Matrices and Linear Equations

Daoping Zhang

The Chinese University of Hong Kong

dpzhang@math.cuhk.edu.hk

May 2, 2020

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

$$\begin{aligned} 2x + y + z &= 1 \\ 5x - y + 7z &= 0. \end{aligned}$$

You have learned to solve such equations by the successive elimination of the variables. In this part, 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.

Outline

1. Matrices
2. Multiplication of Matrices
3. Homogeneous Linear Equations and Elimination
4. Row Operations and Gauss Elimination
5. Row Operations and Elementary Matrices
6. Linear Combinations

Matrices

Let n, m be two integers ≥ 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}), $i = 1, ..., m$ and $j = 1, ..., n$. 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 $(a_{21}, a_{22}, ..., a_{2n})$. We call a_{ij} the *ij*-**entry** or *ij*-**component** of the matrix.

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 $(x_1, ..., x_n)$ 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.

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

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

We define the **zero matrix** to be the matrix such that $a_{ij} = 0$ for all i, j. We shall write it O. We note that we have met so far with the zero number, zero vector, and zero matrix.

We define addition of matrices only when they have the same size. Thus let m, n be fixed integers ≥ 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.

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

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

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

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.

Let A, B be as in the previous example. 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

$$\mathrm{A+(\text{-}A)=O}$$

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

Let $A = (a_{ij})$ be an $m \times n$ matrix. The $n \times m$ matrix $B = (b_{ji})$ such that $b_{ij} = a_{ij}$ is called the **transpose** of A, and is also denoted by A^t. Taking the transpose of a matrix amounts to changing rows into columns and vice versa.

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

Multiplication of Matrices

Let $A = (a_{ij})$, $i = 1, ..., m$ and $j = 1, ..., n$ be an $m \times n$ matrix. Let $B = (b_{jk})$, $j = 1, ..., n$ and $k = 1, ..., s$ be an $n \times s$ matrix:

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

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 $A_1, ..., A_m$ are the row vectors of the matrix A, and if $B^1, ..., B^s$ 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

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

Let

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

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

$$AB = \begin{pmatrix} 15 & 15 \\ 4 & 12 \end{pmatrix}.$$

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

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

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,

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

then $AX = Y$ where Y is also a column vector, whose coordinate 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

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

Matrices give a convenient way of writing linear equations. See a system of two equations in three unknowns

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

Here, 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 the system of two simultaneous equations can be written in the form

$$AX = B.$$

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

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

can be written in the more efficient way

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

by the definition of multiplication of matrices. We say that there are m **equations** and n **unknowns**, or n **variables**.

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:

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

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

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

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

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

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

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

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

$$(AB)^t = B^t A^t.$$

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

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

$$X^t A^t = B^t.$$

If X, B are column vectors, then X^t, B^t are row vectors. It is occasionally convenient to rewrite the system in this fashion.

Let A be an $n \times n$ matrix. An **inverse for** A is a matrix B such that

$$AB = BA = I.$$

Sine 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**)**. In light of this, the inverse is denoted by

$$A^{-1}.$$

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

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

Theorem

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

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

Proof.

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

$$(A^{-1})^t A^t = I^t = I$$

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

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

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

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 then solve for x. We find

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

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

Consider a pair of equations, say

$$\begin{aligned} 2x + 3y - z &= 0, \\ x + y + z &= 0. \end{aligned}$$

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

$$y - 3z = 0.$$

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, x are also solutions of the first equation.

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.

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

$$\begin{array}{ccccc} a_{11}x_1 & + \cdots + & a_{1n}x_n = & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1}x_1 & + \cdots + & a_{mn}x_n = & b_m \end{array}$$

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

Theorem

Let

$$\begin{array}{cccc} a_{11}x_1 & + \cdots + & a_{1n}x_n = & 0 \\ \vdots & & \vdots & \vdots \\ a_{m1}x_1 & + \cdots + & a_{mn}x_n = & 0 \end{array} \tag{1}$$

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 our by induction.
Consider first the case of one equation in n unknowns, $n > 1$:

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

If all coefficients $a_1, ..., a_n$ 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 $x_2, ..., x_n$ arbitrary

Proof

values, for instance we let $x_2 = \cdots = x_n = 1$, and solve for x_1, letting

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

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 (1).

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

Proof

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

which can also be written in the form

$$
\begin{gathered}
A_2 \cdot X - \tfrac{a_{21}}{a_{11}}A_1 \cdot X = 0 \\
\vdots \\
A_m \cdot X - \tfrac{a_{m1}}{a_{11}}A_1 \cdot X = 0.
\end{gathered}
\tag{2}
$$

In this system, the coefficient of x_1 is equal to 0. Hence we may view (2) 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 $(x_2, ..., x_n)$ for this system. We can then solve for x_1 in the first equation, namely

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

Proof.

In that way, we find a solution of $A_1 \cdot X = 0$. But according to (2), we have

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

for $i = 2, ..., m$. Hence $A_i \cdot X = 0$ for $i = 2, ..., m$, and therefore we have found a non-trivial solution to our original system (1).

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. □

Consider the system of linear equations

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

The matrix of coefficient 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:

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

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

$$A'X = B'$$

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

$$\begin{pmatrix} 3 & -2 & 1 & 2 & 1 \\ 1 & 1 & -1 & -1 & -2 \\ 2 & -1 & 3 & 0 & 4 \end{pmatrix} \xrightarrow{R1-3\times R3} \begin{pmatrix} 0 & -5 & 4 & 5 & 7 \\ 1 & 1 & -1 & -1 & -2 \\ 2 & -1 & 3 & 0 & 4 \end{pmatrix} \xrightarrow{R3-2\times R2}$$

$$\begin{pmatrix} 0 & -5 & 4 & 5 & 7 \\ 1 & 1 & -1 & -1 & -2 \\ 0 & -3 & 5 & 2 & 8 \end{pmatrix} \xrightarrow{R1\Leftrightarrow R2,-1\times R2}$$

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 5 & -4 & -5 & -7 \\ 0 & -3 & 5 & 2 & 8 \end{pmatrix} \xrightarrow{3\times R2,5\times R3}$$

$$\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 15 & -12 & -15 & -21 \\ 0 & -15 & 25 & 10 & 40 \end{pmatrix} \xrightarrow{R3+R2} \begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 15 & -12 & -15 & -21 \\ 0 & 0 & 13 & -5 & 19 \end{pmatrix}$$

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. The new system whose augmented matrix is the matrix obtained last can be written in the form:

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

Then we can have

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

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

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 $1, 15, 13$.

The following matrix is in row echelon form:

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

Then, the equations are:

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

The solutions are

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

Theorem

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

Proof.

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. □

Row Operations and Elementary Matrices

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

Let $A = (a_{ij})$ be any $m \times n$ matrix. 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 row 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.

Theorem

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. The EA is the matrix obtained from A by interchanging these two rows.

Proof.

Suppose that we interchange the r-th and s-th row. The 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. □

Theorem

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. □

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 that 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

An elementary matrix is invertible.

Proposition

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

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

Proposition

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.

Theorem

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.

Corollary

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

Proposition

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

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

Linear Combinations

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

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

a **linear combination** of $A^1, ..., A^n$; and we call $x_1, ..., x_n$ 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 $x_1, ..., x_n$ are equal to 0.

Consider once more a system of linear homogenous equations

$$\begin{array}{cccc} a_{11}x_1 & + \cdots + & a_{1n}x_n & = 0 \\ \vdots & & \vdots & \\ a_{m1}x_1 & + \cdots + & a_{mn}x_n & = 0 \end{array} \tag{3}$$

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:

$$x_1 A^1 + \cdots + x_n A^n = O,$$

where $A^1, ..., A^n$ 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 homogenous linear equations is equivalent to finding a non-trivial linear combination of $A^1, ..., A^n$ which is equal to O.

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

$$x_1 A^1 + \cdots + x_n A^n = O.$$

Thus a non-trivial solution $(x_1, ..., x_n)$ is an n-tuple which gives a linear combination of $A^1, ..., A^n$ 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 homogenous linear equations:

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

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

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

$$x_1 A^1 + \cdots + x_n A^n = O,$$

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

The standard unit vectors

$$E_1 = (1, 0, ..., 0), ..., E_n = (0, ..., 0, 1)$$

of $\mathbb{R}^n$ are linearly independent.