$\sigma_2\gamma\sigma_1^{-1}$ lies in $G(r+s, t)$, it follows that τ_1 and τ_2 are in the same left coset of $G(r+s, t)$. Therefore, $\tau_1 = \tau_2$, and $\sigma_1 = \sigma_2\gamma$. But this implies that σ_1 and σ_2 (regarded as elements of S_{r+s}) lie in the same coset of $G(r, s)$; hence $\sigma_1 = \sigma_2$. Therefore, the products $\tau\sigma$ corresponding to the

$$\frac{(r+s+t)!}{(r+s)!t!}\,\frac{(r+s)!}{r!s!}$$

pairs (τ, σ) in $T \times S$ are all distinct and lie in distinct cosets of $G(r, s, t)$. Since there are exactly

$$\frac{(r+s+t)!}{r!s!t!}$$

left cosets of $G(r, s, t)$ in S_{r+s+t}, it follows that $(L \wedge M) \wedge N = E$. By an analogous argument, $L \wedge (M \wedge N) = E$ as well. ∎

Example 13. The exterior product is closely related to certain formulas for evaluating determinants known as the **Laplace expansions.** Let K be a commutative ring with identity and n a positive integer. Suppose that $1 \leq r < n$, and let L be the alternating r-linear form on K^n defined by

$$L(\alpha_1, \ldots, \alpha_r) = \det\begin{bmatrix} A_{11} & \cdots & A_{1r} \\ \vdots & & \vdots \\ A_{r1} & \cdots & A_{rr}\end{bmatrix}.$$

If $s = n - r$ and M is the alternating s-linear form

$$M(\alpha_1, \ldots, \alpha_s) = \det\begin{bmatrix} A_{1(r+1)} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{s(r+1)} & \cdots & A_{sn}\end{bmatrix}$$

then $L \wedge M = D$, the determinant function on K^n. This is immediate from the fact that $L \wedge M$ is an alternating n-linear form and (as can be seen)

$$(L \wedge M)(\epsilon_1, \ldots, \epsilon_n) = 1.$$

If we now describe $L \wedge M$ in the correct way, we obtain one Laplace expansion for the determinant of an $n \times n$ matrix over K.

In the permutation group S_n, let G be the subgroup which permutes the sets $\{1, \ldots, r\}$ and $\{r+1, \ldots, n\}$ within themselves. Each left coset of G contains precisely one permutation σ such that $\sigma 1 < \sigma 2 < \ldots < \sigma r$ and $\sigma(r+1) < \ldots < \sigma n$. The sign of this permutation is given by

$$\operatorname{sgn}\sigma = (-1)^{\sigma 1 + \cdots + \sigma r + (r(r-1)/2)}.$$

The wedge product $L \wedge M$ is given by

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \Sigma\, (\operatorname{sgn}\sigma)L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r})M(\alpha_{\sigma(r+1)}, \ldots, \alpha_{\sigma n})$$

where the sum is taken over a collection of σ's, one from each coset of G. Therefore,

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \sum_{j_1 < \cdots < j_r} e_J \, L(\alpha_{j_1}, \ldots, \alpha_{j_r}) M(\alpha_{k_1}, \ldots, \alpha_{k_s})$$

where

$$\begin{aligned} e_J &= (-1)^{j_1 + \cdots + j_r + (r(r-1)/2)} \\ k_i &= \sigma(r + i). \end{aligned}$$

In other words,

$$\det A = \sum_{j_1 < \cdots < j_r} e_J \begin{vmatrix} A_{j_1,1} & \cdots & A_{j_1,r} \\ \vdots & & \vdots \\ A_{j_r,1} & \cdots & A_{j_r,r} \end{vmatrix} \begin{vmatrix} A_{k_1,r+1} & \cdots & A_{k_1,n} \\ \vdots & & \vdots \\ A_{k_r,r+1} & \cdots & A_{k_r,n} \end{vmatrix}$$

This is one Laplace expansion. Others may be obtained by replacing the sets $\{1, \ldots, r\}$ and $\{r + 1, \ldots, n\}$ by two different complementary sets of indices.

If V is a K-module, we may put the various form modules $\Lambda^r(V)$ together and use the exterior product to define a ring. For simplicity, we shall do this only for the case of a free K-module of rank n. The modules $\Lambda^r(V)$ are then trivial for $r > n$. We define

$$\Lambda(V) = \Lambda^0(V) \oplus \Lambda^1(V) \oplus \cdots \oplus \Lambda^n(V).$$

This is an external direct sum—something which we have not discussed previously. The elements of $\Lambda(V)$ are the $(n + 1)$-tuples $(L_0, \ldots, L_n)$ with L_r in $\Lambda^r(V)$. Addition and multiplication by elements of K are defined as one would expect for $(n + 1)$-tuples. Incidentally, $\Lambda^0(V) = K$. If we identify $\Lambda^r(K)$ with the $(n + 1)$-tuples $(0, \ldots, 0, L, 0, \ldots, 0)$ where L is in $\Lambda^r(K)$, then $\Lambda^r(K)$ is a submodule of $\Lambda(V)$ and the direct sum decomposition

$$\Lambda(V) = \Lambda^0(V) \oplus \cdots \oplus \Lambda^n(V)$$

holds in the usual sense. Since $\Lambda^r(V)$ is a free K-module of rank $\binom{n}{r}$, we see that $\Lambda(V)$ is a free K-module and

$$\begin{aligned} \operatorname{rank} \Lambda(V) &= \sum_{r=0}^{n} \binom{n}{r} \\ &= 2^n. \end{aligned}$$

The exterior product defines a multiplication in $\Lambda(V)$: Use the exterior product on forms and extend it linearly to $\Lambda(V)$. It distributes over the addition of $\Lambda(V)$ and gives $\Lambda(V)$ the structure of a ring. This ring is the **Grassman ring** over V^*. It is not a commutative ring, e.g., if L, M are respectively in Λ^r and Λ^s, then

$$L \wedge M = (-1)^{rs} M \wedge L.$$

But, the Grassman ring is important in several parts of mathematics.

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. *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*

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

(b) *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 1. *Let* T *be a linear operator on a finite-dimensional space* V *and let* c *be a scalar. The following are equivalent.*

(i) c *is a characteristic value of* T.
(ii) *The operator* (T − cI) *is singular (not invertible).*
(iii) *det* (T − cI) = 0.

The determinant criterion (iii) 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. *If* A *is an* n × 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. *Similar matrices have the same characteristic polynomial.*

Proof. If $B = P^{-1}AP$, then

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

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 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 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

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