CHAPTER III

INVARIANT FACTORS AND ELEMENTARY DIVISORS

3.01 Elementary transformations. By an elementary transformation of a matric polynomial $a(\lambda) = \| a_{ij} \|$ is meant one of the following operations on the rows or columns.

Type I. The operation of adding to a row (column) a different row (column) multiplied by a scalar polynomial $\theta(\lambda)$.

Type II. The operation of interchanging two rows (columns).

Type III. The operation of multiplying a row (column) by a constant $k \neq 0$.

These transformations can be performed algebraically as follows.

Type I. Let

$$P_{ij} = 1 + \theta(\lambda)e_{ij} \quad (i \neq j),$$

$\theta(\lambda)$ being a scalar polynomial; then $| P_{ij} | = 1$ and

$$P_{ij}a = \sum_{p,q} a_{pq}e_{pq} + \theta \sum_{q} a_{jq}e_{iq}$$

which is the matrix derived from $a(\lambda)$ by adding θ times the jth row to the ith. The corresponding operation on the columns is equivalent to forming the product aP_{ji}.

Type II. Let Q_{ij} be the matrix

$$Q_{ij} = 1 - e_{ii} - e_{jj} + e_{ij} + e_{ji} \qquad (i \neq j)$$

that is, Q_{ij} is the matrix derived from the identity matrix by inserting 1 in place of 0 in the coefficients of e_{ij} and e_{ji} and 0 in place of 1 in the coefficients of e_{ii} and e_{jj}; then $| Q_{ij} | = -1$ and

$$Q_{ij}a = \sum_{p,q} a_{pq}e_{pq} - \sum_{q} a_{iq}e_{iq} - \sum_{q} a_{jq}e_{jq} + \sum_{q} a_{jq}e_{iq} + \sum_{q} a_{iq}e_{jq},$$

that is, $Q_{ij}a$ is derived from a by interchanging the ith and jth rows. Similarly aQ_{ij} is obtained by interchanging the ith and jth columns.

Since any permutation can be effected by a succession of transpositions, the corresponding transformation in the rows (columns) of a matrix can be produced by a succession[1] of transformations of Type II.

Type III. This transformation is effected on the rth row (column) by multiplying on the left (right) by $R = 1 + (k - 1)e_{rr}$; it is used only when it is convenient to make the leading coefficient in some term equal to 1.

[1] The transformation corresponding to the substitution $\begin{pmatrix} 1 & 2 & \dots & n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$ is $Q = \sum_{i} e_{ip_i}$.

The inverses of the matrices used in these transformations are

$$P_{ij}^{-1} = 1 - \theta e_{ij}, \quad Q_{ij}^{-1} = Q_{ij}, \quad R^{-1} = 1 + (k^{-1} - 1)e_{rr};$$

these inverses are elementary transformations. The transverses are also elementary since $P'_{ij} = P_{ji}$, and Q_{ij} and R are symmetric.[2]

A matric polynomial $b(\lambda)$ which is derived from $a(\lambda)$ by a sequence of elementary transformations is said to be *equivalent* to $a(\lambda)$; every such polynomial has the form $p(\lambda)a(\lambda)q(\lambda)$ where p and q are products of elementary transformations. Since the inverse of an elementary transformation is elementary, $a(\lambda)$ is also equivalent to $b(\lambda)$. Further, the inverses of p and q are polynomials so that these are what we have already called elementary polynomials; we shall see later that every elementary polynomial can be derived from 1 by a sequence of elementary transformations.

In the following sections we require two lemmas whose proofs are almost immediate.

LEMMA 1. *The rank[3] of a matrix is not altered by an elementary transformation.*

For if $|P| \neq 0$, AP and PA have the same rank as A (§1.10).

LEMMA 2. *The highest common factor of the coordinates of a matric polynomial is not altered by an elementary transformation.*

This follows immediately from the definition of elementary transformations.

3.02 The normal form of a matrix. The theorem we shall prove in this section is as follows.

THEOREM 1. *If $a(\lambda)$ is a matric polynomial of rank r, it can be reduced by elementary transformations to a diagonal matrix*

$$(1) \qquad \sum_1^r \alpha_i(\lambda)e_{ji} = \begin{Vmatrix} \alpha_1(\lambda) & & & & & \\ & \alpha_2(\lambda) & & & & \\ & & \ddots & & & \\ & & & \alpha_r(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{Vmatrix} = P(\lambda)a(\lambda)Q(\lambda),$$

[2] The definition of an elementary transformation given above is the most convenient but not the only possible one. All three transformations have the form $T = 1 + xSy$ with the condition that $1 + Sxy$ is not 0 and is independent of λ.

[3] By the rank of a matric polynomial is meant the order of the highest minor which does not vanish identically. For particular values of λ the rank may be smaller than r; there are always values of λ for which it equals r and it cannot be greater.

where the coefficient of the highest power of λ *in each polynomial* $\alpha_i(\lambda)$ *is* 1, α_i *is a factor of* $\alpha_{i+1}, \cdots, \alpha_r$ $(i = 1, 2, \cdots, r-1)$, *and* $P(\lambda)$, $Q(\lambda)$ *are elementary polynomials.*

We shall first show that, if the coordinate of $a(\lambda)$ of minimum degree m, say a_{pq}, is not a factor of every other coordinate, then $a(\lambda)$ is equivalent to a matrix in which the degree of the coordinate of minimum degree is less than m.

Suppose that a_{pq} is not a factor of a_{pi} for some i; then we may set $a_{pi} = \beta a_{pq} + a'_{pi}$ where β is integral and a'_{pi} is not 0 and is of lower degree than m. Subtracting β times the qth column from the ith we have an equivalent matrix in which the coordinate[4] (p, i) is a'_{pi} whose degree is less than m. The same reasoning applies if a_{pq} is not a factor of every coordinate a_{iq} in the qth column.

After a finite number of such steps we arrive at a matrix in which a coordinate of minimum degree, say k_{pq}, is a factor of all the coordinates which lie in the same row or column, but is possibly not a factor of some other coordinate k_{ij}. When this is so, let $k_{pj} = \beta k_{pq}$, $k_{iq} = \gamma k_{pq}$ where β and γ are integral. If we now add $(1 - \beta)$ times the qth column to the jth, (p, j) and (i, j) become respectively

$$k'_{pj} = k_{pj} + (1 - \beta)k_{pq} = k_{pq}, \quad k'_{ij} = k_{ij} + (1 - \beta)k_{iq} = k_{ij} + (1 - \beta)\gamma k_{pq}.$$

Here either the degree of k'_{ij} is less than that of k_{pq}, or k'_{pj} has the minimum degree and is not a factor of k'_{ij}, which lies in the same column, and hence the minimum degree can be lowered as above.

The process just described can be repeated so long as the coordinate of lowest degree is not a factor of every other coordinate and, since each step lowers the minimum degree, we derive in a finite number of steps a matrix $\| b'_{ij} \|$ which is equivalent to $a(\lambda)$ and in which the coordinate of minimum degree is in fact a divisor of every other coordinate; and further we may suppose that $b'_{11} = \alpha_1(\lambda)$ is a coordinate of minimum degree and set $b'_{1i} = \gamma_i b'_{11}$, $b'_{j1} = \delta_j b'_{11}$. Subtracting γ_i times the first column from the ith and then δ_j times the first row from the jth $(i, j = 2, 3, \cdots, n)$ all the coordinates in the first row and column except b'_{11} become 0, and we have an equivalent matrix in the form

$$(2) \qquad \begin{matrix} \alpha_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & b_{22} & b_{23} & \cdots & b_{2n} \\ 0 & b_{32} & b_{33} & \cdots & b_{3n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & b_{n2} & b_{n3} & \cdots & b_{nn} \end{matrix}$$

in which α_1 is a factor of every b_{ij}. The coefficient of the highest power of λ in α_1 may be made 1 by a transformation of type III.

The theorem now follows readily by induction. For, assuming it is true for

[4] That is, the coordinate in the pth row and ith column.

matrices of order $n - 1$, the matrix of this order formed by the b's in (2) can be reduced to the diagonal matrix

$$\begin{pmatrix} \alpha_2(\lambda) & & & & & \\ & \alpha_3(\lambda) & & & & \\ & & \ddots & & & \\ & & & \alpha_s(\lambda) & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix}$$

where the α's satisfy the conditions of the theorem and each has α_1 as a factor (§3.01, Lemma 2). Moreover, the elementary transformations by which this reduction is carried out correspond to transformations affecting the last $n - 1$ rows and columns alone in (2) and, because of the zeros in the first row and column, these transformations when applied to (2) do not affect its first row and column; also, since elementary transformations do not affect the rank (§3.01, Lemma 1), s equals r and $a(\lambda)$ has therefore been reduced to the form required by the theorem.

The theorem is clearly true for matrices of order 1 and hence is true for any order.

Corollary. A matric polynomial whose determinant is independent of λ and is not 0, that is, an elementary polynomial, can be derived from 1 by the product of a finite number of elementary transformations.

The polynomials α_i are called the *invariant factors* of $a(\lambda)$.

3.03 **Determinantal and invariant factors.** The *determinantal factor* of the sth order, D_s, of a matric polynomial $a(\lambda)$ is defined as the highest common factor of all minors of order s, the coefficient of the highest power of λ being taken as 1. An elementary transformation of type I either leaves a given minor unaltered or changes it into the sum of that minor and a multiple of another of the same order, and a transformation of type II simply permutes the minors of a given order among themselves, while one of type III merely multiplies a minor by a constant different from 0. Hence equivalent matrices have the same determinantal factors. Bearing this in mind we see immediately from the form of (1) that the determinantal factors of $a(\lambda)$ are given by

$$D_s = \alpha_1\alpha_2 \cdots \alpha_s \qquad (s = 1, 2, \cdots, r), \qquad D_s = 0 \ (s > r),$$

so that[5]

$$\alpha_s = D_s/D_{s-1}.$$

The invariant factors are therefore known when the determinantal factors are given, and vice versa.

[5] Since α_{s-1} is a factor of α_s, it follows that also D_s^2 is a factor of $D_{s-1}D_{s+1}$.

The definitions of this and the preceding sections have all been made relative to the fundamental basis. But we have seen in §1.08 that, if a_1 is the matrix with the same array of coordinates as a but relative to another basis, then there exists a non-singular constant matrix b such that $a = b^{-1}a_1b$ so that a and a_1 are equivalent matrices. In terms of the new basis a_1 has the same invariant factors as a does in terms of the old and a, being equivalent to a_1, has therefore the same invariant factors in terms of the new basis as it has in the old. Hence the invariant and determinantal factors of a matric polynomial are independent of the (constant) basis in terms of which its coordinates are expressed.

The results of this section may be summarized as follows.

THEOREM 2. *Two matric polynomials are equivalent if, and only if, they have the same invariant factors.*

3.04 **Non-singular linear polynomials.** In the case of linear polynomials Theorem 2 can be made more precise as follows.

THEOREM 3. *If $a\lambda + b$ and $c\lambda + d$ are non-singular linear polynomials which have the same invariant factors, and if $|\, c \,| \neq 0$, there exist non-singular constant matrices p and q such that*

$$p(a\lambda + b)q = c\lambda + d.$$

We have seen in Theorem 2 that there exist elementary polynomials $P(\lambda)$, $Q(\lambda)$ such that

$$c\lambda + d = P(\lambda)(a\lambda + b)Q(\lambda). \tag{3}$$

Since $|\, c \,| \neq 0$, we can employ the division transformation to find matric polynomials p_1, q_1 and constant matrices p, q such that

$$P(\lambda) = (c\lambda + d)p_1 + p, \quad Q(\lambda) = q_1(c\lambda + d) + q.$$

Using this in (3) we have

$$\begin{aligned} c\lambda + d = {} & p(a\lambda + b)q + (c\lambda + d)p_1(a\lambda + b)Q + P(a\lambda + b)q_1(c\lambda + d) \\ & - (c\lambda + d)p_1(a\lambda + b)q_1(c\lambda + d) \end{aligned} \tag{4}$$

and, since from (3)

$$(a\lambda + b)Q = P^{-1}(c\lambda + d), \quad P(a\lambda + b) = (c\lambda + d)Q^{-1},$$

we may write in place of (4)

$$\begin{aligned} p(a\lambda + b)q &= [1 - (c\lambda + d)(p_1P^{-1} + Q^{-1}q_1 - p_1(a\lambda + b)q_1)](c\lambda + d) \\ &= [1 - (c\lambda + d)R](c\lambda + d) \end{aligned} \tag{5}$$

where $R = pP_1^{-1} + Q^{-1}q_1 - p_1(a\lambda + b)q_1$, which is integral in λ since P and Q are elementary. If $R \neq 0$, then, since $|\, c \,| \neq 0$, the degree of the right side

of (5) is at least 2, whereas the degree of the left side is only 1; hence $R = 0$ so that (5) gives $p(a\lambda + b)q = c\lambda + d$. Since $c\lambda + d$ is not singular, neither p nor q can be singular, and hence the theorem is proved.

When $|c| = 0$ (and therefore also $|a| = 0$) the remaining conditions of Theorem 3 are not sufficient to ensure that we can find constant matrices in place of P and Q, but these conditions are readily modified so as to apply to this case also. If we replace λ by λ/μ and then multiply by μ, $a\lambda + b$ is replaced by the homogeneous polynomial $a\lambda + b\mu$; and the definition of invariant factors applies immediately to such polynomials. In fact, if $|a| \neq 0$, the invariant factors of $a\lambda + b\mu$ are simply the homogeneous polynomials which are equivalent to the corresponding invariant factors of $a\lambda + b$. If, however, $|a| = 0$, then $|a\lambda + b\mu|$ is divisible by a power of μ which leads to factors of the form μ^i in the invariant factors of $a\lambda + b\mu$ which have no counterpart in those of $a\lambda + b$.

If $|c| = 0$ but $|c\lambda + d| \not\equiv 0$, there exist values, $\lambda_1 \neq 0$, μ_1, such that $|c\lambda_1 + d\mu_1| \neq 0$ and, if we make the transformation

$$\lambda = \lambda_1\alpha, \quad \mu = \mu_1\alpha + \beta, \tag{6}$$

$a\lambda + b\mu$, $c\lambda + d\mu$ become $a_1\alpha + b_1\beta$, $c_1\alpha + d_1\beta$ where $a_1 = a\lambda_1 + b\mu_1$, $c_1 = c\lambda_1 + d\mu_1$, and therefore $|c_1| \neq 0$. Further, when $a\lambda + b\mu$ and $c\lambda + d\mu$ have the same invariant factors, this is also true of $a_1\alpha + b_1\beta$ and $c_1\alpha + d_1\beta$. Since $|c_1| \neq 0$, the proof of Theorem 3 is applicable, so that there are constant non-singular matrices p, q for which $p(a_1\alpha + b_1\beta)q = c_1\alpha + d_1\beta$, and on reversing the substitution (6) we have

$$p(a\lambda + b\mu)q = c\lambda + d\mu.$$

Theorem 3 can therefore be extended as follows.

THEOREM 4. *If the non-singular polynomials $a\lambda + b\mu$, $c\lambda + d\mu$ have the same invariant factors, there exist non-singular constant matrices p, q such that $p(a\lambda + b\mu)q = c\lambda + d\mu$*

An important particular case of Theorem 3 arises when the polynomials have the form $\lambda - b$, $\lambda - d$. For if $p(\lambda - b)q = \lambda - d$, on equating coefficients we have $pq = 1$, $pbq = d$; hence $b = p^{-1}dp$, that is, b and d are similar. Conversely, if b and d are similar, then $\lambda - b$ and $\lambda - d$ are equivalent, and hence we have the following theorem.

THEOREM 5. *Two constant matrices b, d are similar if, and only if, $\lambda - b$ and $\lambda - d$ have the same invariant factors.*

3.05 **Elementary divisors.** If $D = |a\lambda + b|$ is not identically zero and if $\lambda_1, \lambda_2, \cdots, \lambda_s$ are its distinct roots, say

$$D = (\lambda - \lambda_1)^{\nu_1}(\lambda - \lambda_2)^{\nu_2} \cdots (\lambda - \lambda_s)^{\nu_s},$$

then the invariant factors of $a\lambda + b$, being factors of D, have the form

$$
\begin{aligned}
\alpha_1 &= (\lambda - \lambda_1)^{\nu_{11}}(\lambda - \lambda_2)^{\nu_{12}} \cdots (\lambda - \lambda_s)^{\nu_{1s}} \\
\alpha_2 &= (\lambda - \lambda_1)^{\nu_{12}}(\lambda - \lambda_2)^{\nu_{22}} \cdots (\lambda - \lambda_s)^{\nu_{2s}} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_i &= (\lambda - \lambda_1)^{\nu_{i1}}(\lambda - \lambda_2)^{\nu_{i2}} \cdots (\lambda - \lambda_s)^{\nu_{is}} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_n &= (\lambda - \lambda_1)^{\nu_{n1}}(\lambda - \lambda_2)^{\nu_{n2}} \cdots (\lambda - \lambda_s)^{\nu_{ns}}
\end{aligned} \tag{7}
$$

where $\sum_{j=1}^{n} \nu_{ji} = \nu_i$ and, since α_j is a factor of α_{j+1},

$$\nu_{1i} \leq \nu_{2i} \leq \cdots \leq \nu_{ni} \ (i = 1, 2, \cdots, s). \tag{8}$$

Such of the factors $(\lambda - \lambda_j)^{\nu_{ij}}$ as are not constants, that is, those for which $\nu_{ij} > 0$, are called the *elementary divisors* of $a\lambda + b$. The elementary divisors of $\lambda - b$ are also called the elementary divisors of b. When all the exponents ν_{ij} which are not 0 equal 1, b is said to have *simple* elementary divisors.

For some purposes the degrees of the elementary divisors are of more importance than the divisors themselves and, when this is the case, they are indicated by writing

$$[(\nu_{n1}, \nu_{n-1,1}, \cdots, \nu_{11}), (\nu_{n2}, \nu_{n-1,2}, \cdots, \nu_{12}), \cdots] \tag{9}$$

where exponents belonging to the same linear factor are in the same parenthesis, zero exponents being omitted; (9) is sometimes called the characteristic of $a\lambda + b$. If a root, say λ_1, is zero, it is convenient to indicate this by writing ν_{i1}^0 in place of ν_{i1}.

The maximum degree of $|\, a\lambda + b \,|$ is n and therefore $\sum_{i,j} \nu_{ij} \leq n$ where the equality sign holds only when $|\, a \,| \neq 0$.

The modifications necessary when the homogeneous polynomial $a\lambda + b\mu$ is taken in place of $a\lambda + b$ are obvious and are left to the reader.

3.06 Matrices with given elementary divisors. The direct investigation of the form of a matrix with given elementary divisors is somewhat tedious. It can be carried out in a variety of ways; but, since the form once found is easily verified, we shall here state this form and give the verification, merely saying in passing that it is suggested by the results of §2.07 together with a study of a matrix whose reduced characteristic function is $(\lambda - \lambda_1)^\nu$.

THEOREM 6. *If $\lambda_1, \lambda_2, \cdots, \lambda_s$ are any constants, not necessarily all different and $\nu_1, \nu_2, \cdots, \nu_s$ are positive integers whose sum is n, and if a_i is the array of ν_i rows and columns given by*

$$
\begin{matrix}
\lambda_i & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda_i & 1 & \cdots & 0 & 0 \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
0 & 0 & 0 & \cdots & \lambda_i & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda_i
\end{matrix} \tag{10}
$$

where each coordinate on the main diagonal equals λ_i, *those on the parallel on its right are* 1, *and the remaining ones are* 0, *and if* a *is the matrix of* n *rows and columns given by*

$$a = \begin{Vmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_s \end{Vmatrix} \tag{11}$$

composed of blocks of terms defined by (10) *arranged so that the main diagonal of each lies on the main diagonal of* a, *the other coordinates being* 0, *then* $\lambda - a$ *has the elementary divisors*

$$(\lambda - \lambda_1)^{\nu_1}, (\lambda - \lambda_2)^{\nu_2}, \cdots, (\lambda - \lambda_s)^{\nu_s} \tag{12}$$

In addition to using a_i to denote the block given in (10) we shall also use it for the matrix having this block in the position indicated in (11) and zeros elsewhere. In the same way, if f_i is a block with ν_i rows and columns with 1's in the main diagonal and zeros elsewhere, we may also use f_i for the corresponding matrix of order n. We can then write

$$\lambda - a = \Sigma(\lambda f_i - a_i), \quad f_i a = a_i = a f_i, \quad \Sigma f_i = 1.$$

The block of terms corresponding to $\lambda f_i - a_i$ has then the form

$$\begin{matrix} \lambda - \lambda_i & -1 & & & & \\ & \lambda - \lambda_i & -1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & -1 & \\ & & & & \lambda - \lambda_i & \end{matrix} \quad (\nu_i \text{ rows and columns}) \tag{13}$$

where only the non-zero terms are indicated. The determinant of these ν_i rows and columns is $(\lambda - \lambda_i)^{\nu_i}$ and this determinant has a first minor equal to ± 1; the invariant factors of $\lambda f_i - a_i$, regarded as a matrix of order ν_i, are therefore $1, 1, \cdots, 1, (\lambda - \lambda_i)^{\nu_i}$ and hence it can be reduced by elementary transformation to the diagonal form

$$\begin{matrix} (\lambda - \lambda_i)^{\nu_i} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1. \end{matrix}$$

If we apply the same elementary transformations to the corresponding rows and columns of $\lambda - a$, the effect is the same as regards the block of terms $\lambda f_i - a_i$ (corresponding to a_i in (11)) since all the other coordinates in the rows and columns which contain elements of this block are 0; moreover these trans-

formations do not affect the remaining blocks $\lambda f_j - a_j$ $(j \neq i)$ nor any 0 coordinate. Carrying out this process for $i = 1, 2, \cdots, s$ and permuting rows and columns, if necessary, we arrive at the form

$$\begin{array}{ccccccccc} (\lambda-\lambda_1)^{\nu} & & & & & & & & \\ & (\lambda-\lambda_2)^{\nu_2} & & & & & & & \\ & & \ddots & & & & & & \\ & & & & (\lambda-\lambda_s)^{\nu_s} & & & & \\ & & & & & 1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & & 1. \end{array}$$

Suppose now that the notation is so arranged that

$$\lambda_1 = \lambda_2 = \cdots = \lambda_p = \alpha, \quad \nu_1 \geq \nu_2 \geq \cdots \geq \nu_p,$$

but $\lambda_i \neq \alpha$ for $i > p$. The nth determinantal factor D_n then contains $(\lambda - \alpha)$ to the power $\sum_1^p \nu_i$ exactly. Each minor of order $n - 1$ contains at least $p - 1$ of the factors

$$(\lambda - \alpha)^{\nu_1}, (\lambda - \alpha)^{\nu_2}, \cdots, (\lambda - \alpha)^{\nu_p} \tag{14}$$

and in one the highest power $(\lambda - \alpha)^{\nu}$ is lacking; hence D_{n-1} contains $(\lambda - \alpha)$ to exactly the power $\sum_2^p \nu_i$ and hence the nth invariant factor α_n contains it to exactly the ν_1th power. Similarly the minors of order $n - 2$ each contain at least $p - 2$ of the factors (14) and one lacks the two factors of highest degree; hence $(\lambda - \alpha)$ is contained in D_{n-2} to exactly the power $\sum_3^p \nu_i$ and in α_{n-1} to the power ν_2. Continuing in this way we see that (14) gives the elementary divisors of a which are powers of $(\lambda - \alpha)$ and, treating the other roots in the same way, we see that the complete list of elementary divisors is given by (12) as required by the theorem.

3.07 If A is a matrix with the same elementary divisors as a, it follows from Theorem 5 that there is a matrix P such that $A = PaP^{-1}$ and hence, if we choose in place of the fundamental basis $(e_1, e_2, \cdots, e_n)$ the basis $(Pe_1, Pe_2, \cdots, Pe_n)$, it follows from Theorem 6 of chapter 1 that (11) gives the form of A relative to the new basis. This form is called the canonical form of A. It follows immediately from this that

$$P^{-1}A^kP = \left\| \begin{array}{ccccc} a_1^k & & & & \\ & a_2^k & & & \\ & & \ddots & & \\ & & & & a_s^k \end{array} \right\| \tag{15}$$

where a_i^k is the block of terms derived by forming the kth power of a_i regarded as a matrix of order ν_i.

Since D_n equals $|\lambda - a|$, it is the characteristic function of a (or A) and, since D_{n-1} is the highest common factor of the first minors, it follows from Theorem 3 of chapter 2 that α_n is the reduced characteristic function.

If we add the f's together in groups each group consisting of all the f's that correspond to the same value of λ_i, we get a set of idempotent matrices, say $\varphi_1, \varphi_2, \cdots, \varphi_r$, corresponding to the distinct roots of a, say $\alpha_1, \alpha_2, \cdots, \alpha_r$. These are the principal idempotent elements of a; for (i) $a\varphi_i = \varphi_i a$, (ii) $(a - \alpha_i)\varphi_i$ is nilpotent, (iii) $\Sigma\varphi_i = \Sigma f_i = 1$ and $\varphi_i\varphi_j = 0$ $(i \neq j)$ so that the conditions of §2.11 are satisfied.

When the same root α_i occurs in several elementary divisors, the corresponding f's are called *partial idempotent elements* of a; they are not unique as is seen immediately by taking $a = 1$.

If α is one of the roots of A, the form of $A - \alpha$ is sometimes important. Suppose that $\lambda_1 = \lambda_2 = \cdots = \lambda_p = \alpha$, $\lambda_i \neq \alpha$ $(i > p)$ and set

$$b_i = a_i - \alpha f_i,$$

the corresponding array in the ith block of $a - \alpha$ (cf. (10), (11)) being

$$\begin{matrix} \lambda_i - \alpha & 1 & & & \\ & \lambda_i - \alpha & 1 & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \lambda_i - \alpha. \end{matrix} \tag{16}$$

In the case of the first p blocks $\lambda_i - \alpha = 0$ and the corresponding $b_1, b_2, \cdots, b_p$ are nilpotent, the index of b_i being ν_i and, assuming $\nu_1 \geq \nu_2 > \cdots \geq \nu_p$ as before, $(A - \alpha)^k$ has the form

$$P^{-1}(A - \alpha)^k P = \left\| \begin{matrix} b_1^k & & & \\ & b_2^k & & \\ & & \cdot & \\ & & & \cdot \\ & & & & b_s^k \end{matrix} \right\|$$

or, when $k \geq \nu_1$

$$P^{-1}(A - \alpha)^k P = \left\| \begin{matrix} 0 & & & \\ & \cdot & & \\ & & 0 & \\ & & & b_{p+1}^k \\ & & & & \cdot \\ & & & & & \cdot \\ & & & & & & b_s^k \end{matrix} \right\| \tag{17}$$

Since none of the diagonal coordinates of $b_{p+1}, \cdots, b_s$ are 0, the rank of $(A - \alpha)^k$, when $k \geq \nu_1$, is exactly $n - \sum_1^p \nu_i = \sum_{p+1}^s \nu_i$ and the nullspace of $(A - \alpha)^k$ is then the same as that of $(A - \alpha)^{\nu_1}$. Hence, if there exists a vector z such that $(A - \alpha)^k z = 0$ but $(A - \alpha)^{k-1} z \neq 0$, then (i) $k \leq \nu_1$, (ii) z lies in the nullspace of $(A - \alpha)^{\nu_1}$.

3.08 **Invariant vectors.** If A is a matrix with the elementary divisors given in the statement of Theorem 6, then $\lambda - A$ is equivalent to $\lambda - a$ and by Theorem 5 there is a non-singular matrix P such that $A = PaP^{-1}$. If we denote the unit vectors corresponding to the rows and columns of a_i in (10) by $e_1^i, e_2^i, \cdots, e_{\nu_i}^i$ and set

$$x_j^i = \begin{cases} Pe_j^i (j = 1, 2, \cdots, \nu_i; i = 1, 2, \cdots, s) \\ 0 \ (j < 1 \text{ or } > \nu_i, \text{ or } i < 1 \text{ or } > s) \end{cases} \tag{18}$$

then

$$ae_1^i = \lambda_i e_1^i, \ ae_2^i = \lambda_i e_2^i + e_1^i, \ \cdots, \ ae_{\nu_i}^i = \lambda_i e_{\nu_i}^i + e_{\nu_i - 1}^i$$

and hence

$$Ax_j^i = \lambda_i x_j^i + x_{j-1}^i \ (j = 1, 2, \cdots, \nu_i; i = 1, 2, \cdots, s). \tag{19}$$

The vectors x_j^i are called a set of *invariant vectors*[6] of A.

The matrix A can be expressed in terms of its invariant vectors as follows. We have from (10)

$$a_i = \sum_j (\lambda_i e_j^i + e_{j-1}^i) S e_j^i = \sum_j e_j^i S(\lambda_i e_j^i + e_{j+1}^i)$$

and hence, if

$$y_j^i = (P')^{-1} e_j^i = (PP')^{-1} x_j^i, \tag{20}$$

then

$$A = \sum_{i,j} (\lambda_i x_j^i + x_{j-1}^i) S y_j^i = \sum_{i,j} x_j^i S(\lambda_i y_j^i + y_{j+1}^i) \tag{21}$$

where it should be noted that the y's form a system reciprocal to the x's and that each of these systems forms a basis of the vector space since $|P| \neq 0$.

If we form the transverse of A, we have from (21)

$$A' = \sum_{i,j} (\lambda_i y_j^i + y_{j+1}^i) S x_j^i \tag{22}$$

[6] If homogeneous coordinates are used so that vectors represent points, an invariant vector is usually called a pole.

so that the invariant vectors of A' are obtained by forming the system reciprocal to the x's and inverting the order in each group of vectors corresponding to a given elementary divisor; thus

$$A'y^i_{\nu_i} = \lambda_i y^i_{\nu_i},\ A'y^i_{\nu_i-1} = \lambda_i y^i_{\nu_i-1} + y^i_{\nu_i},\ \cdots,\ A'y^i_1 = \lambda_i y^i_1 + y^i_2.$$

A matrix A and its transverse clearly have the same elementary divisors and are therefore similar. The matrix which transforms A into A' can be given explicitly as follows. Let q_i be the symmetric array

$$\begin{array}{ccccc} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{array} \qquad (\nu_i \text{ rows and columns}).$$

It is easily seen that $q_i a_i = a_i' q_i$ and hence, if Q is the matrix

$$Q = \begin{Vmatrix} q_1 & & & & \\ & q_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & q_s \end{Vmatrix}$$

we have $Qa = a'Q$, and a short calculation gives $A' = R^{-1}AR$ where R is the symmetric matrix

$$R = PQ^{-1}P' = PQP'. \tag{23}$$

If the elementary divisors of A are simple, then $Q = 1$ and $R = PP'$.

If the roots λ_i of the elementary divisors (12) are all different, the nullity of $(A - \lambda_i)$ is 1, and hence x^i_1 is unique to a scalar multiplier. But the remaining x^i_j are not unique. In fact, if the x's denote one choice of the invariant vectors, we may take in place of x^i_j

$$z^i_j = k^i_1 x^i_j + k^i_2 x^i_{j-1} + \cdots + k^i_j x^i_1 \ (j = 1, 2, \cdots, \nu_i)$$

where the k's are any constant scalars subject to the condition $k^i_1 \neq 0$. Suppose now that $\lambda_1 = \lambda_2 = \cdots = \lambda_p = \alpha$, $\lambda_i \neq \alpha\ (i > p)$ and $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_p$ as in §3.07. We shall say that $z_1, z_2, \cdots, z_k$ is a *chain*[7] of invariant vectors *belonging to the exponent* k if

$$\begin{gathered} z_i = (A - \alpha)^{k-i} z_k \neq 0 \qquad (i = 1, 2, \cdots, k) \\ (A - \alpha)^k z_k = 0. \end{gathered} \tag{24}$$

It is also convenient to set $z_i = 0$ for $i < 0$ or $> k$. We have already seen that $k \leq \nu_1$ and that z_k lies in the nullspace of $(A - \alpha)^{\nu_1}$; and from (17) it is

[7] We shall say that the chain is *generated* by z_k.

seen that the nullspace of $(A - \alpha)^{\nu_1}$ has the basis $(x_j^i; j = 1, 2, \cdots, \nu_i, i = 1, 2, \cdots, p)$.

Since z_k belongs to the nullspace of $(A - \alpha)^{\nu_1}$, we may set

$$z_k = \sum_{i=1}^{p} \sum_{j=1}^{\nu_i} \zeta_{ij} x_j^i \tag{25}$$

and therefore by repeated application of (15) with $\lambda_i = \alpha$

$$(A - \alpha)^r z_k = \sum_{i,j} \zeta_{ij} x_{j-2}^i. \tag{26}$$

From this it follows that, in order that $(A - \alpha)^k z_k = 0$, only values of j which are less than or equal to k can actually occur in (25) and in order that $(A - \alpha)^{k-1} z_k \neq 0$ at least one ζ_{ik} must be different from 0; hence

$$\begin{aligned} z_k &= \sum_i (\zeta_{ik} x_k^i + \zeta_{i,k-1} x_{k-1}^i + \cdots) \\ z_{k-1} &= \sum_i (\zeta_{ik} x_{k-1}^i + \zeta_{i,k-1} x_{k-2}^i + \cdots) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ z_1 &= \sum_i \zeta_{ik} x_1^i. \end{aligned} \tag{27}$$

Finally, if we impose the restriction that z_k does not belong to any chain pertaining to an exponent greater than k, it is necessary and sufficient that k be one of the numbers $\nu_1, \nu_2, \cdots, \nu_p$ and that no value of i corresponding to an exponent greater than k occur in (27).

3.09 The actual determination of the vectors x_j^i can be carried out by the processes of §3.02 and §3.04 or alternatively as follows. Suppose that the first s_1 of the exponents ν_i equal n_1, the next s_2 equal n_2, and so on, and finally the last s_q equal n_q. Let $\mathfrak{N}_1$ be the nullspace of $(A - \alpha)^{n_1}$ and $\mathfrak{N}_1'$ the nullspace of $(A - \alpha)^{n_1 - 1}$; then $\mathfrak{N}_1$ contains $\mathfrak{N}_1'$. If $\mathfrak{M}_1$ is a space complementary to $\mathfrak{N}_1'$ in $\mathfrak{N}_1$, then for any vector x in $\mathfrak{M}_1$ we have $(A - \alpha)^r x = 0$ only when $r \geq n_1$. Also, if $x_1, x_2, \cdots, x_{m_1}$ is a basis of $\mathfrak{M}_1$, the vectors

$$(A - \alpha)^r x_i \ (r = 0, 1, \cdots, n_1 - 1) \tag{28}$$

are linearly independent; for, if

$$\sum_{r=s}^{n_1-1} \sum_i \xi_{ir} (A - \alpha)^r x_i = 0,$$

some ξ_{ir} being different from 0, then multiplying by $(A - \alpha)^{n_1 - s - 1}$ we have

$$(A - \alpha)^{n_1 - 1} \sum_i \xi_{is} x_i = 0,$$

which is only possible if every $\xi_{is} = 0$ since $x_1, x_2, \cdots, x_{m_1}$ form a basis of $\mathfrak{M}_1$ and for no other vector of $\mathfrak{M}_1$ is $(A - \alpha)^{n_1 - 1} x = 0$. The space defined by (28) clearly lies in $\mathfrak{N}_1$; we shall denote it by $\mathfrak{L}_1$. If we set $\mathfrak{N}_1 = \mathfrak{N}_2 + \mathfrak{L}_1$ where $\mathfrak{N}_2$ is complementary to $\mathfrak{L}_1$ in $\mathfrak{N}_1$, then $\mathfrak{N}_2$ contains all vectors which are members of sets belonging to the exponents $n_2, n_3, \cdots$ but not lying in sets with the exponent n_1.

We now set $\mathfrak{N}_2 = \mathfrak{N}_2' + \mathfrak{M}_2$ where $\mathfrak{N}_2'$ is the subspace of vectors x in $\mathfrak{N}_2$ such that $(A - \alpha)^{n_2 - 1} x = 0$. As before the elements of $\mathfrak{M}_2$ generate sets with exponent n_2 but are not members of sets with higher exponents; and by a repetition of this process we can determine step by step the sets of invariant vectors corresponding to each exponent n_i.