3. Matrices as functions

Let us review the story so far. We are interested in solving systems of linear equations. We represent such a system in the very compact form $Ax = b$. Here A is the coefficient matrix and x is a vector containing the variables, so that we are trying to solve for x in terms of b and A. If we pick values for the variables $x_1, x_2, \dots, x_n$ we get a vector v and we can compute Av to see if we get b.

But let us look at this problem from a slightly different perspective. We input a value $v \in \mathbb{R}^n$ for the variables x and we output a vector $Av \in \mathbb{R}^m$. In other words the matrix A defines a function, call it

$$\phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \text{by the rule} \qquad v \longrightarrow Av.$$

Here $\phi(v) = Av$ is just simply the vector you get by plugging in the values determined by v for the indeterminates $x_1, x_2, \dots, x_n$. Therefore the vector v is a solution to the linear equation $Ax = \phi(v)$.

Let us interpret solving an equation in terms of the function ϕ.

What are the solutions to the equation $Ax = b$? They are the vectors v such that $\phi(v) = Av = b$. In other words the solution set to the equation $Ax = b$ is the inverse image of b, $\phi^{-1}(b)$.

Now let us ask a slightly different question. What are the vectors b for which we can solve the equation $Ax = b$? They are the vectors in the image of ϕ.

Now ϕ is surjective (onto) if and only if for every $b \in \mathbb{R}^m$ there is a vector $v \in \mathbb{R}^n$ such that $\phi(v) = b$, if and only if for every $b \in \mathbb{R}^m$ there is a solution to the system of equations $Ax = b$.

In other words ϕ is surjective if and only if every equation $Ax = b$ has a solution, for every $b \in \mathbb{R}^m$.

On the other hand ϕ is injective if and only if $\phi(v_1) = \phi(v_2)$ implies $v_1 = v_2$ if and only if whenever $b = \phi(v_1)$ and $b = \phi(b_2)$ then $v_1 = v_2$ if and only if whenever $Av_1 = b$ and $Av_2 = b$ then $v_1 = v_2$ if and only if every equation $Ax = b$ has at most one solution.

In other words ϕ is injective if and only if every equation $Ax = b$ has at one solution solution, for every $b \in \mathbb{R}^m$.

Thus ϕ is bijective if and only if ϕ is injective and surjective (by definition) if and only if every equation $Ax = b$ has at most one solution and at least one solution if and only if every equation has a unique solution.

It is perhaps useful to think of this geometrically. Suppose that $m = 2$, $n = 1$. Geometrically there are two things which can happen. Suppose that ϕ is injective. Then the image of $\mathbb{R} = \mathbb{R}^1$ is a line in $\mathbb{R}^2$. If we pick a point b on this line then the equation $Ax = b$ has a unique solution. For an example when ϕ is injective, take

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

so that $\phi(t) = (t, 0)$. The image is the x-axis. If ϕ is not injective then it must collapse the whole of $\mathbb{R}$ to a single point, necessarily the origin. If $b \neq 0$ then the equation $Ax = b$ has no solutions and otherwise any point in $\mathbb{R}^1$ gives a solution. The only example where ϕ is not injective is when

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

In this case the map is $\phi(t) = (0, 0)$. In particular ϕ is never surjective, if $m = 2$ and $n = 1$.

Now reverse the roles of m and n. Suppose that $m = 1$ and $n = 2$. Suppose that ϕ is surjective. Then every equation has a solution. In fact it has a one dimensional family of solutions. For example if $A = (1, 0)$ then the corresponding map is $\phi(x, y) = x$. If we fix $x = b$, the set of solutions is the line (b, y), where y can take on any value. The only other case is if ϕ is not surjective and in this case if $b = 0$ then we get a two dimensional family of solutions and otherwise we get a one dimensional family. The matrix $A = (0, 0)$ and the map $\phi(x, y) = 0$. In particular ϕ is never injective, if $m = 2$ and $n = 1$.

Finally suppose that $m = n = 2$. In the best case ϕ is bijective. The inverse image of every point is a point. For example consider

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

The corresponding map is $\phi(x, y) = (x, y)$, which is the identity map. But there are other possibilities. If the image of ϕ is a line then either an equation has a solution or it has a line of solutions. For example, consider the matrix

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

The map $\phi(x, y) = (x, 0)$. The image is the x-axis. The fibre over the point $(b, 0)$ is the line (b, y). So much for the geometric picture.

Now let us consider adding and multiplying matrices. Firstly adding matrices is the analogue of adding functions. So to add two matrices A and B together they need to define functions from the same domain to the same range. In other words both A and B need to have the same shape $m \times n$. It turns out that all we do is add corresponding entries:

Definition 3.1. *Let $A = (a_{ij})$ and $B = (b_{ij})$ be two real $m \times n$ matrices. The sum $C = (c_{ij}) = A + B = (a_{ij} + b_{ij})$, so that $c_{ij} = a_{ij} + b_{ij}$.*

A simple example will surely suffice:

$$\begin{pmatrix} -1 & 2 & -5 \\ 3 & 1 & 3 \end{pmatrix} + \begin{pmatrix} 2 & 4 & -5 \\ -1 & 3 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 6 & -10 \\ 2 & 4 & 9 \end{pmatrix}.$$

There are a bunch of trivial observations. Adding matrices is commutative, that is

$$A + B = B + A.$$

Adding matrices is associative, that is

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

The zero matrix 0 of type $m \times n$ is the matrix with a zero in every entry.

$$A + 0 = 0 + A = A.$$

Finally, given a matrix A of type $m \times n$ let $-A$ be the matrix whose entries have the opposite sign to A. Then

$$A + (-A) = (-A) + A = 0.$$

All of these observations can easily be checked using components. Alternatively they follow automatically from the fact that matrices correspond to functions.

We can also multiply a matrix A by a scalar λ. This is the analogue of multiplying a function by a scalar,

Definition 3.2. *Let $A = (a_{ij})$ be a real $m \times n$ matrix and let λ be a real number. Then $\lambda A = (\lambda a_{ij})$.*

Again a simple example will surely suffice:

$$3 \begin{pmatrix} -1 & 2 & -5 \\ 3 & 1 & 3 \end{pmatrix} = \begin{pmatrix} -3 & 6 & -15 \\ 9 & 3 & 9 \end{pmatrix}.$$

Scalar multiplication distributes across sums, which means that if A and B are two $m \times n$ matrices, then

$$\lambda(A + B) = \lambda A + \lambda B.$$

Note that zero times any matrix is the zero matrix and that the additive inverse matrix is nothing but -1 times the original matrix. All of this reflects what is happening at the level of functions.

What about multiplying two matrices? The key point is that multiplying matrices is the analogue of composing functions. In other words if I have a matrix A of type $m \times n$ and a matrix B of type $p \times q$ then I get two functions,

$$\phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \text{and} \qquad \psi\colon \mathbb{R}^q \longrightarrow \mathbb{R}^p.$$

The matrix product AB corresponds to the composition

$$\phi \circ \psi \colon \mathbb{R}^q \longrightarrow \mathbb{R}^m.$$

Now the only way we can possibly compose ϕ and ψ is if $\mathbb{R}^p = \mathbb{R}^n$, which is equivalent to $p = n$. In other words it only makes sense to multiply a matrix of type $m \times n$ with a matrix of type $p \times q$ if $n = p$. Since the resulting function $\phi \circ \psi$ goes from $\mathbb{R}^q$ to $\mathbb{R}^m$ it follows that the matrix product AB has type $m \times q$.

So how do we define AB if $n = p$? Let us start with a known case. Suppose that B is a column vector, so that B has type $n \times 1$. In this case matrix multiplication is the function $\phi \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Suppose that the matrix $AB = C = (c_{ij})$. The ith row of C is a single number c_{i1} which is obtained by taking the ith row of A, thinking of this as an equation and then plugging in the values given by $B = (b_{ij})$. We get

$$c_{i1} = a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1} = \sum_l a_{il}b_{l1}.$$

So now suppose that $q \neq 1$. The rule is to treat the other columns of B just like the first column,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_l a_{il}b_{lj}.$$

Here we take the ith row of A, think of it as defining an equation and plug in the values determined by the jth column of B, to get the (i, j) entry of C. Again, the key point is that this is just what you get by thinking of matrices as defining functions and composing those functions (we will see later when we investigate more the connection between functions and matrices that this is the correct rule).

At this point a few simple examples will probably help.

$$\begin{pmatrix} -1 & 2 & -5 \\ 3 & 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ -1 & 3 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} -2-2-5 & -4+6-15 \\ 6-1+3 & 12+3+9 \end{pmatrix} = \begin{pmatrix} -9 & -13 \\ 8 & 24 \end{pmatrix}.$$

Note that we are multiplying a 2×3 matrix with a 3×2 matrix to get a 2×2 matrix. One curious example is to take a 2×1 matrix and a 1×2 matrix,

$$A = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} -2 & 3 \end{pmatrix}.$$

We first compute AB. This makes sense as the ones match up and the end result is a matrix of type 2×2,

$$AB = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \begin{pmatrix} -2 & 3 \end{pmatrix} = \begin{pmatrix} -4 & 6 \\ 2 & -3 \end{pmatrix}.$$

What about BA? This makes sense as the 2's match up and the end result is a matrix of type 1×1 (a number).

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

Finally let us suppose we take two 2×2 matrices.

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

Then

$$AB = \begin{pmatrix} -1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & -4 \end{pmatrix} = \begin{pmatrix} -1 & -7 \\ 10 & -7 \end{pmatrix},$$

whilst

$$BA = \begin{pmatrix} 3 & -1 \\ 1 & -4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} -6 & 5 \\ -13 & -2 \end{pmatrix}.$$

Note that this gives three reasons why matrix multiplication is not commutative. Sometimes the product AB is defined and yet the product BA is not defined. Sometimes both products are defined and yet they don't have the same shape. Finally even when both products are defined and have the same shape, it is not necessarily the case that the two products are equal.

This might seem surprising until one realises it is really an instance of the fact that composition of functions is not commutative. (For an example in one variable, consider the two functions $f(x) = 2x$ and $g(x) = \sin x$. Then

$$(g \circ f)(x) = \sin 2x \neq 2 \sin x = (f \circ g)(x).)$$

It is interesting to consider how to various functions are represented as matrices. Let's consider the case of functions

$$\phi \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2.$$

We have already seen that the matrix

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

represents the identity function $\phi(x, y) = (x, y)$. It is not hard to see that

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

represents the function $\phi(x, y) = (y, x)$, which switches the x and y-axis. It is given by reflection in the line $y = x$. If we repeat this function twice then we get the identity function. In other words the matrix squares to the identity. The matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

represents the function $\phi(x, y) = (-x, -y)$, rotation through π. For the same reasons as before, it follows that this matrix squares to the identity. Now suppose that θ is an angle and consider the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The corresponding transformation is $\phi(x, y) = (\cos\theta x - \sin\theta y, \sin\theta x + \cos\theta y)$, which, with a little coordinate geometry is seen to be rotation through an angle of θ. For example, suppose $\theta = \pi/2$. The matrix becomes

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and the transformation is $\phi(x, y) = (-y, x)$. It is also instructive to check the identity

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix},$$

which expresses the fact that rotation through θ followed by rotation through ϕ is the same as rotation through $\theta + \phi$.

Definition 3.3. *Let n be a positive integer and let I_n be the $n \times n$ square matrix whose (i, j) entry is*

$$\begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

*If A is a square $n \times n$ matrix then we say that B is the **inverse** of A if $AB = BA = I_n$. We say that A **invertible** if A has an inverse.*

Let A be an $m \times n$ matrix. Note that $I_m A = A$ and $AI_n = A$. For this reason I_n is called the identity matrix (of type $n \times n$).

Note that all the transformations above are bijective and so they have inverses, which are easy to write down.

Theorem 3.4. *Matrix multiplication is associative. That is, if A is a $m \times n$ matrix, B is a $n \times p$ matrix and C is a $p \times q$ matrix then*

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

Proof. If one accepts that matrices correspond to functions then this is clear, since composition of functions is associative. But let us check this using meaningless symbols.

Suppose $A = (a_{ij})$, $B = (b_{ij})$ and $C = (c_{ij})$. The (i, j) entry of AB is then

$$\sum_l a_{il} b_{lj}.$$

It follows that the (i, j) entry of $(AB)C$ is

$$\sum_k \left(\sum_l a_{il} b_{lk} \right) c_{kj} = \sum_{k,l} a_{il} b_{lk} c_{kj}.$$

To get the last equality we distributed multiplication over addition. So much for the LHS.

On the other hand the (i, j) entry of BC is

$$\sum_l b_{il} c_{lj}.$$

It follows that the (i, j) entry of $A(BC)$ is

$$\sum_k a_{ik} \left(\sum_l b_{kl} c_{lj} \right) = \sum_{k,l} a_{ik} b_{kl} c_{lj}.$$

Since k and l are dummy variables in both sums, switching k and l in one sum, we see that both sides are in fact the same. $\square$