

A NEW METHOD IN THE EQUIVALENCE OF PAIRS OF BILINEAR FORMS*

BY

R. G. D. RICHARDSON

INTRODUCTION

In attempting to clear the ground for an attack on the problem of relative maxima and minima of two quadratic or hermitian forms with an infinite number of variables, the author found that various questions insistently arose concerning the equivalence of pairs of forms in a finite number of variables and their reduction to normal type. The methods of attack in this theory as set forth in the literature seem to lack systematic unity. Not only are different devices employed for the treatment of the various phases of the problem, but these devices vary for different cases of the same phase. In investigating the problem of maxima and minima the author uncovered fundamental principles which may serve as a new interpretation of the problem of equivalence and make possible a new and more unified conception of the treatment. This new foundation for the theory of bilinear forms is set forth in this memoir while the problem of relative extrema is developed in the succeeding one.†

To exhibit the formal connection between the problems of relative extrema and of equivalence of pairs of forms, we note that in considering the bilinear forms

$$A(x, y) \equiv \sum_{ij}^{1, n} a_{ij} x_j y_i, \quad B(x, y) \equiv \sum_{ij}^{1, n} b_{ij} x_j y_i,$$

and their equivalence under linear transformations, it is customary for convenience to study the pencil of forms

$$(1) \quad A(x, y) - \lambda B(x, y)$$

with which is intimately connected the matrix $\|a_{ij} - \lambda b_{ij}\|$. This introduction of a parameter λ is on the other hand a familiar device in the problem of relative extrema. For example, on setting up the problem

* Presented to the Society, December 29, 1920.

† *Relative extrema of pairs of quadratic and hermitian forms*, these Transactions, vol. 26, pp. 479-494. This will be referred to as II.

$$(2) \quad A(x, x) = \text{minimum}, \quad B(x, x) = 1,$$

we are led to a consideration of the minimum of

$$(3) \quad A(x, x) - \lambda B(x, x).$$

In the consideration of either problem the linear homogeneous equations obtained by equating to zero the partial derivatives of (3) with respect to x_i (or of (1) with respect to y_i),

$$(4) \quad (a_{i1} - \lambda b_{i1})x_1 + \cdots + (a_{in} - \lambda b_{in})x_n = 0 \quad (i = 1, \dots, n),$$

are of fundamental importance. In addition to these, the equations

$$(4') \quad (a_{1j} - \lambda b_{1j})y_1 + \cdots + (a_{nj} - \lambda b_{nj})y_n = 0 \quad (j = 1, \dots, n),$$

obtained by differentiating (1) with respect to x_j , often come into consideration. The homogeneous equations (4) have a solution when and only when λ is a characteristic number, that is, one of the n zeros $\lambda_1, \dots, \lambda_n$ of the determinant $|a_{ij} - \lambda b_{ij}|$. On setting λ equal to λ_k , denoting the corresponding solutions of (4) by

$$X^{(k)} \equiv (x_1^{(k)}, \dots, x_n^{(k)}),$$

multiplying the equations by $x_1^{(k)}, \dots, x_n^{(k)}$ respectively and adding, there results the fundamental relation

$$(5) \quad A(X^{(k)}, X^{(k)}) - \lambda_k B(X^{(k)}, X^{(k)}) = 0,$$

which may be compared with (1) and (3).

Any finite solution of the minimum problem (2) will be among the n possible solutions of (4). At least in the case where the coefficients $A(x, x)$ and $B(x, x)$ are real and the latter form is positive definite, a minimum exists and we note from (5) that the solution

$$X^{(1)} \equiv (x_1^{(1)}, \dots, x_n^{(1)}),$$

corresponding to the smallest characteristic number λ_1 and normalized to make $B(x, x) = 1$, furnishes to A the minimum value λ_1 .

When B is definite, the other $n-1$ solutions of (4), $X^{(2)}, \dots, X^{(n)}$, also have interpretations as minima, as we shall briefly sketch. Adding to the minimum problem (2) the orthogonal condition

$$(6) \quad B(X^{(1)}, x) \equiv \sum_{ij}^{1,n} b_{ij} x_j^{(1)} x_i = 0,$$

leads in formal fashion (II, § 3) to a consideration of the minimum of the expression

$$A(x, x) - \lambda B(x, x) - 2\mu B(X^{(1)}, x),$$

which, on differentiation with regard to x_i , gives the non-homogeneous linear equations

$$(7) \quad \sum_j (a_{ij} - \lambda b_{ij}) x_j = \mu [b_{i1} x_1^{(1)} + \dots + b_{in} x_n^{(1)}]$$

in the variables x_1, x_2, \dots, x_n . It turns out in the case where B is definite (and even in a more general case) that μ is 0 and that the solution of the extended minimum problem is thus also a solution of (4), which on account of the relation (6) cannot be $X^{(1)}$ and hence from (5) must be

$$X^{(2)} \equiv (x_1^{(2)}, \dots, x_n^{(2)}),$$

corresponding to the second smallest characteristic number λ_2 , and furnishing to $A(x, x)$ the minimum value λ_2 .

One may seek the minima under the successively added conditions

$$(8) \quad B(X^{(2)}, x) = 0, \quad \dots, \quad B(X^{(n-1)}, x) = 0,$$

where $X^{(2)}, \dots, X^{(n-1)}$ are the solutions of the successive problems. Formally this leads to a consideration of the minimum of

$$A(x, x) - \lambda B(x, x) - 2\mu_1 B(X^{(1)}, x) - \dots - 2\mu_{n-1} B(X^{(n-1)}, x),$$

and hence to the linear equations

$$(9) \quad \sum_i (a_{ij} - \lambda b_{ij}) x_i = \mu_1 \sum_i b_{ij} x_i^{(1)} + \dots + \mu_{n-1} \sum_i b_{ij} x_i^{(n-1)}$$

and (6) and (8) together with the quadratic relation $B(x, x) = 1$. Here again it turns out that when B is definite (and even in a more general case), all the μ 's are zero and the solution of the minimum problem is the solution $X^{(n)}$ of the equation (4), corresponding to the largest characteristic number λ_n which on account of (5) gives $A(x, x)$ the value λ_n .

On the other hand the problem of determining equivalence of pairs of forms is closely associated with that of finding linear transformations

The problem of reduction of a pair of bilinear forms is then so to choose $\|X\|$, $\|Y\|$ that this matrix (11) (or (11')) is of simplest possible shape. Since the coefficients of the new bilinear forms are values of $A(\xi, \eta)$, $B(\xi, \eta)$, the ξ , η must be chosen as sets of numbers such that the transformation (10) will reduce A and B to particularly simple forms; to zeros so far as possible, and the remaining ones to unity so far as possible. We shall find that the solutions of (4) and (4') are particularly useful in this connection, since $A(\xi, \eta)$ and $B(\xi, \eta)$ are both zeros when ξ , η are solutions of (4) and (4') corresponding to distinct values of λ . It turns out that in the regular case, which includes that when $B(x, x)$ is definite, there are n linearly independent solutions each of equations (4) and (4'), and that if the rows of the matrices X and Y are chosen as such solutions properly normalized and orthogonalized, the new bilinear forms take on the simplest of all shapes,

$$B'(x, y) = x_1 y_1 + \cdots + x_n y_n, \quad A'(x, y) = \lambda_1 x_1 y_1 + \cdots + \lambda_n x_n y_n.$$

We note further that the n solutions of (4) chosen for the coefficients of the transformations are the solutions of the corresponding minimum problems.

The discussion (§ 5) of the solutions of (7) and (9), which are here forced on our attention, furnishes an interesting chapter in the solution of linear algebraic equations. The situation may be conceived of somewhat as follows.* When in considering the solutions of (4) it is found that there are multiple roots of the determinant $|a_{ij} - \lambda b_{ij}|$, it is possible to alter the elements so as to make all the roots distinct. On passing to the limit so that m of the λ 's become equal, it is fairly obvious that there are various possibilities. Solutions of the linear equations (4) corresponding to the infinitesimally different λ 's may all approach the same limiting solution, in which case there is only one linearly independent solution corresponding to this m -valued λ (we shall say† λ -multiplicity = m , index = 1, solution-multiplicity = m); they may all approach different limiting solutions, in which case there are m linearly independent solutions (λ -multiplicity = m ; index = m , each solution-multiplicity = 1). Between these two extremes a variety of cases arise. If the number of linearly independent solutions of the x 's is p there will be m_1, m_2, \dots, m_p -tuple solutions respectively (λ -multiplicity = m ; index = p ; solution-multiplicity = m_1, m_2, \dots, m_p ; $\sum m_i = m$). These notions have been used by others‡ in proving some of the fundamental theorems of matrices.

* Cf. examples in footnote § 5.

† Sylvester used the terms nullity and vacuity; the terms here chosen are more in line with those used in the theories of differential and integral equations.

‡ For example, Taber, *On the theory of matrices*, American Journal of Mathematics, vol. 12, p. 3.

We have here gone further and set up sets of related equations, called derived equations and similar to (7) and (9), which give solutions to take the place of those lost by the coalescing of distinct ones. The theory of derived equations enables us to compute each solution of (4) and its multiplicity m_i . No details are here presented except those which are pertinent to the problem in hand. The relations of the theories of derived equations and of elementary divisors would furnish material for another memoir.

In the irregular case where the solution-multiplicities are not all unity, the problem of reduction of a pair of bilinear forms is thus more complex than in the regular case. There will be solutions of (4) to fill up s only of the rows of the x and y matrices, but as indicated above there exist $n-s$ solutions of derived equations also and these are used to fill up the gaps. On multiplication by matrices thus formed, (11) takes on the normal types which are well known in the theory of bilinear forms (§ 6).

The procedure in the reduction of pairs of bilinear forms is thus made direct and a simple method of calculation made possible. By this point of view not only is the need made apparent for distinguishing between the cases where the normal types are different, but many related questions are answered. On the other hand, the reductions as worked out by Weierstrass, Kronecker, Darboux and others are not only complex in theory and difficult in technique but they involve the introduction of methods which cannot be classed strictly as algebraic.

The criterion for equivalence of bilinear forms developed in § 6 must naturally be equivalent to that ordinarily expressed in terms of elementary divisors. The new criterion, however, appears essentially simpler in its fundamental conception. It depends only on the notions of the solutions of linear homogeneous and non-homogeneous equations. The necessary and sufficient condition that two pairs of bilinear forms be equivalent is that the λ -polynomials be equivalent and that for any root λ each solution-multiplicity for one set of equations be matched with an equal one for the other.

As an example of the straightforward treatment possible under this new theory, an application is made in § 7 to the reduction to normal type of a single quadratic form, the transformation used being orthogonal.

While this memoir concerns itself only with those families of bilinear forms which are not identically singular in λ , the methods and results may be extended to singular families. In that case the equations (4) or (5) or both will have solutions which involve λ identically. In order completely to characterize the type, certain constants which occur in these

the determinants $|a_{ij}|$, $|b_{ij}|$ are respectively the constant term and the coefficient of λ^n in the expansion of this determinant as a polynomial in λ , there are n zeros $\lambda_1, \dots, \lambda_n$ of the determinant, and if A is non-singular, none of these vanish. Some or all of the λ -zeros may be equal.

For a simple zero λ_k , the determinant $|a_{ij} - \lambda b_{ij}|$ will be of rank $n-1$ and hence there will correspond one linearly independent solution

$$X^{(k)} \equiv (x_1^{(k)}, \dots, x_n^{(k)})$$

of the equations (13) (and one

$$Y^{(k)} \equiv (y_1^{(k)}, \dots, y_n^{(k)})$$

of the equations (14)). For a q -fold zero λ_k the determinant will have rank $n-r$, where r may have any value from 1 to q ; and there will correspond r linearly independent solutions of (13). The linearly independent solutions of (13) may be designated *primary solutions*. The simplest case, where all the λ -roots are distinct, will be treated in § 2. The next simplest case, where some of the λ -roots are multiple but where to such a root there correspond as many linearly independent solutions X as the multiplicity ($r = q$), is treated in § 3. In both these cases the total number of primary solutions is n . The most difficult case, where the number of linearly independent solutions corresponding to any multiple λ -root is less than the multiplicity, is treated in § 5.

2. SETS OF LINEAR EQUATIONS. FIRST REGULAR CASE

In this section we shall discuss the simplest problem connected with the sets of linear equations (13) and (14), viz. when all the zeros of the λ -determinant are distinct. The n primary solutions of (13) and (14) may be designated respectively as follows:

$$(16) \quad X^{(i)} \equiv (x_1^{(i)}, \dots, x_n^{(i)}) \quad (i = 1, \dots, n),$$

$$(16') \quad Y^{(i)} \equiv (y_1^{(i)}, \dots, y_n^{(i)}) \quad (i = 1, \dots, n).$$

If we consider the characteristic number λ_k and multiply equations (13) by any set of numbers $Y \equiv (y_1, \dots, y_n)$ respectively and add, we get

$$(17) \quad A(X^{(k)}, Y) = \lambda_k B(X^{(k)}, Y).$$

On the other hand by considering the characteristic number λ_k and multiplying the equations (14) by any set of numbers $X = (x_1, \dots, x_n)$

respectively and adding, we obtain

$$(18) \quad A(X, Y^{(k')}) = \lambda_{k'} B(X, Y^{(k')}).$$

When we use for multipliers Y and X some solutions of (14) and (13) respectively, the relations (17) and (18) reduce to

$$(19) \quad A(X^{(k)}, Y^{(k')}) = \lambda_k B(X^{(k)}, Y^{(k')}), \quad A(X^{(k)}, Y^{(k')}) = \lambda_{k'} B(X^{(k)}, Y^{(k')}),$$

which, when no ambiguity can result, may be written in the form

$$(19') \quad A(k, k') = \lambda_k B(k, k'), \quad A(k, k') = \lambda_{k'} B(k, k').$$

When the characteristic numbers are different ($\lambda_k \neq \lambda_{k'}$) there follow from (19) the two fundamental relations

$$(20) \quad A(k, k') = 0, \quad B(k, k') = 0,$$

which will be designated as the *orthogonal relations* for $X^{(k)}$ and $Y^{(k')}$ with regard to the matrices $\|a_{ij}\|$ and $\|b_{ij}\|$ respectively. Hence follows

THEOREM I. *The solutions $X^{(k)}$ and $Y^{(k')}$ are orthogonal provided they correspond to unequal characteristic numbers.*

As will be shown in Theorem IV for a more general case, the solutions (16) (and also (16')) are linearly independent.

Given the solutions $X^{(k)}$, $Y^{(k)}$ corresponding to the same characteristic number λ_k , we shall say that $X^{(k)}$, $Y^{(k)}$ are *normalized* provided that one or both solutions are multiplied by constants such as will make $B(k, k)$ equal to unity.

THEOREM II. *Any solutions $X^{(k)}$, $Y^{(k)}$ which correspond to a simple characteristic number can be normalized.*

To prove this we note from Theorem I that when $k \neq k'$, $B(k, k') = 0$. Were $B(k, k)$ also equal to zero, it would follow as in the Introduction (Formula 11) that the product of three matrices $\|Y\|$, $\|b_{ij}\|$, $\|X\|$ of rank n would be a matrix with every element in the k th row zero; that is, three non-vanishing determinants would have a vanishing product. Hence $B(k, k)$ must be different from zero, and on multiplying $X^{(k)}$ or $Y^{(k)}$, or both, by proper constants it may be made equal to unity, and these solutions will from now on be assumed normalized.

Since on setting $k' = k$ in (19') we have

$$(21) \quad A(k, k) = \lambda_k B(k, k),$$

it follows that when the solutions are normalized, $A(k, k)$ has the value λ_k .

Incidentally the following theorem has been established:

THEOREM III. *For solutions (16) and (16'), it is impossible that $B(k, k')$ vanish for all values of k' when k is fixed or for all values of k when k' is fixed.*

The advantages of using the solutions (16) and (16') as the coefficients of the linear transformations to be used in the reduction of the bilinear forms is at once evident if we apply Theorems I and II to the matrix (11). We note that the orthogonal properties make $A(k, k')$, $B(k, k')$ both zero when $k \neq k'$, and that on account of normalization the remaining terms (in the main diagonal) may be written $\lambda_i - \lambda$. This is stated in the form of a theorem in § 4.

3. SETS OF LINEAR EQUATIONS. SECOND REGULAR CASE

When λ is a multiple root of the determinant $|a_{ij} - \lambda b_{ij}|$, the number of corresponding linearly independent solutions (16) of (13) may be equal to or less than the multiplicity. The former case is much simpler and will be treated in this section. If by a slight change of one or more of the coefficients a_{ij} , b_{ij} the multiple root of the determinant is broken up into distinct ones slightly separated, the solutions X, Y of (13) and (14) will be only slightly altered. These altered solutions are subject to the orthogonal and normal properties of Theorems I and II, not only with other solutions but among themselves. It is natural to suppose, on passing back to the multiple λ , that when these solutions X, Y have distinct limits the limits will satisfy the same conditions. That this is true and that this second case is thus intimately related to the first, we shall now proceed to demonstrate.

Concerning the solutions of the equations (13), which may be written in the form

$$(13') \quad a_{i1}x_1 + \cdots + a_{in}x_n = \lambda(b_{i1}x_1 + \cdots + b_{in}x_n),$$

it is possible to prove the following

THEOREM IV. *The n solutions (16) of the equation (13) are linearly independent provided that those corresponding to a multiple characteristic number λ are chosen so as to be linearly independent among themselves.*

For if not, let us consider that the matrix of the x 's in (16) be of rank $n - m$. Then there exist constants c_1, \dots, c_n such that

$$(22) \quad c_1 x_j^{(1)} + \cdots + c_n x_j^{(n)} = 0 \quad (j = 1, \dots, n),$$

where m of the c 's may be assigned at random (except that not all are zero). On the other hand let us pick out any one of the equations (13)

(say the i th), give to λ each of the characteristic values in succession, and consider the corresponding solutions. If the equations are multiplied respectively by c_1, c_2, \dots, c_n and added, the sum of each column on the left vanishes, and we have the resulting equation

$$(23) \quad 0 = b_{i1}(\lambda_1 c_1 x_1^{(1)} + \dots + \lambda_n c_n x_1^{(n)}) + \dots + b_{in}(\lambda_1 c_1 x_n^{(1)} + \dots + \lambda_n c_n x_n^{(n)}).$$

Giving to i its n values, these are n equations with coefficients b_{ij} , the determinant of which is by hypothesis different from zero. It follows then that the corresponding variables are zero:

$$(24) \quad \lambda_1 c_1 x_j^{(1)} + \dots + \lambda_n c_n x_j^{(n)} = 0 \quad (j = 1, \dots, n).$$

Since the rank of the matrix of the $x_i^{(j)}$ is assumed to be $n - m$, the values of m of the $\lambda_i c_i$ may be assigned at random (except that not all are zero), and the values of the remaining $n - m$ can then be expressed linearly in terms of those assigned, the coefficients being formed from certain minors of the x 's. But by hypothesis the c 's satisfy the equations (22) and by the same processes $n - m$ of these can be expressed linearly in terms of m assigned arbitrarily. The coefficients of these linear relations are compounded from the x 's in the same way as they are in the case of the linear relations obtained from (23). Hence the two sets of coefficients are equal, and we have

$$\begin{aligned} \lambda_i c_i &= k_{n-m+1}^{(i)} \lambda_{n-m+1} c_{n-m+1} + \dots + k_n^{(i)} \lambda_n c_n \quad (i = 1, \dots, n-m), \\ c_i &= k_{n-m+1}^{(i)} c_{n-m+1} + \dots + k_n^{(i)} c_n \quad (i = 1, \dots, n-m), \end{aligned}$$

where c_{n-m+1}, \dots, c_n are arbitrary: for example, they may all be zero except one, and this one may be successively c_{n-m+1}, \dots, c_n . But this is only possible when $\lambda_1 = \lambda_2 = \dots = \lambda_n$, that is when λ is an n -tuple root. But even this case is ruled out, for we assumed that those of the solutions (16) corresponding to the same characteristic number were chosen linearly independent. Hence a contradiction is established and the solutions are linearly independent.

It is possible to proceed further and to show that the solutions corresponding to the multiple characteristic numbers may be so chosen that the n solutions (16) and (16') are normalized and mutually orthogonal.

As a typical case, let us assume that there are three characteristic numbers λ_1 , all equal, and no others, and show how to set up three normalized linearly independent solutions having among themselves the

orthogonal property. In the proof the linearly independent $Y^{(1)}, Y^{(2)}, Y^{(3)}$ will be taken at random and the X 's adapted to fulfil the conditions, but other methods may be used (cf. § 7). Denoting by $X^{(k_1)}, X^{(k_2)}, X^{(k_3)}$ three linearly independent solutions corresponding to the value of λ under consideration, we must first show that there exist constants $\alpha_1, \alpha_2, \alpha_3$, such that the solution $X^{(1)} \equiv \alpha_1 X^{(k_1)} + \alpha_2 X^{(k_2)} + \alpha_3 X^{(k_3)}$ satisfies the relations $B(1,1) = 1, B(1,2) = 0, B(1,3) = 0$. The bilinear relation is linear in each of the variables and these equations may be written

$$(25) \quad \begin{aligned} \alpha_1 B(k_1, 1) + \alpha_2 B(k_2, 1) + \alpha_3 B(k_3, 1) &= 1, \\ \alpha_1 B(k_1, 2) + \alpha_2 B(k_2, 2) + \alpha_3 B(k_3, 2) &= 0, \\ \alpha_1 B(k_1, 3) + \alpha_2 B(k_2, 3) + \alpha_3 B(k_3, 3) &= 0. \end{aligned}$$

The determinant of the coefficients B is the product of three determinants, as we note from a formula analogous to (11). Since by hypothesis all three of these determinants $|X|, |Y|$ and $|b_{ij}|$ are non-singular, the determinant of the B 's is different from zero and the constants $\alpha_1, \alpha_2, \alpha_3$ may be determined to satisfy (25).

In similar fashion constants $\beta_1, \beta_2, \beta_3$ may be determined such that

$$X^{(2)} \equiv \beta_1 X^{(k_1)} + \beta_2 X^{(k_2)} + \beta_3 X^{(k_3)}$$

satisfies the relations $B(2, 1) = 0, B(2, 2) = 1, B(2, 3) = 0$, and constants $\gamma_1, \gamma_2, \gamma_3$ such that

$$X^{(3)} \equiv \gamma_1 X^{(k_1)} + \gamma_2 X^{(k_2)} + \gamma_3 X^{(k_3)}$$

satisfies the relations $B(3, 1) = 0, B(3, 2) = 0, B(3, 3) = 1$. That $X^{(1)}, X^{(2)}, X^{(3)}$ are linearly independent is easily shown.

The method of treatment used in this special case is of general application. If besides λ_1 , for which the index and solution-multiplicity are assumed to be q , there are other characteristic numbers $\lambda_2, \lambda_3, \dots$ the number of equations in the q variables corresponding to (25) will be q , the other similar relations being satisfied identically, as may be seen from Theorem II. The determinant of the B 's cannot be zero, for it is a minor of an n -rowed determinant which is the product of three determinants whose values are different from zero; and further all terms of this n -rowed determinant in the first p rows and p columns except this minor composed of the B 's are zero. This completes the proof of the following

THEOREM V. *When each solution-multiplicity of the solutions of (13) is unity, it is possible to choose the linearly independent solutions (16) and (16')*

so that they are normalized with respect to the matrix $\|b_{ij}\|$ and mutually orthogonal with respect to both $\|a_{ij}\|$ and $\|b_{ij}\|$.

We may note further that Theorem III is valid for this case also.

4. SIMULTANEOUS REDUCTION OF TWO MATRICES.

FIRST AND SECOND REGULAR CASES

We are now in a position to discuss the simplest case of the simultaneous reduction of two matrices $\|A\|$ and $\|B\|$ to normal forms. Since the first and second cases treated in §§ 2, 3 are essentially similar they can here be handled together. The number of primary solutions each of (13) and (14) is in both cases n .

In reducing simultaneously the matrices $\|A\|$, $\|B\|$ the λ -matrix $\|A - \lambda B\|$ compounded from them may be considered as playing a fundamental rôle. Denoting the transposed matrix of the sets of solutions (16) by $\|\bar{X}\|$ and the matrix of the sets (16') by $\|Y\|$, and effecting the matrix multiplication

$$\|Y\| \cdot \|A - \lambda B\| \cdot \|\bar{X}\|,$$

we have by Theorems I, II and V and formula (11) this simplest of all normal forms,

$$\left\| \begin{array}{cccccc} \lambda_1 - \lambda & 0 & 0 & \dots & 0 & \\ & 0 & \lambda_2 - \lambda & 0 & \dots & 0 \\ & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & 0 & 0 & 0 & \dots & \lambda_n - \lambda \end{array} \right\|.$$

As was noted in the Introduction, this is equivalent to the following

THEOREM VI. *By means of linear transformations of the x 's and y 's whose coefficients are the columns of (16) and (16') which are respectively the solutions of the equations (13) and (14) associated with the bilinear forms $A(x, y)$, $B(x, y)$, these forms are reduced to*

$$\begin{aligned} B(x, y) &= x_1 y_1 + x_2 y_2 + \dots + x_n y_n, \\ A(x, y) &= \lambda_1 x_1 y_1 + \lambda_2 x_2 y_2 + \dots + \lambda_n x_n y_n. \end{aligned}$$

5. SETS OF LINEAR DERIVED EQUATIONS. IRREGULAR CASE

The simple cases treated in the previous sections may be considered as regular. There are irregular cases branching off at several points in the treatment. Before undertaking a study of these irregular cases let us discuss a possible conception of their relation to the regular case.

In the irregular case there will be for some λ a solution X of the equations (13) which may be regarded as multiple. By a change of one of the coefficients a_{ij} the root of the determinant $|a_{ij} - \lambda b_{ij}|$ can be broken up into two or more. Denoting by $X_{\epsilon}^{(1)}, X_{\epsilon}^{(2)}$ two solutions of (13) corresponding to infinitesimally different λ 's and by $Y_{\epsilon}^{(1)}, Y_{\epsilon}^{(2)}$ those of (14), we have from Theorems I, II

$$\begin{aligned} B(X_{\epsilon}^{(1)}, Y_{\epsilon}^{(2)}) &= 0, & B(X_{\epsilon}^{(1)}, Y_{\epsilon}^{(1)}) &= 1, \\ B(X_{\epsilon}^{(2)}, Y_{\epsilon}^{(1)}) &= 0, & B(X_{\epsilon}^{(2)}, Y_{\epsilon}^{(2)}) &= 1. \end{aligned}$$

Let us now pass back to the limiting case where λ is a multiple root of the determinant. When the limiting solutions $X^{(1)}, X^{(2)}$ are linearly independent, as they were in the regular case, discussed in § 3, the same relations held in the limit:

$$\begin{aligned} B(X^{(1)}, Y^{(2)}) &= 0, & B(X^{(1)}, Y^{(1)}) &= 1, \\ B(X^{(2)}, Y^{(1)}) &= 0, & B(X^{(2)}, Y^{(2)}) &= 1. \end{aligned}$$

But when in the limit these solutions X coalesce, these four relations must reduce to a single one, and this turns out to be $B(X, Y) = 0$. The limit of such expressions as $B(X^{(1)}, Y^{(2)})$ must be zero while in the case of $B(X^{(1)}, Y^{(1)})$ the constant multipliers needed to make it unity have become infinite.

There are not enough linearly independent solutions to make sets such as (16) or (16'), and the previous method of procedure in the reduction of pairs of forms breaks down. The difference of two solutions $X_{\epsilon}^{(1)}, X_{\epsilon}^{(2)}$ is suggested as a substitute for one of them and then the derivative with regard to λ is naturally considered. We shall find that the introduction of the solutions of derived equations in place of those missing from the set of primary solutions furnishes the way out of the difficulty.

As we have noted in § 1, for a q -fold zero λ_k the λ -determinant will have rank $n-r$ where r may have any value from 1 to q and there are r linearly independent solutions of (13). The numbers q and r which are both of fundamental significance may be designated respectively the λ -multiplicity and the index of λ_k and the r linearly independent solutions of (13) have been designated primary solutions. When r is less than q , as it is in the irregular case discussed in this section, corresponding to one or more of these equal λ 's there must be solutions which may be conceived of as multiple solutions of the equations (13). By an alteration of one of the coefficients a_{ij} the q zeros of the λ -determinant become distinct

and the q solutions X will also be distinct. On passing back continuously to the original a_{ij} , each of the q distinct solutions X must have a limit. When the number r of the limiting solutions is less than q , some of them must be multiple solutions. The numbers of solutions X which have united to form one of the limiting solutions may be m_1 , to form another m_2 , and to form the r th it may be m_r . These numbers m_1, \dots, m_r may be called the *solution-multiplicities* of the solutions of the equations in the x 's corresponding to this λ_k and are subject to the relation $m_1 + m_2 + \dots + m_r = q$. Since multiple solutions of equations or sets of equations are often also solutions of the corresponding equations formed by taking the derivatives, it is natural to investigate if something of that sort takes place in this problem.

Let us now take up the detailed discussion of the irregular case of this section. For some characteristic number λ_k the λ -index is less than the λ -multiplicity, and hence not all the solution-multiplicities are unity. By altering one of the coefficients a_{ij} , two roots of the characteristic equation will be made unequal. Calling them λ_1 and $\lambda_1 + \Delta\lambda$, the corresponding solutions of (13) x_1, \dots, x_n and $x_1 + \Delta x_1, \dots, x_n + \Delta x_n$ respectively, and the new a 's, a'_{ij} , and subtracting the two sets of equations we have

$$\sum_j (a'_{ij} - \lambda_1 b_{ij}) \Delta x_j = \Delta \lambda \sum_j [b_{ij} (x_j + \Delta x_j)],$$

while on dividing by $\Delta \lambda$ and passing to the limit the relations

$$(26) \quad \sum_j (a_{ij} - \lambda_1 b_{ij}) \frac{dx_j}{d\lambda} = \sum_j b_{ij} x_j$$

result. This set of non-homogeneous equations* may be obtained formally by taking the derivative with regard to λ of the set (13), but for our purposes they may be best written in the form

$$(26') \quad \sum_j (a_{ij} - \lambda_1 b_{ij}) x_j = \sum_j b_{ij} x_j^{(1)},$$

where $x_j^{(1)}$ are the constants which are the primary solution of (13) for λ_1 .

* That the equations (26) have a solution may also be seen from another standpoint as follows. It will be necessary only to show that if the matrix $\|a_{ij} - \lambda b_{ij}\|$ be augmented by the addition of another column composed of the terms on the right hand side, the rank is not greater than $n - 1$. Now, if the rows are multiplied by $y_1^{(1)}, \dots, y_n^{(1)}$ and each column is added, the result will be zero. For the first n columns this will be true because the y 's are defined to be the solution of such equations and for the last column because of the relation $B(X^{(1)}, Y^{(1)}) = 0$ which we have noted above.

If λ_1 is a triple root, similar procedure shows that the second derivatives of X satisfy the non-homogeneous equations

$$(27) \quad \sum_j (a_{ij} - \lambda_1 b_{ij}) \frac{d^2 x_j}{d\lambda^2} = \sum_j b_{ij} \frac{dx_j}{d\lambda},$$

where $dx_i/d\lambda$ are the constants which represent the solution obtained from (26). And if λ_1 is a p -tuple root there are solutions X , $dX/d\lambda$, \dots , $d^{p-1}X/d\lambda^{p-1}$, the last satisfying the non-homogeneous equations

$$(28) \quad \sum_j (a_{ij} - \lambda_1 b_{ij}) \left(\frac{d^{p-1} x_j}{d\lambda^{p-1}} \right) = \sum_j b_{ij} \frac{d^{p-2} x_j}{d\lambda^{p-2}}.$$

These equations (26–28) may be designated *derived equations* and the solutions will be called solutions* of the derived equations. The non-homogeneous equations (26–28) may for our purposes best be written

* As examples consider each of four sets of linear equations

(α)	(β)	(γ)	(δ)
$(1-\lambda)x_3 = 0,$	$(1-\lambda)x_3 = 0,$	$(1-\lambda)x_3 = 0,$	$(1-\lambda)x_3 = 0,$
$(b-\lambda)x_2 = 0,$	$(1-\lambda)x_2 = 0,$	$(1-\lambda)x_2 + x_3 = 0,$	$(1-\lambda)x_2 + x_3 = 0,$
$(c-\lambda)x_1 = 0;$	$(1-\lambda)x_1 = 0;$	$(1-\lambda)x_1 = 0;$	$(1-\lambda)x_1 + x_2 = 0.$

On equating the determinant of the coefficients of (α) to zero, we have the characteristic numbers $\lambda = 1, b, c$ and corresponding to these are three linearly independent solutions of (α) which may be written $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. In (β) the characteristic numbers are equal but the solutions may be written down as in (α). In fact (β) may be regarded as the limiting case of (α) when b and c approach 1. When λ is set equal to 1 in equations (α) the rank of the matrix is two while in (β) it is zero.

In (γ) and (δ) the determinants again have λ as a triple root, there being two linearly independent solutions, $(1, 0, 0)$, $(0, 1, 0)$; and one, $(1, 0, 0)$, respectively. If one coefficient be modified slightly the characteristic numbers will be separated and there will be three linearly independent solutions of the equations in x . On passing back to the original form, two solutions of (γ) unite in one and the other remains distinct, while in (δ) they all unite in one.

In all three sets of equations (β), (γ), (δ), $n = 3$, $q = 3$; while r is 3, 2, 1, respectively. For (β), $m_1 = m_2 = m_3 = 1$; for (γ), $m_1 = 2$, $m_2 = 1$; for (δ), $m_1 = 3$. In the case of the set (δ) there will be two successive sets of derived equations,

(ϵ)	(ζ)
$(1-\lambda) \frac{dx_3}{d\lambda} = x'_3 (= 0),$	$(1-\lambda) \frac{d^2 x_3}{d\lambda^2} = \left(\frac{dx_3}{d\lambda} \right)' (= 0 + c, 0),$
$(1-\lambda) \frac{dx_2}{d\lambda} + \frac{dx_3}{d\lambda} = x'_2 (= 0),$	$(1-\lambda) \frac{d^2 x_2}{d\lambda^2} + \frac{d^2 x_3}{d\lambda^2} = \left(\frac{dx_2}{d\lambda} \right)' (= 0 + c, 1),$
$(1-\lambda) \frac{dx_1}{d\lambda} + \frac{dx_2}{d\lambda} = x'_1 (= 1);$	$(1-\lambda) \frac{d^2 x_1}{d\lambda^2} + \frac{d^2 x_2}{d\lambda^2} = \left(\frac{dx_1}{d\lambda} \right)' (= 1 + c, 0);$

where the primes on the right indicate that the solutions found in the next previous set of equations are substituted for these symbols. The augmented matrix in (ϵ) consists of

$$(28') \quad \sum_j (a_{ij} - \lambda_1 b_{ij}) x_j = \sum_j b_{ij} \bar{x}_j,$$

where the constants \bar{x}_j are the solution of the next preceding derived equation.

The sum of the solution-multiplicities corresponding to λ_k is v , as we have seen above; the total number for all λ 's is n and the sum of the λ -multiplicities for the various λ 's is also n . The total number of primary solutions (that is, solutions of (13) for all values of λ) is less than n in the case considered here; together with the solutions of the derived equations (26-28 or 28') the total is n . The simplest case, arising when the total number of primary solutions is n , that is, when there is no multiple solution of (13), has been called the regular case, and treated in §§ 2-4.

The systems of equations (13) and (14) are called *adjoint* to one another. They have the same determinant, the same characteristic numbers, and the same solution-multiplicities. When there are m linearly independent solutions of one set, there are the same number of the other. Similar derived equations may be set up for the y 's; corresponding to each λ there will be the same number each of primary solutions and solutions of the derived equations as there are for the x 's.

Corresponding to any λ the solutions may be arranged in as many groups as the index indicates, and within each group the primary solution may be followed by the successive derived solutions. To every non-singular matrix $\|A - \lambda B\|$ corresponds then a total of n primary solutions and solutions of the derived equations in x , and n for the equations in y ; the totality can for convenience be designated as before in (16) and (16') by $X^{(1)}, \dots, X^{(n)}; Y^{(1)}, \dots, Y^{(n)}$. As we shall see in § 6, these solutions when properly chosen with reference to orthogonalization and normalization furnish exactly the matrices $\|X\|, \|Y\|$ needed to make the transformation of pairs of bilinear forms to canonical shape.

It may be well here to interpolate some remarks concerning the number of fundamental solutions on which the primary solutions and the solutions of derived equations depend.

Because it presents the simplest problem, the case where a λ -index is unity will be discussed first.* Let us assume further that the solution-multiplicity is three. Designating by $\Xi^{(1)} = (\xi_1^{(1)}, \dots, \xi_n^{(1)})$ a solution of

four columns and three rows and for $\lambda = 1$ is of rank 2. One solution is $(0, 1, 0)$ and from the theory of linear equations we know that the general solution is $(1, 0, 0) + c_1(0, 1, 0) \equiv (1, c_1, 0)$. For (ζ) the augmented matrix is still of rank 2 while one solution is $(0, 0, 1)$ and the general solution is $(1, 0, 0) + c_1(0, 1, 0) + c_2(0, 0, 1) \equiv (1, c_1, c_2)$. The one solution of the original set (δ) has been augmented by two others and the solutions $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ of (δ) and its derived equations play a fundamental rôle in the discussion that follows.

* In this connection, see last foot-note above.

the equations (13), the rank of the matrix of coefficients on the left of (26) is $n-1$ and since it is known that there are solutions of these non-homogeneous equations, this matrix augmented by a new column made up of the terms on the right (after the set of constants $\Xi^{(1)}$ is substituted for the x 's) cannot be greater than $n-1$. We may denote by $\Xi^{(2)}$ a solution of the non-homogeneous equations (26) and note that by the addition of a constant multiplied by $\Xi^{(1)}$ this still remains a solution of (26). In fact this solution $\Xi^{(2)} + c\Xi^{(1)}$ is the most general possible.

Passing now to (27) we note again that the matrix of the coefficients on the left is of rank $(n-1)$ and that by substituting for the x 's on the right the solutions of (26) and the addition of the column to the already augmented matrix, the rank does not increase.* Since the solution substituted in the right-hand side is linearly dependent on the two sets of constants $\Xi^{(1)}$, $\Xi^{(2)}$ the general solution of (27) is linearly dependent on three. One of these, as may readily be noted from the theory of linear equations, is $\Xi^{(1)}$. Since one of the solutions substituted into the right hand side in place of the x 's is $\Xi^{(1)}$, another of these must be $\Xi^{(2)}$, the third is new and may be denoted by $\Xi^{(3)}$. Since the same considerations apply to equations (28), we have the following

THEOREM VII. *If λ_i has index unity, the solution of the $(p-1)$ th derived equation is linearly dependent on p sets of constants. Of these p sets of constants $p-1$ are solutions also of the $(p-2)$ th derived equation and one is new.*

When, however, the index of λ_i is $m > 1$ there are m linearly independent primary solutions of (13). If the number $p-1$ of solutions of derived equations corresponding to each of these is the same, the m primary solutions may at first be selected at random (except that they must be linearly in-

* If the matrix $\|a_{ij} - \lambda b_{ij}\|$ is extended to the right by the addition successively as columns of the right hand sides of (26), (27), etc., the resulting matrices have some interesting properties. Let us suppose, for example, that $n = 3$ and that there is only one primary solution. The normal type of the matrix $\|a_{ij} - \lambda b_{ij}\|$ will in this case be proved later to be (33), the determinant being divisible by $(\lambda_1 - \lambda)^3$ but being of rank 2. The four-column matrix obtained by augmenting it by the right hand side of (26) is, however, such that all three-column determinants are divisible by $(\lambda_1 - \lambda)^2$ but not all by a higher power. The fifth column made up of the right hand side of (27) introduces determinants which are divisible by $\lambda_1 - \lambda$ but not by a higher power, and this is true even when the fourth column is left out of consideration. The solutions of the corresponding equations in the x 's are all single-valued ones. The addition of a sixth column made up of the solutions of the second derived equations introduces determinants which do not vanish for $\lambda = \lambda_1$ even when the fourth and fifth columns are neglected. There is no solution of the corresponding equations.

The solution of the equations (26) can be regarded as a double one. In fact each determinant containing the fourth column is intimately related to the derivative with respect to λ of the determinant of the first three.

dependent), the solutions of the derived equation corresponding to each one being then built up. Each of the m solutions of the $(p-1)$ th derived equations will be linearly dependent on p sets of constants as in the case where the index is 1, and the total number of different sets of constants to be considered is mp . As will be shown later (Theorem XI) in order to satisfy the conditions the primary solutions must finally be chosen so that each one of any set composed of a primary solution and the solutions of the corresponding derived equations is orthogonal to all those of any other set. In fact the nature of the dependence of the primary solutions finally chosen on the ones selected at random is determined by making these sets mutually orthogonal. Solutions so chosen may be conceived of as the limit of separate solutions which have united and will be designated *proper primary* solutions. Orthogonality would naturally be expected between the various groups which form such limiting cases.

When the index of λ_i is m but the solutions have not all the same multiplicity, some adjustment of the linear dependence of the primary solutions on the set chosen first at random must again be made in order to insure that solutions of the successive derived equations be orthogonal. If one primary solution is to have a greater multiplicity than any of the others, it will necessarily be unique. This proper primary solution will then be entirely determined by the adjustments necessary to obtain solutions of the derived equations corresponding to it. On the other hand the multiplicity of any solution is not lowered by adding to it a solution of greater multiplicity and this principle may be used in adjusting the orthogonalization of the former in making it a proper primary.

Let us proceed now to develop the relations between the various sets of solutions X and Y analogous to those obtained in the regular case in §§ 2-3.

THEOREM VIII. *The totality n of solutions consisting of the primary solutions of (13) or of (14) together with the solutions of the derived equations are linearly independent.*

Following the method of proof of Theorem IV let us again assume the matrix composed of the X 's to be of rank $n-m$. There are then constants c_1, \dots, c_n such that

$$(29) \quad c_1 x_j^{(1)} + \dots + c_n x_j^{(n)} = 0,$$

and of these constants m may be assigned at random. The equations (26')-(28') may for any particular λ_k be written in the form

$$(30) \quad \sum_j a_{ij} x_j^{(k)} = \lambda_k \sum_j b_{ij} x_j^{(k)} + \epsilon_k \sum_j b_{ij} x_j^{(k-1)},$$

where the λ_k are n in number, one for each solution whether primary or of derived equations, and where $\epsilon_k = 0$ for a primary solution and $= 1$ for a solution of a derived equation. Proceeding now as in the proof referred to, we are led to the following equations analogous to (23):

$$0 = \sum_{kj} b_{ij} c_k [\lambda_k x_j^{(k)} + \epsilon_k x_j^{(k-1)}].$$

From them we deduce analogous to (24) the relations

$$\sum_k c_k [\lambda_k x_j^{(k)} + \epsilon_k x_j^{(k-1)}] = 0.$$

We have here and in (29) two essentially different relations between the solutions unless all the λ 's are equal and all ϵ 's zero. But in this event the index is n and each solution-multiplicity is unity and we have seen that the solutions can then be chosen linearly independent. We are thus in every case led to a contradiction and the theorem is established.

To determine the relations between the various primary solutions and solutions of the derived equations, it is desirable to generalize formulas (20) and (21) to cover the new types of solutions. In the first place it may be observed that precisely these same formulas hold if i and j both designate primary solutions corresponding to distinct λ 's. More generally we have by multiplication of the derived equations

$$\sum_j (a_{ij} - \lambda_i b_{ij}) x_j^{(l)} = \sum_j b_{ij} x_j^{(l-1)}$$

by $y^{(j)}$ and addition,

$$(31) \quad A(l, k) - \lambda_l B(l, k) = B(l-1, k);$$

and on multiplication of the derived equations

$$\sum_i (a_{ij} - \lambda_k b_{ij}) y_i^{(k)} = \sum_i b_{ij} y_i^{(k-1)}$$

by $x^{(i)}$ and addition,

$$(32) \quad A(l, k) - \lambda_k B(l, k) = B(l, k-1).$$

The terms on the right of (31) and (32) must be interpreted to be identically zero provided l, k respectively designate primary solutions.

THEOREM IX. *If l, k correspond to distinct characteristic numbers, $B(l, k) = 0$.*

For, this being the case when $l = 1$ and $k = m_p + 1$ designate primary solutions, it follows from (31) and (32) in a manner analogous to that

used in the proof of Theorem I that it is true for $l = 2, 3, \dots$, etc.; and by another application of the same argument, that it is true for $k = m_p + 2, m_p + 3, \dots$; and hence for all corresponding solutions of derived equations.

It is next in order to discover the relations between the primary solutions and solutions of derived equations, corresponding to the same characteristic number λ_1 , whose index is unity and solution-multiplicity p . Whether or not there are other solutions, there is a group of p^2 terms $B(l, k)$ ($l, k = 1, \dots, p$) among which certain equalities exist and certain of which will be zero. From (31) and (32) we note that $B(l-1, k) = B(l, k-1)$ and hence those terms in the group which occur in any one of the cross diagonals are equal. Further since $B(l, 0) \equiv 0$, we have from this same equality $0 = B(2, 0) = B(1, 1)$ and similarly $B(2, 1) = 0, \dots, B(p-1, 1) = 0$. On the other hand $B(p, 1)$ cannot be zero; for, since each solution in this group is orthogonal to those outside, it may be shown as in Theorem 2 that if this were the case, the solutions $X^{(1)}, \dots, X^{(p)}$ would be linearly dependent, which is contrary to Theorem VIII. Hence $B(p, 1)$ may be set equal to 1 and from (31) it follows that $A(p, 1) = \lambda_1 B(p, 1) = \lambda_1$. It is then possible to state the following

THEOREM X. *In the group of p^2 terms $B(l, k)$ corresponding to a characteristic number λ_1 of index 1 and solution-multiplicity p , terms in any cross diagonal are equal, all terms above the main cross diagonal are zero, and by normalization each term in the main cross diagonal takes the form $\lambda_1 - \lambda$.*

In discussing the relations between the solutions corresponding to a λ whose index is greater than unity, let us, for the sake of definiteness, denote a set of primary solutions by $X^{(1)}, Y^{(1)}$ and the solutions of their corresponding derived equations by $X^{(2)}, Y^{(2)}; \dots; X^{(m_1)}, Y^{(m_1)}$ and a second set of primary solutions by $X^{(m_1+1)}, Y^{(m_1+1)}$ and the corresponding solutions of the derived equations by $X^{(m_1+2)}, Y^{(m_1+2)}; \dots; X^{(m_1+m_2)}, Y^{(m_1+m_2)}$ and let $m_1 \leq m_2$. These primary sets when properly determined may be conceived of as respectively m_1 and m_2 -tuple solutions of the linear equations (13) for the same λ .

From (31) and (32) we derive immediately

THEOREM XI. *If l is a value from one of the following sets and k from the other, $1, \dots, m_1; m_1 + 2, \dots, m_1 + m_2$, then $B(l-1, k) = B(l, k-1)$, provided that if l or k is 1 these values of the B 's are understood to be zero; and further $B(l-1, m_1+1) = 0, B(m_1+1, k-1) = 0$.*

COROLLARY. *If these $B(l, k)$ are written down in m_1 rows and m_2 columns, then all terms to the left and above the cross diagonal passing through the upper right hand corner are zero.*

The formulas in Theorems IX–XI automatically hold whatever be the primary solutions (whether proper or not). That is, they are valid no matter what (linearly independent) solutions of (13) are chosen to start with. Proper primary solutions are, however, the limits of solutions which can be regarded as having been separated by a change of one of the coefficients and for such primary solutions and corresponding solutions of derived equations much more drastic conditions are valid. These relations will now be embodied in

THEOREM XII. *If l is one of the numbers from the two sets $1, \dots, m_1; m_1 + 1, \dots, m_1 + m_2$ which correspond to two proper primary solutions and the solutions of the derived equations for the same characteristic number, and k one of the numbers from the other set, then*

$$B(l, k) = 0.$$

A part of this result has already been formulated in the corollary above. As typical of the parts still remaining unproved, let us show that $B(m_1 + m_2 - 1, m_1) = 0$. Since $Y^{(1)}$ is an m_1 -tuple solution, $B(m_1 + m_2 - 1, 1) = 0$ is an m_1 -tuple relation in y_i . Hence it may be shown by a method analogous to that used in the derivation of formulas (26–28) that

$$B\left(X^{(m_1+m_2-1)}, \frac{dY^{(1)}}{d\lambda}\right) \equiv B(m_1 + m_2 - 1, 2) = 0,$$

$$B(m_1 + m_2 - 1, 3) = 0, \quad \dots, \quad B(m_1 + m_2 - 1, m_1) = 0.$$

And in a similar way it may be shown that all the other B 's involving an X from one set and a Y from the other are zero.

6. SIMULTANEOUS REDUCTION OF TWO MATRICES. IRREGULAR CASE

Preparatory to a discussion of the general theory of the reduction of a pair of matrices in the irregular case, it is well to consider a couple of special examples. Let us first treat the case where there is but one characteristic number λ_1 , with multiplicity three. The problem is again that of determining matrices $\|X\|$, $\|Y\|$ so that the product matrix $\|Y\| \|A + \lambda B\| \|\bar{X}\|$ is of a normal type. For this purpose it is necessary to use the primary solutions $X^{(1)}$, $Y^{(1)}$ of the two sets each of three linear homogeneous equations which correspond to the matrix $\|A - \lambda B\|$, together with the solutions of the corresponding derived equations $X^{(2)}$, $X^{(3)}$; $Y^{(2)}$, $Y^{(3)}$. It is proposed to show that by such a multiplication the matrix may be reduced* to

* Cf. second footnote § 5.

$$(33) \quad \left\| \begin{array}{ccc} 0 & 0 & \lambda_1 - \lambda \\ 0 & \lambda_1 - \lambda & 1 \\ \lambda_1 - \lambda & 1 & 0 \end{array} \right\|.$$

In the first place it may be noted from Theorem X that terms in and above the main cross diagonals are as given in (33) and further that it is necessary to discuss only $B(2, 3)$ and $B(3, 3)$. If it can be shown that by adjustment of the solutions of the derived equations these two may be equated to zero, it follows that the product matrix has the desired form. For, by (31) it would follow that $A(2, 3) = B(1, 3) = 1$ and $A(3, 3) = \lambda_1 B(3, 3) = 0$.

Denoting by $X^{(1)} = \Xi^{(1)}$ the primary solution and by $\Xi^{(2)}$ and $\Xi^{(3)}$ the solutions of the derived equations, as discussed in § 5, the problem may be put in the form of determining constants α and β such that

$$X^{(2)} = \Xi^{(2)} + \alpha \Xi^{(1)}, \quad X^{(3)} = \Xi^{(3)} + \beta \Xi^{(1)}$$

satisfy the relations $B(2, 3) = 0$, $B(3, 3) = 0$ or in other words

$$B(\Xi^{(2)}, Y^{(3)}) + \alpha B(\Xi^{(1)}, Y^{(3)}) = 0, \quad B(\Xi^{(3)}, Y^{(3)}) + \beta B(\Xi^{(1)}, Y^{(3)}) = 0,$$

and since, from Theorem X,

$$B(\Xi^{(1)}, Y^{(3)}) = B(\Xi^{(3)}, Y^{(1)}) = 1,$$

this is readily done. It may be observed that this procedure changes the values of none of the terms in the main cross diagonal or above and hence the result is established.*

* From the relation

$$\left\| \begin{array}{ccc} 1 & 0 & 0 \\ \frac{1}{\lambda_1 - \lambda_2} & 1 & 0 \\ \frac{1}{(\lambda_1 - \lambda_3)(\lambda_1 - \lambda_2)} & \frac{1}{\lambda_1 - \lambda_3} & 1 \end{array} \right\| \cdot \left\| \begin{array}{ccc} \lambda_1 - \lambda & 0 & 0 \\ 0 & \lambda_2 - \lambda & 0 \\ 0 & 0 & \lambda_3 - \lambda \end{array} \right\| \cdot \left\| \begin{array}{ccc} 0 & 0 & 1 \\ 0 & 1 & \frac{1}{\lambda_2 - \lambda_1} \\ 1 & \frac{1}{\lambda_3 - \lambda_2} & \frac{1}{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)} \end{array} \right\|$$

$$= \left\| \begin{array}{ccc} 0 & 0 & \lambda_1 - \lambda \\ 0 & \lambda_2 - \lambda & 1 \\ \lambda_3 - \lambda & 1 & 0 \end{array} \right\|,$$

we note that a normal form such as that obtained in Theorem VI may be transformed into one similar to (33) where the λ -roots are distinct. The new shape might be taken as

If there were other characteristic numbers each with the index one, precisely the same procedure would reduce the corresponding square block of matrix to a form similar to (33), each of which blocks in succession would be placed below and to the right of the former ones, while by Theorem XII all the remaining blocks of terms consist of zeros throughout.

The method of procedure is, however, somewhat different when the index of a λ is greater than unity and less than the λ -multiplicity. As a typical example let us consider the case where index is equal to 2 and λ -multiplicity to 5. One of the corresponding solution-multiplicities must then be four or three; which of these it is can be determined by setting up the derived equations and solving. If the solution is a three-fold only, the third derived equations will be found to have no possible solution. Since the case of a three-fold solution and an accompanying two-fold presents more difficulty, let us discuss it here. We shall show that the matrices $\|X\|$ and $\|Y\|$ of primary solutions and solutions of derived equations can be so chosen that the product takes the form

$$(34) \quad \begin{vmatrix} 0 & 0 & \lambda_1 - \lambda & 0 & 0 \\ 0 & \lambda_1 - \lambda & 1 & 0 & 0 \\ \lambda_1 - \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1 - \lambda \\ 0 & 0 & 0 & \lambda_1 - \lambda & 1 \end{vmatrix}.$$

The groups of 2 by 3 terms in the upper right and lower left corners will be zero when the primary solutions are properly chosen (Theorem XII). In each of the other groups the terms in and above the main cross diagonal will be after normalization as they appear in (34). By adding a constant times the primary solution to each of the corresponding solutions of the derived equations the other terms may be adjusted as in the first problem

normal for the regular case, but it is neither as elegant nor so simple to handle as that chosen. However, on passing to the limit and making the roots equal, this becomes the normal shape for this irregular case and is from this standpoint of considerable interest.

It is worthy of note that the multiplying matrices which correspond to these transformations of the x 's and y 's can be regarded as solutions of linear equations connected with the original form. These equations bear a marked resemblance to the primary and derived equations necessary to obtain the matrices $\|X\|$ and $\|Y\|$ used in the reduction to the shape (33). For example, to obtain the second row of the first matrix the equations have the form

$$\sum_i (a_{ij} - \lambda_2 b_{ij}) y_i^{(2)} = \sum_i b_{ij} y_i^{(1)}$$

where the $y_i^{(1)}$ on the right are constants which are solutions of the primary equations for $\lambda = \lambda_2$.

discussed above. Such adjustment does not alter in any way the terms of (34) already determined. The reduction to the form (34) is then completed.

We may now formulate the following

THEOREM XIII. *By means of linear transformations of the x 's and y 's whose coefficients are respectively the n^2 constants made up of: (1) properly chosen primary solutions of the equations (13) and (14) associated with the bilinear forms $A(x, y)$, $B(x, y)$; and (2) the corresponding solutions of the derived equations, reduction takes place as follows:*

$$\begin{aligned} A(x, y) &\equiv \left(\sum_1^{m_1} \lambda_1 x_{m_1-i+1} y_i + \sum_1^{m_1-1} x_{m_1-i+1} y_{i+1} \right) \\ &\quad + \left(\sum_{m_1+1}^{m_1+m_2} \lambda_2 x_{2m_1+m_2-i+1} y_i + \sum_{m_1+1}^{m_1+m_2-1} x_{2m_1+m_2-i+1} y_{i+1} \right) + \cdots \\ &\quad + \left(\sum_{n-m_p+1}^n \lambda_p x_{2n-m_p-i+1} y_i + \sum_{n-m_p+1}^{n-1} x_{2n-m_p-i+1} y_{i+1} \right), \\ B(x, y) &\equiv \sum_1^{m_1} x_{m_1-i+1} y_i + \sum_{m_1+1}^{m_1+m_2} x_{2m_1+m_2-i+1} y_i + \cdots \\ &\quad + \sum_{n-m_p+1}^n x_{2n-m_p-i+1} y_i, \end{aligned}$$

where $\lambda_1, \dots, \lambda_p$ are characteristic numbers, each being repeated a number of times equal to its index, and where the solution-multiplicities are m_1, \dots, m_p respectively ($m_1 + m_2 + \cdots + m_p = n$).

It is now but a step to a solution of the problem of equivalence of pairs of bilinear forms.

THEOREM XIV. *In order that two pairs of bilinear forms $A_1(x, y)$, $B_1(x, y)$ and $A_2(x, y)$, $B_2(x, y)$ be equivalent it is necessary and sufficient that the λ -polynomials be identical and that the corresponding indexes and the corresponding solution-multiplicities be the same.*

7. REDUCTION OF PAIRS OF QUADRATIC AND HERMITIAN FORMS

It may be well to note some of the modifications of the treatment of §§ 1—6 which are necessary in order to specialize our theory for the cases of quadratic and hermitian forms.

In the case of quadratic forms the matrix $A - \lambda B$ is symmetric and the two sets of linear equations associated with it and corresponding to (13) and (14) are identical. When each of the solution-multiplicities is unity, the two sets of solutions (16) and (16') are identical and the discussion proceeds by replacing the y 's by the x 's; but when the solution-multiplicity

is greater than unity the solutions of the derived equations are not unique, and we have seen fit in some cases to leave an originally chosen set of y 's fixed and manipulate the x 's to satisfy the prescribed conditions. However, the methods used can be modified without much difficulty to cover the case of the quadratic forms. As an example of the only type of modification necessary, let us prove Theorem V for solutions associated with quadratic forms.

Denoting as before by $X^{(k_1)}$, $X^{(k_2)}$, $X^{(k_3)}$ the linearly independent solutions originally chosen, it is possible first to show that constants α , β , γ can be determined so that

$$X^{(1)} \equiv \alpha X^{(k_1)} + \beta X^{(k_2)} + \gamma X^{(k_3)}$$

is normalized. For, if not, $B(1, 1)$ would vanish identically in α , β , γ ; and this involves

$$B(k_1, k_1) = B(k_1, k_2) = B(k_1, k_3) = 0,$$

which is a contradiction of Theorem IV.

Let $X^{(k')}$ and $X^{(k'')}$ denote a second and third set which form with $X^{(1)}$ linearly independent solutions of (13) for this value of λ . If $X^{(k')}$ is not already orthogonal to $X^{(1)}$ it may be made so by adding to it $\alpha X^{(1)}$. For, since

$$B(X^{(k')} + \alpha X^{(1)}, X^{(1)}) \equiv B(k', 1) + \alpha B(1, 1),$$

it may be made zero by a proper choice of α . Similarly $X^{(k'')}$, if not already orthogonal to $X^{(1)}$, may be made so by adding to it $\beta X^{(1)}$. The solutions $X^{(1)}$, $X^{(k')} + \alpha X^{(1)}$, $X^{(k'')} + \beta X^{(1)}$ satisfy the conditions of linear independence, for any linear relation between them is also a linear relation between $X^{(1)}$, $X^{(k')}$, $X^{(k'')}$, which were assumed linearly independent.

Let us proceed now with these two solutions which have been built up orthogonal to (1). Any linear combination of them will be orthogonal to $X^{(1)}$ and we can so determine the combination that a normalized solution results. For if not, we would have an identical relation in the multipliers which gives $B(k', k') = 0$, $B(k', k'') = 0$ and from Theorem IV we note that this is not compatible with $B(k', 1) = 0$. Denoting by $X^{(2)}$ this second solution and by $X^{(3)}$ a third solution, linearly independent of $X^{(1)}$, $X^{(2)}$, we can determine α , β , γ so that $X^{(3)} \equiv \alpha X^{(1)} + \beta X^{(2)} + \gamma X^{(3)}$ is normalized and orthogonal to $X^{(1)}$ and $X^{(2)}$. For, the equations involved reduce to

$$\begin{aligned} \alpha B(1, 1) + \beta B(1, 2) + \gamma B(1, l) &= 0, \\ \alpha B(2, 1) + \beta B(2, 2) + \gamma B(2, l) &= 0, \\ \alpha B(l, 1) + \beta B(l, 2) + \gamma B(l, l) &= \frac{1}{\gamma}, \end{aligned}$$

and in solving for α , β , γ the denominator determinant consisting of the coefficients on the left may be proved different from zero, in a manner analogous to that used with (25). Since $B(1, 1) = B(2, 2) \neq 1$, $B(1, 2) = B(2, 1) = 0$, solution for γ has for numerator $1/\gamma$ and hence $\gamma^2 = c \neq 0$, and from this α and β are readily found.

THEOREM XV. *If the form $B(x, x)$ is definite, each solution-multiplicity is unity.*

The λ -determinant in this case has real roots only and (13) has real solutions. In order that the solution-multiplicity be greater than unity we have seen in § 5 that it is necessary and sufficient that $B(k, k) = 0$, where k denotes one of the primary solutions. But this is not possible for a real definite quadratic form.

As a simple example of the comparative elegance of this new theory of the reduction of quadratic forms, let us consider the reduction to normal type of a single quadratic form $A(x, x)$ in n variables and of rank m . In the literature this is done by discussing separately the elimination of the $n-m$ superfluous variables and the reduction of the resulting form by m distinct steps, each time simplifying the problem by separating out one variable. The procedure in each of these steps is different according as there does or does not exist a term in the main diagonal different from zero. On the other hand, from the new point of view let us take the auxiliary form $B(x, x) = \sum_1^n x_i^2$ and consider the pencil $A - \lambda B$. There will be n roots of the λ -determinant of which $n-m$ will be zero. The linear equations (13) have always real solutions of multiplicity unity and the matrix $\|X\|$ obtained by solving them represents an orthogonal transformation which reduces $A(x, x)$ to the form $\sum_1^m \lambda_i x_i^2$.

Let us now elaborate somewhat the developments which concern a pair of hermitian forms. These forms may be considered as a generalization of the real quadratic forms and perhaps their increasing importance in mathematical theory is due in part to the possibility of discussing their minimum properties (which we shall do in II).

Let $\|A\|$ and $\|B\|$ be the matrices of two linearly independent hermitian forms,

$$A(z, \bar{z}) = \sum a_{ij} z_i \bar{z}_j, \quad B(z, \bar{z}) = \sum b_{ij} z_i \bar{z}_j \quad (a_{ij} = \bar{a}_{ji}, \quad b_{ij} = \bar{b}_{ji}),$$

of which B is non-singular. For any real λ , $A + \lambda B$ is also an hermitian form and will take on real values only. The coefficients of λ in the λ -determinant will be real, and hence those characteristic numbers which are not real will occur in conjugate pairs. From the relation $A(k, k) = \lambda_k B(k, k)$ we note that the real hermitian forms must both be zero for a complex characteristic number. If B is definite or more generally

if each solution-multiplicity is unity, all the characteristic numbers must therefore be real. If the form $B(z, \bar{z})$ is definite, each solution-multiplicity is unity, as may be proved in a manner analogous to Theorem XV.

The reduction of pairs of hermitian forms to normal hermitian types proceeds as in §§ 4, 6 with small modifications of the same general character as those indicated above for quadratic forms. The types are not different from those obtained by other methods.* Some further comments concerning the problem may, however, be in order.

For any real λ the solutions of the equations corresponding to (13) will be conjugate to those corresponding to (14); where λ is complex the solution of (13) will be conjugate imaginary to those of (14) for $\bar{\lambda}$, and the same remarks hold for the derived equation, as may be seen by a study of (30) and its conjugate equation. Hence a complete set of solutions such as (16) will be matched with a conjugate set such as (16') but not always in the same order.

We note here, then, that the matrix after the usual multiplication by the matrices of primary and derived solutions may not be hermitian. When the solution-multiplicities are all unity the corresponding elements of the two matrices of solutions $x_i^{(k)}$ and $y_i^{(k)}$ will be conjugate imaginary and the product matrix will be hermitian. But if there are complex λ 's, the product matrix will not be hermitian unless the solutions are rearranged so as to make the corresponding solutions conjugate imaginary. This may, however, be done by interchanging in one of the matrices the two sets of solutions corresponding to λ and $\bar{\lambda}$. To exhibit a special example where the product matrix is made hermitian by such a process we write down the normal form for a matrix where there are two conjugate imaginary characteristic numbers $\lambda_1, \bar{\lambda}_1$, each of which has index one and solution-multiplicity two:

$$\left\| \begin{array}{cccc} 0 & 0 & 0 & \lambda_1 - \lambda \\ 0 & 0 & \lambda_1 - \lambda & 1 \\ 0 & \lambda_1 - \lambda & 0 & 0 \\ \bar{\lambda}_1 - \bar{\lambda} & 1 & 0 & 0 \end{array} \right\|.$$

* See Hilton, *Linear Substitutions*, p. 180. Also Logsdon, *Equivalence and reduction of pairs of hermitian forms*, American Journal of Mathematics, vol. 44, pp. 247-60.