CHAPTER 1

Linear Equations

CHAPTER 2

Vector Spaces

CHAPTER 3

Linear Transformations

CHAPTER 4

Polynomials

CHAPTER 5

Determinants

5.1. Commutative Rings

In this chapter we shall prove the essential facts about determinants of square matrices. We shall do this not only for matrices over a field, but also for matrices with entries which are 'scalars' of a more general type. There are two reasons for this generality. First, at certain points in the next chapter, we shall find it necessary to deal with determinants of matrices with polynomial entries. Second, in the treatment of determinants which we present, one of the axioms for a field plays no role, namely, the axiom which guarantees a multiplicative inverse for each non-zero element. For these reasons, it is appropriate to develop the theory of determinants for matrices, the entries of which are elements from a commutative ring with identity.

DEFINITION 5.1. A *ring* is a set K, together with two operations $(x, y) \to x + y$ and $(x, y) \to xy$ satisfying

(1) K is a commutative group under the operation $(x, y) \to x + y$ (K is a commutative group under addition);
(2) $(xy)\, z = x\, (yz)$ (multiplication is associative);
(3) $x\,(y + z) = xy + xz$; $(y + z)\, x = yx + zx$ (the two distributive laws hold).

If $xy = yx$ for all x and y in K, we say that the ring K is *commutative*. If there is an element 1 in K such that $1x = x1 = x$ for each x, K is said to be a *ring with identity*, and 1 is called the *identity* for K.

We are interested here in commutative rings with identity. Such a ring can be described briefly as a set K, together with two operations which satisfy all the axioms for a field given in Chapter 1, except possibly for axiom **??** and the condition $1 \neq 0$. Thus, a field is a commutative ring with non-zero identity such that to each non-zero x there corresponds an element x^{-1} with $xx^{-1} = 1$. The set of integers, with the usual operations, is a commutative ring with identity which is not a field. Another commutative ring with identity is the set of all polynomials over a field, together with the addition and multiplication which we have defined for polynomials.

If K is a commutative ring with identity, we define an $m \times n$ matrix over K to be a function A from the set of pairs (i, j) of integers, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, into K. As usual we represent such a matrix by a rectangular array having m rows and n columns. The sum and product of matrices over K are defined as for matrices over a field

$$(A + B)_{ij} = A_{ij} + B_{ij},$$
$$(AB)_{ij} = \sum_k A_{ik} B_{kj},$$

the sum being defined when A and B have the same number of rows and the same number of columns, the product being defined when the number of columns of A is equal to the number of rows of B. The basic algebraic properties of these operations are again valid. For example,

$$A(B+C) = AB + AC, \qquad (AB)C = A(BC), \qquad \text{etc.}$$

As in the case of fields, we shall refer to the elements of K as scalars. We may then define linear combinations of the rows or columns of a matrix as we did earlier. Roughly speaking, all that we previously did for matrices over a field is valid for matrices over K, excluding those results which depended upon the ability to 'divide' in K.

5.2. Determinant Functions

Let K be a commutative ring with identity. We wish to assign to each $n \times n$ (square) matrix over K a scalar (element of K) to be known as the determinant of the matrix. It is possible to define the determinant of a square matrix A by simply writing down a formula for this determinant in terms of the entries of A. One can then deduce the various properties of determinants from this formula. However, such a formula is rather complicated, and to gain some technical advantage we shall proceed as follows. We shall define a 'determinant function' on $K^{n\times n}$ as a function which assigns to each $n \times n$ matrix over K a scalar, the function having these special properties. It is linear as a function of each of the rows of the matrix; its value is 0 on any matrix having two equal rows; and its value on the $n \times n$ identity matrix is 1. We shall prove that such a function exists, and then that it is unique, i.e., that there is precisely one such function. As we prove the uniqueness, an explicit formula for the determinant will be obtained, along with many of its useful properties.

This section will be devoted to the definition of 'determinant function' and to the proof that at least one such function exists.

Definition 5.2. Let K be a commutative ring with identity, n a positive integer, and let D be a function which assigns to each $n \times n$ matrix A over K a scalar $D(A)$ in K. We say that D is *n-linear* if for each i, $1 \leqslant i \leqslant n$, D is a linear function of the ith row when the other $(n-1)$ rows are held fixed.

This definition requires some clarification. If D is a function from $K^{n\times n}$ into K, and if $\alpha_1, \ldots, \alpha_n$ are the rows of the matrix A, let us also write

$$D(A) = D(\alpha_1, \ldots, \alpha_n)$$

that is, let us also think of D as the function of the rows of A. The statement that D is n-linear then means

$$\begin{aligned} D(\alpha_1, \ldots, c\alpha_i + \alpha_i', \ldots, \alpha_n) = cD(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) \\ + D(\alpha_1, \ldots, \alpha_i', \ldots, \alpha_n). \end{aligned} \tag{5.1}$$

If we fix all rows except row i and regard D as a function of the ith row, it is often convenient to write $D(\alpha_i)$ for $D(A)$. Thus, we may abbreviate (5.1) to

$$D(c\alpha_i + \alpha_i') = cD(\alpha_i) + D(\alpha_i')$$

so long as it is clear what the meaning is.

Example 5.1. Let $k_1, \ldots, k_n$ be positive integers, $1 \leqslant k_i \leqslant n$, and let a be an element of K. For each $n \times n$ matrix A over K, define

$$D(A) = aA(1, k_1) \cdots A(n, k_n). \tag{5.2}$$

Then the function D defined by (5.2) is n-linear. For, if we regard D as a function of the ith row of A, the others being fixed, we may write

$$D(\alpha_i) = A(i, k_i)\, b$$

where b is some fixed element of K. Let $\alpha_i' = (A_{i1}', \dots, A_{in}')$. Then we have

$$D(c\alpha_i + \alpha_i') = [cA(i, k_i) + A'(i, k_i)]\, b = cD(\alpha_i) + D(\alpha_i').$$

Thus D is a linear function of each of the rows of A.

A particular n-linear function of this type is

$$D(A) = A_{11}A_{22} \cdots A_{nn}.$$

In other words, the 'product of the diagonal entries' is an n-linear function on $K^{n\times n}$.

EXAMPLE 5.2. Let us find all 2-linear functions on 2×2 matrices over K. Let D be such a function. If we denote the rows of the 2×2 identity matrix by ϵ_1, ϵ_2, we have

$$D(A) = D(A_{11}\epsilon_1 + A_{12}\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2).$$

Using the fact that D is 2-linear, (5.1), we have

$$\begin{aligned} D(A) &= A_{11}D(\epsilon_1, A_{21}\epsilon_1 + A_{22}\epsilon_2) + A_{12}D(\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2) \\ &= A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) \\ &\qquad + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2). \end{aligned}$$

Thus D is completely determined by the four scalars

$$D(\epsilon_1, \epsilon_1), \qquad D(\epsilon_1, \epsilon_2), \qquad D(\epsilon_2, \epsilon_1), \qquad \text{and} \qquad D(\epsilon_2, \epsilon_2).$$

The reader should find it easy to verify the following. If a, b, c, d are any four scalars in K and if we define

$$D(A) = A_{11}A_{21}a + A_{11}A_{22}b + A_{12}A_{21}c + A_{12}A_{22}d$$

then D is a 2-linear function on 2×2 matrices over K and

$$\begin{aligned} D(\epsilon_1, \epsilon_1) &= a, & D(\epsilon_1, \epsilon_2) &= b, \\ D(\epsilon_2, \epsilon_1) &= c, & D(\epsilon_2, \epsilon_2) &= d. \end{aligned}$$

LEMMA 5.1. *A linear combination of n-linear functions is n-linear.*

PROOF. It suffices to prove that a linear combination of two n-linear functions is n-linear. Let D and E be n-linear functions. If a and b belong to K, the linear combination $aD + bE$ is of course defined by

$$(aD + bE)(A) = aD(A) + bE(A).$$

Hence, if we fix all rows except row i

$$\begin{aligned} (aD + bE)(c\alpha_i + \alpha_i') &= aD(c\alpha_i + \alpha_i') + bE(c\alpha_i + \alpha_i') \\ &= acD(\alpha_i) + aD(\alpha_i') + bcE(\alpha_i) + bE(\alpha_i') \\ &= c(aD + bE)(\alpha_i) + (aD + bE)(\alpha_i'). \end{aligned}$$ □

If K is a field and V is the set of $n \times n$ matrices over K, the above lemma says the following. The set of n-linear functions on V is a subspace of the space of all functions from V into K.

EXAMPLE 5.3. Let D be the function defined on 2×2 matrices over K by

$$D(A) = A_{11}A_{22} - A_{12}A_{21}. \tag{5.3}$$

Now D is the sum of two functions of the type described in Example 5.1:

$$\begin{aligned} D &= D_1 + D_2, \\ D_1(A) &= A_{11}A_{22}, \\ D_2(A) &= -A_{12}A_{21}. \end{aligned}$$

By the above lemma, D is a 2-linear function. The reader who has had any experience with determinants will not find this surprising, since he will recognize (5.3) as the usual definition of the determinant of a 2×2 matrix. Of course the function D we have just defined is not a typical 2-linear function. It has many special properties. Let us note some of these properties. First, if I is the 2×2 identity matrix, then $D(I) = 1$, i.e., $D(\epsilon_1, \epsilon_2) = 1$. Second, if the two rows of A are equal, then

$$D(A) = A_{11}A_{12} - A_{12}A_{11} = 0.$$

Third, if A' is the matrix obtained from a 2×2 matrix A by interchanging its rows, then $D(A') = -D(A)$; for

$$\begin{aligned} D(A') &= A'_{11}A'_{22} - A'_{12}A'_{21} \\ &= A_{21}A_{12} - A_{22}A_{11} \\ &= -D(A). \end{aligned}$$

Definition 5.3. Let D be an n-linear function. We say D is *alternating* (or *alternate*) if the following two conditions are satisfied:

(1) $D(A) = 0$ whenever two rows of A are equal.
(2) If A' is a matrix obtained from A by interchanging two rows of A, then $D(A') = -D(A)$.

We shall prove below that any n-linear function D which satisfies (1) automatically satisfies (2). We have put both properties in the definition of alternating n-linear function as a matter of convenience. The reader will probably also note that if D satisfies (2) and A is a matrix with two equal rows, then $D(A) = -D(A)$. It is tempting to conclude that D satisfies condition (1) as well. This is true, for example, if K is a field in which $1 + 1 \neq 0$, but in general (1) is not a consequence of (1).

Definition 5.4. Let K be a commutative ring with identity, and let n be a: positive integer. Suppose D is a function from $n \times n$ matrices over K into K. We say that D is a *determinant function* if D is n-linear, alternating, and $D(I) = 1$.

As we stated earlier, we shall ultimately show that there is exactly one determinant function on $n \times n$ matrices over K. This is easily seen for 1×1 matrices $A = [a]$ over K. The function D given by $D(A) = a$ is a determinant function, and clearly this is the only determinant function on 1×1 matrices. We are also in a position to dispose of the case $n = 2$. The function

$$D(A) = A_{11}A_{22} - A_{12}A_{21} \tag{5.4}$$

was shown in Example 5.3 to be a determinant function. Furthermore, the formula exhibited in Example 5.2 shows that D is the only determinant function on 2×2 matrices. For we showed that for any 2-linear function D

$$\begin{aligned} D(A) = A_{11}A_{21}D(\epsilon_1, \epsilon_1) &+ A_{11}A_{22}D(\epsilon_1, \epsilon_2) \\ &+ A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2). \end{aligned}$$

If D is alternating, then

$$D(\epsilon_1, \epsilon_1) = D(\epsilon_2, \epsilon_2) = 0$$

and

$$D(\epsilon_2, \epsilon_1) = -D(\epsilon_1, \epsilon_2) = -D(I).$$

If D also satisfies $D(I) = 1$, then

$$D(A) = A_{11}A_{22} - A_{12}A_{21}.$$

EXAMPLE 5.4. Let F be a field and let D be any alternating 3-linear function on 3×3 matrices over the polynomial ring $F[x]$.

Let

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

If we denote the rows of the 3×3 identity matrix by ϵ_1, ϵ_2, ϵ_3, then

$$D(A) = D\left(x\epsilon_1 - x^2\epsilon_3,\, \epsilon_2,\, \epsilon_1 + x^3\epsilon_3\right).$$

Since D is linear as a function of each row,

$$\begin{aligned} D(A) &= xD(\epsilon_1,\, \epsilon_2,\, \epsilon_1 + x^3\epsilon_3) - x^2 D(\epsilon_3,\, \epsilon_2,\, \epsilon_1 + x^3\epsilon_3) \\ &= xD(\epsilon_1,\, \epsilon_2,\, \epsilon_1) + x^4 D(\epsilon_1,\, \epsilon_2,\, \epsilon_3) - x^2 D(\epsilon_3,\, \epsilon_2,\, \epsilon_1) - x^5 D(\epsilon_3,\, \epsilon_2,\, \epsilon_3). \end{aligned}$$

Because D is alternating it follows that

$$D(A) = \left(x^4 + x^2\right) D(\epsilon_1,\, \epsilon_2,\, \epsilon_3).$$

LEMMA 5.2. *Let D be a 2-linear function with the property that $D(A) = 0$ for all 2×2 matrices A over K having equal rows. Then D is alternating.*

PROOF. What we must show is that if A is a 2×2 matrix and A' is obtained by interchanging the rows of A, then $D(A') = -D(A)$. If the rows of A are α and β, this means we must show that $D(\beta,\, \alpha) = -D(\alpha,\, \beta)$. Since D is 2-linear,

$$D(\alpha + \beta,\, \alpha + \beta) = D(\alpha,\, \alpha) + D(\alpha,\, \beta) + D(\beta,\, \alpha) + D(\beta,\, \beta).$$

By our hypothesis $D(\alpha + \beta,\, \alpha + \beta) = D(\alpha,\, \alpha) = D(\beta,\, \beta) = 0$. So

$$0 = D(\alpha,\, \beta) + D(\beta,\, \alpha).$$

□

LEMMA 5.3. *Let D be an n-linear function on $n \times n$ matrices over K. Suppose D has the property that $D(A) = 0$ whenever two adjacent rows of A are equal. Then D is alternating.*

PROOF. □

CHAPTER 6

Elementary Canonical Forms

6.1. Introduction

We have mentioned earlier that our principal aim is to study linear transformations on finite-dimensional vector spaces. By this time, we have seen many specific examples of linear transformations, and we have proved a few theorems about the general linear transformation. In the finite-dimensional case we have utilized ordered bases to represent such transformations by matrices, and this representation adds to our insight into their behavior. We have explored the vector space $L(V, W)$, consisting of the linear transformations from one space into another, and we have explored the linear algebra $L(V, V)$, consisting of the linear transformations of a space into itself.

In the next two chapters, we shall be preoccupied with linear operators. Our program is to select a single linear operator T on a finite-dimensional vector space V and to 'take it apart to see what makes it tick.' At this early stage, it is easiest to express our goal in matrix language: Given the linear operator T, find an ordered basis for V in which the matrix of T assumes an especially simple form.

Here is an illustration of what we have in mind. Perhaps the simplest matrices to work with, beyond the scalar multiples of the identity, are the diagonal matrices:

$$D = \begin{bmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ 0 & 0 & c_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & c_n \end{bmatrix}. \tag{6.1}$$

Let T be a linear operator on an n-dimensional space V. If we could find an ordered basis $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ for V in which T were represented by a diagonal matrix D (6.1), we would gain considerable information about T. For instance, simple numbers associated with T, such as the rank of T or the determinant of T, could be determined with little more than a glance at the matrix D. We could describe explicitly the range and the null space of T. Since $[T]_{\mathcal{B}} = D$ if and only if

$$T_{\alpha_k} = c_k \alpha_k, \qquad k = 1, \dots, n, \tag{6.2}$$

the range would be the subspace spanned by those α_k's for which $c_k \neq 0$ and the null space would be spanned by the remaining α_k's. Indeed, it seems fair to say that, if we knew a basis $\mathcal{B}$ and a diagonal matrix D such that $[T]_{\mathcal{B}} = D$, we could answer readily any question about T which might arise.

Can each linear operator T be represented by a diagonal matrix in some ordered basis? If not, for which operators T does such a basis exist? How can we find such a basis if there is one? If no such basis exists, what is the simplest type of matrix by which we can represent T? These are some of the questions which we shall attack in this (and the next) chapter. The form of our questions will become more sophisticated as we learn what some of the difficulties are.

6.2. Characteristic Values

The introductory remarks of the previous section provide us with a starting point for our attempt to analyze the general linear operator T. We take our cue from (6.2), which suggests that we should study vectors which are sent by T into scalar multiples of themselves.

DEFINITION 6.1. Let V be a vector space over the field F and let T be a linear operator on V. A *characteristic value* of T is a scalar c in F such that there is a non-zero vector α in V with $T\alpha = c\alpha$. If c is a characteristic value of T, then

(1) any α such that $T\alpha = c\alpha$ is called a *characteristic vector* of T associated with the characteristic value c;
(2) the collection of all α such that $T\alpha = c\alpha$ is called the *characteristic space* associated with c.

Characteristic values are often called characteristic roots, latent roots, eigenvalues, proper values, or spectral values. In this book we shall use only the name 'characteristic values.'

If T is any linear operator and c is any scalar, the set of vectors α such that $T\alpha = c\alpha$ is a subspace of V. It is the null space of the linear transformation $(T - cI)$. We call c a characteristic value of T if this subspace is different from the zero subspace, i.e., if $(T - cI)$ fails to be 1:1. If the underlying space V is finite-dimensional, $(T - cI)$ fails to be 1:1 precisely when its determinant is different from 0. Let us summarize.

THEOREM 6.1. *Let T be a linear operator on a finite-dimensional space V and let c he a scalar. The following are equivalent.*

(1) c is a characteristic value of T.
(2) The operator $(T - cI)$ is singular (not invertible).
(3) $\det(T - cI) = 0$.

The determinant criterion (3) is very important because it tells us where to look for the characteristic values of T. Since $\det(T - cI)$ is a polynomial of degree n in the variable c, we will find the characteristic values as the roots of that polynomial. Let us explain carefully.

If $\mathcal{B}$ is any ordered basis for V and $A = [T]_{\mathcal{B}}$, then $(T - cI)$ is invertible if and only if the matrix $(A - cI)$ is invertible. Accordingly, we make the following definition.

DEFINITION 6.2. If A is an $n \times n$ matrix over the field F, a *characteristic value of A in F* is a scalar c in F such that the matrix $(A - cI)$ is singular (not invertible).

Since c is a characteristic value of A if and only if $\det(A - cI) = 0$, or equivalently if and only if $\det(cI - A) = 0$, we form the matrix $(xI - A)$ with polynomial entries, and consider the polynomial $f = \det(xI - A)$. Clearly the characteristic values of A in F are just the scalars c in F such that $f(c) = 0$. For this reason f is called the *characteristic polynomial* of A. It is important to note that f is a monic polynomial which has degree exactly n. This is easily seen from the formula for the determinant of a matrix in terms of its entries.

LEMMA 6.1. *Similar matrices have the same characteristic polynomial.*

PROOF. It $B = P^{-1}AP$, then

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

$$= \det(xI - A). \qquad \square$$

This lemma enables us to define sensibly the characteristic polynomial of the operator T as the characteristic polynomial of any $n \times n$ matrix which represents T in some ordered basis for V. Just as for matrices, the characteristic values of T will be the roots of the characteristic polynomial for T. In particular, this shows us that T cannot have more than n distinct characteristic values. It is important to point out that T may not have any characteristic values.

EXAMPLE 6.1. Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

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

The characteristic polynomial for T (or for A) is

$$\det(xI - A) = \begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1.$$

Since this polynomial has no real roots, T has no characteristic values. If U is the linear operator on C^2 which is represented by A in the standard ordered basis, then U has two characteristic values, i and $-$i. Here we see a subtle point. In discussing the characteristic values of a matrix A, we must be careful to stipulate the field involved. The matrix A above has no characteristic values in R, but has the two characteristic values i and $-$i in C.

EXAMPLE 6.2. Let A be the (real) 3×3 matrix

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

Then the characteristic polynomial for A is

$$\begin{vmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{vmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

Thus the characteristic values of A are 1 and 2.

Suppose that T is the linear operator on R^3 which is represented by A in the standard basis. Let us find the characteristic vectors of T associated with the characteristic values, 1 and 2. Now

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

It is obvious at a glance that $A - I$ has rank equal to 2 (and hence $T - I$ has nullity equal to 1). So the space of characteristic vectors associated with the characteristic value 1 is one-dimensional. The vector $\alpha_1 = (1, 0, 2)$ spans the null space of $T - I$. Thus $T\alpha = \alpha$ if and only if α is a scalar multiple of α_1. Now consider

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

Evidently $A - 2I$ also has rank 2, so that the space of characteristic vectors associated with the characteristic value 2 has dimension 1. Evidently $T\alpha = 2\alpha$ if and only if α is a scalar multiple of $\alpha_2 = (1, 1, 2)$.

DEFINITION 6.3. Let T be a linear operator on the finite-dimensional space V. We say that T is *diagonalizable* if there is a basis for V each vector of which is a characteristic vector of T.

The reason for the name should be apparent; for, if there is an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V in which each α_i is a characteristic vector of T, then the matrix of T in the ordered basis $\mathcal{B}$ is diagonal. If $T_{\alpha_i} = c_i\alpha_i$, then

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}.$$

We certainly do not require that the scalars $c_1, \ldots, c_n$ be distinct; indeed, they may all be the same scalar (when T is a scalar multiple of the identity operator).

One could also define T to be diagonalizable when the characteristic vectors of T span V. This is only superficially different from our definition, since we can select a basis out of any spanning set of vectors.

For Examples 6.1 and 6.2 we purposely chose linear operators T on R^n which are not diagonalizable. In Example 6.1, we have a linear operator on R^2 which is not diagonalizable, because it has no characteristic values. In Example 6.2, the operator T has characteristic values; in fact, the characteristic polynomial for T factors completely over the real number field: $f = (x-1)(x-2)^2$. Nevertheless T fails to be diagonalizable. There is only a one-dimensional space of characteristic vectors associated with each of the two characteristic values of T. Hence, we cannot possibly form a basis for R^3 which consists of characteristic vectors of T.

Suppose that T is a diagonalizable linear operator. Let $c_1, \ldots, c_k$ be the *distinct* characteristic values of T. Then there is an ordered basis $\mathcal{B}$ in which T is represented by a diagonal matrix which has for its diagonal entries the scalars c_i, each repeated a certain number of times. If c_i is repeated d_i times, then (we may arrange that) the matrix has the block form

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1I_1 & 0 & \cdots & 0 \\ 0 & c_2I_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_kI_k \end{bmatrix} \tag{6.3}$$

where I_j is the $d_j \times d_j$ identity matrix. From that matrix we see two things. First, the characteristic polynomial for T is the product of (possibly repeated) linear factors:

$$f = (x-c_1)^{d_1} \cdots (x-c_k)^{d_k}$$

If the scalar field F is algebraically closed, e.g., the field of complex numbers, every polynomial over F can be so factored (see Section ??); however, if F is not algebraically closed, we are citing a special property of T when we say that its characteristic polynomial has such a factorization. The second thing we see from (6.3) is that d_i, the number of times which c_i is repeated as root of f, is equal to the dimension of the space of characteristic vectors associated with the characteristic value c_i. That is because the nullity of a diagonal matrix is equal to the number of zeros which it has on its main diagonal, and the matrix $[T - c_iI]_{\mathcal{B}}$ has d_i zeros on its main diagonal. This relation between the dimension of the characteristic space and the multiplicity of the characteristic value as a root of f does not seem exciting at first; however, it will provide us with a simpler way of determining whether a given operator is diagonalizable.

LEMMA 6.2. *Suppose that $T\alpha = c\alpha$. If f is any polynomial, then $f(T)\alpha = f(c)\alpha$.*

PROOF. Exercise. $\square$

LEMMA 6.3. *Let T be a linear operator on the finite-dimensional space V. Let c_1, ..., c_k be the distinct characteristic values of T and let W_i be the space of characteristic vectors associated with the characteristic value c_i. If $W = W_1 + \cdots + W_k$, then*

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

In fact, if $\mathcal{B}_i$ is an ordered basis for W_i, then $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ is an ordered basis for W.

PROOF. The space $W = W_1 + \cdots + W_k$ is the subspace spanned by all of the characteristic vectors of T. Usually when one forms the sum W of subspaces W_i, one expects that $\dim W < \dim W_1 + \cdots + \dim W_k$ because of linear relations which may exist between vectors in the various spaces. This lemma states that the characteristic spaces associated with different characteristic values are independent of one another.

Suppose that (for each i) we have a vector β_i in W_i, and assume that $\beta_1 + \cdots + \beta_k = 0$. We shall show that $\beta_i = 0$ for each i. Let f be any polynomial. Since $T\beta_i = c_i\beta_i$, the preceding lemma tells us that

$$\begin{aligned} 0 = f(T)\,0 &= f(T)\,\beta_1 + \cdots + f(T)\,\beta_k \\ &= f(c_1)\,\beta_1 + \cdots + f(c_k)\,\beta_k. \end{aligned}$$

Choose polynomials f_1, ..., f_k such that

$$f_i(c_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Then

$$\begin{aligned} 0 = f_i(T)\,0 &= \sum_j \delta_{ij}\beta_j \\ &= \beta_i. \end{aligned}$$

Now, let $\mathcal{B}_i$ be an ordered basis for W_i, and let $\mathcal{B}$ be the sequence $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. Then $\mathcal{B}$ spans the subspace $W = W_i + \cdots + W_k$. Also, $\mathcal{B}$ is a linearly independent sequence of vectors, for the following reason. Any linear relation between the vectors in $\mathcal{B}$ will have the form $\beta_1 + \cdots + \beta_k = 0$, where β_i is some linear combination of the vectors in $\mathcal{B}_i$. From what we just did, we know that $\beta_i = 0$ for each i. Since each $\mathcal{B}_i$ is linearly independent, we see that we have only the trivial linear relation between the vectors in $\mathcal{B}$. □

THEOREM 6.2. *Let T be a linear operator on a finite-dimensional space V. Let c_1, ..., c_k be the distinct characteristic values of T and let W_i be the null space of $(T - c_iI)$. The following are equivalent.*

(1) T is diagonalizable.
(2) The characteristic polynomial for T is

$$f = (X - c_1)^{d_1} \cdots (X - c_k)^{d_k}$$

and $\dim W_i = d_i$, $i = 1, \ldots, k$.
(3) $\dim W_1 + \cdots + \dim W_k = \dim V$.

PROOF. We have observed that (1) implies (2). If the characteristic polynomial f is the product of linear factors, as in (2), then $d_1 + \cdots + d_k = \dim V$. For, the sum of the d_i's is the degree of the characteristic polynomial, and that degree is $\dim V$. Therefore (2) implies (3). Suppose (3) holds. By the lemma, we must have $V = W_1 + \cdots + W_k$, i.e., the characteristic vectors of T span V. □

The matrix analogue of Theorem 6.2 may be formulated as follows. Let A be an $n \times n$ matrix with entries in a field F, and let c_1, ..., c_k be the distinct characteristic values of A in F. For each i, let W_i be the space of column matrices X (with entries in F) such that

$$(A - c_i I) X = 0,$$

and let $\mathcal{B}_i$ be an ordered basis for W_i. The bases $\mathcal{B}_1$, ..., $\mathcal{B}_k$ collectively string together to form the sequence of columns of a matrix P:

$$P = [P_1, P_2, ...] = (\mathcal{B}_1, \dots, \mathcal{B}_k) .$$

The matrix A is similar over F to a diagonal matrix if and only if P is a square matrix. When P is square, P is invertible and $P^{-1}AP$ is diagonal.