CHAPTER 1

Linear Equations

CHAPTER 2

Vector Spaces

CHAPTER 3

Linear Transformations

CHAPTER 4

Polynomials

CHAPTER 5

Determinants

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 $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ 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

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

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.