ON DIRECT PRODUCT MATRICES[†]

BY W. E. ROTH

1. *Introduction.* If $A = (a_{ij})$, $(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$, is an $m \times n$ matrix and B is a $p \times q$ matrix, then the matrix, $P = (a_{ij}B) = A \langle B \rangle$, of order $mp \times nq$, whose elements occur in mn blocks, $a_{ij}B$, is the direct product of A and B.[‡]

In the present paper we determine the elementary divisors of $A \langle B \rangle - \lambda I$ and of $\rho A \langle I \rangle + \sigma I \langle B \rangle - \lambda I$, where ρ and σ are scalar constants, and where the elementary divisors of $A - \lambda I$ and of $B - \lambda I$ are known. Finally, §3 takes up the discussion of the linear matrix equation

$$A_1 X_1 B_1 + A_2 X_2 B_2 + \cdots + A_r X_r B_r = C.$$

The reduction of this equation to an equation whose solution is known is accomplished by means of direct product matrices and thus perfects a procedure first noted by MacDuffee.[§]

2. *On Elementary Divisors of* $A \langle B \rangle - \lambda I$ *and of* $\rho A \langle I \rangle + \sigma I \langle B \rangle - \lambda I$. In this section it will be convenient to indicate the order of a matrix by subscripts; thus $A_{\alpha,\beta}$ is an $\alpha \times \beta$ matrix, B_p is a square matrix of order p, and I_p is the unit matrix of order p. Matrices will be designated throughout by capitals, whereas lower case letters will be employed to denote scalars, such as parameters, constants, and the elements of matrices. Moreover, all scalars will be regarded as belonging to the complex number field. Hence, for our purpose, $(a_{ij}B) = (Ba_{ij}) = A \langle B \rangle = \langle B \rangle A$.

---

† Presented to the Society, April 6, 1934.

‡ Zehfuss, *Ueber eine gewisse Determinante*, Zeitschrift für Mathematik und Physik, 3te Jahrgang (1858), pp. 298–301, was perhaps the first to study determinants of this form. Rutherford, *On the condition that two Zehfuss matrices be equal*, this Bulletin, vol. 39 (1933), pp. 801–808, called P the Zehfuss matrix of A and B and devised the notation here employed. Dickson, *Algebras and their Arithmetics*, 1923, p. 119, and MacDuffee, *The Theory of Matrices*, 1933, p. 81, employ the term *direct product* to designate P. The reader is also referred to the latter treatise for a more complete discussion of direct product matrices.

§ MacDuffee, loc. cit., p. 89.

If $A_{\alpha,\beta} = (a_{ij})$, $C_{\beta,\gamma} = (c_{ij})$, $B_{a,b}$, and $D_{b,c}$ are given matrices, then

$$A_{\alpha,\beta}\langle B_{a,b}\rangle C_{\beta,\gamma}\langle D_{b,c}\rangle = (a_{ij}B_{a,b})(c_{jk}D_{b,c})$$

(1)
$$= \left(\sum_{j=1}^{\beta} a_{ij}c_{jk}B_{a,b}D_{b,c} \right) = A_{\alpha,\beta}C_{\beta,\gamma}\langle B_{a,b}D_{b,c}\rangle.$$

This identity was first proved by Stéphanos.[†] Evidently, if Q_m and R_n are non-singular matrices, then $Q_m\langle I_n\rangle$ and $I_m\langle R_n\rangle$ are likewise non-singular matrices. Hence, according to (1),

$$Q_m\langle I_n\rangle \cdot I_m\langle R_n\rangle = I_m\langle R_n\rangle \cdot Q_m\langle I_n\rangle = Q_m\langle R_n\rangle;$$

and $Q_m\langle R_n\rangle$, being the product of non-singular matrices, is also non-singular. If Q_m^I is the inverse of Q_m and R_n^I that of R_n,[‡] then $Q_m^I\langle R_n^I\rangle$ is the inverse of $Q_m\langle R_n\rangle$,[§] for by (1),

(2)
$$Q_m\langle R_n\rangle \cdot Q_m^I\langle R_n^I\rangle = I_m\langle I_n\rangle = I_{mn}.$$

Let Q_m and R_n be the non-singular matrices which transform A_m and B_n to their respective normal forms, A_m^* and B_n^*. That is,

(3)
$$Q_m A_m Q_m^I = A_m^*, \qquad R_n B_n R_n^I = B_n^*,$$

where

$$(4) \quad A_m^* = \begin{pmatrix} \mathcal{A}_{m_1} & 0 & \cdots & 0 \\ 0 & \mathcal{A}_{m_2} & \cdots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots} \\ 0 & 0 & \cdots & \mathcal{A}_{m_r} \end{pmatrix}, \quad B_n^* = \begin{pmatrix} \mathcal{B}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathcal{B}_{n_2} & \cdots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots} \\ 0 & 0 & \cdots & \mathcal{B}_{n_s} \end{pmatrix},$$

and

$$\mathcal{A}_{m_i} = \begin{pmatrix} a_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & a_i & 1 & \cdots & 0 & 0 \\ \multicolumn{6}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & 0 & \cdots & a_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & a_i \end{pmatrix}, \quad \mathcal{B}_{n_j} = \begin{pmatrix} b_j & 1 & 0 & \cdots & 0 & 0 \\ 0 & b_j & 1 & \cdots & 0 & 0 \\ \multicolumn{6}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & 0 & \cdots & b_j & 1 \\ 0 & 0 & 0 & \cdots & 0 & b_j \end{pmatrix},$$

---

† Stéphanos, *Sur une extension du calcul des substitutions linéaires*, Journal de Mathématiques, (5), vol. 6 (1900), pp. 73–128.

‡ This notation for inverse is due to MacDuffee, loc. cit.

§ Stéphanos, loc. cit.

$(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$. Consequently $A_m - \lambda I_m$ has the elementary divisors $(a_i - \lambda)^{m_i}$, $(i = 1, 2, \cdots, r)$, and $B_n - \lambda I_n$ has the elementary divisors $(b_j - \lambda)^{n_j}$, $(j = 1, 2, \cdots, s)$. We shall prove the following theorem.

THEOREM 1. *If* $A_m - \lambda I_m$ *and* $B_n - \lambda I_n$ *have the elementary divisors* $(a_i - \lambda)^{m_i}$, $(i = 1, 2, \cdots, r)$, *and* $(b_j - \lambda)^{n_j}$, $(j = 1, 2, \cdots, s)$, *respectively, and if* μ_{ij} *is the lesser of the integers* m_i *and* n_j *or their common value, then the matrix* $A_m \langle B_n \rangle - \lambda I_{mn}$ *has the elementary divisors*

$$(a_i b_j - \lambda)^{m_i + n_j - 2k + 1} \qquad (k = 1, 2, \cdots, \mu_{ij}),$$

if $a_i b_j \neq 0$, λ^{m_i} *occurring* n_j *times, if* $a_i = 0$, $b_j \neq 0$, λ^{n_j} *occurring* m_i *times, if* $a_i \neq 0$, $b_j = 0$, *and* λ, λ^2, $\cdots$, $\lambda^{\mu_{ij}-1}$ *each occurring twice and* $\lambda^{\mu_{ij}}$ *occurring* $m_i + n_j - 2\mu_{ij} + 1$ *times if* $a_i = b_j = 0$, $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$.

Now by (1), (2), and (3), we have

$$Q_m \langle R_n \rangle [A_m \langle B_n \rangle - \lambda I_{mn}] Q_m^I \langle R_n^I \rangle = A_m^* \langle B_n^* \rangle - \lambda I_{mn};$$

consequently the elementary divisors of $A_m \langle B_n \rangle - \lambda I_{mn}$ are identical with those of $A_m^* \langle B_n^* \rangle - \lambda I_{mn}$. Upon writing out the latter matrix, we readily see that its elementary divisors are identical with the aggregate of those of the r matrices

$$\mathscr{A}_{m_i} \langle B_n^* \rangle - \lambda I_{m_i n}, \qquad (i = 1, 2, \cdots, r),$$

each taken separately. By an appropriate interchange of rows and of columns the matrix $\mathscr{A}_{m_i} \langle B_n^* \rangle - \lambda I_{m_i n}$ can be transformed to an equivalent matrix having the $m_i n_j \times m_i n_j$ matrices $\mathscr{A}_{m_i} \langle B_{n_i} \rangle - \lambda I_{m_i n_j}$, $(j = 1, 2, \cdots, s)$, in its principal diagonal and zero matrices elsewhere. Hence the elementary divisors of $A_m \langle B_n \rangle - \lambda I_{mn}$ are identical with the aggregate of those of the rs matrices

$$\mathscr{A}_{m_i} \langle B_{n_j} \rangle - \lambda I_{m_i n_j}, \qquad (i = 1, 2, \cdots r; j = 1 \; 2, \cdots, s),$$

each taken independently.

If $b_j \neq 0$, $\mathscr{B}_{n_j}$ is a non-singular matrix of order n_j; consequently, by the method of Bôcher,[†]

$$(5) \qquad \mathscr{A}_{m_i} \langle \mathscr{B}_{n_j} \rangle - \lambda I_{m_i n_j}$$

[†] Bôcher, *Introduction to Higher Algebra*, 1922, §§91 and 92.

has $(m_i-1)n_j$ unit invariant factors and its elementary divisors in λ are identical with those of $(a_i\mathcal{B}_{n_j}-\lambda I_{n_j})^{m_i}$. If we let μ_{ij} be the lesser of the two integers m_i and n_j or their common value, this matrix has the elementary divisors $(a_ib_j-\lambda)^{m_i+n_j-2k+1}$, $(k=1, 2, \cdots, \mu_{ij})$, if $a_i\neq 0$, and the elementary divisor λ^{m_i} occurring n_j times, if $a_i=0$. This establishes the first two cases under the above theorem.

If $b_j=0$ and $a_i\neq 0, \mathcal{A}_{m_i}\langle\mathcal{B}_{n_j}\rangle$ is a nilpotent matrix such that its n_jth power is zero but no lower power vanishes. Hence, in this case, (5) has only elementary divisors of the form λ^h, where h cannot exceed n_j. But since $a_i\neq 0$, (5) has a non-singular minor of order $(n_j-1)m_i$ which is independent of λ; hence at most m_i elementary divisors can be powers of λ. The determinant of (5) is $\lambda^{m_in_j}$, consequently the m_i elementary divisors must be λ^{n_j}. This establishes the third case under the theorem.

If $a_i=b_j=0, \mathcal{A}_{m_i}\langle\mathcal{B}_{n_j}\rangle$ has $(m_i-1)(n_j-1)$ unit elements in the (n_j+1)th diagonal above its principal diagonal and all remaining elements are zeros. Hence (5) has a non-singular minor of order $(m_i-1)(n_j-1)$ which is independent of λ, and no such minor of higher order. Consequently this matrix has $m_in_j-(m_i-1)(n_j-1)=m_i+n_j-1$ elementary divisors which are powers of λ. Now $\mathcal{A}_{m_i}\langle\mathcal{B}_{n_j}\rangle$ is nilpotent of degree μ_{ij}, for in the present case the μ_{ij}th power of either $\mathcal{A}_{m_i}$ or of $\mathcal{B}_{n_j}$ vanishes but neither vanishes for powers less than μ_{ij}. Consequently (5) has m_i+n_j-1 elementary divisors λ^k, whose degree k does not exceed μ_{ij} and must have at least one of degree μ_{ij}. We may, therefore, assume that (5) has the elementary divisors λ^k occurring ν_k times, $(k=1, 2, \cdots, \mu_{ij})$. Hence its determinant is $-\lambda$ raised to the power,

$$(6) \qquad \nu_1 + 2\nu_2 + \cdots + \mu_{ij}\nu_{\mu_{ij}} = m_im_j.$$

Now let the normal form of $\mathcal{A}_{m_i}\langle\mathcal{B}_{n_j}\rangle$ be P, where P has the form corresponding to that for A_m and B_n as given by (4). Then P^k has

$$\nu_{k+1} + 2\nu_{k+2} + \cdots + (\mu_{ij} - k)\nu_{\mu_{ij}}$$

unities in the kth diagonal above its principal diagonal and zeros elsewhere, and $[\mathcal{A}_{m_i}\langle\mathcal{B}_{n_j}\rangle]^k$ has $(m_i-k)(n_j-k)$ unities in the $k(n_j+1)$th diagonal above its principal diagonal and

zeros elsewhere. Since these two matrices are of the same rank for all values of k, we have

(7) $\quad \nu_{k+1} + 2\nu_{k+2} + \cdots + (\mu_{ij} - k)\nu_{\mu_{ij}} = (m_i - k)(n_j - k),$

$(k = 1, 2, \cdots, \mu_{ij} - 1)$. The μ_{ij} equations (6) and (7) have the solution $\nu_k = 2$, $(k = 1, 2, \cdots, \mu_{ij} - 1)$, and $\nu_{\mu_{ij}} = m_i + n_j - 2\mu_{ij} + 1$. Thus, in case $a_i = b_j = 0$, (5) has the $m_i + n_j - 1$ elementary divisors, stated in the above theorem.

COROLLARY. *Under the hypotheses of the theorem above $A_m\langle I_p\rangle - \lambda I_{mp}$ has the elementary divisors $(a_i - \lambda)^{m_i}$, $(i = 1, 2, \cdots, r)$, each occurring p times, and $I_p\langle B_n\rangle - \lambda I_{np}$ has the elementary divisors $(b_j - \lambda)^{n_j}$, $(j = 1, 2, \cdots, s)$, each occurring p times.*

This corollary follows directly from Theorem 1, for in this case, the matrix $I_p - \lambda I_p$ has the elementary divisor $1 - \lambda$ occurring p times, hence $\mu_{ij} = 1$, $(j = 1, 2, \cdots, p)$, in the present case. Of course, this corollary is obviously true without recourse to the above theorem.

THEOREM 2. *Under the hypotheses of Theorem 1, the matrix*

(8) $\qquad\qquad \rho A_m\langle I_n\rangle + \sigma I_n\langle B_n\rangle - \lambda I_{mn},$

where ρ and σ are scalar constants not zero, has the elementary divisors $(\rho a_i + \sigma b_j - \lambda)^{m_i + n_j - 2k + 1}$, $(k = 1, 2, \cdots, \mu_{ij}; i = 1, 2, \cdots, r; j = 1, 2, \cdots, s)$.

By an argument parallel to that followed in the proof of Theorem 1 we can show that the elementary divisors of $\rho A_m\langle I_n\rangle + \sigma I_m\langle B_n\rangle - \lambda I_{mn}$ are identical with the aggregate of those of the rs matrices

$$[\rho a_i I_{n_j} + \sigma B_{n_j} - \lambda I_{n_j}]^{m_i}, \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s).$$

Hence the elementary divisors of (8) are those given by Theorem 2, if neither ρ nor σ is zero. The case where ρ or σ is zero is covered by the Corollary above.†

---

† Recently the author learned that A. C. Aitken, in a paper that will appear soon in the Proceedings of the London Mathematical Society, gives results quite closely analogous to those of Theorem 1.

3. *The Linear Matrix Equation.* We shall take up briefly the linear matrix equation

(9) $$\sum_{i=1}^{r} A_i X_i B_i = C,$$

where A_i are $m \times n_i$ matrices; B_i, $(i = 1, 2, \cdots, r)$, are $n_i \times n$ matrices; and C is an $m \times n$ matrix, all known; while X_i, $(i = 1, 2, \cdots, r)$, are unknown $m_i \times m_i$ matrices.† We here transform this equation to an equivalent equation by means of the matrix M^R and direct product matrices, where M^R is defined as follows: Let $M = (m_{ij})$ be an $\alpha \times \beta$ matrix and let

$$M_i = (m_{i1}, m_{i2}, \cdots, m_{i\beta}), \qquad (i = 1, 2, \cdots, \alpha);$$

then $M^R = (M_1, M_2, \cdots, M_\alpha)$, and is consequently a one-rowed matrix of $\alpha\beta$ elements. Similarly M^C is the matrix resulting if the second, third, $\cdots$, last column of M be written each below the next preceding column; thus it is an $\alpha\beta \times 1$ matrix. We can then show that $M^{CR} = M^{TR}$, $M^{RT} = M^{TC}$, where M^{CT} is the transpose of M^C, etc. Let

(10) $$Y_i = A_i X_i, \qquad Z_i = Y_i B_i, \qquad (i = 1, 2, \cdots, r);$$

then $Z_i = A_i X_i B_i$. We can readily show from (10) that

$$Y_i^R = X_i^R [A_i^T \langle I_{n_i} \rangle], \qquad Z_i^R = Y_i^R [I_m \langle B_i \rangle],$$

where A_i^T is the transpose of A_i. Hence, by (1),

$$Z_i^R = X_i^R [A_i^T \langle I_{n_i} \rangle] \cdot [I_m \langle B_i \rangle] = X_i^R [A_i^T \langle B_i \rangle],$$

where $A_i^T \langle B_i \rangle$ is then an $m_i n_i \times mn$ matrix. With this transformation, equation (9) becomes $\sum_{i=1}^{r} X_i^R A_i^T \langle B_i \rangle = C^R$, or

(11) $$(X_1^R, X_2^R, \cdots, X_r^R) \begin{pmatrix} A_1^T \langle B_1 \rangle \\ A_2^T \langle B_2 \rangle \\ \vdots \\ A_r^T \langle B_r \rangle \end{pmatrix} = C^R.$$

That is, (9) is thereby transformed to an equivalent equation whose theory is well known, for it is a system of mn linear equa-

---

† MacDuffee, loc. cit., discusses this equation for the case where the A_i and B_i are $n \times n$ matrices and gives adequate references. He extends the theory in the direction we take here, but introduces an error which will be explained below.

tions in $\sum_{i=1}^{r} m_i n_i$ unknowns and has a solution provided the matrix of the coefficients and of this matrix augmented by C^R have the same rank.† Hence we may state the following result.

THEOREM 3. *A necessary and sufficient condition that the equation* (9) *where* A_i *are* $m \times m_i$, B_i *are* $n_i \times n$ *matrices, and* C *is an* $m \times n$ *matrix, all known, while* X_i, $(i = 1, 2, \cdots, r)$ *are unknown* $m_i \times n_i$ *matrices, have a solution, is that the matrices*

$$P = \begin{pmatrix} A_1^T \langle B_1 \rangle \\ A_2^T \langle B_2 \rangle \\ \cdot \\ \cdot \\ \cdot \\ A_r^T \langle B_r \rangle \end{pmatrix} \quad and \quad \begin{pmatrix} C^R \\ P \end{pmatrix}, \quad C \neq 0,$$

have the same rank, and in case $C = 0$, *a necessary and sufficient condition that* (9) *have a solution,* X_i, $(i = 1, 2, \cdots, r)$, *not all zero is that the rank of* P *be less than* $\sum_{i=1}^{r} m_i n_i$.

In case $X_1 = X_i$, $(i = 2, 3, \cdots, r)$, (11) becomes

$$(12) \qquad X_1^R \left[\sum_{i=1}^{r} A_i^T \langle B_i \rangle \right] = C^R,$$

which is a system of mn linear equations in the $m_1 n_1$ unknown elements of X_1. This leads to the following theorem.

THEOREM 4. *A necessary and sufficient condition that the equation*

$$(13) \qquad \sum_{i=1}^{r} A_i X B_i = C,$$

where A_i *are* $m \times n$ *matrices,* B_i *are* $p \times q$ *matrices, and* C *is an* $m \times q$ *matrix, and where* X *is an unknown* $n \times p$ *matrix, have a solution, is that the matrices*

$$Q = \sum_{i=1}^{r} A_i^T \langle B_i \rangle \quad and \quad \begin{pmatrix} C^R \\ Q \end{pmatrix}, \quad C \neq 0,$$

have the same rank; and in case $C = 0$, *a necessary and sufficient condition that the equation have an solution* $X \neq 0$ *is that* Q *be of rank less than* np.

---

† Bôcher, loc. cit.

Equation (12) may be replaced by the equation

$$\left[\sum_{i=1}^{r} \langle A_i \rangle B_i^T \right] X^C = C^C,$$

or the transpose of (12), or that of the present equation. Hence any one of the four matrices,

$$(14) \quad \sum_{i=1}^{r} A_i^T \langle B_i \rangle, \quad \sum_{i=1}^{r} \langle A_i \rangle B_i^T, \quad \sum_{i=1}^{r} A_i \langle B_i^T \rangle, \quad \sum_{i=1}^{r} \langle A_i^T \rangle B_i,$$

may be regarded as that of the coefficients of the elements of X, that is, as Sylvester's nivellateur.† It should be noted that a different arrangement of the elements of X from that given by either X^R or X^C or their transposes will result in another matrix which can be obtained from one of the above by an appropriate interchange of rows or of columns.‡

COROLLARY. *If* $A_1B_1 + A_2B_2 + \cdots + A_rB_r = 0$, *where* A_i' $(i = 1, 2, \cdots, r)$, *are* $m \times n$ *matrices and* B_i *are* $n \times p$ *matrices' then each of the matrices*

$$\sum_{i=1}^{r} A_i^T \langle B_i \rangle, \quad \sum_{i=1}^{r} \langle A_i \rangle B_i$$

is at most of rank $n^2 - 1$.

This corollary is a direct result of Theorem 4, for here $X = I_n$ satisfies the equation (13). It states, perhaps, the most that may be said of such an equation if the B's are not commutative. Phillips§ gives a more restrictive result in case the B's are commutative.

THE UNIVERSITY OF WISCONSIN,
    EXTENSION CENTER, MILWAUKEE, WISCONSIN

---

† Sylvester, Comptes Rendus, vol. 49 (1884), pp. 409–412, 432–436; Mathematical Papers, vol. 4, pp. 199–205.

‡ MacDuffee, loc. cit., was perhaps the first to recognize Sylvester's nivellateur as the sum of direct product matrices, but the matrix 46.2 as he gave it cannot answer the purpose and should be replaced by one of the four matrices in (14). Likewise in his theorem 46.31, the Sylvester nivellateur should be $A\langle I \rangle + I \langle B^T \rangle$ or $\langle A^T \rangle I + \langle I \rangle B$ or the transpose of either of these.

§ Phillips, *Functions of matrices*, American Journal of Mathematics, vol. 41 (1919), pp. 266–278.