CHAPTER 5

Linear Transformations and Matrices

In Section 3.1 we defined matrices by systems of linear equations, and in Section 3.6 we showed that the set of all matrices over a field $\mathcal{F}$ may be endowed with certain algebraic properties such as addition and multiplication. In this chapter we present another approach to defining matrices, and we will see that it also leads to the same algebraic behavior as well as yielding important new properties.

5.1 LINEAR TRANSFORMATIONS

Recall that vector space homomorphisms were defined in Section 2.2. We now repeat that definition using some new terminology. In particular, a mapping $T: U \to V$ of two vector spaces over the same field $\mathcal{F}$ is called a **linear transformation** if it has the following properties for all $x, y \in U$ and $a \in \mathcal{F}$:

(a) $T(x + y) = T(x) + T(y)$
(b) $T(ax) = aT(x)$.

Letting $a = 0$ and -1 shows

$$T(0) = 0$$

and

$$T(-x) = -T(x) .$$

We also see that

$$T(x - y) = T(x + (-y)) = T(x) + T(-y) = T(x) - T(y) .$$

It should also be clear that by induction we have, for any finite sum,

$$T(\Sigma a_i x_i) = \Sigma T(a_i x_i) = \Sigma a_i T(x_i)$$

for any vectors $x_i \in V$ and scalars $a_i \in \mathcal{F}$.

Example 5.1 Let $T: \mathbb{R}^3 \to \mathbb{R}^2$ be the "projection" mapping defined for any $u = (x, y, z) \in \mathbb{R}^3$ by

$$T(u) = T(x, y, z) = (x, y, 0) .$$

Then if $v = (x', y', z')$ we have

$$\begin{aligned} T(u+v) &= T(x+x',\ y+y',\ z+z') \\ &= (x+x',\ y+y',\ 0) \\ &= (x,\ y,\ 0) + (x',\ y',\ 0) \\ &= T(u) + T(v) \end{aligned}$$

and

$$T(au) = T(ax, ay, az) = (ax, ay, 0) = a(x, y, 0) = aT(u) .$$

Hence T is a linear transformation. ∕∕

Example 5.2 Let $P \in M_n(\mathcal{F})$ be a fixed invertible matrix. We define a mapping $S: M_n(\mathcal{F}) \to M_n(\mathcal{F})$ by $S(A) = P^{-1}AP$. It is easy to see that this defines a linear transformation since

$$S(\alpha A + B) = P^{-1}(\alpha A + B)P = \alpha P^{-1}AP + P^{-1}BP = \alpha S(A) + S(B) . \ /\!\!/$$

Example 5.3 Let V be a real inner product space, and let W be any subspace of V. By Theorem 2.22 we have $V = W \oplus W^\perp$, and hence by Theorem 2.12, any $v \in V$ has a unique decomposition $v = x + y$ where $x \in W$ and $y \in W^\perp$. Now define the mapping $T: V \to W$ by $T(v) = x$. Then

$$T(v_1 + v_2) = x_1 + x_2 = T(v_1) + T(v_2)$$

and

$$T(av) = ax = aT(v)$$

so that T is a linear transformation. This mapping is called the **orthogonal projection** of V onto W. ∕∕

Let T: V → W be a linear transformation, and let $\{e_i\}$ be a basis for V. Then for any $x \in V$ we have $x = \sum x_i e_i$, and hence

$$T(x) = T(\sum x_i e_i) = \sum x_i T(e_i) \ .$$

Therefore, if we know all of the $T(e_i)$, then we know T(x) for any $x \in V$. In other words, *a linear transformation is determined by specifying its values on a basis*. Our first theorem formalizes this fundamental observation.

Theorem 5.1 Let U and V be finite-dimensional vector spaces over $\mathcal{F}$, and let $\{e_1, \ldots, e_n\}$ be a basis for U. If $v_1, \ldots, v_n$ are any n arbitrary vectors in V, then there exists a unique linear transformation T: U → V such that $T(e_i) = v_i$ for each $i = 1, \ldots, n$.

Proof For any $x \in U$ we have $x = \sum_{i=1}^n x_i e_i$ for some unique set of scalars x_i (Theorem 2.4, Corollary 2). We define the mapping T by

$$T(x) = \sum_{i=1}^n x_i v_i$$

for any $x \in U$. Since the x_i are unique, this mapping is well-defined (see Exercise 5.1.1). Noting that for any $i = 1, \ldots, n$ we have $e_i = \sum_j \delta_{ij} e_j$, it follows that

$$T(e_i) = \sum_{j=1}^n \delta_{ij} v_j = v_i \ .$$

We show that T so defined is a linear transformation.

If $x = \sum x_i e_i$ and $y = \sum y_i e_i$, then $x + y = \sum (x_i + y_i) e_i$, and hence

$$T(x + y) = \sum (x_i + y_i) v_i = \sum x_i v_i + \sum y_i v_i = T(x) + T(y) \ .$$

Also, if $c \in \mathcal{F}$ then $cx = \sum (cx_i) e_i$, and thus

$$T(cx) = \sum (cx_i) v_i = c \sum x_i v_i = cT(u)$$

which shows that T is indeed a linear transformation.

Now suppose that T′: U → V is any other linear transformation defined by $T'(e_i) = v_i$. Then for any $x \in U$ we have

$$T'(x) = T'(\sum x_i e_i) = \sum x_i T'(e_i) = \sum x_i v_i = \sum x_i T(e_i) = T(\sum x_i e_i) = T(x)$$

and hence $T'(x) = T(x)$ for all $x \in U$. This means that $T' = T$ which thus proves uniqueness. ∎

Example 5.4 Let $T \in L(\mathcal{F}^m, \mathcal{F}^n)$ be a linear transformation from $\mathcal{F}^m$ to $\mathcal{F}^n$, and let $\{e_1, \dots, e_m\}$ be the standard basis for $\mathcal{F}^m$. We may uniquely define T by specifying any m vectors $v_1, \dots, v_m$ in $\mathcal{F}^n$. In other words, we define T by the requirement $T(e_i) = v_i$ for each $i = 1, \dots, m$. Since T is linear, for any $x \in \mathcal{F}^m$ we have $x = \sum_{i=1}^m x_i e_i$ and hence

$$T(x) = \sum_{i=1}^{m} x_i v_i \quad .$$

Now define the matrix $A = (a_{ij}) \in M_{n \times m}(\mathcal{F})$ with column vectors given by $A^i = v_i \in \mathcal{F}^n$. In other words (remember these are columns),

$$A^i = (a_{1i}, \dots, a_{ni}) = (v_{1i}, \dots, v_{ni}) = v_i$$

where $v_i = \sum_{j=1}^n f_j v_{ji}$ and $\{f_1, \dots, f_n\}$ is the standard basis for $\mathcal{F}^n$. Writing out T(x) we have

$$T(x) = \sum_{i=1}^{m} x_i v_i = x_1 \begin{pmatrix} v_{11} \\ \vdots \\ v_{n1} \end{pmatrix} + \cdots + x_m \begin{pmatrix} v_{1m} \\ \vdots \\ v_{nm} \end{pmatrix} = \begin{pmatrix} v_{11}x_1 + \cdots + v_{1m}x_m \\ \vdots \\ v_{n1}x_1 + \cdots + v_{nm}x_m \end{pmatrix}$$

and therefore, in terms of the matrix A, our transformation takes the form

$$T(x) = \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \quad .$$

We have therefore constructed an explicit matrix representation of the transformation T. We shall have much more to say about such matrix representations shortly. ∕∕

Given vector spaces U and V, we claim that the set of all linear transformations from U to V can itself be made into a vector space. To accomplish this we proceed as follows. If U and V are vector spaces over F and f, g: $U \to V$ are mappings, we naturally define

$$(f + g)(x) = f(x) + g(x)$$

and

$$(cf)(x) = cf(x)$$

for $x \in U$ and $c \in \mathcal{F}$. In addition, if $h: V \to W$ (where W is another vector space over $\mathcal{F}$), then we may define the composite mapping $h \circ g: U \to W$ in the usual way by

$$(h \circ g)(x) = h(g(x)) \ .$$

Theorem 5.2 Let U, V and W be vector spaces over $\mathcal{F}$, let $c \in \mathcal{F}$ be any scalar, and let $f, g: U \to V$ and $h: V \to W$ be linear transformations. Then the mappings $f + g$, cf, and $h \circ g$ are all linear transformations.

Proof First, we see that for $x, y \in U$ and $c \in \mathcal{F}$ we have

$$\begin{aligned}(f+g)(x+y) &= f(x+y)+g(x+y)\\ &= f(x)+f(y)+g(x)+g(y)\\ &= (f+g)(x)+(f+g)(y)\end{aligned}$$

and

$$(f+g)(cx) = f(cx)+g(cx) = cf(x)+cg(x) = c[f(x)+g(x)] = c(f+g)(x)$$

and hence $f + g$ is a linear transformation. The proof that cf is a linear transformation is left to the reader (Exercise 5.1.3). Finally, we see that

$$\begin{aligned}(h\circ g)(x+y) &= h(g(x+y)) = h(g(x)+g(y)) = h(g(x))+h(g(y))\\ &= (h\circ g)(x)+(h\circ g)(y)\end{aligned}$$

and

$$(h\circ g)(cx) = h(g(cx)) = h(cg(x)) = ch(g(x)) = c(h\circ g)(x)$$

so that $h \circ g$ is also a linear transformation. ∎

We define the **zero mapping** $0: U \to V$ by $0x = 0$ for all $x \in U$. Since

$$0(x + y) = 0 = 0x + 0y$$

and

$$0(cx) = 0 = c(0x)$$

it follows that the zero mapping is a linear transformation. Next, given a mapping $f: U \to V$, we define its **negative** $-f: U \to V$ by $(-f)(x) = -f(x)$ for all $x \in U$. If f is a linear transformation, then $-f$ is also linear because cf is linear for any $c \in \mathcal{F}$ and $-f = (-1)f$ (by Theorem 2.1(c)). Lastly, we note that

$$\begin{aligned}[f+(-f)](x) &= f(x)+(-f)(x) = f(x)+[-f(x)] = f(x)+f(-x) = f(x-x)\\ &= f(0) = 0\end{aligned}$$

for all $x \in U$ so that $f + (-f) = (-f) + f = 0$ for all linear transformations f.

With all of this algebra out of the way, we are now in a position to easily prove our claim.

Theorem 5.3 Let U and V be vector spaces over $\mathcal{F}$. Then the set of all linear transformations of U to V with addition and scalar multiplication defined as above is a linear vector space over $\mathcal{F}$.

Proof We leave it to the reader to show that the set of all such linear transformations obeys the properties (V1) – (V8) given in Section 2.1 (see Exercise 5.1.4). ∎

We denote the vector space defined in Theorem 5.3 by L(U, V). (Some authors denote this space by Hom(U, V) since a linear transformation is just a vector space homomorphism). The space L(U, V) is often called the space of **linear transformations** (or **mappings**). In the particular case that U and V are finite-dimensional, we have the following important result.

Theorem 5.4 Let dim U = m and dim V = n. Then

$$\dim L(U, V) = (\dim U)(\dim V) = mn\ .$$

Proof We prove the theorem by exhibiting a basis for L(U, V) that contains mn elements. Let $\{e_1, \ldots, e_m\}$ be a basis for U, and let $\{\bar{e}_1, \ldots, \bar{e}_n\}$ be a basis for V. Define the mn linear transformations $E^i{}_j \in L(U, V)$ by

$$E^i{}_j(e_k) = \delta^i{}_k \bar{e}_j$$

where $i, k = 1, \ldots, m$ and $j = 1, \ldots, n$. Theorem 5.1 guarantees that the mappings $E^i{}_j$ are unique. To show that $\{E^i{}_j\}$ is a basis, we must show that it is linearly independent and spans L(U, V).

If

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a^j{}_i E^i{}_j = 0$$

for some set of scalars $a^j{}_i$, then for any e_k we have

$$0 = \textstyle\sum_{i,\,j} a^j{}_i E^i{}_j(e_k) = \sum_{i,\,j} a^j{}_i \delta^i{}_k \bar{e}_j = \sum_j a^j{}_k \bar{e}_j\ .$$

But the $\bar{e}_j$ are a basis and hence linearly independent, and thus we must have $a^j{}_k = 0$ for every $j = 1, \ldots, n$ and $k = 1, \ldots, m$. This shows that the $E^i{}_j$ are linearly independent.

Now suppose $f \in L(U, V)$ and let $x \in U$. Then $x = \sum_i x^i e_i$ and

$$f(x) = f(\textstyle\sum_i x^i e_i) = \sum_i x^i f(e_i) \ .$$

Since $f(e_i) \in V$, we must have $f(e_i) = \sum_j c^j{}_i \bar{e}_j$ for some set of scalars $c^j{}_i$, and hence

$$f(e_i) = \textstyle\sum_j c^j{}_i \bar{e}_j = \sum_{j,k} c^j{}_k \delta^k{}_i \bar{e}_j = \sum_{j,k} c^j{}_k E^k{}_j(e_i) \ .$$

But this means that $f = \sum_{j,k} c^j{}_k E^k{}_j$ (Theorem 5.1), and therefore $\{E^k{}_j\}$ spans $L(U, V)$. ∎

Suppose we have a linear mapping $\phi\colon V \to \mathcal{F}$ of a vector space V to the field of scalars. By definition, this means that

$$\phi(ax + by) = a\phi(x) + b\phi(y)$$

for every $x, y \in V$ and $a, b \in \mathcal{F}$. The mapping ϕ is called a **linear functional** on V.

Example 5.5 Consider the space $M_n(\mathcal{F})$ of n-square matrices over $\mathcal{F}$. Since the trace of any $A = (a_{ij}) \in M_n(\mathcal{F})$ is defined by

$$\operatorname{Tr} A = \sum_{i=1}^{n} a_{ii}$$

(see Exercise 3.6.7), it is easy to show that Tr defines a linear functional on $M_n(\mathcal{F})$ (Exercise 5.1.5). ∕∕

Example 5.6 Let $C[a, b]$ denote the space of all real-valued continuous functions defined on the interval $[a, b]$ (see Exercise 2.1.6). We may define a linear functional L on $C[a, b]$ by

$$L(f) = \int_a^b f(x)\, dx$$

for every $f \in C[a, b]$. It is also left to the reader (Exercise 5.1.5) to show that this does indeed define a linear functional on $C[a, b]$. ∕∕

Let V be a vector space over $\mathcal{F}$. Since $\mathcal{F}$ is also a vector space over itself, we may consider the space $L(V, \mathcal{F})$. This vector space is the set of all linear functionals on V, and is called the **dual space** of V (or the **space of linear functionals** on V). The dual space is generally denoted by V^*. From the proof of Theorem 5.4, we see that if $\{e_i\}$ is a basis for V, then V* has a unique basis $\{\omega^j\}$ defined by

$$\omega^j(e_i) = \delta^j{}_i \ .$$

The basis $\{\omega^j\}$ is referred to as the **dual basis** to the basis $\{e_i\}$. We also see that Theorem 5.4 shows that dim V* = dim V.

(Let us point out that we make no real distinction between subscripts and superscripts. For our purposes, we use whichever is more convenient from a notational standpoint. However, in tensor analysis and differential geometry, subscripts and superscripts are used precisely to distinguish between a vector space and its dual. We shall follow this convention in Chapter 11.)

Example 5.7 Consider the space $V = \mathcal{F}^n$ of all n-tuples of scalars. If we write any $x \in V$ as a column vector, then V* is just the space of row vectors. This is because if $\phi \in V^*$ we have

$$\phi(x) = \phi(\textstyle\sum x_i e_i) = \textstyle\sum x_i \phi(e_i)$$

where the e_i are the standard (column) basis vectors for $V = \mathcal{F}^n$. Thus, since $\phi(e_i) \in \mathcal{F}$, we see that every $\phi(x)$ is the product of some scalar $\phi(e_i)$ times the scalar x_i, summed over $i = 1, \dots, n$. If we write $\phi(e_i) = a_i$, it then follows that we may write

$$\phi(x) = \phi(x_1, \dots, x_n) = (a_1, \dots, a_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad (*)$$

or simply $\phi(x) = \sum a_i x_i$. This expression is in fact the origin of the term "linear form."

Since any row vector in $\mathcal{F}^n$ can be expressed in terms of the basis vectors $\omega^1 = (1, 0, \dots, 0), \dots, \omega^n = (0, 0, \dots, 1)$, we see from (*) that the ω^j do indeed form the basis dual to $\{e_i\}$ since they clearly have the property that $\omega^j(e_i) = \delta^j{}_i$. In other words, the row vector ω^j is just the transpose of the corresponding column vector e_j. ∕∕

Since U* is a vector space, the reader may wonder whether or not we may form the space U** = (U*)*. The answer is "yes," and the space U** is called the **double dual** (or **second dual**) of U. In fact, for finite-dimensional vector spaces, it is essentially true that U** = U (in the sense that U and U** are isomorphic). However, we prefer to postpone our discussion of these matters until a later chapter when we can treat all of this material in the detail that it warrants.

Exercises

1. Verify that the mapping T of Theorem 5.1 is well-defined.

2. Repeat Example 5.4, except now let the matrix $A = (a_{ij})$ have row vectors $A_i = v_i \in \mathcal{F}^n$. What is the matrix representation of the operation $T(x)$?

3. Show that cf is a linear transformation in the proof of Theorem 5.2.

4. Prove Theorem 5.3.

5. (a) Show that the function Tr defines a linear functional on $M_n(\mathcal{F})$ (see Example 5.5).
   (b) Show that the mapping L defined in Example 5.6 defines a linear functional.

6. Explain whether or not each of the following mappings f is linear:
   (a) $f: \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = xy$.
   (b) $f: \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y, z) = 2x - 3y + 4z$.
   (c) $f: \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f(x, y) = (x + 1, 2y, x + y)$.
   (d) $f: \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (|x|, 0)$.
   (e) $f: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (x + y, x)$.
   (f) $f: \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f(x, y, z) = (1, -x, y + z)$.
   (g) $f: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (\sin x, y)$.
   (h) $f: \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = |x - y|$.

7. Let $T: U \to V$ be a bijective linear transformation. Define T^{-1} and show that it is also a linear transformation.

8. Let $T: U \to V$ be a linear transformation, and suppose that we have the set of vectors $u_1, \dots, u_n \in U$ with the property that $T(u_1), \dots, T(u_n) \in V$ is linearly independent. Show that $\{u_1, \dots, u_n\}$ is linearly independent.

9. Let $B \in M_n(\mathcal{F})$ be arbitrary. Show that the mapping $T: M_n(\mathcal{F}) \to M_n(\mathcal{F})$ defined by $T(A) = [A, B]_+ = AB + BA$ is linear. Is the same true for the mapping $T'(A) = [A, B] = AB - BA$?

10. Let T: $\mathcal{F}^2 \to \mathcal{F}^2$ be the linear transformation defined by the system

$$\begin{aligned} y_1 &= -3x_1 + x_2 \\ y_2 &= x_1 - x_2 \end{aligned}$$

and let S be the linear transformation defined by the system

$$\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= x_1 \end{aligned}$$

Find a system of equations that defines each of the following linear transformations:

(a) 2T (b) T − S (c) T^2
(d) TS (e) ST (f) $T^2 + 2S$

11. Does there exist a linear transformation T: $\mathbb{R}^3 \to \mathbb{R}^2$ with the property that T(1, −1, 1) = (1, 0) and T(1, 1, 1) = (0, 1)?

12. Suppose $u_1 = (1, -1)$, $u_2 = (2, -1)$, $u_3 = (-3, 2)$ and $v_1 = (1, 0)$, $v_2 = (0, 1)$, $v_3 = (1, 1)$. Does there exist a linear transformation T: $\mathbb{R}^2 \to \mathbb{R}^2$ with the property that $Tu_i = v_i$ for each i = 1, 2, and 3?

13. Find T(x, y, z) if T: $\mathbb{R}^3 \to \mathbb{R}$ is defined by T(1, 1, 1) = 3, T(0, 1, −2) = 1 and T(0, 0, 1) = −2.

14. Let V be the set of all complex numbers considered as a vector space over the real field. Find a mapping T: V → V that is a linear transformation on V, but is not a linear transformation on the space $\mathbb{C}^1$ (i.e., the set of complex numbers considered as a complex vector space).

15. If V is finite-dimensional and $x_1, x_2 \in V$ with $x_1 \neq x_2$, prove there exists a linear functional $f \in V^*$ such that $f(x_1) \neq f(x_2)$.

5.2 FURTHER PROPERTIES OF LINEAR TRANSFORMATIONS

Suppose $T \in L(U, V)$ where U and V are finite-dimensional over $\mathcal{F}$. We define the **image** of T to be the set

$$\operatorname{Im} T = \{T(x) \in V : x \in U\}$$

and the **kernel** of T to be the set

$$\operatorname{Ker} T = \{x \in U: T(x) = 0\} .$$

(Many authors call Im T the **range** of T, but we use this term to mean the space V in which T takes its values.) Since $T(0) = 0 \in V$, we see that $0 \in \operatorname{Im} T$, and hence $\operatorname{Im} T \neq \varnothing$. Now suppose $x', y' \in \operatorname{Im} T$. Then there exist x, $y \in U$ such that $T(x) = x'$ and $T(y) = y'$. Then for any $a, b \in \mathcal{F}$ we have

$$ax' + by' = aT(x) + bT(y) = T(ax + by) \in \operatorname{Im} T$$

(since $ax + by \in U$), and thus Im T is a subspace of V. Similarly, we see that $0 \in \operatorname{Ker} T$, and if $x, y \in \operatorname{Ker} T$ then

$$T(ax + by) = aT(x) + bT(y) = 0$$

so that Ker T is also a subspace of U. Ker T is frequently called the **null space** of T.

We now restate Theorem 2.5 in our current terminology.

Theorem 5.5 A linear transformation $T \in L(U, V)$ is an isomorphism if and only if $\operatorname{Ker} T = \{0\}$.

For example, the projection mapping T defined in Example 5.1 is not an isomorphism because $T(0, 0, z) = (0, 0, 0)$ for all $(0, 0, z) \in \mathbb{R}^3$. In fact, if x_0 and y_0 are fixed, then we have $T(x_0, y_0, z) = (x_0, y_0, 0)$ independently of z.

If $T \in L(U, V)$, we define the **rank** of T to be the number

$$r(T) = \dim(\operatorname{Im} T)$$

and the **nullity** of T to be the number

$$\operatorname{nul} T = \dim(\operatorname{Ker} T) .$$

We will shortly show that this definition of rank is essentially the same as our previous definition of the rank of a matrix. The relationship between r(T) and nul T is given in the following important result.

Theorem 5.6 If U and V are finite-dimensional over $\mathcal{F}$ and $T \in L(U, V)$, then

$$r(T) + \operatorname{nul} T = \dim U .$$

Proof Let $\{u_1, \ldots, u_n\}$ be a basis for U and suppose that Ker T = {0}. Then for any $x \in U$ we have

$$T(x) = T(\sum x_i u_i) = \sum x_i T(u_i)$$

for some set of scalars x_i, and therefore $\{T(u_i)\}$ spans Im T. If $\sum c_i T(u_i) = 0$, then

$$0 = \sum c_i T(u_i) = \sum T(c_i u_i) = T(\sum c_i u_i)$$

which implies that $\sum c_i u_i = 0$ (since Ker T = {0}). But the u_i are linearly independent so that we must have $c_i = 0$ for every i, and hence $\{T(u_i)\}$ is linearly independent. Since nul T = dim(Ker T) = 0 and r(T) = dim(Im T) = n = dim U, we see that r(T) + nul T = dim U.

Now suppose that Ker T ≠ {0}, and let $\{w_1, \ldots, w_k\}$ be a basis for Ker T. By Theorem 2.10, we may extend this to a basis $\{w_1, \ldots, w_n\}$ for U. Since $T(w_i) = 0$ for each $i = 1, \ldots, k$ it follows that the vectors $T(w_{k+1}), \ldots, T(w_n)$ span Im T. If

$$\sum_{j=k+1}^{n} c_j T(w_j) = 0$$

for some set of scalars c_i, then

$$0 = \sum_{j=k+1}^{n} c_j T(w_j) = \sum_{j=k+1}^{n} T(c_j w_j) = T(\sum_{j=k+1}^{n} c_j w_j)$$

so that $\sum_{j=k+1}^{n} c_j w_j \in$ Ker T. This means that

$$\sum_{j=k+1}^{n} c_j w_j = \sum_{j=1}^{k} a_j w_j$$

for some set of scalars a_i. But this is just

$$\sum_{j=1}^{k} a_j w_j - \sum_{j=k+1}^{n} c_j w_j = 0$$

and hence

$$a_1 = \cdots = a_k = c_{k+1} = \cdots = c_n = 0$$

since the w_j are linearly independent. Therefore $T(w_{k+1}), \ldots, T(w_n)$ are linearly independent and thus form a basis for Im T. We have therefore shown that

$$\dim U = k + (n - k) = \dim(\text{Ker } T) + \dim(\text{Im } T) = \text{nul } T + r(T) \,. \;\blacksquare$$

The reader should carefully compare this theorem with Theorem 3.13 and Exercise 3.6.3.

An extremely important special case of the space L(U, V) is the space L(V, V) of all linear transformations of V into itself. This space is frequently written as L(V), and its elements are usually called **linear operators** on V, or simply **operators**.

Recall that Theorem 5.2 showed that the space L(U, V) is closed with respect to addition and scalar multiplication. Furthermore, in the particular case of L(V), the composition of two functions f, g ∈ L(V) leads naturally to a "multiplication" defined by $fg = f \circ g \in L(V)$. In view of Theorems 5.2 and 5.3, it is now a simple matter to prove the following.

Theorem 5.7 The space L(V) is an associative ring.

Proof All that remains is to verify axioms (R7) and (R8) for a ring as given in Section 1.4. This is quite easy to do, and we leave it to the reader (see Exercise 5.2.1). ∎

In fact, it is easy to see that L(V) is a ring with unit element. In particular, we define the identity mapping I ∈ L(V) by I(x) = x for all x ∈ V, and hence for any T ∈ L(V) we have

$$(IT)(x) = I(T(x)) = T(x) = T(I(x)) = (TI)(x)$$

so that I commutes with every member of L(V). (However L(V) is certainly not a commutative ring in general if dim V > 1.)

An associative ring $\mathcal{A}$ is said to be an **algebra** (or **linear algebra**) over $\mathcal{F}$ if $\mathcal{A}$ is a vector space over $\mathcal{F}$ such that

$$a(ST) = (aS)T = S(aT)$$

for all $a \in \mathcal{F}$ and $S, T \in \mathcal{A}$. Another way to say this is that an algebra is a vector space on which an additional operation, called **vector multiplication**, is defined. This operation associates a new vector to each pair of vectors, and is associative, distributive with respect to addition, and obeys the rule a(ST) = (aS)T = S(aT) given above. Loosely put, an algebra is a vector space in which we can also multiply vectors to obtain a new vector. However note, for example, that the space $\mathbb{R}^3$ with the usual "dot product" defined on it does not define an algebra because $\vec{a} \cdot \vec{b}$ is a scalar. Similarly, $\mathbb{R}^3$ with the usual "cross product" is not an algebra because $(\vec{a} \times \vec{b}) \times \vec{c} \neq \vec{a} \times (\vec{b} \times \vec{c})$.

Theorem 5.8 The space L(V) is an algebra over $\mathcal{F}$.

Proof For any $a \in \mathcal{F}$, any $S, T \in L(V)$ and any $x \in V$ we have

$$(a(ST))x = a(ST)(x) = aS(T(x)) = (aS)T(x) = ((aS)T)x$$

and

$$(a(ST))x = aS(T(x)) = S(aT(x)) = S((aT)x) = (S(aT))x \ .$$

This shows that $a(ST) = (aS)T = S(aT)$ and, together with Theorem 5.7, proves the theorem. ∎

A linear transformation $T \in L(U, V)$ is said to be **invertible** if there exists a linear transformation $T^{-1} \in L(V, U)$ such that $TT^{-1} = T^{-1}T = I$ (note that technically TT^{-1} is the identity on V and $T^{-1}T$ is the identity on U). This is exactly the same definition we had in Section 3.7 for matrices. The unique mapping T^{-1} is called the inverse of T.

Theorem 5.9 A linear transformation $T \in L(U, V)$ is invertible if and only if it is a bijection (i.e., one-to-one and onto).

Proof First suppose that T is invertible. If $T(x_1) = T(x_2)$ for $x_1, x_2 \in U$, then the fact that $T^{-1}T = I$ implies

$$x_1 = T^{-1}T(x_1) = T^{-1}T(x_2) = x_2$$

and hence T is injective. If $y \in V$, then using $TT^{-1} = I$ we have

$$y = I(y) = (TT^{-1})y = T(T^{-1}(y))$$

so that $y = T(x)$ where $x = T^{-1}(y)$. This shows that T is also surjective, and hence a bijection.

Conversely, let T be a bijection. We must define a linear transformation $T^{-1} \in L(V, U)$ with the desired properties. Let $y \in V$ be arbitrary. Since T is surjective, there exists a vector $x \in U$ such that $T(x) = y$. The vector x is unique because T is injective. We may therefore define a mapping $T^{-1}\colon V \to U$ by the rule $T^{-1}(y) = x$ where $y = T(x)$. To show that T^{-1} is linear, let $y_1, y_2 \in V$ be arbitrary and choose $x_1, x_2 \in U$ such that $T(x_1) = y_1$ and $T(x_2) = y_2$. Using the linearity of T we then see that

$$T(x_1 + x_2) = y_1 + y_2$$

and hence

$$T^{-1}(y_1 + y_2) = x_1 + x_2 \; .$$

But then

$$T^{-1}(y_1 + y_2) = x_1 + x_2 = T^{-1}(y_1) + T^{-1}(y_2) \; .$$

Similarly, if $T(x) = y$ and $a \in \mathcal{F}$, then $T(ax) = aT(x) = ay$ so that

$$T^{-1}(ay) = ax = aT^{-1}(y) \; .$$

We have thus shown that $T^{-1} \in L(V, U)$. Finally, we note that for any $y \in V$ and $x \in U$ such that $T(x) = y$ we have

$$TT^{-1}(y) = T(x) = y$$

and

$$T^{-1}T(x) = T^{-1}(y) = x$$

so that $TT^{-1} = T^{-1}T = I$. ∎

A linear transformation $T \in L(U, V)$ is said to be **nonsingular** if $\mathrm{Ker}\, T = \{0\}$. In other words, T is nonsingular if it is one-to-one (Theorem 5.5). As we might expect, T is said to be **singular** if it is not nonsingular. In other words, T is singular if $\mathrm{Ker}\, T \neq \{0\}$.

Now suppose U and V are both finite-dimensional and $\dim U = \dim V$. If $\mathrm{Ker}\, T = \{0\}$, then $\mathrm{nul}\, T = 0$ and Theorem 5.6 shows that $\dim U = \dim(\mathrm{Im}\, T)$. In other words, we must have $\mathrm{Im}\, T = V$, and hence T is surjective. Conversely, if T is surjective then we are forced to conclude that $\mathrm{nul}\, T = 0$, and thus T is also injective. Hence a linear transformation between two finite-dimensional vector spaces of the same dimension is one-to-one if and only if it is onto. Combining this discussion with Theorem 5.9, we obtain the following result and its obvious corollary.

Theorem 5.10 Let U and V be finite-dimensional vector spaces such that $\dim U = \dim V$. Then the following statements are equivalent for any linear transformation $T \in L(U, V)$:

(a) T is invertible.
(b) T is nonsingular.
(c) T is surjective.

Corollary A linear operator $T \in L(V)$ on a finite-dimensional vector space is invertible if and only if it is nonsingular.

Example 5.8 Let $V = \mathcal{F}^n$ so that any $x \in V$ may be written in terms of components as $x = (x_1, \dots, x_n)$. Given any matrix $A = (a_{ij}) \in M_{m\times n}(\mathcal{F})$, we define a linear transformation $T : \mathcal{F}^n \to \mathcal{F}^m$ by $T(x) = y$ which is again given in component form by

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \;, \qquad i = 1, \dots, m \;.$$

We claim that T is one-to-one if and only if the homogeneous system

$$\sum_{j=1}^{n} a_{ij} x_j = 0 \;, \qquad i = 1, \dots, m$$

has only the trivial solution. (Note that if T is one-to-one, this is the same as requiring that the solution of the nonhomogeneous system be unique. It also follows from Corollary 5 of Theorem 3.21 that if T is one-to-one, then A is nonsingular.)

First let T be one-to-one. Clearly $T(0) = 0$, and if $v = (v_1, \dots, v_n)$ is a solution of the homogeneous system, then $T(v) = 0$. But if T is one-to-one, then $v = 0$ is the only solution. Conversely, let the homogeneous system have only the trivial solution. If $T(u) = T(v)$, then

$$0 = T(u) - T(v) = T(u - v)$$

which implies that $u - v = 0$ or $u = v$. ∕∕

Example 5.9 Let $T \in L(\mathbb{R}^2)$ be defined by

$$T(x, y) = (y, 2x - y) \;.$$

If $T(x, y) = (0, 0)$, then we must have $x = y = 0$, and hence $\operatorname{Ker} T = \{0\}$. By the corollary to Theorem 5.10, T is invertible, and we now show how to find T^{-1}.

Suppose we write $(x', y') = T(x, y) = (y, 2x - y)$. Then $y = x'$ and $2x - y = y'$ so that solving for x and y in terms of x' and y' we obtain $x = (1/2)(x' + y')$ and $y = x'$. We therefore see that

$$T^{-1}(x', y') = (x'/2 + y'/2, x') \;.$$

Note this also shows that T is surjective since for any $(x', y') \in \mathbb{R}^2$ we found a point $(x, y) = (x'/2 + y'/2, x')$ such that $T(x, y) = (x', y')$. ∕∕

Our next example shows the importance of finite-dimensionality in Theorem 5.10.

Example 5.10 Let V = $\mathcal{F}$[x], the (infinite-dimensional) space of all polynomials over $\mathcal{F}$ (see Example 2.2). For any v ∈ V with $v = \sum_{i=0}^{n} a_i x^i$ we define T ∈ L(V) by

$$T(v) = \sum_{i=1}^{n} a_i x^{i+1}$$

(this is just a "multiplication by x" operation). We leave it to the reader to show that T is linear and nonsingular (see Exercise 5.2.2). However, it is clear that T can not be surjective (for example, T takes scalars into polynomials of degree 1), so it can not be invertible. However, it is nevertheless possible to find a left inverse T_L^{-1} for T. To see this, we let T_L^{-1} be the operation of subtracting the constant term and then dividing by x:

$$T_L^{-1}(v) = \sum_{i=1}^{n} a_i x^{i-1} \quad .$$

We again leave it to the reader (Exercise 5.2.2) to show that this is a linear transformation, and that $T_L^{-1}T = I$ while $TT_L^{-1} \neq I$.

While the above operation T is an example of a nonsingular linear transformation that is not surjective, we can also give an example of a linear transformation on $\mathcal{F}$[x] that is surjective but not nonsingular. To see this, consider the operation D = d/dx that takes the derivative of every polynomial in $\mathcal{F}$[x]. It is easy to see that D is a linear transformation, but D can not possibly be nonsingular since the derivative of any constant polynomial p(x) = c is zero. Note though, that the image of D is all of $\mathcal{F}$[x], and it is in fact possible to find a right inverse of D. Indeed, if we let $D_R^{-1}(f) = \int_0^x f(t)\, dt$ be the (indefinite) integral operator, then

$$D_R^{-1}\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=0}^{n} \frac{a_i x^{i+1}}{i+1}$$

and hence $DD_R^{-1} = I$. However, it is obvious that $D_R^{-1}D \neq I$ because $D_R^{-1}D$ applied to a constant polynomial yields zero. ∕∕

Exercises

1. Finish the proof of Theorem 5.7.

2. (a) Verify that the mapping A in Example 5.8 is linear.
   (b) Verify that the mapping T in Example 5.9 is linear.
   (c) Verify that the mapping T in Example 5.10 is linear and nonsingular.
   (d) Verify that $T T_L^{-1} \neq I$ in Example 5.10.

3. Find a linear transformation T: $\mathbb{R}^3 \to \mathbb{R}^4$ whose image is generated by the vectors (1, 2, 0, −4) and (2, 0, −1, −3).

4. For each of the following linear transformations T, find the dimension and a basis for Im T and Ker T:
   (a) T: $\mathbb{R}^3 \to \mathbb{R}^3$ defined by T(x, y, z) = (x + 2y − z, y + z, x + y − 2z).
   (b) T: $\mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$T(x, y, z, t) = (x - y + z + t, x + 2z - t, x + y + 3z - 3t) \ .$$

5. Consider the space $M_2(\mathbb{R})$ of real 2 x 2 matrices, and define the matrix

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

Find the dimension and exhibit a specific basis for the kernel of the linear transformation T: $M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined by T(A) = AB − BA = [A, B].

6. Let T: U $\to$ V be a linear transformation with kernel K_T. If T(u) = v, show that $T^{-1}(v)$ is just the coset u + K_T = {u + k: k $\in$ K_T} (see Section 1.5).

7. Show that a linear transformation is nonsingular if and only if it takes linearly independent sets into linearly independent sets.

8. Consider the operator T: $\mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T(x, y, z) = (2x, 4x - y, 2x + 3y - z) \ .$$

   (a) Show that T is invertible.
   (b) Find a formula for T^{-1}.

9. Let E be a projection (or idempotent) operator on a space V, i.e., E^2 = E on V. Define U = Im E and W = Ker E. Show that:
   (a) E(u) = u for every u $\in$ U.
   (b) If E $\neq$ I, then E is singular.
   (c) V = U $\oplus$ W.

10. If $S: U \to V$ and $T: V \to U$ are nonsingular linear transformations, show that $S\,T$ is nonsingular. What can be said if S and/or T is singular?

11. Let $S: U \to V$ and $T: V \to W$ be linear transformations.
(a) Show that $T\,S: U \to W$ is linear.
(b) Show that $r(T\,S) \le r(T)$ and $r(T\,S) \le r(S)$, i.e., $r(TS) \le \min\{r(T), r(S)\}$.

12. If $S, T \in L(V)$ and S is nonsingular, show that $r(ST) = r(TS) = r(T)$.

13. If $S, T \in L(U, V)$, show that $r(S + T) \le r(S) + r(T)$. Give an example of two nonzero linear transformations $S, T \in L(U,V)$ such that $r(S + T) = r(S) + r(T)$.

14. Suppose that $V = U \oplus W$ and consider the linear operators E_1 and E_2 on V defined by $E_1(v) = u$ and $E_2(v) = w$ where $u \in U$, $w \in W$ and $v = u + w$. Show that:
(a) E_1 and E_2 are projection operators on V.
(b) $E_1 + E_2 = I$.
(c) $E_1E_2 = 0 = E_2E_1$.
(d) $V = \mathrm{Im}\,E_1 \oplus \mathrm{Im}\,E_2$.

15. Prove that the nonsingular elements in L(V) form a group.

16. Recall that an operator $T \in L(V)$ is said to be **nilpotent** if $T^n = 0$ for some positive integer n. Suppose that T is nilpotent and $T(x) = \alpha x$ for some nonzero $x \in V$ and some $\alpha \in \mathcal{F}$. Show that $\alpha = 0$.

17. If $\dim V = 1$, show that L(V) is isomorphic to $\mathcal{F}$.

18. Let $V = \mathbb{C}^3$ have the standard basis $\{e_i\}$, and let $T \in L(V)$ be defined by $T(e_1) = (1, 0, i)$, $T(e_2) = (0, 1, 1)$ and $T(e_3) = (i, 1, 0)$. Is T invertible?

19. Let V be finite-dimensional, and suppose $T \in L(V)$ has the property that $r(T^2) = r(T)$. Show that $(\mathrm{Im}\,T) \cap (\mathrm{Ker}\,T) = \{0\}$.

5.3 MATRIX REPRESENTATIONS

By now it should be apparent that there seems to be a definite similarity between Theorems 5.6 and 3.13. This is indeed the case, but to formulate this relationship precisely, we must first describe the representation of a linear transformation by matrices.

Consider a linear transformation $T \in L(U, V)$, and let U and V have bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ respectively. Since $T(u_i) \in V$, it follows from Corollary 2 of Theorem 2.4 that there exists a unique set of scalars $a_{1i}, \dots, a_{mi}$ such that

$$T(u_i) = \sum_{j=1}^{m} v_j a_{ji}$$

for each $i = 1, \dots, n$. Thus, the linear transformation T leads in a natural way to a matrix (a_{ij}) defined with respect to the given bases. On the other hand, if we are given a matrix (a_{ij}), then $\sum_{j=1}^{m} v_j a_{ji}$ is a vector in V for each $i = 1, \dots, n$. Hence, by Theorem 5.1, there exists a unique linear transformation T defined by $T(u_i) = \sum_{j=1}^{m} v_j a_{ji}$.

Now let x be any vector in U. Then $x = \sum_{i=1}^{n} x_i u_i$ so that

$$T(x) = T\left(\sum_{i=1}^{n} x_i u_i\right) = \sum_{i=1}^{n} x_i T(u_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} v_j a_{ji} x_i \ .$$

But $T(x) \in V$ so we may write

$$y = T(x) = \sum_{j=1}^{m} y_j v_j \ .$$

Since $\{v_i\}$ is a basis for V, comparing these last two equations shows that

$$y_j = \sum_{i=1}^{n} a_{ji} x_i$$

for each $j = 1, \dots, m$. The reader should note which index is summed over in this expression for y_j.

If we write out both of the systems $T(u_i) = \sum_{j=1}^{m} v_j a_{ji}$ and $y_j = \sum_{i=1}^{n} a_{ji} x_i$, we have

$$\begin{aligned} T(u_1) &= a_{11}v_1 + \cdots + a_{m1}v_m \\ &\vdots \\ T(u_n) &= a_{1n}v_1 + \cdots + a_{mn}v_m \end{aligned} \tag{1}$$

and

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{aligned} \tag{2}$$

We thus see that the matrix of coefficients in (1) is the transpose of the matrix of coefficients in (2). We shall call the m x n matrix of coefficients in equations (2) the **matrix representation** of the linear transformation T, and we say that T is **represented** by the matrix $A = (a_{ij})$ with respect to the given (ordered) bases $\{u_i\}$ and $\{v_i\}$.

We will sometimes use the notation [A] to denote the matrix corresponding to an operator $A \in L(U, V)$. This will avoid the confusion that may arise when the same letter is used to denote both the transformation and its representation matrix. In addition, if the particular bases chosen are important, then we will write the matrix representation of the above transformation as $[A]^v_u$, and if $A \in L(V)$, then we write simply $[A]_v$.

In order to make these definitions somewhat more transparent, let us make the following observation. If $x \in U$ has coordinates $(x_1, \dots, x_n)$ relative to a basis for U, and $y \in V$ has coordinates $(y_1, \dots, y_m)$ relative to a basis for V, then the expression $y = A(x)$ may be written in matrix form as $Y = [A]X$ where both X and Y are column vectors. In other words, [A]X is the coordinate vector corresponding to the result of the transformation A acting on the vector x. An equivalent way of writing this in a way that emphasizes the bases involved is

$$[y]_v = [A(x)]_v = [A]^v_u[x]_u \quad .$$

If $\{v_j\}$ is a basis for V, then we may clearly write

$$v_i = \textstyle\sum_j v_j \delta_{ji}$$

where the δ_{ji} are now to be interpreted as the components of v_i *with respect to the basis* $\{v_j\}$. In other words, v_1 has components $(1, 0, \dots, 0)$, v_2 has components $(0, 1, \dots, 0)$ and so forth. Hence, writing out $[A(u_1)]_v = \sum_{j=1}^m v_j a_{j1}$, we see that

$$[A(u_1)]_v = \begin{pmatrix} a_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_{21} \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ a_{m1} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

so that $[A(u_1)]_v$ is just the first column of $[A]^v_u$. Similarly, it is easy to see that in general, $[A(u_i)]_v$ is the i*th* column of $[A]^v_u$. In other words, *the matrix representation* $[A]^v_u$ *of a linear transformation* $A \in L(U, V)$ *has columns that are nothing more than the images under* A *of the basis vectors of* U.

We summarize this very important discussion as a theorem for easy reference.

Theorem 5.11 Let U and V have bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ respectively. Then for any $A \in L(U, V)$ the vector

$$[A(u_i)]_v = \sum_{j=1}^{m} v_j a_{ji}$$

is the i*th* column of the matrix $[A]_u^v = (a_{ij})$ that represents A relative to the given bases.

Example 5.11 Let V have a basis $\{v_1, v_2, v_3\}$, and let $A \in L(V)$ be defined by

$$\begin{aligned} A(v_1) &= 3v_1 \phantom{{}-2v_2} + v_3 \\ A(v_2) &= v_1 - 2v_2 - v_3 \\ A(v_3) &= \phantom{v_1 - {}} v_2 + v_3 \end{aligned}$$

Then the representation of A (relative to this basis) is

$$[A]_v = \begin{pmatrix} 3 & 1 & 0 \\ 0 & -2 & 1 \\ 1 & -1 & 1 \end{pmatrix} . \quad //$$

The reader may be wondering why we wrote $A(u_i) = \sum_j v_j a_{ji}$ rather than $A(u_i) = \sum_j a_{ij} v_j$. The reason is that we want the matrix corresponding to a combination of linear transformations to be the product of the individual matrix representations taken in the same order. (The argument that follows is based on what we learned in Chapter 3 about matrix multiplication, even though technically we have not yet defined this operation within the framework of our current discussion. In fact, our present formulation can be taken as the *definition* of matrix multiplication.)

To see what this means, suppose $A, B \in L(V)$. If we had written (note the order of subscripts) $A(v_i) = \sum_j a_{ij} v_j$ and $B(v_i) = \sum_j b_{ij} v_j$, then we would have found that

$$\begin{aligned} (AB)(v_i) &= A(B(v_i)) = A(\textstyle\sum_j b_{ij} v_j) = \sum_j b_{ij} A(v_j) \\ &= \textstyle\sum_{j,\, k} b_{ij} a_{jk} v_k = \sum_k c_{ik} v_k \end{aligned}$$

where $c_{ik} = \sum_j b_{ij} a_{jk}$. As a matrix product, we would then have $[C] = [B][A]$. However, if we write (as we did) $A(v_i) = \sum_j v_j a_{ji}$ and $B(v_i) = \sum_j v_j b_{ji}$, then we obtain

$$\begin{aligned} (AB)(v_i) &= A(B(v_i)) = A(\textstyle\sum_j v_j b_{ji}) = \sum_j A(v_j) b_{ji} \\ &= \textstyle\sum_{j,\, k} v_k a_{kj} b_{ji} = \sum_k v_k c_{ki} \end{aligned}$$

where now $c_{ki} = \sum_j a_{kj} b_{ji}$. Since the matrix notation for this is $[C] = [A][B]$, we see that the order of the matrix representation of transformations is preserved as desired. We have therefore proven the following result.

Theorem 5.12 For any operators $A, B \in L(V)$ we have $[AB] = [A][B]$.

From equations (2) above, we see that any nonhomogeneous system of m linear equations in n unknowns defines an m x n matrix (a_{ij}). According to our discussion, this matrix should also define a linear transformation in a consistent manner.

Example 5.12 Consider the space $\mathbb{R}^2$ with the standard basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

so that any $X \in \mathbb{R}^2$ may be written as

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \; .$$

Suppose we have the system of equations

$$\begin{aligned} y_1 &= 2x_1 - x_2 \\ y_2 &= x_1 + 3x_2 \end{aligned}$$

which we may write in matrix form as $[A]X = Y$ where

$$[A] = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \; .$$

Hence we have a linear transformation $A(x) = [A]X$. In particular,

$$A(e_1) = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2e_1 + e_2$$

$$A(e_2) = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix} = -e_1 + 3e_2 \; .$$

We now see that letting the *ith* column of [A] be $A(e_i)$, we arrive back at the original form [A] that represents the linear transformation $A(e_1) = 2e_1 + e_2$ and $A(e_2) = -e_1 + 3e_2$. ∕∕

Example 5.13 Consider the space $V = \mathbb{R}^2$ with basis vectors $v_1 = (1, 1)$ and $v_2 = (-1, 0)$. Let T be the linear operator on $\mathbb{R}^2$ defined by

$$T(x, y) = (4x - 2y, 2x + y) \;.$$

To find the matrix of T relative to the given basis, all we do is compute the effect of T on each basis vector:

$$\begin{aligned} T(v_1) &= T(1, 1) = (2, 3) = 3v_1 + v_2 \\ T(v_2) &= T(-1, 0) = (-4, -2) = -2v_1 + 2v_2 \;. \end{aligned}$$

Since the matrix of T has columns given by the image of each basis vector, we must have

$$[T] = \begin{pmatrix} 3 & -2 \\ 1 & 2 \end{pmatrix} \;. \;∕∕$$

Theorem 5.13 Let U and V be vector spaces over $\mathcal{F}$ with bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ respectively. Suppose $A \in L(U, V)$ and let [A] be the matrix representation of A with respect to the given bases. Then the mapping $\phi\colon A \to [A]$ is an isomorphism of L(U, V) onto the vector space $M_{m\times n}(\mathcal{F})$ of all m x n matrices over $\mathcal{F}$.

Proof Part of this was proved in the discussion above, but for ease of reference, we repeat it here. Given any $(a_{ij}) \in M_{m\times n}(\mathcal{F})$, we define the linear transformation $A \in L(U, V)$ by

$$A(u_i) = \sum_{j=1}^{m} v_j a_{ji}$$

for each $i = 1, \dots, n$. According to Theorem 5.1, the transformation A is uniquely defined and is in L(U, V). By definition, $[A] = (a_{ij})$, and hence ϕ is surjective. On the other hand, given any $A \in L(U, V)$, it follows from Corollary 2 of Theorem 2.4 that for each $i = 1, \dots, n$ there exists a unique set of scalars $a_{1i}, \dots, a_{mi} \in \mathcal{F}$ such that $A(u_i) = \sum_{j=1}^{m} v_j a_{ji}$. Therefore, any $A \in L(U, V)$ has lead to a unique matrix $(a_{ij}) \in M_{m\times n}(\mathcal{F})$. Combined with the previous result that ϕ is surjective, this shows that ϕ is injective and hence a bijection. Another way to see this is to note that if we also have $B \in L(U, V)$ with $[B] = [A]$, then

$$(B-A)(u_i) = B(u_i) - A(u_i) = \sum_{j=1}^{m} v_j(b_{ji} - a_{ji}) = 0 \;\; .$$

Since B − A is linear (Theorem 5.3), it follows that $(B - A)x = 0$ for all $x \in U$, and hence $B = A$ so that ϕ is one-to-one.

Finally, to show that ϕ is an isomorphism we must show that it is also a vector space homomorphism (i.e., a linear transformation). But this is easy if we simply observe that

$$(A + B)(u_i) \;=\; A(u_i) + B(u_i) \;=\; \textstyle\sum_j v_j a_{ji} + \sum_j v_j b_{ji} \;=\; \sum_j v_j(a_{ji} + b_{ji})$$

and, for any $c \in \mathcal{F}$,

$$(cA)(u_i) \;=\; c(A(u_i)) \;=\; c(\textstyle\sum_j v_j\, a_{ji}) \;=\; \sum_j v_j(ca_{ji}) \;\; .$$

Therefore we have shown that

$$[A + B] \;=\; [A] + [B]$$

and

$$[cA] \;=\; c[A]$$

so that ϕ is a homomorphism. ∎

It may be worth recalling that the space $M_{m\mathrm{x}n}(\mathcal{F})$ is clearly of dimension mn since, for example, we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \;\; .$$

Therefore Theorem 5.13 provides another proof that dim L(U, V) = mn.

Let us return again to the space L(V) = L(V, V) where dim V = n. In this case, each linear operator $A \in L(V)$ will be represented by an n x n matrix, and we then see that the space $M_n(\mathcal{F}) = M_{n\mathrm{x}n}(\mathcal{F})$ of all n x n matrices over $\mathcal{F}$ is closed under addition, multiplication, and scalar multiplication. By Theorem 5.13, L(V) is isomorphic to $M_n(\mathcal{F})$, and this isomorphism preserves addition and scalar multiplication. Furthermore, it also preserves the multiplication of operators since this was the motivation behind how we defined matrix representations (and hence matrix multiplication). Finally, recall that the identity transformation $I \in L(V)$ was defined by $I(x) = x$ for all $x \in V$. In particular

$$I(u_i) \;=\; \textstyle\sum_j u_j \delta_{ji}$$

so that the matrix representation of I is just the usual n x n identity matrix that commutes with every other n x n matrix.

Theorem 5.14 The space $M_n(\mathcal{F})$ of all n x n matrices over $\mathcal{F}$ is a linear algebra.

Proof Since $M_n(\mathcal{F})$ is isomorphic to L(V) where dim V = n, this theorem follows directly from Theorem 5.8. ∎

We now return to the relationship between Theorems 5.6 and 3.13. In particular, we would like to know how the rank of a linear transformation is related to the rank of a matrix. The answer was essentially given in Theorem 5.11.

Theorem 5.15 If $A \in L(U, V)$ is represented by $[A] = (a_{ji}) \in M_{m \times n}(\mathcal{F})$, then $r(A) = r([A])$.

Proof Recall that r(A) = dim(Im A) and r([A]) = cr([A]). For any $x \in U$ we have

$$A(x) = A(\textstyle\sum x_i u_i) = \sum x_i A(u_i)$$

so that the $A(u_i)$ span Im A. But $[A(u_i)]$ is just the *ith* column of [A], and hence the $[A(u_i)]$ also span the column space of [A]. Therefore the number of linearly independent columns of [A] is the same as the number of linearly independent vectors in the image of A (see Exercise 5.3.1). This means that r(A) = cr([A]) = r([A]). ∎

Suppose that we have a system of n linear equations in n unknowns written in matrix form as [A]X = Y where [A] is the matrix representation of the corresponding linear transformation $A \in L(V)$, and dim V = n. If we are to solve this for a unique X, then [A] must be of rank n (Theorem 3.16). Hence r(A) = n also so that nul A = dim(Ker A) = 0 by Theorem 5.6. But this means that Ker A = {0} and thus A is nonsingular. Note also that Theorem 3.13 now says that the dimension of the solution space is zero (which it must be for the solution to be unique) which agrees with Ker A = {0}.

All of this merely shows the various interrelationships between the matrix nomenclature and the concept of a linear transformation that should be expected in view of Theorem 5.13. Our discussion is summarized by the following useful characterization.

Theorem 5.16 A linear transformation $A \in L(V)$ is nonsingular if and only if $\det [A] \neq 0$.

Proof Let dim V = n. If A is nonsingular then nul A = 0, and hence $r([A]) = r(A) = n$ (Theorem 5.6) so that $[A]^{-1}$ exists (Theorem 3.21). But this means that $\det [A] \neq 0$ (Theorem 4.6). The converse follows by an exact reversal of the argument. ∎

Exercises

1. Suppose $A \in L(U, V)$ and let $\{u_i\}$, $\{v_i\}$ be bases for U and V respectively. Show directly that $\{A(u_i)\}$ is linearly independent if and only if the columns of [A] are also linearly independent.

2. Let V be the space of all real polynomials of degree less than or equal to 3. In other words, elements of V are of the form $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ where each $a_i \in \mathbb{R}$.
(a) Show that the derivative mapping $D = d/dx$ is an element of L(V).
(b) Find the matrix of D relative to the ordered basis $\{f_i\}$ for V defined by $f_i(x) = x^{i-1}$.

3. Let $T: \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x, y, z) = (x + y, 2z - x)$.
(a) Find the matrix of T relative to the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$.
(b) Find the matrix of T relative to the basis $\{\alpha_i\}$ for $\mathbb{R}^3$ and $\{\beta_i\}$ for $\mathbb{R}^2$ where $\alpha_1 = (1, 0, -1)$, $\alpha_2 = (1, 1, 1)$, $\alpha_3 = (1, 0, 0)$, $\beta_1 = (0, 1)$ and $\beta_2 = (1, 0)$.

4. Relative to the standard basis, let $T \in L(\mathbb{R}^3)$ have the matrix representation
$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \end{pmatrix}.$$
Find a basis for Im T and Ker T.

5. Let $T \in L(\mathbb{R}^3)$ be defined by $T(x, y, z) = (3x + z, -2x + y, -x + 2y + 4z)$.
(a) Find the matrix of T relative to the standard basis for $\mathbb{R}^3$.
(b) Find the matrix of T relative to the basis $\{\alpha_i\}$ given by $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (-1, 2, 1)$ and $\alpha_3 = (2, 1, 1)$.

(c) Show that T is invertible, and give a formula for T^{-1} similar to that given in part (a) for T.

6. Let $T: \mathcal{F}^n \to \mathcal{F}^m$ be the linear transformation defined by

$$T(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} a_{1i}x_i, \dots, \sum_{i=1}^{n} a_{mi}x_i \right) .$$

(a) Show that the matrix of T relative to the standard bases of $\mathcal{F}^n$ and $\mathcal{F}^m$ is given by

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} .$$

(b) Find the matrix representation of $T: \mathbb{R}^4 \to \mathbb{R}^2$ defined by

$$T(x, y, z, t) = (3x - 4y + 2z - 5t, 5x + 7y - z - 2t)$$

relative to the standard bases of $\mathbb{R}^n$.

7. Suppose that $T \in L(U, V)$ has rank r. Prove that there exists a basis for U and a basis for V relative to which the matrix of T takes the form

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

[*Hint*: Show that Ker T has a basis $\{w_1, \dots, w_{m-r}\}$, and then extend this to a basis $\{u_1, \dots, u_r, w_1, \dots, w_{m-r}\}$ for U. Define $v_i = T(u_i)$, and show that this is a basis for Im T. Now extend this to a basis for V.]

8. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, and let $\{f_i\}$ be the standard basis for $\mathbb{R}^2$.

(a) Define $T: \mathbb{R}^3 \to \mathbb{R}^2$ by $T(e_1) = f_2$, $T(e_2) = f_1$ and $T(e_3) = f_1 + f_2$. Write down the matrix $[T]_e^f$.

(b) Define $S: \mathbb{R}^2 \to \mathbb{R}^3$ by $S(f_1) = (1, 2, 3)$ and $S(f_2) = (2, -1, 4)$. Write down $[S]_f^e$.

(c) Find $ST(e_i)$ for each i = 1, 2, 3, and write down the matrix $[ST]_e$ of the linear operator $ST: \mathbb{R}^3 \to \mathbb{R}^3$. Verify that $[ST] = [S][T]$.

9. Suppose $T \in L(V)$ and let W be a subspace of V. We say that W is **invariant under** T (or **T-invariant**) if $T(W) \subset W$. If dim W = m, show that T has a block matrix representation of the form

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where A is an m x m matrix.

10. Let $T \in L(V)$, and suppose that $V = U \oplus W$ where both U and W are T-invariant (see the previous problem). If dim U = m and dim W = n, show that T has a matrix representation of the form

$$\begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix}$$

where A is an m x m matrix and B is an n x n matrix.

11. Show that $A \in L(V)$ is nonsingular implies $[A^{-1}] = [A]^{-1}$.

5.4 CHANGE OF BASIS

Suppose we have a linear operator $A \in L(V)$. Then, given a basis for V, we can write down the corresponding matrix [A]. If we change to a new basis for V, then we will have a new representation for A. We now investigate the relationship between the matrix representations of A in each of these bases.

Given a vector space V, let us consider two arbitrary bases $\{e_1, \ldots, e_n\}$ and $\{\bar{e}_1, \ldots, \bar{e}_n\}$ for V. Then any vector $x \in V$ may be written as either $x = \sum x_i e_i$ or as $x = \sum \bar{x}_i \bar{e}_i$. (It is important to realize that vectors and linear transformations exist independently of the coordinate system used to describe them, and their components may vary from one coordinate system to another.) Since each $\bar{e}_i$ is a vector in V, we may write its components in terms of the basis $\{e_i\}$. In other words, we define the **transition matrix** $[P] = (p_{ij}) \in M_n(\mathcal{F})$ by

$$\bar{e}_i = \sum_{j=1}^{n} e_j p_{ji}$$

for each i = 1, . . . , n. The matrix [P] must be unique for the given bases according to Corollary 2 of Theorem 2.4.

Note that [P] defines a linear transformation $P \in L(V)$ by $P(e_i) = \bar{e}_i$. Since $\{P(e_i)\} = \{\bar{e}_i\}$ spans Im P and the $\bar{e}_i$ are linearly independent, it follows that

$r(P) = n$ so that P is nonsingular and hence P^{-1} exists. By Theorem 5.13, we conclude that $[P^{-1}] = [P]^{-1}$. (However, it is also quite simple to show directly that if a linear operator A is nonsingular, then $[A^{-1}] = [A]^{-1}$. See Exercise 5.3.11).

Let us emphasize an earlier remark. From Theorem 5.11, we know that $[\bar{e}_i] = [P(e_i)]$ is just the *ith* column vector of [P]. Since relative to the basis $\{e_i\}$ we have $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$ and so on, it follows that the *ith* column of [P] represents the components of $\bar{e}_i$ relative to the basis $\{e_i\}$. In other words, the matrix entry p_{ji} is the *jth* component of the *ith* basis vector $\bar{e}_i$ relative to the basis $\{e_i\}$.

The transition matrix enables us to easily relate the components of any $x \in V$ between the two coordinate systems. To see this, we observe that

$$x = \textstyle\sum_i x_i e_i = \sum_j \bar{x}_j \bar{e}_j = \sum_{i,j} \bar{x}_j e_i p_{ij} = \sum_{i,j} p_{ij} \bar{x}_j e_i$$

and hence the uniqueness of the expansion implies $x_i = \sum_j p_{ij} \bar{x}_j$ so that

$$\bar{x}_j = \textstyle\sum_i p^{-1}{}_{ji} x_i \ .$$

This discussion proves the following theorem.

Theorem 5.17 Let [P] be the transition matrix from a basis $\{e_i\}$ to a basis $\{\bar{e}_i\}$ for a space V. Then for any $x \in V$ we have

$$[x]_{\bar{e}} = [P]^{-1}[x]_e$$

which we sometimes write simply as $\bar{X} = P^{-1}X$.

From now on we will omit the brackets on matrix representations unless they are needed for clarity. Thus we will usually write both a linear transformation $A \in L(U, V)$ and its representation $[A] \in M_{m \times n}(\mathcal{F})$ as simply A. Furthermore, to avoid possible ambiguity, we will sometimes denote a linear transformation by T, and its corresponding matrix representation by $A = (a_{ij})$.

Using the above results, it is now an easy matter for us to relate the representation of a linear operator $A \in L(V)$ in one basis to its representation in another basis. If $A(e_i) = \sum_j e_j a_{ji}$ and $A(\bar{e}_i) = \sum_j \bar{e}_j \bar{a}_{ji}$, then on the one hand we have

$$A(\bar{e}_i) = \textstyle\sum_j \bar{e}_j \bar{a}_{ji} = \sum_{j,k} e_k p_{kj} \bar{a}_{ji}$$

while on the other hand,

$$A(\bar{e}_i) = A(\textstyle\sum_j e_j p_{ji}) = \sum_j A(e_j) p_{ji} = \sum_{j,k} e_k a_{kj} p_{ji} \ .$$

Therefore, since $\{e_k\}$ is a basis for V, we may equate each component in these two equations to obtain $\sum_j p_{kj}\bar{a}_{ji} = \sum_j a_{kj}p_{ji}$ or

$$\bar{a}_{ri} = \sum_{j,\,k} p^{-1}{}_{rk}a_{kj}p_{ji} \quad .$$

In matrix notation, this is just (omitting the brackets on P)

$$[A]_{\bar{e}} = P^{-1}[A]_e P$$

which we will usually write in the form $\bar{A} = P^{-1}AP$ for simplicity.

If A, B $\in M_n(\mathcal{F})$, then B is said to be **similar** to A if there exists a nonsingular matrix S such that $B = S^{-1}AS$, in which case A and B are said to be related by a **similarity transformation**. We leave it to the reader to show that this defines an equivalence relation on $M_n(\mathcal{F})$ (see Exercise 5.4.1).

Since we have shown that in two different bases a linear operator A is represented by two similar matrices, we might wonder whether or not there are any other matrices representing A that are not similar to the others. The answer is given by the following.

Theorem 5.18 If $T \in L(V)$ is represented by A relative to the basis $\{e_i\}$, then a matrix $\bar{A} \in M_n(\mathcal{F})$ represents T relative to some basis $\{\bar{e}_i\}$ if and only if $\bar{A}$ is similar to A. If this is the case, then

$$\bar{A} = P^{-1}AP$$

where P is the transition matrix from the basis $\{e_i\}$ to the basis $\{\bar{e}_i\}$.

Proof The discussion above showed that if A and $\bar{A}$ represent T in two different bases, then $\bar{A} = P^{-1}AP$ where P is the transition matrix from $\{e_i\}$ to $\{\bar{e}_i\}$.

On the other hand, suppose that T is represented by A in the basis $\{e_i\}$, and assume that $\bar{A}$ is similar to A. Then $\bar{A} = P^{-1}AP$ for some nonsingular matrix $P = (p_{ij})$. We define a new basis $\{\bar{e}_i\}$ for V by

$$\bar{e}_i = P(e_i) = \sum_j e_j p_{ji}$$

(where we use the same symbol for both the operator P and its matrix representation). Then

$$T(\bar{e}_i) = T(\textstyle\sum_j e_j p_{ji}) = \sum_j T(e_j)p_{ji} = \sum_{j,\,k} e_k a_{kj} p_{ji}$$

while on the other hand, if T is represented by some matrix $C = (c_{ji})$ in the basis $\{\bar{e}_i\}$, then

$$T(\bar{e}_i) = \sum_j \bar{e}_j c_{ji} = \sum_{j,\, k} e_k p_{kj} c_{ji} \ .$$

Equating the coefficients of e_k in both of these expressions yields

$$\sum_j a_{kj} p_{ji} = \sum_j p_{kj} c_{ji}$$

so that

$$c_{ri} = \sum_{j,\, k} p^{-1}{}_{rk} a_{kj} p_{ji}$$

and hence

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

Therefore $\bar{A}$ represents T in the basis $\{\bar{e}_i\}$. ∎

Note that by Theorem 4.8 and its corollary we have

$$\det \bar{A} = \det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A$$

and hence all matrices which represent a linear operator T have the same determinant. Another way of stating this is to say that the determinant is **invariant** under a similarity transformation. We thus define the **determinant of a linear operator** $T \in L(V)$ as det A, where A is any matrix representing T.

Another important quantity associated with a matrix $A \in M_n(\mathcal{F})$ is the sum $\sum_{i=1}^n a_{ii}$ of its diagonal elements. This sum is called the **trace**, and is denoted by Tr A (see Exercise 3.6.7). A simple but useful result is the following.

Theorem 5.19 If $A, B \in M_n(\mathcal{F})$, then $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

Proof We simply compute

$$\begin{aligned}\mathrm{Tr}(AB) &= \Sigma_i (AB)_{ii} = \Sigma_{i,\, j} a_{ij} b_{ji} = \Sigma_j \Sigma_i b_{ji} a_{ij} = \Sigma_j (BA)_{jj} \\ &= \mathrm{Tr}(BA) \ . \ \blacksquare\end{aligned}$$

From this theorem it is easy to show that the trace is also invariant under a similarity transformation (see Exercise 4.2.14). Because of this, it also makes sense to speak of the trace of a linear operator.

Example 5.14 Consider the space $V = \mathbb{R}^2$ with its standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$, and let $\bar{e}_1 = (1, 2)$, $\bar{e}_2 = (3, -1)$ be another basis. We then see that

$$\begin{aligned}\bar{e}_1 &= e_1 + 2e_2\\ \bar{e}_2 &= 3e_1 - e_2\end{aligned}$$

and consequently the transition matrix P from $\{e_i\}$ to $\{\bar{e}_i\}$ and its inverse P^{-1} are given by

$$P = \begin{pmatrix} 1 & 3\\ 2 & -1\end{pmatrix} \qquad \text{and} \qquad P^{-1} = \begin{pmatrix} 1/7 & 3/7\\ 2/7 & -1/7\end{pmatrix}\ .$$

Note that P^{-1} may be found either using Theorem 4.11, or by solving for $\{e_i\}$ in terms of $\{\bar{e}_i\}$ to obtain

$$\begin{aligned}e_1 &= (1/7)\bar{e}_1 + (2/7)\bar{e}_2\\ e_2 &= (3/7)\bar{e}_1 - (1/7)\bar{e}_2\end{aligned}$$

Now let T be the operator defined by

$$\begin{aligned}T(e_1) &= (20/7)e_1 - (2/7)e_2\\ T(e_2) &= (-3/7)e_1 + (15/7)e_2\end{aligned}$$

so that relative to the basis $\{e_i\}$ we have

$$A = \begin{pmatrix} 20/7 & -3/7\\ -2/7 & 15/7\end{pmatrix}\ .$$

We thus find that

$$\bar{A} = P^{-1}AP = \begin{pmatrix} 1/7 & 3/7\\ 2/7 & -1/7\end{pmatrix}\begin{pmatrix} 20/7 & -3/7\\ -2/7 & 15/7\end{pmatrix}\begin{pmatrix} 1 & 3\\ 2 & -1\end{pmatrix} = \begin{pmatrix} 2 & 0\\ 0 & 3\end{pmatrix}\ .$$

Alternatively, we have

$$\begin{aligned}T(\bar{e}_1) &= T(e_1 + 2e_2) = T(e_1) + 2T(e_2) = 2e_1 + 4e_2 = 2\bar{e}_1\\ T(\bar{e}_2) &= T(3e_1 - e_2) = 3T(e_1) - T(e_2) = (63/7)e_1 - 3e_2 = 3\bar{e}_2\end{aligned}$$

so that again we find

$$\bar{A} = \begin{pmatrix} 2 & 0\\ 0 & 3\end{pmatrix}\ .$$

We now see that

$$\mathrm{Tr}\,A = 20/7 + 15/7 = 5 = \mathrm{Tr}\,\bar{A}$$

and also

$$\det A = 6 = \det \bar{A}$$

as they should. ∕∕

We point out that in this example, $\bar{A}$ turns out to be a diagonal matrix. In this case the basis $\{\bar{e}_i\}$ is said to **diagonalize** the operator T. While it is certainly *not* true that there always exists a basis in which every operator is diagonal, we will spend a considerable amount of time in Chapters 7 and 8 investigating the various standard forms (called **normal** or **canonical**) that a matrix representation of an operator can take.

Let us make one related additional comment about our last example. While it is true that (algebraically speaking) a linear operator is completely determined once its effect on a basis is known, there is no real geometric interpretation of this when the matrix representation of an operator is of the same form as A in Example 5.14. However, if the representation is diagonal as it is with $\bar{A}$, then in this basis the operator represents a magnification factor in each direction. In other words, we see that $\bar{A}$ represents a multiplication of any vector in the $\bar{e}_1$ direction by 2, and a multiplication of any vector in the $\bar{e}_2$ direction by 3. This is the physical interpretation that we will attach to eigenvalues (see Chapter 7).

Exercises

1. Show that the set of similar matrices defines an equivalence relation on $M_n(\mathcal{F})$.

2. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, and consider the basis $f_1 = (1, 1, 1)$, $f_2 = (1, 1, 0)$ and $f_3 = (1, 0, 0)$.
   (a) Find the transition matrix P from $\{e_i\}$ to $\{f_i\}$.
   (b) Find the transition matrix Q from $\{f_i\}$ to $\{e_i\}$.
   (c) Verify that $Q = P^{-1}$.
   (d) Show that $[v]_f = P^{-1}[v]_e$ for any $v \in \mathbb{R}^3$.
   (e) Define $T \in L(\mathbb{R}^3)$ by $T(x, y, z) = (2y + z, x - 4y, 3x)$. Show that $[T]_f = P^{-1}[T]_e P$.

3. Let $\{e_1, e_2\}$ be a basis for V, and define $T \in L(V)$ by $T(e_1) = 3e_1 - 2e_2$ and $T(e_2) = e_1 + 4e_2$. Define the basis $\{f_i\}$ for V by $f_1 = e_1 + e_2$ and $f_2 = 2e_1 + 3e_2$. Find $[T]_f$.

4. Consider the field $\mathbb{C}$ as a vector space over $\mathbb{R}$, and define the linear "conjugation operator" $T \in L(\mathbb{C})$ by $T(z) = z^*$ for each $z \in \mathbb{C}$.
(a) Find the matrix of T relative to the basis $\{e_j\} = \{1, i\}$.
(b) Find the matrix of T relative to the basis $\{f_j\} = \{1 + i, 1 + 2i\}$.
(c) Find the transition matrices P and Q that go from $\{e_j\}$ to $\{f_j\}$ and from $\{f_j\}$ to $\{e_j\}$ respectively.
(d) Verify that $Q = P^{-1}$.
(e) Show that $[T]_f = P^{-1}[T]_e P$.
(f) Verify that $\mathrm{Tr}\, [T]_f = \mathrm{Tr}\, [T]_e$ and $\det [T]_f = \det [T]_e$.

5. Let $\{e_i\}$, $\{f_i\}$ and $\{g_i\}$ be bases for V, and let P and Q be the transition matrices from $\{e_i\}$ to $\{f_i\}$ and from $\{f_i\}$ to $\{g_i\}$ respectively. Show that PQ is the transition matrix from $\{e_i\}$ to $\{g_i\}$.

6. Let A be a 2 x 2 matrix such that only A is similar to itself. Show that A has the form

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

7. Show that similar matrices have the same rank.

8. Let A, B and C be linear operators on $\mathbb{R}^2$ with the following matrices relative to the standard basis $\{e_i\}$:

$$[A]_e = \begin{pmatrix} 4 & 6 \\ -2 & -3 \end{pmatrix} \qquad [B]_e = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \qquad [C]_e = \begin{pmatrix} 7 & 3 \\ -10 & -4 \end{pmatrix} .$$

(a) If $f_1 = (2, -1)$ and $f_2 = (3, -2)$, show that $A(f_1) = f_1$ and $A(f_2) = 0$.
(b) Find $[A]_f$.
(c) What is the geometric effect of A?
(d) Show that B is a rotation about the origin of the xy-plane, and find the angle of rotation (see Example 1.2).
(e) If $f_1 = (1, -2)$ and $f_2 = (3, -5)$, find $C(f_1)$ and $C(f_2)$.
(f) Find $[C]_f$.
(g) What is the geometric effect of C?

9. (a) Let $\{e_i\}$ be the standard basis for $\mathbb{R}^n$, and let $\{f_i\}$ be any other ortho– normal basis (relative to the standard inner product). Show that the transition matrix P from $\{e_i\}$ to $\{f_i\}$ is **orthogonal**, i.e., $P^T = P^{-1}$.

(b) Let $T \in L(\mathbb{R}^3)$ have the following matrix relative to the standard basis:

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

Find the matrix of T relative to the basis $f_1 = (2/3, 2/3, -1/3)$, $f_2 = (1/3, -2/3, -2/3)$ and $f_3 = (2/3, -1/3, 2/3)$.

10. Let $T \in L(\mathbb{R}^2)$ have the following matrix relative to the standard basis $\{e_i\}$ for $\mathbb{R}^2$:

$$[T]_e = \begin{pmatrix} a & b \\ c & d \end{pmatrix} .$$

(a) Suppose there exist two linearly independent vectors f_1 and f_2 in $\mathbb{R}^2$ with the property that $T(f_1) = \lambda_1 f_1$ and $T(f_2) = \lambda_2 f_2$ (where $\lambda_i \in \mathbb{R}$). If P is the transition matrix from the basis $\{e_i\}$ to the basis $\{f_i\}$, show that

$$[T]_f = P^{-1}[T]_e P = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} .$$

(b) Prove there exists a nonzero vector $x \in \mathbb{R}^2$ with the property that $T(x) = x$ if and only if

$$\begin{vmatrix} a-1 & b \\ c & d-1 \end{vmatrix} = 0$$

(c) Prove there exists a one-dimensional T-invariant subspace of $\mathbb{R}^2$ if and only if

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = 0$$

for some scalar λ. (Recall that a subspace W is T-invariant if $T(W) \subset W$.)

11. If $\theta \in \mathbb{R}$, show that the matrices

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

and

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

are similar over the complex field. [*Hint*: Suppose $T \in L(\mathbb{C}^2)$ has the first matrix as its representation relative to the standard basis. Find a new basis $\{v_1, v_2\}$ such that $T(v_1) = \exp(i\theta)v_1$ and $T(v_2) = \exp(-i\theta)v_2$.]

12. Let $V = \mathbb{R}^2$ have basis vectors $e_1 = (1, 1)$ and $e_2 = (1, -1)$. Suppose we define another basis for V by $\bar{e}_1 = (2, 4)$ and $\bar{e}_2 = (3, 1)$. Define the transition operator $P \in L(V)$ as usual by $\bar{e}_i = Pe_i$. Write down the matrix $[P]_e^{\bar{e}}$.

13. Let U have bases $\{u_i\}$ and $\{\bar{u}_i\}$ and let V have bases $\{v_i\}$ and $\{\bar{v}_i\}$. Define the transition operators $P \in L(U)$ and $Q \in L(V)$ by $\bar{u}_i = Pu_i$ and $\bar{v}_i = Qv_i$. If $T \in L(U, V)$, express $[T]_u^v$ in terms of $[T]_{\bar{u}}^{\bar{v}}$.

14. Show that the transition matrix defined by the Gram-Schmidt process is upper-triangular with strictly positive determinant.