CHAPTER XX

ELEMENTARY DIVISORS AND THE EQUIVALENCE OF λ-MATRICES

91. λ-Matrices and their Elementary Transformations. The theory of elementary divisors, invented by Sylvester, H. J. S. Smith, and, more particularly, Weierstrass, and perfected in important respects by Kronecker, Frobenius, and others, has, in the form in which we will present it,* for its immediate purpose the study of matrices (which without loss of generality we assume to be square) whose elements are polynomials in a single variable λ. Such matrices we will call λ-matrices.† The determinant of a λ-matrix is a polynomial in λ, and if this determinant vanishes identically, we will call the matrix a singular λ-matrix. By the rank of a λ-matrix we understand the order of the largest determinant of the matrix which is not identically zero.

We have occasion here, as in § 19, to consider certain *elementary transformations* which we define as follows:

Definition 1. *By an elementary transformation of a λ-matrix we understand a transformation of any one of the following forms:*

(*a*) *The interchange of two rows or of two columns.*

(*b*) *The multiplication of each element of a row (or of a column) by the same constant not zero.*

(*c*) *The addition to the elements of a row (or column) of the products of the corresponding elements of another row (or column) by one and the same polynomial in λ.*

* Various modifications of the point of view here adopted are possible and important. First, we may consider matrices whose elements are polynomials in any number of variables. Secondly, we may confine ourselves to polynomials whose coefficients lie in a certain domain of rationality. Thirdly, we may approach the subject from the side of the theory of numbers, assuming that the coefficients of the polynomials are integers. The simplest case here would be that in which the elements of the matrix are themselves integers; see Exercise 2, § 91, Exercise 3, § 92, and Exercise 2, § 94.

† The matrix of a pencil of quadratic forms is an important example of a λ-matrix to which the general theory will be applied in Chapter XXII.

If we pass from a first matrix to a second by an elementary transformation, it is clear that we can pass back from the second to the first by an elementary transformation. Thus the following definition is justified:

Definition 2. *Two λ-matrices are said to be equivalent if it is possible to pass from one to the other by means of a finite number of elementary transformations.*

We see here that all λ-matrices equivalent to a given matrix are equivalent to each other; and, as in § 19, that two equivalent λ-matrices always have the same rank.

The rank of a λ-matrix is not, however, the only thing which is left unchanged by every elementary transformation. In order to show this we begin with

Lemma 1. *If the polynomial $\phi(\lambda)$ is a factor of all the i-rowed determinants of a λ-matrix* a, *it will be a factor of all the i-rowed determinants of every λ-matrix obtained from* a *by means of an elementary transformation.*

If the transformation is of the type (*a*) or (*b*) of Definition 1, this lemma is obviously true, since these transformations have no effect on the i-rowed determinants of a except to multiply them by constants which are not zero. If it is of the type (*c*), let us suppose it consists in adding to the elements of the pth column of a the corresponding elements of the qth column, each multiplied by the polynomial $\psi(\lambda)$. Any i-rowed determinant of a which either does not involve the pth column, or involves both the pth and the qth, will be unaffected by this transformation. An i-rowed determinant which involves the pth column but not the qth may be written after the transformation in the form $A \pm \psi(\lambda)B$, where A and B are i-rowed determinants of a; so that here also our lemma is true.

Theorem 1. *If* a *and* b *are equivalent λ-matrices of rank r, and $D_i(\lambda)$ is the greatest common divisor of the i-rowed determinants $(i \leqq r)$ of* a, *then it is also the greatest common divisor of the i-rowed determinants of* b.

For by our lemma, $D_i(\lambda)$ is a factor of all the i-rowed determinants of b; and if these determinants had a common factor of higher degree, this factor would, by our lemma, be a factor of all the i-rowed determinants of a; which is contrary to hypothesis.

The theorem just proved shows that the greatest common divisors $D_1(\lambda), \cdots D_r(\lambda)$ are invariants with regard to elementary transformations, or, more generally, that they are invariants with regard to all transformations which can be built up from a finite number of elementary transformations. In point of fact they form, along with the rank r, a complete system of invariants. To prove this we now proceed to show how, by means of elementary transformations, a λ-matrix may be reduced to a very simple normal form.

LEMMA 2. *If the first element* $f(\lambda)$ of a λ-matrix is not identically zero and is not a factor of all the other elements, then an equivalent matrix can be formed whose first element is not identically zero and is of lower degree than f.*

Suppose first there is an element $f_1(\lambda)$ in the first row which is not divisible by $f(\lambda)$ and let j denote the number of the column in which it lies. Dividing f_1 by f and calling the quotient q and the remainder r, we have
$$f_1(\lambda) \equiv q(\lambda)f(\lambda) + r(\lambda).$$
Accordingly, if to the elements of the jth column we add those of the first, each multiplied by $-q(\lambda)$, we get an equivalent matrix in which the first element of the jth column is $r(\lambda)$, which is a polynomial of degree lower than $f(\lambda)$. If now we interchange the first and jth columns, the truth of our lemma is established in the case we are considering.

A similar proof obviously applies if there is an element in the first column which is not divisible by $f(\lambda)$.

Finally, suppose every element of the first row and column is divisible by $f(\lambda)$, but that there is an element, say in the ith row and jth column, which is not divisible by $f(\lambda)$. Let us suppose the element in the first row and jth column is $\psi(\lambda)f(\lambda)$, and form an equivalent matrix by adding to the elements of the jth column $-\psi(\lambda)$ times the corresponding elements of the first column. In this matrix, $f(\lambda)$ still stands in the upper left-hand corner, the first element of the jth column is zero; the first element of the ith row has not been changed and is therefore divisible by $f(\lambda)$; and the element in the ith row and jth column is still not divisible by f. Now form another equivalent matrix by adding to the elements of the first column the corresponding elements of the jth column. The upper left-hand element is still $f(\lambda)$, while the first element of the

* By the first element of a matrix we will understand the element in the upper left-hand corner

ith row is not divisible by $f(\lambda)$. This matrix, therefore, comes under the case already treated in which there is an element in the first column which is not divisible by $f(\lambda)$, and our lemma is established.

LEMMA 3. *If we have a λ-matrix whose elements are not all identically zero, an equivalent matrix can be formed which has the following three properties:*

(*a*) *The first element $f(\lambda)$ is not identically zero.*

(*b*) *All the other elements of the first row and of the first column are identically zero.*

(*c*) *Every element neither in the first row nor in the first column is divisible by $f(\lambda)$.*

For we may first, by an interchange of rows and of columns, bring into the first place an element which is not identically zero. If this is not a factor of all the other elements, we can, by Lemma 2, find an equivalent matrix whose first element is of lower degree and is not identically zero. If this element is not a factor of all the others, we may repeat the process. Since at each step we lower the degree of the first element, there must, after a finite number of steps, come a point where the process stops, that is, where the first element is a factor of all the others. We can then, by using transformations of type (c) (Definition 1), reduce all the elements in the first row and in the first column except this first one to zero, while the other elements remain divisible by the first one. Thus our lemma is established.

Finally, we note that since $f(\lambda)$ in the lemma just proved is a factor of all the other elements of the simplified matrix, it must, by Theorem 1, be the greatest common divisor of all the elements of the original matrix.

The lemma just proved tells us that the λ-matrix of the nth order of rank $r > 0$

$$(1) \qquad \left\| \begin{matrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ a_{n1} & \cdots & a_{nn} \end{matrix} \right\|$$

can be reduced by means of elementary transformations to the form

$$(2) \qquad \left\| \begin{matrix} f_1(\lambda) & 0 & \cdots & 0 \\ 0 & b_{11} & \cdots & b_{1,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & b_{n-1,1} & \cdots & b_{n-1,n-1} \end{matrix} \right\|,$$

where $f_1(\lambda) \not\equiv 0$ and where $f_1(\lambda)$ is a factor of all the b's. The last written matrix being necessarily of rank r, the matrix of the $(n-1)$th order

$$
(3) \qquad \left\| \begin{matrix} b_{11} & \cdots & b_{1,\,n-1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ b_{n-1,\,1} & \cdots & b_{n-1,\,n-1} \end{matrix} \right\|
$$

is of rank $r-1$. Consequently, if $r>1$, (3) may be reduced by means of elementary transformations to the form

$$
(4) \qquad \left\| \begin{matrix} f_2(\lambda) & 0 & \cdots & 0 \\ 0 & c_{11} & \cdots & c_{1,\,n-2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & c_{n-2,\,1} & \cdots & c_{n-2,\,n-2} \end{matrix} \right\|,
$$

where $f_2(\lambda) \not\equiv 0$ and where $f_2(\lambda)$ is a factor of all the c's. By Theorem 1, $f_2(\lambda)$, being the greatest common divisor of all the elements of (4), is also the greatest common divisor of all the b's, and is therefore divisible by $f_1(\lambda)$.

Now it is important to notice that the elementary transformations which carry over (3) into (4) may be regarded as elementary transformations of (2) which leave the first row and column of this matrix unchanged. Thus by a succession of elementary transformations, we have reduced (1) to the form

$$
(5) \qquad \left\| \begin{matrix} f_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & f_2(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & c_{11} & \cdots & c_{1,\,n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & c_{n-2,\,1} & \cdots & c_{n-2,\,n-2} \end{matrix} \right\|,
$$

where neither f_1 nor f_2 vanishes identically, f_1 is a factor of f_2, and f_2 is a factor of all the c's.

If $r > 2$, we may treat the $(n-2)$-rowed matrix of the c's, which is clearly of rank $r-2$, in a similar manner. Proceeding in this way, we finally reduce our matrix (1) to the form

$$
(6) \qquad \left\| \begin{matrix} f_1(\lambda) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & f_2(\lambda) & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & f_r(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{matrix} \right\|,
$$

where none of the f's is identically zero, and each is a factor of the next following one.

So far we have used merely elementary transformations of the forms (a) and (c), Definition 1. By means of transformations of the form (b) we can simplify (6) still further by reducing the coefficient of the highest power of λ in each of the polynomials $f_i(\lambda)$ to unity. We have thus proved the theorem:

Theorem 2. *Every λ-matrix of the nth order and of rank r can be reduced by elementary transformations to the normal form*

$$
(7) \qquad \left\| \begin{matrix} E_1(\lambda) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & E_2(\lambda) & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & E_r(\lambda) & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{matrix} \right\|,
$$

where the coefficient of the highest power of λ in each of the polynomials $E_i(\lambda)$ is unity, and $E_i(\lambda)$ is a factor of $E_{i+1}(\lambda)$ for $i = 1, 2, \cdots r-1$.

By Theorem 1, the greatest common divisor of the i-rowed determinants ($i \leqq r$) of the original matrix is the same as the greatest common divisor of the i-rowed determinants of the normal form (7) to which it is reduced. These last mentioned i-rowed determinants are, however, all identically zero except those which are the product of i of the E's. Let

$$
(8) \qquad E_{k_1}(\lambda)E_{k_2}(\lambda) \cdots E_{k_i}(\lambda)
$$

be any one of these, and suppose the integers $k_1, k_2, \cdots k_i$ to have been arranged in order of increasing magnitude. We obviously have $k_1 \geqq 1, k_2 \geqq 2, \cdots k_i \geqq i$. Consequently E_1 is a factor of E_{k_1}, E_2 of E_{k_2}, etc. Thus

$$
E_1(\lambda)E_2(\lambda) \cdots E_i(\lambda)
$$

is seen to be a factor of (8), and, being itself one of the i-rowed determinants of (7), it is their greatest common divisor. That is,

Theorem 3. *The greatest common divisor of the i-rowed determinants of a λ-matrix of rank r, when $i \leqq r$, is*

$$
D_i(\lambda) \equiv E_1(\lambda)E_2(\lambda) \cdots E_i(\lambda),
$$

where the E's are the elements of the normal form (7) to which the given matrix is equivalent.

It may be noticed that this greatest common divisor is so determined that the coefficient of the highest power of λ in it is unity.

We come now to the fundamental theorem:

THEOREM 4. *A necessary and sufficient condition for the equivalence of two λ-matrices of the nth order is that they have the same rank r, and that for every value of i from 1 to r inclusive, the i-rowed determinants of one matrix have the same greatest common divisor as the i-rowed determinants of the other.*

To say that this is a necessary condition is merely to restate Theorem 1. To prove it sufficient, suppose both matrices to be reduced to the normal form (7), where we will distinguish the normal form for the second matrix by attaching accents to the E's in it. If the conditions of our theorem are fulfilled, we have, by Theorem 3,

$$E_1'(\lambda) \equiv E_1(\lambda),$$
$$E_1'(\lambda)E_2'(\lambda) \equiv E_1(\lambda)E_2(\lambda),$$
$$E_1'(\lambda)E_2'(\lambda)E_3'(\lambda) \equiv E_1(\lambda)E_2(\lambda)E_3(\lambda),$$
$$\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots$$

and, since none of these E's are identically zero, it follows that

$$E_i'(\lambda) \equiv E_i(\lambda) \qquad (i = 1, 2, \cdots r).$$

Thus the normal forms to which the two λ-matrices can be reduced are identical, and hence the matrices are equivalent, since two λ-matrices equivalent to a third are equivalent to each other.

EXERCISES

1. Reduce the matrix

$$\begin{Vmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda-1 & 0 \\ 0 & 0 & 0 & 0 & \lambda-1 \end{Vmatrix}$$

by means of elementary transformations to the normal form of Theorem 2.

Verify the result by finding the greatest common divisors $D_i(\lambda)$ first directly, and secondly from the normal form.

2. By an elementary transformation of a matrix all of whose elements are integers is understood a transformation of any one of the following forms:

(*a*) The interchange of two rows or of two columns.

(*b*) The change of sign of all the elements of any row or column.

(*c*) The addition to the elements of one row (or column) of the products of the corresponding elements of another row (or column) by one and the same integer.

Starting from this definition, develop the theory of matrices whose elements are integers along the same lines as the theory of λ-matrices was developed in this section.

92. Invariant Factors and Elementary Divisors. In place of the invariants $D_i(\lambda)$ of the last section, it is, for most purposes, more convenient to introduce certain other invariants to which we will give the technical name *invariant factors.* As a basis for the definition of these invariants we state the following theorem, which is merely an immediate consequence of Theorem 3, § 91:

THEOREM 1. *The greatest common divisor of the i-rowed determinants (i = 2, 3, ⋯ r) of a λ-matrix of rank r is divisible by the greatest common divisor of the (i − 1)-rowed determinants of this matrix.*

DEFINITION 1. *If* a *is a λ-matrix of rank r, and*

$$D_i(\lambda) \qquad (i = 1, 2, \cdots r)$$

the greatest common divisor of its i-rowed determinants so determined that the coefficient of the highest power of λ is unity; and if $D_0(\lambda) \equiv 1$; then the polynomial

$$(1) \qquad E_i(\lambda) \equiv \frac{D_i(\lambda)}{D_{i-1}(\lambda)} \qquad (i = 1, 2, \cdots r)$$

is called the ith invariant factor of a.

This definition shows that these E's are really invariants since they are completely determined by the D's which we proved to be invariants in § 91. Moreover, by multiplying together the first i of the relations (1), we get the formula

$$(2) \qquad D_i(\lambda) \equiv E_1(\lambda)E_2(\lambda) \cdots E_i(\lambda) \qquad (i = 1, 2, \cdots r).$$

This shows us that the E's completely determine the D's, and since these latter were seen in § 91 to form, together with the rank, a complete system of invariants, the same is true of the E's. That is,

THEOREM 2. *A necessary and sufficient condition that two λ-matrices be equivalent is that they have the same rank r, and that the invariant factors of one be identical respectively with the corresponding invariant factors of the other.*

Since, in the case of a non-singular matrix of the nth order, $D_n(\lambda)$ differs from the determinant of the matrix only by a constant factor, we see that in this case the determinant of the matrix is, except for a constant factor, precisely the product of all the invariant factors. This is the case which is of by far the greatest importance, and the term *invariant factor* comes from the fact that the E's are really factors of the determinant of the matrix in this case.

A reference to Theorem 3, § 91, shows that our invariant factors are precisely the polynomials E_i which occur in the normal form of Theorem 2, § 91; and, since in that normal form each E is a factor of the next following one, we have the important result,

THEOREM 3. *If $E_1(\lambda), \cdots E_r(\lambda)$ are the successive invariant factors of a λ-matrix of rank r, then each of these E's is a factor of the next following one.*

This theorem enables us to arrange the invariant factors of a λ-matrix in the proper order by simply arranging them in the order of increasing degree, two E's of the same degree being necessarily identical.

The invariant factors (like the D's of the last section) may be spoken of as *rational* invariants of our λ-matrix since they are formed from the elements of the λ-matrix by purely rational processes, namely the elementary transformations of § 91, which involve only the rational operations of addition, subtraction, multiplication, and divison. In distinction to these the *elementary divisors*, first introduced by Weierstrass, are, in general, irrational invariants.* These we now proceed to define.

DEFINITION 2. *If* $\mathbf{a}$ *is a λ-matrix of rank r, and $D_r(\lambda)$ is the greatest common divisor of the r-rowed determinants of* $\mathbf{a}$, *then the linear factors*

$$\lambda-\alpha,\ \lambda-\alpha',\ \lambda-\alpha'',\ \cdots$$

of $D_r(\lambda)$ are called the linear factors of $\mathbf{a}$.†

Since, by formula (2), $D_r(\lambda)$ is the product of all the invariant factors of $\mathbf{a}$, it is clear that each invariant factor is merely the product of certain integral powers, positive or zero, of the linear factors of $\mathbf{a}$. We may therefore lay down the following definition:

* German writers, following Frobenius, use the term *elementary divisor* to cover both kinds of invariants. This is somewhat confusing, and necessitates the use of modifying adjectives such as *simple elementary divisors* for the elementary divisors as originally defined by Weierstrass, *composite elementary divisors* for the E's. On the other hand Bromwich (*Quadratic Forms and their Classification by Means of Invariant-factors*, Cambridge, England, 1906) proposes to substitute the term *invariant factor* for the term *elementary divisor*. Inasmuch as this latter term is wholly appropriate, it seems clear that it should be retained in English as well as in German in the sense in which Weierstrass first used it.

† It will be noticed that if $\mathbf{a}$ is non-singular, the linear factors of $\mathbf{a}$ are simply the linear factors of the determinant of $\mathbf{a}$.

DEFINITION 3. *Let* $\mathbf{a}$ *be a λ-matrix of rank r, and*

$$\lambda-\alpha,\ \lambda-\alpha',\ \lambda-\alpha'',\ \cdots$$

its distinct linear factors. Then if

$$E_i(\lambda)\equiv(\lambda-\alpha)^{e_i}(\lambda-\alpha')^{e_i'}(\lambda-\alpha'')^{e_i''}\cdots \qquad (i=1,2,\cdots r),$$

are the invariant factors of $\mathbf{a}$, *such of the factors*

$$\begin{array}{llll}
(\lambda-\alpha)^{e_1}, & (\lambda-\alpha)^{e_2}, & \cdots\cdots & (\lambda-\alpha)^{e_r}, \\
(\lambda-\alpha')^{e_1'}, & (\lambda-\alpha')^{e_2'}, & \cdots\cdots & (\lambda-\alpha')^{e_r'}, \\
(\lambda-\alpha'')^{e_1''}, & (\lambda-\alpha'')^{e_2''}, & \cdots\cdots & (\lambda-\alpha'')^{e_r''}, \\
\cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\
\cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot
\end{array}$$

as are not mere constants are called the elementary divisors of $\mathbf{a}$, *each elementary divisor being said to correspond to the linear factor of which it is a power.**

Since the invariant factors completely determine the elementary divisors and *vice versa*, it is clear that the elementary divisors are not merely invariants, but that, together with the rank, they form a complete system of invariants. That is,

THEOREM 4. *A necessary and sufficient condition that two λ-matrices be equivalent is that they have the same rank and that the elementary divisors of one be identical respectively with the corresponding elementary divisors of the other.*

By means of Theorem 3 we infer the important result:

THEOREM 5. *The degrees e_i of the elementary divisors corresponding to any particular linear factor satisfy the inequalities*

$$e_i \geqq e_{i-1} \qquad (i=2,3,\cdots r).$$

By means of this theorem we can arrange the elementary divisors corresponding to any given linear factor in the proper order by simply noticing their degrees.

* It will be seen that the definition just given is equivalent to the following one, in which the conception of invariant factors is not introduced:

DEFINITION. *Let $\lambda-\alpha$ be a linear factor of the λ-matrix* $\mathbf{a}$ *of rank r, and let l_i be the exponent of the highest power of $\lambda-\alpha$ which is a factor of all the i-rowed determinants ($i\leqq r$) of* $\mathbf{a}$. *If the integers e_i (which are necessarily positive or zero) are defined by the formula*

$$e_i=l_i-l_{i-1} \qquad (i=1,2,\cdots r),$$

then such of the expressions

$$(\lambda-\alpha)^{e_1},\ (\lambda-\alpha)^{e_2},\ \cdots\ (\lambda-\alpha)^{e_r}$$

as are not constants are called the elementary divisors of $\mathbf{a}$ *which correspond to the linear factor $\lambda-\alpha$.*

EXERCISES

1. If $\phi = 0$ and $\psi = 0$ are two conics of which the second is non-singular, show how the number and kind of singular conics contained in the pencil $\phi - \lambda\psi = 0$ depends on the nature of the elementary divisors of the matrix of the quadratic form $\phi - \lambda\psi$.

2. Extend Exercise 1 to the case of three dimensions.

3. Apply the considerations of this section to matrices whose elements are integers. (Cf. Exercise 2, §91).

93. The Practical Determination of Invariant Factors and Elementary Divisors. The easiest general method for determining the invariant factors of a particular λ-matrix is to reduce it by means of elementary transformations to the normal form of Theorem 2, §91, following out step by step the reduction used in the proof of that theorem. From this normal form the invariant factors may be read off; and from these the elementary divisors may be computed, although only, in general, by the solution of equations of more or less high degree.

There are, however, many cases of great importance in which the elementary divisors may more easily be obtained by other methods. The most obvious of these is to apply the definition of elementary divisors directly to the case in hand. As an illustration, we mention a matrix of the nth order which has $\alpha - \lambda$ as the element in each place of the principal diagonal, while all the other elements are zero except those which lie immediately to the right of or above the elements of the principal diagonal, these being all constants different from zero:

$$(1) \qquad \left\| \begin{matrix} \alpha-\lambda & c_1 & 0 & \cdots & 0 & 0 \\ 0 & \alpha-\lambda & c_2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \alpha-\lambda & c_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & \alpha-\lambda \end{matrix} \right\| \qquad (c_1c_2\cdots c_{n-1} \neq 0).$$

The determinant of this matrix is $(\alpha-\lambda)^n$. The determinant obtained by striking out the first column and the last row is $c_1c_2\cdots c_{n-1}$. Accordingly

$$D_n(\lambda) \equiv (\lambda-\alpha)^n, \; D_{n-1}(\lambda) \equiv 1, \; E_n(\lambda) \equiv (\lambda-\alpha)^n.$$

Thus we see that $(\lambda-\alpha)^n$ is the only elementary divisor of this matrix, while the invariant factors are $(\lambda-\alpha)^n$ and $n-1$ 1's.

This direct method may sometimes be employed to advantage in conjunction with the method of reduction by elementary transformations. Cf. Exercise 1 at the end of this section.

A further means of recognizing the elementary divisors in some special cases is furnished by the following theorems whose proofs, which present no difficulty, we leave to the reader:

THEOREM 1. *If all the elements of a λ-matrix are zeros except those in the principal diagonal, and if each element of this diagonal which is not a constant is resolved into the product of a constant by powers of distinct linear factors of the form $\lambda-\alpha$, $\lambda-\alpha'$, $\cdots$, then these powers of linear factors will be precisely the elementary divisors of the matrix.*

THEOREM 2. *If all the elements of a λ-matrix are zeros except those which lie in a certain number of non-overlapping principal minors, then the elementary divisors of the matrix may be found by taking the elementary divisors of all these principal minors.*

The proof of this theorem consists in reducing the given matrix to the form referred to in Theorem 1 by means of elementary transformations each of which may be regarded as an elementary transformation of one of the principal minors in question.

It should be noticed that this theorem would not be true if the words *invariant factors* were substituted in it for *elementary divisors;* cf. Exercise 3 below. The invariant factors may, however, be computed from the elementary divisors when these have been found.

EXERCISES

1. Prove that the matrix

$$\left\| \begin{array}{ccc:ccc} \lambda-\alpha & 0 & 0 & -1 & 0 & 0 \\ 0 & \lambda-\alpha & 0 & 0 & -1 & 0 \\ 0 & 0 & \lambda-\alpha & 0 & 0 & -1 \\ \hdashline \beta^2 & 1 & 0 & \lambda-\alpha & 0 & 0 \\ 0 & \beta^2 & 1 & 0 & \lambda-\alpha & 0 \\ 0 & 0 & \beta^2 & 0 & 0 & \lambda-\alpha \end{array} \right\|$$

is equivalent to

$$\left\| \begin{array}{ccc:ccc} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & (\lambda-\alpha)^2+\beta^2 & 1 & 0 \\ 0 & 0 & 0 & 0 & (\lambda-\alpha)^2+\beta^2 & 1 \\ 0 & 0 & 0 & 0 & 0 & (\lambda-\alpha)^2+\beta^2 \end{array} \right\|,$$

and hence that its elementary divisors are

$$[\lambda-(\alpha+\beta i)]^3,\ [\lambda-(\alpha-\beta i)]^3.$$

2. Generalize Exercise 1 to matrices of order $2n$.

3. Find (a) the elementary divisors, and (b) the invariant factors of the matrix

$$\left\|\begin{matrix} \lambda^2(\lambda-1)^2 & 0 & 0 & 0 \\ 0 & \lambda(\lambda-1)^3 & 0 & 0 \\ 0 & 0 & \lambda-1 & 0 \\ 0 & 0 & 0 & \lambda \end{matrix}\right\|.$$

4. Determine the invariant factors and the elementary divisors of the matrix

$$\left\|\begin{matrix} 2\lambda & 3 & 0 & 1 & \lambda \\ 4\lambda & 3(\lambda+2) & 0 & \lambda+2 & 2\lambda \\ 0 & 6\lambda & \lambda & 2\lambda & 0 \\ \lambda-1 & 0 & \lambda-1 & 0 & 0 \\ 3(\lambda-1) & 1-\lambda & 2(\lambda-1) & 0 & 0 \end{matrix}\right\|.$$

Is this matrix equivalent to the matrix in the exercise at the end of § 91?

5. Devise a convenient rational process for computing the invariant factors of matrices of the kinds considered in Theorems 1 and 2.

94. A Second Definition of the Equivalence of λ-Matrices. The definition of equivalence of λ-matrices which we have used so far rests on the elementary transformations. These transformations are of such a special character that this definition is not convenient for most purposes. We now give a new definition which we will prove to be coextensive with the old one.

Definition. *Two n-rowed λ-matrices* **a** *and* **b** *are said to be equivalent if there exist two non-singular n-rowed λ-matrices* **c** *and* **d** *whose determinants are independent of* λ, *and such that*

$$(1)\qquad \mathbf{b} \equiv \mathbf{cad}.^*$$

Since the matrices **c** and **d** have, by hypothesis, constant determinants, the inverse matrices $\mathbf{c}^{-1}$ and $\mathbf{d}^{-1}$ will also be λ-matrices, and not matrices whose coefficients are fractional rational functions of λ as would in general be the case for the inverse of λ-matrices. Consequently, if we write (1) in the form

$$(2)\qquad \mathbf{a} \equiv \mathbf{c}^{-1}\mathbf{b}\mathbf{d}^{-1},$$

we see that the relation established by our definition between the matrices **a** and **b** is a reciprocal one, as is implied in the wording of the definition.

* We use here and in what follows the sign ≡ between two λ-matrices to denote that every element of one matrix is identically equal to the corresponding element of the other.

In order to justify the definition just given, we begin by establishing the

Lemma. *If* **a** *and* **b** *are n-rowed λ-matrices, and the polynomial φ(λ) is a factor of all the i-rowed determinants of* **a**, *it is a factor of all the i-rowed determinants of* **ab** *and also of* **ba**.

For, by Theorem 5, § 25, every i-rowed determinant of **ab** and also of **ba** is a homogeneous linear combination of certain i-rowed determinants of **a**.

Theorem 1. *If* **a** *and* **b** *are equivalent according to the definition of this section, they are also equivalent according to the definition of* § 91.

For in this case there exist two non-singular λ-matrices, **c** and **d**, whose determinants are constants, such that relation (1) holds. Consequently, by Theorem 7, § 25,* **a** and **b** have the same rank r. Let $D_i(\lambda)$ be the greatest common divisor of the i-rowed determinants of **a**, where $i \leqq r$. By our lemma, $D_i(\lambda)$ is a factor of all the i-rowed determinants of **ca**, and therefore, applying the lemma again, it is a factor of all the i-rowed determinants of **cad**, that is, of **b**.

We can infer further that $D_i(\lambda)$ is the *greatest* common divisor of the i-rowed determinants of **b**. For applying to relation (2) the reasoning just used, we see that the greatest common divisor of the i-rowed determinants of **b** is a factor of all the i-rowed determinants of **a**, and cannot therefore be of higher degree than $D_i(\lambda)$.

A reference to Theorem 4, § 91, now shows us that **a** and **b** are equivalent according to the definition of that section.

Theorem 2. *If* **a** *and* **b** *are equivalent according to the definition of* § 91, *they are also equivalent according to the definition of the present section.*

We begin by showing that if we can pass from a matrix **a** to a matrix $\mathbf{a}_1$ by means of an elementary transformation, one of the following relations always holds:

$$(3)\qquad \mathbf{a}_1 \equiv \mathbf{ca} \quad\text{or}\quad \mathbf{a}_1 \equiv \mathbf{ad}$$

where **c** and **d** are non-singular matrices whose determinants are independent of λ. To prove this we consider in succession the elementary transformations of the forms which were called (a), (b), (c), in Definition 1, § 91.

* How is it that we have a right to apply this theorem to λ-matrices?

(*a*) Suppose we interchange the *p*th and *q*th rows. This can be effected by forming the product **ca** where the matrix **c** may be obtained by interchanging the *p*th and *q*th rows (or columns) in the unit matrix

$$\begin{Vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{Vmatrix}.$$

Similarly the interchange of the *p*th and *q*th columns of **a** may be effected by forming the product **ac**, where **c** has the same meaning as before.

In each of these cases, **c** may be regarded as a non-singular λ-matrix with constant determinant, since its elements are constants and its determinant is -1.

(*b*) To multiply the *p*th row of **a** by a constant k, we may form the product **ca**, where **c** differs from the unit matrix only in having k instead of 1 as the *p*th element of the principal diagonal.

Similarly, we multiply the *p*th column of **a** by k, by forming the product **ac**, where **c** has the same meaning as before.

If we take the constant k different from zero, **c** may be regarded as a non-singular λ-matrix with constant determinant.

(*c*) We can add to the *p*th row of **a** $\phi(\lambda)$ times the *q*th row by forming the product **ca**, where **c** differs from the unit matrix only in having $\phi(\lambda)$ instead of zero as the element in the *p*th row and *q*th column.

Similarly we add to the *q*th column $\phi(\lambda)$ times the *p*th column by forming the product **ac** where **c** has the same meaning as before.

The matrix **c**, whose determinant is 1, is a non-singular λ-matrix.

It being thus established that one of the relations (3) holds between any two λ-matrices which can be obtained from one another by an elementary transformation, it follows that two matrices **a** and **b** which are equivalent according to the definition of § 91 will satisfy a relation of the form

$$\mathbf{b} \equiv \mathbf{c}_p \mathbf{c}_{p-1} \cdots \mathbf{c}_1 \mathbf{a} \mathbf{d}_1 \mathbf{d}_2 \cdots \mathbf{d}_q$$

where each of the **c**'s and **d**'s is a non-singular λ-matrix of constant determinant which corresponds to one of the elementary transformations we use in passing from **a** to **b**. This last relation being of the form

$$\mathbf{b} \equiv \mathbf{cad},$$

where **c** and **d** are non-singular λ-matrices with constant determinants, our theorem is proved.

We have now completed the proof that our two definitions of the equivalence of λ-matrices are coextensive.

EXERCISES

1. If **a** denotes the matrix in Exercise 1, § 91, and **b** the normal form of Theorem 2, § 91, for this matrix, determine two λ-matrices, **c** and **d**, such that relation (1) holds.

Verify your result by showing that the determinants of **c** and **d** are constants

2. Apply the considerations of this section to matrices whose elements are integers. Cf. Exercise 2, § 91, and Exercise 3, § 92.

95. Multiplication and Division of λ-Matrices. We close this chapter by giving a few developments of what might be called the elementary algebra of λ-matrices.

Definition. *By the degree of a λ-matrix is understood the highest degree in λ of any one of its elements.*

For a λ-matrix of the kth degree, the element in the ith row and jth column may be written $a_{ij}\lambda^k + a'_{ij}\lambda^{k-1} + \cdots + a_{ij}^{[k]}$,

and at least one of the coefficients of λ^k (*i.e.* one of the a_{ij}'s) must be different from zero. If, then, we denote by $\mathbf{a}_p$ the matrix of which $a_{ij}^{[p]}$ is the element which stands in the ith row and jth column, we get the theorem

Theorem 1. *Every λ-matrix of the kth degree may be written in the form*

$$\mathbf{a}_0\lambda^k + \mathbf{a}_1\lambda^{k-1} + \cdots + \mathbf{a}_k \qquad (\mathbf{a}_0 \neq 0) \tag{1}$$

where $\mathbf{a}_0, \cdots \mathbf{a}_k$ *are matrices with constant elements; and conversely, every expression* (1) *is a λ-matrix of degree k.*

Theorem 2. *The product of two λ-matrices of degrees k and l*

$$\mathbf{a}_0\lambda^k + \mathbf{a}_1\lambda^{k-1} + \cdots + \mathbf{a}_k \qquad (\mathbf{a}_0 \neq 0)$$
$$\mathbf{b}_0\lambda^l + \mathbf{b}_1\lambda^{l-1} + \cdots + \mathbf{b}_l \qquad (\mathbf{b}_0 \neq 0)$$

is a λ-matrix of degree $k + l$ provided at least one of the matrices $\mathbf{a}_0$ *and* $\mathbf{b}_0$ *is non-singular.*

For this product is a λ-matrix of the form

$$c_0\lambda^{k+l} + c_1\lambda^{k+l-1} + \cdots + c_{k+l}$$

where c_0 has the value a_0b_0 or b_0a_0 according to the order in which the two given matrices are multiplied together. By Theorem 7, § 25, neither a_0b_0 nor b_0a_0 is zero if a_0 and b_0 are not both singular.

The next theorem relates to what we may call the division of λ-matrices.

THEOREM 3. *If* a *and* b *are two λ-matrices and if* b, *when written in the form* (1), *has as the coefficient of the highest power of* λ *a non-singular matrix, then there exists one, and only one, pair of λ-matrices* q_1 *and* r_1 *for which*

$$a \equiv q_1 b + r_1$$

and such that either $r_1 \equiv 0$, *or* r_1 *is a λ-matrix of lower degree than* b; *and also one and only one pair of λ-matrices* q_2 *and* r_2 *for which*

$$a \equiv bq_2 + r_2$$

and such that either $r_2 \equiv 0$, *or* r_2 *is a λ-matrix of lower degree than* b.

The proof of this theorem is practically identical with the proof of Theorem 1, § 63.

EXERCISE

DEFINITION. *By a real matrix is understood a matrix whose elements are real; by a real λ-matrix, a matrix whose elements are real polynomials in* λ; *and by a real elementary transformation, an elementary transformation in which the constant in* (*b*) *and the polynomial in* (*c*), *Definition* 1, § 91, *are real.*

Show that all the results of this chapter still hold if we interpret the words *matrix*, *λ-matrix*, and *elementary transformation* to mean *real matrix*, *real λ-matrix*, and *real elementary transformation*, respectively.

CHAPTER XXI

THE EQUIVALENCE AND CLASSIFICATION OF PAIRS OF BILINEAR FORMS AND OF COLLINEATIONS

96. The Equivalence of Pairs of Matrices. The applications of the theory of elementary divisors with which we shall be concerned in this chapter and the next have reference to problems in which λ-matrices occur only indirectly. A typical problem is the theory of a pair of bilinear forms. The matrices a and b of these two forms have constant elements, and we get our λ-matrix only by considering the matrix $a - \lambda b$ of the pencil of forms determined by the two given forms. It will be noticed that this matrix is of the first degree, and in fact we shall deal, from now on, exclusively with λ-matrices of the first degree.

By the side of this simplification, a new difficulty is introduced, as will be clear from the following considerations. We shall subject the two sets of variables in the bilinear forms to two non-singular linear transformations whose coefficients we naturally assume to be constants, that is, independent of λ. These transformations have the effect of multiplying the λ-matrix, $a - \lambda b$, by certain non-singular matrices whose elements are constants (cf. § 36) and therefore, by § 94, carry it over into an equivalent λ-matrix which is evidently of the first degree. The transformations of § 94, however, were far more general than those just referred to, so that it is not at all obvious whether every λ-matrix of the first degree equivalent to the given one can be obtained by transformations of the sort just referred to or not.

These considerations show the importance of the following theorem:

THEOREM 1. *If* a_1, a_2, b_1, b_2 *are matrices with constant elements of which the last two are non-singular, and if the λ-matrices of the first degree*

$$m_1 \equiv a_1 - \lambda b_1, \qquad m_2 \equiv a_2 - \lambda b_2$$