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

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) $$T\left(\sum_i c_i v_i\right) = \sum_i c_i T(v_i).$$

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) $$F^n \xrightarrow{\text{left mult. by } A} F^m$$
$$X \rightsquigarrow AX.$$

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

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

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

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

(1.5) $$\ker T = \textit{kernel} \text{ of } T = \{v \in V \mid T(v) = 0\}$$
$$\operatorname{im} T = \textit{image} \text{ of } T = \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

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 $T: V \longrightarrow W$ 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

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

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 $(u_1, \dots, u_k)$ be a basis for the subspace $\ker T$, and extend it to a basis of V [Chapter 3 (3.15)]:

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

Let $w_i = T(v_i)$ for $i = 1, \dots, n - k$. If we prove that $(w_1, \dots, w_{n-k}) = S$ 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

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

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 $(u_1, \dots, u_k)$ 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 $-a_1 = 0, \dots, -a_k = 0$, and $c_1 = 0, \dots, c_{n-k} = 0$. 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:

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

and we form the $m \times n$ matrix $A = (a_{ij})$ having these vectors as its columns. We can write an arbitrary vector $X = (x_1, \ldots, x_n)^t$ 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 $\mathbf{B} = (v_1, \ldots, v_n)$ and $\mathbf{C} = (w_1, \ldots, w_m)$ be bases of V and of W, and let us use the shorthand notation $T(\mathbf{B})$ to denote the hypervector

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

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

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

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

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

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:

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

The relationship between T and A can be explained in terms of the isomorphisms ψ: $F^n \longrightarrow V$ and ψ': $F^m \longrightarrow W$ 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:

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

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

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

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. □

Proposition (2.9) is the prototype for a number of results which will be proved later. It shows the power of working in vector spaces without fixed bases (or coordinates), because the structure of an arbitrary linear transformation is related to the very simple matrix (2.10). It also tells us something remarkable about matrix multiplication, because left multiplication by A on F^m is a linear transformation. Namely, it says that left multiplication by A is the same as left multiplication by a matrix of the form (2.10), but with reference to different coordinate systems. Since multiplication by the matrix (2.10) is easy to describe, we have learned something new.

3. LINEAR OPERATORS AND EIGENVECTORS

Let us now consider the case of a linear transformation $T: V \longrightarrow V$ of a vector space to itself. Such a linear transformation is called a *linear operator* on V. Left multiplication by an $n \times n$ matrix with entries in F defines a linear operator on the space F^n of column vectors.

For example, a rotation ρ_θ of the plane through an angle θ is a linear operator on $\mathbb{R}^2$, whose matrix with respect to the standard basis is

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3.1}$$

To verify that this matrix represents a rotation, we write a vector $X \in \mathbb{R}^2$ in polar coordinates, as $X = (r, \alpha)$. Then in rectangular coordinates, $X = \begin{bmatrix} r\cos\alpha \\ r\sin\alpha \end{bmatrix}$. The addition formulas for sine and cosine show that $RX = \begin{bmatrix} r\cos(\alpha+\theta) \\ r\sin(\alpha+\theta) \end{bmatrix}$. So in polar coordinates, $RX = (r, \alpha + \theta)$. This shows that RX is obtained from X by rotation through the angle θ.

The discussion of the previous section must be changed slightly when we are dealing with linear operators. It is clear that we want to pick only one basis $\mathbf{B} = (v_1, \ldots, v_n)$ for V, and use it in place of both of the bases $\mathbf{B}$ and $\mathbf{C}$ considered in Section 2. In other words, we want to write

$$T(\mathbf{B}) = \mathbf{B}A \tag{3.2}$$

or

$$T(v_j) = \sum_i v_i a_{ij} = v_1 a_{1j} + \cdots + v_n a_{nj}.$$

This defines the matrix $A = (a_{ij})$ of T. It is a square matrix whose jth column is the coordinate vector of $T(v_j)$ with respect to the basis $\mathbf{B}$. Formula (2.4) is unchanged, provided that W and $\mathbf{C}$ are replaced by V and $\mathbf{B}$. As in the previous section, if X and Y denote the coordinate vectors of v and $T(v)$ respectively, then

$$Y = AX. \tag{3.3}$$

The new feature arises when we study the effect of a change of basis on V. Suppose that $\mathbf{B}$ is replaced by a new basis $\mathbf{B}' = (v_1', \dots, v_n')$. Then formula (2.7) shows that the new matrix A' has the form

$$A' = PAP^{-1}, \tag{3.4}$$

where P is the matrix of change of basis. Thus the rule for change of basis in a linear transformation gets replaced by the following rule:

(3.5) **Proposition.** Let A be the matrix of a linear operator T with respect to a basis $\mathbf{B}$. The matrices A' which represent T for different bases are those of the form

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

for arbitrary $P \in GL_n(F)$. □

In general, we say that a square matrix A is *similar* to A' if $A' = PAP^{-1}$ for some $P \in GL_n(F)$. We could also use the word *conjugate* [see Chapter 2 (3.4)].

Now given A, it is natural to ask for a similar matrix A' which is particularly simple. One may hope to get a result somewhat like (2.10). But here our allowable change is much more restricted, because we have only one basis, and therefore one matrix P, to work with.

We can get some insight into the problem by writing the hypothetical matrix P as a product of elementary matrices: $P = E_r \cdots E_1$. Then

$$PAP^{-1} = E_r \cdots E_1 A E_1^{-1} \cdots E_r^{-1}.$$

In terms of elementary operations, we are allowed to change A by a sequence of steps $A \rightsquigarrow EAE^{-1}$. In other words, we may perform an arbitrary row operation E, but then we must also make the inverse column operation E^{-1}. Unfortunately, the row and column operations interfere with each other, and this makes the direct analysis of such operations confusing. I don't know how to use them. It is remarkable that a great deal can be done by another method.

The main tools for analyzing linear operators are the concepts of eigenvector and invariant subspace.

Let $T: V \longrightarrow V$ be a linear operator on a vector space. A subspace W of V is called an *invariant subspace* or a *T-invariant subspace* if it is carried to itself by the operator:

$$TW \subset W. \tag{3.6}$$

In other words, W is T-invariant if $T(w) \in W$ for all $w \in W$. When this is so, T defines a linear operator on W, called the *restriction* of T to W.

Let W be a T-invariant subspace, and let us choose a basis $\mathbf{B}$ of V by appending some vectors to a basis $(w_1, \dots, w_k)$ of W:

$$\mathbf{B} = (w_1, \dots, w_k, v_1, \dots, v_{n-k}).$$

Then the fact that W is invariant can be read off from the matrix M of T. For, the columns of this matrix are the coordinate vectors of the image vectors [see (2.3)], and $T(w_j)$ is in the subspace W, so it is a linear combination of the basis $(w_1,\dots,w_k)$. So when we write $T(w_j)$ in terms of the basis $\mathbf{B}$, the coefficients of the vectors $v_1,\dots,v_{n-k}$ are zero. It follows that M has the block form

$$(3.7)\qquad M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix},$$

where A is a $k \times k$ matrix. Moreover, A is the matrix of the restriction of T to W.

Suppose that $V = W_1 \oplus W_2$ is the direct sum of two T-invariant subspaces, and let $\mathbf{B}_i$ be a basis of W_i. Then we can make a basis $\mathbf{B}$ of V by listing the elements of $\mathbf{B}_1$ and $\mathbf{B}_2$ in succession [Chapter 3 (6.6a)]. In this case the matrix of T will have the block diagonal form

$$(3.8)\qquad M = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$$

where A_i is the matrix of T restricted to W_i.

The concept of an eigenvector is closely related to that of an invariant subspace. An *eigenvector* v for a linear operator T is a nonzero vector such that

$$(3.9)\qquad T(v) = cv$$

for some scalar $c \in F$. Here c is allowed to take the value 0, but the vector v can not be zero. Geometrically, if $V = \mathbb{R}^n$, an eigenvector is a nonzero vector v such that v and $T(v)$ are parallel.

The scalar c appearing in (3.9) is called the *eigenvalue* associated to the eigenvector v. When we speak of an *eigenvalue* of a linear operator T, we mean a scalar $c \in F$ which is the eigenvalue associated to some eigenvector.

For example, the standard basis vector e_1 is an eigenvector for left multiplication by the matrix

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

The eigenvalue associated to the eigenvector e_1 is 3. Or, the vector $(0,1,1)^t$ is an eigenvector for multiplication by the matrix

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

on the space $\mathbb{R}^3$ of column vectors, and its eigenvalue is 2.

Sometimes eigenvectors and eigenvalues are called *characteristic vectors* and *characteristic values*.

Let v be an eigenvector for a linear operator T. The subspace W spanned by v is T-invariant, because $T(av) = acv \in W$ for all $a \in F$. Conversely, if this subspace is invariant, then v is an eigenvector. So an eigenvector can be described as a basis of a one-dimensional T-invariant subspace. If v is an eigenvector, and if we extend it to a basis $(v = v_1, \ldots, v_n)$ of V, then the matrix of T will have the block form

$$\begin{bmatrix} c & B \\ 0 & D \end{bmatrix} = \left[\begin{array}{c|ccc} c & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & * & \\ 0 & & & \end{array}\right],$$

where c is the eigenvalue associated to v_1. This is the block decomposition (3.7) in the case of an invariant subspace of dimension 1.

When we speak of an *eigenvector* for an $n \times n$ matrix A, we mean a vector which is an eigenvector for left multiplication by A, a nonzero column vector such that

$$AX = cX, \quad \text{for some } c \in F.$$

As before, the scalar c is called an *eigenvalue*. Suppose that A is the matrix of T with respect to a basis $\mathbf{B}$, and let X denote the coordinate vector of a vector $v \in V$. Then $T(v)$ has coordinates AX (2.4). Hence X is an eigenvector for A if and only if v is an eigenvector for T. Moreover, if so, then the eigenvalues are the same: T and A have the same eigenvalues.

(3.10) **Corollary.** Similar matrices have the same eigenvalues.

This follows from the fact (3.5) that similar matrices represent the same linear transformation. □

Eigenvectors aren't always easy to find, but it is easy to tell whether or not a given vector X is an eigenvector for a matrix A. We need only check whether or not AX is a multiple of X. So we can tell whether or not a given vector v is an eigenvector for a linear operator T, provided that the coordinate vector of v and the matrix of T with respect to a basis are known. If we do this for one of the basis vectors, we find the following criterion:

(3.11) *The basis vector v_j is an eigenvector of T, with eigenvalue c, if and only if the jth column of A has the form ce_j.*

For the matrix A is defined by the property $T(v_j) = v_1 a_{1j} + \cdots + v_n a_{nj}$. So if $T(v_j) = cv_j$, then $a_{jj} = c$ and $a_{ij} = 0$ if $i \neq j$. □

(3.12) **Corollary.** With the above notation, A is a diagonal matrix if and only if every basis vector v_j is an eigenvector. □

(3.13) **Corollary.** The matrix A of a linear transformation is similar to a diagonal matrix if and only if there is a basis $\mathbf{B}' = (v_1', \ldots, v_n')$ of V made up of eigenvectors. □

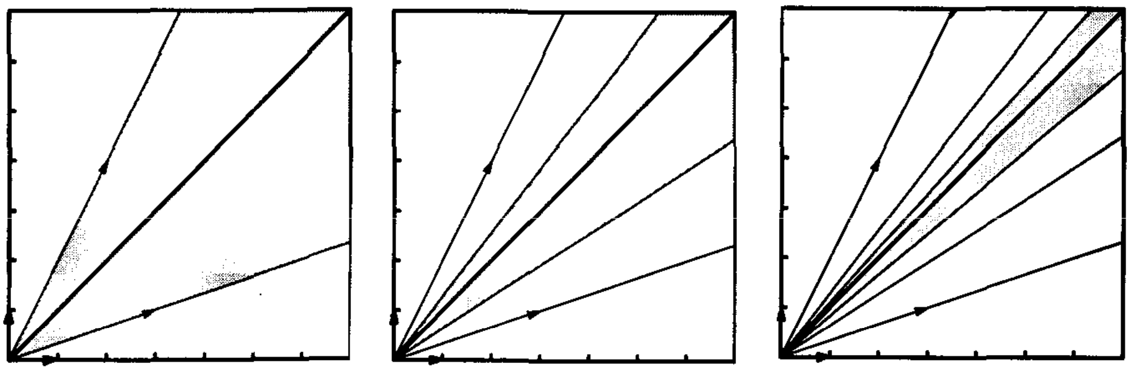
This last corollary shows that we can represent a linear operator very simply by a diagonal matrix, provided that it has enough eigenvectors. We will see in Section 4 that every linear operator on a *complex* vector space has at least one eigenvector, and in Section 6 that in most cases the eigenvectors form a basis. But a linear operator on a real vector space needn't have an eigenvector. For example, the rotation ρ_θ (3.1) of the plane does not carry any vector to a parallel one, unless $\theta = 0$ or π. So ρ_θ has no eigenvector unless $\theta = 0$ or π.

The situation is quite different for real matrices having positive entries. Such matrices are sometimes called *positive* matrices. They occur often in applications, and one of their most important properties is that they always have an eigenvector whose coordinates are positive (a *positive* eigenvector). Instead of proving this fact, let us illustrate it in the case of two variables by examining the effect of multiplication by a positive 2×2 matrix A on $\mathbb{R}^2$.

Let $w_i = Ae_i$. The parallelogram law for vector addition shows that A sends the first quadrant S to the sector bounded by the vectors w_1, w_2. And the coordinate vector of w_i is the ith column of A. Since the entries of A are positive, the vectors w_i lie in the first quadrant. So A carries the first quadrant to itself: $S \supset AS$. Applying A again, we find $AS \supset A^2S$, and so on:

$$S \supset AS \supset A^2S \supset A^3S \supset \ldots, \tag{3.14}$$

as illustrated below in Figure (3.15) for the matrix $A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$.



(3.15) **Figure.** Images of the first quadrant under repeated multiplication by a positive matrix.

Now the intersection of a nested set of sectors is either a sector or a half line. In our case, the intersection $Z = \bigcap A^r S$ turns out to be a half line. This is intuitively plausible, and it can be shown in various ways. The proof is left as an exercise. We multiply the relation $Z = \bigcap A^r S$ on both sides by A:

$$AZ = A\left(\bigcap_0^\infty A^r S\right) = \bigcap_1^\infty A^r S = Z.$$

Hence $Z = AZ$. This shows that the nonzero vectors in Z are eigenvectors. □

4. THE CHARACTERISTIC POLYNOMIAL

In this section we determine the eigenvectors of an arbitrary linear operator T. Recall that an eigenvector for T is a nonzero vector v such that

$$T(v) = cv, \tag{4.1}$$

for some c in F. At first glance, it seems difficult to find eigenvectors if the matrix of the linear operator is complicated. The trick is to solve a different problem, namely to determine the *eigenvalues* first. Once an eigenvalue c is determined, equation (4.1) becomes linear in the coordinates of v, and solving it presents no problem.

We begin by writing (4.1) in the form

$$[T - cI](v) = 0, \tag{4.2}$$

where I stands for the identity operator and $T - cI$ is the linear operator defined by

$$[T - cI](v) = T(v) - cv. \tag{4.3}$$

It is easy to check that $T - cI$ is indeed a linear operator. If A is the matrix of T with respect to some basis, then the matrix of $T - cI$ is $A - cI$.

We can restate (4.2) as follows:

$$v \text{ is in the kernel of } T - cI. \tag{4.4}$$

(4.5) **Lemma.** The following conditions on a linear operator $T: V \longrightarrow V$ on a finite-dimensional vector space are equivalent:

(a) $\ker T > 0$.
(b) $\operatorname{im} T < V$.
(c) If A is the matrix of the operator with respect to an arbitrary basis, then $\det A = 0$.
(d) 0 is an eigenvalue of T.

Proof. The dimension formula (1.6) shows that $\ker T > 0$ if and only if $\operatorname{im} T < V$. This is true if and only if T is not an isomorphism, or, equivalently, if and only if A is not an invertible matrix. And we know that the square matrices A which are not invertible are those with determinant zero. This shows the equivalence of (a), (b), and (c). Finally, the nonzero vectors in the kernel of T are the eigenvectors with eigenvalue zero. Hence (a) is equivalent to (d). □

The conditions (4.5a) and (4.5b) are not equivalent for infinite-dimensional vector spaces. For example, let $V = \mathbb{R}^\infty$ be the space of infinite row vectors $(a_1, a_2, \ldots)$, as in Section 5 of Chapter 3. The *shift operator*, defined by

$$T(a_1, a_2, \ldots) = (0, a_1, a_2, \ldots), \tag{4.6}$$

is a linear operator on V. For this operator, $\ker T = 0$ but $\operatorname{im} T < V$.

(4.7) **Definition.** A linear operator T on a finite-dimensional vector space V is called *singular* if it satisfies any of the equivalent conditions of (4.5). Otherwise, T is *nonsingular*.

We know that c is an eigenvalue for the operator T if and only if $T - cI$ has a nonzero kernel (4.4). So, if we replace T by $T - cI$ in the lemma above, we find:

(4.8) **Corollary.** The eigenvalues of a linear operator T are the scalars $c \in F$ such that $T - cI$ is singular. □

If A is the matrix of T with respect to some basis, then the matrix of $T - cI$ is $A - cI$. So $T - cI$ is singular if and only if $\det(A - cI) = 0$. This determinant can be computed explicitly, and doing so provides us with a concrete method for determining the eigenvalues and eigenvectors.

Suppose for example that A is the matrix

$$\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \tag{4.9}$$

whose action on $\mathbb{R}^2$ is illustrated in Figure (3.15). Then

$$A - cI = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} - \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} = \begin{bmatrix} 3-c & 2 \\ 1 & 4-c \end{bmatrix}$$

and

$$\det(A - cI) = c^2 - 7c + 10 = (c-5)(c-2).$$

This determinant vanishes if $c = 5$ or 2, so we have shown that the eigenvalues of A are 5 and 2. To find the eigenvectors, we solve the two systems of linear equations $[A - 5I]X = 0$ and $[A - 2I]X = 0$. The solutions are unique up to scalar factor:

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}. \tag{4.10}$$

Note that the eigenvector v_1 with eigenvalue 5 is in the first quadrant. It lies on the half line Z which is illustrated in Figure (3.15).

We now make the same computation with an arbitrary matrix. It is convenient to change sign. Obviously $\det(cI - A) = 0$ if and only if $\det(A - cI) = 0$. Also, it is customary to replace the symbol c by a variable t. We form the matrix $tI - A$:

$$tI - A = \begin{bmatrix} (t-a_{11}) & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & (t-a_{22}) & \cdots & -a_{2n} \\ \vdots & & & \vdots \\ -a_{n1} & \cdots & \cdots & (t-a_{nn}) \end{bmatrix}. \tag{4.11}$$

Then the complete expansion of the determinant [Chapter 1 (4.11)] shows that $\det(tI - A)$ is a polynomial of degree n in t, whose coefficients are scalars.

(4.12) **Definition.** The *characteristic polynomial* of a linear operator T is the polynomial

$$p(t) = \det(tI - A),$$

where A is the matrix of T with respect to some basis.

The eigenvalues of T are determined by combining (4.8) and (4.12): c is an eigenvalue if and only if $p(c) = 0$.

(4.13) **Corollary.** The eigenvalues of a linear operator are the roots of its characteristic polynomial. □

(4.14) **Corollary.** The eigenvalues of an upper or lower triangular matrix are its diagonal entries.

Proof. If A is an upper triangular matrix, then so is $tI - A$. The determinant of a triangular matrix is the product of its diagonal entries, and the diagonal entries of $tI - A$ are $t - a_{ii}$. Therefore the characteristic polynomial is $p(t) = (t - a_{11})(t - a_{22})\cdots(t - a_{nn})$, and its roots, the eigenvalues, are $a_{11}, \dots, a_{nn}$. □

We can compute the characteristic polynomial of an arbitrary 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

without difficulty. It is

(4.15) $$\det(tI - A) = \det\begin{bmatrix} t-a & -b \\ -c & t-d \end{bmatrix} = t^2 - (a + d)t + (ad - bc).$$

The discriminant of this polynomial is

(4.16) $$(a + d)^2 - 4(ad - bc) = (a - d)^2 + 4bc.$$

If the entries of A are positive real numbers, then the discriminant is also positive, and therefore the characteristic polynomial has real roots, as predicted at the end of Section 3.

(4.17) **Proposition.** The characteristic polynomial of an operator T does not depend on the choice of a basis.

Proof. A second basis leads to a matrix $A' = PAP^{-1}$ [see (3.4)]. We have

$$tI - A' = tI - PAP^{-1} = P(tI)P^{-1} - PAP^{-1} = P(tI - A)P^{-1}.$$

Thus

$$\det(tI - A') = \det(P(tI - A)P^{-1}) = \det P \det(tI - A) \det P^{-1} = \det(tI - A).$$

So the characteristic polynomials computed with A and A' are equal, as was asserted. □

(4.18) **Proposition.** The characteristic polynomial $p(t)$ has the form

$$p(t) = t^n - (\operatorname{tr} A)t^{n-1} + (\textit{intermediate terms}) + (-1)^n(\det A),$$

where $\operatorname{tr} A$, the *trace* of A, is the sum of the diagonal entries:

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

All coefficients are independent of the basis. For instance $\operatorname{tr} PAP^{-1} = \operatorname{tr} A$.

This is proved by computation. The independence of the basis follows from (4.17). □

Since the characteristic polynomial, the trace, and the determinant are independent of the basis, they depend only on the operator T. So we may define the terms *characteristic polynomial, trace,* and *determinant* of a linear operator T to be those obtained using the matrix of T with respect to an arbitrary basis.

(4.19) **Proposition.** Let T be a linear operator on a finite-dimensional vector space V.

(a) If V has dimension n, then T has at most n eigenvalues.
(b) If F is the field of complex numbers and $V \neq 0$, then T has at least one eigenvalue, and hence it has an eigenvector.

Proof.

(a) A polynomial of degree n can have at most n different roots. This is true for any field F, though we have not proved it yet [see Chapter 11, (1.8)]. So we can apply (4.13).
(b) Every polynomial of positive degree with complex coefficients has at least one complex root. This fact is called the Fundamental Theorem of Algebra. There is a proof in Chapter 13 (9.1). □

For example, let A be the rotation (3.1) of the real plane $\mathbb{R}^2$ by an angle θ. Its characteristic polynomial is

$$p(t) = t^2 - (2\cos\theta)t + 1, \tag{4.20}$$

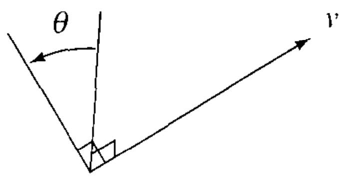
which has no real root unless $\cos\theta = \pm 1$. But if we view A as an operator on $\mathbb{C}^2$, there are two complex eigenvalues.

5. ORTHOGONAL MATRICES AND ROTATIONS

In this section we describe the rotations of two- and three-dimensional spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ about the origin as linear operators. We have already noted (3.1) that a rotation of $\mathbb{R}^2$ through an angle θ is represented as multiplication by the matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

A rotation of $\mathbb{R}^3$ about the origin can be described by a pair (v, θ) consisting of a *unit vector* v, a vector of length 1, which lies in the axis of rotation, and a nonzero angle θ, the angle of rotation. The two pairs (v, θ) and $(-v, -\theta)$ represent the same rotation. We also consider the identity map to be a rotation, though its axis is indeterminate.



(5.1) **Figure.**

The matrix representing a rotation through the angle θ about the vector e_1 is obtained easily from the 2×2 rotation matrix. It is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}. \tag{5.2}$$

Multiplication by A fixes the first coordinate x_1 of a vector and operates by rotation on $(x_2, x_3)^t$. All rotations of $\mathbb{R}^3$ are linear operators, but their matrices can be fairly complicated. The object of this section is to describe these rotation matrices.

A real $n \times n$ matrix A is called *orthogonal* if $A^t = A^{-1}$, or, equivalently, if $A^t A = I$. The orthogonal $n \times n$ matrices form a subgroup of $GL_n(\mathbb{R})$ denoted by O_n and called the *orthogonal group:*

$$O_n = \{A \in GL_n(\mathbb{R}) \mid A^t A = I\}. \tag{5.3}$$

The determinant of an orthogonal matrix is ± 1, because if $A^t A = I$, then

$$(\det A)^2 = (\det A^t)(\det A) = 1.$$

The orthogonal matrices having determinant $+1$ form a subgroup called the *special orthogonal group* and denoted by SO_n:

$$SO_n = \{A \in GL_n(\mathbb{R}) \mid A^t A = I, \det A = 1\}. \tag{5.4}$$

This subgroup has one coset in addition to SO_n, namely the set of elements with determinant -1. So it has index 2 in O_n.

The main fact which we will prove about rotations is stated below:

(5.5) **Theorem.** The rotations of $\mathbb{R}^2$ or $\mathbb{R}^3$ about the origin are the linear operators whose matrices with respect to the standard basis are orthogonal and have determinant 1. In other words, a matrix A represents a rotation of $\mathbb{R}^2$ (or $\mathbb{R}^3$) if and only if $A \in SO_2$ (or SO_3).

Note the following corollary:

(5.6) **Corollary.** The composition of two rotations of $\mathbb{R}^3$ about the origin is also a rotation.

This corollary follows from the theorem because the matrix representing the composition of two linear operators is the product matrix, and because SO_3, being a subgroup of $GL_3(\mathbb{R})$, is closed under products. It is far from obvious geometrically. Clearly, the composition of two rotations about the same axis is also a rotation about that axis. But imagine composing rotations about different axes. What is the axis of rotation of the composed operator?

Because their elements represent rotations, the groups SO_2 and SO_3 are called the two- and three-dimensional *rotation groups*. Things become more complicated in dimension > 3. For example, the matrix

$$(5.7) \qquad \begin{bmatrix} \cos\theta & -\sin\theta & & \\ \sin\theta & \cos\theta & & \\ & & \cos\eta & -\sin\eta \\ & & \sin\eta & \cos\eta \end{bmatrix}$$

is an element of SO_4. Left multiplication by this matrix is the composition of a rotation through the angle θ on the first two coordinates and a rotation through the angle η on the last two. Such an operation can not be realized as a single rotation.

The proof of Theorem (5.5) is not very difficult, but it would be clumsy if we did not first introduce some terminology. So we will defer the proof to the end of the section.

To understand the relationship between orthogonal matrices and rotations, we will need the dot product of vectors. By definition, the *dot product* of column vectors X and Y is

$$(5.8) \qquad (X \cdot Y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

It is sometimes useful to write the dot product in matrix form as

$$(5.9) \qquad (X \cdot Y) = X^tY.$$

There are two main properties of the dot product of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. The first is that $(X \cdot X)$ is the square of the length of the vector:

$$|X|^2 = x_1^2 + x_2^2 \quad \text{or} \quad x_1^2 + x_2^2 + x_3^2,$$

according to the case. This property, which follows from Pythagoras's theorem, is the basis for the definition of length of vectors in $\mathbb{R}^n$: The *length* of X is defined by the formula

$$(5.10) \qquad |X|^2 = (X \cdot X) = x_1^2 + \cdots + x_n^2.$$

The *distance* between two vectors X, Y is defined to be the length $|X - Y|$ of $X - Y$.

The second important property of dot product in $\mathbb{R}^2$ and $\mathbb{R}^3$ is the formula

$$(5.11) \qquad (X \cdot Y) = |X|\,|Y| \cos\theta,$$

where θ is the angle between the vectors. This formula is a consequence of the law of cosines

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

for the side lengths a, b, c of a triangle, where θ is the angle subtended by the sides a, b. To derive (5.11), we apply the law of cosines to the triangle with vertices $0, X, Y$. Its side lengths are $|X|$, $|Y|$ and $|X - Y|$, so the law of cosines can be written as

$$(X - Y \cdot X - Y) = (X \cdot X) + (Y \cdot Y) - 2|X|\,|Y| \cos\theta.$$

The left side expands to

$$(X - Y \cdot X - Y) = (X \cdot X) - 2(X \cdot Y) + (Y \cdot Y),$$

and formula (5.11) is obtained by comparing terms.

The most important application of (5.11) is that two vectors X and Y are orthogonal, meaning that the angle θ is $\pi/2$, if and only if $(X \cdot Y) = 0$. This property is taken as the definition of orthogonality of vectors in $\mathbb{R}^n$:

(5.12) *X is orthogonal to Y if $(X \cdot Y) = 0$.*

(5.13) **Proposition.** The following conditions on a real $n \times n$ matrix A are equivalent:

(a) A is orthogonal.
(b) Multiplication by A preserves dot product, that is, $(AX \cdot AY) = (X \cdot Y)$ for all column vectors X, Y.
(c) The columns of A are mutually orthogonal unit vectors.

A basis consisting of mutually orthogonal unit vectors is called an *orthonormal* basis. An orthogonal matrix is one whose columns form an orthonormal basis.

Left multiplication by an orthogonal matrix is also called an *orthogonal operator*. Thus the orthogonal operators on $\mathbb{R}^n$ are the ones which preserve dot product.

Proof of Proposition (5.13). We write $(X \cdot Y) = X^t Y$. If A is orthogonal, then $A^t A = I$, so

$$(X \cdot Y) = X^t Y = X^t A^t A Y = (AX)^t (AY) = (AX \cdot AY).$$

Conversely, suppose that $X^t Y = X^t A^t A Y$ for all X and Y. We rewrite this equality as $X^t B Y = 0$, where $B = I - A^t A$. For any matrix B,

(5.14) $$e_i^t B e_j = b_{ij}.$$

So if $X^t B Y = 0$ for all X, Y, then $e_i^t B e_j = b_{ij} = 0$ for all i, j, and $B = 0$. Therefore $I = A^t A$. This proves the equivalence of (a) and (b). To prove that (a) and (c) are equivalent, let A_j denote the jth column of the matrix A. The (i, j) entry of the product matrix $A^t A$ is $(A_i \cdot A_j)$. Thus $A^t A = I$ if and only if $(A_i \cdot A_i) = 1$ for all i, and $(A_i \cdot A_j) = 0$ for all $i \neq j$, which is to say that the columns have length 1 and are orthogonal. □

The geometric meaning of multiplication by an orthogonal matrix can be explained in terms of rigid motions. A *rigid motion* or *isometry* of $\mathbb{R}^n$ is a map $m: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ which is distance preserving; that is, it is a map satisfying the following condition: If X, Y are points of $\mathbb{R}^n$, then the distance from X to Y is equal to the distance from $m(X)$ to $m(Y)$:

$$(5.15) \qquad |m(X) - m(Y)| = |X - Y|.$$

Such a rigid motion carries a triangle to a congruent triangle, and therefore it preserves angles and shapes in general.

Note that the composition of two rigid motions is a rigid motion, and that the inverse of a rigid motion is a rigid motion. Therefore the rigid motions of $\mathbb{R}^n$ form a group M_n, with composition of operations as its law of composition. This group is called the *group of motions*.

(5.16) **Proposition.** Let m be a map $\mathbb{R}^n \longrightarrow \mathbb{R}^n$. The following conditions on m are equivalent:

(a) m is a rigid motion which fixes the origin.
(b) m preserves dot product; that is, for all $X, Y \in \mathbb{R}^n$, $(m(X) \cdot m(Y)) = (X \cdot Y)$.
(c) m is left multiplication by an orthogonal matrix.

(5.17) **Corollary.** A rigid motion which fixes the origin is a linear operator.

This follows from the equivalence of (a) and (c).

Proof of Proposition (5.16). We will use the shorthand $'$ to denote the map m, writing $m(X) = X'$. Suppose that m is a rigid motion fixing 0. With the shorthand notation, the statement (5.15) that m preserves distance reads

$$(5.18) \qquad (X' - Y' \cdot X' - Y') = (X - Y \cdot X - Y)$$

for all vectors X, Y. Setting $Y = 0$ shows that $(X' \cdot X') = (X \cdot X)$ for all X. We expand both sides of (5.18) and cancel $(X \cdot X)$ and $(Y \cdot Y)$, obtaining $(X' \cdot Y') = (X \cdot Y)$. This shows that m preserves dot product, hence that (a) implies (b).

To prove that (b) implies (c), we note that the only map which preserves dot product and which also fixes each of the basis vectors e_i is the identity. For, if m preserves dot product, then $(X \cdot e_j) = (X' \cdot e_j')$ for any X. If $e_j' = e_j$ as well, then

$$x_j = (X \cdot e_j) = (X' \cdot e_j') = (X' \cdot e_j) = x_j'$$

for all j. Hence $X = X'$, and m is the identity.

Now suppose that m preserves dot product. Then the images $e_1', \ldots, e_n'$ of the standard basis vectors are orthonormal: $(e_i' \cdot e_i') = 1$ and $(e_i' \cdot e_j') = 0$ if $i \neq j$. Let $\mathbf{B}' = (e_1', \ldots, e_n')$, and let $A = [\mathbf{B}']$. According to Proposition (5.13), A is an orthogonal matrix. Since the orthogonal matrices form a group, A^{-1} is also orthogonal. This being so, multiplication by A^{-1} preserves dot product too. So the composed motion $A^{-1}m$ preserves dot product, and it fixes each of the basis vectors e_i. Therefore $A^{-1}m$ is the identity map. This shows that m is left multiplication by A, as required.

Finally, if m is a linear operator whose matrix A is orthogonal, then $X' - Y' = (X - Y)'$ because m is linear, and $|X' - Y'| = |(X - Y)'| = |X - Y|$ by (5.13b). So m is a rigid motion. Since a linear operator also fixes 0, this shows that (c) implies (a). □

One class of rigid motions which do not fix the origin, and which are therefore not linear operators, is the translations. Given any fixed vector $b = (b_1,\dots,b_n)^t$ in $\mathbb{R}^n$, *translation* by b is the map

(5.19)
$$t_b(X) = X + b = \begin{bmatrix} x_1 + b_1 \\ \vdots \\ x_n + b_n \end{bmatrix}.$$

This map is a rigid motion because $t_b(X) - t_b(Y) = (X + b) - (Y + b) = X - Y$, and hence $|t_b(X) - t_b(Y)| = |X - Y|$.

(5.20) **Proposition.** Every rigid motion m is the composition of an orthogonal linear operator and a translation. In other words, it has the form $m(X) = AX + b$ for some orthogonal matrix A and some vector b.

Proof. Let $b = m(0)$. Then $t_{-b}(b) = 0$, so the composed operation $t_{-b}m$ is a rigid motion which fixes the origin: $t_{-b}(m(0)) = 0$. According to Proposition (5.16), $t_{-b}m$ is left multiplication by an orthogonal matrix A: $t_{-b}m(X) = AX$. Applying t_b to both sides of this equation, we find $m(X) = AX + b$.

Note that both the vector b and the matrix A are uniquely determined by m, because $b = m(0)$ and A is the operator $t_{-b}m$. □

Recall that the determinant of an orthogonal matrix is ± 1. An orthogonal operator is called *orientation-preserving* if its determinant is $+1$, and *orientation-reversing* if its determinant is -1. Similarly, let m be a rigid motion. We write $m(X) = AX + b$ as above. Then m is called *orientation-preserving* if $\det A = 1$, and *orientation-reversing* if $\det A = -1$. A motion of $\mathbb{R}^2$ is orientation-reversing if it flips the plane over, and orientation-preserving if it does not.

Combining Theorem (5.5) with Proposition (5.16) gives us the following characterization of rotations:

(5.21) **Corollary.** The rotations of $\mathbb{R}^2$ and $\mathbb{R}^3$ are the orientation-preserving rigid motions which fix the origin. □

We now proceed to the proof of Theorem (5.5), which characterizes the rotations of $\mathbb{R}^2$ and $\mathbb{R}^3$ about the origin. Every rotation ρ is a rigid motion, so Proposition (5.16) tells us that ρ is multiplication by an orthogonal matrix A. Also, the determinant of A is 1. This is because $\det A = \pm 1$ for any orthogonal matrix, and because the determinant varies continuously with the angle of rotation. When the angle is zero, A is the identity matrix, which has determinant 1. Thus the matrix of a rotation is an element of SO_2 or SO_3.

Conversely, let $A \in SO_2$ be an orthogonal 2×2 matrix of determinant 1. Let v_1 denote the first column Ae_1 of A. Since A is orthogonal, v_1 is a unit vector. There is a rotation R (3.1) such that $Re_1 = v_1$ too. Then $B = R^{-1}A$ fixes e_1. Also, A and R are elements of SO_2, and this implies that B is in SO_2. So the columns of B form an orthonormal basis of $\mathbb{R}^2$, and the first column is e_1. Being of length 1 and orthogonal to e_1, the second column must be either e_2 or $-e_2$, and the second case is ruled out by the fact that $\det B = 1$. It follows that $B = I$ and that $A = R$. So A is a rotation.

To prove that an element A of SO_3 represents a rotation, we'd better decide on a definition of a rotation ρ of $\mathbb{R}^3$ about the origin. We will require the following:

(5.22)

(i) ρ is a rigid motion which fixes the origin;
(ii) ρ also fixes a nonzero vector v;
(iii) ρ operates as a rotation on the plane P orthogonal to v.

According to Proposition (5.16), the first condition is equivalent to saying that ρ is an orthogonal operator. So our matrix $A \in SO_3$ satisfies this condition. Condition (ii) can be stated by saying that v is an eigenvector for the operator ρ, with eigenvalue 1. Then since ρ preserves orthogonality, it sends the orthogonal space P to itself. In other words, P is an invariant subspace. Condition (iii) says that the restriction of ρ to this invariant subspace is a rotation.

Notice that the matrix (5.2) does satisfy these conditions, with $v = e_1$.

(5.23) **Lemma.** Every element $A \in SO_3$ has the eigenvalue 1.

Proof. We will show that $\det(A - I) = 0$. This will prove the lemma [see (4.8)]. This proof is tricky, but efficient. Recall that $\det A = \det A^t$ for any matrix A, so $\det A^t = 1$. Since A is orthogonal, $A^t(A - I) = (I - A)^t$. Then

$$\det(A - I) = \det A^t(A - I) = \det(I - A)^t = \det(I - A).$$

On the other hand, for any 3×3 matrix B, $\det(-B) = -\det B$. Therefore $\det(A - I) = -\det(I - A)$, and it follows that $\det(A - I) = 0$. $\square$

Now given a matrix $A \in SO_3$, the lemma shows that left multiplication by A fixes a nonzero vector v_1. We normalize its length to 1, and we choose orthogonal unit vectors v_2, v_3 lying in the plane P orthogonal to v_1. Then $\mathbf{B} = (v_1, v_2, v_3)$ is an orthonormal basis of $\mathbb{R}^3$. The matrix $P = [\mathbf{B}]^{-1}$ is orthogonal because $[\mathbf{B}]$ is orthogonal, and $A' = PAP^{-1}$ represents the same operator as A does, with respect to the basis $\mathbf{B}$. Since A and P are orthogonal, so is A'. Also $\det A' = \det A = 1$. So $A' \in SO_3$.

Since v_1 is an eigenvector with eigenvalue 1, the first column of A' is e_1. Since A' is orthogonal, the other columns are orthogonal to e_1, and A' has the block form

$$\left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & R \end{array}\right].$$

Using the fact that $A' \in SO_3$, one finds that $R \in SO_2$. So R is a rotation. This shows that A' has the form (5.2) and that it represents a rotation. Hence A does too. This completes the proof of Theorem (5.5). □

(5.24) *Note.* To keep the new basis separate from the old basis, we denoted it by $\mathbf{B}'$ in Chapter 3. The prime is not needed when the old basis is the standard basis, and since it clutters the notation, we will often drop it, as we did here.

6. DIAGONALIZATION

In this section we show that for "most" linear operators on a *complex* vector space, there is a basis such that the matrix of the operator is diagonal. The key fact, which we already noted at the end of Section 4, is that every complex polynomial of positive degree has a root. This tells us that every linear operator has an eigenvector.

(6.1) **Proposition.**

(a) *Vector space form:* Let T be a linear operator on a finite-dimensional complex vector space V. There is a basis $\mathbf{B}$ of V such that the matrix A of T is upper triangular.

(b) *Matrix form:* Every complex $n \times n$ matrix A is similar to an upper triangular matrix. In other words, there is a matrix $P \in GL_n(\mathbb{C})$ such that PAP^{-1} is upper triangular.

Proof. The two assertions are equivalent, because of (3.5). We begin by applying (4.19b), which shows the existence of an eigenvector, call it v_1'. Extend to a basis $\mathbf{B}' = (v_1', \ldots, v_n')$ for V. Then by (3.11), the first column of the matrix A' of T with respect to $\mathbf{B}'$ will be $(c_1, 0, \ldots, 0)^t$, where c_1 is the eigenvalue of v_1'. Therefore A' has the form

$$A' = \left[\begin{array}{c|c} c_1 & * \cdots * \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right]$$

where B is an $(n-1) \times (n-1)$ matrix. The matrix version of this reduction is this: Given any $n \times n$ matrix A, there is a $P \in GL_n(\mathbb{C})$ such that $A' = PAP^{-1}$ has the above form. Now apply induction on n. By induction, we may assume that the existence of some $Q \in GL_{n-1}(\mathbb{C})$ such that QBQ^{-1} is triangular has been proved. Let Q_1 be the $n \times n$ matrix

$$\left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & Q & \\ 0 & & & \end{array}\right].$$

Then

$$(Q_1P)A(Q_1P)^{-1} = Q_1(PAP^{-1})Q_1^{-1} = Q_1A'Q_1^{-1}$$

has the form

$$\left[\begin{array}{c|ccc} c_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & QBQ^{-1} & \\ 0 & & & \end{array}\right],$$

which is triangular. □

As we mentioned, the important point in the proof is that every complex polynomial has a root. The same proof will work for any field F, provided that all the roots of the characteristic polynomial are in the field.

(6.2) **Corollary.** Let F be a field.

(a) *Vector space form:* Let T be a linear operator on a finite-dimensional vector space V over F, and suppose that the characteristic polynomial of T factors into linear factors in the field F. Then there is a basis **B** of V such that the matrix A of T is triangular.

(b) *Matrix form:* Let A be an $n \times n$ matrix whose characteristic polynomial factors into linear factors in the field F. There is a matrix $P \in GL_n(F)$ such that PAP^{-1} is triangular.

Proof. The proof is the same, except that to make the induction step one has to check that the characteristic polynomial of the matrix B is $p(t)/(t - c_1)$, where $p(t)$ is the characteristic polynomial of A. This is true because $p(t)$ is also the characteristic polynomial of A' (4.17), and because $\det(tI - A') = (t - c_1)\det(tI - B)$.

So our hypothesis that the characteristic polynomial factors into linear factors carries over from A to B. □

Let us now ask which matrices A are similar to *diagonal* matrices. As we saw in (3.12), these are the matrices A which have a basis of eigenvectors. Suppose again that $F = \mathbb{C}$, and look at the roots of the characteristic polynomial $p(t)$. Each root is the eigenvalue associated to some eigenvector, and an eigenvector has only one eigenvalue. Most complex polynomials of degree n have n distinct roots. So most complex matrices have n eigenvectors with different eigenvalues, and it is reasonable to suppose that these eigenvectors may form a basis. This is true.

(6.3) **Proposition.** Let $v_1,\dots,v_r \in V$ be eigenvectors for a linear operator T, with distinct eigenvalues $c_1,\dots,c_r$. Then the set $(v_1,\dots,v_r)$ is linearly independent.

Proof. Induction on r: Suppose that a dependence relation

$$0 = a_1v_1 + \cdots + a_rv_r$$

is given. We must show that $a_i = 0$ for all i, and to do so we apply the operator T:

$$0 = T(0) = a_1T(v_1) + \cdots + a_rT(v_r) = a_1c_1v_1 + \cdots + a_rc_rv_r.$$

This is a second dependence relation among $(v_1,\dots,v_r)$. We eliminate v_r from the two relations, multiplying the first relation by c_r and subtracting the second:

$$0 = a_1(c_r - c_1)v_1 + \cdots + a_{r-1}(c_r - c_{r-1})v_{r-1}.$$

Applying the principle of induction, we assume that $(v_1,\dots,v_{r-1})$ are independent. Then the coefficients $a_1(c_r - c_1),\dots,a_{r-1}(c_r - c_{r-1})$ are all zero. Since the c_i's are distinct, $c_r - c_i \neq 0$ if $i < r$. Thus $a_1 = \dots = a_{r-1} = 0$, and the original relation is reduced to $0 = a_rv_r$. Since an eigenvector can not be zero, $a_r = 0$ too. □

The next theorem follows by combining (3.12) and (6.3):

(6.4) **Theorem.** Let T be a linear operator on a vector space V of dimension n over a field F. Assume that its characteristic polynomial has n distinct roots in F. Then there is a basis for V with respect to which the matrix of T is diagonal. □

Note that the diagonal entries are determined, except for their order, by the linear operator T. They are the eigenvalues.

When $p(t)$ has multiple roots, there is usually no basis of eigenvectors, and it is harder to find a nice matrix for T. The study of this case leads to what is called the *Jordan canonical form* for a matrix, which will be discussed in Chapter 12.

As an example of diagonalization, consider the matrix

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$$

whose eigenvectors were computed in (4.10). These eigenvectors form a basis $\mathbf{B} = (v_1, v_2)$ of $\mathbb{R}^2$. According to [Chapter 3 (4.20), see also Note (5.24)], the matrix relating the standard basis $\mathbf{E}$ to this basis $\mathbf{B}$ is

$$P = [\mathbf{B}]^{-1} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}^{-1} = -\frac{1}{3}\begin{bmatrix} -1 & -2 \\ -1 & 1 \end{bmatrix}, \tag{6.5}$$

and $PAP^{-1} = A'$ is diagonal:

$$-\frac{1}{3}\begin{bmatrix} -1 & -2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & \\ & 2 \end{bmatrix} = A'. \tag{6.6}$$

The general rule is stated in Corollary (6.7):

(6.7) **Corollary.** If a basis $\mathbf{B}$ of eigenvectors of A in F^n is known and if $P = [\mathbf{B}]^{-1}$, then $A' = PAP^{-1}$ is diagonal. $\square$

The importance of Theorem (6.4) comes from the fact that it is easy to compute with diagonal matrices. For example, if $A' = PAP^{-1}$ is diagonal, then we can compute powers of the matrix A using the formula

$$A^k = (P^{-1}A'P)^k = P^{-1}A'^kP. \tag{6.8}$$

Thus if A is the matrix (4.9), then

$$A^k = -\frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 5 & \\ & 2 \end{bmatrix}^k\begin{bmatrix} -1 & -2 \\ -1 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 5^k + 2 \cdot 2^k & 2(5^k - 2^k) \\ 5^k - 2^k & 2 \cdot 5^k + 2^k \end{bmatrix}.$$

7. *SYSTEMS OF DIFFERENTIAL EQUATIONS*

We learn in calculus that the solutions to the first-order linear differential equation

$$\frac{dx}{dt} = ax \tag{7.1}$$

are $x(t) = ce^{at}$, c being an arbitrary constant. Indeed, ce^{at} obviously solves (7.1). To show that every solution has this form, let $x(t)$ be an arbitrary differentiable function which is a solution. We differentiate $e^{-at}x(t)$ using the product rule:

$$\frac{d}{dt}(e^{-at}x(t)) = -ae^{-at}x(t) + e^{-at}ax(t) = 0.$$

Thus $e^{-at}x(t)$ is a constant c, and $x(t) = ce^{at}$.

As an application of diagonalization, we will extend this solution to systems of differential equations. In order to write our equations in matrix notation, we use the following terminology. A *vector-valued function* $X(t)$ is a vector whose entries are functions of t. Similarly, a *matrix-valued function* $A(t)$ is a matrix whose entries are functions:

$$X(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad A(t) = \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & & \vdots \\ a_{m1}(t) & \cdots & a_{mn}(t) \end{bmatrix}. \tag{7.2}$$

The calculus operations of taking limits, differentiating, and so on are extended to vector-valued and matrix-valued functions by performing the operations on each entry separately. Thus by definition

$$\lim_{t \to t_0} X(t) = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}, \quad \text{where } \xi_i = \lim_{t \to t_0} x_i(t). \tag{7.3}$$

So this limit exists if and only if $\lim x_i(t)$ exists for each i. Similarly, the derivative of a vector-valued or matrix-valued function is the function obtained by differentiating each entry separately:

$$\frac{dX}{dt} = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \quad \frac{dA}{dt} = \begin{bmatrix} a_{11}'(t) & \cdots & a_{1n}'(t) \\ \vdots & & \vdots \\ a_{m1}'(t) & \cdots & a_{mn}'(t) \end{bmatrix},$$

where $x_i'(t)$ is the derivative of $x_i(t)$, and so on. So dX/dt is defined if and only if each of the functions $x_i(t)$ is differentiable. The derivative can also be described in vector notation, as

$$\frac{dX}{dt} = \lim_{h \to 0} \frac{X(t+h) - X(t)}{h}. \tag{7.4}$$

Here $X(t+h) - X(t)$ is computed by vector addition and the h in the denominator stands for scalar multiplication by h^{-1}. The limit is obtained by evaluating the limit of each entry separately, as above. So the entries of (7.4) are the derivatives $x_i'(t)$. The same is true for matrix-valued functions.

A system of homogeneous first-order linear, constant-coefficient differential equations is a matrix equation of the form

$$\frac{dX}{dt} = AX, \tag{7.5}$$

where A is an $n \times n$ real or complex matrix and $X(t)$ is an n-dimensional vector-valued function. Writing out such a system, we obtain a system of n differential equations, of the form

$$\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1(t) + \cdots + a_{1n}x_n(t) \\ &\vdots \qquad\quad \vdots \qquad\qquad \vdots \\ \frac{dx_n}{dt} &= a_{n1}x_1(t) + \cdots + a_{nn}x_n(t). \end{aligned} \tag{7.6}$$

The $x_i(t)$ are unknown functions, and the a_{ij} are scalars. For example, if we substitute the matrix $\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$ for A, (7.5) becomes a system of two equations in two unknowns:

$$\begin{aligned} \frac{dx_1}{dt} &= 3x_1 + 2x_2 \\ \frac{dx_2}{dt} &= x_1 + 4x_2. \end{aligned} \tag{7.7}$$

The simplest systems (7.5) are those in which A is a diagonal matrix. Let the diagonal entries be a_i. Then equation (7.6) reads

$$\frac{dx_i}{dt} = a_i x_i(t), \quad i = 1, \ldots, n. \tag{7.8}$$

Here the unknown functions x_i are not mixed up by the equations, so we can solve for each one separately:

$$x_i = c_i e^{a_i t}, \tag{7.9}$$

for some constant c_i.

The observation which allows us to solve the differential equation (7.5) in most cases is this: If v is an eigenvector for A with eigenvalue a, then

$$X = e^{at} v \tag{7.10}$$

is a particular solution of (7.5). Here $e^{at}v$ is to be interpreted as the scalar product of the function e^{at} and the vector v. Differentiation operates on the scalar function, fixing the constant vector v, while multiplication by A operates on the vector v, fixing the scalar function e^{at}. Thus $\frac{d}{dt} e^{at} v = a e^{at} v = A e^{at} v$. For example, $(2, -1)^t$ is an eigenvector with eigenvalue 2 of the matrix $\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$, and $\begin{bmatrix} 2e^{2t} \\ -e^{2t} \end{bmatrix}$ solves the system of differential equations (7.7).

This observation allows us to solve (7.5) whenever the matrix A has distinct real eigenvalues. In that case every solution will be a linear combination of the special solutions (7.10). To work this out, it is convenient to diagonalize. Let us replace the notation $'$ used in the previous section by $\tilde{}$ here, to avoid confusion with differentiation. Let P be an invertible matrix such that $PAP^{-1} = \tilde{A}$ is diagonal. So $P = [\mathbf{B}]^{-1}$, where $\mathbf{B}$ is a basis of eigenvectors. We make the linear change of variable

(7.11) $$X = P^{-1}\tilde{X}.$$

Then

(7.12) $$\frac{dX}{dt} = P^{-1}\frac{d\tilde{X}}{dt}.$$

Substituting into (7.5), we find

(7.13) $$\frac{d\tilde{X}}{dt} = PAP^{-1}\tilde{X} = \tilde{A}\tilde{X}.$$

Since $\tilde{A}$ is diagonal, the variables $\tilde{x}_i$ have been separated, so the equation can be solved in terms of exponentials. The diagonal entries of $\tilde{A}$ are the eigenvalues $\lambda_1,\ldots,\lambda_n$ of A, so the solution of the system (7.13) is

(7.14) $$\tilde{x}_i = c_i e^{\lambda_i t}, \quad \text{for some } c_i.$$

Substituting back,

(7.15) $$X = P^{-1}\tilde{X}$$

solves the original system (7.5). This proves the following:

(7.16) **Proposition.** Let A be an $n \times n$ matrix, and let P be an invertible matrix such that $PAP^{-1} = \tilde{A}$ is diagonal, with diagonal entries $\lambda_1,\ldots,\lambda_n$. The general solution of the system $\frac{dX}{dt} = AX$ is $X = P^{-1}\tilde{X}$, where $\tilde{x}_i = c_i e^{\lambda_i t}$, for some arbitrary constants c_i. □

The matrix which diagonalizes A in example (7.7) was computed in (6.5):

(7.17) $$P^{-1} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}, \quad \text{and} \quad \tilde{A} = \begin{bmatrix} 5 & \\ & 2 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} c_1 e^{5t} \\ c_2 e^{2t} \end{bmatrix} \quad \text{and}$$

(7.18) $$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} c_1 e^{5t} \\ c_2 e^{2t} \end{bmatrix} = \begin{bmatrix} c_1 e^{5t} + 2c_2 e^{2t} \\ c_1 e^{5t} - \ c_2 e^{2t} \end{bmatrix}.$$

In other words, every solution is a linear combination of the two basic solutions

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} e^{5t} \\ e^{5t} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2e^{2t} \\ -e^{2t} \end{bmatrix}.$$

These are the solutions (7.10) corresponding to the eigenvectors $(1, 1)^t$ and $(2, -1)^t$. The coefficients c_i appearing in these solutions are arbitrary. They are usually determined by assigning *initial conditions,* meaning the value of X at some particular t_0.

Let us now consider the case that the coefficient matrix A has distinct eigenvalues, but that they are not all real. To copy the method which we used above, we must first consider differential equations of the form (7.1), in which a is a complex number. Properly interpreted, the solutions of such a differential equation still have the form ce^{at}. The only thing to remember is that e^{at} will now be a complex-valued function of t. In order to focus attention, we restrict the variable t to real values here, although this is not the most natural choice when working with complex-valued functions. Allowing t to take on complex values would not change things very much.

The definition of the derivative of a complex-valued function is the same as for real-valued functions:

$$\frac{dx}{dt} = \lim_{h\to 0}\frac{x(t+h) - x(t)}{h}, \tag{7.19}$$

provided that this limit exists. There are no new features. We can write any such function $x(t)$ in terms of its real and imaginary parts, which will be real-valued functions:

$$x(t) = u(t) + iv(t). \tag{7.20}$$

Then x is differentiable if and only if u and v are differentiable, and if they are, the derivative of x is $x' = u' + iv'$. This follows directly from the definition. The usual rules for differentiation, such as the product rule, hold for complex-valued functions. These rules can be proved by applying the corresponding theorem for real functions to u and v, or else by carrying the proof for real functions over to the complex case.

Recall the formula

$$e^{r+si} = e^r(\cos s + i \sin s). \tag{7.21}$$

Differentiation of this formula shows that $de^{at}/dt = ae^{at}$ for all complex numbers $a = r + si$. Therefore ce^{at} solves the differential equation (7.1), and the proof given at the beginning of the section shows that these are the only solutions.

Having extended the case of one equation to complex coefficients, we can now use the method of diagonalization to solve a system of equations (7.5) when A is an arbitrary complex matrix with distinct eigenvalues.

For example, let $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. The vectors $v_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $v_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}$ are eigenvectors, with eigenvalues $1 + i$ and $1 - i$ respectively. Let $\mathbf{B} = (v_1, v_2)$. According to (6.7), A is diagonalized by the matrix P, where

$$P^{-1} = [\mathbf{B}] = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}. \tag{7.22}$$

Formula (7.14) tells us that $\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} c_1 e^{t+it} \\ c_2 e^{t-it} \end{bmatrix}$. The solutions of (7.5) are

(7.23)
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P^{-1}\tilde{X} = \begin{bmatrix} c_1 e^{t+it} + ic_2 e^{t-it} \\ ic_1 e^{t+it} + c_2 e^{t-it} \end{bmatrix},$$

where c_1, c_2 are arbitrary complex numbers. So every solution is a linear combination of the two basic solutions

(7.24)
$$\begin{bmatrix} e^{t+it} \\ ie^{t+it} \end{bmatrix} \text{ and } \begin{bmatrix} ie^{t-it} \\ e^{t-it} \end{bmatrix}.$$

However, these solutions are not completely satisfactory, because we began with a system of differential equations with real coefficients, and the answer we obtained is complex. When the original matrix is real, we want to have real solutions. We note the following lemma:

(7.25) **Lemma.** Let A be a real $n \times n$ matrix, and let $X(t)$ be a complex-valued solution of the differential equation (7.5). The real and imaginary parts of $X(t)$ solve the same equation. □

Now *every* solution of the original equation (7.5), whether real or complex, has the form (7.23) for some complex numbers c_i. So the real solutions are among those we have found. To write them down explicitly, we may take the real and imaginary parts of the complex solutions.

The real and imaginary parts of the basic solutions (7.24) are determined using (7.21). They are

(7.26)
$$\begin{bmatrix} e^t \cos t \\ -e^t \sin t \end{bmatrix} \text{ and } \begin{bmatrix} e^t \sin t \\ e^t \cos t \end{bmatrix}.$$

Every real solution is a real linear combination of these particular solutions.

8. THE MATRIX EXPONENTIAL

Systems of first-order linear, constant-coefficient differential equations can also be solved formally, using the *matrix exponential*. The exponential of an $n \times n$ real or complex matrix A is obtained by substituting a matrix into the Taylor's series

(8.1)
$$1 + x/1! + x^2/2! + x^3/3! + \cdots$$

for e^x. Thus by definition,

(8.2)
$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots.$$

This is an $n \times n$ matrix.

(8.3) **Proposition.** The series (8.2) converges absolutely for all complex matrices A.

In order not to break up the discussion, we have collected the proofs together at the end of the section.

Since matrix multiplication is relatively complicated, it isn't easy to write down the matrix entries of e^A directly. In particular, the entries of e^A are usually not obtained by exponentiating the entries of A. But one case in which they are, and in which the exponential is easily computed, is when A is a diagonal matrix, say with diagonal entries a_i. Inspection of the series shows that e^A is also diagonal in this case and that its diagonal entries are e^{a_i}.

The exponential is also relatively easy to compute for a triangular 2×2 matrix. For example, let

$$A = \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix}. \tag{8.4}$$

Then

$$e^A = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 3 \\ & 4 \end{bmatrix} + \cdots = \begin{bmatrix} e & * \\ & e^2 \end{bmatrix}. \tag{8.5}$$

The diagonal entries are exponentiated to obtain the diagonal entries of e^A. It is a good exercise to calculate the missing entry $*$ directly from the definition.

The exponential of a matrix A can also be determined whenever we know a matrix P such that PAP^{-1} is diagonal. Using the rule $PA^kP^{-1} = (PAP^{-1})^k$ and the distributive law for matrix multiplication, we find

$$Pe^AP^{-1} = PIP^{-1} + (PAP^{-1}) + \frac{1}{2!}(PAP^{-1})^2 + \cdots = e^{PAP^{-1}}. \tag{8.6}$$

Suppose that $PAP^{-1} = \tilde{A}$ is diagonal, with diagonal entries λ_i. Then $e^{\tilde{A}}$ is also diagonal, and its diagonal entries are e^{λ_i}. Therefore we can compute e^A explicitly:

$$e^A = P^{-1}e^{\tilde{A}}P. \tag{8.7}$$

In order to use the matrix exponential to solve systems of differential equations, we need to extend some of the properties of the ordinary exponential to it. The most fundamental property is $e^{x+y} = e^xe^y$. This property can be expressed as a formal identity between the two infinite series which are obtained by expanding

$$\begin{aligned} e^{x+y} &= 1 + (x+y)/1! + (x+y)^2/2! + \cdots \quad \text{and} \\ e^xe^y &= (1 + x/1! + x^2/2! + \cdots)(1 + y/1! + y^2/2! + \cdots). \end{aligned} \tag{8.8}$$

We can not substitute matrices into this identity because the commutative law is needed to obtain equality of the two series. For instance, the quadratic terms of (8.8), computed without the commutative law, are $\frac{1}{2}(x^2 + xy + yx + y^2)$ and $\frac{1}{2}x^2 + xy + \frac{1}{2}y^2$. They are not equal unless $xy = yx$. So there is no reason to expect

e^{A+B} to equal $e^A e^B$ in general. However, if two matrices A and B happen to commute, the formal identity can be applied.

(8.9) **Proposition.**

(a) The formal expansions of (8.8), with commuting variables x,y, are equal.
(b) Let A, B be complex $n \times n$ matrices which commute: $AB = BA$. Then $e^{A+B} = e^A e^B$.

The proof is at the end of the section. □

(8.10) **Corollary.** For any $n \times n$ complex matrix A, the exponential e^A is invertible, and its inverse is e^{-A}.

This follows from the proposition because A and $-A$ commute, and hence $e^A e^{-A} = e^{A-A} = e^0 = I$. □

As a sample application of Proposition (8.9b), consider the matrix

$$A = \begin{bmatrix} 2 & 3 \\ & 2 \end{bmatrix}. \tag{8.11}$$

We can compute its exponential by writing it in the form $A = 2I + B$, where $B = 3e_{12}$. Since $2I$ commutes with B, Proposition (8.9b) applies: $e^A = e^{2I}e^B$, and from the series expansion we read off the values $e^{2I} = e^2 I$ and $e^B = I + B$. Thus

$$e^A = \begin{bmatrix} e^2 & \\ & e^2 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ & 1 \end{bmatrix} = \begin{bmatrix} e^2 & 3e^2 \\ & e^2 \end{bmatrix}.$$

We now come to the main result relating the matrix exponential to differential equations. Given an $n \times n$ matrix A, we consider the exponential e^{tA}, t being a variable scalar, as a matrix-valued function:

$$e^{tA} = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots. \tag{8.12}$$

(8.13) **Proposition.** e^{tA} is a differentiable function of t, and its derivative is Ae^{tA}.

The proof is at the end of the section. □

(8.14) **Theorem.** Let A be a real or complex $n \times n$ matrix. The columns of the matrix e^{tA} form a basis for the vector space of solutions of the differential equation

$$\frac{dX}{dt} = AX.$$

We will need the following lemma, whose proof is an exercise:

(8.15) **Lemma.** *Product rule:* Let $A(t)$ and $B(t)$ be differentiable matrix-valued functions of t, of suitable sizes so that their product is defined. Then the matrix product $A(t)B(t)$ is differentiable, and its derivative is

$$\frac{d}{dt}(A(t)B(t)) = \frac{dA}{dt}B + A\frac{dB}{dt}. \square$$

Proof of Theorem (8.14). Proposition (8.13) shows that the columns of A solve the differential equation, because differentiation and multiplication by A act independently on the columns of the matrix e^{tA}. To show that every solution is a linear combination of the columns, we copy the proof given at the beginning of Section 7. Let $X(t)$ be an arbitrary solution of (7.5). We differentiate the matrix product $e^{-tA}X(t)$, obtaining

$$\frac{d}{dt}(e^{-tA}X(t)) = -Ae^{-tA}X(t) + e^{-tA}AX(t).$$

Fortunately, A and e^{-tA} commute. This follows directly from the definition of the exponential. So the derivative is zero. Therefore, $e^{-tA}X(t)$ is a constant column vector, say $C = (c_1,\dots,c_n)^t$, and $X(t) = e^{tA}C$. This expresses $X(t)$ as a linear combination of the columns of e^{tA}. The expression is unique because e^{tA} is an invertible matrix. $\square$

According to Theorem (8.14), the matrix exponential always solves the differential equation (7.5). Since direct computation of the exponential can be quite difficult, this theorem may not be easy to apply in a concrete situation. But if A is a diagonalizable matrix, then the exponential can be computed as in (8.7): $e^A = P^{-1}e^{\tilde{A}}P$. We can use this method of evaluating e^{tA} to solve equation (7.5), but of course it gives the same result as before. Thus if A is the matrix used in example (7.7), so that $P, \tilde{A}$ are as in (7.17), then

$$e^{t\tilde{A}} = \begin{bmatrix} e^{5t} & \\ & e^{2t} \end{bmatrix}$$

and

$$e^{tA} = P^{-1}e^{t\tilde{A}}P = -\frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} e^{5t} & \\ & e^{2t} \end{bmatrix}\begin{bmatrix} -1 & -2 \\ -1 & 1 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} e^{5t} + 2e^{2t} & 2e^{5t} - 2e^{2t} \\ e^{5t} - e^{2t} & 2e^{5t} + e^{2t} \end{bmatrix}.$$

The columns we have obtained form a second basis for the general solution (7.18).

On the other hand, the matrix $A = \begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix}$, which represents the system of equations

(8.16) $$\frac{dx}{dt} = x, \quad \frac{dy}{dt} = x + y,$$

is not diagonalizable. So the method of Section 7 can not be applied. To solve it, we write $At = It + Bt$, where $B = e_{21}$, and find, as in the discussion of (8.11),

$$e^{At} = e^{It}e^{Bt} = \begin{bmatrix} e^t & \\ te^t & e^t \end{bmatrix}. \tag{8.17}$$

Thus the solutions of (8.16) are linear combinations of the columns

$$\begin{bmatrix} e^t \\ te^t \end{bmatrix}, \quad \begin{bmatrix} 0 \\ e^t \end{bmatrix}. \tag{8.18}$$

To compute the exponential explicitly in all cases requires putting the matrix into Jordan form (see Chapter 12).

We now go back to prove Propositions (8.3), (8.9), and (8.13). For want of a more compact notation, we will denote the i, j-entry of a matrix A by A_{ij} here. So $(AB)_{ij}$ will stand for the entry of the product matrix AB, and $(A^k)_{ij}$ for the entry of A^k. With this notation, the i, j-entry of e^A is the sum of the series

$$(e^A)_{ij} = I_{ij} + A_{ij} + \frac{1}{2!}(A^2)_{ij} + \frac{1}{3!}(A^3)_{ij} + \cdots. \tag{8.19}$$

In order to prove that the series for the exponential converges, we need to show that the entries of the powers A^k of a given matrix do not grow too fast, so that the absolute values of the i, j-entries form a bounded (and hence convergent) series. Let us define the *norm* of an $n \times n$ matrix A to be the maximum absolute value of the matrix entries:

$$\|A\| = \max_{i,j} |A_{ij}|. \tag{8.20}$$

In other words, $\|A\|$ is the smallest real number such that

$$|A_{ij}| \le \|A\| \quad \text{for all } i,j. \tag{8.21}$$

This is one of several possible definitions of the norm. Its basic property is as follows:

(8.22) **Lemma.** Let A, B be complex $n \times n$ matrices. Then $\|AB\| \le n\|A\|\|B\|$, and $\|A^k\| \le n^{k-1}\|A\|^k$ for all $k > 0$.

Proof. We estimate the size of the i, j-entry of AB:

$$|(AB)_{ij}| = \left|\sum_{\nu} A_{i\nu}B_{\nu j}\right| \le \sum_{\nu=1}^{n} |A_{i\nu}||B_{\nu j}| \le n\|A\|\|B\|.$$

Thus $\|AB\| \le n\|A\|\|B\|$. The second inequality follows by induction from the first inequality. □

Proof of Proposition (8.3). To prove that the matrix exponential converges absolutely, we estimate the series as follows: Let $a = n\|A\|$. Then

$$(8.23) \qquad |(e^A)_{ij}| \le |I_{ij}| + |A_{ij}| + \frac{1}{2!}|(A^2)_{ij}| + \frac{1}{3!}|(A^3)_{ij}| + \cdots$$
$$\le 1 + \|A\| + \frac{1}{2!}n\|A\|^2 + \frac{1}{3!}n^2\|A\|^3 + \cdots$$
$$= 1 + (a + \frac{1}{2!}a^2 + \frac{1}{3!}a^3 + \cdots)/n = 1 + (e^a - 1)/n. \ \square$$

Proof of Proposition (8.9).

(a) The terms of degree k in the expansions of (8.8) are

$$(x+y)^k/k! = \sum_{r+s=k}\binom{k}{r}x^r y^s/k! \quad \text{and} \quad \sum_{r+s=k}\frac{x^r}{r!}\frac{y^s}{s!}.$$

To show that these terms are equal, we have to show that

$$\binom{k}{r}/k! = \frac{1}{r!s!} \quad \text{or} \quad \binom{k}{r} = \frac{k!}{r!s!},$$

for all k and all r, s such that $r + s = k$. This is a standard formula for binomial coefficients.

(b) Denote by $S_n(x)$ the partial sum $1 + x/1! + x^2/2! + \cdots + x^n/n!$. Then

$$S_n(x)S_n(y) = (1 + x/1! + x^2/2! + \cdots + x^n/n!)(1 + y/1! + y^2/2! + \cdots + y^n/n!)$$
$$= \sum_{r,s=0}^{n}\frac{x^r}{r!}\frac{y^s}{s!},$$

while

$$S_n(x+y) = (1 + (x+y)/1! + (x+y)^2/2! + \cdots + (x+y)^n/n!)$$
$$= \sum_{k=0}^{n}\sum_{r+s=k}\binom{k}{r}x^r y^s/k! = \sum_{k=0}^{n}\sum_{r+s=k}\frac{x^r}{r!}\frac{y^s}{s!}.$$

Comparing terms, we find that the expansion of the partial sum $S_n(x + y)$ consists of the terms in $S_n(x)S_n(y)$ such that $r + s \le n$. The same is true when we substitute commuting matrices A, B for x, y. We must show that the sum of the remaining terms tends to zero as $k \longrightarrow \infty$.

(8.24) **Lemma.** The series $\displaystyle\sum_k \sum_{r+s=k}\left|\left(\frac{A^r}{r!}\frac{B^s}{s!}\right)_{ij}\right|$ converges for all i, j.

Proof. Let $a = n\|A\|$ and $b = n\|B\|$. We estimate the terms in the sum. According to (8.22), $|(A^rB^s)_{ij}| \le n(n^{r-1}\|A\|^r)(n^{s-1}\|B\|^s) \le a^r b^s$. Therefore

$$\sum_k \sum_{r+s=k}\left|\left(\frac{A^r}{r!}\frac{B^s}{s!}\right)_{ij}\right| \le \sum_k\sum_{r+s=k}\frac{a^r}{r!}\frac{b^s}{s!} = e^{a+b}.$$

The proposition follows from this lemma because, on the one hand, the i, j-entry of

$$(S_k(A)S_k(B) - S_k(A+B))_{ij} \text{ is bounded by } \sum_{r+s>k} \left| \left(\frac{A^r}{r!} \frac{B^s}{s!} \right)_{ij} \right|.$$

According to the lemma, this sum tends to zero as $k \longrightarrow \infty$. And on the other hand,

$$(S_k(A)S_k(B) - S_k(A+B)) \longrightarrow (e^A e^B - e^{A+B}). \;\square$$

Proof of Proposition (8.13). By definition,

$$\frac{d}{dt}(e^{tA}) = \lim_{h \to 0} \frac{e^{(t+h)A} - e^{tA}}{h}.$$

Since the matrices tA and hA commute, the Proposition (8.9) shows that

$$\frac{e^{(t+h)A} - e^{tA}}{h} = \left(\frac{e^{hA} - I}{h} \right) e^{tA}.$$

So our proposition follows from this lemma:

(8.25) **Lemma.** $\lim_{h \to 0} \dfrac{e^{hA} - I}{h} = A.$

Proof. The series expansion for the exponential shows that

$$\frac{e^{hA} - I}{h} - A = \frac{h}{2!}A^2 + \frac{h^2}{3!}A^3 + \cdots. \tag{8.26}$$

We estimate this series: Let $a = |h| n \|A\|$. Then

$$\left| \left(\frac{h}{2!}A^2 + \frac{h^2}{3!}A^3 + \cdots \right)_{ij} \right| \leq \left| \frac{h}{2!}(A^2)_{ij} \right| + \left| \frac{h^2}{3!}(A^3)_{ij} \right| + \cdots$$

$$\leq \frac{1}{2!}|h| n \|A\|^2 + \frac{1}{3!}|h|^2 n^2 \|A\|^3 + \cdots = \|A\| \left(\frac{1}{2!}a + \frac{1}{3!}a^2 + \cdots \right)$$

$$= \frac{\|A\|}{a}(e^a - 1 - a) = \|A\| \left(\frac{e^a - 1}{a} - 1 \right).$$

Note that $a \longrightarrow 0$ as $h \longrightarrow 0$. Since the derivative of e^x is e^x,

$$\lim_{a \to 0} \frac{e^a - 1}{a} = \frac{d}{dx} e^x \bigg|_{x=0} = e^0 = 1.$$

So (8.26) tends to zero with h. $\square$

We will use the remarkable properties of the matrix exponential again, in Chapter 8.

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 $\varphi(X) = v_1 x_1 + \cdots + v_n x_n$ 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 $W^i = \operatorname{im} T^i$.
   (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 $\frac{d}{dx}$, considered as a linear operator on P.

(a) Find the matrix of D with respect to a convenient basis, and prove that D is a nilpotent operator.

(b) Determine all the D-invariant subspaces.

10. Prove that the matrices $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$ and $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ $(b \neq 0)$ are similar if and only if $a \neq d$.

11. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a real 2×2 matrix. Prove that A can be reduced to a matrix $\begin{bmatrix} 0 & * \\ 1 & * \end{bmatrix}$ by row and column operations of the form $A \longrightarrow EAE^{-1}$, unless $b = c = 0$ and $a = d$. Make a careful case analysis to take care of the possibility that b or c is zero.

12. Let T be a linear operator on $\mathbb{R}^2$ with two linearly independent eigenvectors v_1, v_2. Assume that the eigenvalues c_1, c_2 of these operators are positive and that $c_1 > c_2$. Let ℓ_i be the line spanned by v_i.

(a) The operator T carries every line ℓ through the origin to another line. Using the parallelogram law for vector addition, show that every line $\ell \neq \ell_2$ is shifted away from ℓ_2 toward ℓ_1.

(b) Use (a) to prove that the only eigenvectors are multiples of v_1 or v_2.

(c) Describe the effect on lines when there is a single line carried to itself, with positive eigenvalue.

13. Consider an arbitrary 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The condition that a column vector X be an eigenvector for left multiplication by A is that $Y = AX$ be parallel to X, which means that the slopes $s = x_2/x_1$ and $s' = y_2/y_1$ are equal.

(a) Find the equation in s which expresses this equality.

(b) For which A is $s = 0$ a solution? $s = \infty$?

(c) Prove that if the entries of A are positive real numbers, then there is an eigenvector in the first quadrant and also one in the second quadrant.

4. *The Characteristic Polynomial*

1. Compute the characteristic polynomials, eigenvalues, and eigenvectors of the following complex matrices.

(a) $\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$

2. (a) Prove that the eigenvalues of a real symmetric 2×2 matrix are real numbers.

(b) Prove that a real 2×2 matrix whose off-diagonal entries are positive has real eigenvalues.

3. Find the complex eigenvalues and eigenvectors of the notation matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

4. Prove that a real 3×3 matrix has at least one real eigenvalue.

5. Determine the characteristic polynomial of the matrix

$$\begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & 0 \end{bmatrix}.$$

6. Prove Proposition (4.18).
7. (a) Let T be a linear operator having two linearly independent eigenvectors with the same eigenvalue λ. Is it true that λ is a multiple root of the characteristic polynomial of T?
   (b) Suppose that λ is a multiple root of the characteristic polynomial. Does T have two linearly independent eigenvectors with eigenvalue λ?
8. Let V be a vector space with basis $(v_1,\dots,v_n)$ over a field F, and let $a_1,\dots,a_{n-1}$ be elements of F. Define a linear operator on V by the rules $T(v_i) = v_{i+1}$ if $i < n$ and $T(v_n) = a_1v_1 + a_2v_2 + \cdots + a_{n-1}v_{n-1}$.
   (a) Determine the matrix of T with respect to the given basis.
   (b) Determine the characteristic polynomial of T.
9. Do A and A^t have the same eigenvalues? the same eigenvectors?
10. (a) Use the characteristic polynomial to prove that a 2×2 real matrix P all of whose entries are positive has two distinct real eigenvalues.
    (b) Prove that the larger eigenvalue has an eigenvector in the first quadrant, and the smaller eigenvalue has an eigenvector in the second quadrant.
11. (a) Let A be a 3×3 matrix, with characteristic polynomial

$$p(t) = t^3 - (\operatorname{tr} A)t^2 + s_1 t - (\det A).$$

Prove that s_1 is the sum of the symmetric 2×2 subdeterminants:

$$s_1 = \det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \det\begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} + \det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

   *(b) Generalize to $n \times n$ matrices.
12. Let T be a linear operator on a space of dimension n, with eigenvalues $\lambda_1,\dots,\lambda_n$.
    (a) Prove that $\operatorname{tr} T = \lambda_1 + \cdots + \lambda_n$ and that $\det T = \lambda_1 \cdots \lambda_n$.
    (b) Determine the other coefficients of the characteristic polynomial in terms of the eigenvalues.

*13. Consider the linear operator of left multiplication of an $n \times n$ matrix A on the space $F^{n \times n}$ of all $n \times n$ matrices. Compute the trace and the determinant of this operator.

*14. Let P be a real matrix such that $P^t = P^2$. What are the possible eigenvalues of P?

15. Let A be a matrix such that $A^n = I$. Prove that the eigenvalues of A are powers of nth root of unity $\zeta_n = e^{2\pi i/n}$.

5. Orthogonal Matrices and Rotations

1. What is the matrix of the three-dimensional rotation through the angle θ about the axis e_2?
2. Prove that every orthonormal set of n vectors in $\mathbb{R}^n$ is a basis.
3. Prove algebraically that a real 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ represents a rotation if and only if it is in SO_2.
4. (a) Prove that O_n and SO_n are subgroups of $GL_n(\mathbb{R})$, and determine the index of SO_n in O_n.
   (b) Is O_2 isomorphic to the product group $SO_2 \times \{\pm I\}$? Is O_3 isomorphic to $SO_3 \times \{\pm I\}$?

5. What are the eigenvalues of the matrix A which represents the rotation of $\mathbb{R}^3$ by θ about an axis v?
6. Let A be a matrix in O_3 whose determinant is -1. Prove that -1 is an eigenvalue of A.
7. Let A be an orthogonal 2×2 matrix whose determinant is -1. Prove that A represents a reflection about a line through the origin.
8. Let A be an element of SO_3, with angle of rotation θ. Show that $\cos\theta = \frac{1}{2}(\operatorname{tr} A - 1)$.
9. Every real polynomial of degree 3 has a real root. Use this fact to give a less tricky proof of Lemma (5.23).

*10. Find a geometric way to determine the axis of rotation for the composition of two three-dimensional rotations.

11. Let v be a vector of unit length, and let P be the plane in $\mathbb{R}^3$ orthogonal to v. Describe a bijective correspondence between points on the unit circle in P and matrices $P \in SO_3$ whose first column is v.
12. Describe geometrically the action of an orthogonal matrix with determinant -1.
13. Prove that a rigid motion, as defined by (5.15), is bijective.

*14. Let A be an element of SO_3. Show that if it is defined, the vector

$$((a_{23} + a_{32})^{-1}, (a_{13} + a_{31})^{-1}, (a_{12} + a_{21})^{-1})^t$$

is an eigenvector with eigenvalue 1.

6. Diagonalization

1. (a) Find the eigenvectors and eigenvalues of the matrix
$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$
(b) Find a matrix P such that PAP^{-1} is diagonal.
(c) Compute $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{30}$.
2. Diagonalize the rotation matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, using complex numbers.
3. Prove that if A, B are $n \times n$ matrices and if A is nonsingular, then AB is similar to BA.
4. Let A be a complex matrix having zero as its only eigenvalue. Prove or disprove: A is nilpotent.
5. In each case, if the matrix is diagonalizable, find a matrix P such that PAP^{-1} is diagonal.
(a) $\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$
6. Can the diagonalization (6.1) be done with a matrix $P \in SL_n$?
7. Prove that a linear operator T is nilpotent if and only if there is a basis of V such that the matrix of T is upper triangular, with diagonal entries zero.
8. Let T be a linear operator on a space of dimension 2. Assume that the characteristic polynomial of T is $(t - a)^2$. Prove that there is a basis of V such that the matrix of T has one of the two forms $\begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}, \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$.

9. Let A be a nilpotent matrix. Prove that $\det(I + A) = 1$.

10. Prove that if A is a nilpotent $n \times n$ matrix, then $A^n = 0$.

11. Find all real 2×2 matrices such that $A^2 = I$, and describe geometrically the way they operate by left multiplication on $\mathbb{R}^2$.

12. Let M be a matrix made up of two diagonal blocks: $M = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$. Prove that M is diagonalizable if and only if A and D are.

13. (a) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix with eigenvalue λ. Show that $(b, \lambda - a)^t$ is an eigenvector for A.

(b) Find a matrix P such that PAP^{-1} is diagonal, if A has two distinct eigenvalues $\lambda_1 \neq \lambda_2$.

14. Let A be a complex $n \times n$ matrix. Prove that there is a matrix B arbitrarily close to A (meaning that $|b_{ij} - a_{ij}|$ can be made arbitrarily small for all i, j) such that B has n distinct eigenvalues.

***15.** Let A be a complex $n \times n$ matrix with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Assume that λ_1 is the largest eigenvalue, that is, that $|\lambda_1| > |\lambda_i|$ for all $i > 1$. Prove that for most vectors X the sequence $X_k = \lambda_1^{-k} A^k X$ converges to an eigenvector Y with eigenvalue λ_1, and describe precisely what the conditions on X are for this to be the case.

16. (a) Use the method of the previous problem to compute the largest eigenvalue of the matrix $\begin{bmatrix} 3 & 1 \\ 3 & 4 \end{bmatrix}$ to three-place accuracy.

(b) Compute the largest eigenvalue of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ to three-place accuracy.

***17.** Let A be $m \times m$ and B be $n \times n$ complex matrices, and consider the linear operator T on the space $F^{m \times n}$ of all complex matrices defined by $T(M) = AMB$.

(a) Show how to construct an eigenvector for T out of a pair of column vectors X, Y, where X is an eigenvector for A and Y is an eigenvector for B^t.

(b) Determine the eigenvalues of T in terms of those of A and B.

***18.** Let A be an $n \times n$ complex matrix.

(a) Consider the linear operator T defined on the space $F^{n \times n}$ of all complex $n \times n$ matrices by the rule $T(B) = AB - BA$. Prove that the rank of this operator is at most $n^2 - n$.

(b) Determine the eigenvalues of T in terms of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A.

7. Systems of Differential Equations

1. Let v be an eigenvector for the matrix A, with eigenvalue c. Prove that $e^{ct}v$ solves the differential equation $\frac{dX}{dt} = AX$.

2. Solve the equation $\frac{dX}{dt} = AX$ for the following matrices A:

(a) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ (e) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

3. Explain why diagonalization gives the general solution.

4. (a) Prove Proposition (7.16).
   (b) Why is it enough to write down the real and imaginary parts to get the general solution?
5. Prove Lemma (7.25).
6. Solve the inhomogeneous differential equation $\frac{dX}{dt} = AX + B$ in terms of the solutions to the homogeneous equation $\frac{dX}{dt} = AX$.
7. A differential equation of the form $d^nx/dt^n + a_{n-1}d^{n-1}x/dt^{n-1} + \cdots + a_1dx/dt + a_0x = 0$ can be rewritten as a system of first-order equations by the following trick: We introduce unknown functions $x_0, x_1, \ldots, x_{n-1}$ with $x = x_0$, and we set $dx_i/dt = x_{i+1}$ for $i = 0, \ldots, n-2$. The original equation can be rewritten as the system $dx_i/dt = x_{i+1}$, $i = 0, \ldots, n-2$, and $dx_{n-1}/dt = -(a_{n-1}x_{n-1} + \cdots + a_1x_1 + a_0x)$. Determine the matrix which represents this system of equations.
8. (a) Rewrite the second-order linear equation in one variable

$$\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = 0$$

   as a system of two first-order equations in two unknowns $x_0 = x$, $x_1 = dx/dt$.
   (b) Solve the system when $b = -4$ and $c = 3$.
9. Let A be an $n \times n$ matrix, and let $B(t)$ be a column vector of continuous functions on the interval $[\alpha, \beta]$. Define $F(t) = \int_\alpha^t e^{-tA} B(t)\, dt$.
   (a) Prove that $X = F(t)$ is a solution of the differential equation $X' = AX + B(t)$ on the interval (α, β).
   (b) Determine all solutions of this equation on the interval.

8. The Matrix Exponential

1. Compute e^A for the following matrices A:
   (a) $\begin{bmatrix} & 1 \\ -1 & \end{bmatrix}$ (b) $\begin{bmatrix} a & b \\ & \end{bmatrix}$
2. Let $A = \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix}$.
   (a) Compute e^A directly from the expansion.
   (b) Compute e^A by diagonalizing the matrix.
3. Compute e^A for the following matrices A:
   (a) $\begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 1 & & & \\ & \cdot & 1 & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & 0 \end{bmatrix}$
4. Compute e^A for the following matrices A:
   (a) $\begin{bmatrix} 2\pi i & 2\pi i \\ & 2\pi i \end{bmatrix}$ (b) $\begin{bmatrix} 6\pi i & 4\pi i \\ 2\pi i & 8\pi i \end{bmatrix}$
5. Let A be an $n \times n$ matrix. Prove that the map $t \rightsquigarrow e^{tA}$ is a homomorphism from the additive group $\mathbb{R}^+$ to $GL_n(\mathbb{C})$.

6. Find two matrices A, B such that $e^{A+B} \neq e^A e^B$.
7. Prove the formula $e^{\text{trace } A} = \det(e^A)$.
8. Solve the differential equation $\frac{dX}{dt} = AX$, when $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$.
9. Let $f(t)$ be a polynomial, and let T be a linear operator. Prove that $f(T)$ is a linear operator.
10. Let A be a symmetric matrix, and let $f(t)$ be a polynomial. Prove that $f(A)$ is symmetric.
11. Prove the product rule for differentiation of matrix-valued functions.
12. Let $A(t)$, $B(t)$ be differentiable matrix-valued functions of t. Compute the following.
    (a) $d/dt\,(A(t)^3)$
    (b) $d/dt(A(t)^{-1})$, assuming that $A(t)$ is invertible for all t
    (c) $d/dt(A(t)^{-1}B(t))$
13. Let X be an eigenvector of an $n \times n$ matrix A, with eigenvalue λ.
    (a) Prove that if A is invertible then X is also an eigenvector for A^{-1}, and that its eigenvalue is λ^{-1}.
    (b) Let $p(t)$ be a polynomial. Then X is an eigenvector for $p(A)$, with eigenvalue $p(\lambda)$.
    (c) Prove that X is an eigenvector for e^A, with eigenvalue e^λ.
14. For an $n \times n$ matrix A, define $\sin A$ and $\cos A$ by using the Taylor's series expansions for $\sin x$ and $\cos x$.
    (a) Prove that these series converge for all A.
    (b) Prove that $\sin tA$ is a differentiable function of t and that $d(\sin tA)/dt = A \cos tA$.
15. Discuss the range of validity of the following identities.
    (a) $\cos^2 A + \sin^2 A = I$
    (b) $e^{iA} = \cos A + i \sin A$
    (c) $\sin(A + B) = \sin A \cos B + \cos A \sin B$
    (d) $\cos(A + B) = \cos A \cos B - \sin A \sin B$
    (e) $e^{2\pi i A} = I$
    (f) $d(e^{A(t)})/dt = e^{A(t)} A'(t)$, where $A(t)$ is a differentiable matrix-valued function of t.
16. (a) Derive the product rule for differentiation of complex-valued functions in two ways: directly, and by writing $x(t) = u(t) + iv(t)$ and applying the product rule for real-valued functions.
    (b) Let $f(t)$ be a complex-valued function of a real variable t, and let $\varphi(u)$ be a real-valued function of u. State and prove the chain rule for $f(\varphi(u))$.
17. (a) Let B_k be a sequence of $m \times n$ matrices which converges to a matrix B, and let P be an $m \times m$ matrix. Prove that PB_k converges to PB.
    (b) Prove that if $m = n$ and P is invertible, then PB_kP^{-1} converges to PBP^{-1}.
18. Let $f(x) = \Sigma c_k x^k$ be a power series such that $\Sigma c_k A^k$ converges when A is a sufficiently small $n \times n$ matrix. Prove that A and $f(A)$ commute.
19. Determine $\frac{d}{dt} \det A(t)$, when $A(t)$ is a differentiable matrix function of t.

Miscellaneous Problems

1. What are the possible eigenvalues of a linear operator T such that (a) $T^r = I$, (b) $T^r = 0$, (c) $T^2 - 5T + 6 = 0$?

2. A linear operator T is called nilpotent if some power of T is zero.
  (a) Prove that T is nilpotent if and only if its characteristic polynomial is t^n, $n = \dim V$.
  (b) Prove that if T is a nilpotent operator on a vector space of dimension n, then $T^n = 0$.
  (c) A linear operator T is called *unipotent* if $T - I$ is nilpotent. Determine the characteristic polynomial of a unipotent operator. What are its possible eigenvalues?

3. Let A be an $n \times n$ complex matrix. Prove that if trace $A^i = 0$ for all i, then A is nilpotent.

***4.** Let A, B be complex $n \times n$ matrices, and let $C = AB - BA$. Prove that if C commutes with A then C is nilpotent.

5. Let $\lambda_1, \ldots, \lambda_n$ be the roots of the characteristic polynomial $p(t)$ of a complex matrix A. Prove the formulas trace $A = \lambda_1 + \cdots + \lambda_n$ and $\det A = \lambda_1 \cdots \lambda_n$.

6. Let T be a linear operator on a real vector space V such that $T^2 = I$. Define subspaces as follows:

$$W^+ = \{v \in V \mid T(v) = v\}, \quad W^- = \{v \in V \mid T(v) = -v\}.$$

Prove that V is isomorphic to the direct sum $W^+ \oplus W^-$.

7. The *Frobenius norm* $|A|$ of an $n \times n$ matrix A is defined to be the length of A when it is considered as an n^2-dimensional vector: $|A|^2 = \Sigma\, |a_{ij}|^2$. Prove the following inequalities: $|A + B| \leq |A| + |B|$ and $|AB| \leq |A|\,|B|$.

8. Let $T: V \longrightarrow V$ be a linear operator on a finite-dimensional vector space V. Prove that there is an integer n so that $(\ker T^n) \cap (\operatorname{im} T^n) = 0$.

9. Which infinite matrices represent linear operators on the space Z [Chapter 3 (5.2d)]?

***10.** The $k \times k$ *minors* of an $m \times n$ matrix A are the square submatrices obtained by crossing out $m - k$ rows and $n - k$ columns. Let A be a matrix of rank r. Prove that some $r \times r$ minor is invertible and that no $(r + 1) \times (r + 1)$ minor is invertible.

11. Let $\varphi: F^n \longrightarrow F^m$ be left multiplication by an $m \times n$ matrix A. Prove that the following are equivalent.
  (a) A has a right inverse, a matrix B such that $AB = I$.
  (b) φ is surjective.
  (c) There is an $m \times m$ minor of A whose determinant is not zero.

12. Let $\varphi: F^n \longrightarrow F^m$ be left multiplication by an $m \times n$ matrix A. Prove that the following are equivalent.
  (a) A has a left inverse, a matrix B such that $BA = I$.
  (b) φ is injective.
  (c) There is an $n \times n$ *minor* of A whose determinant is not zero.

***13.** Let A be an $n \times n$ matrix such that $A^r = I$. Prove that if A has only one eigenvalue ζ, then $A = \zeta I$.

14. **(a)** Without using the characteristic polynomial, prove that a linear operator on a vector space of dimension n can have at most n different eigenvalues.
  (b) Use (a) to prove that a polynomial of degree n with coefficients in a field F has at most n roots in F.

15. Let A be an $n \times n$ matrix, and let $p(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$ be its characteristic polynomial. The *Cayley–Hamilton Theorem* asserts that

$$p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = 0.$$

  (a) Prove the Cayley–Hamilton Theorem for 2×2 matrices.
  (b) Prove it for diagonal matrices.

(c) Prove it for diagonalizable matrices.

*(d) Show that every complex $n \times n$ matrix is arbitrarily close to a diagonalizable matrix, and use this fact to extend the proof for diagonalizable matrices to all complex matrices by continuity.

16. (a) Use the Cayley–Hamilton Theorem to give an expression for A^{-1} in terms of A, $(\det A)^{-1}$, and the coefficients of the characteristic polynomial.

(b) Verify this expression in the 2×2 case by direct computation.

***17.** Let A be a 2×2 matrix. The Cayley–Hamilton Theorem allows all powers of A to be written as linear combinations of I and A. Therefore it is plausible that e^A is also such a linear combination.

(a) Prove that if a, b are the eigenvalues of A and if $a \neq b$, then

$$e^A = \frac{ae^b - be^a}{a - b} I + \frac{e^a - e^b}{a - b} A.$$

(b) Find the correct formula for the case that A has two equal eigenvalues.

18. The Fibonacci numbers $0, 1, 1, 2, 3, 5, 8, \ldots$ are defined by the recursive relations $f_n = f_{n-1} + f_{n-2}$, with the initial conditions $f_0 = 0, f_1 = 1$. This recursive relation can be written in matrix form as

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} f_{n-2} \\ f_{n-1} \end{bmatrix} = \begin{bmatrix} f_{n-1} \\ f_n \end{bmatrix}.$$

(a) Prove the formula

$$f_n = \frac{1}{\alpha}\left[\left(\frac{1 + \alpha}{2}\right)^n - \left(\frac{1 - \alpha}{2}\right)^n\right],$$

where $\alpha = \sqrt{5}$.

(b) Suppose that the sequence a_n is defined by the relation $a_n = \frac{1}{2}(a_{n-1} + a_{n-2})$. Compute $\lim a_n$ in terms of a_0, a_1.

***19.** Let A be an $n \times n$ real positive matrix, and let $X \in \mathbb{R}^n$ be a column vector. Let us use the shorthand notation $X > 0$ or $X \geq 0$ to mean that all entries of the vector X are positive or nonnegative, respectively. By "positive quadrant" we mean the set of vectors $X \geq 0$. (But note that $X \geq 0$ and $X \neq 0$ do not imply $X > 0$ in our sense.)

(a) Prove that if $X \geq 0$ and $X \neq 0$ then $AX > 0$.

(b) Let C denote the set of pairs (X, t), $t \in \mathbb{R}$, such that $X \geq 0$, $|X| = 1$, and $(A - tI)X \geq 0$. Prove that C is a compact set in $\mathbb{R}^{n+1}$.

(c) The function t takes on a maximum value on C, say at the point (X_0, t_0). Then $(A - t_0 I)X_0 \geq 0$. Prove that $(A - t_0 I)X_0 = 0$.

(d) Prove that X_0 is an eigenvector with eigenvalue t_0 by showing that otherwise the vector $AX_0 = X_1$ would contradict the maximality of t_0.

(e) Prove that t_0 is the eigenvalue of A with largest absolute value.

***20.** Let $A = A(t)$ be a matrix of functions. What goes wrong when you try to prove that, in analogy with $n = 1$, the matrix

$$\exp\left(\int_{t_0}^{t} A(u)du\right)$$

is a solution of the system $dX/dt = AX$? Can you find conditions on the matrix function $A(t)$ which will make this a solution?