Matrices and Linear Functions

1 Matrices and Vectors

An $m \times n$ *matrix* is a rectangular array of mn numbers:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}. \tag{1}$$

Another way to write this is $A = (a_{ij})$.

An $n \times 1$ vector is just a column of numbers, called a *column vector*. For these matrices there is no need for two subscripts on the entries, and if A is an $n \times 1$ matrix we just write

$$A = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}.$$

A column vector can be identified with a vector in $\mathbb{R}^n$ in the obvious way:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} \longleftrightarrow (a_1, a_2, a_3, \ldots, a_n).$$

Likewise, any vector in $\mathbb{R}^n$ can be identified with a column vector.

From now on we will use the above identification and think of every $\mathbf{x} \in \mathbb{R}^n$ as a column vector. Really, all we're doing is writing our vectors "standing up." Since we have vector addition, scalar multiplication and the dot product in $\mathbb{R}^n$ we have the same operations on column vectors (e.g. to multiply a column vector by a scalar α just multiply all its entries by α).

Our identification of $\mathbb{R}^n$ with the set of column vectors also gives us some shorthand for writing matrices. Using the matrix A of (1) we set

$$\mathbf{a}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \\ \vdots \\ a_{mi} \end{pmatrix}.$$

Then we can write

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \cdots & \mathbf{a}_n \end{pmatrix}. \tag{2}$$

With this notation it is easy to describe the multiplication of a matrix by a column vector. Let A be an $m \times n$ matrix and let $\mathbf{x} \in \mathbb{R}^n$ (remember, think of $\mathbf{x}$ as a column vector). Write A as in (2) and let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}.$$

Then we define

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + \cdots + x_n\mathbf{a}_n.$$

That is, the result of the multiplication of $\mathbf{x}$ by A is a column vector in $\mathbb{R}^m$, and we obtain this vector by forming the weighted sum of the columns of A using the entries of $\mathbf{x}$ as the weights (note that in order for $A\mathbf{x}$ to make sense $\mathbf{x}$ must have as many entries as A has columns!). Matrix multiplication has two very nice properties, which we now illustrate.

Let A be an $m \times n$ matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} A(\mathbf{x} + \mathbf{y}) &= A\mathbf{x} + A\mathbf{y} \\ A(\alpha\mathbf{x}) &= \alpha(A\mathbf{x}). \end{aligned}$$

We only prove the first equality. The proof of the second follows the same lines. Write

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

so that

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

Writing A as in (2), the definition above and properties of vector addition and scalar multiplication give us

$$\begin{aligned}
A(\mathbf{x}+\mathbf{y}) &= (x_1+y_1)\mathbf{a}_1 + (x_2+y_2)\mathbf{a}_2 + (x_3+y_3)\mathbf{a}_3 + \cdots + (x_n+y_n)\mathbf{a}_n \\
&= x_1\mathbf{a}_1 + y_1\mathbf{a}_1 + x_2\mathbf{a}_2 + y_2\mathbf{a}_2 + x_3\mathbf{a}_3 + y_3\mathbf{a}_3 + \cdots + x_n\mathbf{a}_n + y_n\mathbf{a}_n \\
&= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + \cdots + x_n\mathbf{a}_n + y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + y_3\mathbf{a}_3 + \cdots + y_n\mathbf{a}_n \\
&= A\mathbf{x} + A\mathbf{y}.
\end{aligned}$$

2 Linear Functions and Matrix Multiplication

We can now define linear functions. Our primary reason for doing so is to motivate the general definition of matrix multiplication, which at first sight can seem unnecessarily complicated (I guess my point is that the complication *is* actually necessary!). A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear function* if there is an $m \times n$ matrix A so that

$$f(\mathbf{x}) = A\mathbf{x}.$$

Hence, every $m \times n$ matrix A gives us a linear function $l_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$l_A(\mathbf{x}) = A\mathbf{x}.$$

The letter l is used to suggest the name for this function: *l*eft multiplication by A.

By attempting to compose two linear functions we will be led to the notion of matrix multiplication. We start with two linear functions $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^l$ given by the matrices B and A, respectively. Thus, B is an $m \times n$ matrix and A is $l \times m$. We'll compute the value of $g \circ f(\mathbf{x})$ for any vector $\mathbf{x} \in \mathbb{R}^n$. First let's write

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \cdots & \mathbf{a}_m \end{pmatrix}$$

and

$$B = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \cdots & \mathbf{b}_n \end{pmatrix}$$

and

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}.$$

Then

$$\begin{aligned}
g \circ f(\mathbf{x}) &= g(f(\mathbf{x})) \\
&= g(B\mathbf{x}) \\
&= g(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3 + \cdots + x_n\mathbf{b}_n) \\
&= A(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3 + \cdots + x_n\mathbf{b}_n) \\
&= x_1(A\mathbf{b}_1) + x_2(A\mathbf{b}_2) + x_3(A\mathbf{b}_3) + \cdots + x_n(A\mathbf{b}_n).
\end{aligned}$$

So, if we define a new matrix C by

$$C = \begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 & \cdots & A\mathbf{b}_n \end{pmatrix} \tag{3}$$

(i.e. the columns of C are the columns of B multiplied by A) then we may continue the series of equalities above with

$$= \quad C\mathbf{x}.$$

That is, *the composition of two linear functions is another linear function* and we can compute the matrix of the composition from the matrices of the functions being composed. This is our motivation for defining the *product* of the matrices A and B, AB, by

$$AB = C$$

where C is as given in (3).

It is important to note that in order for the product of two matrices to be defined their dimensions must agree in a certain way. The *second* dimension of A must equal the *first* dimension of B (i.e. A must be $l \times m$ and B must be $m \times n$). Thus, even though AB might make sense it is possible that BA does not. For example, if A is 3×5 and B is 5×4 then AB makes sense (and is 3×4) but BA is meaningless.

Another thing to note is that even if we have two matrices A and B so that AB *and* BA both make sense, it is often the case that $AB \neq BA$. A simple example is if A is 3×5 and B is 5×3, for then AB is 3×3 but BA is 5×5. But even if AB and BA are the same size they can be different. To see this, find 2×2 matrices A and B so that $AB \neq BA$. Do the same with 3×3 matrices.

There's a whole lot more that can be said about matrices and linear maps. So much, in fact, that there is an entire subject dedicated to their study: linear algebra.