chapter III Determinants, eigenvalues, and similarity transformations

This chapter is devoted to the study of matrices representing linear transformations of a vector space into itself. We have seen that if A represents a linear transformation σ of V into itself with respect to a basis A, and P is the matrix of transition from A to a new basis A', then $P^{-1}AP = A'$ is the matrix representing σ with respect to A'. In this case A and A' are said to be similar and the mapping of A onto $A' = P^{-1}AP$ is called a similarity transformation (on the set of matrices, not on V).

Given σ, we seek a basis for which the matrix representing σ is particularly simple. In practice σ is given only implicitly by giving a matrix A representing σ. The problem, then, is to determine the matrix of transition P so that $P^{-1}AP$ has the desired form. The matrix representing σ has its simplest form whenever σ maps each basis vector onto a multiple of itself; that is, whenever for each basis vector α there exists a scalar λ such that $\sigma(\alpha) = \lambda\alpha$. It is not always possible to find such a basis, but there are some rather general conditions under which it is possible. These conditions include most cases of interest in the applications of this theory to physical problems.

The problem of finding non-zero α such that $\sigma(\alpha) = \lambda\alpha$ is equivalent to the problem of finding non-zero vectors in the kernel of $\sigma - \lambda$. This is a linear problem and we have given practical methods for solving it. But there is no non-zero solution to this problem unless $\sigma - \lambda$ is singular. Thus we are faced with the problem of finding those λ for which $\sigma - \lambda$ is singular. The values of λ for which $\sigma - \lambda$ is singular are called the eigenvalues of σ, and the non-zero vectors α for which $\sigma(\alpha) = \lambda\alpha$ are called eigenvectors of σ.

We introduce some topics from the theory of determinants solely for the purpose of finding the eigenvalues of a linear transformation. Were it not for this use of determinants we would not discuss them in this book. Thus, the treatment given them here is very brief.

Whenever a basis of eigenvectors exists, the use of determinants will provide a method for finding the eigenvalues and, knowing the eigenvalues, use of the Hermite normal form will enable us to find the eigenvectors. This method is convenient only for vector spaces of relatively small dimension. For numerical work with large matrices other methods are required.

The chapter closes with a discussion of what can be done if a basis of eigenvectors does not exist.

1 | Permutations

To define determinants and handle them we have to know something about permutations. Accordingly, we introduce permutations in a form most suitable for our purposes and develop their elementary properties.

A *permutation* π of a set S is a one-to-one mapping of S onto itself. We are dealing with permutations of finite sets and we take S to be the set of the first n integers; $S = \{1, 2, \ldots, n\}$. Let $\pi(i)$ denote the element which π associates with i. Whenever we wish to specify a particular permutation we describe it by writing the elements of S in two rows; the first row containing the elements of S in any order and the second row containing the element $\pi(i)$ directly below the element i in the first row. Thus for $S = \{1, 2, 3, 4\}$, the permutation π for which $\pi(1) = 2$, $\pi(2) = 4$, $\pi(3) = 3$, and $\pi(4) = 1$, can conveniently be described by the notations

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 2 & 4 & 1 & 3 \\ 4 & 1 & 2 & 3 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 4 & 1 & 3 & 2 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Two permutations acting on the same set of elements can be combined as functions. Thus, if π and σ are two permutations, $\sigma\pi$ will denote that permutation mapping i onto $\sigma[\pi(i)]$; $(\sigma\pi)(i) = \sigma[\pi(i)]$. As an example, let π denote the permutation described above and let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}.$$

Then

$$\sigma\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}.$$

Notice particularly that $\sigma\pi \neq \pi\sigma$.

If π and σ are two given permutations, there is a unique permutation ρ such that $\rho\pi = \sigma$. Since ρ must satisfy the condition that $\rho[\pi(i)] = \sigma(i)$, ρ can be described in our notation by writing the elements $\pi(i)$ in the first row and the elements $\sigma(i)$ in the second row. For the π and σ described above,

$$\rho = \begin{pmatrix} 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & 2 \end{pmatrix}.$$

The permutation that leaves all elements of S fixed is called the *identity permutation* and will be denoted by ϵ. For a given π the unique permutation π^{-1} such that $\pi^{-1}\pi = \epsilon$ is called the *inverse of* π.

If for a pair of elements $i < j$ in S we have $\pi(i) > \pi(j)$, we say that π performs an *inversion*. Let $k(\pi)$ denote the total number of inversions performed by π; we then say that π contains $k(\pi)$ inversions. For the permutation π described above, $k(\pi) = 4$. The number of inversions in π^{-1} is equal to the number of inversions in π.

For a permutation π, let sgn π denote the number $(-1)^{k(\pi)}$. "*Sgn*" is an abbreviation for "signum" and we use the term "sgn π" to mean "the sign of π." If sgn $\pi = 1$, we say that π is *even*; if sgn $\pi = -1$, we say that π is *odd*.

Theorem 1.1. Sgn $\sigma\pi = \text{sgn}\,\sigma \cdot \text{sgn}\,\pi$.

PROOF. σ can be represented in the form

$$\sigma = \begin{pmatrix} \cdots & \pi(i) & \cdots & \pi(j) & \cdots \\ \cdots & \sigma\pi(i) & \cdots & \sigma\pi(j) & \cdots \end{pmatrix}$$

because every element of S appears in the top row. Thus, in counting the inversions in σ it is sufficient to compare $\pi(i)$ and $\pi(j)$ with $\sigma\pi(i)$ and $\sigma\pi(j)$. For a given $i < j$ there are four possibilities:

1. $i < j$; $\pi(i) < \pi(j)$; $\sigma\pi(i) < \sigma\pi(j)$: no inversions.
2. $i < j$; $\pi(i) < \pi(j)$; $\sigma\pi(i) > \sigma\pi(j)$: one inversion in σ, one in $\sigma\pi$.
3. $i < j$; $\pi(i) > \pi(j)$; $\sigma\pi(i) > \sigma\pi(j)$: one inversion in π, one in $\sigma\pi$.
4. $i < j$; $\pi(i) > \pi(j)$; $\sigma\pi(i) < \sigma\pi(j)$: one inversion in π, one in σ, and none in $\sigma\pi$.

Examination of the above table shows that $k(\sigma\pi)$ differs from $k(\sigma) + k(\pi)$ by an even number. Thus sgn $\sigma\pi = \text{sgn}\,\sigma \cdot \text{sgn}\,\pi$. □

Theorem 1.2. *If a permutation π leaves an element of S fixed, the inversions involving that element need not be considered in determining whether π is even or odd.*

PROOF. Suppose $\pi(j) = j$. There are $j - 1$ elements of S less than j and $n - j$ elements of S larger than j. For $i < j$ an inversion occurs if and only if $\pi(i) > \pi(j) = j$. Let k be the number of elements i in S preceding j for which $\pi(i) > j$. Then there must also be exactly k elements i of S following j for which $\pi(i) < j$. It follows that there are $2k$ inversions involving j. Since their number is even they may be ignored in determining sgn π. □

Theorem 1.3. *A permutation which interchanges exactly two elements of S and leaves all other elements of S fixed is an odd permutation.*

PROOF. Let π be a permutation which interchanges the elements i and j and leaves all other elements of S fixed. According to Theorem 1.2, in determining sgn π we can ignore the inversions involving all elements of S other than i and j. There is just one inversion left to consider and sgn $\pi = -1$. □

Among other things, this shows that there is at least one odd permutation. In addition, there is at least one even permutation. From this it is but a step to show that the number of odd permutations is equal to the number of even permutations.

Let σ be a fixed odd permutation. If π is an even permutation, $\sigma\pi$ is odd. Furthermore, σ^{-1} is also odd so that to each odd permutation τ there corresponds an even permutation $\sigma^{-1}\tau$. Since $\sigma^{-1}(\sigma\pi) = \pi$, the mapping of the set of even permutations into the set of odd permutations defined by $\pi \to \sigma\pi$ is one-to-one and onto. Thus the number of odd permutations is equal to the number of even permutations.

EXERCISES

1. Show that there are $n!$ permutations of n objects.

2. There are six permutations of three objects. Determine which of them are even and which are odd.

3. There are 24 permutations of four objects. By use of Theorem 1.2 and Exercise 2 we can determine the parity (evenness or oddness) of 15 of these permutations without counting inversions. Determine the parity of these 15 permutations by this method and the parity of the remaining nine by any other method.

4. The nine permutations of four objects that leave no object fixed can be divided into two types of permutations, those that interchange two pairs of objects and those that permut the four objects in some cyclic order. There are three permutations of the first type and six of the second. Find them. Knowing the parity of the 15 permutations that leave at least one object fixed, as in Exercise 3, and that exactly half of the 24 permutations must be even, determine the parity of these nine.

5. By counting the inversions determine the parity of

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 5 & 1 & 3 \end{pmatrix}.$$

Notice that π permutes the objects in $\{1, 2, 4\}$ among themselves and the objects in $\{3, 5\}$ among themselves. Determine the parity of π on each of these subsets separately and deduce the parity of π on all of S.

2 | Determinants

Let $A = [a_{ij}]$ be a square $n \times n$ matrix. We wish to associate with this matrix a scalar that will in some sense measure the "size" of A and tell us whether or not A is non-singular.

Definition. The *determinant* of the matrix $A = [a_{ij}]$ is defined to be the scalar $\det A = |a_{ij}|$ computed according to the rule

$$\det A = |a_{ij}| = \sum_{\pi} (\operatorname{sgn} \pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}, \tag{2.1}$$

where the sum is taken over all permutations of the elements of $S = \{1, \ldots, n\}$. Each term of the sum is a product of n elements, each taken from a different row of A and from a different column of A, and $\operatorname{sgn} \pi$. The number n is called the *order* of the determinant.

As a direct application of this definition we see that

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{2.2}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{aligned} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} \\ - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32}. \end{aligned} \tag{2.3}$$

In general, a determinant of order n will be the sum of $n!$ products. As n increases, the amount of computation increases astronomically. Thus it is very desirable to develop more efficient ways of handling determinants.

Theorem 2.1. $\det A^T = \det A$.

PROOF. In the expansion of $\det A$ each term is of the form

$$(\operatorname{sgn} \pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}.$$

The factors of this term are ordered so that the indices of the rows appear in the usual order and the column indices appear in a permuted order. In the expansion of $\det A^T$ the same factors will appear but they will be ordered according to the row indices of A^T, that is, according to the column indices of A. Thus this same product will appear in the form

$$(\operatorname{sgn} \pi^{-1}) a_{\pi^{-1}(1),1} a_{\pi^{-1}(2),2} \cdots a_{\pi^{-1}(n),n}.$$

But since $\operatorname{sgn} \pi^{-1} = \operatorname{sgn} \pi$, this term is identical to the one given above. Thus, in fact, all the terms in the expansion of $\det A^T$ are equal to corresponding terms in the expansion of $\det A$, and $\det A^T = \det A$. □

A consequence of this discussion is that any property of determinants developed in terms of the rows (or columns) of A will also imply a corresponding property in terms of the columns (or rows) of A.

Theorem 2.2. *If A' is the matrix obtained from A by multiplying a row (or column) of A by a scalar c, then* $\det A' = c \det A$.

PROOF. Each term of the expansion of $\det A$ contains just one element from each row of A. Thus multiplying a row of A by c introduces the factor c into each term of $\det A$. Thus $\det A' = c \det A$. □

Theorem 2.3. *If A' is the matrix obtained from A by interchanging any two rows (or columns) of A, then* $\det A' = -\det A$.

PROOF. Interchanging two rows of A has the effect of interchanging two row indices of the elements appearing in A. If σ is the permutation interchanging these two indices, this operation has the effect of replacing each permutation π by the permutation $\pi\sigma$. Since σ is an odd permutation, this has the effect of changing the sign of every term in the expansion of $\det A$. Therefore, $\det A' = -\det A$. □

Theorem 2.4. *If A has two equal rows,* $\det A = 0$.

PROOF. The matrix obtained from A by interchanging the two equal rows is identical to A, and yet, by Theorem 2.3, this operation must change the sign of the determinant. Since the only number equal to its negative is 0 $\det A = 0$. □

Note: There is a minor point to be made here. If $1 + 1 = 0$, the proof of this theorem is not valid, but the theorem is still true. To see this we return our attention to the definition of a determinant. $\operatorname{Sgn} \pi = 1$ for both even and odd permutations. Then the terms in (2.1) can be grouped into pairs of equal terms. Since the sum of each pair is 0, the determinant is 0.

Theorem 2.5. *If A' is the matrix obtained from A by adding a multiple of one row (or column) to another, then* $\det A' = \det A$.

PROOF. Let A' be the matrix obtained from A by adding c times row k to row j. Then

$$\begin{aligned}\det A' &= \sum_{\pi} (\operatorname{sgn} \pi) a_{1\pi(1)} \cdots (a_{j\pi(j)} + c a_{k\pi(j)}) \cdots a_{k\pi(k)} \cdots a_{n\pi(n)} \\ &= \sum_{\pi} (\operatorname{sgn} \pi) a_{1\pi(1)} \cdots a_{j\pi(j)} \cdots a_{k\pi(k)} \cdots a_{n\pi(n)} \\ &\quad + c \sum_{\pi} (\operatorname{sgn} \pi) a_{1\pi(1)} \cdots a_{k\pi(j)} \cdots a_{k\pi(k)} \cdots a_{n\pi(n)}. \end{aligned} \tag{2.4}$$

The second sum on the right side of this equation is, in effect, the determinant of a matrix in which rows j and k are equal. Thus it is zero. The first term is just the expansion of $\det A$. Therefore, $\det A' = \det A$. □

It is evident from the definition that, if I is the identity matrix, $\det I = 1$.

If E is an elementary matrix of type I, $\det E = c$ where c is the scalar factor employed in the corresponding elementary operation. This follows from Theorem 2.2 applied to the identity matrix.

If E is an elementary matrix of type II, $\det E = 1$. This follows from Theorem 2.5 applied to the identity matrix.

If E is an elementary matrix of type III, $\det E = -1$. This follows from Theorem 2.3 applied to the identity matrix.

Theorem 2.6. *If E is an elementary matrix and A is any matrix, then* $\det EA = \det E \cdot \det A = \det AE$.

PROOF. This is an immediate consequence of Theorems 2.2, 2.5, 2.3, and the values of the determinants of the corresponding elementary matrices. □

Theorem 2.7. $\det A = 0$ *if and only if A is singular.*

PROOF. If A is non-singular, it is a product of elementary matrices (see Chapter II, Theorem 6.1). Repeated application of Theorem 2.6 shows that $\det A$ is equal to the product of the determinants of the corresponding elementary matrices, and hence is non-zero.

If A is singular, the rows are linearly dependent and one row is a linear combination of the others. By repeated application of elementary operations of type II we can obtain a matrix with a row of zeros. The determinant of this matrix is zero, and by Theorem 2.5 so also is $\det A$. □

Theorem 2.8. *If A and B are any two matrices of order n, then* $\det AB = \det A \cdot \det B = \det BA$.

PROOF. If A and B are non-singular, the theorem follows by repeated application of Theorem 2.6. If either matrix is singular, then AB and BA are also singular and all terms are zero. □

EXERCISES

1. If all elements of a matrix below the main diagonal are zero, the matrix is said to be in *superdiagonal* form; that is, $a_{ij} = 0$ for $i > j$. If $A = [a_{ij}]$ is in superdiagonal form, compute $\det A$.

2. Theorem 2.6 provides an effective and convenient way to evaluate determinants. Verify the following sequence of steps.

$$\begin{vmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & 4 & 1 \\ 3 & 2 & 2 \\ -2 & -4 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & 4 & 1 \\ 0 & -10 & -1 \\ 0 & 4 & 1 \end{vmatrix}$$

$$= -\begin{vmatrix} 1 & 4 & 1 \\ 0 & -2 & 1 \\ 0 & 4 & 1 \end{vmatrix} = -\begin{vmatrix} 1 & 4 & 1 \\ 0 & -2 & 1 \\ 0 & 0 & 3 \end{vmatrix}.$$

Now use the results of Exercise 1 to evaluate the last determinant.

3. Actually, to compute a determinant there is no need to obtain a superdiagonal form. And elementary column operations can be used as well as elementary row operations. Any sequence of steps that will result in a form with a large number of zero elements will be helpful. Verify the following sequence of steps.

$$\begin{vmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{vmatrix} = \begin{vmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -1 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 3 & 0 & 2 \\ 1 & 3 & 1 \\ -1 & 0 & 0 \end{vmatrix}.$$

This last determinant can be evaluated by direct use of the definition by computing just one product. Evaluate this determinant.

4. Evaluate the determinants:

$$(a) \begin{vmatrix} 1 & -2 & 2 \\ -1 & 3 & 1 \\ 2 & 5 & -1 \end{vmatrix} \qquad (b) \begin{vmatrix} 1 & 2 & 0 & 1 \\ 1 & 3 & 4 & 0 \\ 0 & 1 & 5 & 6 \\ 1 & 2 & 3 & 4 \end{vmatrix}$$

5. Consider the real plane R^2. We agree that the two points (a_1, a_2), (b_1, b_2) suffice to describe a quadrilateral with corners at $(0, 0)$, (a_1, a_2), (b_1, b_2), and $(a_1 + b_1, a_2 + b_2)$. (See Fig. 2.) Show that the area of this quadrilateral is

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

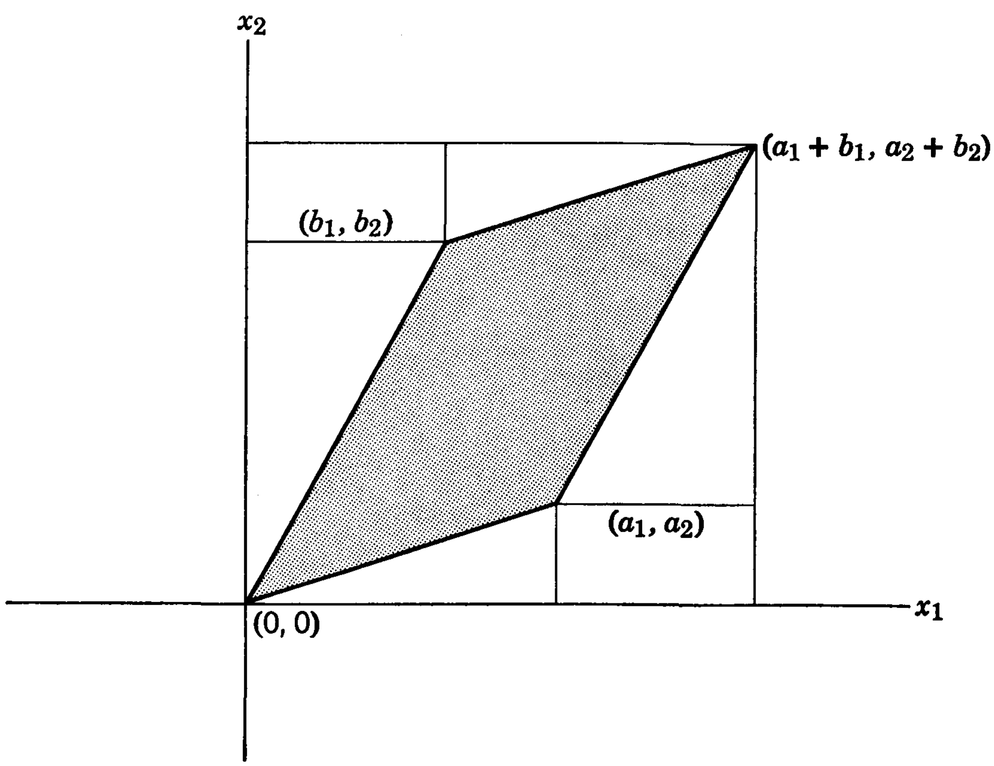

Fig. 2

Notice that the determinant can be positive or negative, and that it changes sign if the first and second rows are interchanged. To interpret the value of the determinant as an area, we must either use the absolute value of the determinant or give an interpretation to a negative area. We make the latter choice since to take the absolute value is to discard information. Referring to Fig. 2, we see that the direction of rotation from (a_1, a_2) to (b_1, b_2) across the enclosed area is the same as the direction of rotation from the positive x_1-axis to the positive x_2-axis. To interchange (a_1, a_2) and (b_1, b_2) would be to change the sense of rotation and the sign of the determinant. Thus the sign of the determinant determines an orientation of the quadrilateral on the coordinate system. Check the sign of the determinant for choices of (a_1, a_2) and (b_1, b_2) in various quadrants and various orientations.

6. (Continuation) Let E be an elementary transformation of R^2 onto itself. E maps the vertices of the given quadrilateral onto the vertices of another quadrilateral. Show that the area of the new quadrilateral is det E times the area of the old quadrilateral.

7. Let $x_1, \dots, x_n$ be a set of indeterminates. The determinant

$$V = \begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix}$$

is called the Vandermonde determinant of order n.

(*a*) Show that V is a polynomial of degree $n - 1$ in each indeterminate separately and of degree $n(n - 1)/2$ in all the indeterminates together.

(*b*) Show that, for each $i < j$, V is divisible by $x_j - x_i$.

(*c*) Show that $\prod_{1 \le i < j \le n} (x_j - x_i)$ is a polynomial of degree $n - 1$ in each indeterminate separately, and of degree $n(n - 1)/2$ in all the indeterminates together.

(*d*) Show that $V = \prod_{1 \le i < j \le n} (x_j - x_i)$.

3 | Cofactors

For a given pair i, j, consider in the expansion for det A those terms which have a_{ij} as a factor. Det A is of the form det $A = a_{ij}A_{ij} +$ (terms which do not contain a_{ij} as a factor). The scalar A_{ij} is called the *cofactor* of a_{ij}.

In particular, we see that $A_{11} = \sum_\pi (\operatorname{sgn} \pi) a_{2\pi(2)} \cdots a_{n\pi(n)}$ where this sum includes all permutations π that leave 1 fixed. Each such π defines a permutation π' on $S' = \{2, \dots, n\}$ which coincides with π on S. Since no inversion of π involves the element 1, we see that $\operatorname{sgn} \pi = \operatorname{sgn} \pi'$. Thus A_{ij} is a determinant, the determinant of the matrix obtained from A by crossing out the first row and the first column of A.

A similar procedure can be used to compute the cofactors A_{ij}. By a sequence of elementary row and column operations of type III we can obtain a matrix in which the element a_{ij} is moved into row 1, column 1. By applying the observation of the previous paragraph we see that the cofactor A_{ij} is essentially the determinant of the matrix obtained by crossing out the row and column containing the element a_{ij}. Furthermore, we can keep the other rows and columns in the same relative order if the sequence of operations we use interchanges only adjacent rows or columns. It takes $i - 1$ interchanges to move the element a_{ij} into the first row, and it takes $j - 1$ interchanges to move it into the first column. Thus A_{ij} is $(-1)^{i-1+j-1} = (-1)^{i+j}$ times the determinant of the matrix obtained by crossing out the ith row and the jth column of A.

Each term in the expansion of det A contains exactly one factor from each row and each column of A. Thus, for any given row of A each term of det A contains exactly one factor from that row. Hence, for any given i,

$$\det A = \sum_j a_{ij}A_{ij}. \tag{3.1}$$

Similarly, for any given column of A each term of det A contains exactly one factor from that column. Hence, for any given k,

$$\det A = \sum_j a_{jk}A_{jk}. \tag{3.2}$$

These expansions of a determinant according to the cofactors of a row or column reduce the problem of computing an nth order determinant to that of computing n determinants of order $n - 1$. We have already given explicit expansions for determinants of orders 2 and 3, and the technique of expansions according to cofactors enables us to compute determinants of higher orders. The labor of evaluating a determinant of even quite modest order is still quite formidable, however, and we make some suggestions as to how the work can be minimized.

First, observe that if any row or column has several zeros in it, expansion according to cofactors of that row or column will require the evaluation of only those cofactors corresponding to non-zero elements. It is clear that the presence of several zeros in any row or column would considerably reduce the labor. If we are not fortunate enough to find such a row or column, we can produce a row or column with a large number of zeros by applying some elementary operations of type II. For example, consider the determinant

$$\det A = \begin{vmatrix} 3 & 2 & -2 & 10 \\ 3 & 1 & 1 & 2 \\ -2 & 2 & 3 & 4 \\ 1 & 1 & 5 & 2 \end{vmatrix}.$$

If the numbers appearing in the array were unwieldy, there would be no choice but to wade in and make the best of it. The numbers in our example are all integers, and we will not introduce fractions if we take advantage of the 1's that appear in the array. By Theorem 2.5, a sequence of elementary operations of type II will not change the value of the determinant. Thus we can obtain

$$\det A = \begin{vmatrix} 0 & -1 & -17 & 4 \\ 0 & -2 & -14 & -4 \\ 0 & 4 & 13 & 8 \\ 1 & 1 & 5 & 2 \end{vmatrix} = -\begin{vmatrix} -1 & -17 & 4 \\ -2 & -14 & -4 \\ 4 & 13 & 8 \end{vmatrix}.$$

Now we face several options. We can expand the 3rd order determinants as it stands; we can try the same technique again; or we can try to remove a common factor from some row or column. We can remove the common factor -1 from the second row and the common factor 4 from the third column. Although 2 is factor of the second row, we cannot remove both a 2 from the second row and a 4 from the third column. Thus we can obtain

$$\det A = 4 \cdot \begin{vmatrix} -1 & -17 & 1 \\ 2 & 14 & 1 \\ 4 & 13 & 2 \end{vmatrix} = 4 \cdot \begin{vmatrix} -1 & -17 & 1 \\ 3 & 31 & 0 \\ 6 & 47 & 0 \end{vmatrix}$$

$$= 4 \cdot \begin{vmatrix} 3 & 31 \\ 6 & 47 \end{vmatrix} = -180.$$

If we multiply the elements in row i by the cofactors of the elements in row $k \neq i$, we get the same result as we would if the elements in row k were equal to the elements in row i. Hence,

$$\sum_j a_{ij} A_{kj} = 0 \qquad \text{for} \qquad i \neq k, \tag{3.3}$$

and

$$\sum_i a_{ij} A_{ik} = 0 \qquad \text{for} \qquad j \neq k. \tag{3.4}$$

The various relations we have developed between the elements of a matrix and their cofactors can be summarized in the form

$$\sum_j a_{ij} A_{kj} = \delta_{ik} \det A, \tag{3.5}$$

$$\sum_i a_{ij} A_{ik} = \delta_{jk} \det A. \tag{3.6}$$

If $A = [a_{ij}]$ is any square matrix and A_{ij} is the cofactor of a_{ij}, the matrix $[A_{ij}]^T = \operatorname{adj} A$ is called the *adjunct* of A. What we here call the "adjunct" is traditionallly called the "adjoint." Unfortunately, the term "adjoint" is also used to denote a linear transformation that is not represented by the adjoint (or adjunct) matrix. A new term is badly needed. We shall have a use for the adjunct matrix only in this chapter. Thus, this unconventional terminology will cause only a minor inconvenience and help to avoid confusion.

Theorem 3.1. $A \cdot \operatorname{adj} A = (\operatorname{adj} A) \cdot A = (\det A) \cdot I.$

PROOF.

$$A \cdot \operatorname{adj} A = [a_{ij}] \cdot [A_{kl}]^T = \left[\sum_j a_{ij}A_{kj}\right] = (\det A) \cdot I. \tag{3.7}$$

$$(\operatorname{adj} A) \cdot A = [A_{kl}]^T \cdot [a_{ij}] = \left[\sum_i A_{ik}a_{ij}\right] = (\det A) \cdot I. \ \Box \tag{3.8}$$

Theorem 3.1 provides us with an effective technique for computing the inverse of a non-singular matrix. However, it is effective only in the sense that the inverse can be computed by a prescribed sequence of steps. The number of steps is large for matrices of large order, and it is not sufficiently small for matrices of low order to make it a preferred technique. The method described in Section 6 of Chapter II is the best method that is developed in this text. In numerical analysis where matrices of large order are inverted, highly specialized methods are available. But a discussion of such methods is beyond the scope of this book.

A matrix A is non-singular if and only if $\det A \neq 0$, and in this case we can see from the theorem that

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A. \tag{3.9}$$

This is illustrated in the following example.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ -2 & 1 & -1 \end{bmatrix}, \qquad \operatorname{adj} A = \begin{bmatrix} -3 & 5 & 1 \\ -2 & 5 & 4 \\ 4 & -5 & -3 \end{bmatrix},$$

$$A^{-1} = \tfrac{1}{5}\begin{bmatrix} -3 & 5 & 1 \\ -2 & 5 & 4 \\ 4 & -5 & -3 \end{bmatrix}.$$

The relations between the elements of a matrix and their cofactors lead to a method for solving a system of n simultaneous equations in n unknowns when the equations are independent. Suppose we are given the system of equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \qquad (i = 1, 2, \ldots, n). \tag{3.10}$$

The assumption that the equations are independent is expressed in the condition that $\det A \neq 0$, where $A = [a_{ij}]$. Let A_{ij} be the cofactor of a_{ij}. Then for a given k

$$\begin{aligned}\sum_{i=1}^{n} A_{ik}\left(\sum_{j=1}^{n} a_{ij}x_j\right) &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} A_{ik}a_{ij}\right)x_j \\ &= \sum_{j=1}^{n} \det A\ \delta_{kj}x_j \\ &= \det A\ x_k = \sum_{i=1}^{n} A_{ik}b_i.\end{aligned} \tag{3.11}$$

Since $\det A \neq 0$ we see that

$$x_k = \frac{\sum_{i=1}^{n} A_{ik}b_i}{\det A}. \tag{3.12}$$

The numerator can be interpreted as the cofactor expansion of the determinant of the matrix obtained by replacing the kth column of A by the column of the b_i. In this form the method is known as *Cramer's rule.*

Cramer's rule is convenient for systems of equations of low order, but it fails if the system of equations is dependent or the number of equations is different from the number of unknowns. Even in these cases Cramer's rule can be modified to provide solutions. However, the methods we have already developed are usually easier to apply, and the balance in their favor increases as the order of the system of equations goes up and the nullity increases.

EXERCISES

1. In the determinant

$$\begin{vmatrix} 2 & 7 & 5 & 8 \\ 7 & -1 & 2 & 5 \\ 1 & 0 & 4 & 2 \\ -3 & 6 & -1 & 2 \end{vmatrix},$$

find the cofactor of the "8"; find the cofactor of the "-3."

2. The expansion of a determinant in terms of a row or column, as in formulas (3.1) and (3.2), provides a convenient method for evaluating determinants. The

amount of work involved can be reduced if a row or column is chosen in which some of the elements are zeros. Expand the determinant

$$\begin{vmatrix} 1 & 3 & 4 & -1 \\ 2 & 2 & 0 & 1 \\ 0 & -1 & 1 & 3 \\ -3 & 0 & 1 & 2 \end{vmatrix}$$

in terms of the cofactors of the third row.

3. It is even more convenient to combine an expansion in terms of cofactors with the method of elementary row and column operations described in Section 2. Subtract appropriate multiples of column 2 from the other columns to obtain

$$\begin{vmatrix} 1 & 3 & 7 & 8 \\ 2 & 2 & 2 & 7 \\ 0 & -1 & 0 & 0 \\ -3 & 0 & 1 & 2 \end{vmatrix}$$

and expand this determinant in terms of cofactors of the third row.

4. Show that $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

5. Show that a matrix is non-singular if and only if its adj A is also non-singular.

6. Let $A = [a_{ij}]$ be an arbitrary $n \times n$ matrix and let adj A be the adjunct of A. If $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ show that

$$Y^T(\operatorname{adj} A)X = -\begin{vmatrix} a_{11} & \cdots & a_{1n} & x_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{n1} & \cdots & a_{nn} & x_n \\ y_1 & \cdots & y_n & 0 \end{vmatrix}$$

For notation see pages 42 and 55.

4 | The Hamilton-Cayley Theorem

Let $p(x) = a_m x^m + \cdots + a_0$ be a polynomial in an indeterminate x with scalar coefficients a_i. If A is an $n \times n$ matrix, by $p(A)$ we mean the matrix $a_m A^m + a_{m-1}A^{m-1} + \cdots + a_0 I$. Notice particularly that the constant term a_0 must be replaced by $a_0 I$ so that each term of $p(A)$ will be a matrix. No particular problem is encountered with matric polynomials of this form since all powers of a single matrix commute with each other. Any polynomial identity will remain valid if the indeterminate is replaced by a matrix, provided any scalar terms are replaced by corresponding scalar multiples of the identity matrix.

We may also consider polynomials with matric coefficients. To make sense, all coefficients must be matrices of the same order. We consider only the possibility of substituting scalars for the indeterminate, and in all manipulations with such polynomials the matric coefficients commute with the powers of the indeterminate. Polynomials with matric coefficients can be added and multiplied in the usual way, but the order of the factors is important in multiplication since the coefficients may not commute. The algebra of polynomials of this type is not simple, but we need no more than the observation that two polynomials with matric coefficients are equal if and only if they have exactly the same coefficients.

We avoid discussing the complications that can occur for polynomials with matric coefficients in a matric variable.

Now we should like to consider matrices for which the elements are polynomials. If F is the field of scalars for the set of polynomials in the indeterminate x, let K be the set of all rational functions in x; that is, the set of all permissible quotients of polynomials in x. It is not difficult to show that K is a field. Thus a matrix with polynomial components is a special case of a matrix with elements in K.

From this point of view a polynomial with matric coefficients can be expressed as a single matrix with polynomial components. For example,

$$\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} x^2 + \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} x + \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} x^2 + 2 & 2x - 1 \\ -x^2 - 2x + 1 & 2x^2 + 1 \end{bmatrix}.$$

Conversely, a matrix in which the elements are polynomials in an indeterminate x can be expanded into a polynomial with matric coefficients. Since polynomials with matric coefficients and matrices with polynomial components can be converted into one another, we refer to both types of expressions as *polynomial matrices*.

Definition. If A is any square matrix, the polynomial matrix $A - xI = C$ is called the *characteristic matrix* of A.

C has the form

$$\begin{bmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{bmatrix}. \tag{4.1}$$

The determinant of C is a polynomial $\det C = f(x) = k_n x^n + k_{n-1}x^{n-1} + \cdots + k_0$ of degree n; it is called the *characteristic polynomial* of A. The equation $f(x) = 0$ is called the *characteristic equation* of A. First, we should observe that the coefficient of x^n in the characteristic polynomial is $(-1)^n$, the coefficient of x^{n-1} is $(-1)^{n-1} \sum_{i=1}^{n} a_{ii}$, and the constant term $k_0 = \det A$.

Theorem 4.1. *(Hamilton-Cayley theorem). If A is a square matrix and $f(x)$ is its characteristic polynomial, then $f(A) = 0$.*

PROOF. Since C is of order n, adj C will contain polynomials in x of degree not higher than $n - 1$. Hence adj C can be expanded into a polynomial with matric coefficients of degree at most $n - 1$:

$$\operatorname{adj} C = C_{n-1}x^{n-1} + C_{n-2}x^{n-2} + \cdots + C_1 x + C_0 \tag{4.2}$$

where each C_i is a matrix with scalar elements.

By Theorem 3.1 we have

$$\begin{aligned}\operatorname{adj} C \cdot C &= \det C \cdot I = f(x)I \\ &= \operatorname{adj} C \cdot (A - xI) = (\operatorname{adj} C)A - (\operatorname{adj} C)x.\end{aligned} \tag{4.3}$$

Hence,

$$\begin{aligned}k_n I x^n &+ k_{n-1} I x^{n-1} + \cdots + k_1 I x + k_0 I \\ &= -C_{n-1}x^n - C_{n-2}x^{n-1} - \cdots - C_0 x \\ &\qquad + C_{n-1}Ax^{n-1} + \cdots + C_1 A x + C_0 A.\end{aligned} \tag{4.4}$$

The expressions on the two sides of this equality are $n \times n$ polynomial matrices. Since two polynomial matrices are equal if and only if the corresponding coefficients are equal, (4.4) is equivalent to the following set of matric equations:

$$\begin{aligned}k_n I &= -C_{n-1} \\ k_{n-1} I &= -C_{n-2} + C_{n-1}A \\ &\;\;\vdots \\ k_1 I &= -C_0 + C_1 A \\ k_0 I &= \phantom{-C_0 + {}} C_0 A.\end{aligned} \tag{4.5}$$

Multiply each of these equations by $A^n, A^{n-1}, \ldots, A, I$ from the right, respectively, and add them. The terms on the right side will cancel out leaving the zero matrix. The terms on the left add up to

$$k_n A^n + k_{n-1}A^{n-1} + \cdots + k_1 A + k_0 I = f(A) = 0. \;\square \tag{4.6}$$

The equation $m(x) = 0$ of lowest degree which A satisfies is called the *minimum equation* (or *minimal equation*) for A; $m(x)$ is called the *minimum polynomial* for A. Since A satisfies its characteristic equation the degree of $m(x)$ is not more than n. Since a linear transformation and any matrix representing it satisfy the same relations, similar matrices satisfy the same set of polynomial equations. In particular, similar matrices have the same minimum polynomials.

Theorem 4.2. *If $g(x)$ is any polynomial with coefficients in F such that $g(A) = 0$, then $g(x)$ is divisible by the minimum polynomial for A. The minimum polynomial is unique except for a possible non-zero scalar factor.*

PROOF. Upon dividing $g(x)$ by $m(x)$ we can write $g(x)$ in the form

$$g(x) = m(x) \cdot q(x) + r(x), \tag{4.7}$$

where $q(x)$ is the quotient polynomial and $r(x)$ is the remainder, which is either identically zero or is a polynomial of degree less than the degree of $m(x)$. If $g(x)$ is a polynomial such that $g(A) = 0$, then

$$g(A) = 0 = m(A) \cdot q(A) + r(A) = r(A). \tag{4.8}$$

This would contradict the selection of $m(x)$ as the minimum polynomial for A unless the remainder $r(x)$ is identically zero. Since two polynomials of the same lowest degree must divide each other, they must differ by a scalar factor. □

As we have pointed out, the elements of adj C are polynomials of degree at most $n - 1$. Let $g(x)$ be the greatest common divisor of the elements of adj C. Since adj $C \cdot C = f(x)I$, $g(x)$ divides $f(x)$.

Theorem 4.3. $h(x) = \dfrac{f(x)}{g(x)}$ *is the minimum polynomial for A.*

PROOF. Let adj $C = g(x)B$ where the elements of B have no non-scalar common factor. Since adj $C \cdot C = f(x)I$ we have $h(x) \cdot g(x)I = g(x)BC$. Since $g(x) \not\equiv 0$ this yields

$$BC = h(x)I. \tag{4.9}$$

Using B in place of adj C we can repeat the argument used in the proof of the Hamilton-Cayley theorem to deduce that $h(A) = 0$. Thus $h(x)$ is divisible by $m(x)$.

On the other hand, consider the polynomial $m(x) - m(y)$. Since it is a sum of terms of the form $c_i(x^i - y^i)$, each of which is divisible by $y - x$, $m(x) - m(y)$ is divisible by $y - x$:

$$m(x) - m(y) = (y - x) \cdot k(x, y). \tag{4.10}$$

Replacing x by xI and y by A we have

$$m(xI) - m(A) = m(x)I = (A - xI) \cdot k(xI, A) = C \cdot k(xI, A). \tag{4.11}$$

Multiplying by adj C we have

$$m(x) \operatorname{adj} C = (\operatorname{adj} C)C \cdot k(xI, A) = f(x) \cdot k(xI, A). \tag{4.12}$$

Hence,

$$m(x) \cdot g(x)B = h(x) \cdot g(x) \cdot k(xI, A), \tag{4.13}$$

or

$$m(x)B = h(x) \cdot k(xI, A). \tag{4.14}$$

Since $h(x)$ divides every element of $m(x)B$ and the elements of B have no non-scalar common factor, $h(x)$ divides $m(x)$. Thus, $h(x)$ and $m(x)$ differ at most by a scalar factor. □

Theorem 4.4. *Each irreducible factor of the characteristic polynomial $f(x)$ of A is also an irreducible factor of the minimum polynomial $m(x)$.*

PROOF. As we have seen in the proof of the previous theorem

$$m(x)I = C \cdot k(xI, A).$$

Thus

$$\det m(x)I = [m(x)]^n = \det C \cdot \det k(xI, A)$$
$$= f(x) \cdot \det k(xI, A). \tag{4.15}$$

We see then that every irreducible factor of $f(x)$ divides $[m(x)]^n$, and therefore $m(x)$ itself. □

Theorem 4.4 shows that a characteristic polynomial without repeated factors is also the minimum polynomial. As we shall see, it is the case in which the characteristic polynomial has repeated factors that generally causes trouble.

We now ask the converse question. Given the polynomial $f(x) = (-1)^n x^n + k_{n-1}x^{n-1} + \cdots + k_0$, does there exist an $n \times n$ matrix A for which $f(x)$ is the minimum polynomial?

Let $\mathsf{A} = \{\alpha_1, \ldots, \alpha_n\}$ be any basis and define the linear transformation σ by the rules

$$\sigma(\alpha_i) = \alpha_{i+1} \qquad \text{for} \qquad i < n, \tag{4.16}$$

and

$$(-1)^n \sigma(\alpha_n) = -k_0\alpha_1 - k_1\alpha_2 - \cdots - k_{n-1}\alpha_n.$$

It follows directly from the definition of σ that

$$f(\sigma)(\alpha_1) = (-1)^n \sigma(\alpha_n) + k_{n-1}\alpha_n + \cdots + k_1\alpha_2 + k_0\alpha_1 = 0. \tag{4.17}$$

For any other basis element we have

$$f(\sigma)(\alpha_j) = f(\sigma)[\sigma^{j-1}(\alpha_1)] = \sigma^{j-1}[f(\sigma)(\alpha_1)] = 0. \tag{4.18}$$

Since $f(\sigma)$ vanishes on the basis elements $f(\sigma) = 0$ and any matrix representing σ satisfies the equation $f(x) = 0$.

On the other hand, σ cannot satisfy an equation of lower degree because the corresponding polynomial in σ applied to α_1 could be interpreted as a relation among the basis elements. Thus, $f(x)$ is a minimum polynomial for σ and for any matrix representing σ. Since $f(x)$ is of degree n, it must also be the characteristic polynomial of any matrix representing σ.

With respect to the basis A the matrix representing σ is

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 & -(-1)^n k_0 \\ 1 & 0 & \cdots & 0 & -(-1)^n k_1 \\ 0 & 1 & \cdots & 0 & -(-1)^n k_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & & 1 & -(-1)^n k_{n-1} \end{bmatrix}. \tag{4.19}$$

A is called the *companion matrix* of $f(x)$.

Theorem 4.5. *$f(x)$ is a minimum polynomial for its companion matrix.* □

EXERCISES

1. Show that $-x^3 + 39x - 90$ is the characteristic polynomial for the matrix

$$\begin{bmatrix} 0 & 0 & -90 \\ 1 & 0 & 39 \\ 0 & 1 & 0 \end{bmatrix}.$$

2. Find the characteristic polynomial for the matrix

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

and show by direct substitution that this matrix satisfies its characteristic equation.

3. Find the minimum polynomial for the matrix

$$\begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix}.$$

4. Write down a matrix which has $x^4 + 3x^3 + 2x^2 - x + 6 = 0$ as its minimum equation.

5. Show that if the matrix A satisfies the equation $x^2 + x + 1 = 0$, then A is non-singular and the inverse A^{-1} is expressible as a linear combination of A and I.

6. Show that no real 3×3 matrix satisfies $x^2 + 1 = 0$. Show that there are complex 3×3 matrices which do. Show that there are real 2×2 matrices that satisfy the equation.

7. Find a 2×2 matrix with integral elements satisfying the equation $x^3 - 1 = 0$, but not satisfying the equation $x - 1 = 0$.

8. Show that the characteristic polynomial of

$$\begin{bmatrix} 7 & 4 & -4 \\ 4 & -8 & -1 \\ -4 & -1 & -8 \end{bmatrix}$$

is not its minimum polynomial. What is the minimum polynomial?

5 | Eigenvalues and Eigenvectors

Let σ be a linear transformation of V into itself. It is often useful to find subspaces of V in which σ also acts as a linear transformation. If W is such a subspace, this means that $\sigma(W) \subset W$. A subspace with this property is called an *invariant subspace* of V under σ. Generally, the problem of determining the properties of σ on V can be reduced to the problem of determining the properties of σ on the invariant subspaces.

The simplest and most restricted case occurs when an invariant subspace W is of dimension 1. In that case, let $\{\alpha_1\}$ be a basis for W. Then, since $\sigma(\alpha_1) \in W$, there is a scalar λ_1 such that $\sigma(\alpha_1) = \lambda_1\alpha_1$. Also for any $\alpha \in W$, $\alpha = a_1\alpha_1$ and hence $\sigma(\alpha) = a_1\sigma(\alpha_1) = a_1\lambda_1\alpha_1 = \lambda_1\alpha$. In some sense the scalar λ_1 is characteristic of the invariant subspace W; σ stretches every vector in W by the factor λ_1.

In general, a problem of finding those scalars λ and associated vectors ξ for which $\sigma(\xi) = \lambda\xi$ is called an *eigenvalue problem.* A non-zero vector ξ is called an *eigenvector* of σ if there exists a scalar λ such that $\sigma(\xi) = \lambda\xi$. A scalar λ is called an *eigenvalue* of σ if there exists a non-zero vector ξ such that $\sigma(\xi) = \lambda\xi$. Notice that the equation $\sigma(\xi) = \lambda\xi$ is an equation in two variables, one of which is a vector and the other a scalar. The solution $\xi = 0$ and λ any scalar is a solution we choose to ignore since it will not lead to an invariant subspace of positive dimension. Without further conditions we have no assurance that the eigenvalue problem has any other solutions.

A typical and very important eigenvalue problem occurs in the solution of partial differential equations of the form

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

subject to the boundary conditions that $u(0, y) = u(\pi, y) = 0$,

$$\lim_{y \to \infty} u(x, y) = 0, \quad \text{and} \quad u(x, 0) = f(x)$$

where $f(x)$ is a given function. The standard technique of separation of variables leads us to try to construct a solution which is a sum of functions of the form XY where X is a function of x alone and Y is a function of y alone. For this type of function, the partial differential equation becomes

$$\frac{d^2X}{dx^2} \cdot Y + \frac{d^2Y}{dy^2} \cdot X = 0.$$

Since

$$\frac{1}{Y} \cdot \frac{d^2Y}{dy^2} = -\frac{1}{X} \cdot \frac{d^2X}{dx^2}$$

is a function of x alone and also a function of y alone, it must be a constant (scalar) which we shall call k^2. Thus we are trying to solve the equations

$$\frac{d^2X}{dx^2} = -k^2X, \qquad \frac{d^2Y}{dy^2} = k^2Y.$$

These are eigenvalue problems as we have defined the term. The vector space is the space of infinitely differentiable functions over the real numbers and the linear transformation is the differential operator d^2/dx^2.

For a given value of $k^2 (k > 0)$ the solutions would be

$$X = a_1 \cos kx + a_2 \sin kx,$$
$$Y = a_3e^{-ky} + a_4e^{ky}.$$

The boundary conditions $u(0, y) = 0$ and $\lim_{y \to \infty} u(x, y) = 0$ imply that $a_1 = a_4 = 0$. The most interesting condition for the purpose of this example is that the boundary condition $u(\pi, y) = 0$ implies that k is an integer. Thus, the eigenvalues of this eigenvalue problem are the integers, and the corresponding eigenfunctions (eigenvectors) are of the form $a_ke^{-ky} \sin kx$. The fourth boundary condition leads to a problem in Fourier series; the problem of determining the a_k so that the series

$$\sum_{n=1}^{\infty} a_k \sin kx$$

represents the given function $f(x)$ for $0 \leq x \leq \pi$.

Although the vector space in this example is of infinite dimension, we restrict our attention to the eigenvalue problem in finite dimensional vector spaces. In a finite dimensional vector space there exists a simple necessary and sufficient condition which determines the eigenvalues of an eigenvalue problem.

The eigenvalue equation can be written in the form $(\sigma - \lambda)(\xi) = 0$. We know that there exists a non-zero vector ξ satisfying this condition if and only if $\sigma - \lambda$ is singular. Let $\mathsf{A} = \{\alpha_1, \ldots, \alpha_n\}$ be any basis of V and let $A = [a_{ij}]$ be the matrix representing σ with respect to this basis. Then $A - \lambda I = C(\lambda)$ is the matrix representing $\sigma - \lambda$. Since $A - \lambda I$ is singular if and only if det $(A - \lambda I) = f(\lambda) = 0$, we see that we have proved

Theorem 5.1. *A scalar λ is an eigenvalue of σ if and only if it is a solution of the characteristic equation of a matrix representing σ.* □

Notice that Theorem 5.1 applies only to scalars. In particular a solution of the characteristic equation which is not a scalar is not an eigenvalue. For example, if the field of scalars is the field of real numbers, then non-real complex solutions of the characteristic equation are not eigenvalues. In the published literature on matrices the terms "proper values" and "characteristic values" are also used to denote what we have called eigenvalues. But, unfortunately, the same terms are often also applied to the solutions of the characteristic equation. We call the solutions of the characteristic equation *characteristic values*. Thus, a characteristic value is an eigenvalue if and only if it is also in the given field of scalars. This distinction between eigenvalues and characteristic values is not standard in the literature on matrices, but we hope this or some other means of distinguishing between these concepts will become conventional.

In abstract algebra a field is said to be *algebraically closed* if every polynomial with coefficients in the field factors into linear factors in the field. The field of complex numbers is algebraically closed. Though many proofs of this assertion are known, none is elementary. It is easy to show that algebraically closed fields exist, but it is not easy to show that a specific field is algebraically closed.

Since for most applications of concepts using eigenvalues or characteristic values the underlying field is either rational, real or complex, we content ourselves with the observation that the concepts, eigenvalue and characteristic, value, coincide if the underlying field is complex, and do not coincide if the underlying field is rational or real.

The procedure for finding the eigenvalues and eigenvectors of σ is fairly direct. For some basis $\mathsf{A} = \{\alpha_1, \ldots, \alpha_n\}$, let A be the matrix representing σ. Determine the characteristic matrix $C(x) = A - xI$ and the characteristic equation det $(A - \lambda I) = f(x) = 0$. Solve the characteristic equation. (It is this step that presents the difficulties. The characteristic equation may have no solution in F. In that event the eigenvalue problem has no solution. Even in those cases where solutions exists, finding them can present practical difficulties.) For each solution λ of $f(x) = 0$, solve the system of homogeneous equations

$$(A - \lambda I)X = C(\lambda) \cdot X = 0. \tag{5.1}$$

Since this system of equations has positive nullity, non-zero solutions exist and we should use the Hermite normal form to find them. All solutions are the representations of eigenvectors corresponding to σ.

Generally, we are given the matrix A rather than σ itself, and in this case we regard the problem as solved when the eigenvalues and the representations of the eigenvectors are obtained. We refer to the eigenvalues and eigenvectors of σ as eigenvalues and eigenvectors, respectively, of A.

Theorem 5.2. *Similar matrices have the same eigenvalues and eigenvectors.*

PROOF. This follows directly from the definitions since the eigenvalues and eigenvectors are associated with the underlying linear transformation. □

Theorem 5.3. *Similar matrices have the same characteristic polynomial.*

PROOF. Let A and $A' = P^{-1}AP$ be similar. Then

$$\det(A' - xI) = \det(P^{-1}AP - xI) = \det\{P^{-1}(A - xI)P\} = \det P^{-1} \det(A - xI)\det P = \det(A - xI) = f(x). \ \square$$

We call the characteristic polynomial of any matrix representing σ the *characteristic polynomial* of σ. Theorem 5.3 shows that the characteristic polynomial of a linear transformation is uniquely defined.

Let $S(\lambda)$ be the set of all eigenvectors of σ corresponding to λ, together with 0.

Theorem 5.4. *$S(\lambda)$ is a subspace of V.*

PROOF. If α and $\beta \in S(\lambda)$, then

$$\begin{aligned} \sigma(a\alpha + b\beta) &= a\sigma(\alpha) + b\sigma(\beta) \\ &= a\lambda\alpha + b\lambda\beta \\ &= \lambda(a\alpha + b\beta). \end{aligned} \tag{5.2}$$

Hence, $a\alpha + b\beta \in S(\lambda)$ and $S(\lambda)$ is a subspace. □

We call $S(\lambda)$ *the eigenspace* of σ corresponding to λ, and any subspace of $S(\lambda)$ is called *an eigenspace* of σ.

The dimension of $S(\lambda)$ is equal to the nullity of $C(\lambda)$, the characteristic matrix of A with λ substituted for the indeterminate x. The dimension of $S(\lambda)$ is called the *geometric multiplicity* of λ. We have shown that λ is also a solution of the characteristic equation $f(x) = 0$. Hence, $(x - \lambda)$ is a factor of $f(x)$. If $(x - \lambda)^k$ is a factor of $f(x)$ while $(x - \lambda)^{k+1}$ is not, λ is a root of $f(x) = 0$ of multiplicity k. We refer to this multiplicity as the *algebraic multiplicity* of λ.

Theorem 5.5. *The geometric multiplicity of λ does not exceed the algebraic multiplicity of λ.*

PROOF. Since the geometric multiplicity of λ is defined independently of any matrix representing σ and the characteristic equation is the same for all matrices representing σ it will be sufficient to prove the theorem for any particular matrix representing σ. We shall choose the matrix representing σ so that the assertion of the theorem is evident. Let r be the dimension of $S(\lambda)$ and let $\{\xi_1, \ldots, \xi_r\}$ be a basis of $S(\lambda)$. This linearly independent set can be extended to a basis $\{\xi_1, \ldots, \xi_n\}$ of V. Since $\sigma(\xi_i) = \lambda\xi_i$ for $i \leq r$, the matrix A representing σ with respect to this basis has the form

$$A = \begin{bmatrix} \lambda & 0 & \cdots & 0 & a_{1,r+1} & \cdots & a_{1n} \\ 0 & \lambda & \cdots & 0 & a_{2,r+1} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \lambda & a_{r,r+1} & \cdots & \cdot \\ 0 & 0 & \cdots & 0 & a_{r+1,r+1} & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 0 & a_{n,r+1} & \cdots & a_{nn} \end{bmatrix}. \tag{5.3}$$

From the form of A it is evident that $\det(A - xI) = f(x)$ is divisible by $(x - \lambda)^r$. Therefore, the algebraic multiplicity of λ is at least r, which is the geometric multiplicity. □

Theorem 5.6. *If the eigenvalues $\lambda_1, \ldots, \lambda_s$ are all different and $\{\xi_1, \ldots, \xi_s\}$ is a set of eigenvectors, ξ_i corresponding to λ_i, then the set $\{\xi_1, \ldots, \xi_s\}$ is linearly independent.*

PROOF. Suppose the set is dependent and that we have reordered the eigenvectors so that the first k eigenvectors are linearly independent and the last $s - k$ are dependent on them. Then

$$\xi_s = \sum_{i=1}^{k} a_i\xi_i$$

where the representation is unique. Not all $a_i = 0$ since $\xi_s \neq 0$. Upon applying the linear transformation σ we have

$$\lambda_s\xi_s = \sum_{i=1}^{k} a_i\lambda_i\xi_i.$$

There are two possibilities to be considered. If $\lambda_s = 0$, then none of the λ_i for $i \leq k$ is zero since the eigenvalues are distinct. This would imply that $\{\xi_1, \ldots, \xi_k\}$ is linearly dependent, contrary to assumption. If $\lambda_s \neq 0$, then

$$\xi_s = \sum_{i=1}^{k} a_i \frac{\lambda_i}{\lambda_s} \xi_i.$$

Since not all $a_i = 0$ and $\lambda_i/\lambda_s \neq 1$, this would contradict the uniqueness of the representation of ξ_s. Since we get a contradiction in any event, the set $\{\xi_1, \ldots, \xi_s\}$ must be linearly independent. □

EXERCISES

1. Show that $\lambda = 0$ is an eigenvalue of a matrix A if and only if A is singular.

2. Show that if ξ is an eigenvector of σ, then ξ is also an eigenvector of σ^n for each $n \geq 0$. If λ is the eigenvalue of σ corresponding to ξ, what is the eigenvalue of σ^n corresponding to ξ?

3. Show that if ξ is an eigenvector of both σ and τ, then ξ is also an eigenvector of $a\sigma$(for $a \in F$) and $\sigma + \tau$. If λ_1 is the eigenvalue of σ and λ_2 is the eigenvalue of τ corresponding to ξ, what are the eigenvalues of $a\sigma$ and $\sigma + \tau$?

4. Show, by producing an example, that if λ_1 and λ_2 are eigenvalues of σ_1 and σ_2, respectively, it is not necessarily true that $\lambda_1 + \lambda_2$ is an eigenvalue of $\sigma_1 + \sigma_2$.

5. Show that if ξ is an eigenvector of σ, then it is also an eigenvector of $p(\sigma)$ where $p(x)$ is a polynomial with coefficients in F. If λ is an eigenvalue of σ corresponding to ξ, what is the eigenvalue of $p(\sigma)$ corresponding to ξ?

6. Show that if σ is non-singular and λ is an eigenvalue of σ, then λ^{-1} is an eigenvalue of σ^{-1}. What is the corresponding eigenvector?

7. Show that if every vector in V is an eigenvector of σ, then σ is a scalar transformation.

8. Let P_n be the vector space of polynomials of degree at most $n - 1$, and let D be the differentiation operator; that is $D(t^k) = kt^{k-1}$. Determine the characteristic polynomial for D. From your knowledge of the differentiation operator and net using Theorem 4.3, determine the minimum polynomial for D. What kind of differential equation would an eigenvector of D have to satisfy? What are the eigenvectors of D?

9. Let $A = [a_{ij}]$. Show that if $\sum_j a_{ij} = c$ independent of i, then $\xi = (1, 1, \ldots, 1)$ is an eigenvector. What is the corresponding eigenvalue?

10. Let W be an invariant subspace of V under σ, and let $A = \{\alpha_1, \ldots, \alpha_n\}$ be a basis of V such that $\{\alpha_1, \ldots, \alpha_k\}$ is a basis of W. Let $A = [a_{ij}]$ be the matrix representing σ with respect to the basis A. Show that all elements in the first k columns below the kth row are zeros.

11. Show that if λ_1 and $\lambda_2 \neq \lambda_1$ are eigenvalues of σ_1 and ξ_1 and ξ_2 are eigenvectors corresponding to λ_1 and λ_2, respectively, then $\xi_1 + \xi_2$ is not an eigenvector.

12. Assume that $\{\xi_1, \ldots, \xi_{1r}\}$ are eigenvectors with distinct eigenvalues. Show that $\sum_{i=1}^{r} a_i\xi_i$ is never an eigenvector unless precisely one coefficient is non-zero.

13. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Show that if Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & & \lambda_n \end{bmatrix},$$

and $P = [p_{ij}]$ is the matrix in which column j is the n-tuple representing an eigenvector corresponding to λ_j, then $AP = P\Lambda$.

14. Use the notation of Exercise 13. Show that if A has n linearly independent eigenvalues, then eigenvectors can be chosen so that P is non-singular. In this case $P^{-1}AP = \Lambda$.

6 | Some Numerical Examples

Since we are interested here mainly in the numerical procedures, we start with the matrices representing the linear transformations and obtain their eigenvalues and the representations of the eigenvectors.

Example 1. Let

$$A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}.$$

The first step is to obtain the characteristic matrix

$$C(x) = \begin{bmatrix} -1-x & 2 & 2 \\ 2 & 2-x & 2 \\ -3 & -6 & -6-x \end{bmatrix}$$

and then the characteristic polynomial

$$\det C(x) = -(x+2)(x+3)x.$$

Thus the eigenvalues of A are $\lambda_1 = -2$, $\lambda_2 = -3$, and $\lambda_3 = 0$. The next steps are to substitute, successively, the eigenvalues for x in the characteristic matrix. Thus we have

$$C(-2) = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 2 \\ -3 & -6 & -4 \end{bmatrix}.$$

The Hermite normal form obtained from $C(-2)$ is

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

The components of the eigenvector corresponding to $\lambda_1 = -2$ are found by solving the equations

$$\begin{aligned} x_1 + 2x_2 \quad &= 0 \\ x_3 &= 0. \end{aligned}$$

Thus $(2, -1, 0)$ is the representation of an eigenvector corresponding to λ_1; for simplicity we shall write $\xi_1 = (2, -1, 0)$, identifying the vector with its representation.

In a similar fashion we obtain

$$C(-3) = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & 2 \\ -3 & -6 & -3 \end{bmatrix}.$$

From $C(-3)$ we obtain the Hermite normal form

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

and hence the eigenvector $\xi_2 = (1, 0, -1)$.

Similarly, from

$$C(0) = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}$$

we obtain the eigenvector $\xi_3 = (0, 1, -1)$.

By Theorem 5.6 the three eigenvectors obtained for the three different eigenvalues are linearly independent.

Example 2. Let

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

From the characteristic matrix

$$C(x) = \begin{bmatrix} 1-x & 1 & -1 \\ -1 & 3-x & -1 \\ -1 & 2 & -x \end{bmatrix}$$

we obtain the characteristic polynomial $\det C(x) = -(x-1)^2(x-2)$. Thus we have just two distinct eigenvalues; $\lambda_1 = \lambda_2 = 1$ with algebraic multiplicity two, and $\lambda_3 = 2$.

Substituting λ_1 for x in the characteristic matrix we obtain

$$C(1) = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}.$$

The corresponding Hermite normal form is

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

Thus it is seen that the nullity of $C(1)$ is 1. The eigenspace $S(1)$ is of dimension 1 and the geometric multiplicity of the eigenvalue 1 is 1. This shows that the geometric multiplicity can be lower than the algebraic multiplicity. We obtain $\xi_1 = (1, 1, 1)$.

The eigenvector corresponding to $\lambda_3 = 2$ is $\xi_3 = (0, 1, 1)$.

EXERCISES

For each of the following matrices find all the eigenvalues and as many linearly independent eigenvectors as possible.

1. $\begin{bmatrix} 2 & 4 \\ 5 & 3 \end{bmatrix}$

2. $\begin{bmatrix} 3 & 2 \\ -2 & 3 \end{bmatrix}$

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

4. $\begin{bmatrix} 1 & -\sqrt{2} \\ \sqrt{2} & 4 \end{bmatrix}$

5. $\begin{bmatrix} 4 & 9 & 0 \\ 0 & -2 & 8 \\ 0 & 0 & 7 \end{bmatrix}$

6. $\begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix}$

7. $\begin{bmatrix} 7 & 4 & -4 \\ 4 & -8 & -1 \\ -4 & -1 & -8 \end{bmatrix}$

8. $\begin{bmatrix} 2 & -i & 0 \\ i & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

7 | Similarity

Generally, for a given linear transformation σ we seek a basis for which the matrix representing σ has as simple a form as possible. The simplest form is that in which the elements not on the main diagonal are zero, a *diagonal* matrix. Not all linear transformations can be represented by diagonal matrices, but relatively large classes of transformations can be represented by diagonal matrices, and we seek conditions under which such a representation exists.

Theorem 7.1. *A linear transformation σ can be represented by a diagonal matrix if and only if there exists a basis consisting of eigenvectors of σ.*

PROOF. Suppose there is a linearly independent set $X = \{\xi_1, \ldots, \xi_n\}$ of eigenvectors and that $\{\lambda_1, \ldots, \lambda_n\}$ are the corresponding eigenvalues. Then $\sigma(\xi_i) = \lambda_i \xi_i$ so that the matrix representing σ with respect to the basis X has the form

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \tag{7.1}$$

that is, σ is represented by a diagonal matrix.

Conversely, if σ is represented by a diagonal matrix, the vectors in that basis are eigenvectors. □

Usually, we are not given the linear transformation σ directly. We are given a matrix A representing σ with respect to an unspecified basis. In this case Theorem 7.1 is usually worded in the form: A matrix A is similar to a diagonal matrix if and only if there exist n linearly independent eigenvectors of A. In this form a computation is required. We must find the matrix P such that $P^{-1}AP$ is a diagonal matrix.

Let the matrix A be given; that is, A represents σ with respect to some basis $\mathsf{A} = \{\alpha_1, \ldots, \alpha_n\}$. Let $\xi_j = \sum_{i=1}^{n} p_{ij}\alpha_j$ be the representations of the eigenvectors of A with respect to A. Then the matrix A' representing σ with respect to the basis $X = \{\xi_1, \ldots, \xi_n\}$ is $P^{-1}AP = A'$. By Theorem 7.1, A' is a diagonal matrix.

In Example 1 of Section 6, the matrix

$$A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}$$

has three linearly independent eigenvectors, $\xi_1 = (2, -1, 0)$, $\xi_2 = (1, 0, -1)$, and $\xi_3 = (0, 1, -1)$. The matrix of transition P has the components of these vectors written in its columns:

$$P = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}, \qquad P^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -2 \\ 1 & 2 & 1 \end{bmatrix}.$$

The reader should check that $P^{-1}AP$ is a diagonal matrix with the eigenvalues appearing in the main diagonal.

In Example 2 of Section 6, the matrix

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

has one linearly independent eigenvector corresponding to each of its two eigenvalues. As there are no other eigenvalues, there does not exist a set of three linearly independent eigenvectors. Thus, the linear transformation represented by this matrix cannot be represented by a diagonal matrix; A is not similar to a diagonal matrix.

Corollary 7.2. *If σ can be represented by a diagonal matrix D, the elements in the main diagonal of D are the eigenvalues of σ.* □

Theorem 7.3. *If an $n \times n$ matrix has n distinct eigenvalues, then A is similar to a diagonal matrix.*

PROOF. By Theorem 5.6 the n eigenvectors corresponding to the n eigenvalues of A are linearly independent and form a basis. By Theorem 7.1 the matrix representing the underlying linear transformation with respect to this basis is a diagonal matrix. Hence, A is similar to a diagonal matrix. □

Theorem 7.3 is quite practical because we expect the eigenvalues of a randomly given matrix to be distinct; however, there are circumstances under which the theorem does not apply. There may not be n distinct eigenvalues, either because some have algebraic multiplicity greater than one or because the characteristic equation does not have enough solutions in the field. The most general statement that can be made without applying more conditions to yield more results is

Theorem 7.4. *A necessary and sufficient condition that a matrix A be similar to a diagonal matrix is that its minimum polynomial factor into distinct linear factors with coefficients in F.*

PROOF. Suppose first that the matrix A is similar to a diagonal matrix D. By Theorem 5.3, A and D have the same characteristic polynomial. Since D is a diagonal matrix the elements in the main diagonal are the solutions of the chracteristic equation and the characteristic polynomial must factor into linear factors. By Theorem 4.4 the minimum polynomial for A must contain each of the linear factors of $f(x)$, although possibly with lower multiplicity. It can be seen, however, either from Theorem 4.3 or by direct substitution, that D satisfies an equation without repeated factors. Thus, the minimum polynomial for A has distinct linear factors.

On the other hand, suppose that the minimum polynomial for A is $m(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_p)$ with distinct linear factors. Let M_i be the kernel of $\sigma - \lambda_i$. The non-zero vectors in M_i are the eigenvectors of σ corresponding to λ_i. It follows from Theorem 5.6 that a non-zero vector in M_i cannot be expressed as a sum of vectors in $\sum_{j \neq i} M_j$. Hence, the sum $M_1 + M_2 + \cdots + M_p$ is direct.

Let $\nu_i = \dim M_i$, that is, ν_i is the nullity of $\sigma - \lambda_i$. Since $M_1 \oplus \cdots \oplus M_p \subset V$ we have $\nu_1 + \cdots + \nu_p \leq n$. By Theorem 1.5 of Chapter II $\dim (\sigma - \lambda_i)V = n - \nu_i = \rho_i$, By another application of the same theorem we have $\dim (\sigma - \lambda_j)\{(\sigma - \lambda_i)V\} \geq \rho_i - \nu_j = n - (\nu_i + \nu_j)$.

Finally, by repeated application of the same ideas we obtain $0 = \dim m(\sigma)V \geq n - (\nu_1 + \cdots + \nu_p)$. Thus, $\nu_1 + \cdots + \nu_p = n$. This shows that $M_1 \oplus \cdots \oplus M_p = V$. Since every vector in V is a linear combination of eigenvectors, there exists a basis of eigenvectors. Thus, A is similar to a diagonal matrix. □

Theorem 7.4 is important in the theory of matrices, but it does not provide the most effective means for deciding whether a particular matrix is similar to diagonal matrix. If we can solve the characteristic equation, it is easier to try to find the n linearly independent eigenvectors than it is to use Theorem 7.4 to ascertain that they do or do not exist. If we do use this theorem and are able to conclude that a basis of eigenvectors does exist, the work done in making this conclusion is of no help in the attempt to find the eigenvectors. The straightforward attempt to find the eigenvectors is always conclusive On the other hand, if it is not necessary to find the eigenvectors, Theorem 7.4 can help us make the necessary conclusion without solving the characteristic equation.

For any square matrix $A = [a_{ij}]$, $\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}$ is called the *trace* of A. It is the sum of the elements in the diagonal of A. Since $\mathrm{Tr}(AB) = \sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ij}b_{ji}) = \sum_{j=1}^{n}(\sum_{i=1}^{n} b_{ji}a_{ij}) = \mathrm{Tr}(BA)$,

$$\mathrm{Tr}(P^{-1}AP) = \mathrm{Tr}(APP^{-1}) = \mathrm{Tr}(A). \tag{7.2}$$

This shows that the trace is invariant under similarity transformations;

that is, similar matrices have the same trace. For a given linear transformation σ of V into itself, all matrices representing σ have the same trace. Thus we can define $\mathrm{Tr}(\sigma)$ to be the trace of any matrix representing σ.

Consider the coefficient of x^{n-1} in the expansion of the determinant of the characteristic matrix,

$$\begin{bmatrix} a_{11} - x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - x \end{bmatrix}. \tag{7.3}$$

The only way an x^{n-1} can be obtained is from a product of $n - 1$ of the diagonal elements, multiplied by the scalar from the remaining diagonal element. Thus, the coefficient of x^{n-1} is $(-1)^{n-1} \sum_{i=1}^{n} a_{ii}$, or $(-1)^{n-1}\mathrm{Tr}(A)$.

If $f(x) = \det (A - xI)$ is the characteristic polynomial of A, then $\det A = f(0)$ is the constant term of $f(x)$. If $f(x)$ is factored into linear factors in the form

$$f(x) = (-1)^n(x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p}, \tag{7.4}$$

the constant term is $\prod_{i=1}^{p} \lambda_i^{r_i}$. Thus $\det A$ is the product of the characteristic values (each counted with the multiplicity with which it is a factor of the characteristic polynomials). In a similar way it can be seen that $\mathrm{Tr}(A)$ is the sum of the characteristic values (each counted with multiplicity).

We have now shown the existence of several objects associated with a matrix, or its underlying linear transformation, which are independent of the coordinate system. For example, the characteristic polynomial, the determinant, and the trace are independent of the coordinate system. Actually, this list is redundant since $\det A$ is the constant term of the characteristic polynomial, and $\mathrm{Tr}(A)$ is $(-1)^{n-1}$ times the coefficient of x^{n-1} of the characteristic polynomial. Functions of this type are of interest because they contain information about the linear transformation, or the matrix, and they are sometimes rather easy to evaluate. But this raises a host of questions. What information do these invariants contain? Can we find a complete list of non-redundant invariants, in the sense that any other invariant can be computed from those in the list? While some partial answers to these questions will be given, a systematic discussion of these questions is beyond the scope of this book.

Theorem 7.5. *Let V be a vector space with a basis consisting of eigenvectors of σ. If W is any subspace of V invariant under σ, then W also has a basis consisting of eigenvectors of σ.*

PROOF. Let α be any vector in W. Since V has a basis of eigenvectors of σ, α can be expressed as a linear combination of eigenvectors of σ. By disregarding terms with zero coefficients, combining terms corresponding to the same eigenvalue, and renaming a term like $a_i\xi_i$, where ξ_i is an eigenvector and $a_i \neq 0$, as an eigenvector with coefficient 1, we can represent α in the form

$$\alpha = \sum_{i=1}^{r} \xi_i,$$

where the ξ_i are eigenvectors of σ with distinct eigenvalues. Let λ_i be the eigenvalue corresponding to ξ_i. We will show that each $\xi_i \in W$.

$(\sigma - \lambda_2)(\alpha - \lambda_3) \cdots (\sigma - \lambda_r)(\alpha)$ is in W since W is invariant under σ, and hence invariant under $\sigma - \lambda$ for any scalar λ. But then $(\sigma - \lambda_2)(\sigma - \lambda_3) \cdots (\sigma - \lambda_r)(\alpha) = (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_r)\xi_1 \in W$, and $\xi_1 \in W$ since $(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_r) \neq 0$. A similar argument shows that each $\xi_i \in W$.

Since this argument applies to any $\alpha \in W$, W is spanned by eigenvectors of σ. Thus, W has a basis of eigenvectors of σ. □

Theorem 7.6. *Let V be a vector space over C, the field of complex numbers. Let σ be a linear transformation of V into itself. V has a basis of eigenvectors for σ if and only if for every subspace S invariant under σ there is a subspace T invariant under σ such that $V = S \oplus T$.*

PROOF. The theorem is obviously true if V is of dimension 1. Assume the assertions of the theorem are correct for spaces of dimension less than n, where n is the dimension of V.

Assume first that for every subspace S invariant under σ there is a complementary subspace T also invariant under σ. Since V is a vector space over the complex numbers σ has at least one eigenvalue λ_1. Let α_1 be an eigenvector corresponding to λ_1. The subspace $S_1 = \langle\alpha_1\rangle$ is then invariant under σ. By assumption there is a subspace T_1 invariant under σ such that $V = S_1 \oplus T_1$.

Every subspace S_2 of T_1 invariant under $R\sigma$ is also invariant under σ. Thus there exists a subspace T_2 of V invariant under σ such that $V = S_2 \oplus T_2$. Now $S_2 \subset T_1$ and $T_1 = S_2 \oplus (T_2 \cap T_1)$. (See Exercise 15, Section I-4.) Since $T_2 \cap T_1$ is invariant under σ, and therefore under $R\sigma$, the induction assumption holds for the subspace T_1. Thus, T_1 has a basis of eigenvectors, and by adjoining α_1 to this basis we obtain a basis of eigenvectors of V.

Now assume there is a basis of V consisting of eigenvectors of σ. By theorem 7.5 any invariant subspace S has a basis of eigenvectors. The method of proof of Theorem 2.7 of Chapter I (the Steinitz replacement theorem) will yield a basis of V consisting of eigenvectors of σ, and this basis will contain the basis of S consisting of eigenvectors. The eigenvectors adjoined will span a subspace T, and this subspace will be invariant under σ and complementary to S. □

EXERCISES

1. For each matrix A given in the exercises of Section 6 find, when possible, a non-singular matrix P for which $P^{-1}AP$ is diagonal.

2. Show that the matrix $\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}$ where $c \neq 0$ is not similar to a diagonal matrix

3. Show that any 2×2 matrix satisfying $x^2 + 1 = 0$ is similar to the matrix
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

4. Show that if A is non-singular, then AB is similar to BA.

5. Show that any two projections of the same rank are similar.

*8 | The Jordan Normal Form

A normal form that is obtainable in general when the field of scalars is the field of complex numbers is known as the *Jordan normal form.* An application of the Jordan normal form to power series of matrices and systems of linear differential equations is given in the chapter on applications. Except for these applications this section can be skipped without penalty.

We assume that the field of scalars is the field of complex numbers. Thus for any square matrix A the characteristic polynomial $f(x)$ factors into linear factors, $f(x) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p}$ where $\lambda_i \neq \lambda_j$ for $i \neq j$ and r_i is the algebraic multiplicity of the eigenvalue λ_i. The minimum polynomial $m(x)$ for A is of the form $m(x) = (x - \lambda_1)^{s_1}(x - \lambda_2)^{s_2} \cdots (x - \lambda_p)^{s_p}$ where $1 \leq s_i \leq r_i$.

In the theorems about the diagonalization of matrices we sought bases made up of eigenvectors. Because we are faced with the possibility that such bases do not exist, we must seek proper generalizations of the eigenvectors. It is more fruitful to think of the eigenspaces rather than the eigenvectors themselves. An eigenvalue is a scalar λ for which the linear transformation $\sigma - \lambda$ is singular. An eigenspace is the kernel (of positive dimension) of the linear transformation $\sigma - \lambda$. The proper generalization of eigenspaces turns out to be the kernels of higher powers of $\sigma - \lambda$. For a given eigenvalue λ, let M^k be the kernel of $(\sigma - \lambda)^k$. Thus, $M^0 = \{0\}$ and M^1

is the eigenspace of λ. For $\alpha \in M^k$, $(\sigma - \lambda)^{k+1}(\alpha) = (\sigma - \lambda)(\sigma - \lambda)^k(\alpha) = (\sigma - \lambda)(0) = 0$. Hence, $M^k \subset M^{k+1}$. Also, for $\alpha \in M^{k+1}$, $(\sigma - \lambda)^k(\sigma - \lambda)(\alpha) = (\sigma - \lambda)^{k+1}(\alpha) = 0$ so that $(\sigma - \lambda)(\alpha) \in M^k$. Hence, $(\sigma - \lambda)M^{k+1} \subset M^k$.

Since all $M^k \subset V$ and V is finite dimensional, the sequence of subspaces $M^0 \subset M^1 \subset M^2 \subset \cdots$ must eventually stop increasing. Let t be the smallest index such that $M^k = M^t$ for all $k \geq t$, and denote M^t by $M_{(\lambda)}$. Let m_k be the dimension of M^k and m_t the dimension of $M_{(\lambda)}$.

Let $(\sigma - \lambda)^k V = W^k$. Then $W^{k+1} = (\sigma - \lambda)^{k+1}V = (\sigma - \lambda)^k\{(\sigma - \lambda)V\} \subset (\sigma - \lambda)^k V = W^k$. Thus, the subspaces W^k form a decreasing sequence $W^0 \supset W^1 \supset W^2 \supset \cdots$. Since the dimension of W^k is $n - m_k$, we see that $W^k = W^t$ for all $k \geq t$. Denote W^t by $W_{(\lambda)}$.

Theorem 8.1. *V is the direct sum of $M_{(\lambda)}$ and $W_{(\lambda)}$.*

PROOF. Since $(\sigma - \lambda)W^t = (\sigma - \lambda)^{t+1}V = W^{t+1} = W^t$ we see that $\sigma - \lambda$ is non-singular on $W^t = W_{(\lambda)}$. Now let α be any vector in V. Then $(\sigma - \lambda)^t(\alpha) = \beta$ is an element of $W_{(\lambda)}$. Because $(\sigma - \lambda)^t$ is non-singular on $W_{(\lambda)}$ there is a unique vector $\gamma \in W_{(\lambda)}$ such that $(\sigma - \lambda)^t(\gamma) = \beta$. Let $\alpha - \gamma$ be denoted by δ. It is easily seen that $\delta \in M_{(\lambda)}$. Hence $V = M_{(\lambda)} + W_{(\lambda)}$. Finally, since $\dim M_{(\lambda)} = m_t$ and $\dim W_{(\lambda)} = n - m_t$, the sum is direct. □

In the course of defining M^k and W^k we have shown that

(1) $(\sigma - \lambda)M^{k+1} \subset M^k \subset M^{k+1}$,

(2) $(\sigma - \lambda)W^k = W^{k+1} \subset W^k$.

This shows that each M^k and W^k is invariant under $\sigma - \lambda$. It follows immediately that each is invariant under any polynomial in $\sigma - \lambda$, and hence also under any polynomial in σ. The use we wish to make of this observation is that if μ is any other eigenvalue, then $\sigma - \mu$ also maps $M_{(\lambda)}$ and $W_{(\lambda)}$ into themselves.

Let $\lambda_1, \ldots, \lambda_p$ be the distinct eigenvalues of σ. Let M_i be a simpler notation for the subspace $M_{(\lambda_i)}$ defined as above, and let W_i be a simpler notation for $W_{(\lambda_i)}$.

Theorem 8.2. *For $\lambda_i \neq \lambda_j$, $M_i \subset W_j$.*

PROOF. Suppose $\alpha \in M_i$ is in the kernel of $\sigma - \lambda_j$. Then

$$(\lambda_j - \lambda_i)^{t_i}\alpha = \{(\sigma - \lambda_i) - (\sigma - \lambda_j)\}^{t_i}(\alpha)$$
$$= (\sigma - \lambda_i)^{t_i}(\alpha) + \sum_{k=1}^{t_i}(-1)^k\binom{t_i}{k}(\sigma - \lambda_i)^{t_i-k}(\sigma - \lambda_j)^k(\alpha).$$

The first term is zero because $\alpha \in M_i$, and the others are zero because α is in the kernel of $\sigma - \lambda_j$. Since $\lambda_j - \lambda_i \neq 0$, it follows that $\alpha = 0$. This means that $\sigma - \lambda_j$ is non-singular on M_i; that is, $\sigma - \lambda_j$ maps M_i onto itself. Thus M_i is contained in the set of images under $(\sigma - \lambda_j)^{t_j}$, and hence $M_i \subset W_j$. □

Theorem 8.3. $V = M_1 \oplus M_2 \oplus \cdots \oplus M_p$.

PROOF. Since $V = M_2 \oplus W_2$ and $M_2 \subset W_1$, we have $V = M_1 \oplus W_1 = M_1 \oplus \{M_2 \oplus (W_1 \cap W_2)\}$. Continuing in the same fashion, we get $V = M_1 \oplus \cdots \oplus M_p \oplus \{W_1 \cap \cdots \cap W_p\}$. Thus the theorem will follow if we can show that $W = W_1 \cap \cdots \cap W_p = \{0\}$. By an extension of remarks already made $(\sigma - \lambda_1) \cdots (\sigma - \lambda_p) = q(\sigma)$ is non-singular on W; that is, $q(\sigma)$ maps W onto itself. For arbitrarily large k, $[q(\sigma)]^k$ also maps W onto itself. But $q(x)$ contains each factor of the characteristic polynomial $f(x)$ so that for large enough k, $[q(x)]^k$ is divisible by $f(x)$. This implies that $W = \{0\}$. □

Corollary 8.4. $t_i = s_i$ *for* $i = 1, \ldots, p$.

PROOF. Since $V = M_1 \oplus \cdots \oplus M_p$ and $(\sigma - \lambda_i)^{t_i}$ vanishes on M_i, it follows that $(\sigma - \lambda_1)^{t_1} \cdots (\sigma - \lambda_p)^{t_p}$ vanishes on all of V. Thus $(x - \lambda_1)^{t_1} \cdots (x - \lambda_p)^{t_p}$ is divisible by the minimum polynomial and $s_i \leq t_i$.

On the other hand, if for a single i we have $s_i < t_i$, there is an $\alpha \in M_i$ such that $(\sigma - \lambda_i)^{s_i}(\alpha) \neq 0$. For all $\lambda_j \neq \lambda_i$, $\sigma - \lambda_j$ is non-singular on M_i. Hence $m(\sigma) \neq 0$. This is a contradiction so that $t_i = s_i$. □

Let us return to the situation where, for the single eigenvalue λ, M^k is the kernel of $(\sigma - \lambda)^k$ and $W^k = (\sigma - \lambda)^k V$. In view of Corollary 8.4 we let s be the smallest index such that $M^k = M^s$ for all $k \geq s$. By induction we can construct a basis $\{\alpha_1, \ldots, \alpha_{m_s}\}$ of $M_{(\lambda)}$ such that $\{\alpha_1, \ldots, \alpha_{m_k}\}$ is a basis of M^k.

We now proceed step by step to modify this basis. The set $\{\alpha_{m_{s-1}+1}, \ldots, \alpha_{m_s}\}$ consists of those basis elements in M^s which are not in M^{s-1}. These elements do not have to be replaced, but for consistency of notation we change their names; let $\alpha_{m_{s-1}+\nu} = \beta_{m_{s-1}+\nu}$. Now set $(\sigma - \lambda)(\beta_{m_{s-1}+\nu}) = \beta_{m_{s-2}+\nu}$ and consider the set $\{\alpha_1, \ldots, \alpha_{m_{s-2}}\} \cup \{\beta_{m_{s-2}+1}, \ldots, \beta_{m_{s-2}+m_s-m_{s-1}}\}$. We wish to show that this set is linearly independent.

If this set were linearly dependent, a non-trivial relation would exist and it would have to involve at least one of the β_i with a non-zero coefficient since the set $\{\alpha_1, \ldots, \alpha_{m_{s-2}}\}$ is linearly independent. But then a non-trivial linear combination of the β_i would be an element of M^{s-2}, and $(\sigma - \lambda)^{s-2}$ would map this linear combination onto 0. This would mean that $(\sigma - \lambda)^{s-1}$ would map a non-trivial linear combination of $\{\alpha_{m_{s-1}+1}, \ldots, \alpha_{m_s}\}$ onto 0. Then this non-trivial linear combination would be in M^{s-1}, which would contradict the linear independence of $\{\alpha_1, \ldots, \alpha_{m_s}\}$. Thus the set $\{\alpha_1, \ldots, \alpha_{m_{s-2}}\} \cup \{\beta_{m_{s-2}+1}, \ldots, \beta_{m_{s-2}+m_s-m_{s-1}}\}$ is linearly independent.

This linearly independent subset of M^{s-1} can be expanded to a basis of M^{s-1} We use β's to denote these additional elements of this basis, if any additional elements are required. Thus we have the new basis $\{\alpha_1, \ldots, \alpha_{m_{s-2}}\} \cup \{\beta_{m_{s-2}+1}, \ldots, \beta_{m_{s-1}}\}$ of M^{s-1}.

We now set $(\sigma - \lambda)(\beta_{m_{s-2}+\nu}) = \beta_{m_{s-3}+\nu}$ and proceed as before to obtain a new basis $\{\alpha_1, \ldots, \alpha_{m_{s-3}}\} \cup \{\beta_{m_{s-3}+1}, \ldots, \beta_{m_{s-2}}\}$ of M^{s-2}.

Proceeding in this manner we finally get a new basis $\{\beta_1, \ldots, \beta_{m_s}\}$ of $M_{(\lambda)}$ such that $\{\beta_1, \ldots, \beta_{m_k}\}$ is a basis of M^k and $(\sigma - \lambda)(\beta_{m_k+\nu}) = \beta_{m_{k-1}+\nu}$ for $k \geq 1$. This relation can be rewritten in the form

$$
\begin{aligned}
\sigma(\beta_{m_k+\nu}) &= \lambda\beta_{m_k+\nu} + \beta_{m_{k-1}+\nu} \qquad && \text{for} \quad k \geq 1, \\
\sigma(\beta_\nu) &= \lambda\beta_\nu && \text{for} \quad \nu \leq m_1.
\end{aligned}
\tag{8.1}
$$

Thus we see that in a certain sense $\beta_{m_k+\nu}$ is "almost" an eigenvector.

This suggests reordering the basis vectors so that $\{\beta_1, \beta_{m_1+1}, \ldots, \beta_{m_{s-1}+1}\}$ are listed first. Next we should like to list the vectors $\{\beta_2, \beta_{m_1+2}, \ldots\}$, etc. The general idea is to list each of the first elements from each section of the β's, then each of the second elements from each section, and continue until a new ordering of the basis is obtained.

With the basis of $M_{(\lambda)}$ listed in this order (and assuming for the moment that that $M_{(\lambda)}$ is all of V) the matrix representing σ takes the form

$$
\begin{array}{r}
s \text{ rows} \left\{ \vphantom{\begin{matrix} 1\\1\\1\\1\\1\\1\\1\\1 \end{matrix}} \right. \\
\le s \text{ rows} \left\{ \vphantom{\begin{matrix} 1\\1\\1\\1\\1\\1 \end{matrix}} \right. \\
\vphantom{1}
\end{array}
\left[\begin{array}{c|c|c}
\begin{matrix}
\lambda & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda & 1 & \cdots & 0 & 0 \\
0 & 0 & \lambda & \cdots & 0 & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & \cdot \\
0 & 0 & 0 & \cdots & \lambda & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda
\end{matrix} & \text{all zeros} & \text{all zeros} \\
\hline
\text{all zeros} &
\begin{matrix}
\lambda & 1 & \cdots & 0 \\
0 & \lambda & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
0 & 0 & \cdots & \lambda
\end{matrix} & \text{all zeros} \\
\hline
\text{all zeros} & \text{all zeros} & \text{etc.}
\end{array}\right].
$$

Theorem 8.5. *Let A be a matrix with characteristic polynomial $f(x) = (x - \lambda_1)^{r_1} \cdots (x - \lambda_p)^{r_p}$ and minimum polynomial $m(x) = (x - \lambda_1)^{s_1} \cdots (x - \lambda_p)^{s_p}$. A is similar to a matrix J with submatrices of the form*

$$B_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}$$

along the main diagonal. All other elements of J are zero. For each λ_i there at least one B_i of order s_i. All other B_i corresponding to this λ_i are of order less than or equal to s_i. The number of B_i corresponding to this λ_i is equal to the geometric multiplicity of λ_i. The sum of the orders of all the B_i corresponding to λ_i is r_i. While the ordering of the B_i along the main diagonal of J is not unique, the number of B_i of each possible order is uniquely determined by A. J is called the Jordan normal form corresponding to A.

PROOF. From Theorem 8.3 we have $V = M_1 \oplus \cdots \oplus M_p$. In the discussion preceding the statement of Theorem 8.5 we have shown that each M_i has a basis of a special type. Since V is the sum of the M_i, the union of these bases spans V. Since the sum is direct, the union of these bases is linearly independent and, hence, a basis for V. This shows that a matrix J of the type described in Theorem 8.5 does represent σ and is therefore similar to A.

The discussion preceding the statement of the theorem also shows that the dimensions $m_{i,k}$ of the kernels M_i^k of the various $(\sigma - \lambda_i)^k$ determine the orders of the B_i in J. Since A determines σ and σ determines the subspace M_i^k independently of the bases employed, the B_i are uniquely determined.

Since the λ_i appear along the main diagonal of J and all other non-zero elements of J are above the main diagonal, the number of times $x - \lambda_i$ appears as a factor of the characteristic polynomial of J is equal to the number of times λ_i appears in the main diagonal. Thus the sum of the orders of the B_i corresponding to λ_i is exactly r_i. This establishes all the statements of Theorem 8.5. □

Let us illustrate the workings of the theorems of this section with some examples. Unfortunately, it is a little difficult to construct an interesting example of low order. Hence, we give two examples, The first example illustrates the choice of basis as described for the space $M_{(\lambda)}$. The second example illustrates the situation described by Theorem 8.3.

Example 1. Let

$$A = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ -4 & 1 & -3 & 2 & 1 \\ -2 & -1 & 0 & 1 & 1 \\ -3 & -1 & -3 & 4 & 1 \\ -8 & -2 & -7 & 5 & 4 \end{bmatrix}.$$

The first step is to obtain the characteristic matrix

$$C(x) = \begin{bmatrix} 1-x & 0 & -1 & 1 & 0 \\ -4 & 1-x & -3 & 2 & 1 \\ -2 & -1 & -x & 1 & 1 \\ -3 & -1 & -3 & 4-x & 1 \\ -8 & -2 & -7 & 5 & 4-x \end{bmatrix}.$$

Although it is tedious work we can obtain the characteristic polynomial $f(x) = (x-2)^5$. We have one eigenvalue with algebraic multiplicity 5. What is the geometric multiplicity and what is the minimum equation for A? Although there is an effective method for determining the minimum equation, it is less work and less wasted effort to proceed directly with determining the eigenvectors. Thus, from

$$C(2) = \begin{bmatrix} -1 & 0 & -1 & 1 & 0 \\ -4 & -1 & -3 & 2 & 1 \\ -2 & -1 & -2 & 1 & 1 \\ -3 & -1 & -3 & 2 & 1 \\ -8 & -2 & -7 & 5 & 2 \end{bmatrix},$$

we obtain by elementary row operations the Hermite normal form

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

From this we learn that there are two linearly independent eigenvectors corresponding to 2. The dimension of M^1 is 2. Without difficulty we find the eigenvectors

$$\alpha_1 = (0, -1, 1, 1, 0)$$
$$\alpha_2 = (0, 1, 0, 0, 1).$$

Now we must compute $(A - 2I)^2 = (C(2))^2$, and obtain

$$(A - 2I)^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 \end{bmatrix}.$$

The rank of $(A - 2I)^2$ is 1 and hence M^2 is of dimension 4. The α_1 and α_2 we already have are in M^2 and we must obtain two more vectors in M^2 which, together with α_1 and α_2, will form an independent set. There is quite a bit of freedom for choice and

$$\alpha_3 = (0, 1, 0, 0, 0)$$
$$\alpha_4 = (-1, 0, 1, 0, 0)$$

appear to be as good as any.

Now $(A - 2I)^3 = 0$, and we know that the minimum polynomial for A is $(x - 2)^3$. We have this knowledge and quite a bit more more for less work than would be required to find the minimum polynomial directly. We see, then, that $M^3 = V$ and we have to find another vector independent of α_1, α_2, α_3, and α_4. Again, there are many possible choices. Some choices will lead to a simpler matrix of transition than will others, and there seems to be no very good way to make the choice that will result in the simplest matrix of transition. Let us take

$$\alpha_5 = (0, 0, 0, 1, 0).$$

We now have the basis of $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$ such that $\{\alpha_1, \alpha_2\}$ is a basis of M^1, $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ is a basis of M^2, and $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$ is a basis of M^3. Following our instructions, we set $\beta_5 = \alpha_5$. Then

$$(A - 2I)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 5 \end{bmatrix}.$$

Hence, we set $\beta_3 = (1, 2, 1, 2, 5)$. Now we must choose β_4 so that $\{\alpha_1, \alpha_2, \beta_3, \beta_4\}$ is a basis for M^2. We can choose $\beta_4 = (-1, 0, 1, 0, 0)$. Then

$$(A - 2I)\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad (A - 2I)\begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Hence, we choose $\beta_1 = (0, 0, 1, 1, 1)$ and $\beta_2 = (0, 1, 0, 0, 1)$. Thus,

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & -1 \\ 0 & 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 \\ 1 & 5 & 0 & 1 & 0 \end{bmatrix}$$

is the matrix of transition that will transform A to the Jordan normal form

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

Example 2. Let

$$A = \begin{bmatrix} 5 & -1 & -3 & 2 & -5 \\ 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & -2 \\ 0 & -1 & 0 & 3 & 1 \\ 1 & -1 & -1 & 1 & 1 \end{bmatrix}.$$

The characteristic polynomial is $f(x) = -(x - 2)^3(x - 3)^2$. Again we have repeated eigenvalues, one of multiplicity 3 and one of multiplicity 2.

$$C(2) = \begin{bmatrix} 3 & -1 & -3 & 2 & -5 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & -2 \\ 0 & -1 & 0 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 \end{bmatrix},$$

from which we obtain the Hermite normal form

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

Again, the geometric multiplicity is less than the algebraic multiplicity. We obtain the eigenvectors

$$\alpha_1 = (1, 0, 1, 0, 0)$$
$$\alpha_2 = (2, 1, 0, 0, 1).$$

Now we must compute $(A - 2I)^2$. We find

$$(A - 2I)^2 = \begin{bmatrix} 1 & 0 & -1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & -1 & 2 & 0 \\ 1 & -1 & -1 & 1 & -1 \end{bmatrix},$$

from which we obtain the Hermite normal form

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

For the third basis vector we can choose

$$\alpha_3 = (0, 1, 0, 1, 0).$$

Then

$$(A - 2I)\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix};$$

hence, we have $\beta_3 = \alpha_3$, $\beta_1 = \alpha_1$, and we can choose $\beta_2 = \alpha_2$.

In a similar fashion we find $\beta_4 = (-1, 0, 0, 1, 0)$ and $\beta_5 = (2, 0, 0, 0, 1)$ corresponding to the eigenvalue 3. β_4 is an eigenvector and $(A - 3I)\beta_5 = \beta_4$.