Chapter 4

Matrices and Linear Maps

In this chapter, all vector spaces are defined over an arbitrary field K. For the sake of concreteness, the reader may safely assume that $K = \mathbb{R}$.

4.1 Representation of Linear Maps by Matrices

Proposition 3.18 shows that given two vector spaces E and F and a basis $(u_j)_{j\in J}$ of E, every linear map $f\colon E \to F$ is uniquely determined by the family $(f(u_j))_{j\in J}$ of the images under f of the vectors in the basis $(u_j)_{j\in J}$.

If we also have a basis $(v_i)_{i\in I}$ of F, then every vector $f(u_j)$ can be written in a unique way as

$$f(u_j) = \sum_{i\in I} a_{i\,j} v_i,$$

where $j \in J$, for a family of scalars $(a_{i\,j})_{i\in I}$. Thus, with respect to the two bases $(u_j)_{j\in J}$ of E and $(v_i)_{i\in I}$ of F, the linear map f is completely determined by a "$I \times J$-matrix" $M(f) = (a_{i\,j})_{(i,j)\in I\times J}$.

Remark: Note that we intentionally assigned the index set J to the basis $(u_j)_{j\in J}$ of E, and the index set I to the basis $(v_i)_{i\in I}$ of F, so that the rows of the matrix $M(f)$ associated with $f\colon E \to F$ are indexed by I, and the columns of the matrix $M(f)$ are indexed by J. Obviously, this causes a mildly unpleasant reversal. If we had considered the bases $(u_i)_{i\in I}$ of E and $(v_j)_{j\in J}$ of F, we would obtain a $J \times I$-matrix $M(f) = (a_{j\,i})_{(j,i)\in J\times I}$. No matter what we do, there will be a reversal! We decided to stick to the bases $(u_j)_{j\in J}$ of E and $(v_i)_{i\in I}$ of F, so that we get an $I \times J$-matrix $M(f)$, knowing that we may occasionally suffer from this decision!

When I and J are finite, and say, when $|I| = m$ and $|J| = n$, the linear map f is determined by the matrix $M(f)$ whose entries in the j-th column are the components of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, that is, the matrix

$$M(f) = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix}$$

whose entry on Row i and Column j is $a_{i\,j}$ ($1 \leq i \leq m$, $1 \leq j \leq n$).

We will now show that when E *and* F *have finite dimension, linear maps can be very conveniently represented by matrices, and that composition of linear maps corresponds to matrix multiplication.* We will follow rather closely an elegant presentation method due to Emil Artin.

Let E and F be two vector spaces, and assume that E has a finite basis $(u_1, \ldots, u_n)$ and that F has a finite basis $(v_1, \ldots, v_m)$. Recall that we have shown that every vector $x \in E$ can be written in a unique way as

$$x = x_1 u_1 + \cdots + x_n u_n,$$

and similarly every vector $y \in F$ can be written in a unique way as

$$y = y_1 v_1 + \cdots + y_m v_m.$$

Let $f\colon E \to F$ be a linear map between E and F. Then for every $x = x_1 u_1 + \cdots + x_n u_n$ in E, by linearity, we have

$$f(x) = x_1 f(u_1) + \cdots + x_n f(u_n).$$

Let

$$f(u_j) = a_{1\,j} v_1 + \cdots + a_{m\,j} v_m,$$

or more concisely,

$$f(u_j) = \sum_{i=1}^{m} a_{i\,j} v_i,$$

for every j, $1 \leq j \leq n$. This can be expressed by writing the coefficients $a_{1j}, a_{2j}, \ldots, a_{mj}$ of $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, as the jth column of a matrix, as shown below:

$$\begin{array}{c c} & \begin{array}{cccc} f(u_1) & f(u_2) & \ldots & f(u_n) \end{array} \\ \begin{array}{c} v_1 \\ v_2 \\ \vdots \\ v_m \end{array} & \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}.$$

Then substituting the right-hand side of each $f(u_j)$ into the expression for $f(x)$, we get

$$f(x) = x_1(\sum_{i=1}^{m} a_{i\,1} v_i) + \cdots + x_n(\sum_{i=1}^{m} a_{i\,n} v_i),$$

which, by regrouping terms to obtain a linear combination of the v_i, yields

$$f(x) = (\sum_{j=1}^{n} a_{1\,j} x_j) v_1 + \cdots + (\sum_{j=1}^{n} a_{m\,j} x_j) v_m.$$

Thus, letting $f(x) = y = y_1 v_1 + \cdots + y_m v_m$, we have

$$y_i = \sum_{j=1}^{n} a_{i\,j} x_j \tag{1}$$

for all i, $1 \leq i \leq m$.

To make things more concrete, let us treat the case where $n = 3$ and $m = 2$. In this case,

$$\begin{aligned} f(u_1) &= a_{11} v_1 + a_{21} v_2 \\ f(u_2) &= a_{12} v_1 + a_{22} v_2 \\ f(u_3) &= a_{13} v_1 + a_{23} v_2, \end{aligned}$$

which in matrix form is expressed by

$$\begin{array}{c c} & \begin{array}{ccc} f(u_1) & f(u_2) & f(u_3) \end{array} \\ \begin{array}{c} v_1 \\ v_2 \end{array} & \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \end{array},$$

and for any $x = x_1 u_1 + x_2 u_2 + x_3 u_3$, we have

$$\begin{aligned} f(x) &= f(x_1 u_1 + x_2 u_2 + x_3 u_3) \\ &= x_1 f(u_1) + x_2 f(u_2) + x_3 f(u_3) \\ &= x_1 (a_{11} v_1 + a_{21} v_2) + x_2 (a_{12} v_1 + a_{22} v_2) + x_3 (a_{13} v_1 + a_{23} v_2) \\ &= (a_{11} x_1 + a_{12} x_2 + a_{13} x_3) v_1 + (a_{21} x_1 + a_{22} x_2 + a_{23} x_3) v_2. \end{aligned}$$

Consequently, since

$$y = y_1 v_1 + y_2 v_2,$$

we have

$$\begin{aligned} y_1 &= a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \\ y_2 &= a_{21} x_1 + a_{22} x_2 + a_{23} x_3. \end{aligned}$$

This agrees with the matrix equation

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

We now formalize the representation of linear maps by matrices.

Definition 4.1. Let E and F be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for E, and $(v_1, \ldots, v_m)$ be a basis for F. Each vector $x \in E$ expressed in the basis $(u_1, \ldots, u_n)$ as $x = x_1 u_1 + \cdots + x_n u_n$ is represented by the column matrix

$$M(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and similarly for each vector $y \in F$ expressed in the basis $(v_1, \ldots, v_m)$.

Every linear map $f \colon E \to F$ is represented by the matrix $M(f) = (a_{i\,j})$, where $a_{i\,j}$ is the i-th component of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, i.e., where

$$f(u_j) = \sum_{i=1}^{m} a_{i\,j} v_i, \quad \text{for every } j,\ 1 \leq j \leq n.$$

The coefficients $a_{1j}, a_{2j}, \ldots, a_{mj}$ of $f(u_j)$ over the basis $(v_1, \ldots, v_m)$ form the jth column of the matrix $M(f)$ shown below:

$$\begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_m \end{array} \begin{array}{c} \begin{array}{cccc} f(u_1) & f(u_2) & \ldots & f(u_n) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}.$$

The matrix $M(f)$ associated with the linear map $f \colon E \to F$ is called the *matrix of f with respect to the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$*. When $E = F$ and the basis $(v_1, \ldots, v_m)$ is identical to the basis $(u_1, \ldots, u_n)$ of E, the matrix $M(f)$ associated with $f \colon E \to E$ (as above) is called the *matrix of f with respect to the basis $(u_1, \ldots, u_n)$*.

Remark: As in the remark after Definition 3.12, there is no reason to assume that the vectors in the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ are ordered in any particular way. However, it is often convenient to assume the natural ordering. When this is so, authors sometimes refer to the matrix $M(f)$ as the matrix of f with respect to the *ordered bases* $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$.

Let us illustrate the representation of a linear map by a matrix in a concrete situation. Let E be the vector space $\mathbb{R}[X]_4$ of polynomials of degree at most 4, let F be the vector space $\mathbb{R}[X]_3$ of polynomials of degree at most 3, and let the linear map be the derivative map d: that is,

$$\begin{aligned} d(P + Q) &= dP + dQ \\ d(\lambda P) &= \lambda dP, \end{aligned}$$

with $\lambda \in \mathbb{R}$. We choose $(1, x, x^2, x^3, x^4)$ as a basis of E and $(1, x, x^2, x^3)$ as a basis of F. Then the 4×5 matrix D associated with d is obtained by expressing the derivative dx^i of each basis vector x^i for $i = 0, 1, 2, 3, 4$ over the basis $(1, x, x^2, x^3)$. We find

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

If P denotes the polynomial

$$P = 3x^4 - 5x^3 + x^2 - 7x + 5,$$

we have

$$dP = 12x^3 - 15x^2 + 2x - 7.$$

The polynomial P is represented by the vector $(5, -7, 1, -5, 3)$, the polynomial dP is represented by the vector $(-7, 2, -15, 12)$, and we have

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ -7 \\ 1 \\ -5 \\ 3 \end{pmatrix} = \begin{pmatrix} -7 \\ 2 \\ -15 \\ 12 \end{pmatrix},$$

as expected! The kernel (nullspace) of d consists of the polynomials of degree 0, that is, the constant polynomials. Therefore $\dim(\operatorname{Ker} d) = 1$, and from

$$\dim(E) = \dim(\operatorname{Ker} d) + \dim(\operatorname{Im} d)$$

(see Theorem 6.16), we get $\dim(\operatorname{Im} d) = 4$ (since $\dim(E) = 5$).

For fun, let us figure out the linear map from the vector space $\mathbb{R}[X]_3$ to the vector space $\mathbb{R}[X]_4$ given by integration (finding the primitive, or anti-derivative) of x^i, for $i = 0, 1, 2, 3$). The 5×4 matrix S representing $\int$ with respect to the same bases as before is

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}.$$

We verify that $DS = I_4$,

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This is to be expected by the fundamental theorem of calculus since the derivative of an integral returns the function. As we will shortly see, the above matrix product corresponds to this functional composition. The equation $DS = I_4$ shows that S is injective and has D as a left inverse. However, $SD \neq I_5$, and instead

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

because constant polynomials (polynomials of degree 0) belong to the kernel of D.

4.2 Composition of Linear Maps and Matrix Multiplication

Let us now consider how the composition of linear maps is expressed in terms of bases.

Let E, F, and G, be three vectors spaces with respective bases $(u_1, \ldots, u_p)$ for E, $(v_1, \ldots, v_n)$ for F, and $(w_1, \ldots, w_m)$ for G. Let $g\colon E \to F$ and $f\colon F \to G$ be linear maps. As explained earlier, $g\colon E \to F$ is determined by the images of the basis vectors u_j, and $f\colon F \to G$ is determined by the images of the basis vectors v_k. We would like to understand how $f \circ g\colon E \to G$ is determined by the images of the basis vectors u_j.

Remark: Note that we are considering linear maps $g\colon E \to F$ and $f\colon F \to G$, instead of $f\colon E \to F$ and $g\colon F \to G$, which yields the composition $f \circ g\colon E \to G$ instead of $g \circ f\colon E \to G$. Our perhaps unusual choice is motivated by the fact that if f is represented by a matrix $M(f) = (a_{i\,k})$ and g is represented by a matrix $M(g) = (b_{k\,j})$, then $f \circ g\colon E \to G$ is represented by the product AB of the matrices A and B. If we had adopted the other choice where $f\colon E \to F$ and $g\colon F \to G$, then $g \circ f\colon E \to G$ would be represented by the product BA. Personally, we find it easier to remember the formula for the entry in Row i and Column j of the product of two matrices when this product is written by AB, rather than BA. Obviously, this is a matter of taste! We will have to live with our perhaps unorthodox choice.

Thus, let

$$f(v_k) = \sum_{i=1}^{m} a_{i\,k} w_i,$$

for every k, $1 \leq k \leq n$, and let

$$g(u_j) = \sum_{k=1}^{n} b_{k\,j} v_k,$$

for every j, $1 \leq j \leq p$; in matrix form, we have

$$\begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{cccc} f(v_1) & f(v_2) & \ldots & f(v_n) \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}$$

and

$$\begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{cccc} g(u_1) & g(u_2) & \ldots & g(u_p) \end{array} \\ \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1p} \\ b_{21} & b_{22} & \ldots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{np} \end{pmatrix} \end{array}.$$

By previous considerations, for every

$$x = x_1u_1 + \cdots + x_pu_p,$$

letting $g(x) = y = y_1v_1 + \cdots + y_nv_n$, we have

$$y_k = \sum_{j=1}^{p} b_{k\,j}x_j \tag{2}$$

for all k, $1 \leq k \leq n$, and for every

$$y = y_1v_1 + \cdots + y_nv_n,$$

letting $f(y) = z = z_1w_1 + \cdots + z_mw_m$, we have

$$z_i = \sum_{k=1}^{n} a_{i\,k}y_k \tag{3}$$

for all i, $1 \leq i \leq m$. Then if $y = g(x)$ and $z = f(y)$, we have $z = f(g(x))$, and in view of (2) and (3), we have

$$\begin{aligned} z_i &= \sum_{k=1}^{n} a_{i\,k}(\sum_{j=1}^{p} b_{k\,j}x_j) \\ &= \sum_{k=1}^{n}\sum_{j=1}^{p} a_{i\,k}b_{k\,j}x_j \\ &= \sum_{j=1}^{p}\sum_{k=1}^{n} a_{i\,k}b_{k\,j}x_j \\ &= \sum_{j=1}^{p}(\sum_{k=1}^{n} a_{i\,k}b_{k\,j})x_j. \end{aligned}$$

Thus, defining $c_{i\,j}$ such that

$$c_{i\,j} = \sum_{k=1}^{n} a_{i\,k} b_{k\,j},$$

for $1 \leq i \leq m$, and $1 \leq j \leq p$, we have

$$z_i = \sum_{j=1}^{p} c_{i\,j} x_j \tag{4}$$

Identity (4) shows that the composition of linear maps corresponds to the product of matrices.

Then given a linear map $f\colon E \to F$ represented by the matrix $M(f) = (a_{i\,j})$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, by Equation (1), namely

$$y_i = \sum_{j=1}^{n} a_{i\,j} x_j \quad 1 \leq i \leq m,$$

and the definition of matrix multiplication, the equation $y = f(x)$ corresponds to the matrix equation $M(y) = M(f)M(x)$, that is,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{m\,1} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Recall that

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{1\,1} \\ a_{2\,1} \\ \vdots \\ a_{m\,1} \end{pmatrix} + x_2 \begin{pmatrix} a_{1\,2} \\ a_{2\,2} \\ \vdots \\ a_{m\,2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1\,n} \\ a_{2\,n} \\ \vdots \\ a_{m\,n} \end{pmatrix}.$$

Sometimes, it is necessary to incorporate the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ in the notation for the matrix $M(f)$ expressing f with respect to these bases. This turns out to be a messy enterprise!

We propose the following course of action:

Definition 4.2. Write $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_m)$ for the bases of E and F, and denote by $M_{\mathcal{U},\mathcal{V}}(f)$ the *matrix of f with respect to the bases $\mathcal{U}$ and $\mathcal{V}$*. Furthermore, write $x_{\mathcal{U}}$ for the coordinates $M(x) = (x_1, \ldots, x_n)$ of $x \in E$ w.r.t. the basis $\mathcal{U}$ and write $y_{\mathcal{V}}$ for the coordinates $M(y) = (y_1, \ldots, y_m)$ of $y \in F$ w.r.t. the basis $\mathcal{V}$. Then

$$y = f(x)$$

is expressed in matrix form by

$$y_{\mathcal{V}} = M_{\mathcal{U},\mathcal{V}}(f)\, x_{\mathcal{U}}.$$

When $\mathcal{U} = \mathcal{V}$, we abbreviate $M_{\mathcal{U},\mathcal{V}}(f)$ as $M_{\mathcal{U}}(f)$.

The above notation seems reasonable, but it has the slight disadvantage that in the expression $M_{\mathcal{U},\mathcal{V}}(f)x_{\mathcal{U}}$, the input argument $x_{\mathcal{U}}$ which is fed to the matrix $M_{\mathcal{U},\mathcal{V}}(f)$ does not appear next to the subscript $\mathcal{U}$ in $M_{\mathcal{U},\mathcal{V}}(f)$. We could have used the notation $M_{\mathcal{V},\mathcal{U}}(f)$, and some people do that. But then, we find a bit confusing that $\mathcal{V}$ comes before $\mathcal{U}$ when f maps from the space E with the basis $\mathcal{U}$ to the space F with the basis $\mathcal{V}$. So, we prefer to use the notation $M_{\mathcal{U},\mathcal{V}}(f)$.

Be aware that other authors such as Meyer [125] use the notation $[f]_{\mathcal{U},\mathcal{V}}$, and others such as Dummit and Foote [54] use the notation $M_{\mathcal{U}}^{\mathcal{V}}(f)$, instead of $M_{\mathcal{U},\mathcal{V}}(f)$. This gets worse! You may find the notation $M_{\mathcal{V}}^{\mathcal{U}}(f)$ (as in Lang [109]), or ${}_{\mathcal{U}}[f]_{\mathcal{V}}$, or other strange notations.

Definition 4.2 shows that the function which associates to a linear map $f\colon E \to F$ the matrix $M(f)$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ has the property that matrix multiplication corresponds to composition of linear maps. This allows us to transfer properties of linear maps to matrices. Here is an illustration of this technique:

Proposition 4.1. *(1) Given any matrices $A \in \mathrm{M}_{m,n}(K)$, $B \in \mathrm{M}_{n,p}(K)$, and $C \in \mathrm{M}_{p,q}(K)$, we have*

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

that is, matrix multiplication is associative.

(2) Given any matrices $A, B \in \mathrm{M}_{m,n}(K)$, and $C, D \in \mathrm{M}_{n,p}(K)$, for all $\lambda \in K$, we have

$$\begin{aligned} (A + B)C &= AC + BC \\ A(C + D) &= AC + AD \\ (\lambda A)C &= \lambda(AC) \\ A(\lambda C) &= \lambda(AC), \end{aligned}$$

so that matrix multiplication $\cdot\colon \mathrm{M}_{m,n}(K) \times \mathrm{M}_{n,p}(K) \to \mathrm{M}_{m,p}(K)$ is bilinear.

Proof. (1) Every $m \times n$ matrix $A = (a_{i\,j})$ defines the function $f_A\colon K^n \to K^m$ given by

$$f_A(x) = Ax,$$

for all $x \in K^n$. It is immediately verified that f_A is linear and that the matrix $M(f_A)$ representing f_A over the canonical bases in K^n and K^m is equal to A. Then Formula (4) proves that

$$M(f_A \circ f_B) = M(f_A)M(f_B) = AB,$$

so we get

$$M((f_A \circ f_B) \circ f_C) = M(f_A \circ f_B)M(f_C) = (AB)C$$

and

$$M(f_A \circ (f_B \circ f_C)) = M(f_A)M(f_B \circ f_C) = A(BC),$$

and since composition of functions is associative, we have $(f_A \circ f_B) \circ f_C = f_A \circ (f_B \circ f_C)$, which implies that

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

(2) It is immediately verified that if $f_1, f_2 \in \mathrm{Hom}_K(E, F)$, $A, B \in \mathrm{M}_{m,n}(K)$, $(u_1, \ldots, u_n)$ is any basis of E, and $(v_1, \ldots, v_m)$ is any basis of F, then

$$\begin{aligned} M(f_1 + f_2) &= M(f_1) + M(f_2) \\ f_{A+B} &= f_A + f_B. \end{aligned}$$

Then we have

$$\begin{aligned} (A + B)C &= M(f_{A+B})M(f_C) \\ &= M(f_{A+B} \circ f_C) \\ &= M((f_A + f_B) \circ f_C)) \\ &= M((f_A \circ f_C) + (f_B \circ f_C)) \\ &= M(f_A \circ f_C) + M(f_B \circ f_C) \\ &= M(f_A)M(f_C) + M(f_B)M(f_C) \\ &= AC + BC. \end{aligned}$$

The equation $A(C + D) = AC + AD$ is proven in a similar fashion, and the last two equations are easily verified. We could also have verified all the identities by making matrix computations. $\square$

Note that Proposition 4.1 implies that the vector space $\mathrm{M}_n(K)$ of square matrices is a (noncommutative) ring with unit I_n. (It even shows that $\mathrm{M}_n(K)$ is an associative *algebra*.)

The following proposition states the main properties of the mapping $f \mapsto M(f)$ between $\mathrm{Hom}(E, F)$ and $\mathrm{M}_{m,n}$. In short, it is an isomorphism of vector spaces.

Proposition 4.2. *Given three vector spaces E, F, G, with respective bases $(u_1, \ldots, u_p)$, $(v_1, \ldots, v_n)$, and $(w_1, \ldots, w_m)$, the mapping $M\colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ that associates the matrix $M(g)$ to a linear map $g\colon E \to F$ satisfies the following properties for all $x \in E$, all $g, h\colon E \to F$, and all $f\colon F \to G$:*

$$\begin{aligned} M(g(x)) &= M(g)M(x) \\ M(g + h) &= M(g) + M(h) \\ M(\lambda g) &= \lambda M(g) \\ M(f \circ g) &= M(f)M(g), \end{aligned}$$

where $M(x)$ is the column vector associated with the vector x and $M(g(x))$ is the column vector associated with $g(x)$, as explained in Definition 4.1.

Thus, $M\colon \mathrm{Hom}(E,F) \to \mathrm{M}_{n,p}$ is an isomorphism of vector spaces, and when $p = n$ and the basis $(v_1,\ldots,v_n)$ is identical to the basis $(u_1,\ldots,u_p)$, $M\colon \mathrm{Hom}(E,E) \to \mathrm{M}_n$ is an isomorphism of rings.

Proof. That $M(g(x)) = M(g)M(x)$ was shown by Definition 4.2 or equivalently by Formula (1). The identities $M(g+h) = M(g) + M(h)$ and $M(\lambda g) = \lambda M(g)$ are straightforward, and $M(f \circ g) = M(f)M(g)$ follows from Identity (4) and the definition of matrix multiplication. The mapping $M\colon \mathrm{Hom}(E,F) \to \mathrm{M}_{n,p}$ is clearly injective, and since every matrix defines a linear map (see Proposition 4.1), it is also surjective, and thus bijective. In view of the above identities, it is an isomorphism (and similarly for $M\colon \mathrm{Hom}(E,E) \to \mathrm{M}_n$, where Proposition 4.1 is used to show that M_n is a ring). $\square$

In view of Proposition 4.2, it seems preferable to represent vectors from a vector space of finite dimension as column vectors rather than row vectors. *Thus, from now on, we will denote vectors of $\mathbb{R}^n$ (or more generally, of K^n) as column vectors.*

We explained in Section 3.9 that if the space E is finite-dimensional and has a finite basis $(u_1,\ldots,u_n)$, then a linear form $f^*\colon E \to K$ is represented by the *row vector* of coefficients

$$\begin{pmatrix} f^*(u_1) & \cdots & f^*(u_n) \end{pmatrix}, \tag{1}$$

over the bases $(u_1,\ldots,u_n)$ and 1 (in K), and that over the dual basis $(u_1^*,\ldots,u_n^*)$ of E^*, the linear form f^* is represented by the same coefficients, but as the *column vector*

$$\begin{pmatrix} f^*(u_1) \\ \vdots \\ f^*(u_n) \end{pmatrix}, \tag{2}$$

which is the transpose of the row vector in (1).

This is a special case of a more general phenomenon. A linear map $f\colon E \to F$ induces a map $f^\top\colon F^* \to E^*$ called the *transpose* of f (note that $f^\top$ maps F^* to E^*, *not* E^* to F^*), and if $(u_1\ldots,u_n)$ is a basis of E, $(v_1\ldots,v_m)$ is a basis of F, and if f is represented by the $m \times n$ matrix A over these bases, then over the dual bases $(v_1^*,\ldots,v_m^*)$ and $(u_1^*,\ldots,u_n^*)$, the linear map $f^\top$ is represented by $A^\top$, the transpose of the matrix A.

This is because over the basis $(v_1,\ldots,v_m)$, a linear form $\varphi \in F^*$ is represented by the row vector

$$\lambda = \begin{pmatrix} \varphi(v_1) & \cdots & \varphi(v_m) \end{pmatrix},$$

and we define $f^\top(\varphi)$ as the linear form represented by the row vector

$$\lambda A$$

over the basis $(u_1, \ldots, u_n)$. Since φ is represented by the column vector $\lambda^\top$ over the dual basis $(v_1^*, \ldots, v_m^*)$, we see that $f^\top(\varphi)$ is represented by the column vector

$$(\lambda A)^\top = A^\top \lambda^\top$$

over the dual basis $(u_1^*, \ldots, u_n^*)$. The matrix defining $f^\top$ over the dual bases $(v_1^*, \ldots, v_m^*)$ and $(u_1^*, \ldots, u_n^*)$ is indeed $A^\top$.

Conceptually, we will show later (see Section 30.1) that the linear map $f^\top \colon F^* \to E^*$ is defined by

$$f^\top(\varphi) = \varphi \circ f,$$

for all $\varphi \in F^*$ (remember that $\varphi \colon F \to K$, so composing $f \colon E \to F$ and $\varphi \colon F \to K$ yields a linear form $\varphi \circ f \colon E \to K$).

4.3 Change of Basis Matrix

It is important to observe that the isomorphism $M \colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ given by Proposition 4.2 depends on the choice of the bases $(u_1, \ldots, u_p)$ and $(v_1, \ldots, v_n)$, and similarly for the isomorphism $M \colon \mathrm{Hom}(E, E) \to \mathrm{M}_n$, which depends on the choice of the basis $(u_1, \ldots, u_n)$. Thus, it would be useful to know how a change of basis affects the representation of a linear map $f \colon E \to F$ as a matrix. The following simple proposition is needed.

Proposition 4.3. *Let E be a vector space, and let $(u_1, \ldots, u_n)$ be a basis of E. For every family $(v_1, \ldots, v_n)$, let $P = (a_{i\,j})$ be the matrix defined such that $v_j = \sum_{i=1}^n a_{i\,j} u_i$. The matrix P is invertible iff $(v_1, \ldots, v_n)$ is a basis of E.*

Proof. Note that we have $P = M(f)$, the matrix (with respect to the basis $(u_1, \ldots, u_n)$) associated with the unique linear map $f \colon E \to E$ such that $f(u_i) = v_i$. By Proposition 3.18, f is bijective iff $(v_1, \ldots, v_n)$ is a basis of E. Furthermore, it is obvious that the identity matrix I_n is the matrix associated with the identity $\mathrm{id} \colon E \to E$ w.r.t. any basis. If f is an isomorphism, then $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}$, and by Proposition 4.2, we get $M(f)M(f^{-1}) = M(f^{-1})M(f) = I_n$, showing that P is invertible and that $M(f^{-1}) = P^{-1}$. $\square$

An important corollary of Proposition 4.3 yields the following criterion for a square matrix to be invertible. This criterion was already proven in Proposition 3.14 but Proposition 4.3 yields a shorter proof.

Proposition 4.4. *A square matrix $A \in \mathrm{M}_n(K)$ is invertible iff its columns $(A^1, \ldots, A^n)$ are linearly independent.*

Proof. First assume that A is invertible. If $\lambda_1 A^1 + \cdots + \lambda_n A^n = 0$ for some $\lambda_1, \ldots, \lambda_n \in K$, then

$$A\lambda = \lambda_1 A^1 + \cdots + \lambda_n A^n = 0,$$

where λ is the column vector $\lambda = (\lambda_1, \ldots, \lambda_n)$. Since A has an inverse A^{-1}, by multiplying both sides of the equation $A\lambda = 0$ by A^{-1} we obtain

$$A^{-1}A\lambda = I_n\lambda = \lambda = A^{-1}0 = 0,$$

which shows that the columns $(A^1, \ldots, A^n)$ are linearly independent.

Conversely, assume that the columns $(A^1, \ldots, A^n)$ are linearly independent. Since the vector space $E = K^n$ has dimension n, the vectors $(v_1, \ldots, v_n) = (A^1, \ldots, A^n)$ form a basis of K^n. By definition, the matrix A is defined by expressing each vector $v_j = A^j$ as the linear combination $A^j = \sum_{i=1}^n a_{ij}e_i$, where $(e_1, \ldots, e_n)$ is the canonical basis of K^n, and since $(v_1, \ldots, v_n)$ is a basis, by Proposition 4.3, the matrix A is invertible. $\square$

Proposition 4.3 suggests the following definition.

Definition 4.3. Given a vector space E of dimension n, for any two bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ of E, let $P = (a_{i\,j})$ be the invertible matrix defined such that

$$v_j = \sum_{i=1}^n a_{i\,j} u_i,$$

which is also the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, *in that order*. Indeed, we express each $\mathrm{id}(v_j) = v_j$ over the basis $(u_1, \ldots, u_n)$. The coefficients $a_{1j}, a_{2j}, \ldots, a_{nj}$ of v_j over the basis $(u_1, \ldots, u_n)$ form the jth column of the matrix P shown below:

$$\begin{array}{c c}
 & \begin{array}{cccc} v_1 & v_2 & \ldots & v_n \end{array} \\
\begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} & \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}
\end{array}.$$

The matrix P is called the *change of basis matrix from* $(u_1, \ldots, u_n)$ *to* $(v_1, \ldots, v_n)$.

Clearly, the change of basis matrix from $(v_1, \ldots, v_n)$ to $(u_1, \ldots, u_n)$ is P^{-1}. Since $P = (a_{i\,j})$ is the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, given any vector $x \in E$, if $x = x_1u_1 + \cdots + x_nu_n$ over the basis $(u_1, \ldots, u_n)$ and $x = x'_1v_1 + \cdots + x'_nv_n$ over the basis $(v_1, \ldots, v_n)$, from Proposition 4.2, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{n\,1} & \ldots & a_{n\,n} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix},$$

showing that the *old* coordinates (x_i) of x (over $(u_1, \ldots, u_n)$) are expressed in terms of the *new* coordinates (x'_i) of x (over $(v_1, \ldots, v_n)$). This fact may seem wrong but it is correct as we can reassure ourselves by doing the following computation. Suppose that $n = 2$, so that

$$\begin{aligned} v_1 &= a_{11}u_1 + a_{21}u_2 \\ v_2 &= a_{12}u_1 + a_{22}u_2, \end{aligned}$$

and our matrix is

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The same vector x is written as

$$x = x_1u_1 + x_2u_2 = x'_1v_1 + x'_2v_2,$$

so by substituting the expressions for v_1 and v_2 as linear combinations of u_1 and u_2, we obtain

$$\begin{aligned} x_1u_1 + x_2u_2 &= x'_1v_1 + x'_2v_2 \\ &= x'_1(a_{11}u_1 + a_{21}u_2) + x'_2(a_{12}u_1 + a_{22}u_2) \\ &= (a_{11}x'_1 + a_{12}x'_2)u_1 + (a_{21}x'_1 + a_{22}x'_2)u_2, \end{aligned}$$

and since u_1 and u_2 are linearly independent, we must have

$$\begin{aligned} x_1 &= a_{11}x'_1 + a_{12}x'_2 \\ x_2 &= a_{21}x'_1 + a_{22}x'_2, \end{aligned}$$

namely

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix},$$

as claimed.

If the vectors $u_1, \ldots, u_n$ and the vectors $v_1, \ldots, v_n$ are vectors in K^n, then we can form the $n \times n$ matrix $U = (u_1 \ \cdots \ u_n)$ whose columns are $u_1, \ldots, u_n$ and the $n \times n$ matrix $V = (v_1 \ \cdots \ v_n)$ whose columns are $v_1, \ldots, v_n$. Then we can express the change of basis P from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$ in terms of U and V. Indeeed, the equation

$$v_j = \sum_{i=1}^{n} a_{ij}u_i$$

can be expressed in matrix form as

$$v_j = UA^j,$$

where

$$A^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{n1} \end{pmatrix}$$

is the jth column of P, so we get

$$V = UP,$$

which yields

$$P = U^{-1}V.$$

Now we face the painful task of assigning a "good" notation incorporating the bases $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ into the notation for the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$. Because the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ is the matrix of the identity map id_E *with respect to the bases* $\mathcal{V}$ *and* $\mathcal{U}$ *in that order*, we could denote it by $M_{\mathcal{V},\mathcal{U}}(\mathrm{id})$ (Meyer [125] uses the notation $[I]_{\mathcal{V},\mathcal{U}}$). We prefer to use an abbreviation for $M_{\mathcal{V},\mathcal{U}}(\mathrm{id})$.

Definition 4.4. The *change of basis matrix from* $\mathcal{U}$ *to* $\mathcal{V}$ is denoted

$$P_{\mathcal{V},\mathcal{U}}.$$

Note that

$$P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}.$$

Then, if we write $x_{\mathcal{U}} = (x_1, \ldots, x_n)$ for the *old* coordinates of x with respect to the basis $\mathcal{U}$ and $x_{\mathcal{V}} = (x'_1, \ldots, x'_n)$ for the *new* coordinates of x with respect to the basis $\mathcal{V}$, we have

$$x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}}\, x_{\mathcal{V}}, \quad x_{\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}\, x_{\mathcal{U}}.$$

The above may look backward, but remember that the matrix $M_{\mathcal{U},\mathcal{V}}(f)$ takes input expressed over the basis $\mathcal{U}$ to output expressed over the basis $\mathcal{V}$. Consequently, $P_{\mathcal{V},\mathcal{U}}$ takes input expressed over the basis $\mathcal{V}$ to output expressed over the basis $\mathcal{U}$, and $x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}}\, x_{\mathcal{V}}$ matches this point of view!

Beware that some authors (such as Artin [7]) define the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ as $P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}$. Under this point of view, the old basis $\mathcal{U}$ is expressed in terms of the new basis $\mathcal{V}$. We find this a bit unnatural. Also, in practice, it seems that the new basis is often expressed in terms of the old basis, rather than the other way around.

Since the matrix $P = P_{\mathcal{V},\mathcal{U}}$ expresses the *new* basis $(v_1, \ldots, v_n)$ in terms of the *old* basis $(u_1, \ldots, u_n)$, we observe that the coordinates (x_i) of a vector x vary in the *opposite direction* of the change of basis. For this reason, vectors are sometimes said to be *contravariant*. However, this expression does not make sense! Indeed, a vector in an intrinsic quantity that does not depend on a specific basis. What makes sense is that the *coordinates* of a vector vary in a contravariant fashion.

Let us consider some concrete examples of change of bases.

Example 4.1. Let $E = F = \mathbb{R}^2$, with $u_1 = (1, 0)$, $u_2 = (0, 1)$, $v_1 = (1, 1)$ and $v_2 = (-1, 1)$. The change of basis matrix P from the basis $\mathcal{U} = (u_1, u_2)$ to the basis $\mathcal{V} = (v_1, v_2)$ is

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and its inverse is

$$P^{-1} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

The old coordinates (x_1, x_2) with respect to (u_1, u_2) are expressed in terms of the new coordinates (x_1', x_2') with respect to (v_1, v_2) by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix},$$

and the new coordinates (x_1', x_2') with respect to (v_1, v_2) are expressed in terms of the old coordinates (x_1, x_2) with respect to (u_1, u_2) by

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Example 4.2. Let $E = F = \mathbb{R}[X]_3$ be the set of polynomials of degree at most 3, and consider the bases $\mathcal{U} = (1, x, x^2, x^3)$ and $\mathcal{V} = (B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x))$, where $B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x)$ are the *Bernstein polynomials* of degree 3, given by

$$B_0^3(x) = (1-x)^3 \qquad B_1^3(x) = 3(1-x)^2 x \qquad B_2^3(x) = 3(1-x)x^2 \qquad B_3^3(x) = x^3.$$

By expanding the Bernstein polynomials, we find that the change of basis matrix $P_{\mathcal{V},\mathcal{U}}$ is given by

$$P_{\mathcal{V},\mathcal{U}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{pmatrix}.$$

We also find that the inverse of $P_{\mathcal{V},\mathcal{U}}$ is

$$P_{\mathcal{V},\mathcal{U}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Therefore, the coordinates of the polynomial $2x^3 - x + 1$ over the basis $\mathcal{V}$ are

$$\begin{pmatrix} 1 \\ 2/3 \\ 1/3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix},$$

and so

$$2x^3 - x + 1 = B_0^3(x) + \frac{2}{3} B_1^3(x) + \frac{1}{3} B_2^3(x) + 2B_3^3(x).$$

4.4 The Effect of a Change of Bases on Matrices

The effect of a change of bases on the representation of a linear map is described in the following proposition.

Proposition 4.5. *Let E and F be vector spaces, let $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{U}' = (u'_1, \ldots, u'_n)$ be two bases of E, and let $\mathcal{V} = (v_1, \ldots, v_m)$ and $\mathcal{V}' = (v'_1, \ldots, v'_m)$ be two bases of F. Let $P = P_{\mathcal{U}',\mathcal{U}}$ be the change of basis matrix from $\mathcal{U}$ to $\mathcal{U}'$, and let $Q = P_{\mathcal{V}',\mathcal{V}}$ be the change of basis matrix from $\mathcal{V}$ to $\mathcal{V}'$. For any linear map $f \colon E \to F$, let $M(f) = M_{\mathcal{U},\mathcal{V}}(f)$ be the matrix associated to f w.r.t. the bases $\mathcal{U}$ and $\mathcal{V}$, and let $M'(f) = M_{\mathcal{U}',\mathcal{V}'}(f)$ be the matrix associated to f w.r.t. the bases $\mathcal{U}'$ and $\mathcal{V}'$. We have*

$$M'(f) = Q^{-1}M(f)P,$$

or more explicitly

$$M_{\mathcal{U}',\mathcal{V}'}(f) = P^{-1}_{\mathcal{V}',\mathcal{V}} M_{\mathcal{U},\mathcal{V}}(f) P_{\mathcal{U}',\mathcal{U}} = P_{\mathcal{V},\mathcal{V}'} M_{\mathcal{U},\mathcal{V}}(f) P_{\mathcal{U}',\mathcal{U}}.$$

Proof. Since $f \colon E \to F$ can be written as $f = \mathrm{id}_F \circ f \circ \mathrm{id}_E$, since $P = P_{\mathcal{U}',\mathcal{U}}$ is the matrix of id_E w.r.t. the bases $(u'_1, \ldots, u'_n)$ and $(u_1, \ldots, u_n)$, and $Q^{-1} = P^{-1}_{\mathcal{V}',\mathcal{V}} = P_{\mathcal{V},\mathcal{V}'}$ is the matrix of id_F w.r.t. the bases $(v_1, \ldots, v_m)$ and $(v'_1, \ldots, v'_m)$ as illustrated by the following diagram

$$\begin{tikzcd}
\mathcal{U}, E \arrow[r, "f", "M_{\mathcal{U},\mathcal{V}}(f)"'] & \mathcal{V}, F \arrow[d, "P^{-1}_{\mathcal{V}',\mathcal{V}}"', "\mathrm{id}_F"] \\
\mathcal{U}', E \arrow[u, "P_{\mathcal{U}',\mathcal{U}}", "\mathrm{id}_E"'] \arrow[r, "M_{\mathcal{U}',\mathcal{V}'}(f)", "f"'] & \mathcal{V}', F,
\end{tikzcd}$$

by Proposition 4.2, we have $M'(f) = Q^{-1}M(f)P$. ☐

As a corollary, we get the following result.

Corollary 4.6. *Let E be a vector space, and let $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{U}' = (u'_1, \ldots, u'_n)$ be two bases of E. Let $P = P_{\mathcal{U}',\mathcal{U}}$ be the change of basis matrix from $\mathcal{U}$ to $\mathcal{U}'$. For any linear map $f \colon E \to E$, let $M(f) = M_{\mathcal{U}}(f)$ be the matrix associated to f w.r.t. the basis $\mathcal{U}$, and let $M'(f) = M_{\mathcal{U}'}(f)$ be the matrix associated to f w.r.t. the basis $\mathcal{U}'$. We have*

$$M'(f) = P^{-1}M(f)P,$$

or more explicitly,

$$M_{\mathcal{U}'}(f) = P^{-1}_{\mathcal{U}',\mathcal{U}} M_{\mathcal{U}}(f) P_{\mathcal{U}',\mathcal{U}} = P_{\mathcal{U},\mathcal{U}'} M_{\mathcal{U}}(f) P_{\mathcal{U}',\mathcal{U}},$$

as illustrated by the following diagram

$$
\begin{tikzcd}
\mathcal{U}, E \arrow[r, "f", "M_{\mathcal{U},}(f)"'] & \mathcal{U}, E \arrow[d, "P^{-1}_{\mathcal{U}',\mathcal{U}}"', "\mathrm{id}_E"] \\
\mathcal{U}', E \arrow[u, "P_{\mathcal{U}',\mathcal{U}}", "\mathrm{id}_E"'] \arrow[r, "M_{\mathcal{U}'}(f)", "f"'] & \mathcal{U}', E.
\end{tikzcd}
$$

Example 4.3. Let $E = \mathbb{R}^2$, $\mathcal{U} = (e_1, e_2)$ where $e_1 = (1, 0)$ and $e_2 = (0, 1)$ are the canonical basis vectors, let $\mathcal{V} = (v_1, v_2) = (e_1, e_1 - e_2)$, and let

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.$$

The change of basis matrix $P = P_{\mathcal{V},\mathcal{U}}$ from $\mathcal{U}$ to $\mathcal{V}$ is

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

and we check that

$$P^{-1} = P.$$

Therefore, in the basis $\mathcal{V}$, the matrix representing the linear map f defined by A is

$$A' = P^{-1}AP = PAP = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = D,$$

a diagonal matrix. In the basis $\mathcal{V}$, it is clear what the action of f is: it is a stretch by a factor of 2 in the v_1 direction and it is the identity in the v_2 direction. Observe that v_1 and v_2 are not orthogonal.

What happened is that we *diagonalized* the matrix A. The diagonal entries 2 and 1 are the *eigenvalues* of A (and f), and v_1 and v_2 are corresponding *eigenvectors*. We will come back to eigenvalues and eigenvectors later on.

The above example showed that the same linear map can be represented by different matrices. This suggests making the following definition:

Definition 4.5. Two $n \times n$ matrices A and B are said to be *similar* iff there is some invertible matrix P such that

$$B = P^{-1}AP.$$

It is easily checked that similarity is an equivalence relation. From our previous considerations, *two $n \times n$ matrices A and B are similar iff they represent the same linear map with respect to two different bases.* The following surprising fact can be shown: Every square matrix A is similar to its transpose $A^\top$. The proof requires advanced concepts (the Jordan form or similarity invariants).

If $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ are two bases of E, the change of basis matrix

$$P = P_{\mathcal{V},\mathcal{U}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$ is the matrix whose jth column consists of the coordinates of v_j over the basis $(u_1, \ldots, u_n)$, which means that

$$v_j = \sum_{i=1}^n a_{ij} u_i.$$

It is natural to extend the matrix notation and to express the vector $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ in E^n as the product of a matrix times the vector $\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ in E^n, namely as

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

but notice that the matrix involved is not P, but its *transpose* $P^\top$.

This observation has the following consequence: if $\mathcal{U} = (u_1, \ldots, u_n)$ and $\mathcal{V} = (v_1, \ldots, v_n)$ are two bases of E and if

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = A \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

that is,

$$v_i = \sum_{j=1}^n a_{ij} u_j,$$

for any vector $w \in E$, if

$$w = \sum_{i=1}^n x_i u_i = \sum_{k=1}^n y_k v_k = \sum_{k=1}^n y_k \left(\sum_{j=1}^n a_{kj} u_j \right) = \sum_{j=1}^n \left(\sum_{k=1}^n a_{kj} y_k \right) u_j,$$

so

$$x_i = \sum_{k=1}^{n} a_{kj} y_k,$$

which means (note the inevitable transposition) that

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A^\top \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and so

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (A^\top)^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It is easy to see that $(A^\top)^{-1} = (A^{-1})^\top$. Also, if $\mathcal{U} = (u_1, \ldots, u_n)$, $\mathcal{V} = (v_1, \ldots, v_n)$, and $\mathcal{W} = (w_1, \ldots, w_n)$ are three bases of E, and if the change of basis matrix from $\mathcal{U}$ to $\mathcal{V}$ is $P = P_{\mathcal{V},\mathcal{U}}$ and the change of basis matrix from $\mathcal{V}$ to $\mathcal{W}$ is $Q = P_{\mathcal{W},\mathcal{V}}$, then

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

so

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = (PQ)^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

which means that the change of basis matrix $P_{\mathcal{W},\mathcal{U}}$ from $\mathcal{U}$ to $\mathcal{W}$ is PQ. This proves that

$$P_{\mathcal{W},\mathcal{U}} = P_{\mathcal{V},\mathcal{U}} P_{\mathcal{W},\mathcal{V}}.$$

Remark: In order to avoid the transposition involved in writing

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

as a more convenient notation we may write

$$\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} P.$$

Here we are defining the product

$$\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} p_{1j} \\ \vdots \\ p_{nj} \end{pmatrix} \tag{$*$}$$

of a row of vectors $\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ by the jth column of P as the linear combination

$$\sum_{i=1}^{n} p_{ij} u_i.$$

Such a definition is needed since scalar multiplication of a vector by a scalar is only defined if the scalar is on the left of the vector, but in the matrix expression $(*)$ above, the vectors are on the left of the scalars!

Even though matrices are indispensable since they are *the* major tool in applications of linear algebra, one should not lose track of the fact that

linear maps are more fundamental because they are intrinsic objects that do not depend on the choice of bases. Consequently, we advise the reader to try to think in terms of linear maps rather than reduce everything to matrices.

In our experience, this is particularly effective when it comes to proving results about linear maps and matrices, where proofs involving linear maps are often more "conceptual." These proofs are usually more general because they do not depend on the fact that the dimension is finite. Also, instead of thinking of a matrix decomposition as a purely algebraic operation, it is often illuminating to view it as a *geometric decomposition*. This is the case of the SVD, which in geometric terms says that every linear map can be factored as a rotation, followed by a rescaling along orthogonal axes and then another rotation.

After all,

a matrix is a representation of a linear map,

and most decompositions of a matrix reflect the fact that with a *suitable choice of a basis (or bases)*, the linear map is a represented by a matrix having a special shape. The problem is then to find such bases.

Still, for the beginner, matrices have a certain irresistible appeal, and we confess that it takes a certain amount of practice to reach the point where it becomes more natural to deal with linear maps. We still recommend it! For example, try to translate a result stated in terms of matrices into a result stated in terms of linear maps. Whenever we tried this exercise, we learned something.

Also, always try to keep in mind that

linear maps are geometric in nature; they act on space.

4.5 Summary

The main concepts and results of this chapter are listed below:

- The representation of linear maps by *matrices*.
- The *matrix representation mapping* $M\colon \mathrm{Hom}(E, F) \to \mathrm{M}_{n,p}$ and the representation isomorphism (Proposition 4.2).
- *Change of basis matrix* and Proposition 4.5.

4.6 Problems

Problem 4.1. Prove that the column vectors of the matrix A_1 given by

$$A_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ 1 & 3 & 1 \end{pmatrix}$$

are linearly independent.

Prove that the coordinates of the column vectors of the matrix B_1 over the basis consisting of the column vectors of A_1 given by

$$B_1 = \begin{pmatrix} 3 & 5 & 1 \\ 1 & 2 & 1 \\ 4 & 3 & -6 \end{pmatrix}$$

are the columns of the matrix P_1 given by

$$P_1 = \begin{pmatrix} -27 & -61 & -41 \\ 9 & 18 & 9 \\ 4 & 10 & 8 \end{pmatrix}.$$

Give a nontrivial linear dependence of the columns of P_1. Check that $B_1 = A_1 P_1$. Is the matrix B_1 invertible?

Problem 4.2. Prove that the column vectors of the matrix A_2 given by

$$A_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 3 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 3 \end{pmatrix}$$

are linearly independent.