Matrices and Linear Mappings

Suppose A is an $n\times k$ matrix over $\mathbf{F}$ and $X=[x_1,x_2, \dots ,x_k]$ is a single-row matrix representing a vector from $\mathbf{F}^k$. By the definition of the matrix multiplication, AX^T is a single-column matrix representing a vector from $\mathbf{F}^n$. Hence if we fix A and let X vary over $\mathbf{F}^k$ we have a function φ_A: $\mathbf{F}^k \to \mathbf{F}^n$ defined as $\varphi_A(X)=AX^T$.

Theorem 1.1. For every $n\times k$ matrix A the function $\varphi_A(X)=AX^T$ is a linear mapping of $\mathbf{F}^k$ into $\mathbf{F}^n$.

Proof. The i-th component of the vector $\varphi_A(pX)$ is $a_{i,1}px_1+a_{i,2}px_2+ \dots +a_{i,k}px_k = p(a_{i,1}x_1+a_{i,2}x_2+ \dots +a_{i,k}x_k)$. Since $a_{i,1}x_1+a_{i,2}x_2+ \dots +a_{i,k}x_k$ is the i-th component of $\varphi_A(X)$, we have proved that φ_A preserves scaling. Suppose $Y=[y_1,y_2, \dots ,y_k]$ and consider $\varphi_A(X+Y) = A((X+Y)^T) = A[x_1+y_1, x_2+y_2, \dots , x_k+y_k]^T$. By the definition of the matrix product, the i-th component of $\varphi_A(X+Y)$ is equal to $a_{i,1}(x_1+y_1)+a_{i,2}(x_2+y_2) \dots +a_{i,k}(x_k+y_k) = a_{i,1}x_1+a_{i,2}x_2+ \dots +a_{i,k}x_k + a_{i,1}y_1+a_{i,2}y_2+ \dots +a_{i,k}y_k$ and this is the i-th component of $\varphi_A(X)+ \varphi_A(Y)$. This shows that φ_A preserves addition.□

In case of coordinate vectors we will often leave out the transposition symbol ^{T} allowing the context to decide whether the sequence of scalars should be written horizontally or vertically.

Corollary. For every $n\times k$ matrix A and for every two $k\times s$ matrices B and C, A(B+C)=AB+AC, i.e. matrix multiplication is distributive with respect to matrix addition.

Proof. From the definition of matrix multiplication the i-th column of A(B+C) is the product of A and the i-th column of B+C. Since the i-th column of B+C is the sum of the i-th columns of B and C, by Theorem 1.1 we have that the i-th column of A(B+C) is equal to the i-th column of AB+AC.□

Theorem 1.1 says that with every $n\times k$ matrix we can associate a linear mapping from $\mathbf{F}^k$ into $\mathbf{F}^n$. The following definition shows how to assign a matrix to a linear mapping.

Definition 1.2. Let us consider a linear mapping $\varphi:\mathbf{F}^k\to\mathbf{F}^n$ and let us choose bases $R=\{v_1,v_2, \dots ,v_k\}$ for $\mathbf{F}^k$ and $S=\{w_1,w_2, \dots ,w_n\}$ for $\mathbf{F}^n$. For each vector $v_i\in R$ there exist unique scalars

$a_{1,i}$, $a_{2,i}$, … , $a_{n,i}$ such that $\varphi(v_i)=a_{1,i}w_1+a_{2,i}w_2 + \ldots + a_{n,i}w_n$. The $k\times n$ matrix $M_S^R(\varphi) = [a_{i,j}]$ is called the matrix of φ in bases R to S.

Remark 1. In other words, for each i, the i-th column of the matrix $M_S^R(\varphi)$ is equal to $[\varphi(v_i)]_S$.

Remark 2. The matrix assigned to a linear mapping depends on the choice of the bases R and S but the size of the matrix depends only on the dimension of the domain and the range of φ, namely the number of columns is equal to the dimension of the domain and the number of rows to the dimension of the range of φ. If we choose another pair of bases we will obtain another matrix for the same linear mapping. Also, two different linear mappings may have the same matrix but with respect to two pairs of bases.

Example 1.2. Let $\varphi:\mathbf{R}^3\to\mathbf{R}^2$, φ(x,y,z) = (x+2y−2z,3x−y+2z). We will construct the matrix for φ in bases R={(0,1,1),(1,0,1),(1,1,0)} and S={(1,1),(1,0)}. We calculate the values of φ on vectors from R and represent them as linear combinations of vectors from S.

φ(0,1,1) = (0,1) =1(1,1)+(−1)(1,0),

φ(1,0,1) = (−1,5) = 5(1,1)+(−6)(1,0) and

φ(1,1,0) = (3,2) = 2(1,1)+1(1,0).

Finally we form the matrix placing the coefficients of the linear combinations in consecutive columns $M_S^R(\varphi) = \begin{bmatrix} 1 & 5 & 2 \\ -1 & -6 & 1 \end{bmatrix}$.

Recall that $[v]_R$ denotes the coordinate vector of v with respect to a basis R. The following theorem, states that matrices and linear mappings are essentially the same thing, or rather, there is a 1-1 correspondence between matrices and linear mappings.

Theorem 1.2. Given a linear mapping $\varphi:\mathbf{F}^k\to\mathbf{F}^n$ and bases S={v_1,v_2, … ,v_k} for $\mathbf{F}^k$, and R={w_1,w_2, … ,w_n} for $\mathbf{F}^n$, for every vector x from $\mathbf{F}^k$ and for every k×n matrix A we have $A[x]_S = [\varphi(x)]_R$ if and only if $A = M_R^S(\varphi)$

Proof. (⇐) Suppose that $A = M_R^S(\varphi)$=[$a_{i,j}$] and $[x]_S$=[x_1,x_2, … ,x_k]. We have φ(x) = $\varphi(\sum_{s=1}^{k} x_s v_s) = \sum_{s=1}^{k} x_s\varphi(v_s) = \sum_{s=1}^{k} x_s(\sum_{t=1}^{n} a_{t,s}w_t) = \sum_{s=1}^{k}\sum_{t=1}^{n} x_s a_{t,s} w_t = \sum_{t=1}^{n}(\sum_{s=1}^{k} x_s a_{t,s})w_t$ (the proof of the last equality is left to the reader) which means that the t-th coordinate of $[\varphi(x)]_R$ = $\sum_{s=1}^{k} x_s a_{t,s}$. On the other hand, the t-th coordinate of $M_R^S(\varphi)[x]_S$ is also equal to $\sum_{r=1}^{k} a_{t,r} x_r$.

($\Rightarrow$) Now suppose that for every x from $\mathbf{F}^k$, $A[x]_S = [\varphi(x)]_R$. Let us put x=v_i. Then $[v_i]_S$ = [0, … ,1, … ,0]. The only 1 in the sequence is in the i-th place. Hence $A[v_i]_S$ is equal to the i-th column of A. By the first remark following Definition 1.2 we are done.□

Theorem 1.3. If φ:$\mathbf{F}^k \to \mathbf{F}^n$ and ψ:$\mathbf{F}^n \to \mathbf{F}^p$ are linear mappings, and S,R,T are bases for $\mathbf{F}^k$,$\mathbf{F}^n$,$\mathbf{F}^p$, respectively, then $M_T^S(\psi \circ \varphi) = M_T^R(\psi) M_R^S(\varphi)$.

Proof. This is an immediate consequence of Theorem 1.2.□

Theorem 1.4. Matrix multiplication is associative, i.e. for every three matrices A,B and C for whom the products exist A(BC) = (AB)C.

Proof. It follows immediately from Theorem 1.3 and the fact that function composition is associative.□

Theorem 1.5. For every two bases R and S of $\boldsymbol{F}^n$ and for every linear operator φ on $\boldsymbol{F}^n$ (a linear operator is a linear mapping which maps a vector space into itself) we have $M_R^R(\varphi) = M_R^S(id) M_S^S(\varphi) M_S^R(id)$.

Proof. Applying twice Theorem 1.3 to $M_R^S(id) M_S^S(\varphi) M_S^R(id)$ we get $M_R^R((id \circ \varphi) \circ id)$ which is the same as $M_R^R(\varphi)$. □