Basics: Matrices and vectors

An $M \times N$ **matrix** A is written as

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,N} \end{bmatrix}.$$

The **dimension** of the matrix is said to be $M \times N$, where the number of rows is always given first, and the number of columns is given second. In general, the coefficients of the matrix are real or complex numbers, so they support all the standard arithmetic operations. We write the matrix more compactly as

$$A = [a_{i,j}, 1 \leq i \leq M, 1 \leq j \leq N]$$

or, even more simply as $A = [a_{i,j}]$ when the dimension of A is understood. The matrix entries $a_{i,j}$ are called the coefficients of the matrix.

A **vector** is a matrix with only one row or only one column. An N-dimensional **column vector** s is written as an $N \times 1$ matrix

$$s = \begin{bmatrix} s_1 \\ s_2 \\ \cdots \\ s_N \end{bmatrix}$$

An N-dimensional **row vector** z is written as a $1 \times N$ matrix

$$z = [z_1, z_2, \cdots, z_N]$$

The **transpose** of a $M \times N$ matrix $A = [a_{i,j}]$ is the $N \times M$ matrix $A^T = [a_{j,i}]$. Therefore, the transpose of an N-dimensional column vector s is the N-dimensional row vector s^T, and the transpose of a N-dimensional row vector z is the N-dimensional column vector z^T.

From now on, unless explicitly stated otherwise, all vectors denoted s, x, y, b, c etc. *without* the transpose notation are column vectors, and vectors denoted s^T, x^T, y^T, b^T, c^T *with* the transpose notation are row vectors.

A tuple of numeric values is often represented as a vector. A tuple, however, is neither a “row” nor a “column.” Thus, the representation as a vector carries the additional information that it is either a row or a column vector.

Basics: Matrix arithmetic

Two matrices (or vectors, since they are also matrices) can be added or subtracted provided that they have the same dimension. Just as with adding or subtracting tuples, the elements are added or subtracted. Thus if $A = [a_{i,j}]$ and $B = [b_{i,j}]$ and both have dimension $M \times N$, then

$$A + B = [a_{i,j} + b_{i,j}].$$

Under certain circumstances, matrices can also be multiplied. If A has dimension $M \times N$ and B has dimension $N \times P$, then the product AB is defined. The number of columns of A must match the number of rows of B. Suppose the matrices are given by

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,N} \end{bmatrix} \quad B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,P} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,P} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N,1} & b_{N,2} & \cdots & b_{N,P} \end{bmatrix},$$

Then the i, j element of the product $C = AB$ is

$$c_{i,j} = \sum_{m=1}^{N} a_{i,m} b_{m,j}. \tag{5.23}$$

The product has dimension $M \times P$.

Of course, matrix multiplication also works if one of the matrices is a vector. If b is a column vector of dimension N, then $c = Ab$ as defined by (5.23) is a column vector of dimension M. If on the other hand b^T is a row vector of dimension M, then $c^T = b^T A$ as defined by (5.23) is a row vector of dimension N. By convention, we write x^T to indicate a row vector, and x to indicate a column vector. Also by convention, we (usually) use lower case variable names for vectors and upper case variable names for matrices.

Multiplying a matrix by a vector can be interpreted as applying a function to a tuple. The vector is the tuple and the matrix (together with the definition of matrix multiplication) defines the function. Thus, in introducing matrix multiplication into our systems, we are doing nothing new except introducing a more compact notation for defining a particular class of functions.

A matrix A is a **square matrix** if it has the same number of rows and columns. A square matrix may be multiplied by itself. Thus, A^n for some integer $n > 0$ is defined to be A multiplied by itself n times. A^0 is defined to be the **identity matrix**, which has ones along the diagonal and zeros everywhere else.