j=1}^{n} a_{ij}x_j = a_{i1}x_1 + \cdots + a_{in}x_n.$$

Visually, the multiplication $AX = Y$ looks like

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Example. Linear equations. Matrices give a convenient way of writing linear equations. You should already have considered systems of linear equations. For instance, one equation like:

$$3x - 2y + 3z = 1,$$

with three unknowns x, y, z. Or a system of two equations in three unknowns

$$(*) \qquad \begin{aligned} 3x - 2y + 3z &= 1, \\ -x + 7y - 4z &= -5. \end{aligned}$$

In this example we let the **matrix of coefficients** be

$$A = \begin{pmatrix} 3 & -2 & 3 \\ -1 & 7 & -4 \end{pmatrix}.$$

Let B be the column vector of the numbers appearing on the right-hand side, so

$$B = \begin{pmatrix} 1 \\ -5 \end{pmatrix}.$$

Let the vector of unknowns be the column vector.

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Then you can see that the system of two simultaneous equations can be written in the form

$$AX = B.$$

Example. The first equation of (*) represents equality of the first component of AX and B; whereas the second equation of (*) represents equality of the second component of AX and B.

In general, let $A = (a_{ij})$ be an $m \times n$ matrix, and let B be a column vector of size m. Let

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

be a column vector of size n. Then the system of linear equations

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

can be written in the more efficient way

$$\boxed{AX = B,}$$

by the definition of multiplication of matrices. We shall see later how to solve such systems. We say that there are m **equations** and n **unknowns**, or n **variables**.

Example. Markov matrices. A matrix can often be used to represent a practical situation. Suppose we deal with three cities, say Los Angeles, Chicago, and Boston, denoted by LA, Ch, and Bo. Suppose that any given year, some people leave each one of these cities to go to one of the others. The peıcentages of people leaving and going is given as follows, for each year.

$\frac{1}{4}$ LA goes to Bo	and	$\frac{1}{7}$ LA goes to Ch.
$\frac{1}{5}$ Ch goes to LA	and	$\frac{1}{3}$ Ch goes to Bo.
$\frac{1}{6}$ Bo goes to LA	and	$\frac{1}{8}$ Bo goes to Ch.

Let x_n, y_n, z_n be the populations of LA, Ch, and Bo, respectively, in the n-th year. Then we can express the population in the $(n+1)$-th year as follows.

In the $(n+1)$-th year, $\frac{1}{4}$ of the LA population leaves for Boston, and $\frac{1}{7}$ leaves for Chicago. The total fraction leaving LA during the year is therefore

$$\tfrac{1}{4} + \tfrac{1}{7} = \tfrac{11}{28}.$$

Hence the total fraction remaining in LA is

$$1 - \tfrac{11}{28} = \tfrac{17}{28}.$$

Hence the population in LA for the $(n+1)$-th year is

$$x_{n+1} = \tfrac{17}{28}x_n + \tfrac{1}{5}y_n + \tfrac{1}{6}z_n.$$

Similarly the fraction leaving Chicago each year is

$$\tfrac{1}{5} + \tfrac{1}{3} = \tfrac{8}{15},$$

so the fraction remaining is $\frac{7}{15}$. Finally, the fraction leaving Boston each year is

$$\tfrac{1}{6} + \tfrac{1}{8} = \tfrac{7}{24},$$

so the fraction remaining in Boston is $\frac{17}{24}$. Thus

$$\begin{aligned} y_{n+1} &= \tfrac{1}{7}x_n + \tfrac{7}{15}y_n + \tfrac{1}{8}z_n, \\ z_{n+1} &= \tfrac{1}{4}x_n + \tfrac{1}{3}\,y_n + \tfrac{17}{24}z_n. \end{aligned}$$

Let A be the matrix

$$A = \begin{pmatrix} \frac{17}{28} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{7} & \frac{7}{15} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{3} & \frac{17}{24} \end{pmatrix}.$$

Then we can write down more simply the population shift by the expression

$$X_{n+1} = AX_n \qquad \text{where} \qquad X_n = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}.$$

The change from X_n to X_{n+1} is called a **Markov process**. This is due to the special property of the matrix A, all of whose components are $\geqq 0$, and such that the sum of all the elements in each column is equal to 1. Such a matrix is called a **Markov matrix**.

If A is a square matrix, then we can form the product AA, which will be a square matrix of the same size as A. It is denoted by A^2. Similarly, we can form A^3, A^4, and in general, A^n for any positive integer n. Thus A^n is the product of A with itself n times.

We can define the **unit** $n \times n$ matrix to be the matrix having diagonal components all equal to 1, and all other components equal to 0. Thus the unit $n \times n$ matrix, denoted by I_n, looks like this:

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

We can then define $A^0 = I$ (the unit matrix of the same size as A). Note that for any two integers r, $s \geqq 0$ we have the usual relation

$$A^r A^s = A^s A^r = A^{r+s}.$$

For example, in the Markov process described above, we may express the population vector in the $(n+1)$-th year as

$$X_{n+1} = A^n X_1,$$

where X_1 is the population vector in the first year.

Warning. It is **not always true** that $AB = BA$. For instance, compute AB and BA in the following cases:

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

You will find two different values. This is expressed by saying that multiplication of matrices is not necessarily commutative. Of course, in some *special* cases, we do have $AB = BA$. For instance, powers of A commute, i.e. we have $A^r A^s = A^s A^r$ as already pointed out above.

We now prove other basic properties of multiplication.

Distributive law. *Let A, B, C be matrices. Assume that A, B can be multiplied, and A, C can be multiplied, and B, C can be added. Then A, $B + C$ can be multiplied, and we have*

$$A(B + C) = AB + AC.$$

If x is a number, then

$$A(xB) = x(AB).$$

Proof. Let A_i be the i-th row of A and let B^k, C^k be the k-th column of B and C, respectively.... Then $B^k + C^k$ is the k-th column of $B + C$. By definition, the ik-component of $A(B + C)$ is $A_i \cdot (B^k + C^k)$. Since

$$A_i \cdot (B^k + C^k) = A_i \cdot B^k + A_i \cdot C^k,$$

our first assertion follows. As for the second, observe that the k-th column of xB is xB^k. Since

$$A_i \cdot xB^k = x(A_i \cdot B^k),$$

our second assertion follows.

Associative law. *Let A, B, C be matrices such that A, B can be multiplied and B, C can be multiplied. Then A, BC can be multiplied. So can AB, C, and we have*

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

Proof. Let $A = (a_{ij})$ be an $m \times n$ matrix, let $B = (b_{jk})$ be an $n \times r$ matrix, and let $C = (c_{kl})$ be an $r \times s$ matrix. The product AB is an $m \times r$ matrix, whose ik-component is equal to the sum

$$a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}.$$

We shall abbreviate this sum using our $\sum$ notation by writing

$$\sum_{j=1}^{n} a_{ij}b_{jk}.$$

By definition, the il-component of $(AB)C$ is equal to

$$\sum_{k=1}^{r}\left[\sum_{j=1}^{n} a_{ij}b_{jk}\right]c_{kl} = \sum_{k=1}^{r}\left[\sum_{j=1}^{n} a_{ij}b_{jk}c_{kl}\right].$$

The sum on the right can also be described as the sum of all terms

$$\sum a_{ij}b_{jk}c_{kl},$$

where j, k range over all integers $1 \leqq j \leqq n$ and $1 \leqq k \leqq r$, respectively.

If we had started with the jl-component of BC and then computed the il-component of $A(BC)$ we would have found exactly the same sum, thereby proving the desired property.

The above properties are very similar to those of multiplication of numbers, except that the commutative law does not hold.

We can also relate multiplication with the transpose:

Let A, B be matrices of a size such that AB is defined. Then

$${}^t(AB) = {}^tB\,{}^tA.$$

In other words, the transpose of the product is equal to the product of the transpose in reverse order.

Proof. Let $A = (a_{ij})$ and $B = (b_{jk})$. Then $AB = C = (c_{ik})$ where

$$\begin{aligned} c_{ik} &= a_{i1}b_{1k} + \cdots + a_{in}b_{nk} \\ &= b_{1k}a_{i1} + \cdots + b_{nk}a_{in}. \end{aligned}$$

Let ${}^tA = (a'_{ji})$, ${}^tB = (b'_{kj})$, and ${}^tC = (c'_{ki})$. Then

$$a'_{ji} = a_{ij}, \qquad b'_{kj} = b_{jk}, \qquad c'_{ki} = c_{ik}.$$

Hence we can reread the above relation as

$$c'_{ki} = b'_{k1}a'_{1i} + \cdots + b'_{kn}a'_{ni},$$

which shows that ${}^tC = {}^tB\,{}^tA$, as desired.

Example. Instead of writing the system of linear equations $AX = B$ in terms of column vectors, we can write it by taking the transpose, which gives

$${}^tX\,{}^tA = {}^tB.$$

If X, B are column vectors, then ${}^tX$, ${}^tB$ are row vectors. It is occasionally convenient to rewrite the system in this fashion.

Unlike division with non-zero numbers, **we cannot divide by a matrix**, any more than we could divide by a vector (n-tuple). Under certain circumstances, we can define an inverse as follows. We do this only for square matrices. Let A be an $n \times n$ matrix. An **inverse for** A is a matrix B such that

$$AB = BA = I.$$

Since we multiplied A with B on both sides, the only way this can make sense is if B is also an $n \times n$ matrix. Some matrices do not have inverses. However, **if an inverse exists, then there is only one** (we say that **the inverse is unique, or uniquely determined by** A). This is easy to prove. Suppose that B, C are inverses, so we have

$$AB = BA = I \qquad \text{and} \qquad AC = CA = I.$$

Multiply the equation $BA = I$ on the right with C. Then

$$BAC = IC = C$$

and we have assumed that $AC = I$, so $BAC = BI = B$. This proves that $B = C$. In light of this, the inverse is denoted by

$$A^{-1}.$$

Then A^{-1} is the unique matrix such that

$$\boxed{A^{-1}A = I \qquad \text{and} \qquad AA^{-1} = I.}$$

We shall prove later that if A, B are square matrices of the same size such that $AB = I$ then it follows that also

$$BA = I.$$

In other words, if B is a right inverse for A, then it is also a left inverse. You may assume this for the time being. Thus in verifying that a matrix is the inverse of another, you need only do so on one side.

We shall also find later a way of computing the inverse when it exists. It can be a tedious matter.

Let c be a number. Then the matrix

$$cI = \begin{pmatrix} c & 0 & \cdots & \cdots & 0 \\ 0 & c & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & c \end{pmatrix}$$

having component c on each diagonal entry and 0 otherwise is called a **scalar matrix**. We can also write it as cI, where I is the unit $n \times n$ matrix. Cf. Exercise 6.

As an application of the formula for the transpose of a product, we shall now see that:

The transpose of an inverse is the inverse of the transpose, that is

$${}^t(A^{-1}) = ({}^tA)^{-1}.$$

Proof. Take the transpose of the relation $AA^{-1} = I$. Then by the rule for the transpose of a product, we get

$${}^t(A^{-1})\,{}^tA = {}^tI = I$$

because I is equal to its own transpose. Similarly, applying the transpose to the relation $A^{-1}A = I$ yields

$${}^tA\,{}^t(A^{-1}) = {}^tI = I.$$

Hence ${}^t(A^{-1})$ is an inverse for ${}^tA$, as was to be shown.

In light of this result, it is customary to omit the parentheses, and to write

$${}^tA^{-1}$$

for the inverse of the transpose, which we have seen is equal to the transpose of the inverse.

We end this section with an important example of multiplication of matrices.

Example. Rotations. A special type of 2×2 matrix represents rotations. For each number θ, let $R(\theta)$ be the matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Let $X = \begin{pmatrix} x \\ y \end{pmatrix}$ be a point on the unit circle. We may write its coordinates x, y in the form

$$x = \cos\varphi, \qquad y = \sin\varphi$$

for some number φ. Then we get, by matrix multiplication:

$$R(\theta)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\theta+\varphi) \\ \sin(\theta+\varphi) \end{pmatrix}.$$

This follows from the **addition formula for sine and cosine**, namely

$$\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi,$$
$$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi.$$

An arbitrary point in $\mathbf{R}^2$ can be written in the form

$$rX = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix},$$

where r is a number $\geqq 0$. Since

$$R(\theta)rX = rR(\theta)X,$$

we see that multiplication by $R(\theta)$ also has the effect of rotating rX by an angle θ. Thus rotation by an angle θ can be represented by the matrix $R(\theta)$.

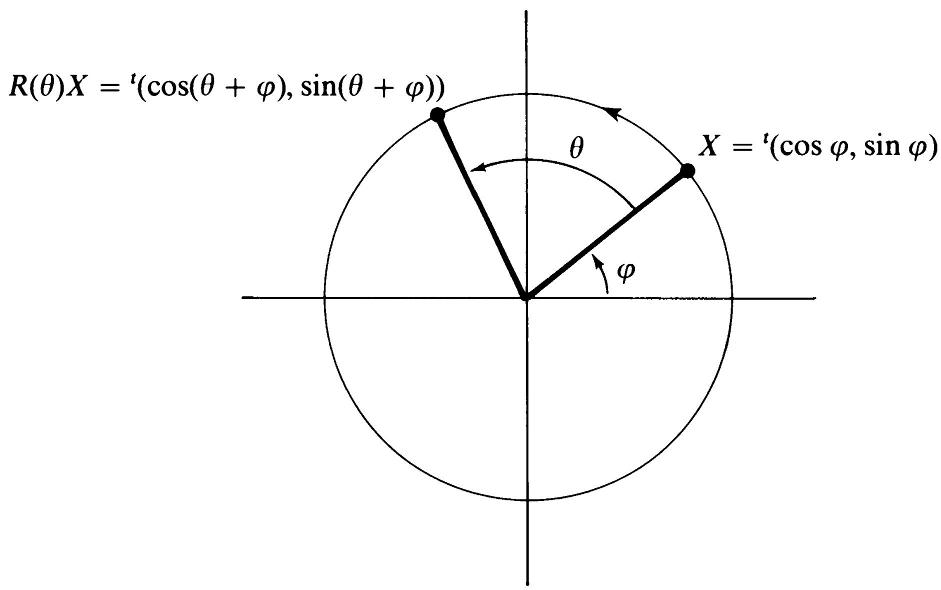

Figure 1

Note that for typographical reasons, we have written the vector ${}^tX$ horizontally, but have put a little t on the upper left superscript, to denote transpose, so X is a column vector.

Example. The matrix corresponding to rotation by an angle of $\pi/3$ is given by

$$R(\pi/3) = \begin{pmatrix} \cos \pi/3 & -\sin \pi/3 \\ \sin \pi/3 & \cos \pi/3 \end{pmatrix}$$

$$= \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}.$$

Example. Let $X = {}^t(2, 5)$. If you rotate X by an angle of $\pi/3$, find the coordinates of the rotated vector.

These coordinates are:

$$R(\pi/3)X = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

$$= \begin{pmatrix} 1 - 5\sqrt{3}/2 \\ \sqrt{3} + 5/2 \end{pmatrix}.$$

Warning. Note how we multiply the column vector on the left with the matrix $R(\theta)$. If you want to work with row vectors, then take the transpose and verify directly that

$$(2, 5)\begin{pmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{pmatrix} = (1 - 5\sqrt{3}/2, \sqrt{3} + 5/2).$$

So the matrix $R(\theta)$ gets transposed. The minus sign is now in the lower left-hand corner.

Exercises II, §2

The following exercises give mostly routine practice in the multiplication of matrices. However, they also illustrate some more theoretical aspects of this multiplication. Therefore they should be all worked out. Specifically:

Exercises 7 through 12 illustrate multiplication by the standard unit vectors.

Exercises 14 through 19 illustrate multiplication of triangular matrices.

Exercises 24 through 27 illustrate how addition of numbers is transformed into multiplication of matrices.

Exercises 27 through 32 illustrate rotations.

Exercises 33 through 37 illustrate elementary matrices, *and should be worked out before studying* §5.

1. Let I be the unit $n \times n$ matrix. Let A be an $n \times r$ matrix. What is IA? If A is an $m \times n$ matrix, what is AI?

2. Let O be the matrix all of whose coordinates are 0. Let A be a matrix of a size such that the product AO is defined. What is AO?

3. In each one of the following cases, find $(AB)C$ and $A(BC)$.

(a) $A = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}$, $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$

(b) $A = \begin{pmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -1 \end{pmatrix}$, $C = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

(c) $A = \begin{pmatrix} 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & -1 \\ 3 & 1 & 5 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ -1 & 4 \end{pmatrix}$

4. Let A, B be square matrices of the same size, and assume that $AB = BA$. Show that

$$(A + B)^2 = A^2 + 2AB + B^2, \qquad \text{and} \qquad (A + B)(A - B) = A^2 - B^2,$$

using the distributive law.

5. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}.$$

Find AB and BA.

6. Let

$$C = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}.$$

Let A, B be as in Exercise 5. Find CA, AC, CB, and BC. State the general rule including this exercise as a special case.

7. Let $X = (1, 0, 0)$ and let

$$A = \begin{pmatrix} 3 & 1 & 5 \\ 2 & 0 & 1 \\ 1 & 1 & 7 \end{pmatrix}.$$

What is XA?

8. Let $X = (0, 1, 0)$, and let A be an arbitrary 3×3 matrix. How would you describe XA? What if $X = (0, 0, 1)$? Generalize to similar statements concerning $n \times n$ matrices, and their products with unit vectors.

9. Let

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

Find AX for each of the following values of X.

(a) $X = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ (b) $X = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ (c) $X = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

10. Let

$$A = \begin{pmatrix} 3 & 7 & 5 \\ 1 & -1 & 4 \\ 2 & 1 & 8 \end{pmatrix}.$$

Find AX for each of the values of X given in Exercise 9.

11. Let

$$X = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} a_{11} & \cdots & a_{14} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{m4} \end{pmatrix}.$$

What is AX?

12. Let X be a column vector having all its components equal to 0 except the j-th component which is equal to 1. Let A be an arbitrary matrix, whose size is such that we can form the product AX. What is AX?

13. Let X be the indicated column vector, and A the indicated matrix. Find AX as a column vector.

(a) $X = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$, $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 0 & -1 \end{pmatrix}$ (b) $X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $A = \begin{pmatrix} 2 & 1 & 5 \\ 0 & 1 & 1 \end{pmatrix}$

(c) $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (d) $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

14. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Find the product AS for each one of the following matrices S. Describe in words the effect on A of this product.

(a) $S = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ (b) $S = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}$.

15. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ again. Find the product SA for each one of the following matrices S. Describe in words the effect of this product on A.

(a) $S = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ (b) $S = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}$.

16. (a) Let A be the matrix

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Find A^2, A^3. Generalize to 4×4 matrices.

(b) Let A be the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Compute A^2, A^3, A^4.

17. Let

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

Find A^2, A^3, A^4.

18. Let A be a diagonal matrix, with diagonal elements $a_1, \ldots, a_n$. What is A^2, A^3, A^k for any positive integer k?

19. Let

$$A = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}.$$

Find A^3

20. (a) Find a 2×2 matrix A such that $A^2 = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.
(b) Determine all 2×2 matrices A such that $A^2 = O$.

21. Let A be a square matrix.
(a) If $A^2 = O$ show that $I - A$ is invertible.
(b) If $A^3 = O$, show that $I - A$ is invertible.
(c) In general, if $A^n = O$ for some positive integer n, show that $I - A$ is invertible. [*Hint*: Think of the geometric series.]
(d) Suppose that $A^2 + 2A + I = O$. Show that A is invertible.
(e) Suppose that $A^3 - A + I = O$. Show that A is invertible.

22. Let A, B be two square matrices of the same size. We say that A is **similar** to B if there exists an invertible matrix T such that $B = TAT^{-1}$. Suppose this is the case. Prove:
(a) B is similar to A.
(b) A is invertible if and only if B is invertible.
(c) ${}^tA$ is similar to ${}^tB$.
(d) Suppose $A^n = O$ and B is an invertible matrix of the same size as A. Show that $(BAB^{-1})^n = O$.

23. Let A be a square matrix which is of the form

$$\begin{pmatrix} a_{11} & * & \cdots\cdots & * \\ 0 & a_{22} & * \cdots & * \\ \vdots & & \ddots & \vdots \\ & & & * \\ 0 & \cdots\cdots & 0 & a_{nn} \end{pmatrix}.$$

The notation means that all elements below the diagonal are equal to 0, and the elements above the diagonal are arbitrary. One may express this property by saying that

$$a_{ij} = 0 \quad \text{if} \quad i > j.$$

Such a matrix is called **upper triangular**. If A, B are upper triangular matrices (of the same size) what can you say about the diagonal elements of AB?

Exercises 24 through 27 give examples where addition of numbers is transformed into multiplication of matrices.

24. Let a, b be numbers, and let

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

What is AB? What is A^2, A^3? What is A^n where n is a positive integer?

25. Show that the matrix A in Exercise 24 has an inverse. What is this inverse?

26. Show that if A, B are $n \times n$ matrices which have inverses, then AB has an inverse.

27. **Rotations.** Let $R(\theta)$ be the matrix given by

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

(a) Show that for any two numbers θ_1, θ_2 we have

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2).$$

[You will have to use the addition formulas for sine and cosine.]
(b) Show that the matrix $R(\theta)$ has an inverse, and write down this inverse.
(c) Let $A = R(\theta)$. Show that

$$A^2 = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}.$$

(d) Determine A^n for any positive integer n. Use induction.

28. Find the matrix $R(\theta)$ associated with the rotation for each of the following values of θ.
(a) $\pi/2$ (b) $\pi/4$ (c) π (d) $-\pi$ (e) $-\pi/3$
(f) $\pi/6$ (g) $5\pi/4$

29. In general, let $\theta > 0$. What is the matrix associated with the rotation by an angle $-\theta$ (i.e. clockwise rotation by θ)?

30. Let $X = {}^t(1, 2)$ be a point of the plane. If you rotate X by an angle of $\pi/4$, what are the coordinates of the new point?

31. Same question when $X = {}^t(-1, 3)$ and the rotation is by an angle of $\pi/2$.

32. For any vector X in $\mathbf{R}^2$ let $Y = R(\theta)X$ be its rotation by an angle θ. Show that $\|Y\| = \|X\|$.

The following exercises on elementary matrices should be done before studying §5.

33. **Elementary matrices.** Let

$$A = \begin{pmatrix} 2 & 3 & -1 & 1 \\ 1 & 4 & 2 & -2 \\ -1 & 1 & 3 & -5 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

Let U be the matrix as shown. In each case find UA.

(a) $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ (d) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

(e) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ (f) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

34. Let E be the matrix as shown. Find EA where A is the same matrix as in the preceding exercise.

(a) $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

35. Let E be the matrix as shown. Find EA where A is the same matrix as in the preceding exercise and Exercise 33.

(a) $\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

36. Let $A = (a_{ij})$ be an $m \times n$ matrix,

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Let $1 \leqq r \leqq m$ and $1 \leqq s \leqq m$. Let I_{rs} be the matrix whose rs-component is 1 and such that all other components are equal to 0.

(a) What is $I_{rs}A$?

(b) Suppose $r \neq s$. What is $(I_{rs} + I_{sr})A$?

(c) Suppose $r \neq s$. Let I_{jj} be the matrix whose jj-component is 1 and such that all other components are 0. Let

$$E_{rs} = I_{rs} + I_{sr} + \text{sum of all } I_{jj} \text{ for } \quad j \neq r, \quad j \neq s.$$

What is $E_{rs}A$?

37. Again let $r \neq s$.

(a) Let $E = I + 3I_{rs}$. What is EA?

(b) Let c be any number. Let $E = I + cI_{rs}$. What is EA?

The rest of the chapter will be mostly concerned with linear equations, and especially homogeneous ones. We shall find three ways of interpreting such equations, illustrating three dfferent ways of thinking about matrices and vectors.

II, §3. Homogeneous Linear Equations and Elimination

In this section, we look at linear equations by one method of elimination. In the next section, we shall discuss another method.

We shall be interested in the case when the number of unknowns is greater than the number of equations, and we shall see that in that case, there always exists a non-trivial solution.

Before dealing with the general case, we shall study examples.

Example 1. Suppose that we have a single equation, like

$$2x + y - 4z = 0.$$

We wish to find a solution with not all of x, y, z equal to 0. An equivalent equation is

$$2x = -y + 4z.$$

To find a non-trivial solution, we give all the variables except the first a special value $\neq 0$, say $y = 1$, $z = 1$. We than solve for x. We find

$$2x = -y + 4z = 3,$$

whence $x = \frac{3}{2}$.

Example 2. Consider a pair of equations, say

$$2x + 3y - z = 0, \tag{1}$$

$$x + y + z = 0. \tag{2}$$

We reduce the problem of solving these simultaneous equations to the preceding case of one equation, by eliminating one variable. Thus we multiply the second equation by 2 and subtract it from the first equation, getting

$$y - 3z = 0. \tag{3}$$

Now we meet one equation in more than one variable. We give z any value $\neq 0$, say $z = 1$, and solve for y, namely $y = 3$. We then solve for x from the second equation, namely $x = -y - z$, and obtain $x = -4$. The values which we have obtained for x, y, z are also solutions of the first equation, because the first equation is (in an obvious sense) the sum of equation (2) multiplied by 2, and equation (3).

Example 3. We wish to find a solution for the system of equations

$$3x - 2y + z + 2w = 0,$$

$$x + y - z - w = 0,$$

$$2x - 2y + 3z = 0.$$

Again we use the elimination method. Multiply the second equation by 2 and subtract it from the third. We find

$$-4y + 5z + 2w = 0.$$

Multiply the second equation by 3 and subtract it from the first. We find

$$-5y + 4z + 5w = 0.$$

We have now eliminated x from our equations, and find two equations in three unknowns, y, z, w. We eliminate y from these two equations as follows: Multiply the top one by 5, multiply the bottom one by 4, and subtract them. We get

$$9z - 10w = 0.$$

Now give an arbitrary value $\neq 0$ to w, say $w = 1$. Then we can solve for z, namely

$$z = 10/9.$$

Going back to the equations before that, we solve for y, using

$$4y = 5z + 2w.$$

This yields

$$y = 17/9.$$

Finally we solve for x using say the second of the original set of three equations, so that

$$x = -y + z + w,$$

or numerically,

$$x = -49/9.$$

Thus we have found:

$$w = 1, \qquad z = 10/9, \qquad y = 68/9, \qquad x = -49/9.$$

Note that we had three equations in four unknowns. By a successive elimination of variables, we reduced these equations to two equations in three unknowns, and then one equation in two unknowns.

Using precisely the same method, suppose that we start with three equations in five unknowns. Eliminating one variable will yield two equations in four unknowns. Eliminating another variable will yield one equation in three unknowns. We can then solve this equation, and proceed backwards to get values for the previous variables just as we have shown in the examples.

In general, suppose that we start with m equations with n unknowns, and $n > m$. We eliminate one of the variables, say x_1, and obtain a system of $m - 1$ equations in $n - 1$ unknowns. We eliminate a second variable, say x_2, and obtain a system of $m - 2$ equations in $n - 2$ unknowns. Proceeding stepwise, we eliminate $m - 1$ variables, ending up with 1 equation in $n - m + 1$ unknowns. We then give non-trivial arbitrary values to all the remaining variables but one, solve for this last variable, and then proceed backwards to solve successively for each one of the eliminated variables as we did in our examples. Thus we have an effective way of finding a non-trivial solution for the original system.

We shall phrase this in terms of induction in a precise manner.

Let $A = (a_{ij})$, $i = 1,\dots,m$ and $j = 1,\dots,n$ be a matrix. Let $b_1,\dots,b_m$ be numbers. Equations like

$$(*)\qquad \begin{matrix} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \qquad\qquad \vdots \quad\ \ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{matrix}$$

are called linear equations. We also say that $(*)$ is a system of linear equations. The system is said to be **homogeneous** if all the numbers $b_1,\dots,b_m$ are equal to 0. The number n is called the number of **unknowns**, and m is the number of equations.

The system of equations

$$(**)\qquad \begin{matrix} a_{11}x_1 + \dots + a_{1n}x_n = 0 \\ \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n = 0 \end{matrix}$$

will be called the **homogeneous system associated with** $(*)$. In this section, we study the homogeneous system $(**)$.

The system $(**)$ always has a solution, namely the solution obtained by letting all $x_i = 0$. This solution will be called the **trivial** solution. A solution $(x_1,\dots,x_n)$ such that some x_i is $\neq 0$ is called **non-trivial.**

Consider our system of homogeneous equations $(**)$. Let $A_1,\dots,A_m$ be the row vectors of the matrix (a_{ij}). Then we can rewrite our equations $(**)$ in the form

$$(**)\qquad \begin{matrix} A_1 \cdot X = 0 \\ \vdots \\ A_m \cdot X = 0. \end{matrix}$$

Therefore a solution of the system of linear equations can be interpreted as the set of all n-tuples X which are perpendicular to the row vectors of the matrix A. Geometrically, to find a solution of $(**)$ amounts to finding a vector X which is perpendicular to $A_1,\dots,A_m$. Using the notation of the dot product will make it easier to formulate the proof of our main theorem, namely:

Theorem 3.1. *Let*

$$(**)\qquad \begin{matrix} a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\ \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0 \end{matrix}$$

be a system of m linear equations in n unknowns, and assume that $n > m$. Then the system has a non-trivial solution.

Proof. The proof will be carried out by induction.

Consider first the case of one equation in n unknowns, $n > 1$:

$$a_1x_1 + \cdots + a_nx_n = 0.$$

If all coefficients $a_1,\dots,a_n$ are equal to 0, then any value of the variables will be a solution, and a non-trivial solution certainly exists. Suppose that some coefficient a_i is $\neq 0$. After renumbering the variables and the coefficients, we may assume that it is a_1. Then we give $x_2,\dots,x_n$ arbitrary values, for instance we let $x_2 = \cdots = x_n = 1$, and solve for x_1, letting

$$x_1 = \frac{-1}{a_1}(a_2 + \cdots + a_n).$$

In that manner, we obtain a non-trivial solution for our system of equations.

Let us now assume that our theorem is true for a system of $m - 1$ equations in more than $m - 1$ unknowns. We shall prove that it is true for m equations in n unknowns when $n > m$. We consider the system (**).

If all coefficients (a_{ij}) are equal to 0, we can give any non-zero value to our variables to get a solution. If some coefficient is not equal to 0, then after renumbering the equations and the variables, we may assume that it is a_{11}. We shall subtract a multiple of the first equation from the others to eliminate x_1. Namely, we consider the system of equations

$$\begin{gathered} \left(A_2 - \frac{a_{21}}{a_{11}}A_1\right)\cdot X = 0 \\ \vdots \\ \left(A_m - \frac{a_{m1}}{a_{11}}A_1\right)\cdot X = 0, \end{gathered}$$

which can also be written in the form

$$(***)\qquad \begin{gathered} A_2\cdot X - \frac{a_{21}}{a_{11}}A_1\cdot X = 0 \\ \vdots \\ A_m\cdot X - \frac{a_{m1}}{a_{11}}A_1\cdot X = 0. \end{gathered}$$

In this system, the coefficient of x_1 is equal to 0. Hence we may view (***) as a system of $m-1$ equations in $n-1$ unknowns, and we have $n-1 > m-1$.

According to our assumption, we can find a non-trivial solution $(x_2,\dots,x_n)$ for this system. We can then solve for x_1 in the first equation, namely

$$x_1 = \frac{-1}{a_{11}}(a_{12}x_2 + \cdots + a_{1n}x_n).$$

In that way, we find a solution of $A_1 \cdot X = 0$. But according to (***), we have

$$A_i \cdot X = \frac{a_{i1}}{a_{11}} A_1 \cdot X$$

for $i = 2,\dots,m$. Hence $A_i \cdot X = 0$ for $i = 2,\dots,m$, and therefore we have found a non-trivial solution to our original system (**).

The argument we have just given allows us to proceed stepwise from one equation to two equations, then from two to three, and so forth. This concludes the proof.

Exercises II, §3

1. Let

$$E_1 = (1, 0,\dots,0), \quad E_2 = (0, 1, 0,\dots,0), \quad \dots, \quad E_n = (0,\dots,0, 1)$$

be the standard unit vectors of $\mathbf{R}^n$. Let X be an n-tuple. If $X \cdot E_i = 0$ for all i, show that $X = O$.

2. Let $A_1,\dots,A_m$ be vectors in $\mathbf{R}^n$. Let X, Y be solutions of the system of equations

$$X \cdot A_i = 0 \qquad \text{and} \qquad Y \cdot A_i = 0 \qquad \text{for} \quad i = 1,\dots,m.$$

Show that $X + Y$ is also a solution. If c is a number, show that cX is a solution.

3. In Exercise 2, suppose that X is perpendicular to each one of the vectors $A_1,\dots,A_m$. Let $c_1,\dots,c_m$ be numbers. A vector

$$c_1A_1 + \cdots + c_mA_m$$

is called a **linear combination** of $A_1,\dots,A_m$. Show that X is perpendicular to such a vector.

4. Consider the inhomogeneous system (*) consisting of all X such that $X \cdot A_i = b_i$ for $i = 1, \ldots, m$. If X and X' are two solutions of this system, show that there exists a solution Y of the homogeneous system (**) such that $X' = X + Y$. Conversely, if X is any solution of (*), and Y a solution of (**), show that $X + Y$ is a solution of (*).

5. Find at least one non-trivial solution for each one of the following systems of equations. Since there are many choices involved, we don't give answers.

(a) $3x + y + z = 0$

(b) $\begin{aligned} 3x + y + z &= 0 \\ x + y + z &= 0 \end{aligned}$

(c) $\begin{aligned} 2x - 3y + 4z &= 0 \\ 3x + y + z &= 0 \end{aligned}$

(d) $\begin{aligned} 2x + y + 4z + w &= 0 \\ -3x + 2y - 3z + w &= 0 \\ x + y + z &= 0 \end{aligned}$

(e) $\begin{aligned} -x + 2y - 4z + w &= 0 \\ x + 3y + z - w &= 0 \end{aligned}$

(f) $\begin{aligned} -2x + 3y + z + 4w &= 0 \\ x + y + 2z + 3w &= 0 \\ 2x + y + z - 2w &= 0 \end{aligned}$

6. Show that the only solutions of the following systems of equations are trivial.

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

(b) $\begin{aligned} 4x + 5y &= 0 \\ -6x + 7y &= 0 \end{aligned}$

(c) $\begin{aligned} 3x + 4y - 2z &= 0 \\ x + y + z &= 0 \\ -x - 3y + 5z &= 0 \end{aligned}$

(d) $\begin{aligned} 4x - 7y + 3z &= 0 \\ x + y &= 0 \\ y - 6z &= 0 \end{aligned}$

(e) $\begin{aligned} 7x - 2y + 5z + w &= 0 \\ x - y + z &= 0 \\ y - 2z + w &= 0 \\ x + z + w &= 0 \end{aligned}$

(f) $\begin{aligned} -3x + y + z &= 0 \\ x - y + z - 2w &= 0 \\ x - z + w &= 0 \\ -x + y - 3w &= 0 \end{aligned}$

II, §4. Row Operations and Gauss Elimination

Consider the system of linear equations

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

The matrix of coefficients is

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

By the **augmented matrix** we shall mean the matrix obtained by inserting the column

$$\begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}$$

as a last column, so the augmented matrix is

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

In general, let $AX = B$ be a system of m linear equations in n unknowns, which we write in full:

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

Then we define the **augmented matrix** to be the m by $n + 1$ matrix:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

In the examples of homogeneous linear equations of the preceding section, you will notice that we performed the following operations, called **elementary row operations**:

Multiply one equation by a non-zero number.
Add one equation to another.
Interchange two equations.

These operations are reflected in operations on the augmented matrix of coefficients, which are also called **elementary row operations**:

Multiply one row by a non-zero number.
Add one row to another.
Interchange two rows.

Suppose that a system of linear equations is changed by an elementary row operation. Then the solutions of the new system are exactly the same as the solutions of the old system. By making row operations, we can hope to simplify the shape of the system so that it is easier to find the solutions.

Let us define two matrices to be **row equivalent** if one can be obtained from the other by a succession of elementary row operations. If A is the matrix of coefficients of a system of linear equations, and B the column vector as above, so that

$$(A, B)$$

is the augmented matrix, and if (A', B') is row-equivalent to (A, B) then the solutions of the system

$$AX = B$$

are the same as the solutions of the system

$$A'X = B'.$$

To obtain an equivalent system (A', B') as simple as possible we use a method which we first illustrate in a concrete case.

Example. Consider the augmented matrix in the above example. We have the following row equivalences:

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

Subtract 3 times second row from first row

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

Subtract 2 times second row from third row

$$\begin{pmatrix} 0 & -5 & 4 & 5 & 7 \\ 1 & 1 & -1 & -1 & -2 \\ 0 & -3 & 5 & 2 & 8 \end{pmatrix}$$

Interchange first and second row; multiply second row by -1.

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 5 & -4 & -5 & -7 \\ 0 & -3 & 5 & 2 & 8 \end{pmatrix}$$

Multiply second row by 3; multiply third row by 5.

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 15 & -12 & -15 & -21 \\ 0 & -15 & 25 & 10 & 40 \end{pmatrix}$$

Add second row to third row.

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 15 & -12 & -15 & -21 \\ 0 & 0 & 13 & -5 & 19 \end{pmatrix}$$

What we have achieved is to make each successive row start with a non-zero entry at least one step further than the preceding row. This makes it very simple to solve the equations. The new system whose augmented matrix is the matrix obtained last can be written in the form:

$$\begin{aligned} x + y - z - w &= -2, \\ 15y - 12z - 15w &= -21, \\ 13z - 5w &= 19. \end{aligned}$$

This is now in a form where we can solve by giving w an arbitrary value in the third equation, and solve for z from the third equation. Then we solve for y from the second, and x from the first. With the formulas, this gives:

$$z = \frac{19 + 5w}{13},$$

$$y = \frac{-21 + 12z + 15w}{15},$$

$$x = -1 - y + z + w.$$

We can give w any value to start with, and then determine values for x, y, z. Thus we see that the solutions depend on one free parameter. Later we shall express this property by saying that the set of solutions has dimension 1.

For the moment, we give a general name to the above procedure. Let M be a matrix. We shall say that M is in **row echelon form** if it has the following property:

Whenever two successive rows do not consist entirely of zeros, then the second row starts with a non-zero entry at least one step further to the right than the first row. All the rows consisting entirely of zeros are at the bottom of the matrix.

In the previous example we transformed a matrix into another which is in row echelon form. The non-zero coefficients occurring furthest to the left in each row are called the **leading coefficients**. In the above example, the leading coefficients are 1, 15, 13. One may perform one more change by dividing each row by the leading coefficient. Then the above matrix is row equivalent to

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -1 \\ 0 & 1 & -\frac{4}{5} & -1 & -\frac{7}{5} \\ 0 & 0 & 1 & -\frac{5}{13} & \frac{19}{13} \end{pmatrix}.$$

In this last matrix, the leading coefficient of each row is equal to 1. One could make further row operations to insert further zeros, for instance subtract the second row from the first, and then subtract $\frac{2}{5}$ times the third row from the second. This yields:

$$\begin{pmatrix} 1 & 0 & -\frac{7}{5} & -\frac{6}{5} & -2 \\ 0 & 1 & 0 & \frac{1}{5}+\frac{2}{13} & 1-\frac{38}{65} \\ 0 & 0 & 1 & -\frac{5}{13} & \frac{19}{13} \end{pmatrix}.$$

Unless the matrix is rigged so that the fractions do not look too horrible, it is usually a pain to do this further row equivalence by hand, but a machine would not care.

Example. The following matrix is in row echelon form.

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

Suppose that this matrix is the augmented matrix of a system of linear equations, then we can solve the linear equations by giving some variables an arbitrary value as we did. Indeed, the equations are:

$$\begin{aligned} 2y - 3z + 4w + \ \ t &= 7, \\ 5w + 2t &= -4, \\ -3t &= 1. \end{aligned}$$

Then the solutions are

$$\begin{aligned} t &= -1/3, \\ w &= \frac{-4-2t}{5}, \\ z &= \text{any arbitrarily given value}, \\ y &= \frac{7+3z-4w-t}{2}, \\ x &= \text{any arbitrarily given value}. \end{aligned}$$

The method of changing a matrix by row equivalences to put it in row echelon form works in general.

Theorem 4.1. *Every matrix is row equivalent to a matrix in row echelon form.*

Proof. Select a non-zero entry furthest to the left in the matrix. If this entry is not in the first column, this means that the matrix consists entirely of zeros to the left of this entry, and we can forget about them. So suppose this non-zero entry is in the first column. After an interchange of rows, we can find an equivalent matrix such that the upper left-hand corner is not 0. Say the matrix is

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

and $a_{11} \neq 0$. We multiply the first row by a_{21}/a_{11} and subtract from the second row. Similarly, we multiply the first row by a_{i1}/a_{11} and subtract it from the i-th row. Then we obtain a matrix which has zeros in the first column except for a_{11}. Thus the original matrix is row equivalent to a matrix of the form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{pmatrix}.$$

We then repeat the procedure with the smaller matrix

$$\begin{pmatrix} a'_{22} & \cdots & a'_{2n} \\ \vdots & & \vdots \\ a'_{m2} & \cdots & a'_{mn} \end{pmatrix}.$$

We can continue until the matrix is in row echelon form (formally by induction). This concludes the proof.

Observe that the proof is just another way of formulating the elimination argument of §3.

We give another proof of the fundamental theorem:

Theorem 4.2. *Let*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0, \\ \vdots \qquad\qquad & \ \ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0, \end{aligned}$$

be a system of m homogeneous linear equations in n unknowns with $n > m$. Then there exists a non-trivial solution.

Proof. Let $A = (a_{ij})$ be the matrix of coefficients. Then A is equivalent to A' in row echelon form:

$$\begin{aligned}
a_{k_1}x_{k_1} + S_{k_1}(x) \qquad\qquad\qquad &= 0,\\
a_{k_2}x_{k_2} + S_{k_2}(x) \qquad\quad &= 0,\\
\cdots\cdots\cdots\cdots \quad &\\
a_{k_r}x_{k_r} + S_{k_r}(x) &= 0,
\end{aligned}$$

where $a_{k_1} \neq 0, \ldots, a_{k_r} \neq 0$ are the non-zero coefficients of the variables occurring furthest to the left in each successive row, and $S_{k_1}(x), \ldots, S_{k_r}(x)$ indicate sums of variables with certain coefficients, but such that if a variable x_j occurs in $S_{k_1}(x)$, then $j > k_1$ and similarly for the other sums. If x_j occurs in S_{k_i} then $j > k_i$. Since by assumption the total number of variables n is strictly greater than the number of equations, we must have $r < n$. Hence there are $n - r$ variables other than $x_{k_1}, \ldots, x_{k_r}$ and $n - r > 0$. We give these variables arbitrary values, which we can of course select not all equal to 0. Then we solve for the variables x_{k_r}, $x_{k_{r-1}}, \ldots, x_{k_1}$ starting with the bottom equation and working back up, for instance

$$\begin{aligned}
x_{k_r} &= -S_{k_r}(x)/a_{k_r},\\
x_{k_{r-1}} &= -S_{k_{r-1}}(x)/a_{k_{r-1}}, \quad \text{and so forth.}
\end{aligned}$$

This gives us the non-trivial solution, and proves the theorem.

Observe that the pattern follows exactly that of the examples, but with a notation dealing with the general case.

Exercises II, §4

In each of the following cases find a row equivalent matrix in row echelon form.

1. (a) $\begin{pmatrix} 6 & 3 & -4 \\ -4 & 1 & -6 \\ 1 & 2 & -5 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix}$.

2. (a) $\begin{pmatrix} 1 & -2 & 3 & -1 \\ 2 & -1 & 2 & 2 \\ 3 & 1 & 2 & 3 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 & 3 & -2 \\ 2 & 1 & -4 & 3 \\ 2 & 3 & 2 & -1 \end{pmatrix}$.

3. (a) $\begin{pmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & -2 & 3 \\ 3 & 6 & 2 & -6 & 5 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 11 & -5 & 3 \\ 2 & -5 & 3 & 1 \\ 4 & 1 & 1 & 5 \end{pmatrix}$.

4. Write down the coefficient matrix of the linear equations of Exercise 5 in §3, and in each case give a row equivalent matrix in echelon form. Solve the linear equations in each case by this method.

II, §5. Row Operations and Elementary Matrices

Before reading this section, work out the numerical examples given in Exercises 33 *through* 37 *of* §2.

The row operations which we used to solve linear equations can be represented by matrix operations. Let $1 \leqq r \leqq m$ and $1 \leqq s \leqq m$. Let I_{rs} be the square $m \times m$ matrix which has component 1 in the rs place, and 0 elsewhere:

$$I_{rs} = \begin{pmatrix} 0 & \cdots\cdots\cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots 1_{rs} \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots\cdots\cdots & 0 \end{pmatrix}.$$

Let $A = (a_{ij})$ be any $m \times n$ matrix. What is the effect of multiplying $I_{rs}A$?

$$r\left\{\underbrace{\begin{pmatrix} 0 & \cdots\cdots\cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots 1_{rs} \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots\cdots\cdots & 0 \end{pmatrix}}_{s}\right. \left.\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ a_{s1} & \cdots & a_{sn} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\right\}s = \left.\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ a_{s1} & \cdots & a_{sn} \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}\right\}r.$$

The definition of multiplication of matrices shows that $I_{rs}A$ is the matrix obtained by putting the s-th row of A in the r-th row, and zeros elsewhere.

If $r = s$ then I_{rr} has a component 1 on the diagonal place, and 0 elsewhere. Multiplication by I_{rr} then leaves the r-th row fixed, and replaces all the other rows by zeros.

If $r \neq s$ let

$$J_{rs} = I_{rs} + I_{sr}.$$

Then

$$J_{rs}A = I_{rs}A + I_{sr}A.$$

Then $I_{rs}A$ puts the s-th row of A in the r-th place, and $I_{sr}A$ puts the r-th row of A in the s-th place. All other rows are replaced by zero. Thus J_{rs} interchanges the r-th row and the s-th row, and replaces all other rows by zero.

Example. Let

$$J = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 3 & 2 & -1 \\ 1 & 4 & 2 \\ -2 & 5 & 1 \end{pmatrix}.$$

If you perform the matrix multiplication, you will see directly that JA interchanges the first and second row of A, and replaces the third row by zero.

On the other hand, let

$$E = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then EA is the matrix obtained from A by interchanging the first and second row, and leaving the third row fixed. We can express E as a sum:

$$E = I_{12} + I_{21} + I_{33}$$

where I_{rs} is the matrix which has rs-component 1, and all other components 0 as before. Observe that E is obtained from the unit matrix by interchanging the first two rows, and leaving the third row unchanged. Thus the operation of interchanging the first two rows of A is carried out by multiplication with the matrix E obtained by doing this operation on the unit matrix.

This is a special case of the following general fact.

Theorem 5.1. *Let E be the matrix obtained from the unit $n \times n$ matrix by interchanging two rows. Let A be an $n \times n$ matrix. Then EA is the matrix obtained from A by interchanging these two rows.*

Proof. The proof is carried out according to the pattern of the example, it is only a question of which symbols are used. Suppose that we interchange the r-th and s-th row. Then we can write

$$E = I_{rs} + I_{sr} + \text{sum of the matrices } I_{jj} \text{ with } j \neq r,\ j \neq s.$$

Thus E differs from the unit matrix by interchanging the r-th and s-th rows. Then

$$EA = I_{rs}A + I_{sr}A + \text{sum of the matrices } I_{jj}A,$$

with $j \neq r$, $j \neq s$. By the previous discussion, this is precisely the matrix obtained by interchanging the r-th and s-th rows of A, and leaving all the other rows unchanged.

The same type of discussion also yields the next result.

Theorem 5.2. *Let E be the matrix obtained from the unit $n \times n$ matrix by multiplying the r-th row with a number c and adding it to the s-th row, $r \neq s$. Let A be an $n \times n$ matrix. Then EA is obtained from A by multiplying the r-th row of A by c and adding it to the s-th row of A.*

Proof. We can write

$$E = I + cI_{sr}.$$

Then $EA = A + cI_{sr}A$. We know that $I_{sr}A$ puts the r-th row of A in the s-th place, and multiplication by c multiplies this row by c. All other rows besides the s-th row in $cI_{sr}A$ are equal to 0. Adding $A + cI_{sr}A$ therefore has the effect of adding c times the r-th row of A to the s-th row of A, as was to be shown.

Example. Let

$$E = \begin{pmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then E is obtained from the unit matrix by adding 4 times the third row to the first row. Take any $4 \times n$ matrix A and compute EA. You will find that EA is obtained by multiplying the third row of A by 4 and adding it to the first row of A.

More generally, we can let $E_{rs}(c)$ for $r \neq s$ be the elementary matrix.

$$E_{rs}(c) = I + cI_{rs}.$$

$$
\begin{array}{r@{}l}
 & \quad\overbrace{\hphantom{1 \quad \cdots \quad 0}}^{r} \\
s\left\{\vphantom{\begin{matrix}1\\ \vdots\\0\\ \vdots\\0\end{matrix}}\right. &
\left(\begin{matrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots\\ 0 & \cdots & c & \cdots & 1 & \cdots & 0\\ \vdots & & \vdots & & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1\end{matrix}\right).
\end{array}
$$

It differs from the unit matrix by having rs-component equal to c. The effect of multiplication on the left by $E_{rs}(c)$ is to add c times the s-th row to the r-th row.

By an **elementary matrix**, we shall mean any one of the following three types:

(a) A matrix obtained from the unit matrix by multiplying the r-th diagonal component with a number $c \neq 0$.
(b) A matrix obtained from the unit matrix by interchanging two rows (say the r-th and s-th row, $r \neq s$).
(c) A matrix $E_{rs}(c) = I + cI_{rs}$ with $r \neq s$ having rs-component c for $r \neq s$, and all other components 0 except the diagonal components which are equal to 1.

These three types reflect the row operations discussed in the preceding section.

Multiplication by a matrix of type (a) multiplies the r-th row by the number c.

Multiplication by a matrix of type (b) interchanges the r-th and s-th row.

Multiplication by a matrix of type (c) adds c times the s-th row to the r-th row.

Proposition 5.3. *An elementary matrix is invertible.*

Proof. For type (a), the inverse matrix has r-th diagonal component c^{-1}, because multiplying a row first by c and then by c^{-1} leaves the row unchanged.

For type (b), we note that by interchanging the r-th and s-th row twice we return to the same matrix we started with.

For type (c), as in Theorem 5.2, let E be the matrix which adds c times the s-th row to the r-th row of the unit matrix. Let D be the matrix which adds $-c$ times the s-th row to the r-th row of the unit matrix (for $r \neq s$). Then DE is the unit matrix, and so is ED, so E is invertible.

Example. The following elementary matrices are inverse to each other:

$$E = \begin{pmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad E^{-1} = \begin{pmatrix} 1 & 0 & -4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We shall find an effective way of finding the inverse of a square matrix if it has one. This is based on the following properties.

If A, B are square matrices of the same size and have inverses, then so does the product AB, and

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

This is immediate, because

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

Similarly, for any number of factors:

Proposition 5.4. *If $A_1, \ldots, A_k$ are invertible matrices of the same size, then their product has an inverse, and*

$$(A_1 \cdots A_k)^{-1} = A_k^{-1} \cdots A_1^{-1}.$$

Note that in the right-hand side, we take the product of the inverses in reverse order. Then

$$A_1 \cdots A_k A_k^{-1} \cdots A_1^{-1} = I$$

because we can collapse $A_k A_k^{-1}$ to I, then $A_{k-1}A_{k-1}^{-1}$ to I and so forth.

Since an elementary matrix has an inverse, we conclude that any product of elementary matrices has an inverse.

Proposition 5.5. *Let A be a square matrix, and let A' be row equivalent to A. Then A has an inverse if and only if A' has an inverse.*

Proof. There exist elementary matrices $E_1, \ldots, E_k$ such that

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

Suppose that A has an inverse. Then the right-hand side has an inverse by Proposition 5.4 since the right-hand side is a product of invertible matrices. Hence A' has an inverse. This proves the proposition.

We are now in a position to find an inverse for a square matrix A if it has one. By Theorem 4.1 we know that A is row equivalent to a matrix A' in echelon form. If one row of A' is zero, then by the definition of echelon form, the last row must be zero, and A' is not invertible, hence A is not invertible. If all the rows of A' are non-zero, then A' is a triangular matrix with non-zero diagonal components. It now suffices to find an inverse for such a matrix. In fact, we prove:

Theorem 5.6. *A square matrix A is invertible if and only if A is row equivalent to the unit matrix. Any upper triangular matrix with non-zero diagonal elements is invertible.*

Proof. Suppose that A is row equivalent to the unit matrix. Then A is invertible by Proposition 5.5. Suppose that A is invertible. We have just seen that A is row equivalent to an upper triangular matrix with non-zero elements on the diagonal. Suppose A is such a matrix:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

By assumption we have $a_{11} \cdots a_{nn} \neq 0$. We multiply the i-th row with a_{ii}^{-1}. We obtain a triangular matrix such that all the diagonal components are equal to 1. Thus to prove the theorem, it suffices to do it in this case, and we may assume that A has the form

$$\begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & 1 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

We multiply the last row by a_{in} and subtract it from the i-th row for $i = 1, \ldots, n - 1$. This makes all the elements of the last column equal to 0 except for the lower right-hand corner, which is 1. We repeat this procedure with the next to the last row, and continue upward. This means that by row equivalences, we can replace all the components which lie strictly above the diagonal by 0. We then terminate with the unit matrix, which is therefore row equivalent with the original matrix. This proves the theorem.

Corollary 5.7. *Let A be an invertible matrix. Then A can be expressed as a product of elementary matrices.*

Proof. This is because A is row equivalent to the unit matrix, and row operations are represented by multiplication with elementary matrices, so there exist $E_1,\dots,E_k$ such that

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

Then $A = E_1^{-1} \cdots E_k^{-1}$, thus proving the corollary.

When A is so expressed, we also get an expression for the inverse of A, namely

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

The elementary matrices $E_1,\dots,E_k$ are those which are used to change A to the unit matrix.

Example. Let

$$A = \begin{pmatrix} 2 & -3 & 1 \\ 1 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix}, \qquad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We want to find an inverse for A. We perform the following row operations, corresponding to the multiplication by elementary matrices as shown.

Interchange first two rows.

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & -3 & 1 \\ 2 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Subtract 2 times first row from second row.
Subtract 2 times first row from third row.

$$\begin{pmatrix} 1 & 1 & -1 \\ 0 & -5 & 3 \\ 0 & -2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & -2 & 1 \end{pmatrix}.$$

Subtract 2/5 times second row from third row.

$$\begin{pmatrix} 1 & 1 & -1 \\ 0 & -5 & 3 \\ 0 & 0 & 9/5 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ -2/5 & -6/5 & 1 \end{pmatrix}.$$

Subtract 5/3 of third row from second row.
Add 5/9 of third row to first row.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 9/5 \end{pmatrix}, \quad \begin{pmatrix} -2/9 & 1/3 & 5/9 \\ 5/3 & 0 & -5/3 \\ -2/5 & -6/5 & 1 \end{pmatrix}.$$

Add 1/5 of second row to first row.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 9/5 \end{pmatrix}, \quad \begin{pmatrix} 1/9 & 1/3 & 2/9 \\ 5/3 & 0 & -5/3 \\ -2/5 & -6/5 & 1 \end{pmatrix}.$$

Multiply second row by $-1/5$.
Multiply third row by 5/9.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1/9 & 1/3 & 2/9 \\ -1/3 & 0 & 1/3 \\ -2/9 & -2/3 & 5/9 \end{pmatrix}.$$

Then A^{-1} is the matrix on the right, that is

$$A^{-1} = \begin{pmatrix} 1/9 & 1/3 & 2/9 \\ -1/3 & 0 & 1/3 \\ -2/9 & -2/3 & 5/9 \end{pmatrix}.$$

You can check this by direct multiplication with A to find the unit matrix.

If A is a square matrix and we consider an inhomogeneous system of linear equations

$$AX = B,$$

then we can use the inverse to solve the system, if A is invertible. Indeed, in this case, we multiply both sides on the left by A^{-1} and we find

$$X = A^{-1}B.$$

This also proves:

Proposition 5.8. *Let $AX = B$ be a system of n linear equations in n unknowns. Assume that the matrix of coefficients A is invertible. Then there is a unique solution X to the system, and*

$$X = A^{-1}B.$$

Exercises II, §5

1. Using elementary row operations, find inverses for the following matrices.

(a) $\begin{pmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 3 & -1 & 5 \\ -1 & 2 & 1 \\ -2 & 4 & 3 \end{pmatrix}$

(c) $\begin{pmatrix} 2 & 4 & 3 \\ -1 & 3 & 0 \\ 0 & 2 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 2 & 7 \end{pmatrix}$

(e) $\begin{pmatrix} -1 & 5 & 3 \\ 4 & 0 & 0 \\ 2 & 7 & 8 \end{pmatrix}$ (f) $\begin{pmatrix} 3 & 1 & 2 \\ 4 & 5 & 1 \\ -1 & 2 & -1 \end{pmatrix}$

Note: For another way of finding inverses, see the chapter on determinants.

2. Let $r \neq s$. Show that $I_{rs}^2 = O$.

3. Let $r \neq s$. Let $E_{rs}(c) = I + cI_{rs}$. Show that

$$E_{rs}(c)E_{rs}(c') = E_{rs}(c + c').$$

II, §6. Linear Combinations

Let $A^1, \ldots, A^n$ be m-tuples in $\mathbf{R}^m$. Let $x_1, \ldots, x_n$ be numbers. Then we call

$$x_1 A^1 + \cdots + x_n A^n$$

a **linear combination** of $A^1, \ldots, A^n$; and we call $x_1, \ldots, x_n$ the **coefficients** of the linear combination. A similar definition applies to a linear combination of row vectors.

The linear combination is called **non-trivial** if not all the coefficients $x_1, \ldots, x_n$ are equal to 0.

Consider once more a system of linear homogeneous equations

$$(**) \qquad \begin{matrix} a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\ \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0. \end{matrix}$$

Our system of homogeneous equations can also be written in the form

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or more concisely:

$$x_1A^1 + \cdots + x_nA^n = O,$$

where $A^1,\ldots,A^n$ are the column vectors of the matrix of coefficients, which is $A = (a_{ij})$. Thus the problem of finding a non-trivial solution for the system of homogeneous linear equations is equivalent to finding a non-trivial linear combination of $A^1,\ldots,A^n$ which is equal to O.

Vectors $A^1,\ldots,A^n$ are called **linearly dependent** if there exist numbers $x_1,\ldots,x_n$ not all equal to 0 such that

$$x_1A^1 + \cdots + x_nA^n = O.$$

Thus a non-trivial solution $(x_1,\ldots,x_n)$ is an n-tuple which gives a linear combination of $A^1,\ldots,A^n$ equal to O, i.e. a relation of linear dependence between the columns of A. *We may thus summarize the description of the set of solutions of the system of homogeneous linear equations in a table.*

(a) *It consists of those vectors X giving linear relations*

$$x_1A^1 + \cdots + x_nA^n = O$$

between the columns of A.

(b) *It consists of those vectors X perpendicular to the rows of A, that is $X \cdot A_i = 0$ for all i.*

(c) *It consists of those vectors X such that $AX = O$.*

Vectors $A^1,\ldots,A^n$ are called **linearly independent** if, given any linear combination of them which is equal to O, i.e.

$$x_1A^1 + \cdots + x_nA^n = O,$$

then we must necessarily have $x_j = 0$ for all $j = 1,\ldots,n$. This means that there is no non-trivial relation of linear dependence among the vectors $A^1,\ldots,A^n$.

Example. The standard unit vectors

$$E_1 = (1, 0,\ldots,0),\ldots,E_n = (0,\ldots,0, 1)$$

of $\mathbf{R}^n$ are linearly independent. Indeed, let $x_1,\ldots,x_n$ be numbers such that

$$x_1E_1 + \cdots + x_nE_n = O.$$

The left-hand side is just the n-tuple $(x_1, \dots, x_n)$. If this n-tuple is O, then all components are 0, so $x_i = 0$ for all i. This proves that $E_1, \dots, E_n$ are linearly independent.

We shall study the notions of linear dependence and independence more systematically in the next chapter. They were mentioned here just to have a complete table for the three basic interpretations of a system of linear equations, and to introduce the notion in a concrete special case before giving the general definitions in vector spaces.

Exercise II, §6

1. (a) Let $A = (a_{ij})$, $B = (b_{jk})$ and let $AB = C$ with $C = (c_{ik})$. Let C^k be the k-th column of C. Express C^k as a linear combination of the columns of A. Describe precisely which are the coefficients, coming from the matrix B.
   (b) Let $AX = C^k$ where X is some column of B. Which column is it?