

# Chapter 3

## 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  $K = \mathbb{R}$ .

### 3.1 Representation of Linear Maps by Matrices

Proposition 2.17 shows that given two vector spaces  $E$  and  $F$  and a basis  $(u_j)_{j \in J}$  of  $E$ , every linear map  $f: E \rightarrow F$  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

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

where  $j \in J$ , for a family of scalars  $(a_{ij})_{i \in I}$ . 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”  $M(f) = (a_{ij})_{i \in I, j \in J}$ .

**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  $f: E \rightarrow F$  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  $M(f) = (a_{ji})_{j \in J, i \in I}$ . 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  $(v_1, \dots, v_m)$ , that is, the matrix

$$M(f) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

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

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

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

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

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

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

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

Let

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

or more concisely,

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

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

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

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

$$f(x) = x_1 \left( \sum_{i=1}^m a_{i1} v_i \right) + \dots + x_n \left( \sum_{i=1}^m a_{in} v_i \right),$$

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

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

Thus, letting  $f(x) = y = y_1v_1 + \cdots + y_mv_m$ , we have

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

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

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

and for any  $x = x_1u_1 + x_2u_2 + x_3u_3$ , we have

$$\begin{aligned} f(x) &= f(x_1u_1 + x_2u_2 + x_3u_3) \\ &= x_1f(u_1) + x_2f(u_2) + x_3f(u_3) \\ &= x_1(a_{11}v_1 + a_{21}v_2) + x_2(a_{12}v_1 + a_{22}v_2) + x_3(a_{13}v_1 + a_{23}v_2) \\ &= (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)v_1 + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3)v_2. \end{aligned}$$

Consequently, since

$$y = y_1v_1 + y_2v_2,$$

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 3.1.** Let  $E$  and  $F$  be two vector spaces, and let  $(u_1, \dots, u_n)$  be a basis for  $E$ , and  $(v_1, \dots, v_m)$  be a basis for  $F$ . Each vector  $x \in E$  expressed in the basis  $(u_1, \dots, u_n)$  as  $x = x_1u_1 + \dots + x_nu_n$  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  $(v_1, \dots, v_m)$ .

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

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

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

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

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

**Remark:** As in the remark after Definition 2.14, there is no reason to assume that the vectors in the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  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*  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$ .

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  $4 \times 5$  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  $\dim(\text{Ker } d) = 1$ , and from

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

(see Theorem 5.8), we get  $\dim(\text{Im } d) = 4$  (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  $5 \times 4$  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  $D$ .

## 3.2 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  $(u_1, \dots, u_p)$  for  $E$ ,  $(v_1, \dots, v_n)$  for  $F$ , and  $(w_1, \dots, w_m)$  for  $G$ . Let  $g: E \rightarrow F$  and  $f: F \rightarrow G$  be linear maps. As explained earlier,  $g: E \rightarrow F$  is determined by the images of the basis vectors  $u_j$ , and  $f: F \rightarrow G$  is determined by the images of the basis vectors  $v_k$ . We would like to understand how  $f \circ g: E \rightarrow G$  is determined by the images of the basis vectors  $u_j$ .

**Remark:** Note that we are considering linear maps  $g: E \rightarrow F$  and  $f: F \rightarrow G$ , instead of  $f: E \rightarrow F$  and  $g: F \rightarrow G$ , which yields the composition  $f \circ g: E \rightarrow G$  instead of  $g \circ f: E \rightarrow G$ . Our perhaps unusual choice is motivated by the fact that if  $f$  is represented by a matrix  $M(f) = (a_{ik})$  and  $g$  is represented by a matrix  $M(g) = (b_{kj})$ , then  $f \circ g: E \rightarrow G$  is represented by the product  $AB$  of the matrices  $A$  and  $B$ . If we had adopted the other choice where  $f: E \rightarrow F$  and  $g: F \rightarrow G$ , then  $g \circ f: E \rightarrow G$  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

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

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

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

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

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

and

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

By previous considerations, for every

$$x = x_1 u_1 + \dots + x_p u_p,$$

letting  $g(x) = y = y_1 v_1 + \dots + y_n v_n$ , we have

$$y_k = \sum_{j=1}^p b_{kj} x_j \quad (2)$$

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

$$y = y_1 v_1 + \dots + y_n v_n,$$

letting  $f(y) = z = z_1 w_1 + \dots + z_m w_m$ , we have

$$z_i = \sum_{k=1}^n a_{ik} y_k \quad (3)$$

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

$$\begin{aligned} z_i &= \sum_{k=1}^n a_{ik} \left( \sum_{j=1}^p b_{kj} x_j \right) \\ &= \sum_{k=1}^n \sum_{j=1}^p a_{ik} b_{kj} x_j \\ &= \sum_{j=1}^p \sum_{k=1}^n a_{ik} b_{kj} x_j \\ &= \sum_{j=1}^p \left( \sum_{k=1}^n a_{ik} b_{kj} \right) x_j. \end{aligned}$$

Thus, defining  $c_{ij}$  such that

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj},$$

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

$$z_i = \sum_{j=1}^p c_{ij} x_j \quad (4)$$

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

Then given a linear map  $f: E \rightarrow F$  represented by the matrix  $M(f) = (a_{ij})$  w.r.t. the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$ , by Equation (1), namely

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

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

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Recall that

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Sometimes, it is necessary to incorporate the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  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 3.2.** Write  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_m)$  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  $M(x) = (x_1, \dots, x_n)$  of  $x \in E$  w.r.t. the basis  $\mathcal{U}$  and write  $y_{\mathcal{V}}$  for the coordinates  $M(y) = (y_1, \dots, y_m)$  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 [48] use the notation  $[f]_{\mathcal{U},\mathcal{V}}$ , and others such as Dummit and Foote [19] 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 [41]), or  ${}_{\mathcal{U}}[f]_{\mathcal{V}}$ , or other strange notations.

Definition 3.2 shows that the function which associates to a linear map  $f: E \rightarrow F$  the matrix  $M(f)$  w.r.t. the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  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 3.1.** (1) *Given any matrices  $A \in M_{m,n}(K)$ ,  $B \in M_{n,p}(K)$ , and  $C \in M_{p,q}(K)$ , we have*

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

*that is, matrix multiplication is associative.*

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

$$(A + B)C = AC + BC$$

$$A(C + D) = AC + AD$$

$$(\lambda A)C = \lambda(AC)$$

$$A(\lambda C) = \lambda(AC),$$

*so that matrix multiplication  $\cdot: M_{m,n}(K) \times M_{n,p}(K) \rightarrow M_{m,p}(K)$  is bilinear.*

*Proof.* (1) Every  $m \times n$  matrix  $A = (a_{ij})$  defines the function  $f_A: K^n \rightarrow K^m$  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  $f_1, f_2 \in \text{Hom}_K(E, F)$ ,  $A, B \in M_{m,n}(K)$ ,  $(u_1, \dots, u_n)$  is any basis of  $E$ , and  $(v_1, \dots, v_m)$  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 3.1 implies that the vector space  $M_n(K)$  of square matrices is a (noncommutative) ring with unit  $I_n$ . (It even shows that  $M_n(K)$  is an associative *algebra*.)

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

**Proposition 3.2.** *Given three vector spaces  $E, F, G$ , with respective bases  $(u_1, \dots, u_p)$ ,  $(v_1, \dots, v_n)$ , and  $(w_1, \dots, w_m)$ , the mapping  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  that associates the matrix  $M(g)$  to a linear map  $g: E \rightarrow F$  satisfies the following properties for all  $x \in E$ , all  $g, h: E \rightarrow F$ , and all  $f: F \rightarrow G$ :*

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

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 3.1.

Thus,  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  is an isomorphism of vector spaces, and when  $p = n$  and the basis  $(v_1, \dots, v_n)$  is identical to the basis  $(u_1, \dots, u_p)$ ,  $M: \text{Hom}(E, E) \rightarrow M_n$  is an isomorphism of rings.

*Proof.* That  $M(g(x)) = M(g)M(x)$  was shown by Definition 3.2 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  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  is clearly injective, and since every matrix defines a linear map (see Proposition 3.1), it is also surjective, and thus bijective. In view of the above identities, it is an isomorphism (and similarly for  $M: \text{Hom}(E, E) \rightarrow M_n$ , where Proposition 3.1 is used to show that  $M_n$  is a ring).  $\square$

In view of Proposition 3.2, 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.*

### 3.3 Change of Basis Matrix

It is important to observe that the isomorphism  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  given by Proposition 3.2 depends on the choice of the bases  $(u_1, \dots, u_p)$  and  $(v_1, \dots, v_n)$ , and similarly for the isomorphism  $M: \text{Hom}(E, E) \rightarrow M_n$ , which depends on the choice of the basis  $(u_1, \dots, u_n)$ . Thus, it would be useful to know how a change of basis affects the representation of a linear map  $f: E \rightarrow F$  as a matrix. The following simple proposition is needed.

**Proposition 3.3.** *Let  $E$  be a vector space, and let  $(u_1, \dots, u_n)$  be a basis of  $E$ . For every family  $(v_1, \dots, v_n)$ , let  $P = (a_{ij})$  be the matrix defined such that  $v_j = \sum_{i=1}^n a_{ij}u_i$ . The matrix  $P$  is invertible iff  $(v_1, \dots, v_n)$  is a basis of  $E$ .*

*Proof.* Note that we have  $P = M(f)$ , the matrix associated with the unique linear map  $f: E \rightarrow E$  such that  $f(u_i) = v_i$ . By Proposition 2.17,  $f$  is bijective iff  $(v_1, \dots, v_n)$  is a basis of  $E$ . Furthermore, it is obvious that the identity matrix  $I_n$  is the matrix associated with the identity  $\text{id}: E \rightarrow E$  w.r.t. any basis. If  $f$  is an isomorphism, then  $f \circ f^{-1} = f^{-1} \circ f = \text{id}$ , and by Proposition 3.2, we get  $M(f)M(f^{-1}) = M(f^{-1})M(f) = I_n$ , showing that  $P$  is invertible and that  $M(f^{-1}) = P^{-1}$ .  $\square$

Proposition 3.3 suggests the following definition.

**Definition 3.3.** Given a vector space  $E$  of dimension  $n$ , for any two bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_n)$  of  $E$ , let  $P = (a_{ij})$  be the invertible matrix defined such that

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

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

matrix  $P$  shown below:

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

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

Clearly, the change of basis matrix from  $(v_1, \dots, v_n)$  to  $(u_1, \dots, u_n)$  is  $P^{-1}$ . Since  $P = (a_{ij})$  is the matrix of the identity  $\text{id}: E \rightarrow E$  with respect to the bases  $(v_1, \dots, v_n)$  and  $(u_1, \dots, u_n)$ , given any vector  $x \in E$ , if  $x = x_1 u_1 + \dots + x_n u_n$  over the basis  $(u_1, \dots, u_n)$  and  $x = x'_1 v_1 + \dots + x'_n v_n$  over the basis  $(v_1, \dots, v_n)$ , from Proposition 3.2, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix},$$

showing that the *old* coordinates  $(x_i)$  of  $x$  (over  $(u_1, \dots, u_n)$ ) are expressed in terms of the *new* coordinates  $(x'_i)$  of  $x$  (over  $(v_1, \dots, v_n)$ ).

Now we face the painful task of assigning a “good” notation incorporating the bases  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  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  $\text{id}_E$  with respect to the bases  $\mathcal{V}$  and  $\mathcal{U}$  in that order, we could denote it by  $M_{\mathcal{V},\mathcal{U}}(\text{id})$  (Meyer [48] uses the notation  $[I]_{\mathcal{V},\mathcal{U}}$ ). We prefer to use an abbreviation for  $M_{\mathcal{V},\mathcal{U}}(\text{id})$ .

**Definition 3.4.** 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  $x_{\mathcal{U}} = (x_1, \dots, x_n)$  for the *old* coordinates of  $x$  with respect to the basis  $\mathcal{U}$  and  $x_{\mathcal{V}} = (x'_1, \dots, x'_n)$  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 [3]) 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  $(v_1, \dots, v_n)$  in terms of the *old* basis  $(u_1, \dots, u_n)$ , 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 is 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 3.1.** 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 3.2.** 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

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

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

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

### 3.4 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 3.4.** *Let  $E$  and  $F$  be vector spaces, let  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{U}' = (u'_1, \dots, u'_n)$  be two bases of  $E$ , and let  $\mathcal{V} = (v_1, \dots, v_m)$  and  $\mathcal{V}' = (v'_1, \dots, v'_m)$  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  $f: E \rightarrow F$ , 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

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

*Proof.* Since  $f: E \rightarrow F$  can be written as  $f = \text{id}_F \circ f \circ \text{id}_E$ , since  $P$  is the matrix of  $\text{id}_E$  w.r.t. the bases  $(u'_1, \dots, u'_n)$  and  $(u_1, \dots, u_n)$ , and  $Q^{-1}$  is the matrix of  $\text{id}_F$  w.r.t. the bases  $(v_1, \dots, v_m)$  and  $(v'_1, \dots, v'_m)$ , by Proposition 3.2, we have  $M'(f) = Q^{-1}M(f)P$ .  $\square$

As a corollary, we get the following result.

**Corollary 3.5.** *Let  $E$  be a vector space, and let  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{U}' = (u'_1, \dots, u'_n)$  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  $f: E \rightarrow E$ , 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,

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

**Example 3.3.** 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 3.5.** 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  $A^T$** . The proof requires advanced concepts (the Jordan form or similarity invariants).

If  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  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  $(u_1, \dots, u_n)$  to  $(v_1, \dots, v_n)$  is the matrix whose  $j$ th column consists of the coordinates of  $v_j$  over the basis  $(u_1, \dots, u_n)$ , 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  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  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

$$w = \sum_{i=1}^n x_i u_i = \sum_{k=1}^n y_k v_k,$$

then

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



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  $\mathcal{U} = (u_1, \dots, u_n)$ ,  $\mathcal{V} = (v_1, \dots, v_n)$ , and  $\mathcal{W} = (w_1, \dots, w_n)$  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

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

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 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.*

### 3.5 Summary

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

- The representation of linear maps by *matrices*.
- The *matrix representation mapping*  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  and the representation isomorphism (Proposition 3.2).
- *Change of basis matrix* and Proposition 3.4.

### 3.6 Problems

**Problem 3.1.** 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 3.2.** 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.

Prove that the column vectors of the matrix  $B_2$  given by

$$B_2 = \begin{pmatrix} 1 & -2 & 2 & -2 \\ 0 & -3 & 2 & -3 \\ 3 & -5 & 5 & -4 \\ 3 & -4 & 4 & -4 \end{pmatrix}$$

are linearly independent.

Prove that the coordinates of the column vectors of the matrix  $B_2$  over the basis consisting of the column vectors of  $A_2$  are the columns of the matrix  $P_2$  given by

$$P_2 = \begin{pmatrix} 2 & 0 & 1 & -1 \\ -3 & 1 & -2 & 1 \\ 1 & -2 & 2 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}.$$

Check that  $A_2 P_2 = B_2$ . Prove that

$$P_2^{-1} = \begin{pmatrix} -1 & -1 & -1 & 1 \\ 2 & 1 & 1 & -2 \\ 2 & 1 & 2 & -3 \\ -1 & -1 & 0 & -1 \end{pmatrix}.$$

What are the coordinates over the basis consisting of the column vectors of  $B_2$  of the vector whose coordinates over the basis consisting of the column vectors of  $A_1$  are  $(2, -3, 0, 0)$ ?

**Problem 3.3.** Consider the polynomials

$$\begin{array}{lll} B_0^2(t) = (1-t)^2 & B_1^2(t) = 2(1-t)t & B_2^2(t) = t^2 \\ B_0^3(t) = (1-t)^3 & B_1^3(t) = 3(1-t)^2t & B_2^3(t) = 3(1-t)t^2 \quad B_3^3(t) = t^3, \end{array}$$

known as the *Bernstein polynomials* of degree 2 and 3.

(1) Show that the Bernstein polynomials  $B_0^2(t), B_1^2(t), B_2^2(t)$  are expressed as linear combinations of the basis  $(1, t, t^2)$  of the vector space of polynomials of degree at most 2 as follows:

$$\begin{pmatrix} B_0^2(t) \\ B_1^2(t) \\ B_2^2(t) \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}.$$

Prove that

$$B_0^2(t) + B_1^2(t) + B_2^2(t) = 1.$$

(2) Show that the Bernstein polynomials  $B_0^3(t), B_1^3(t), B_2^3(t), B_3^3(t)$  are expressed as linear combinations of the basis  $(1, t, t^2, t^3)$  of the vector space of polynomials of degree at most 3 as follows:

$$\begin{pmatrix} B_0^3(t) \\ B_1^3(t) \\ B_2^3(t) \\ B_3^3(t) \end{pmatrix} = \begin{pmatrix} 1 & -3 & 3 & -1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ t^2 \\ t^3 \end{pmatrix}.$$

Prove that

$$B_0^3(t) + B_1^3(t) + B_2^3(t) + B_3^3(t) = 1.$$

(3) Prove that the Bernstein polynomials of degree 2 are linearly independent, and that the Bernstein polynomials of degree 3 are linearly independent.

**Problem 3.4.** Recall that the *binomial coefficient*  $\binom{m}{k}$  is given by

$$\binom{m}{k} = \frac{m!}{k!(m-k)!},$$

with  $0 \leq k \leq m$ .

For any  $m \geq 1$ , we have the  $m+1$  *Bernstein polynomials* of degree  $m$  given by

$$B_k^m(t) = \binom{m}{k} (1-t)^{m-k} t^k, \quad 0 \leq k \leq m.$$

(1) Prove that

$$B_k^m(t) = \sum_{j=k}^m (-1)^{j-k} \binom{m}{j} \binom{j}{k} t^j. \quad (*)$$

Use the above to prove that  $B_0^m(t), \dots, B_m^m(t)$  are linearly independent.

(2) Prove that

$$B_0^m(t) + \dots + B_m^m(t) = 1.$$

(3) What can you say about the symmetries of the  $(m+1) \times (m+1)$  matrix expressing  $B_0^m, \dots, B_m^m$  in terms of the basis  $1, t, \dots, t^m$ ?

Prove your claim (beware that in equation  $(*)$  the coefficient of  $t^j$  in  $B_k^m$  is the entry on the  $(k+1)$ th row of the  $(j+1)$ th column, since  $0 \leq k, j \leq m$ . Make appropriate modifications to the indices).

What can you say about the sum of the entries on each row of the above matrix? What about the sum of the entries on each column?

(4) The purpose of this question is to express the  $t^i$  in terms of the Bernstein polynomials  $B_0^m(t), \dots, B_m^m(t)$ , with  $0 \leq i \leq m$ .

First, prove that

$$t^i = \sum_{j=0}^{m-i} t^i B_j^{m-i}(t), \quad 0 \leq i \leq m.$$

Then prove that

$$\binom{m}{i} \binom{m-i}{j} = \binom{m}{i+j} \binom{i+j}{i}.$$

Use the above facts to prove that

$$t^i = \sum_{j=0}^{m-i} \frac{\binom{i+j}{i}}{\binom{m}{i}} B_{i+j}^m(t).$$

Conclude that the Bernstein polynomials  $B_0^m(t), \dots, B_m^m(t)$  form a basis of the vector space of polynomials of degree  $\leq m$ .

Compute the matrix expressing  $1, t, t^2$  in terms of  $B_0^2(t), B_1^2(t), B_2^2(t)$ , and the matrix expressing  $1, t, t^2, t^3$  in terms of  $B_0^3(t), B_1^3(t), B_2^3(t), B_3^3(t)$ .

You should find

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1/2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

and

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

(5) A *polynomial curve*  $C(t)$  of degree  $m$  in the plane is the set of points  $C(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$  given by two polynomials of degree  $\leq m$ ,

$$\begin{aligned} x(t) &= \alpha_0 t^{m_1} + \alpha_1 t^{m_1-1} + \dots + \alpha_{m_1} \\ y(t) &= \beta_0 t^{m_2} + \beta_1 t^{m_2-1} + \dots + \beta_{m_2}, \end{aligned}$$

with  $1 \leq m_1, m_2 \leq m$  and  $\alpha_0, \beta_0 \neq 0$ .

Prove that there exist  $m+1$  points  $b_0, \dots, b_m \in \mathbb{R}^2$  so that

$$C(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = B_0^m(t)b_0 + B_1^m(t)b_1 + \dots + B_m^m(t)b_m$$

for all  $t \in \mathbb{R}$ , with  $C(0) = b_0$  and  $C(1) = b_m$ . Are the points  $b_1, \dots, b_{m-1}$  generally on the curve?

We say that the curve  $C$  is a *Bézier curve* and  $(b_0, \dots, b_m)$  is the list of *control points* of the curve (control points need not be distinct).

**Remark:** Because  $B_0^m(t) + \cdots + B_m^m(t) = 1$  and  $B_i^m(t) \geq 0$  when  $t \in [0, 1]$ , the curve segment  $C[0, 1]$  corresponding to  $t \in [0, 1]$  belongs to the convex hull of the control points. This is an important property of Bézier curves which is used in geometric modeling to find the intersection of curve segments. Bézier curves play an important role in computer graphics and geometric modeling, but also in robotics because they can be used to model the trajectories of moving objects.

**Problem 3.5.** Consider the  $n \times n$  matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & 0 & \cdots & 0 & -a_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & -a_2 \\ 0 & 0 & 0 & \cdots & 1 & -a_1 \end{pmatrix},$$

with  $a_n \neq 0$ .

(1) Find a matrix  $P$  such that

$$A^\top = P^{-1}AP.$$

What happens when  $a_n = 0$ ?

*Hint.* First, try  $n = 3, 4, 5$ . Such a matrix must have zeros above the “antidiagonal,” and identical entries  $p_{ij}$  for all  $i, j \geq 0$  such that  $i + j = n + k$ , where  $k = 1, \dots, n$ .

(2) Prove that if  $a_n = 1$  and if  $a_1, \dots, a_{n-1}$  are integers, then  $P$  can be chosen so that the entries in  $P^{-1}$  are also integers.

**Problem 3.6.** For any matrix  $A \in M_n(\mathbb{C})$ , let  $R_A$  and  $L_A$  be the maps from  $M_n(\mathbb{C})$  to itself defined so that

$$L_A(B) = AB, \quad R_A(B) = BA, \quad \text{for all } B \in M_n(\mathbb{C}).$$

(1) Check that  $L_A$  and  $R_A$  are linear, and that  $L_A$  and  $R_B$  commute for all  $A, B$ .

Let  $\text{ad}_A: M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$  be the linear map given by

$$\text{ad}_A(B) = L_A(B) - R_A(B) = AB - BA = [A, B], \quad \text{for all } B \in M_n(\mathbb{C}).$$

Note that  $[A, B]$  is the Lie bracket.

(2) Prove that if  $A$  is invertible, then  $L_A$  and  $R_A$  are invertible; in fact,  $(L_A)^{-1} = L_{A^{-1}}$  and  $(R_A)^{-1} = R_{A^{-1}}$ . Prove that if  $A = PBP^{-1}$  for some invertible matrix  $P$ , then

$$L_A = L_P \circ L_B \circ L_P^{-1}, \quad R_A = R_P^{-1} \circ R_B \circ R_P.$$

(3) Recall that the  $n^2$  matrices  $E_{ij}$  defined such that all entries in  $E_{ij}$  are zero except the  $(i, j)$ th entry, which is equal to 1, form a basis of the vector space  $M_n(\mathbb{C})$ . Consider the partial ordering of the  $E_{ij}$  defined such that for  $i = 1, \dots, n$ , if  $n \geq j > k \geq 1$ , then  $E_{ij}$  precedes  $E_{ik}$ , and for  $j = 1, \dots, n$ , if  $1 \leq i < h \leq n$ , then  $E_{ij}$  precedes  $E_{hj}$ .

Draw the Hasse diagram of the partial order defined above when  $n = 3$ .

There are total orderings extending this partial ordering. How would you find them algorithmically? Check that the following is such a total order:

$$(1, 3), (1, 2), (1, 1), (2, 3), (2, 2), (2, 1), (3, 3), (3, 2), (3, 1).$$

(4) Let the total order of the basis  $(E_{ij})$  extending the partial ordering defined in (2) be given by

$$(i, j) < (h, k) \quad \text{iff} \quad \begin{cases} i = h \text{ and } j > k \\ \text{or } i < h. \end{cases}$$

Let  $R$  be the  $n \times n$  permutation matrix given by

$$R = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Observe that  $R^{-1} = R$ . Prove that for any  $n \geq 1$ , the matrix of  $L_A$  is given by  $A \otimes I_n$ , and the matrix of  $R_A$  is given by  $I_n \otimes RA^T R$  (over the basis  $(E_{ij})$  ordered as specified above), where  $\otimes$  is the *Kronecker product* (also called *tensor product*) of matrices defined in Definition 4.4. *Hint.* Figure out what are  $R_B(E_{ij}) = E_{ij}B$  and  $L_B(E_{ij}) = BE_{ij}$ .

(5) Prove that if  $A$  is upper triangular, then the matrices representing  $L_A$  and  $R_A$  are also upper triangular.

Note that if instead of the ordering

$$E_{1n}, E_{1n-1}, \dots, E_{11}, E_{2n}, \dots, E_{21}, \dots, E_{nn}, \dots, E_{n1},$$

that I proposed you use the standard lexicographic ordering

$$E_{11}, E_{12}, \dots, E_{1n}, E_{21}, \dots, E_{2n}, \dots, E_{n1}, \dots, E_{nn},$$

then the matrix representing  $L_A$  is still  $A \otimes I_n$ , but the matrix representing  $R_A$  is  $I_n \otimes A^T$ . In this case, if  $A$  is upper-triangular, then the matrix of  $R_A$  is *lower triangular*. This is the motivation for using the first basis (avoid upper becoming lower).

