Notes 1 – Matrix addition and multiplication

L1.1 Matrices and their entries. A *matrix* is a rectangular array of numbers. Examples:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 9 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 1 & 4 \end{pmatrix}$$

The individual numbers are called the *entries, elements,* or *components* of the matrix. If the matrix has m rows and n columns, we say that it has size 'm by n' or $m \times n$. The above examples have respective sizes 2×3, 4×1, 5×5, 2×2. If $m = n$ (as in the last two cases) the matrix is obviously *square.*

The set of matrices of size $m \times n$ whose entries are real numbers is denoted by $\mathbb{R}^{m,n}$; the first superscript is always the number of rows. Sometimes we use symbols to represent unspecified numbers, so the statement

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2,2}$$

is tantamount to saying that a, b, c, d are real numbers.

For matrices with more than about 4 entries, it is convenient to use subscripts to label the entries. Given a matrix A, we typically denote by a_{ij} the entry in the ith row and jth column (lower case to emphasize that the entry is a number).

    Example. In this notation, the generic 3×4 matrix is

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}.$$

    Mathematicians like to deal in generalities and will even write a matrix as $A = (a_{ij})$ without specifying its size.

Definition. *The transpose of a matrix A is the matrix, indicated $^t A$ or $A^\top$, is obtained by interchanging its rows and columns.*

For example

$$(1, 2, -7)^\top = \begin{pmatrix} 1 \\ 2 \\ -7 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

If $A \in \mathbb{R}^{m,n}$ then $A^\top \in \mathbb{R}^{n,m}$. In subscript notation, we have

$$(A^\top)_{ij} = a_{ji}.$$

Notice that $(A^\top)^\top = A$, so the operation of taking the transpose is self-inverse.

L1.2 Vectors. Of special importance are matrices that have only one row or column; they are called *row* and *column vectors.* In writing a row vector with digits, it is useful to use commas to separate the entries. For example, both the matrices

$$A = (1, 2, -7) \in \mathbb{R}^{1,3}, \qquad B = \begin{pmatrix} 1 \\ 2 \\ -7 \end{pmatrix} \in \mathbb{R}^{3,1}$$

can be used to represent the point in space with Cartesian coordinates $x = 1$, $y = 2$, $z = -7$. (Sometimes commas are used to distinguish between matrices and row vectors, but it is simpler to regard them as the same object.)

One can switch between row and column vectors by observing that $A = B^\top$ or $B = A^\top$. For this reason, the distinction between a row vector and a column vector is often unimportant, and the sets $\mathbb{R}^{1,n}$ and $\mathbb{R}^{n,1}$ can be written more simply $\mathbb{R}^n$, and we can refer to both $A \in \mathbb{R}^3$ and $B \in \mathbb{R}^3$ as 'vectors' of *length* 3. We shall use such vectors to study analytic geometry later in the course.

Whenever we write '$\mathbb{R}^n$' the reader is free to use row or column vectors as he or she prefers; if such a choice is not possible, we shall use the other notation to specify either rows or columns. Actually, vectors tend to be given lower-case names, and a vector of unspecified length n is more likely to be written

$$\mathbf{u} = (u_1, \ldots, u_n) \qquad \text{or} \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ v_n \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}.$$

Row and column vectors are not merely special cases of matrices. *Any* matrix can be regarded as an ordered list of both row vectors and column vectors. Given a matrix $A \in \mathbb{R}^{m,n}$, we shall denote its rows (thought of as matrices in their own right) by

$$\mathbf{r}_1, \ldots, \mathbf{r}_m \in \mathbb{R}^{1,n}$$

and its columns

$$\mathbf{c}_1, \ldots, \mathbf{c}_n \in \mathbb{R}^{m,1}$$

More informally (ignoring parentheses in a way that would be spell disaster in a computer program), we may write

$$\begin{pmatrix} \leftarrow \mathbf{r}_1 \rightarrow \\ \cdots \\ \leftarrow \mathbf{r}_m \rightarrow \end{pmatrix} = A = \begin{pmatrix} \uparrow & \cdot & \cdot & \uparrow \\ \mathbf{c}_1 & \cdot & \cdot & \mathbf{c}_n \\ \downarrow & \cdot & \cdot & \downarrow \end{pmatrix}.$$

Much of the study of matrices is ultimately be based on one or other of these two descriptions.

L1.3 Addition of matrices. A matrix is much more than an array of data. It is an algebraic object that is subject to operations generalizing the more familar ones applicable to numbers and vectors.

Definition. *To form the* sum *of two matrices* A, B*, they must have the same size. The entries are then added component-wise.*

For example

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 9 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -2 & 4 \\ 2 & -6 & 1 \end{pmatrix} \quad \Rightarrow \quad A + B = \begin{pmatrix} 1 & 0 & 7 \\ 2 & 0 & 10 \end{pmatrix}$$

Of course, the result is still a matrix of size 2×3.

In particular, if we add a matrix to itself, we merely double every entry and it is reasonable to call the result $2A$:

$$A + A = \begin{pmatrix} 2 & 4 & 6 \\ 0 & 12 & 18 \end{pmatrix} = 2A$$

Definition. *If $c \in \mathbb{R}$ and $A \in \mathbb{R}^{m,n}$ then cA is the matrix formed by multiplying every entry of A by c.*

If c is zero, we get a *null* matrix

$$0A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{0}.$$

A null matrix is denoted by $\mathbf{0}$ or $\underline{0}$ (or even 0 like the number), provided the context makes clear its size. Of course, these definitions apply equally to vectors, so for example

$$2(x, y, z) = (2x, 2y, 2z).$$

We denote $(-1)B$ by $-B$, so that matrices can be subtracted in the obvious way:

$$A - B = \begin{pmatrix} 1 & 4 & -1 \\ -2 & 12 & 8 \end{pmatrix} = A + (-B).$$

Exercise. Explain why $A + B = B + A$ and $(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}}$.

L1.4 Matrix multiplication. First we define a numerical product between two vectors $\mathbf{u}, \mathbf{v}$ of the same length. For this it does not really matter whether they are row or column vectors, but for egalitarian purposes we shall suppose that the first is a row vector and the second a column vector. Thus, we consider

$$\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{R}^{1,n}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ v_n \end{pmatrix} \in \mathbb{R}^{n,1}.$$

Definition. *The* dot *or* scalar *product of $\mathbf{u}$ and $\mathbf{v}$, written $\mathbf{u} \cdot \mathbf{v}$ is the number*

$$\sum_{i=1}^{n} u_i v_i = u_1 v_1 + \cdots + u_n v_n.$$

(We shall not use the summation symbol much in this course, but students should be familiar with its use.) The dot product provides the basis for multiplying matrices:

Definition. *The product of two matrices A, B is only defined if the number of columns of A equals the number of rows of B. If $A \in \mathbb{R}^{m,n}$ has rows $\mathbf{r}_1, \ldots \mathbf{r}_m$ and $B \in \mathbb{R}^{n,p}$ has columns $\mathbf{c}_1, \ldots, \mathbf{c}_p$ then AB is the matrix with entries $\mathbf{r}_i \cdot \mathbf{c}_j$ and has size $m \times p$.*

More explicitly,

$$
AB = \begin{pmatrix} \leftarrow \mathbf{r}_1 \rightarrow \\ \cdots \\ \leftarrow \mathbf{r}_m \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \cdot & \cdot & \uparrow \\ \mathbf{c}_1 & \cdot & \cdot & \mathbf{c}_p \\ \downarrow & \cdot & \cdot & \downarrow \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 \cdot \mathbf{c}_1 & \cdot & \cdot & \mathbf{r}_1 \cdot \mathbf{c}_p \\ & \cdots & & \cdots \\ \mathbf{r}_m \cdot \mathbf{c}_1 & \cdot & \cdot & \mathbf{r}_m \cdot \mathbf{c}_p \end{pmatrix}.
$$

One should imagine taking each row of A, rotating it and placing it on top of each column of B in turn so as to perform the dot product.

> **Example.** A very special case is the product $\mathbf{r}_1 \mathbf{c}_1 = (\mathbf{r}_1 \cdot \mathbf{c}_1)$ of a single row and a column. Strictly speaking, this is a 1×1 matrix, but (again ignoring parentheses) we shall regard it as a number, i.e. the dot product. With this convention, if $\mathbf{v} = (x, y, z)$ then
>
> $$
> \mathbf{v}\mathbf{v}^T = (x, y, z) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 + y^2 + z^2.
> $$
>
> Later, we shall refer to the square root of this quantity as the *norm* of the vector $\mathbf{v}$ (it is the distance from the corresponding point to the origin). By contrast, note that
>
> $$
> \mathbf{v}^T \mathbf{v} = \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix}
> $$
>
> is a 3×3 matrix.

An intermediate case of the matrix product is that in which the second factor is a single column $\mathbf{v} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^{n,1}$, so that

$$
A\mathbf{v} = \begin{pmatrix} \mathbf{r}_1 \cdot \mathbf{v} \\ \cdot \\ \mathbf{r}_m \cdot \mathbf{v} \end{pmatrix}.
$$

In general we can say that

$$
AB = \begin{pmatrix} \uparrow & \cdot & \cdot & \uparrow \\ A\mathbf{c}_1 & \cdot & \cdot & A\mathbf{c}_p \\ \downarrow & \cdot & \cdot & \downarrow \end{pmatrix}.
$$

This shows clearly that each column of the product is obtained by premultiplying the corresponding column of B.

The rule for manipulating the sizes can be remembered by the scheme

$$
m \times \underbrace{n \qquad n} \times p \quad \rightsquigarrow \quad m \times p,
$$

and matrix multiplication defines a mapping

$$
\mathbb{R}^{m,n} \times \mathbb{R}^{n,p} \longrightarrow \mathbb{R}^{m,p}.
$$

Even if AB is defined, it will often be the case that BA is not. The situation is much more symmetrical if $m = n$, and we investigate this next.