Chapter 4

LINEAR TRANSFORMATIONS AND THEIR MATRICES

4.1 LINEAR TRANSFORMATIONS

The central objective of linear algebra is the analysis of linear functions defined on a finite-dimensional vector space. For example, analysis of the shear transformation is a problem of this sort. First we define the concept of a linear function or transformation.

Definition 4.1.1. Let V and W be real vector spaces (their dimensions can be different), and let T be a function with domain V and range in W (written $T\colon\ V \to W$). We say T is a **linear transformation** if

(a) For all $\mathbf{x}, \mathbf{y} \in V$, $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$($T$ is **additive**).

(b) For all $\mathbf{x} \in V, r \in \mathbb{R}, T(r\mathbf{x}) = rT(\mathbf{x})$($T$ is **homogeneous**).

If V and W are complex vector spaces, the definition is the same except in (**b**), $r \in \mathbb{C}$. If $V = W$, then T can be called a **linear operator**.

Example 1. Let $V = W = E^1$. Define $T(x) = mx$, where m is a fixed real number. Show that T is a linear transformation.

Solution We must show that T is additive and homogeneous. For the additivity, we let x and y be in E^1 and calculate

$$T(x+y) = m(x+y) = mx + my$$
$$T(x) + T(y) = mx + my$$

Since $T(x+y) = T(x) + T(y)$, we know that T is additive. Also T is homogeneous since

$$T(rx) = m(rx) = (mr)x = r(mx) = rT(x)$$

Thus T is a linear transformation.

Example 2. Let $V = W = E^1$. For $x \in V$, define $F(x) = mx + b$, where m and b are real numbers and $b \neq 0$. Show that F is **not** a linear transformation.

Solution First we check additivity, noting $F(\cdot) = m(\cdot) + b$:

$$F(x+y) = m(x+y) + b = mx + my + b$$

However,

$$F(x) + F(y) = (mx+b) + (my+b) = mx + my + 2b$$

Since $b \neq 0, 2b \neq b$ so $F(x+y) \neq F(x) + F(y)$ for all $x, y \in V$, and F is not linear.

Example 3. Let $V = \mathcal{P}_n$ and $W = \mathcal{P}_{n-1}$, and define, for f in $V, T\colon\ V \to W$ by $(T(f)) = f'(x)x$ in $\mathbb{R}$. That is, T is differentiation. From calculus, we know that for differentiable functions f and $g, (f+g)' = f' + g'$ and $(rf)' = rf'$, so T is linear.

Example 4. Let $V = \{$real-valued functions defined and continuous on $[a,b]\} = C[a,b]$. Let $W = E^1$ and define $T\colon\ V \to W$ by $T(f) = \int_a^b f(x)\ dx$. Then T is linear because from calculus we know that for integrable functions f and g,

$\int_a^b f(x) + g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$ and $\int_a^b rf(x)\ dx = r\int_a^b f(x)\ dx$, for r in $\mathbb{R}$.

Example 5. Let $V = \mathcal{C}_{nn}$ and let $W = \mathbb{C}^1$. Define $T\colon\ V \to W$ by $T(A) = \text{tr}\ A$, for A in $\mathcal{C}_{nn}$. So T is linear by properties of the trace of a matrix.

Example 6. Let $V = \mathcal{M}_{mn}$ and $W = \mathcal{M}_{nm}$, and define $T(A) = A^T$ for A in $\mathcal{M}_{mn}$. Then T is linear by properties of the transpose.

In Examples 1 and 2, the functions T and F have graphs as straight lines, yet in Example 2 we found F was not linear. The difference between T and F is in the constant term. If $b = 0$, we have linearity; if not, we do not have linearity. In examples 3 through 6, $T(\theta) = \theta$. This gives us a clue to the first property of linear transformations.

Theorem 4.1.1. *Let V and W be vector spaces. If $T\colon\ V \to W$ is a linear transformation, then $T(\theta_V) = \theta_W$. (The subscripts emphasize the vector space that the zero vector comes from.)*

Proof. Since $\theta_V + \theta_V = \theta_V$,

$$\underbrace{T(\theta_V)}_{\text{In } W} = \underset{\text{By additivity}\nearrow}{T(\theta_V + \theta_V)} = \underbrace{T(\theta_V) + T(\theta_V)}_{\text{In } W}$$

and so

$$T(\theta_V) = T(\theta_V) + T(\theta_V)$$

By uniqueness of θ_W in W, the only way the last equation can hold is if $T(\theta_V) = \theta_W$. □

This theorem can sometimes be used to show transformations are nonlinear. A logical consequence of the theorem is

If $T(\theta_V) \neq \theta_W$, then T is not linear.

Example 7. Show that $T\colon\ E^2 \to E^2$, defined by

$$T((x_1, x_2)) = (x_1 + x_2, x_1 - x_2 + 1)$$

is not linear

Solution In $E^2, T(\theta) = T((0,0)) = (0,1) \neq \theta$. Therefore, T is not linear.

Example 8. Let $T\colon\ E^2 \to E^1$ be defined by

$$T((x_1, x_2)) = {x_1}^2 + {x_2}^2$$

Show that T is not linear even though $T(\theta) = \theta$.

Solution We have $T(\theta) = T((0,0)) = 0^2 + 0^2 = 0$, which is the zero of E^1. This allows no conclusion; the definition of linearity must be used. To check additivity we calculate

$$\begin{aligned} T(\mathbf{x}+\mathbf{y}) &= T((x_1,x_2)+(y_1,y_2)) \\ &= T((x_1+y_1, x_2+y_2)) = (x_1+y_2)^2 + (x_2+y_2)^2 \\ &= {x_1}^2 + 2x_1y_1 + {y_1}^2 + {x_2}^2 + 2x_2y_2 + {y_2}^2 \end{aligned}$$

and

$$T(\mathbf{x}) + T(\mathbf{y}) = T((x_1,x_2)) + T((y_1,y_2)) = {x_1}^2 + {x_2}^2 + {y_1}^2 + {y_2}^2$$

Since $T(\mathbf{x}+\mathbf{y}) \neq T(\mathbf{x})+T(\mathbf{y})$, we know that T is not linear. In most cases, to determine linearity or nonlinearity of a transformation, we use the definition.

Example 9. Show that the following transformation are linear.

(**a**) $T\colon\ E^3 \to E^3$ defined by

$$T((x_1,x_2,x_3)) = (x_1+x_2, x_2+x_3, x_3+x_1)$$

(**b**) $T\colon\ E^3 \to E^3$ defined by

$$T((x_1,x_2,x_3)) = \mathbf{v} \times (x_1,x_2,x_3)$$

where $\mathbf{v}$ is a fixed vector in E^3

(**c**) $T\colon\ E^3 \to E^1$ defined by

$$T((x_1,x_2,x_3)) = ax_1 + bx_2 + cx_3$$

where a, b, and c are fixed real numbers

(**d**) $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{22}$ defined by

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

(**e**) $T\colon\ \mathcal{P}_1 \to \mathcal{P}_2$ defined by

$$T(ax+b) = \frac{ax^2}{2} + bx$$

(**f**) $T\colon\ \mathbb{C}^2 \to \mathbb{C}^2$ defined by

$$T((z_1, z_2)) = (z_1 + z_2, z_1 - 2z_2)$$

Solution Parts (**a**) through (**e**) are left to the problems.

(**f**)
$$\begin{aligned}
T((z_1, z_2) + (u_1, u_2)) &= T(z_1 + u_1, z_2 + u_2)\\
&= (z_1 + u_1 + z_2 + u_2, z_1 + u_1 - 2z_2 - 2u_2)\\
&= (z_1 + z_2, z_1 - 2z_2) + (u_1 + u_2, u_1 - 2u_2)\\
&= T(z_1, z_2) + T(u_1, u_2)\\
T(e(z_1, z_2)) = T(cz_1, cz_2) &= (cz_1 + cz_2, cz_1 - 2cz_2)\\
&= c(z_1 + z_2, z_1 - 2z_2)\\
&= cT(z_1, z_2)
\end{aligned}$$

Thus T is linear.

Example 10. Show that $T\colon\ \mathcal{C}_{22} \to \mathcal{C}_{22}$ defined by $T(A) = \bar{A}$ is not linear.

Solution We know that $T(cA) = \overline{cA} = \bar{c}\bar{A} = \bar{c}T(A) \neq cT(A)$ unless $c \in \mathbb{R}$, but c can have a nonzero imaginary part. So T is not linear. [However, T is called **conjugate linear** because $T(cA) = \bar{c}T(A)$ and $T(A + B) = T(A) + T(B)$.]

Example 11. Let $V = \mathcal{M}_{n1}$ and $W = \mathcal{M}_{m1}$. Let M be an $m \times n$ real matrix. Define $T\colon\ V \to W$ by

$$T(X) = MX$$

T is linear because by matrix algebra

$$\begin{aligned}
T(X + Y) &= M(X + Y) = MX + MY\\
T(cX) &= M(cX) = c(MX)
\end{aligned}$$

Example 12. Let $V = \mathcal{C}_{n1}$ and $W = \mathcal{C}_{m1}$ and let Z be an $m \times n$ matrix from $\mathcal{C}_{mn}$. Define $T\colon\ V \to W$ by $T(X) = ZX$. Then T is linear because by matrix algebra

$$\begin{aligned}
Z(X + Y) &= ZX + ZY\\
Z(cX) &= c(ZX)
\end{aligned}$$

Some special linear transformations must be noted for future use. The **zero** transformation T_θ from V to W is defined as

$$T(\mathbf{x}) = \theta_W \qquad \text{for all } \mathbf{x} \text{ in } V$$

The **identity** transformation I from V to V is defined as

$$I(\mathbf{x}) = \mathbf{x} \qquad \text{for all } \mathbf{x} \text{ in } V$$

The **contraction** transformation T_α from V to V is

$$T_\alpha(\mathbf{x}) = \alpha\mathbf{x} \qquad 0 < \alpha < 1, \text{ for all } \mathbf{x} \in V$$

The **dilation** transformation T_β from V to V is

$$T_\beta\mathbf{x} = \beta\mathbf{x} \qquad 1 < \beta, \text{ for all } \mathbf{x} \in V$$

Verification that these are linear transformations is left to the problems.

Although several examples of linear transformations have now been given, we have not yet begun to analyze linear transformations. In algebra, analysis of functions was done with graphs of the functions. **In our present situation we must usually be satisfied without the types of graphs we drew in algebra**. Usually we draw "graphs," as indicated in Fig. 4.1.1, whenever possible. Ordinarily this can be done only when V and W are versions of E^n, $n = 1, 2$, or 3. Other cases require considerable imagination. Consider Example 13.

Example 13. "Graph" the transformation $T\colon\ E^2 \to E^3$, defined by $T(x_1, x_2) = \frac{1}{2}(x_1, x_2, 0)$.

Solution The visualizations of E^2 and E^3 as well as some special vectors are shown in Fig. 4.1.2. The image of these vectors after T acts on them is also shown in that figure. If we put more vectors of length 1 in the circle in Fig. 4.1.2**a**, the terminal points of the images lie on the circle of radius $\frac{1}{2}$, as in Fig. 4.2.2**b**. This supports our intuitive feeling that T "shrinks" all vectors in the domain, much like a contraction transformation.

Since graphs are not simple for linear transformations, we must be able to analyze them without graphs as well. Unless specified otherwise, all vector spaces from now on are assumed to be finite-dimensional. One of the basic tools for the analysis of linear transformations is the following:

Kernel problem

Given $T\colon\ V \to W$, find all $\mathbf{x}$ in V such that $T(x) = \theta$. The set of all such $\mathbf{x}$ is called the **kernel of** T and written $\ker T$.

Roughly speaking, the kernel problem is very much like the problem from algebra of solving the equation $f(x) = 0$, for example, solving $x^2 - 2x - 3 = 0$. In algebra this problem is solved by factoring or using the quadratic formula. In linear algebra the solution to the kernel problem many times reduces to solving m equations in n unknowns (the "first basic problem of linear algebra").

Example 14. Find $\ker T$, where $T\colon\ E^3 \to E^2$ is defined by $T((x_1, x_2, x_3)) = (x_1 + x_2, x_2 - x_3)$.

Solution Since $\ker T = \{\mathbf{x} | T(\mathbf{x}) = \theta\}$, we must solve $T((x_1, x_2, x_3)) = (0, 0)$, that is,

$$(x_1 + x_2, x_2 - x_3) = (0, 0)$$

The resulting equations are

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_2 - x_3 &= 0 \end{aligned}$$

which have solution $(-k, k, k)$. Therefore

$$\ker T = \{\mathbf{v} \in E^3 | \mathbf{v} = k(-1, 1, 1)\} = \operatorname{span}\{(-1, 1, 1)\}$$

In example 14 the kernel of the given linear transformation was a subspace of the domain. In fact, a basis for $\ker T$ was $\{(-1, 1, 1)\}$. The kernel of a linear transformation is always a vector space.

Theorem 4.1.2. *Let V and W be vector spaces, and let $T\colon\ V \to W$ be a linear transformation. The set* $\ker T$ *is a subspace of V.*

Proof. The kernel of T is nonempty because $T(\theta) = \theta$. We need to show that $\ker T$ is closed under addition and scalar multiplication. Recall that $\mathbf{x} \in \ker T$ if and only if $T\mathbf{x} = \theta$. Let $\mathbf{x}$ and $\mathbf{y}$ be in $\ker T$, and let c be a number. By the linearity of T,

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) = \theta + \theta = \theta$$

and

$$T(c\mathbf{x}) = cT(\mathbf{x}) = c\theta = \theta$$

so $\mathbf{x} + \mathbf{y} \in \ker T$ and $c\mathbf{x} \ker T$. Thus $\ker T$ is a subspace of V. $\square$

Since $\ker T$ is a subspace of V, it has dimension. The dimension of $\ker T$ is called the **nullity** of T. Thus for the linear transformation in Example 14 the nullity is 1. We write this

$$\eta(T) = 1$$

Example 15. Calculate $\eta(T)$ for the linear transformation $T\colon\ E^3 \to E^2$ defined by

$$T((a, b, c)) = (a + 2b + c, -a + 3b + c)$$

Find a basis for $\ker T$.

Solution We must find the set of all vectors (a, b, c) in E^3 that $T(a, b, c) = (0, 0)$. That is, the equation

$$\begin{pmatrix} a + 2b + c \\ -a + 3b + c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

must be solved. The solution is

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -k \\ -2k \\ 5k \end{pmatrix}$$

and $\ker T = \text{span}\{(-1, -2, 5)\}$. Therefore $\dim(\ker T) = 1$, so $\eta(T) = 1$. A basis is $\{(-1, -2, 5)\}$.

To continue the analysis of linear transformations, we consider the **range of** T. In algebra finding the range of a function f is important in graphing $y = f(x)$. For example, $y = x^2 - 2x - 3$ has range $\{y| - 4 \leq y \leq \infty\}$. The solutions of $x^2 - 2x - 3 = 0$ are $x = 3$ and $x = -1$. (That is, the kernel of f is $\{-1, 3\}$.) All this information is shown in Fig. 4.1.3. The range of a linear transformation cannot always be used to obtain a graph of T, but it is quite useful in other ways.

Definition 4.1.2. Let $T\colon\ V \to W$ be a linear transformation. The **range of** T is the set of all possible $\mathbf{v}$ in W such that $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x}$ in V. The range of T is written range T. The range of T is a subspace of W (see the problems).

Example 16. Define T from E^3 to E^3 by $T((a,b,c)) = (a-b+c, 2a+b-c, -a-2b+2c)$. Determine range T and dim(range T). Find two vectors in range T and two vectors not in range T. Find a basis for range T. Find $\ker T$. Graph $\ker T$ and range T. Attempt a graph of T.

Solution Let $\mathbf{y} = (y_1, y_2, y_3)$ be in range T. Thus $\mathbf{y} = T((a,b,c))$ for some vector (a,b,c) in E^3. That is, the equation $\mathbf{y} = T((a,b,c))$ **must be consistent**. We reduce the equations and see what conditions the consistency forces. The equations are

$$\begin{aligned} a - \ \ b + \ \ c &= y_1 \\ 2a + \ \ b - \ \ c &= y_2 \\ -a - 2b + 2c &= y_3 \end{aligned}$$

and they reduce to

$$\begin{aligned} a - b + \ \ c &= \ \ y_1 \\ 3b - 3c &= \ \ y_2 - 2y_1 \\ 0 &= -y_1 + y_2 + y_3 \end{aligned}$$

So if $\mathbf{y} = (y_1, y_2, y_3)$ is to be the range T, then $-y_1 + y_2 + y_3 = 0$. That is,

$$\text{range } T = \{(y_1, y_2, y_3) | y_1 = y_2 + y_3\}$$

The condition on y_1, y_2, and y_3 gives a criterion for inclusion in range T. Some vectors in range T are $(-2,-1,-1)$ and $(0,-1,1)$. Some vectors not in range T are (1,1,1) and (1,0,0). The dimension of range T is 2, since the equation $-y_1 + y_2 + y_3 = 0$ allows the assignment of arbitrary values to any **two** of the values of y_k.

To obtain a basis, we can use $(-2,-1,-1)$ and $(0,-1,1)$ as above, since they are linearly independent in the range and dim(range T) = 2. In fact, any two linearly independent vectors in range T form a basis for range T. The kernel of T is found by setting $y_1 = y_2 = y_3 = 0$ in the linear equations above. We obtain $\ker T = \text{span}\{(0,1,1)\}$. Graphs are shown in Fig. 4.1.4.

The dimension of the range of a linear transformation T is called the **rank** of T and written $\mathcal{R}(T)$. That is,

$$\mathcal{R}(T) = \dim(\text{range } T)$$

The rank and nullity of a linear transformation are related to each other by the equation

$$\text{rank } T + \text{nullity } T = \dim(\text{domain})$$

This is the result of the following basic theorem, one of the most important in linear algebra:

Theorem 4.1.3. *If* $T\colon\ V \to W$ *is a linear transformation and* $\dim V = n$*, then*

$$\mathcal{R}(T) + \eta(T) = n \tag{4.1.1}$$

Before proving this theorem, we consider an example of its use.

Example 17. Find the nullity of the linear transformation in Example 16.

Solution We had $T\colon\ E^3 \to E^3$ and found $\mathcal{R}(T) = 2$. Since $\dim V = \dim E^3 = 3$, Eq. (4.1.1) leads to

$$2 + \eta(T) = 3$$

Therefore $\eta(T) = 1$

Proof of Theorem 4.1.3. Since $\ker T$ and range T are vector spaces, $\mathcal{R}(T)$ and $\eta(T)$ are defined. We consider three cases: $\eta(T) = 0$, $\eta(T) = n$, and $1 \le \eta(T) \le n - 1$.

Case 1: $\eta(T) = 0$. Suppose $\mathcal{R}(T) = k < n$. That is, suppose that $\eta(T) + \mathcal{R}(T) < n$. We will obtain a contradiction. Since $\mathcal{R}(T) = k$, any set of more than k vectors in range T is linearly dependent. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V. Since $k < n$, $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ must be linearly dependent and so there exist $c_1, \ldots, c_n$, not all zero, with

$$c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n) = \theta$$

Thus by linearity $T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = \theta$ and $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in \ker T$. Since $\ker T = \{\theta\}$ and not all the c_i's are zero, we have $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ being linearly dependent, which is a contradiction. Therefore, $\mathcal{R}(T) = n$ and $\eta(T) + \mathcal{R}(T) = 0 + n = n$.

Case 2: $\eta(T) = n$. Since $\ker T$ is a subspace of V and $\dim(\ker T) = \dim V$, we actually have $\ker T = V$and $T(\mathbf{x}) = \theta$ for all $\mathbf{x} \in V$. Therefore range $T = \{\theta\}$ and $\dim(\text{range } T) = 0$. Thus $\mathcal{R}(T) + \eta(T) = 0 + n = n$.

Case 3: $\eta(T) = k, 1 \leq k \leq n-1$. Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for $\ker T$. By a previous result, B can be extended to a basis $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ of V, since $\dim V = n$. We will show that $\mathcal{T} = \{T(\mathbf{u}_{k+1}), \ldots, T(\mathbf{u}_n)\}$ is a basis for range T. Then we will have $\mathcal{R}(T) = n-k$ and

$$\mathcal{R}(T) + \eta(T) = (n-k) + k = n$$

$\mathcal{T}$ is Linearly Independent Consider

$$c_{k+1}T(\mathbf{u}_{k+1}) + \cdots + c_nT(\mathbf{u}_n) = \theta$$

By the linearity of T,

$$T(c_{k+1}, \mathbf{u}_{k+1} + \cdots + c_n\mathbf{u}_n) = \theta$$

and so $c_{k+1}\mathbf{u}_{k+1} + \cdots + c_n\mathbf{u}_n$ is in $\ker T$. So there exist $c_1, \ldots, c_k$ with

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{u}_n$$

That is,

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k - c_{k+1}\mathbf{u}_{k+1} - \cdots - c_n\mathbf{u}_n = \theta$$

and since S is a basis for V, we know that $c_1 = \cdots = c_k = c_{k+1} = \cdots = c_n = 0$. Thus $\mathcal{T}$ is linearly independent. $\square$

$\mathcal{T}$ Spans range T Let $\mathbf{y} \in$ range T, so that $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x} \in \mathbf{V}$. Since S is a basis for V, we can write $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{u}_{k+1} + \cdots + c_n\mathbf{u}_n$, and so

$$\begin{aligned} \mathbf{y} = T(\mathbf{x}) &= T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) + c_{k+1}T(\mathbf{u}_{k+1}) + \cdots + c_nT(\mathbf{u}_n) \\ &= \qquad\qquad \theta \qquad\qquad + c_{k+1}T(\mathbf{u}_{k+1}) + \cdots + c_nT(\mathbf{u}_n) \end{aligned}$$

Thus $\mathbf{y} \in$ span $\mathcal{T}$.

Example 18. "Graph" $T\colon\ E^3 \to E^3$, defined by $T((x_1, x_2, x_3)) = (-x_1 + x_2 + x_3, 2x_1 - x_2, x_1 + x_2 + 3x_3)$, indicating $\ker T$ and range T.

Solution Solving $T(\mathbf{x}) = \mathbf{y}$, we find

$$\left(\begin{array}{rrr|c} -1 & 1 & 1 & y_1 \\ 0 & 1 & 2 & y_2 + 2y_1 \\ 0 & 0 & 0 & y_3 - 2y_2 - 3y_1 \end{array}\right) \tag{4.1.2}$$

To determine $\ker T$, set $\mathbf{y} = \theta$. The solution of the resulting equations is $\mathbf{x} = (-k, -2k, k)$, so $\ker T = \text{span}\,\{(-1, -2, 1)\}$. From Eq. (4.1.2) we see that range $T = \{\mathbf{y}|y_3 - 2y_2 - 3y_1 = 0\} = \{\mathbf{y}|\mathbf{y} = (s, t, 3s + 2t)\} = \text{span}\,\{(1, 0, 3), (0, 1, 2)\}$. A graph is shown in Fig. 4.1.5.

Two more ways to view the action of a linear transformation are to determine the images under T of geometric figures such as squares and circles.

Example 19. It can be shown (see the problems) that if $T\colon\ \mathcal{M}_{21} \to \mathcal{M}_{21}$ is a linear transformation defined by

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

where A is a nonsingular matrix, then the image of a straight-line segment from P to Q in E^2 will be a straight-line segment from $T(P)$ to $T(Q)$. Let T be defined as

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Find the image under T of the "unit square" shown in Fig. 4.1.6.

Solution We find the images of the vertices. Since each side of the square is a straight-line segment, the image of the square will be the figure generated by joining the images of the vertice with a straight-line segment. Now

$$T\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$
$$T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad T\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Therefore the image is the parallelogram shown in Fig. 4.1.7. Note that points are associated with terminal points of the vectors naturally associated with the elements of $\mathcal{M}_{21}$.

Example 20. Show that $T\colon\ \mathcal{M}_{21} \to \mathcal{M}_{21}$ defined by

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

transforms the unit circle ${x_1}^2 = {x_2}^2 = 1$ to an ellipse.

Solution Since the image of

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

under T is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 3x_2 \end{pmatrix}$$

we have

$$\frac{{y_1}^2}{4} + \frac{{y_2}^2}{9} = \frac{4{x_1}^2}{4} + \frac{9{x_2}^2}{9} = {x_1}^2 + {x_2}^2 = 1$$

Therefore the image of the circle is an ellipse. This action of T is shown in Fig. 4.1.8.

An interpretation of Fig. 4.1.8 is that T dilates with constant 3 in the x_2 direction and constant 2 in the x_1 direction.

Example 21. Show that $T\colon\ \mathcal{M}_{31} \to \mathcal{M}_{31}$ defined by

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = T\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad a, b, c > 0$$

transforms the sphere ${x_1}^2 + {x_2}^2 + {x_3}^2 = R^2$ to an ellipsoid.

Solution The image of

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \text{is} \qquad \begin{pmatrix} ax_1 \\ bx_2 \\ cx_3 \end{pmatrix}$$

and

$$\frac{(ax_1)^2}{a^2} + \frac{(bx_2)^2}{b^2} + \frac{(cx_3)^2}{c^2} = {x_1}^2 + {x_2}^2 + {x_3}^2 = R^2$$

Division by R^2 yields

$$\frac{{y_1}^2}{(aR)^2} + \frac{{y_2}^2}{(bR)^2} + \frac{{y_3}^2}{(cR)^2} = 1$$

which is an equation for the ellipsoid shown in Fig. 4.1.9.

Important examples of linear transformations exist which cannot be analyzed geometrically except in some generalized way. One example is $T\colon\ \mathcal{P}_n \to \mathcal{P}_{n+1}$ defined by

$$T(a_0 + a_1 x + \cdots + a_n x^n) = a_0 x + a_1 x^2 + \cdots + a_n x^{n+1}$$

That is, for f in $\mathcal{P}_n$, $T(f)$ is the function obtained at each x by multiplying $f(x)$ by x. That is,

$$(T(f))(x) = xf(x)$$

This linear transformation is a special case of the **coordinate operator** in quantum mechanics.

If we allow complex vector spaces and consider the set of nth-degree polynomials with complex coefficients with the same operations as $\mathcal{P}_n$, we will have a vector space ${\mathcal{P}_n}^{\mathbb{C}}$. We can define a transformation

$$T(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = -i(a_1 + 2a_2 x + \cdots + n a_n x^{n-1})$$

that is,

$$(T(f))(x) = -i\frac{d}{dx}(f(x))$$

Now T is linear, and the rule for T is the same as the rule for the **momentum operator** in quantum mechanics. However, the momentum operator has a different domain.

PROBLEMS 4.1

1. Determine whether the following transformations $T\colon\ E^3 \to E^3$ are linear.

   (**a**) $T((x_1, x_2, x_3)) = (x_1, x_1 - x_2, x_2 + x_3)$
   (**b**) $T((x_1, x_2, x_3)) = (x_1, x_2, x_2 x_3)$
   (**c**) $T((x_1, x_2, x_3)) = (x_1, 0, 0)$
   (**d**) $T((x_1, x_2, x_3)) = (1, 0, 0)$
   (**e**) $T((x_1, x_2, x_3)) = (3x_1 + 2x_2, x_3, |x_2|)$
   (**f**) $T((x_1, x_2, x_3)) = (x_1 - x_2, x_1 + x_2, x_3)$

2. Determine whether the following transformations defined on $\mathcal{M}_{22}$ are linear.

(**a**) $T(A) = A^T$

(**b**) $T(A) = A + A^T$

(**c**) $T(A) = A^T A$

(**d**) $T(A) = A + I$ (where I is the identity matrix)

(**e**) $T(A) = kA$ (where k is a real number), $k \neq 0$

(**f**) $T(A) = A^2$

(**g**) $T(A) = \det A$

(**h**) $T(A) = \operatorname{tr} A$ (trace of $A =$ sum of the diagonal elements)

3. Determine whether the following transformations $T\colon\ \mathcal{P}_2 \to \mathcal{P}_2$ are linear.

(**a**) $T(a + bx + cx^2) = ax + bx^2$

(**b**) $T(a + bx + cx^2) = x - c$

(**c**) $T(a + bx + cx^2) = 2$

(**d**) $T(a + bx + cx^2) = \dfrac{1}{a + bx + cx^2}$

(**e**) $T(a + bx + cx^2) = (a - b) + (b + c)x + (a - c)x^2$

4. Determine whether the following transformation $\mathcal{C}_{22} \to \mathcal{C}_{22}$ are linear.

(**a**) $T(a) = A^*$

(**b**) $T(A) = A^* A$

(**c**) $T(A) = A + \bar{A}$

(**d**) $T(A) = iA$

5. For each **linear** transformation from Prob. 1, determine $\ker T$ and range T, find bases for $\ker T$ and range T, and verify the equation $\eta(T) + \mathcal{R}(T) = \dim(\text{domain } T)$. Sketch a "graph" of T as in Example 13 or 19.

6. For each **linear** transformation from Prob. 2, determine $\ker T$ and range T, find bases for $\ker T$ and range T, and verify the equation $\eta(T) + \mathcal{R}(T) = \dim(\text{domain } T)$.

7. For each **linear** transformation from Prob. 3, determine $\ker T$ and range T, find bases for $\ker T$ and range T, and verify the equation $\eta(T) + \mathcal{R}(T) = \dim(\text{domain } T)$.

8. Show that the zero transformation is linear.

9. Show that the identity transformation is linear.

10. Show that a contraction operator is linear.

11. Show that a dilation transformation is linear.

12. Show that a linear transformation $T\colon\ E^2 \to E^2$ defined by

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

transforms the circle ${x_1}^2 + {x_2}^2 = 1$ to the ellipse ${y_1}^2/a^2 + {y_2}^2/b^2 = 1$.

13. Draw the image under T of the unit square in E^2 for T defined by

$$(a)\quad T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (b)\quad T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ 2x_1 + 3x_2 \end{pmatrix}$$

14. Show that the additivity condition $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ implies that $T(n\mathbf{x}) = nT(\mathbf{x})$ for any positive integer n.

15. Show that the additivity condition $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x})+T(\mathbf{y})$ implies that $T((p/q)\mathbf{x}) = (p/q)T(\mathbf{x})$, where p and q are positive integers. [**Hint**: By Prob. 14, $pT(\mathbf{x}) = T(p\mathbf{x}) = T(q(p/q)\mathbf{x})$.]

16. Show that this definition is equivalent to Definition 4.1.1:

    Let V and W be vectors spaces, and let T be a function with domain V and range in W. Then T is a linear transformation if for all $a, b \in \mathbb{R}$, $\mathbf{x}, \mathbf{y} \in V$, $T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y})$.

17. Let $T\colon\ V \to W$ be a linear transformation. Show that range T is a subspace of W.

18. Complete the details of Example 9**a** through **e**.

19. Verify that the coordinate operator as defined in this section is linear.

20. Verify that the momentum operator as defined in this section is linear.

21. Let $T\colon\ E^3 \to E^3$ be linear, and suppose that $T((1,0,1)) = (1,-1,3)$ and $T((2,1,0)) = (0,2,1)$. Determine $T((8,3,2))$. [**Hint**: write (8,3,2) as a linear combination of (1,0,1) and (2,1,0), and use the linearity of T.]

22. Regarding T as in Prob. 21, calculate $T((1,2,-3))$ and $T((4,-4,12))$.

23. Regarding T as in Prob. 21, why can $T((3,0,4))$ not be calculated from the information given?

24. Define $T\colon\ \mathcal{P}_2 \to \mathcal{M}_{22}$ for each f in $\mathcal{P}_2$ by

$$T(f) = \begin{pmatrix} f(1) & f(0) \\ f(0) & f(2) \end{pmatrix}$$

Is T a linear transformation?

25. Define $T\colon\ \mathcal{M}_{21} \to \mathcal{M}_{21}$ by

$$T\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = M\begin{pmatrix} a \\ b \end{pmatrix}$$

where M is an invertible 2×2 real matrix. Show that the image of a line under T is again a line (**Hint**: Describe a line in the domain by its **vector equation**; then use the linearity of T.)

4.2 MATRIX REPRESENTATION PROBLEM FOR LINEAR TRANSFORMATIONS

The methods of Sec. 4.1 can be used to discuss systems of linear equations. For a set of equations

$$A_{m\times n}X_{n\times 1} = B_{m\times 1}$$

the matrix on the left-hand side represents a linear transformation from $\mathcal{M}_{n1}$ to $\mathcal{M}_{m1}$ because, by the laws of matrix algebra, $A(cX + dY) = cAX + dAY$. The kernel of the linear transformation is the solution set for the homogeneous equation $AX = \theta$. The range of the linear transformation is the set of all vectors B for which $AX = B$ has a solution. The "rank-kernel equation" from theorem 4.1.3 of Sec. 4.1 means that

$$\dim(\text{solution space of } AX = \theta) + \dim(\text{range}) = n$$

However from previous chapters we know that the dimension of the solution space is the number of zero rows of the reduced row echelon form of A. So the last equation can be rewritten

$$(\dim \text{ of solution space of } AX = \theta) + (\text{rank } A) = n = \text{no. of columns of } A$$

Example 1. Find the dimension of the solution space of $AX = \theta$, where

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

Solution Since $n = 4$, if we find A, then $4-$ rank of A is the number we desire. Since rank $A =$ row rank A, we row-reduce A:

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

to find rank $A = 2$. Therefore the dimension of the solution space of $AX = \theta$ is 2. Notice that this is also the number of unknowns which can be arbitrarily set. Thus the terminology 2 **degrees of freedom**.

Example 1 illustrates this general principle:

> If T is a linear transformation generated by a matrix A, then $\eta(T)$ **and** $\mathcal{R}(T)$ **can be found by row-reducing matrix** A. **That is information about a linear transformation can be gained by analyzing a matrix.**

For this reason (and others which appear later), representation of a linear transformation by a matrix is important. Thus we come to the third basic problem of linear algebra.

Third Basic Problem of Linear Algebra

Given a linear transformation $T\colon\ V \to W$, where $\dim V = n$ and $\dim W = m$, find an $m \times n$ matrix A which "represents" T.

Before stating precisely what the word **represents** means, we consider some simple examples.

Example 2. Consider the identity transformation $T\colon\ E^3 \to E^3$, defined by $T(\mathbf{x}) = \mathbf{x}$. Let $X = (\mathbf{x})_{\mathcal{E}}$, where $\mathcal{E}$ is the standard ordered basis $\{(1,0,0),(0,1,0),$
$(0,0,1)\}$. Then

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

when $\mathbf{x} = (x_1, x_2, x_3)$. So

$$I_{3\times 3}X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = X$$

The action of the **identity transformation** is represented by matrix multiplication of coordinate matrices by the **identity matrix** I.

Example 3. Consider the projection $P\colon\ E^3 \to E^3$ defined by $P(x_1, x_2, x_3) = (x_1, x_2, 0)$. For $\mathcal{E}$ as in Example 2,

$$MX = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$$

we see that the action of P is represented by matrix multiplication by

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Note that $P(P(X)) = P(X)$; also $MM = M$. We have the projection represented by an idempotent matrix.

Example 4. Consider differentiation $D\colon\ \mathcal{P}_1 \to \mathcal{P}_1$ defined by $D(a+bx) = b$. If we use the standard ordered basis $\mathcal{E} = \{1, x\}$, then

$$(a+bx)_{\mathcal{E}} = \begin{pmatrix} a \\ b \end{pmatrix}$$

and we can write

$$D\colon\ \begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} b \\ 0 \end{pmatrix}$$

Now the matrix

$$M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

satisfies

$$M \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

that is, the action of D is represented by multiplication by M. We note that

$$D(D(a+bx)) = D(b+0x) = 0 + 0x = \theta$$

and

$$MM = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The transformation D is represented by a matrix which is nilpotent of exponent 2. This just means that the second derivative of a first-degree polynomial is zero.

Examples 2 to 4 depend on the fact that we used the standard basis to represent the vectors in each vector space. We will see that in general the **representing matrix depends on the bases used for the domain and range**.

Solution of Representation Problem The basic principle which leads to the solution of the basis problem for a linear transformation $T\colon\ V \to W$ is as follows:

> If $\dim V = n$ and $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, then the range of T is completely describable in terms of the images $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ of the basis vectors.

To see this, let $\mathbf{x}$ be any vector in V. There exist constants $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Therefore, $T(\mathbf{x}) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2 + \cdots + c_nT(\mathbf{v}_n)$ and we see that every element $T(\mathbf{x})$ in the range is a linear combination of the images of basis elements. That is $T(\mathbf{x}) \in \operatorname{span}\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$.

The setup and procedure for solving the representation problem are as follows:

> Suppose $\dim V = n, \mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an ordered basis for V, and suppose $\dim W = m$ and $\mathcal{T} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ is an ordered basis for W.
>
> 1. Calculate $T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)$.
> 2. Find the coordinate vectors $(T(\mathbf{v}_1))_{\mathcal{T}}, (T(\mathbf{v}_2))_{\mathcal{T}}, \ldots, (T(\mathbf{v}_m))_{\mathcal{T}}$.
> 3. Write the matrix with columns as the column vectors calculated in Step 2:
>
> $$M = ((T(\mathbf{v}_1))_{\mathcal{T}}(T(\mathbf{v}_2))_{\mathcal{T}} \cdots (T(\mathbf{v}_n))_{\mathcal{T}})$$

The $m \times n$ matrix M represents T, as indicated in Fig. 4.2.1. Whenever necessary, we write M_T to denote the matrix of T. The diagram gives the content of the theorem that we will state for the solution of the representation problem. Before spelling out the theorem, we consider several examples. In these examples, we write $(V, \mathcal{S})$ to indicate the vector space V with basis $\mathcal{S}$

Example 5. Let $T\colon\ (E^2, \mathcal{S}) \to (E^2, \mathcal{T})$ be defined by $T((x_1, x_2)) = (x_1 + 2x_2, x_1 - x_2)$. Find the matrix M representing T when

(**a**) $\mathcal{S} = \mathcal{T} = \{\mathbf{e}_1, \mathbf{e}_2\}$, the standard basis

(**b**) $\mathcal{S} = \mathcal{T} = \{(1, 2), (3, -1)\}$

(**c**) $\mathcal{S} = \mathcal{T} = \{(3, -1), (1, 2)\}$

(**d**) $\mathcal{S} = \{(1, -1), (1, 1)\}, \mathcal{T} = \{(1, 0), (0, -1)\}$

(**e**) $\mathcal{S} = \{(1, -1), (1, 1)\}, \mathcal{T} = \{(0, -1), (1, 0)\}$

In each case, calculate $T((3, 2))$ directly and by using M.

Solution (**a**) $T(\mathbf{e}_1) = (1,1) = 1(1,0) + 1(0,1)$ so $(T(\mathbf{e}_1))_{\mathcal{T}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$T(\mathbf{e}_2) = (2,-1) = 2(1,0) + (-1)(0,1)$ so $(T(\mathbf{e}_2))_{\mathcal{T}} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$

The matrix is
$$M = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$
Now $T((3,2)) = (7,1)$ from the definition of T. But
$$((3,2))_{\mathcal{S}} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$
so
$$(T(3,2))_{\mathcal{T}} = M \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$$
and finally,
$$\begin{pmatrix} 7 \\ 1 \end{pmatrix} \longrightarrow 7(1,0) + 1(0,1) = (7,1)$$
In this case, since $\mathcal{S} = \mathcal{T}$, we say that M is the **matrix of** T **with respect to** $\mathcal{S}$. Also, since $\mathcal{S} = \mathcal{T} =$ the standard basis, M is called the **standard matrix of** T.

(**b**) We work as in (**a**):

$T((1,2)) = (5,-1) = \frac{2}{7}(1,2) + \frac{11}{7}(3,-1)$ so $(T(1,2))_{\mathcal{T}} = \begin{pmatrix} \frac{2}{7} \\ \frac{11}{7} \end{pmatrix}$

$T((3,-1)) = (1,4) = \frac{13}{7}(1,2) + (-\frac{2}{7})(3,-1)$ so $(T(3,-1))_{\mathcal{T}} = \begin{pmatrix} \frac{13}{7} \\ -\frac{2}{7} \end{pmatrix}$

So we have the matrix of T with respect to $\mathcal{S}$:
$$M = \begin{pmatrix} \frac{2}{7} & \frac{13}{7} \\ \frac{11}{7} & -\frac{2}{7} \end{pmatrix}$$
From the definition of $T, T(3,2) = (7,1)$; since
$$((3,2))_{\mathcal{S}} = \begin{pmatrix} \frac{9}{7} \\ \frac{9}{7} \\ \frac{4}{7} \end{pmatrix}$$

using the matrix M we have

$$(T(3,2))_{\mathcal{T}} = M \begin{pmatrix} \frac{9}{7} \\ \frac{4}{7} \end{pmatrix} = \begin{pmatrix} \frac{2}{7} & \frac{13}{7} \\ \frac{11}{7} & -\frac{2}{7} \end{pmatrix} \begin{pmatrix} \frac{9}{7} \\ \frac{4}{7} \end{pmatrix} = \begin{pmatrix} \frac{10}{7} \\ \frac{13}{7} \end{pmatrix}$$

Finally

$$\begin{pmatrix} \frac{10}{7} \\ \frac{13}{7} \end{pmatrix} \longrightarrow \tfrac{10}{7}(1,2) + \tfrac{13}{7}(3,-1) = (7,1)$$

(**c**) The case differs from (**b**) in that the order of $\mathcal{S}$ has been reversed. The calculations, however, are similar:

$$T((3,-1)) = (1,4) = -\tfrac{2}{7}(3,-1) + \tfrac{13}{7}(1,2) \quad \text{so} \quad (T(3,-1))_{\mathcal{T}} = \begin{pmatrix} -\frac{2}{7} \\ \frac{13}{7} \end{pmatrix}$$

$$T((1,2)) = (5,-1) = \tfrac{11}{7}(3,-1) + \tfrac{2}{7}(1,2) \quad \text{so} \quad (T(1,2))_{\mathcal{T}} = \begin{pmatrix} \frac{11}{7} \\ \frac{2}{7} \end{pmatrix}$$

Therefore the matrix of T with respect to $\mathcal{S}$ is

$$M = \begin{pmatrix} -\frac{2}{7} & \frac{11}{7} \\ \frac{13}{7} & \frac{2}{7} \end{pmatrix}$$

Note that this matrix differs from the representing matrix in (**b**) in that an interchange of rows and columns has occurred. Calculations of $T(3,2)$ is left to the reader.

(**d**) In this case $\mathcal{S} \neq \mathcal{T}$, but our usual procedure can be used.

$$T((1,-1)) = (-1,2) = -1(1,0) + (-2)(0,-1) \quad \text{so} \quad (T(1,-1))_{\mathcal{T}} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

$$T((1,1)) = (3,0) = 3(1,0) + 0(0,-1) \quad \text{so} \quad (T(1,1))_{\mathcal{T}} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

The **matrix of T with respect to $\mathcal{S}$** and $\mathcal{T}$ is

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

Now to calculate $T(3,2)$ in two ways, $T(3,2) = (7,1)$ by definition, but we also have

$$((3,2))_{\mathcal{S}} = \begin{pmatrix} \frac{1}{2} \\ \frac{5}{2} \end{pmatrix}$$

and

$$M\begin{pmatrix}\frac{1}{2}\\ \frac{5}{2}\end{pmatrix}=\begin{pmatrix}7\\ -1\end{pmatrix}\longrightarrow 7(1,0)+(-1)(0,-1)=(7,1)$$

(**e**) The details of this case are left to the reader. The matrix in this case is

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

Note that the difference between (**d**) and (**e**) is that the order of the basis in the range is reversed. This resulted in an interchange of rows in the representing matrix.

Having solved some examples, we now state and prove the theorem which furnishes our procedure for the solution of the representation problem

Theorem 4.2.1. *(Solution of the representation problem) Let* $T\colon\ V\to W$, *where* $\dim V=n$ *and* $\dim W=m$, *be a linear transformation. Let* $\mathcal{S}=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ *and* $\mathcal{T}=\{\mathbf{w}_1,\dots,\mathbf{w}_m\}$ *be bases for* V *and* W, *respectively. There exists an* $m\times n$ *matrix* M *(unique to the ordered bases* $\mathcal{S}$ *and* $\mathcal{T}$*) with the property that for any* $\mathbf{x}\in V$, $(T(\mathbf{x}))_{\mathcal{T}}=M(\mathbf{x})_{\mathcal{S}}$.

Proof. Let $\mathbf{x}\in V$. Because $\mathcal{S}$ is a basis, there exists a unique set $\{c_1,\dots,c_n\}$ of constants such that $\mathbf{x}=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_n\mathbf{v}_n$. Now $T(\mathbf{x})=c_1(T(\mathbf{v}_1))+c_2(T(\mathbf{v}_2))+\cdots+c_n(T(\mathbf{v}_n))$ by the linearity of T. For each k $(1\le k\le n)$, $T(\mathbf{v}_k)$ is in W and can be represented by the basis elements for W:

$$T(\mathbf{v}_k)=a_{1k}\mathbf{w}_1+a_{2k}\mathbf{w}_2+\cdots+a_{mk}\mathbf{w}_m$$

Therefore,

$$\begin{aligned}T(\mathbf{x})=\ &c_1(a_{11}\mathbf{w}_1+a_{21}\mathbf{w}_2+\cdots+a_{m1}\mathbf{w}_m)\\&+c_2(a_{12}\mathbf{w}_1+a_{22}\mathbf{w}_2+\cdots+a_{m2}\mathbf{w}_m)+\cdots\\&+c_n(a_{1n}\mathbf{w}_1+a_{2n}\mathbf{w}_2+\cdots+a_{mn}\mathbf{w}_m)\end{aligned}$$

and after collecting terms we have

$$\begin{aligned}T(\mathbf{x})=\ &(a_{11}c_1+a_{12}c_2+\cdots+a_{1n}c_n)\mathbf{w}_1\\&+(a_{21}c_1+a_{22}c_2+\cdots+a_{2n}c_n)\mathbf{w}_2+\cdots\\&+(a_{m1}c_1+a_{m2}c_2+\cdots+a_{mn}c_n)\mathbf{w}_m\end{aligned}$$

The coefficients of $\mathbf{w}_1, \ldots, \mathbf{w}_m$ in the last expression are exactly the row-column products from

$$M \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

$$\begin{array}{ccc} \nearrow & \uparrow & \nearrow \\ (T(\mathbf{v}_1))_{\mathcal{T}} & (T(\mathbf{v}_2))_{\mathcal{T}} \cdots & (T(\mathbf{v}_n))_{\mathcal{T}} \end{array}$$

Therefore the $m \times n$ matrix M, the columns of which are the coordinate vectors of $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$, is the desired matrix. The matrix is unique to the pair of bases since coordinate vectors are unique in a given basis. $\square$

Example 6. Let $T\colon\ \mathcal{P}_1 \to \mathcal{P}_2$ be defined by $T(a+bx) = ax + (b/2)x^2$. Give $\mathcal{P}_1$ and $\mathcal{P}_2$ the standard bases $\mathcal{S} = \{1, x\}$ and $\mathcal{T} = \{1, x, x^2\}$, respectively. Find the matrix of T with respect to these bases. Do the same for $L\colon\ \mathcal{P}_2 \to \mathcal{P}_1$ defined by $L(a + bx + cx^2) = b + 2cx$.

Solution Now $T(1) = x$, so

$$(T(1))_{\mathcal{T}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Likewise, $T(x) = \frac{1}{2}x^2$, so

$$(T(x))_{\mathcal{T}} = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}$$

Therefore the matrix M_T representing T is

$$M_T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

For the transformation L,

$$L(1) = 0, \qquad L(x) = 1 \qquad L(x^2) = 2x$$

Thus

$$(L(1))_{\mathcal{S}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (L(x))_{\mathcal{S}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad (L(x^2))_{\mathcal{S}} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

and the matrix M_L representing L is

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

Note that

$$M_L M_T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

so that we could call M_L a **left inverse** of M_T. However,

$$M_T M_L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq I_3$$

and M_T is **not** a left inverse of M_L. Note that T is just antidifferentiation with arbitrary constant set to zero. When we antidifferentiate and then differentiate, we get the original function back. This is reflected by $M_L M_T = I$.

Example 7. Let $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{22}$ be defined by $T(A) = A - A^T$. Give $\mathcal{M}_{22}$ the standard basis

$$\mathcal{S} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$$

and find the matrix for T with respect to $\mathcal{S}$.

Solution First we calculate the images of the basis vectors:

$$\begin{aligned}
T\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3 + 0\mathbf{e}_4 \\
T\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = 0\mathbf{e}_1 + \mathbf{e}_2 - \mathbf{e}_3 + 0\mathbf{e}_4 \\
T\left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = 0\mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3 + 0\mathbf{e}_4 \\
T\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right) &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3 + 0\mathbf{e}_4
\end{aligned}$$

Therefore,

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

Example 8. Define $T\colon\ \mathbb{C}^2$ to $\mathbb{C}^2$ by $T((z_1, z_2)) = (iz_1, (1+i)z_2 - z_1)$. Let $\mathcal{C}^2$ have the basis $\mathcal{S} = \{(i, 0), (0, 1)\}$. Calculate M_T.

Solution

$$\begin{aligned} T(i, 0) &= (-1, -i) = i(i, 0) + (-i)(0, 1) \\ T(0, 1) &= (0, 1+i) = 0(i, 0) + (1+i)(0, 1) \end{aligned}$$

Therefore

$$M_T = \begin{pmatrix} i & 0 \\ -i & 1+i \end{pmatrix}$$

Some Algebra of Linear Transformations Let V and W be vector spaces, and let L and T be linear transformations from V to W. We can define the scalar multiple rL of L and the sum $L + T$ of L and T as linear transformations from V to W by the rules

$$\begin{aligned} (rL)(v) &= r(L(v)) && r \text{ a number, } v \text{ in } V \\ (L+T)(v) &= L(v) + T(v) && v \text{ in } V \end{aligned}$$

If M_L and M_T are the representing matrices with respect to bases $\mathcal{S}$ and $\mathcal{T}$, respectively, then rM_L represents rL and $M_L + M_T$ represents $L + T$. This can be shown by the method of proof in Theorem 4.2.1.

Example 9. Consider T as defined in Example 5**a**, and define $L\colon\ E^2 \to E^2$ by $L((x_1, x_2)) = (x_2, x_1)$. The standard matrices for T and L are

$$M_T = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \qquad M_L = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

So $T + L$ is defined as

$$\begin{aligned} (T+L)((x_1, x_2)) &= T((x_1, x_2)) + L((x_1, x_2)) \\ &= (x_1 + 2x_2, x_1 - x_2) + (x_2, x_1) \\ &= (x_1 + 3x_2, 2x_1 - x_2) \end{aligned}$$

and rT is defined as

$$\begin{aligned}(rT)((x_1, x_2)) = r(T((x_1, x_2))) &= r(x_1 + 2x_2, x_1 - x_2) \\ &= (rx_1 + 2rx_2, rx_1 - rx_2)\end{aligned}$$

The standard matrix for $T + L$ is

$$\begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$$

and the standard matrix for rT is

$$\begin{pmatrix} r & 2r \\ r & -r \end{pmatrix}$$

Direct calculation shows that

$$\begin{aligned} M_{T+L} &= M_T + M_L \\ M_{rL} &= rM_L \end{aligned}$$

PROBLEMS 4.2

1. For the following sets of homogeneous equations $AX = 0$, find rank A, dim (solution space), and verify that

$$\dim(\text{solution space}) + \text{rank } A = \text{no. of columns of } A$$

(a) $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ **(b)** $A = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 3 \end{pmatrix}$

(c) $A = \begin{pmatrix} 1 & 2 & -1 & 0 \\ 4 & -2 & 3 & 1 \\ 6 & 2 & 1 & 1 \\ 3 & -4 & 4 & 1 \end{pmatrix}$

2. For the following transformations $T\colon\ V \to W$, find the matrix of T, **assuming the standard basis in both** V and W.

(a) $T\colon\ E^2 \to E^2$, $T((x_1, x_2)) = (3x_1, 2x_2 - 5x_2)$

(b) $T\colon\ \mathcal{P}_1 \to E^2$, $T(ax + b) = (3a, 2a - 5b)$ for $\mathcal{P}_1, S = \{1, x\}$

(**c**) $T\colon\ E^2 \to E^3$, $\qquad T((x_1, x_2)) = (x_1, x_1 + x_2, 3x_1 - x_2)$

(**d**) $T\colon\ \mathcal{P}_2 \to \mathcal{P}_1 \longrightarrow T(ax^2 + bx + c) = cx + b$ for $\mathcal{P}_2, S = \{1, x, x^2\}$

(**e**) $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{22}$, $\qquad T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} c & d \\ a & b \end{pmatrix}$

(**f**) $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{21}$, $\qquad T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a \\ b \end{pmatrix}$

3. Let $T\colon\ E^3 \to E^2$ be defined by $T((x_1, x_2, x_3)) = (x_1 - x_2 + x_3, x_2 - x_3)$, let E^3 have the standard basis, and let E^2 have the basis $\mathcal{S} = \{(1, 1), (1, -1)\}$.

   (**a**) Find the matrix of T with respect to these bases.

   (**b**) Calculate $T((1, -1, 2))$ directly and by using the matrix of T.

4. Let $T\colon\ E^3 \to E^3$ be defined by $T((x_1, x_2, x_3)) = (x_1 + x_2, x_2 + x_3, x_3 + x_1)$, and let E^3 have the basis $\mathcal{S} = \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$.

   (**a**) Find the matrix of T with respect to $\mathcal{S}$.

   (**b**) Calculate $T((1, 1, 1))$ directly and by using the matrix of T.

5. Let $T\colon\ E^2 \to E^2$ be a linear transformation with the property $T(1, 1) = (1, 0)$ and $T(1, -1) = (0, 1)$. Find a matrix representation of T. [**Hint**: For the domain use $\mathcal{S} = \{(1, 1), (1, -1)\}$ as a basis.]

6. Define $L\colon\ E^n \to E^n$ by $L((x_1, x_2, x_3, \ldots, x_n)) = (x_2, x_3, \ldots, x_n, 0)$ and define $R\colon\ E^n \to E^n$ by $R((x_1, x_2, x_3, \ldots, x_n)) = (0, x_1, x_2, x_3, \ldots, x_{n-1})$. Find the matrices for L and R with respect to the standard basis.

7. Show that the matrix representing the zero transformation is the zero matrix regardless of basis.

8. Show that a contraction or dilation transformation from V to V has a diagonal matrix representation regardless of the basis given to V (same basis in domain and range).

9. Let T be defined by $T\colon\ \mathcal{P}_n \to \mathcal{P}_{n+1}$ as the coordinate operator

$$(T(f))(x) = xf(x)$$

Show that the standard matrix of T is

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & & \cdots & 0 & 1 \end{pmatrix}_{(n+2)\times(n+1)}$$

10. Let T be the momentum operator defined at the end of Sec. 4.1. Given that $\mathcal{P}_n{}^{\mathbb{C}}$ has the standard ordered basis $\{1, x, \ldots, x^n\}$, show that the standard matrix of T is

$$\begin{pmatrix} 0 & -i & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -2i & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -3i & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & \ddots & 0 \\ 0 & 0 & & \cdots & 0 & \cdots & -ni \end{pmatrix}_{n\times(n+1)}$$

11. For the following transformations $T\colon V \to W$, find the standard matrix of T.

 (**a**) $T\colon \mathbb{C}^2 \to \mathbb{C}^2, T((z_1, z_2)) = (z_1 + z_2, iz_2)$

 (**b**) $T\colon \mathbb{C}^3 \to \mathbb{C}^2, T((z_1, z_2, z_3)) = (iz_2, iz_3, 0)$

 (**c**) $T\colon \mathcal{C}_{22} \to \mathcal{C}_{22}, T(A) = A + iA^T$

4.3 SIMILAR MATRICES AND CHANGE OF BASIS

The purpose of a matrix representation M for a linear transformation T is to enable us to analyze T by working with M. If M is easy to work with, we have gained an advantage; if not, we have no advantage. Since different bases lead to different matrices, the "right" choice of basis to obtain a simple matrix M is important. The right choice of basis is not obvious, as Example 1 shows.

Example 1. Show that $T\colon\ E^2 \to E^2$ defined by $T(x_1, x_2) = (x_1+6x_2, 3x_1+4x_2)$ has standard matrix

$$\begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix}$$

Then show that, with respect to the basis $\mathcal{T} = \{(2,-1),(1,1)\}, T$ has a diagonal matrix representation.

Solution For the standard matrix we have

$$T((1,0)) = (1,3) = 1(1,0) + 3(0,1) \qquad \text{so} \qquad (T((1,0)))_{\text{std}} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$T((0,1)) = (6,4) = 6(1,0) + 4(0,1) \qquad \text{so} \qquad (T((0,1)))_{\text{std}} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

and

$$M_{\text{std}} = \begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix}$$

But with respect to $\mathcal{T}$,

$$T((2,-1)) = (-4,2) = -2(2,-1) + 0(1,1) \qquad \text{so} \qquad [T((1,0))]_{\mathcal{T}} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$T((1,1)) = (7,7) = 0(2,-1) + 7(1,1) \qquad \text{so} \qquad [T((1,1))]_{\mathcal{T}} = \begin{pmatrix} 0 \\ 7 \end{pmatrix}$$

and the matrix with respect to $\mathcal{T}$ is

$$M = \begin{pmatrix} -2 & 0 \\ 0 & 7 \end{pmatrix}$$

We recall from matrix algebra that diagonal matrices are easy to work with for certain operations: inversion, determinants, and multiplication, to name three. As the dimension of the vector spaces (and size of the matrices) grows, this is even more the case. We need, then, to find a way of getting the simplest possible matrix to represent a transformation T. To solve this problem (a solution is presented in Chap. 5), we must discover how to relate different matrix representations for the given linear transformation. We restrict our attention to the case $V = W$ with the same basis in V and W. This case occurs most frequently in applications.

To discover the relationship, suppose that $M_{(\mathcal{S})}$ is the matrix representing $T\colon\ (V, \mathcal{S}) \to (V, \mathcal{S})$ and that $M_{(\mathcal{T})}$ represents $T\colon\ (V, \mathcal{T}) \to (V, \mathcal{T})$. Let P be the transition matrix from basis $\mathcal{T}$ to basis $\mathcal{S}$, so that for any x in V,

$$\begin{aligned} M_{(\mathcal{T})}(\mathbf{x})_{\mathcal{T}} = (\mathbf{x})_{\mathcal{T}} = P^{-1}(T\mathbf{x})_{\mathcal{S}} &= P^{-1}(M_{(\mathcal{S})}(\mathbf{x}_{\mathcal{S}}) \\ &= P^{-1}M_{(\mathcal{S})}I(\mathbf{x})_{\mathcal{S}} \\ &= P^{-1}M_{\mathcal{S}}(PP^{-1})(\mathbf{x})_{\mathcal{S}} \\ &= (P^{-1}M_{(\mathcal{S})}P)(P^{-1}(\mathbf{x})_{\mathcal{S}}) \\ &= (P^{-1}M_{(\mathcal{S})}P)(\mathbf{x})_{\mathcal{T}} \end{aligned}$$

Therefore

$$M_{(\mathcal{T})}(\mathbf{x})_{\mathcal{T}} = (P^{-1}M_{(\mathcal{S})}P)(\mathbf{x})_{\mathcal{T}} \quad \text{for all } \mathbf{x} \text{ in } V$$

so that $M_{(\mathcal{T})} = P^{-1}M_{(\mathcal{S})}P$. These equations actually give a proof of the basic result.

Theorem 4.3.1. *Let $T\colon\ V \to V$ be a linear transformation with matrix $M_{(\mathcal{S})}$ with respect to a basis $\mathcal{S}$ and with matrix $M_{(\mathcal{T})}$ with respect to a basis $\mathcal{T}$. If P is the transition matrix from basis $\mathcal{T}$ to basis $\mathcal{S}$, then*

$$M_{(\mathcal{T})} = P^{-1}M_{(\mathcal{S})}P$$

The relation $M_{(\mathcal{T})} = P^{-1}M_{(\mathcal{S})}P$ is important enough to be given a name.

Definition 4.3.1. Two $n \times n$ matrices A and B are **similar** if there exists an invertible matrix P such that $B = P^{-1}AP$.

Note that the definition in no way tells us how to find the **similarity transform** P. In the case of two representation matrices for a linear transformation, P is a transition matrix from one basis to another, as we saw in Theorem 4.3.1.

An important restatement of Theorem 4.3.1 is as follows:

> Let $T\colon\ V \to V$ be a linear transformation. Any two representing matrices of T are similar.

Example 2. In Example 1, denote the standard basis by $\mathcal{S}$. Illustrate Theorem 4.3.1 for T, $\mathcal{S}$, and $\mathcal{T}$, as given in Example 1.

Solution The standard matrix, as before, is

$$M_{(\mathcal{S})} = \begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix}$$

Now calculate the transition matrix from $\mathcal{T}$ to $\mathcal{S}$:

$$\begin{aligned} (2,-1) &= 2(1,0) + (-1)(0,1) \\ (1,1) &= 1(1,0) + 1(0,1) \\ P = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} &\Rightarrow P^{-1} = \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \end{aligned}$$

Then

$$\begin{aligned} P^{-1}M_{(\mathcal{S})}P = P^{-1}\begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix} P &= \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} -4 & 7 \\ 2 & 7 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 0 \\ 0 & 7 \end{pmatrix} = M_{(\mathcal{T})} \end{aligned}$$

One way to remember how to relate matrices with respect to $\mathcal{S}$ and $\mathcal{T}$ is to use the diagram

$$(V,\Im)@>T>M_{\Im}>(V,\Im)@VIVP_{\mathcal{S}\leftarrow\Im}V@AP_{\Im\leftarrow\mathcal{S}}AIA(V,\mathcal{S})@>M_{\mathcal{S}}>T>(V,\mathcal{S})$$

The point is that the transition matrix is just the matrix which represents the identity transformation. The notation $P_{\mathcal{A}\leftarrow\mathcal{B}}$ indicates the transition matrix from $\mathcal{B}$ to $\mathcal{A}$. Note that

$$\begin{array}{c} M_{(\mathcal{T})} = P_{\mathcal{T}\leftarrow\mathcal{S}} M_{(\mathcal{S})} \quad P_{\mathcal{S}\leftarrow\mathcal{T}} \\ \nwarrow \quad \uparrow \quad \nearrow \\ \text{All the same} \end{array}$$

and remember that $P_{\mathcal{T}\leftarrow\mathcal{S}} = (P_{\mathcal{S}\leftarrow\mathcal{T}})^{-1}$

To exploit fully the result of Theorem 4.3.1, we will later use some properties of similar matrices.

Theorem 4.3.2. *The following statements regarding similarity are true. In all cases the matrices are $n \times n$.*

(**a**) *Matrix A is similar to A.*

(**b**) *If A is similar to B, then B is similar to A.*

(**c**) *If A is similar to B and B is similar to C, then A is similar to C.*

(**d**) *If A is similar to B, then* $\det A = \det B$.

(**e**) *If A is similar to B, then tr $A =$ tr B.*

(**f**) *If A is similar to B, then A^m is similar to B^m for any positive integer m.*

(**g**) *If A is similar to B, then A is invertible if and only if B is invertible. In that case A^{-1} is similar to B^{-1}.*

[**Note**: Parts (**d**), (**e**), and (**g**) state, respectively, that the determinant, trace, and invertibility are **invariant under similarity**.]

Proof. (**a**) Since $A = IAI = I^{-1}AI$, A is similar to A.

(**b**) If A is similar to B, then $B = P^{-1}AP$. Thus $PBP^{-1} = P(P^{-1}AP)P^{-1} = A$. Thus $A = (P^{-1})^{-1}B(P^{-1})$. Calling P^{-1} by the name $\mathcal{P}$, we have $A = \mathcal{P}^{-1}B\mathcal{P}$, and B is similar to A.

(**c**) If A is similar to B and B is similar to C, we have $B = P^{-1}AP$ and $C = Q^{-1}BQ$, where P and Q are, in general, not equal. Now we have

$$C = Q^{-1}(B)Q = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) = (PQ)^{-1}A(PQ)$$

Since PQ is invertible, C is similar to A.

(**d**) If $B = P^{-1}AP$, then

$$\begin{aligned}\det B &= \det(P^{-1}AP)\\ &= (\det P^{-1})(\det A)(\det P) = (\det A)(\det P^{-1})(\det P)\\ &= (\det A)\left(\frac{1}{\det P}\right)(\det P) = \det A\end{aligned}$$

(**e**) If $B = P^{-1}AP$, since tr $(AB) =$ tr (BA) we have tr $B =$ tr $(P^{-1}AP) =$ tr $((P^{-1}A)P) =$ tr $(P(P^{-1}A)) =$ tr A.

(**f**) If $B = P^{-1}AP$, then

$$\begin{aligned} B^2 = (P^{-1}AP)^2 &= (P^{-1}AP)(P^{-1}AP) \\ &= (P^{-1}A)(PP^{-1})(AP) \\ &= P^{-1}A^2P \end{aligned}$$

Thus B^2 is similar to A^2 and has the same similarity transform P.

Now we use induction. Suppose for the induction hypothesis that A^k is similar to B^k with $B^k = P^{-1}A^kP$. Now we show that B^{k+1} is similar to A^{k+1}. By the induction hypothesis

$$B^k = P^{-1}A^kP$$

Now

$$\begin{aligned} B^{k+1} &= BP^{-1}A^kP = (P^{-1}AP)P^{-1}A^kP \\ &= P^{-1}AIA^kP \\ &= P^{-1}A^{k+1}P \end{aligned}$$

Therefor B^{k+1} is similar to A^{k+1} with similarity transform P. By the principle of mathematical induction, (**f**) is proved.

(**g**) Since $\det A = \det B$, we know that $\det A \neq 0$ if and only if $\det B \neq 0$. For the second part note that

$$A = P^{-1}BP, A^{-1} = (P^{-1}BP)^{-1} = P^{-1}B^{-1}(P^{-1})^{-1} = P^{-1}B^{-1}P.$$

□

Parts (**d**) and (**e**) are illustrated by the similar matrices

$$\begin{pmatrix} 1 & 6 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -2 & 0 \\ 0 & 7 \end{pmatrix}$$

from Example 1. Both matrices have trace 5 and determinant -14.

The problem of determining whether two given matrices are similar is generally difficult. But parts (**d**) and (**e**) can be used to rule out similarity: If tr $A \neq$ tr B or $\det A \neq \det B$, then A cannot be similar to B.

Example 3. For the following pairs of matrices, determine whether A is similar to B.

(**a**) $A = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$ $\qquad B = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix}$

(**b**) $A = \begin{pmatrix} 2 & 6 & 2 \\ 5 & 1 & -1 \\ 4 & 1 & 3 \end{pmatrix}$ $\qquad B = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 2 & 2 & 5 \end{pmatrix}$

(**c**) $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ $\qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Solution

(**a**) Although tr $A =$ tr $B, \det A \neq \det B$, so A is not similar to B.

(**b**) Since tr $A \neq$ tr B, matrix A is not similar to matrix B.

(**c**) In this case, the traces and determinants of A and B coincide, so similarity cannot be easily ruled out. Since A and B are small, we can check the equation

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

If this is to hold, we must have

$$PB = AP$$

for a nonsingular matrix p. Let

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then the required equation is

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

which leads to

$$\begin{aligned} a &= a + c \Rightarrow 0 = c \\ b &= b + d \Rightarrow 0 = d \\ c &= c \qquad \Rightarrow c = c \\ d &= d \qquad \Rightarrow d = d \end{aligned}$$

Therefore a and b are arbitrarily, and $c = d = 0$ which leads to

$$P = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$$

which is singular. Therefore A is not similar to B. For large matrices, this method of solution is not practical.

PROBLEMS 4.3

In Probs. 1 to 9 a linear transformation $T\colon\ V \to V$ and bases $\mathcal{S}$ and $\mathcal{T}$ are given.

(**a**) Find the matrix of T with respect to $\mathcal{S}$.

(**b**) Find the matrix of T with respect to $\mathcal{T}$ by using the transition matrix.

(**c**) Find the matrix of T with respect to $\mathcal{T}$ directly.

1. $T\colon\ E^2 \to E^2, T((x_1, x_2)) = (x_1, 0)$
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1), (1, -1)\}$

2. $T\colon\ E^2 \to E^2, T((x_1, x_2)) = (2x_1, 2x_2)$
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1), (1, -1)\}$
   (This is a dilation. Compare the result of this problem with Prob. 8 of the last section.)

3. $T\colon\ E^2 \to E^2, T((x_1, x_2)) = (x_1 + x_2, 2x_1 - 3x_2)$
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1), (1, -1)\}$

4. T as in Prob. 3
   $\mathcal{S} = \{(1, 1), (1, -1)\} \qquad \mathcal{T} = \{(1, 1), (1, 2)\}$

5. $T\colon\ E^2 \to E^2$, where T is rotation through $\pi/4$ radians counterclockwise.
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1), (1, 2)\}$

6. $T\colon\ E^3 \to E^3, T((x_1, x_2, x_3)) = (x_1, x_1 + x_2, x_2 - x_3)$
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1, 0), (-1, 1, 0), (0, 0, 1)\}$

7. $T\colon\ E^3 \to E^3, T((x_1, x_2, x_3)) = (x_1 + 2x_2, x_1 + x_2 + x_3, x_3)$
   $\mathcal{S} =$ standard basis $\qquad \mathcal{T} = \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$

8. $T\colon\ \mathcal{P}_1 \to \mathcal{P}_1, T(a+bx) = a+b+(2a-3b)x$
   $\mathcal{S}$ = standard basis $\qquad \mathcal{T} = \{1+x, 1-x\}$

9. $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{22}, T(A) = A^T$

   $$\mathcal{S} = \text{standard basis} \qquad \mathcal{T} = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

10. Show that if A is similar to B and A is invertible, then A^{-k} is similar to B^{-k} for $k = 1, 2, \ldots$

11. Pairs of matrices A and B are given. In each case show that A and B are not similar.
    (**a**) $A = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & 1 \\ 3 & 2 \end{pmatrix}$
    (**b**) $A = \begin{pmatrix} 9 & 3 & 7 \\ 0 & 5 & 6 \\ 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$
    (**c**) $A = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

12. Show that if A and B are similar matrices, then rank A = rank B. (**Hint**: Use the fact that similar matrices represent the same linear transformation.)

13. Compare the results of Probs. 3 and 8. Is there a reasonable identification of the transformations (and V) in these problems?

14. Let M be an $m \times n$ matrix. Let V be an n-dimensional vector space with basis S, and let W be an m-dimensional vector space with basis T. Let $\mathbf{x}$ be in V. Define a transformation L from V to W by this rule: L takes $\mathbf{x}$ in V to $\mathbf{y}$ in W as follows:

    1. Replace x by $(\mathbf{x})_S$, the coordinate matrix of x.
    2. Calculate $M\ (\mathbf{x})_S$.
    3. Let y be the vector in W with $(\mathbf{y})_T = M(\mathbf{x})_S$.

Show that L so defined is linear. This means that any matrix generates a linear transformation.

4.4 INVERTIBLE TRANSFORMATIONS AND CLASSIFYING TRANSFORMATIONS

One of the most important problems in applied mathematics is to solve

$$T(\mathbf{x}) = \mathbf{y} \tag{4.4.1}$$

where $T\colon\ V \to V$ is a linear transformation, V is a vector space, $\mathbf{y}$ is given, and $\mathbf{x}$ is to be found. If V is given a basis $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and M is the matrix of T with respect to $\mathcal{S}$, then the corresponding matrix problem is

$$M(\mathbf{x})_{\mathcal{S}} = (\mathbf{y})_{\mathcal{S}} \tag{4.4.2}$$

which is the first fundamental problem of linear algebra. So we can see the importance of the representation problem in reducing transformation equations in applied mathematics to matrix equations. If we solve Eq. (4.4.2), we have essentially inverted the transformation T. This concept must now be developed.

Definition 4.4.1. Let $T\colon\ V \to W$ and $L\colon\ W \to Z$, where V, W, and Z are vector spaces. The composition of L and T, denoted $L \circ T$, is defined for $\mathbf{x} \in V$ by

$$(L \circ T)(\mathbf{x}) = L(T(\mathbf{x}))$$

Pictorially the situation in Definition 4.4.1 is shown in Fig. 4.4.1. It is not hard to define longer strings of compositions $L(T(U(\mathbf{x}))$ and to see that $((L \circ T) \circ U)(\mathbf{x}) = (L \circ (T \circ U))(\mathbf{x})$.

Example 1. Let $T\colon\ E^2 \to E^3$ and $L\colon\ E^3 \to E^2$ be defined by $T((x_1, x_2)) = (x_1, x_1 + x_2, x_2)$ and $L((x_1, x_2, x_3)) = (x_1 - x_2, x_2 - x_3)$. Calculate $(L \circ T)((x_1, x_2))$. Show that $L \circ T$ is a linear transformation. Determine the standard matrices for L, T, and $L \circ T$. Show that the matrix of $L \circ T$ is the product of the matrix of L and the matrix of T.

Solution By Definition, $(L \circ T)((x_1, x_2)) = L(T(x_1, x_2)) = L((x_1, x_1 + x_2, x_2)) = (x_1 - (x_1 + x_2), (x_1 + x_2) - x_2) = (-x_2, x_1)$. Now let (x_1, x_2) and (y_1, y_2) be two vectors in E^2. We have $(L \circ T)((x_1, x_2) + (y_1, y_2)) = (L \circ T)((x_1 + y_1, x_2 + y_2)) = (-(x_2 + y_2), x_1 + y_1)$. Also $(L \circ T)((x_1, x_2)) + (L \circ T)((y_1, y_2)) = (-x_2, x_1) + (-y_2, y_1) = (-x_2 - y_2, x_1 + y_1)$. It is easy to show that $(L \circ T)(r(x_1, x_2)) = r(L \circ T)((x_1, x_2))$; therefore $L \circ T$ is linear. By the methods of Sec. 4.3 we have the following:

TRANSFORMATION	**STANDARD MATRIX**
L	$M_L = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$
T	$M_T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$
$L \circ T$	$M_{L\circ T} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$

Furthermore,

$$M_L M_T = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = M_{L\circ T}$$

Note that $M_T M_L \neq M_{L\circ T}$. This reflects something we already know: Matrix multiplication is, in general, not commutative.

Example 1 illustrates the following theorem.

Theorem 4.4.1. *If $T\colon\ (V, \mathcal{S}) \to (W, \mathcal{T})$ and $L\colon\ (W, \mathcal{T}) \to (Z, \mathcal{B})$ are linear, then $(L \circ T) : (V, \mathcal{S}) \to (Z, \mathcal{B})$ is linear. If M_T, M_L, and $M_{L\circ T}$ are the matrices representing T, L, and $L \circ T$, respectively, then*

$$M_{L\circ T} = M_L M_T$$

Proof. For the linearity, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V, and let r and s be numbers. By the linearity of L and T separately, $(L\circ T)(r\mathbf{u}+s\mathbf{v}) = L(T(r\mathbf{u}+s\mathbf{v})) = L(rT(\mathbf{u})+sT(\mathbf{v})) = rL(T(\mathbf{u}))+sL(T(\mathbf{v})) = r(L\circ T)(\mathbf{u})+s(L\circ T)(\mathbf{v})$. Thus $L \circ T$ is linear. To show that the matrix representation, let $\mathbf{x} \in V$, $T(\mathbf{x}) = \mathbf{y}$, and $L(T(\mathbf{x})) = \mathbf{z}$. Now $(\mathbf{y})_{\mathcal{T}} = M_T(\mathbf{x})_{\mathcal{S}}$ and $(\mathbf{z})_{\mathcal{B}} = M_L(\mathbf{y})_{\mathcal{T}}$. Therefore $(\mathbf{z})_{\mathcal{B}} = M_L(M_T(\mathbf{x})_{\mathcal{S}}) = (M_L M_T)(\mathbf{x})_{\mathcal{S}}$. But we also have

$$(\mathbf{z})_{\mathcal{B}} = M_{L\circ T}(\mathbf{x})_{\mathcal{S}}$$

By the uniqueness of matrix representation,

$$M_{L\circ T} = M_L M_T$$

The idea of the composition of transformation is set; we can define the inverse of a transformation. $\square$

Definition 4.4.2. Let $T\colon\ V \to V$ be a linear transformation. The (two-sided) inverse of T is a transformation $T^{-1}\colon\ V \to V$ for which

$$(T^{-1} \circ T)(\mathbf{x}) = \mathbf{x} \qquad \text{for all } \mathbf{x} \in V$$

and

$$(T \circ T^{-1})(\mathbf{x}) = \mathbf{x} \qquad \text{for all } \mathbf{x} \in V$$

If T^{-1} exists, the T is called **invertible**.

We note that T^{-1} is linear. In fact, $T^{-1}(\mathbf{x}+\mathbf{y}) = T^{-1}(T(T^{-1}(\mathbf{x})) + T(T^{-1}(\mathbf{y}))) = T^{-1}(T(T^{-1}(\mathbf{x}) + T^{-1}(\mathbf{y}))) = (T^{-1}T)(T^{-1}(\mathbf{x}) + T^{-1}(\mathbf{y})) = T^{-1}(\mathbf{x})\ + T^{-1}(\mathbf{y})$ and $T^{-1}(rx) = T^{-1}(rT(T^{-1}(\mathbf{x}))) = T^{-1}(T(rT^{-1}(\mathbf{x}))) = (T^{-1}T)$
$(rT^{-1}(\mathbf{x})) = rT^{-1}(\mathbf{x})$. Other important properties of inverses are contained in Theorem 4.4.2.

Theorem 4.4.2. *Let $T\colon\ (V,\mathcal{S}) \to (V,\mathcal{S})$ be a linear transformation.*

(**a**) *Let M be the matrix of T with respect to $\mathcal{S}$. Then T is invertible if and only if M^{-1} exists. Moreover, the matrix for T^{-1} in this case is precisely M^{-1}.*

(**b**) *T is invertible if and only if $\eta(T) = 0$ [that is, $\dim(\ker T) = 0$].*

(**c**) *T is invertible if and only if $T(\mathbf{x}) = T(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in V$. (That is, T is a one-to-one function.)*

Proof. (**a**) $\Rightarrow$ Since $(T \circ T^{-1})(\mathbf{x}) = \mathbf{x}$, we have

$$(\text{matrix of } T \text{ times matrix of } T^{-1})(\mathbf{x})_s = (\mathbf{x})_s$$

If $M_{T^{-1}}$ denotes the matrix for T^{-1}, the last equation is

$$MM_{T^{-1}}(\mathbf{x})_s = (\mathbf{x})_s$$

which means that $MM_{T^{-1}} = I$. Therefore $M_{T^{-1}} = M^{-1}$.

($\Leftarrow$) Suppose M^{-1} exists. Since it represents a linear transformation L and $MM^{-1} = M^{-1}M = I$, we know that $(L \circ T)(\mathbf{x}) = \mathbf{x}$ and $(T \circ L)(\mathbf{x}) = \mathbf{x}$. Thus $L = T^{-1}$, and T is invertible.

(**b**) Let $\dim V = n$ and M be the matrix of T.

($\Rightarrow$) Suppose T is invertible and $\eta(T) > 0$. Then since $\eta(T)+$ rank $T = n$, rank $T < n$. Therefore rank $M < n$ and M^{-1} does not exist, a contradiction since if T^{-1} exists, M^{-1} must exist.[1]

($\Leftarrow$) If $\eta(T) = 0$, then

$$\eta(T) + \mathcal{R}(T) = n$$

we have $\mathcal{R}(T) =$ rank $T = n$. Thus rank $M = n$ and M exists. Reasoning as in (**a**) now, we see that T^{-1} must exist.

(**c**) ($\Rightarrow$) Suppose T is invertible and there exist $\mathbf{x}'$ and $\mathbf{y}'$ in V with $\mathbf{x}' \neq \mathbf{y}'$ and $T(\mathbf{x}') = T(\mathbf{y}')$. Then $\mathbf{x}' - \mathbf{y}' \neq \theta$ and $T(\mathbf{x}' - \mathbf{y}') = \theta$. But then $\dim(\ker T) \neq 0$. That is, $\eta(T) \neq 0$, which contradicts (**b**).

($\Leftarrow$) Suppose $T(\mathbf{x}) = T(\mathbf{y})$ implies that $\mathbf{x} = \mathbf{y}$ for all $\mathbf{x}$,$\mathbf{y}$ in V and that T is not invertible. Then by (**b**) $\dim \ker T \neq 0$. Therefore $\ker T$ contains at least θ and some $\mathbf{z} \neq \theta$. Now since $\ker T$ is a subspace, $\mathbf{z} - \theta \in \ker T$. Thus $T(\mathbf{z} - \theta) = T(\mathbf{z}) - T(\theta) = \theta$. So we have $T(\mathbf{z}) = T(\theta)$, but $\mathbf{z} \neq \theta$, which contradicts our assumption.

Part (**a**) of Theorem 4.4.2 tells us that we can determine the invertibility of a transformation by determining the invertibility of any representing matrix. This is so because if A and B are any two representing matrices, then they are similar: $A = P^{-1}BP$. Now since $\det A = \det B$, matrix A is invertible if and only if matrix B is invertible. So if one representing matrix is invertible, all are. $\square$

Example 2. Define $T\colon\ E^3 \to E^3$ by $T(\mathbf{x}) = T((x_1, x_2, x_3)) = (bx_3 - cx_2, cx_1 - ax_3, ax_2 - bx_1)$. This transformation maps (x_1, x_2, x_3) to $(a, b, c) \times (x_1, x_2, x_3)$, the cross product of the fixed vector (a, b, c) with $\mathbf{x}$. Show that T is not invertible, thereby showing that we cannot determine a vector from its cross product with a known vector (a, b, c).

Solution We will find a matrix M representing T and show that M^{-1} does not exist. Since invertibility is preserved by similarity, we may use any representing matrix to determine invertibility. We use the standard matrix since it is easiest to compute. We have

$$\begin{aligned} T(1,0,0) &= (0, c, -b) \\ T(0,1,0) &= (-c, 0, a) \\ T(0,0,1) &= (b, -a, 0) \end{aligned}$$

[1]Prob. 12 of Sec. 4.3 shows that rank T is defined as rank M_T.

so the matrix is

$$M = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$$

Note that M is antisymmetric. Now $\det M = abc - abc = 0$, so M^{-1} does not exist. Therefore T is not invertible.

Example 3. (Shear revisited) Imagine a cube of gelatin held between the hands as viewed from the side in Fig. 4.4.2**a**. Suppose the upper hand is moved to the right k units (k is small). Then the height of the gelatin will not change much, and the side view will look like a parallelogram. This action in mechanics leads to shear. In Fig. 4.4.2**b**, some vectors are imposed on the face of the gelatin. **Assuming that the shear is a linear transformation** S, find a matrix representing S and show that S is invertible. (Invertibility is reasonable: To "undo" the shear, we can just move the top hand back to its original position.)

Solution As stated before, the basic principle for finding matrix representations is that a linear transformation is determined by its action or basis elements. Since $\mathcal{S} = \{(1,0),(0,1)\}$ is a basis for E^2, the actions of S on (1,0) and (0,1) can be used:

$$\begin{aligned} S((1,0)) &= (1,0) = 1(1,0) + 0(0,1) \\ S((0,1)) &= (k,1) = k(1,0) + 1(0,1) \end{aligned}$$

Therefore the matrix M of S with respect to $\mathcal{S}$ is

$$M = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

Matrix M is invertible; therefore by Theorem 4.4.2**a**, S is invertible. Note that

$$M^{-1} = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix}$$

which represents shear with the upper face of the cube moving $-k$ units to the right, which means k units to the left. This, of course, makes sense: To undo moving the upper hand to the right, simply move it to the left the same distance.

Example 4. Consider a square as shown in Fig. 4.4.3. Let R be a counterclockwise rotation about c of $90°$. Show, using matrices, that four successive applications of R give the identity transformation.

Solution We find the standard matrix of R and calculate its fourth power. Since $R(1,0) = (0,1) = 0(1,0) + 1(0,1)$ and $R(0,1) = (-1,0) = -1(1,0) + 0(0,1)$, the standard matrix is

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

Now

$$M^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad M^4 = M^2M^2 = I$$

Classifying Linear Transformations We have seen that if a matrix A is invertible and a matrix B is similar to A, then B is invertible also. That is, invertibility is preserved under similarity. As a result of Theorem 4.2.2**a**, we say that a linear transformation T is invertible if any matrix representation of T is an invertible matrix. Because other properties of matrices are preserved under similarity, we make the following definition.

Definition 4.4.3. Let $\mathcal{P}$ be a property of matrices which is preserved under similarity. We classify a linear transformation $T\colon\ V \to V$ as having property $\mathcal{P}$ if in some basis the matrix representing T has property $\mathcal{P}$.

Example 5. Show that the property of idempotency is preserved under similarity.

Solution Let A be idempotent (that is, $A^2 = A$), and let B be similar to A: $B = P^{-1}AP$. Now $B^2 = (P^{-1}AP)(P^{-1}AP) = P^{-1}AIAP = P^{-1}A^2P = P^{-1}AP = B$. Therefore B is idempotent.

Example 6. Show that the linear transformation $T\colon\ E^3 \to E^3$ of projection $T((x_1, x_2, x_3)) = (x_1, 0, x_3)$ is idempotent.

Solution By Example 5, idempotency is preserved under similarity. All we need to do is to find a matrix representation of T which is an idempotent matrix. We look at the standard matrix, which is easiest to compute. It is

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now $M^2 = M$, so T is idempotent.

Another property preserved under similarity is nilpotency.

Example 7. Show that nilpotency is preserved under similarity. Then show that the **left-shift** linear transformation $T\colon\ E^4 \to E^4$ defined by $T(x_1, x_2, x_3, x_4) = (x_2, x_3, x_4, 0)$ is a nilpotent transformation.

Solution Let A be nilpotent of exponent k, so that $A^k = 0$. If B is similar to A, then $B^k = P^{-1}A^kP$ by Theorem 4.3.2**f**. Thus $B^k = P^{-1}0P = 0$, and B is nilpotent of exponent k. The standard matrix for T is

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The powers of M are

$$M^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad M^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad M^4 = 0$$

Therefore T is nilpotent of exponent 4. So four successive applications of T produce the zero vector. Note that T is not invertible because M is not invertible. The noninvertibility makes sense because when we shift left, we lose entirely the information from the first component.

Linear Equations (Reprise) As we said at the beginning of this section, the problem of solving

$$L(\mathbf{x}) = \mathbf{y}$$

for $\mathbf{x}$ in V, given the linear transformation $L\colon\ V \to V$ and $\mathbf{y}$ in V, is a generalization of the first basic problem of linear algebra. When V is finite-dimensional, the problem reduces to the first basic problem of solving linear equations once a basis is assigned to V and a matrix representing L is found. In this case the equation $L(\mathbf{x}) = \mathbf{y}$ is uniquely solvable if and only if M_L is invertible. When M_L is not invertible, $\dim(\ker L) \neq 0$ and the general solution is of the form $\mathbf{p} + \mathbf{h}$ where $\mathbf{p}$ is a particular solution of $L(\mathbf{x}) = \mathbf{y}$ and $\mathbf{h}$ is a solution of the associated homogeneous problem

$L(\mathbf{x}) = \theta$. (See Theorem 1.5.5.) To illustrate this, consider $L\colon\ \mathcal{P}_2 \to \mathcal{P}_2$, where L is differentiation. We wish to solve $L(f) = g$. The standard matrix for L is

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

The matrix is not invertible, so we cannot expect a unique solution. Let

$$(g)_{\mathcal{S}} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Then

$$M(f)_{\mathcal{S}} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

can have a solution if and only if $c = 0$. When $c \neq 0, g$ is not in the range of L. Now when $c = 0, g$ is in the range of L, and using the matrix equation

$$M(f)_{\mathcal{S}} = (g)_{\mathcal{S}}$$

we find

$$(f)_{\mathcal{S}} = \begin{pmatrix} k \\ a \\ b/2 \end{pmatrix}$$

where k is arbitrary. That is, $f(x) = k + ax + bx^2/2$, where the constant k is arbitrary (remember antiderivatives and the arbitrary "constant of integration"?). Note that $p(x) = ax + bx^2/2$ is a particular solution of $L(f) = g$, and $h(x) = k$ is the most general solution to $L(f) = \theta$. Thus the general solution, which exists only when g is not a quadratic ($c = 0$), is of the form $p + h$, where h is in the kernel of L.

For a pure matrix problem consider

$$\begin{aligned} x + y - z + w &= 1 \\ 2x + y \phantom{{}-2z} - w &= -2 \\ x \phantom{{}+y} + z - 2w &= -3 \\ y - 2z + 3w &= 4 \end{aligned}$$

The equations are equivalent to

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & -3 \\ 0 & 1 & -2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Thus the solution exists but is not unique. A particular solution is obtained by putting $z = 1$ and $w = 1$, which leads to $y = 3$ and $x = -2$. Looking at the associated homogeneous system

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 0 \\ 0 & 1 & -2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

we find a solution

$$\mathbf{h} = \begin{pmatrix} 2s - r \\ 2r - 3s \\ r \\ s \end{pmatrix}$$

Thus the most general solution is

$$\mathbf{p} + \mathbf{h} = \begin{pmatrix} -2 \\ 3 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2s - r \\ 2r - 3s \\ r \\ s \end{pmatrix}$$

The preceding examples illustrate the following theorem, which is stated without proof.

Theorem 4.4.3. *Let V be a vector space with $\dim V \geq 1$, and consider the linear equation*

$$L(\mathbf{x}) = \mathbf{y}$$

where $L\colon\ V \to V$ is linear, $\dim(\text{range } L) \geq 1$, and $\mathbf{y}$ is in V. If $\mathbf{y}$ is in the range of L, then either (a) the equation has a unique solution or (b) the equation has an infinite number of solutions which are of the form $\mathbf{p} + \mathbf{h}$, where $\mathbf{h}$ is the general solution of $L(\mathbf{x}) = \theta$ and p is any particular solution of $L(\mathbf{x}) = \mathbf{y}$. If $\mathbf{y}$ is not in the range of L, then the equation has no solution.

Finally let us consider a case which can arise in applications: attempting to solve $L(\mathbf{x}) = \mathbf{y}$ when $\mathbf{y}$ is not in the range of L: Theorem 4.4.3 tells us that a solution in the usual sense cannot be expected but if an inner product can be defined on V, we could call a vector $\tilde{\mathbf{x}}$ an **approximate solution** if

$$\|L(\tilde{\mathbf{x}}) - \mathbf{y}\|^2 \leq \|L(\mathbf{x}) - \mathbf{y}\|^2 \qquad \text{for all } \mathbf{x} \text{ in } V$$

Consider, the problem of solving

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 + x_2 &= 1 \end{aligned}$$

which has no solution. Defining $L\colon\ \mathcal{M}_{21} \to \mathcal{M}_{21}$ by

$$L\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

we see that

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

the right-hand side of the equations, is not in the range of L. Using the standard inner product on $\mathcal{M}_{21}$, we have

$$\left\| L\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) - \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\|^2 = (x_1 + x_2 - 2)^2 + (x_1 + x_2 - 1)^2$$

Methods of calculus show that this last expression is minimized if $x_1 + x_2 = \frac{3}{2}$; this does not yield a unique approximate solution. One way to avoid this problem is to require that the approximate solution be of **minimum norm** (in this case, the one closest to the origin). With this requirement we obtain $\tilde{x} = \left(\frac{3}{4}, \frac{3}{4}\right)$ (see Fig. 4.4.4). Finding minimum norm solutions leads to the concepts of **generalized inverses and regularization**, which the interested reader can find in texts such as **Regression and the Moore-Penrose Pseudoinverse**, by A.E. Albert (Academic Press, New York, 1972).

PROBLEMS 4.4

In Probs. 1 to 6, a linear transformation and a property or properties (similarity preserved) are given. Determine whether the given transformation has the given property.

1. $T\colon\ E^3 \to E^3, T((x_1, x_2, x_3)) = (x_1, x_2, 0)$; invertibility, idempotency
2. $T\colon\ \mathcal{P}_2 \to \mathcal{P}_2, T(ax^2 + bx + c) = 2xa + b$; invertibility, nilpotency
3. $T\colon\ \mathcal{M}_{22} \to \mathcal{M}_{22}, T(A) = A^T$; invertibility, idempotency
4. $T\colon\ E^2 \to E^2, T$ is rotation by $180°$; invertibility
5. $T\colon\ E^3 \to E^3, T((x_1, x_2, x_3)) = (x_1, x_1 + x_2, x_1 + x_2 + x_3)$; invertibility, nilpotency
6. $T\colon\ \mathcal{P}_1 \to \mathcal{P}_1, T(a + bx) = b + ax$; invertibility, idempotency
7. Consider an equilateral triangle with center C. Let R be a counterclockwise rotation about C of $120°$. Show that three successive applications of R yield the identity transformation. (Use matrices.)
8. Let T be a counterclockwise rotation about C of $240°$. Show that $T \circ T \circ T = I$.
9. Consider $T\colon\ \mathbb{C}^n \to \mathbb{C}^n$ defined by

   $$T((z_1, z_2, \dots, z_n)) = (0, z_1, z_2, \dots, z_{n-1})$$

   (T is called a **right shift**.) Is T invertible? Is T nilpotent? Is T idempotent?
10. Consider $T\colon\ \mathbb{C}^n \to \mathbb{C}^n$ defined by

    $$T((z_1, z_2, \dots, z_n)) = (z_2, z_3, \dots, z_n, z_1)$$

    Is T invertible? Is T nilpotent? Is T idempotent?
11. If A is an involutory matrix ($A^2 = I$) and B is similar to A, is B involutory?
12. If A is an orthogonal matrix and B is similar to A, is B orthogonal?
13. If A is a symmetric matrix, P is orthogonal, and $B = P^T AP$, is B symmetric?
14. Let us say that a matrix $A_{n\times n}$ is diagonalizable if it is similar to a diagonal matrix D. If B is similar to A, is B similar to a diagonal matrix?

15. Let L be an invertible linear operator from V to V, and let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a linearly independent set in V. Show that the set $\{L(\mathbf{v}_1), L(\mathbf{v}_2), \ldots, L(\mathbf{v}_k)\}$ is linearly independent in the range of L.

4.5 CALCULUS REVISITED

Two of the most important linear transformations in applied mathematics are differentiation and definite integration, which are studied in calculus. The familiar rules[2]

$$\frac{d}{dx}(f(x)+g(x)) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$$
$$\frac{d}{dx}(cf(x)) = c\frac{d}{dx}f(x)$$
$$\int_a^b [f(x)+g(x)]\ dx = \int_a^b f(x)dx + \int_a^b g(x)\ dx$$
$$\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx$$

are among the most useful for calculating derivatives and integrals. These rules are just statements of linearity. In fact, if we denote differentiation by D, the first two rules above are just

$$D(f+g) = D(f) + D(g) \quad \text{and} \quad D(cf) = cD(f)$$

Example 1. Let D be the transformation of differentiation. Show that $D\colon\ \mathcal{P}_3 \to \mathcal{P}_3$ is not invertible. Give $\mathcal{P}_3$ the standard basis $\mathcal{S} = \{1, x, x^2, x^3\}$. Find the matrix of D with respect to $\mathcal{S}$. Show that D is nilpotent.

Solution To show that D is not invertible, we show that $\dim(\ker D) \neq 0$. Now $\ker D$ is the set of polynomials f for which $D(f) = 0$ (the zero function). That is, the kernel is the set of all polynomials which have derivative zero. From calculus we know that this is the set of constant functions. Thus $\ker = D =$ span 1, and $\eta(D) = 1$. Therefore D is not invertible. To find the

[2] Which, of course, come from theorems in which f and g must satisfy certain conditions.

matrix, we calculate

$$\begin{aligned}
D(x^3) &= 3x^2 = 0x^3 + 3x^2 + 0x + 0 \\
D(x^2) &= 2x = 0x^3 + 0x^2 + 2x + 0 \\
D(x) &= 1 = 0x^3 + 0x^2 + 0x + 1 \\
D(1) &= 0 = 0x^3 + 0x^2 + 0x + 0
\end{aligned}$$

to find

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

Notice that M is not invertible, as we would expect since D is not invertible. Now since $M^4 = 0$, matrix M is nilpotent and thus D is nilpotent of order 4. This is just a linear algebra way of saying that if we differentiate a cubic 4 times or more, we obtain the zero function.

Some calculus students would object to the statement "D is not invertible" because "Everyone knows integration and differentiation are just opposite operations." We must be careful since the process of antidifferentiation involves the addition of an arbitrary constant. If we let A denote antidifferentiation, then, for example, $A(2x) = x^3 + c$, where c is an arbitrary constant. If A were to be an inverse to D, we would have $A(D(x^2)) = x^2$; however, $A(D(x^2)) = A(2x) = x^2 + c$, which is equal to x^2 only if $c = 0$. So in general $A \circ D \neq I$: therefore, A and D are not inverses. The next example shows, however, that for some vector spaces antidifferentiation **is** the inverse of differentiation.

Example 2. Let V be the subspace of $\mathcal{P}_3$ defined by $V = \{f(x) \text{ in } \mathcal{P}_3 | f(0) = 0\}$. Show that $\dim V = 3$. Show that the range of D with domain restricted to V is $\mathcal{P}_2$. Define $A_0\colon\ \mathcal{P}_2 \to V$ as antidifferentiation with arbitrary constant zero. Show that $(A_0 \circ D)(f) = f = I(f)$ for any f in $\mathcal{P}_3$.

Solution A natural basis for V is $\mathcal{S} = \{x, x^2, x^3\}$ since any cubic $p(x) = ax^3 + bx^2 + cx + d$ has the property $p(0) = 0$ if and only if $d = 0$. Therefore $\dim V = 3$. Now we calculate the range of D. The range of D is spanned by the images under D of the basis vectors for V; we have

$$\begin{aligned} D(x) &= 1 \\ D(x^2) &= 2x \\ D(x^3) &= 3x^2 \end{aligned}$$

so that range $D = \text{span}\,\{1, 2x, 3x^2\}$. The range of D is all linear combinations of $1, 2x$, and $3x^2$ which is $\mathcal{P}_2$. To show that $A_0 \circ D = I$, we calculate the matrices for A_0 and D, using the natural bases for V and $\mathcal{P}_2$. We find

$$\begin{aligned} D(x) &= 1 + 0x + 0x^2 \\ D(x^2) &= 0 + 2x + 0x^2 \\ D(x^3) &= 0 + 0x + 3x^2 \end{aligned}$$

and

$$\begin{aligned} A_0(1) &= 1x + 0x^2 + 0x^3 \\ A_0(x) &= 0x + \tfrac{1}{2}x^2 + 0x^3 \\ A_0(x_2) &= 0x + 0x^2 + \tfrac{1}{3}x^3 \end{aligned}$$

Therefore

$$M_D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad M_{A_0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \quad M_{A_0}M_D = I$$

Also we see that $M_D M_{A_0} = I$, so that $D \circ A = I$. In this example we have forced the arbitrary constant of antidifferentiation to be zero.

Example 2 shows that general principle that to "invert" differentiation, we need some conditions on the domain of D. In Example 2 the condition was $f(0) = 0$. In general, such conditions can be called **initial conditions**; they arise in problems in the area of mathematics known as **differential equations**. In some textbooks the solving of a differential equation is referred to as **integrating** the equation. This terminology comes up because of the inverse relation we have just seen.

Example 3. Show that the transformation $\mathcal{A}$ defined on $C[0, 1]$ by

$$(\mathcal{A}(f))(x) \equiv \int_{x_0}^{x} f(t)\; dt \quad \text{for all } x \text{ in } [0, 1]$$

is linear.

Solution Let f and g be functions in $C[0,1]$. We have

$$\begin{aligned}(\mathcal{A}(f+g))(x) &\equiv \int_{x_0}^{x} [f(t)+g(t)]\ dt \\ &= \int_{x_0}^{x} f(t)\ dt + \int_{x_0}^{x} g(t)\ dt \\ &= (\mathcal{A}(f))(x) + (\mathcal{A}(g))(x)\end{aligned}$$

and

$$\begin{aligned}(\mathcal{A}(cf))(x) &\equiv \int_{x_0}^{x} cf(t)\ dt \\ &= c\int_{x_0}^{x} f(t)\ dt \\ &= (c\mathcal{A}(f))(x)\end{aligned}$$

and $\mathcal{A}$ is linear. Notice that $(\mathcal{A}(f))(x_0) = 0$ for all f.

The fundamental theorem of calculus may be stated in this form:

> If $f'(x)$ is continuous on [0,1], then for a fixed $x_0 \in [0,1]$ and any $x \in [0,1]$,
> $$\int_{x_0}^{x} f'(t)\ dt = f(x) - f(x_0)$$

Using our notation of D and $\mathcal{A}$, we see that the last equation is

$$\mathcal{A}(D(f))(x) = f(x) - f(x_0)$$

Therefore $\mathcal{A}$ is the inverse of D if and only if $f(x_0) = 0$. Again we see an extra condition to guarantee invertibility of differentiation.

To illustrate this last point, we consider a simple differential equation: Find a function $y(x)$, in $C^1(\mathbb{R})$, which satisfies

$$y'(x) = y(x) \qquad x \in \mathbb{R}$$

Now we know that **any** function of the form $y(x) = Ce^x$ satisfies this equation. Since C is arbitrary, we have not **uniquely solved** the problem. However, if we require further a condition such as $y(0) = 3$, we find

$y(x) = 3e^x$ as the only[3] solution of the problem. Unique solvability of a differential equation and invertibility of the differentiation operator are closely connected.

Linear transformations arise in several-variable calculus also. Recall the gradient: Let $f\colon \mathbb{R}^3 \to \mathbb{R}$ with $w = f(x, y, z)$. We have

$$(\text{grad } f)(x, y, z) = \left(\frac{\partial f}{\partial x}(x, y, z), \frac{\partial f}{\partial y}(x, y, z), \frac{\partial f}{\partial z}(x, y, z)\right)$$

and the properties

$$\begin{aligned} \text{grad } (f + g) &= \text{grad } f + \text{grad } g \\ \text{grad } (cf) &= c(\text{grad } f) \end{aligned}$$

are not hard to see. Therefore the gradient operation is a linear transformation from the space of all real-valued functions of three variables with continuous derivatives to the vector space of ordered triples of continuous functions of three variables.

The **Jacobian** is another linear operator which is studied in several-variable calculus. It is generated by a matrix. If $\mathbf{f}$ is a function defined by

$$\mathbf{f}(x, y) = (g(x, y), h(x, y))$$

then the Jacobian of $\mathbf{f}$ is the matrix of functions

$$J(x, y) = \begin{pmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\ \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{pmatrix}$$

At each point (x_0, y_0), $J(x_0, y_0)$ is a fixed 2×2 matrix. If $J(x_0, y_0)$ is an invertible matrix, then $\mathbf{f}$ is invertible in some neighborhood of (x_0, y_0). In this way the local invertibility of a **nonlinear** function $\mathbf{f}$ is studied by determining the invertibility of an associated **linear** transformation (generated by the Jacobian).

Example 4. Consider the nonlinear transformation $\mathbf{f}\colon\ E^2 \to E^2$ defined by $\mathbf{f}(x, y) = (x \cos y, x \sin y)$. Find the Jacobian of $\mathbf{f}$. Where is $\mathbf{f}$ locally invertible?

[3] This is proved in differential equations courses.

Solution

$$J(x,y) = \begin{pmatrix} \dfrac{\partial(x\cos y)}{\partial x} & \dfrac{\partial(x\cos y)}{\partial y} \\ \dfrac{\partial x(\sin y)}{\partial x} & \dfrac{\partial(x\sin y)}{\partial y} \end{pmatrix}$$
$$= \begin{pmatrix} \cos y & x\sin y \\ \sin y & x\cos y \end{pmatrix}$$

This matrix is invertible if and only if

$$\det J(x,y) \neq 0$$

That is, $x(\cos^2 y + \sin^2 y) = x \neq 0$. Therefore we say $\mathbf{f}$ is **locally invertible** in a neighborhood of any point (x_0, y_0) with $x_0 \neq 0$.

These examples illustrate the fact that the operations of differentiation and integration are the source of many linear transformations in applied mathematics.

PROBLEMS 4.5

1. Let V be the set of all functions $f(x)$ for which $\lim_{x\to a} f(x)$ exists. Show that V, with the usual definitions of addition and scalar multiplication for functions, is a vector space. How does the vector space structure depend on the linearity of the transformation $L_a\colon\ V \to \mathbb{R}$ defined by $L_a(f) = \lim_{x\to a} f(x)$?

2. Let f be a fixed function in $C[0,\pi]$. Define $L_f\colon\ C[0,\pi] \to \mathbb{R}$ by $L_f(g) = \int_0^\pi f(x)g(x)\,dx$. Show that L_f is a linear operator.

3. Consider the linear operator in Prob. 2. The kernel of L_f is the set of all functions g orthogonal to f [where the dot product is $\langle f, g\rangle = \int_0^\pi f(x)g(x)\,dx$]. Show that if $f(x) = \sin x$, then $g(x) = \sin nx$, $n \neq 1, n$ a positive integer, is in $\ker L_f$.

4. Let $\mathbf{f}(x,y) = (x^2, y^2)$ be a nonlinear operator mapping E^2 into E^2. Use the Jacobian to determine the local invertibility of $\mathbf{f}$.

5. Show that $D^2\colon\ \mathcal{P}_3 \to \mathcal{P}_3$ defined by $D^2(p(x)) = p''(x)$ is a linear transformation. Find the standard matrix of D^2. Show that this matrix is the square of the matrix for D found in Example 1.

6. Let V be the set of all functions $f\colon\ \mathbb{R} \to \mathbb{R}$ with Maclaurin series convergent to the function for all x. Let

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots \\ g(x) &= b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n + \cdots \end{aligned}$$

and define

$$\begin{aligned} (f+g)(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n + \cdots \\ (cf)(x) &= ca_0 + ca_1 x + ca_2 x^2 + \cdots + ca_n x^n + \cdots \end{aligned}$$

Show that V with these operations is a vector space (recall theorems about convergent power series).

7. Consider the vector space V from Prob. 6. From calculus we know that if $f \in V$, and f is differentiable on $\mathbb{R}$ then

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + na_n x^{n-1} + \cdots$$

Therefor, $D\colon\ V \to V$. What would be an inverse operator to D?

8. Regarding Prob. 7, show how we might consider the "infinite matrix"

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 2 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 3 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & & & \\ \vdots & \vdots & \vdots & \vdots & & n & 0 & \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & & & \cdots \end{pmatrix}$$

to be the matrix of D. [**Hint**: What would be a good choice for a "basis" of V?]

9. Recall that for $\mathbf{f}\colon\ E^3 \to E^3$ the curl of $\mathbf{f}$ can be defined, when $\mathbf{f}(x, y, z) = \langle f_1(x, y, z), f_2(x, y, z), f_3(x, y, z)\rangle$, as

$$\operatorname{curl} \mathbf{f} = \text{“det”} \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ f_1 & f_2 & f_3 \end{pmatrix}$$

Show that curl is a linear transformation from the space of ordered triples of continuously differentiable functions of three variables to the space of ordered triples of continuous functions of three variables.

10. For a function $\mathbf{f}\colon\ E^3 \to E^3$ given by

$$\mathbf{f}(x, y, z) = (f_1(x, y, z), f_2(x, y, z), f_3(x, y, z))$$

the divergence of f is defined by

$$\operatorname{div}\ \mathbf{f} = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}$$

Show that div is a linear transformation from the space of ordered triples of continuously differentiable functions of three variables to the space of real-valued continuous functions of three variables.

11. Let V be the space of all real-valued functions $f\colon\ [0, 1] \to \mathbb{R}$, along with the standard operations of addition and scalar multiplication.

(**a**) Show that $T\colon\ V \to V$ defined by

$$T(f)(x) = xf(x) \qquad \text{for all } x \text{ in } [0, 1]$$

is linear.

(**b**) Show that g defined by

$$g(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$$

is in $\ker T$.

(**c**) Is T invertible?

12. Let $V = C[0, 1]$, and define T as in Prob. 11**a**. Is T invertible with this domain? (Check the kernel.)

SUMMARY

Linear transformations, central to applied mathematics, were defined and then analyzed by considering the **kernel and range** of the transformation as well as the **matrix representing the transformation**. To find the matrix representing a transformation is the **third basic problem of linear algebra**.

Linear transformations from a vector space to itself **can be classified** in a definite way because certain properties of matrices are invariant under similarity (invertibility, nilpotency, and idempotency are three) and **all representing matrices of a transformation from** $(V, \mathcal{S})$ **to** $(V, \mathcal{S})$ **are similar**. The similarity transform P in these cases is simply a transition matrix, as defined in Chap. 3.

Having the ideas of matrix algebra, linear transformations, and vector spaces in our repertoire, we are ready to tackle two extremely important problems of linear algebra: the eigenvalue-eigenvector problem and the diagonalization problem. In Chap. 5, determinants, inverses, systems of homogeneous equations, vector spaces, bases, dimension, linear dependence, linear independence, similarity, and orthogonality are all used to solve these problems. Thus the results from the first four chapters will at last come together to solve important theoretical and practical problems.

ADDITIONAL PROBLEMS

1. Define $T\colon\ \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ by $T(A) = A - A^T$. Show that T is linear. Describe the kernel of T.

2. Define $L\colon\ \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ by $L(A) = A + A^T$. Show that L is linear. Describe the kernel of L.

3. Compare the dimension of the vector space of $n \times n$ symmetric matrices and the dimension of the vector space of $n \times n$ upper triangular matrices.

4. Let A be an $n \times n$ invertible matrix. Define $T\colon\ \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ by $T(B) = A^{-1}BA$. Is T linear? Is T one-to-one?

5. Let $P \neq 0$ be a matrix with $P^2 = P$, and define $T\colon\ \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ by $T(A) = PA$. Show that T is linear. Is T one-to-one?

6. Let c be in $\mathbb{C}$ and A be in $\mathcal{C}_{nn}$. Define $T\colon\ \mathcal{C}_{n1} \to \mathcal{C}_{n1}$ by $T(X) = AX - cIX$. Show that T is linear.

7. Let A be a square matrix. Consider the matrix $B = I + A$. Multiply $B[I - A + A^2 - A^3 + \cdots + (-1)^n A^n]$ for different values of n. If A is

nilpotent of exponent k, show how the sum in the brackets can be used to compute $(I+A)^{-1}$.

8. Let A be a square matrix such that A^n tends to the zero matrix as n increases without bound. How can you "construct" $(I+A)^{-1}$ by using a sum such as in Prob. 7?

9. Describe the linear transformation $L \circ T$, where L and T come from Probs. 1 and 2. Is $L \circ T = T \circ L$?

10. Define $N\colon \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ by $N(A) = A^T$. Using L and T from Probs. 1 and 2, calculate $T \circ N$ and $L \circ N$. Compare $T \circ N$ and $N \circ T$. Compare $L \circ N$ and $N \circ L$.

11. Linear transformations can be indexed by a variable. For example, if a particle is rotating about the origin of the plane with constant angular velocity ω, the position of the particle is defined by

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \cos\omega t & -\sin\omega t \\ \sin\omega t & \cos\omega t \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

where

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

is the initial position vector. Notice that the 2×2 matrix $A(t)$ in the equation is a matrix function. Show that for all $t, A(t)$ is invertible. Show that for all $t, A(t)$ is orthogonal.

12. The matrix

$$A = \begin{pmatrix} b_1 & b_2 & b_3 & & \cdots & b_n \\ 1-d_1 & 0 & 0 & & \cdots & 0 \\ 0 & 1-d_2 & 0 & & \cdots & 0 \\ 0 & 0 & 1-d_3 & & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1-d_{n-1} & 0 \end{pmatrix}$$

is used in difference equations for determining age distributions in populations; that is, determining the numbers of individuals in different age brackets. The matrix represents a linear transformation from the state of age distributions at one observation time to the next. The b_k are birth rates, and the d_k are death rates for the kth age bracket. Calculate $\det A$ for various values of n. Can you generate a formula for $\det A$ for arbitrary n?

13. In mechanics, when a body is subjected to forces, some change in relative positions of particles in the body may result. For example, when a rod is bent, the relative positions of points on the curved surface change. This change in relative position is known as **strain**. The term **homogeneous strain** refers to strain in which the new coordinates of a material point, given by Y in $\mathcal{M}_{31}$, are related to the old coordinates, given by X in $\mathcal{M}_{31}$, by $Y = AX$, where A is a 3×3 real matrix. Show that for homogeneous strain straight lines remain straight.

14. Show that in the case of homogeneous strain, parallel straight lines remain parallel.

15. Newton's method for solving $f(x) = 0$ can be extended to the problem of solving

$$f(x, y) = 0 \qquad g(x, y) = 0$$

Recall that the iterative step for Newton's method is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

For the case of the two equations involving $f(x, y)$ and $g(x, y)$, we have

$$\begin{pmatrix} x_{n+1} \\ \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ \\ y_n \end{pmatrix} - \begin{pmatrix} \dfrac{\partial f(x_n, y_n)}{\partial x} & \dfrac{\partial f(x_n, y_n)}{\partial y} \\ \dfrac{\partial g(x_n, y_n)}{\partial x} & \dfrac{\partial g(x_n, y_n)}{\partial y} \end{pmatrix}^{-1} \begin{pmatrix} f(x_n, y_n) \\ \\ g(x_n, y_n) \end{pmatrix}$$

Apply this to

$$\begin{aligned} x^2 + y^2 - 2 &= 0 \\ x - y &= 0 \end{aligned}$$

and start with

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Complete three steps of the iteration. Actual solutions are

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$