Chapter 4

Linear Transformations

That confusions of thought and errors of reasoning
still darken the beginnings of Algebra,
is the earnest and just complaint of sober and thoughtful men.

Sir William Rowan Hamilton

1. THE DIMENSION FORMULA

The analogue for vector spaces of a homomorphism of groups is a map

$$T: V \longrightarrow W$$

from one vector space over a field F to another, which is compatible with addition and scalar multiplication:

$$(1.1) \qquad T(v_1 + v_2) = T(v_1) + T(v_2) \quad \text{and} \quad T(cv) = cT(v),$$

for all v_1, v_2 in V and all $c \in F$. It is customary to call such a map a *linear transformation*, rather than a homomorphism. However, use of the word *homomorphism* would be correct too. Note that a linear transformation is compatible with linear combinations:

$$(1.2) \qquad T\Big(\sum_i c_i v_i\Big) = \sum_i c_i T(v_i).$$

This follows from (1.1) by induction. Note also that the first of the conditions of (1.1) says that T is a homomorphism of additive groups $V^+ \longrightarrow W^+$.

We already know one important example of a linear transformation, which is in fact the main example: left multiplication by a matrix. Let A be an $m \times n$ matrix with entries in F, and consider A as an operator on column vectors. It defines a linear transformation

$$(1.3) \qquad \begin{aligned} F^n &\xrightarrow{\text{left mult. by } A} F^m \\ X &\rightsquigarrow AX. \end{aligned}$$

Indeed, $A(X_1 + X_2) = AX_1 + AX_2$, and $A(cX) = cAX$.

Another example: Let P_n be the vector space of real polynomial functions of degree $\leq n$, of the form

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0. \tag{1.4}$$

The derivative $\frac{d}{dx}$ is a linear transformation from P_n to P_{n-1}.

Let $T\colon V \longrightarrow W$ be any linear transformation. We introduce two subspaces

$$\ker T = \textit{kernel of } T = \{v \in V \mid T(v) = 0\} \tag{1.5}$$

$$\operatorname{im} T = \textit{image of } T = \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

As one may guess from the similar case of group homomorphisms (Chapter 2, Section 4), $\ker T$ is a subspace of V and $\operatorname{im} T$ is a subspace of W.

It is interesting to interpret the kernel and image in the case that T is left multiplication by a matrix A. In that case the kernel T is the set of solutions of the homogeneous linear equation $AX = 0$. The image of T is the set of vectors $B \in F^m$ such that the linear equation $AX = B$ has a solution.

The main result of this section is the *dimension formula*, given in the next theorem.

(1.6) Theorem Let $T\colon V \longrightarrow W$ be a linear transformation, and assume that V is finite-dimensional. Then

$$\dim V = \dim(\ker T) + \dim(\operatorname{im} T).$$

The dimensions of $\operatorname{im} T$ and $\ker T$ are called the *rank* and *nullity* of T, respectively. Thus (1.6) reads

$$\dim V = \text{rank} + \text{nullity}. \tag{1.7}$$

Note the analogy with the formula $|G| = |\ker \varphi|\,|\operatorname{im} \varphi|$ for homomorphisms of groups [Chapter 2 (6.15)].

The *rank* and *nullity* of an $m \times n$ matrix A are defined to be the dimensions of the image and kernel of left multiplication by A. Let us denote the rank by r and the nullity by k. Then k is the dimension of the space of solutions of the equation $AX = 0$. The vectors B such that the linear equation $AX = B$ has a solution form the image, a space whose dimension is r. The sum of these two dimensions is n.

Let B be a vector in the image of multiplication by A, so that the equation $AX = B$ has at least one solution $X = X_0$. Let K denote the space of solutions of the homogeneous equation $AX = 0$, the kernel of multiplication by A. Then the set of solutions of $AX = B$ is the additive coset $X_0 + K$. This restates a familiar fact: Adding any solution of the homogeneous equation $AX = 0$ to a particular solution X_0 of the inhomogeneous equation $AX = B$, we obtain another solution of the inhomogeneous equation.

Suppose that A is a square $n \times n$ matrix. If $\det A \neq 0$, then, as we know, the system of equations $AX = B$ has a unique solution for every B, because A is invertible. In this case, $k = 0$ and $r = n$. On the other hand, if $\det A = 0$ then the space K has dimension $k > 0$. By the dimension formula, $r < n$, which implies that the image is not the whole space F^n. This means that not all equations $AX = B$ have solutions. But those that do have solutions have more than one, because the set of solutions of $AX = B$ is a coset of K.

Proof of Theorem (1.6). Say that $\dim V = n$. Let $(u_1,\ldots,u_k)$ be a basis for the subspace $\ker T$, and extend it to a basis of V [Chapter 3 (3.15)]:

(1.8)
$$(u_1,\ldots,u_k; v_1,\ldots,v_{n-k}).$$

Let $w_i = T(v_i)$ for $i = 1,\ldots,n-k$. If we prove that $(w_1,\ldots,w_{n-k}) = S$ is a basis for $\operatorname{im} T$, then it will follow that $\operatorname{im} T$ has dimension $n-k$. This will prove the theorem.

So we must show that S spans $\operatorname{im} T$ and that it is a linearly independent set. Let $w \in \operatorname{im} T$ be arbitrary. Then $w = T(v)$ for some $v \in V$. We write v in terms of the basis (1.8):

$$v = a_1u_1 + \cdots + a_ku_k + b_1v_1 + \cdots + b_{n-k}v_{n-k},$$

and apply T, noting that $T(u_i) = 0$:

$$w = 0 + \cdots + 0 + b_1w_1 + \cdots + b_{n-k}w_{n-k}.$$

Thus w is in the span of S, and so S spans $\operatorname{im} T$.

Next, suppose a linear relation

(1.9)
$$c_1w_1 + \cdots + c_{n-k}w_{n-k} = 0$$

is given, and consider the linear combination $v = c_1v_1 + \cdots + c_{n-k}v_{n-k}$, where v_i are the vectors (1.8). Applying T to v gives

$$T(v) = c_1w_1 + \cdots + c_{n-k}w_{n-k} = 0.$$

Thus $v \in \ker T$. So we may write v in terms of the basis $(u_1,\ldots,u_k)$ of $\ker T$, say $v = a_1u_1 + \cdots + a_ku_k$. Then

$$-a_1u_1 + \cdots + -a_ku_k + c_1v_1 + \cdots + c_{n-k}v_{n-k} = 0.$$

But (1.8) is a basis. So $-a_1 = 0,\ldots,-a_k = 0$, and $c_1 = 0,\ldots,c_{n-k} = 0$. Therefore the relation (1.9) was trivial. This shows that S is linearly independent and completes the proof.

2. THE MATRIX OF A LINEAR TRANSFORMATION

It is not hard to show that every linear transformation $T: F^n \longrightarrow F^m$ is left multiplication by some $m \times n$ matrix A. To see this, consider the images $T(e_j)$ of the standard basis vectors e_j of F^n. We label the entries of these vectors as follows:

(2.1)
$$T(e_j) = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

and we form the $m \times n$ matrix $A = (a_{ij})$ having these vectors as its columns. We can write an arbitrary vector $X = (x_1, \dots, x_n)^t$ from F^n in the form $X = e_1x_1 + \cdots + e_nx_n$, putting scalars on the right. Then

$$T(X) = \sum_j T(e_j)x_j = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = AX.$$

For example, the linear transformation $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ such that

$$T(e_1) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad T(e_2) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

is left multiplication by the matrix

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

If $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = e_1x_1 + e_2x_2$, then

$$T(X) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} -1 \\ 0 \end{bmatrix} x_2 = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ 2x_1 \end{bmatrix}.$$

Using the notation established in Section 4 of Chapter 3, we can make a similar computation with an arbitrary linear transformation $T: V \longrightarrow W$, once bases of the two spaces are given. Let $\mathbf{B} = (v_1, \dots, v_n)$ and $\mathbf{C} = (w_1, \dots, w_m)$ be bases of V and of W, and let us use the shorthand notation $T(\mathbf{B})$ to denote the hypervector

$$T(\mathbf{B}) = (T(v_1), \dots, T(v_n)).$$

Since the entries of this hypervector are in the vector space W, and since $\mathbf{C}$ is a basis for that space, there is an $m \times n$ matrix A such that

$$T(\mathbf{B}) = \mathbf{C}A \quad \text{or} \quad (T(v_1), \dots, T(v_n)) = (w_1, \dots, w_m)\begin{bmatrix} & A & \end{bmatrix} \tag{2.2}$$

[Chapter 3 (4.13)]. Remember, this means that for each j,

$$T(v_j) = \sum_i w_i a_{ij} = w_1 a_{1j} + \cdots + w_m a_{mj}. \tag{2.3}$$

So A is the matrix whose jth column is the coordinate vector of $T(v_j)$. This $m \times n$ matrix $A = (a_{ij})$ is called the *matrix of T with respect to the bases* $\mathbf{B}, \mathbf{C}$. Different choices of the bases lead to different matrices.

In the case that $V = F^n$, $W = F^m$, and the two bases are the standard bases, A is the matrix constructed as in (2.1).

The matrix of a linear transformation can be used to compute the coordinates of the image vector $T(v)$ in terms of the coordinates of v. To do this, we write v in terms of the basis, say

$$v = \mathbf{B}X = v_1x_1 + \cdots + v_nx_n.$$

Then

$$T(v) = T(v_1)x_1 + \cdots + T(v_n)x_n = T(\mathbf{B})X = \mathbf{C}AX.$$

Therefore the coordinate vector of $T(v)$ is

$$Y = AX,$$

meaning that $T(v) = \mathbf{C}Y$. Recapitulating, the matrix A of the linear transformation has two dual properties:

(2.4)
$$T(\mathbf{B}) = \mathbf{C}A \quad \text{and} \quad Y = AX.$$

The relationship between T and A can be explained in terms of the isomorphisms $\psi: F^n \longrightarrow V$ and $\psi': F^m \longrightarrow W$ determined by the two bases [Chapter 3 (4.14)]. If we use ψ and ψ' to identify V and W with F^n and F^m, then T corresponds to left multiplication by A.

$$\begin{tikzcd} V \arrow[r, "T"] & W \\ F^n \arrow[u, "\psi"] \arrow[r, "\text{mult by } A"'] & F^m \arrow[u, "\psi'"'] \end{tikzcd}, \qquad \begin{tikzcd} \mathbf{B}X \arrow[r, squiggly] & \mathbf{C}AX \\ X \arrow[u, squiggly] \arrow[r, squiggly] & AX \arrow[u, squiggly] \end{tikzcd}$$

Going around this square in the two directions gives the same answer: $T \circ \psi = \psi' \circ A$.

Thus any linear transformation between finite-dimensional vector spaces V and W can be identified with matrix multiplication, once bases for the two spaces are chosen. But if we study changes of basis in V and W, we can do much better. Let us ask how the matrix A changes when we make other choices of bases for V and W. Let $\mathbf{B}' = (v_1', \ldots, v_n')$, $\mathbf{C}' = (w_1', \ldots, w_m')$ be new bases for these spaces. We can relate the new basis $\mathbf{B}'$ to the old basis $\mathbf{B}$ by a matrix $P \in GL_n(F)$, as in Chapter 3 (4.19). Similarly, $\mathbf{C}'$ is related to $\mathbf{C}$ by a matrix $Q \in GL_m(F)$. These matrices have the following properties:

Find from here

(2.6)
$$PX = X' \quad \text{and} \quad QY = Y'.$$

Here X and X' denote the coordinate vectors of a vector $v \in V$ with respect to the bases $\mathbf{B}$ and $\mathbf{B}'$, and similarly Y and Y' denote the coordinate vectors of a vector $w \in W$ with respect to $\mathbf{C}$ and $\mathbf{C}'$.

Let A' denote the matrix of T with respect to the new bases, defined as above (2.4), so that $A'X' = Y'$. Then $QAP^{-1}X' = QAX = QY = Y'$. Therefore

(2.7)
$$A' = QAP^{-1}.$$

Note that P and Q are arbitrary invertible $n \times n$ and $m \times m$ matrices [Chapter 3 (4.23)]. Hence we obtain the following description of the matrices of a given linear transformation:

(2.8) Proposition. Let A be the matrix of a linear transformation T with respect to some given bases $\mathbf{B}, \mathbf{C}$. The matrices A' which represent T with respect to other bases are those of the form

$$A' = QAP^{-1},$$

where $Q \in GL_m(F)$ and $P \in GL_n(F)$ are arbitrary invertible matrices. □

Now given a linear transformation $T: V \longrightarrow W$, it is natural to look for bases $\mathbf{B}, \mathbf{C}$ of V and W such that the matrix of T becomes especially nice. In fact the matrix can be simplified remarkably.

(2.9) Proposition.

(a) *Vector space form:* Let $T: V \longrightarrow W$ be a linear transformation. Bases $\mathbf{B}, \mathbf{C}$ can be chosen so that the matrix of T takes the form

(2.10)
$$A = \left[\begin{array}{c|c} I_r & \\ \hline & 0 \end{array}\right],$$

where I_r is the $r \times r$ identity matrix, and $r = \operatorname{rank} T$.

(b) *Matrix form:* Given any $m \times n$ matrix A, there are matrices $Q \in GL_m(F)$ and $P \in GL_n(F)$ so that QAP^{-1} has the form (2.10).

It follows from our discussion that these two assertions amount to the same thing. To derive (a) from (b), choose arbitrary bases $\mathbf{B}, \mathbf{C}$ to start with, and let A be the matrix of T with respect to these bases. Applying (b), we can find P, Q so that QAP^{-1} has the required form. Let $\mathbf{B}' = \mathbf{B}P^{-1}$ and $\mathbf{C}' = \mathbf{C}Q^{-1}$ be the new bases, as in Chapter 3 (4.22). Then the matrix of T with respect to the bases $\mathbf{B}', \mathbf{C}'$ is QAP^{-1}. So these new bases are the required ones. Conversely, to derive (b) from (a) we view an arbitrary matrix A as the matrix of the linear transformation "left multiplication by A", with respect to the standard bases. Then (a) and (2.7) guarantee the existence of P, Q so that QAP^{-1} has the required form.

Note that we can interpret QAP^{-1} as the matrix obtained from A by a succession of row and column operations: We write P and Q as products of elementary matrices: $P = E_p \cdots E_1$ and $Q = E_q' \cdots E_1'$ [Chapter 1 (2.18)]. Then $QAP^{-1} = E_q' \cdots E_1' A E_1^{-1} \cdots E_p^{-1}$. Because of the associative law, it does not matter whether the row operations or the column operations are done first. The equation $(E'A)E = E'(AE)$ tells us that row operations commute with column operations.

It is not hard to prove (2.9b) by matrix manipulation, but let us prove (2.9a) using bases instead. Let $(u_1, \dots, u_k)$ be a basis for $\ker T$. Extend to a basis $\mathbf{B}$ for V: $(v_1, \dots, v_r; u_1, \dots, u_k)$, where $r + k = n$. Let $w_i = T(v_i)$. Then, as in the proof of (1.6), $(w_1, \dots, w_r)$ is a basis for $\operatorname{im} T$. Extend to a basis $\mathbf{C}$ of W: $(w_1, \dots, w_r; x_1, \dots, x_s)$. The matrix of T with respect to these bases has the required form. □

I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case.

Arthur Cayley

EXERCISES

1. The Dimension Formula

1. Let T be left multiplication by the matrix $\begin{bmatrix} 1 & 2 & 0 & -1 & 5 \\ 2 & 0 & 2 & 0 & 1 \\ 1 & 1 & -1 & 3 & 2 \\ 0 & 3 & -3 & 2 & 6 \end{bmatrix}$. Compute $\ker T$ and $\operatorname{im} T$ explicitly by exhibiting bases for these spaces, and verify (1.7).
2. Determine the rank of the matrix $\begin{bmatrix} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{bmatrix}$.
3. Let $T: V \longrightarrow W$ be a linear transformation. Prove that $\ker T$ is a subspace of V and that $\operatorname{im} T$ is a subspace of W.
4. Let A be an $m \times n$ matrix. Prove that the space of solutions of the linear system $AX = 0$ has dimension at least $n - m$.
5. Let A be a $k \times m$ matrix and let B be an $n \times p$ matrix. Prove that the rule $M \rightsquigarrow AMB$ defines a linear transformation from the space $F^{m \times n}$ of $m \times n$ matrices to the space $F^{k \times p}$.
6. Let $(v_1, \ldots, v_n)$ be a subset of a vector space V. Prove that the map $\varphi: F^n \longrightarrow V$ defined by $\varphi(X) = v_1x_1 + \cdots + v_nx_n$ is a linear transformation.
7. When the field is one of the fields $\mathbb{F}_p$, finite-dimensional vector spaces have finitely many elements. In this case, formula (1.6) and formula (6.15) from Chapter 2 both apply. Reconcile them.
8. Prove that every $m \times n$ matrix A of rank 1 has the form $A = XY^t$, where X,Y are m- and n-dimensional column vectors.
9. (a) The *left shift* operator S^- on $V = \mathbb{R}^\infty$ is defined by $(a_1, a_2, \ldots,) \rightsquigarrow (a_2, a_3, \ldots)$. Prove that $\ker S^- > 0$, but $\operatorname{im} S^- = V$.
   (b) The *right shift* operator S^+ on $V = \mathbb{R}^\infty$ is defined by $(a_1, a_2, \ldots) \rightsquigarrow (0, a_1, a_2, \ldots)$. Prove that $\ker S^+ = 0$, but $\operatorname{im} S^+ < V$.

2. The Matrix of a Linear Transformation

1. Determine the matrix of the differentiation operator $\frac{d}{dx}: P_n \longrightarrow P_{n-1}$ with respect to the natural bases (see (1.4)).
2. Find all linear transformations $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ which carry the line $y = x$ to the line $y = 3x$.
3. Prove Proposition (2.9b) using row and column operations.

4. Let $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be the linear transformation defined by the rule $T(x_1, x_2, x_3)^t = (x_1 + x_2, 2x_3 - x_1)^t$. What is the matrix of T with respect to the standard bases?
5. Let A be an $n \times n$ matrix, and let $V = F^n$ denote the space of row vectors. What is the matrix of the linear operator "right multiplication by A" with respect to the standard basis of V?
6. Prove that different matrices define different linear transformations.
7. Describe left multiplication and right multiplication by the matrix (2.10), and prove that the rank of this matrix is r.
8. Prove that A and A^t have the same rank.
9. Let T_1, T_2 be linear transformations from V to W. Define $T_1 + T_2$ and cT by the rules $[T_1 + T_2](v) = T_1(v) + T_2(v)$ and $[cT](v) = cT(v)$.
   (a) Prove that $T_1 + T_2$ and cT_1 are linear transformations, and describe their matrices in terms of the matrices for T_1, T_2.
   (b) Let L be the set of all linear transformations from V to W. Prove that these laws make L into a vector space, and compute its dimension.

3. Linear Operators and Eigenvectors

1. Let V be the vector space of real 2×2 symmetric matrices $X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$, and let $A = \begin{bmatrix} 2 & 1 \\ & 1 \end{bmatrix}$. Determine the matrix of the linear operator on V defined by $X \rightsquigarrow AXA^t$, with respect to a suitable basis.
2. Let $A = (a_{ij})$, $B = (b_{ij})$ be 2×2 matrices, and consider the operator $T: M \rightsquigarrow AMB$ on the space $F^{2\times 2}$ of 2×2 matrices. Find the matrix of T with respect to the basis $(e_{11}, e_{12}, e_{21}, e_{22})$ of $F^{2\times 2}$.
3. Let $T: V \longrightarrow V$ be a linear operator on a vector space of dimension 2. Assume that T is not multiplication by a scalar. Prove that there is a vector $v \in V$ such that $(v, T(v))$ is a basis of V, and describe the matrix of T with respect to that basis.
4. Let T be a linear operator on a vector space V, and let $c \in F$. Let W be the set of eigenvectors of T with eigenvalue c, together with 0. Prove that W is a T-invariant subspace.
5. Find all invariant subspaces of the real linear operator whose matrix is as follows.
   (a) $\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & & \\ & 2 & \\ & & 3 \end{bmatrix}$
6. An operator on a vector space V is called *nilpotent* if $T^k = 0$ for some k. Let T be a nilpotent operator, and let $W^i = \text{im } T^i$.
   (a) Prove that if $W^i \neq 0$, then $\dim W^{i+1} < \dim W^i$.
   (b) Prove that if V is a space of dimension n and if T is nilpotent, then $T^n = 0$.
7. Let T be a linear operator on $\mathbb{R}^2$. Prove that if T carries a line ℓ to ℓ, then it also carries every line parallel to l to another line parallel to l.
8. Prove that the composition $T_1 \circ T_2$ of linear operators on a vector space is a linear operator, and compute its matrix in terms of the matrices A_1, A_2 of T_1, T_2.
9. Let P be the real vector space of polynomials $p(x) = a_0 + a + \cdots + a_n x^n$ of degree $\leq n$, and let D denote the derivative $\dfrac{d}{dx}$, considered as a linear operator on P.