SIAM REVIEW
Vol. 12, No. 4, October 1970

EXPLICIT SOLUTIONS OF LINEAR MATRIX EQUATIONS*

PETER LANCASTER†

1. Introduction. Let $A_1, A_2, \cdots, A_k$ be $m \times m$ complex matrices (members of $C_{m \times m}$) and $B_1, B_2, \cdots, B_k \in C_{n \times n}$. In this paper we review some techniques for obtaining explicit representations for a solution matrix $X \in C_{m \times n}$ for equations of the form

$$A_1XB_1 + A_2XB_2 + \cdots + A_kXB_k = C, \tag{1}$$

where $C \in C_{m \times n}$ is given, and such a solution is known to exist. Certain special cases of this equation arise frequently in applications and will be considered in some detail. The problem in its full generality is far from tractable, although the transformation to a matrix-vector equation, which we shall discuss in a moment, allows us to use the considerable arsenal of numerical weapons currently available for the solution of such problems.

Our objective is to bring together some results concerning solutions of our equation and present them in a connected treatment which in parts allows us to improve the theory and in others allows us to make simplifications. In particular, we wish to present and make available to a wider audience some of the elegant results of Krein [7]. By presenting these in a matrix context we can significantly abbreviate some of the theory.

2. The equivalent matrix-vector equation. If $A \in C_{m \times m}$ and $B \in C_{n \times n}$, then the *direct product* (or tensor product) of A and B, written $A \otimes B$, is defined to be the partitioned matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix} \in C_{mn \times mn}.$$

It must be assumed that the reader has some familiarity with this concept. An account of the uses and applications of this operation can be found in [8, Chap. 8] or in MacDuffee [10]. With any matrix D of $C_{m \times n}$ we associate a column vector $\mathbf{d} \in C_{mn}$ whose elements are just those of D written out in a particular order. We start with d_{11}, work across the first row, then the second, and so on, down to the mth row when all the elements of D have been exhausted.

With these definitions it is a routine matter to verify that if $G = A_1 \otimes B'_1 + \cdots + A_k \otimes B'_k$ (the prime denotes transposition), then (1) is equivalent to

$$G\mathbf{x} = \mathbf{c}. \tag{2}$$

In fact, by viewing (1) as a set of mn scalar equations for the elements of X, (2) is precisely the same set of equations written out in a different way. However, in

* Received by the editors August 22, 1969, and in revised form February 16, 1970.

† Department of Mathematics, Faculty of Arts and Sciences, University of Calgary, Calgary 44, Alberta, Canada. This work was supported by the National Research Council of Canada.

order to make pronouncements about existence, uniqueness and techniques for the solution of (2), we generally need some information about the eigenvalues of G. We call the set of complex numbers which are eigenvalues of a matrix M the *spectrum* of M and denote this set by $\sigma(M)$.

In the general case it is difficult to say anything very useful about $\sigma(G)$, even if we know $\sigma(A_j)$ and $\sigma(B_j)$ for $j = 1, 2, \cdots, k$. Indeed, we cannot say when $0 \in \sigma(G)$; that is, when G will be singular. Our first special case, then, is one which still admits some generality, but allows us to determine $\sigma(G)$ in terms of $\sigma(A_j)$ and $\sigma(B_j)$, $j = 1, 2, \cdots, k$. We consider the case in which each A_j is a scalar polynomial in a fixed matrix $A \in C_{m \times m}$ and each B_j is a scalar polynomial in a fixed $B \in C_{n \times n}$. Equation (1) then takes the form

$$\sum_{j,k=0}^{l} \alpha_{jk} A^j X B^k = C, \tag{3}$$

where the α_{jk} are complex numbers. If we define the polynomial in two scalar variables: $\phi(x, y) = \sum \alpha_{jk} x^j y^k$, then we define the polynomial ϕ with matrix arguments by

$$\phi(A; B) = \sum_{j,k=0}^{l} \alpha_{jk} A^j \otimes B^k.$$

Then we see that (3) is equivalent to an equation of the form (2) with $G = \phi(A; B')$. The remarkable thing about this class of problems is the following result.

THEOREM 1. *The spectrum of the matrix $\phi(A; B)$ is the set of all numbers of the form $\phi(\lambda, \mu)$, where $\lambda \in \sigma(A)$, $\mu \in \sigma(B)$.*

The reader is referred to [8] or [10] for a proof of this important theorem. Thus, given $\sigma(A)$ or $\sigma(B)$ we can immediately say, for example, that G is singular if and only if $\phi(\lambda, \mu) = 0$ for some eigenvalues λ, μ or A, B, respectively.

3. Special cases. A special case which will occupy us is the much studied equation

$$AX + XB = C, \tag{4}$$

in which case $\phi(\lambda, \mu) = \lambda + \mu$, $\phi(A; B') = G = A \otimes I + I \otimes B'$ and $\sigma(G) = \{\lambda + \mu: \lambda \in \sigma(A), \mu \in \sigma(B)\}$. A further important special case is obtained by putting $B = A^*$, the conjugate transpose of A. We then have $m = n$ and seek an $X \in C_{n \times n}$ for which

$$AX + XA^* = C. \tag{5}$$

We shall refer to this as the *Lyapunov equation* and it is of particular importance because properties of a solution matrix X can sometimes be used to decide whether A is a stable matrix or not. (The matrix A is said to be *stable* if and only if all of its eigenvalues have negative real parts.) We mention also the special case of (4) in which $B = -A$ and $C = 0$. We then include the problem of finding all matrices X which *commute* with the given matrix A, i.e., for which $AX = XA$.

Another case arising frequently in practice can be derived by a transformation of (4). Suppose that a is a nonzero real number and let $f(z) = (z + a)(z - a)^{-1}$.

Define[1] $U = f(A)$ and $V = f(B)$. It is easily verified that (4) is equivalent to

$$(aI + A)X(aI + B) - (aI - A)X(aI - B) = 2aC.$$

If we now suppose that a is chosen so that $(aI - A)^{-1}$ and $(aI - B)^{-1}$ exist, then on pre- and postmultiplying by these matrices, respectively, we obtain

$$UXV - X = 2a(aI - A)^{-1}C(aI - B)^{-1}.$$

Observing that $f(z) - 1 = 2a(z - a)^{-1}$ we obtain

$$X - UXV = -\frac{1}{2a}(U - I)C(V - I). \tag{6}$$

Thus, with an a chosen as indicated, X satisfies (4) if and only if it satisfies (6).

The transformed equation is particularly useful in connection with the stability problem. Thus, suppose it is known that A and B are stable. The function f maps the imaginary axis in the z-plane ($\operatorname{Re}(z) = 0$) onto the unit circle, and the left half-plane ($\operatorname{Re}(z) < 0$) onto the interior of this circle. Thus, if A and B are stable, then $U = f(A)$ and $V = f(B)$ have all their eigenvalues inside the unit circle. If we define the *spectral radius* of a square matrix M, $\rho(M)$, to be the maximum of the moduli of the eigenvalues of M, then we have $\rho(U)$, $\rho(V) < 1$.

Note that in the application of Theorem 1 to (6) we have $\phi(\lambda, \mu) = 1 - \lambda\mu$ and $G = \phi(U; V') = I - U \otimes V'$. Thus, when $\rho(U)\rho(V) < 1$, G is nonsingular, and (6) has a unique solution.

4. Remarks on existence and uniqueness theorems. For the general problem (1), the best we can do for existence and uniqueness results is to formulate them in terms of the equivalent matrix-vector problem (2). Thus, there exists a solution if and only if the vector $\mathbf{c}$ in (2) is in $\mathscr{R}(G)$, the *range* of G. Or, what is equivalent, the rank of the augmented matrix $[G, \mathbf{c}]$ is equal to the rank of G. A third formulation is: There exists a solution of (1) if and only if $\mathbf{y}'\mathbf{c} = \mathbf{0}$ for all vectors $\mathbf{y}$ in $\mathscr{N}(G')$, the *null space* of G', i.e., all $\mathbf{y}$ for which $\mathbf{y}'G = \mathbf{0}'$. Finally, we have the uniqueness theorem: There exists a *unique* solution of (1) if and only if G is nonsingular.

We have noted that we cannot say in general when G will be singular and so it seems unlikely that these results can be related more directly to the spectral properties of the coefficient matrices, A_j and B_j. In the case of (3) we can say when G will be singular, and hence when we have uniqueness. But in analyzing the problem completely and trying to characterize the set of all solutions, (4) seems to be the first tractable case, and even this is far from pretty. We shall not pursue this question. Our interest is in cases where explicit solutions can be written down in matrix form and hence primarily in the case of a unique solution, but also in the nonunique case when special assumptions will simplify the nature of the problem.

There are several good accounts of the general solution of (4) to be found in the literature, among which we mention one of the earlier accounts by Rutherford [15] which has been reformulated by Ma [9], and the account given by Gantmacher [2]. Givens [3] has given a detailed discussion of the structure of $G = A \otimes I - I \otimes B$ in terms of its elementary divisors. Trampus goes further

[1] Refer to § 6 for more information on the definition of $f(A)$, $f(B)$.

and in [18] and [19] he shows how complete sets of generalized eigenvectors can be formed for operators of the form $L_1(X) = AX + XB$ (cf. (4)) and $L_2(X) = AXB$. He avoids our use of the direct product of matrices by observing that eigenelements (or vectors) of L_1 are matrices from $C_{m \times n}$ of rank one. (This divergence in methods of attack follows from two distinct identifications of the tensor product of the spaces C_m and C_n.)

Much the most elegant criteria for existence of a solution of (4) in the case $m = n$ was proved by Roth [14]. He proves that a solution exists if and only if the following partitioned $2n \times 2n$ matrices are similar:

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}, \qquad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

A simple, but significant, necessary condition for the existence of a solution of (3) can be formulated as follows.

THEOREM 2. *A necessary condition for the existence of a solution of* (3) *is that, for every $\lambda_r \in \sigma(A)$ and $\mu_s \in \sigma(B)$ with $\phi(\lambda_r, \mu_s) = 0$, we have $\mathbf{a}'C\mathbf{b} = 0$ for all $\mathbf{a} \in \mathcal{N}(A' - \lambda_r I)$ and all $\mathbf{b} \in \mathcal{N}(B - \mu_s I)$.*

This is easily proved on multiplying (3) on the left by $\mathbf{a}'$, on the right by $\mathbf{b}$ and noting that $\mathbf{a}'A^j = \lambda_r^j\mathbf{a}'$, $B^k\mathbf{b} = \mu_s^k\mathbf{b}$. However, it is instructive to view this another way. We have noted that a necessary and sufficient condition for existence is that, whenever $\phi(\lambda_r, \mu_s) = 0$, we must have $\mathbf{y}'\mathbf{c} = 0$ for all $\mathbf{y} \in \mathcal{N}(G')$. We claim that, in the above notation, $\mathbf{y} = \mathbf{a} \otimes \mathbf{b} \in \mathcal{N}(G')$, for

$$\begin{aligned}(\mathbf{a} \otimes \mathbf{b})'G &= (\mathbf{a}' \otimes \mathbf{b}')\Big(\sum_{j,k} \alpha_{jk}A^j \otimes B'^k\Big) \\ &= \sum_{j,k} \alpha_{jk}(\mathbf{a}'A^j) \otimes (B^k\mathbf{b})' \\ &= \phi(\lambda_r, \mu_s)(\mathbf{a} \otimes \mathbf{b})' = \mathbf{0}.\end{aligned}$$

Thus, when a solution exists, we must also have $(\mathbf{a} \otimes \mathbf{b})'\mathbf{c} = 0$, and this is another way of writing $\mathbf{a}'C\mathbf{b} = 0$.

Using this approach we can easily see that if d_1, d_2 are the dimensions of $\mathcal{N}(A' - \lambda_r I)$ and $\mathcal{N}(B - \mu_s I)$, respectively, then the dimension of $\mathcal{N}(G')$ is not less than $\sum d_1 d_2$, where the summation is over all pairs λ_r, μ_s for which $\phi(\lambda_r, \mu_s) = 0$. Unfortunately it *can* exceed this so that not all the left eigenvectors of G are necessarily formed in this way.

A first insight into the difficulties of the problem is obtained on simplifying (3) by reducing A and B to Jordan normal form. Thus, let

$$A = SJ_1S^{-1}, \qquad B = TJ_2T^{-1},$$

where J_1 and J_2 are matrices in the Jordan normal form. It is then a simple verification to see that (3) is equivalent to $\sum_{j,k} \alpha_{jk} J_1^j \hat{X} J_2^k = \hat{C}$, where $\hat{X} = S^{-1}XT$, $\hat{C} = S^{-1}CT$. Then G takes the form

$$\hat{G} = \phi(J_1; J_2') = \sum_{j,k} a_{jk} J_1^j \otimes J_2'^k.$$

Thus, we see that G is considerably simplified if either A or B has only linear elementary divisors (in which case J_1 or J_2 is diagonal). If A and B both have only linear elementary divisors, then we may write $J_1 = \operatorname{diag}\{\lambda_1, \cdots, \lambda_m\}$,

$J_2 = \operatorname{diag}\{\mu_1, \cdots, \mu_n\}$ and G is diagonal with diagonal elements $\phi(\lambda_r, \mu_s)$, $r = 1, 2, \cdots, m$, $s = 1, 2, \cdots, n$. If, in addition, $\phi(\lambda_r, \mu_s) \neq 0$ for all r and s, then the unique solution is given by

$$\hat{x}_{rs} = \frac{\hat{c}_{rs}}{\phi(\lambda_r, \mu_s)}.$$

5. A solution in series. We consider equations which can be written in the form

$$X - \sum_{j=1}^{t} A_j X B_j = C.$$

The corresponding vector equation is $\mathbf{x} - G\mathbf{x} = \mathbf{c}$, where $G = \sum_j A_j \otimes B_j'$, or $(I - G)\ \mathbf{x} = \mathbf{c}$. When $\det(I - G) \neq 0$ we have the unique solution $\mathbf{x} = (I - G)^{-1}\mathbf{c}$. Now it is well known [8, Theorem 7.1.1] that we may write

$$(I - G)^{-1} = I + G + G^2 + \cdots$$

if and only if $\rho(G) < 1$. Thus, in this case we have the series representation for $\mathbf{x}$,

$$\mathbf{x} = \mathbf{c} + G\mathbf{c} + G^2\mathbf{c} + \cdots, \tag{7}$$

and the bound

$$\|\mathbf{x}\| \leqq \frac{1}{1 - v}\|\mathbf{c}\|, \tag{8}$$

where $v = \|G\| < 1$. Here we use any vector norm together with its induced matrix norm (the Euclidean vector norm and spectral matrix norm, for example). Now the vector series (7) has a matrix representation

$$X = C + \sum_j A_j C B_j + \sum_{j,k} A_j A_k C B_k B_j + \cdots$$

which is just the solution proposed by Wedderburn [20]. Wedderburn arrived at his series in the following way. Put $\mathbf{x} = \mathbf{c} + G\mathbf{x}$ in the second term of $\mathbf{x} - G\mathbf{x} = \mathbf{c}$ to obtain

$$\mathbf{x} - G^2\mathbf{x} = \mathbf{c} + G\mathbf{c}.$$

Substitute in a similar way $n - 1$ times, and we get

$$\begin{aligned} \mathbf{x} - G^3\mathbf{x} &= \mathbf{c} + G\mathbf{c} + G^2\mathbf{c}, \\ &\vdots \\ \mathbf{x} - G^n\mathbf{x} &= \mathbf{c} + G\mathbf{c} + \cdots + G^{n-1}\mathbf{c}. \end{aligned}$$

Thus, if x_n is the nth partial sum of (7), we have

$$\mathbf{x} - \mathbf{x}_n = G^n\mathbf{x}, \tag{9}$$

and combining this with (8) we obtain the error estimate

$$\|\mathbf{x} - \mathbf{x}_n\| \leqq \frac{v^n}{1 - v}\|c\|.$$

In the application of the technique one problem is likely to be that of deciding when $\rho(G) < 1$. If we are in luck, we may be able to apply Theorem 1 to estimate $\rho(G)$. The important equation (6) is such a case which we may write in the form

$$X - UXV = C_1. \tag{10}$$

If $G = U \otimes V'$, then $G^v = U^v \otimes V'^v$. We see from Theorem 1 that $\rho(G) = \rho(U)\rho(V)$ and so we obtain the following theorem.

THEOREM 3. *Equation* (10) *has a unique solution X with the series representation*

$$X = \sum_{j=1}^{\infty} U^{j-1}C_1V^{j-1}$$

if and only if $\rho(U)\rho(V) < 1$.

When (10) is obtained by transformation from $AX + XB = C$, we can certainly use the series solution if it is known that A and B are stable. But note that we may have $\rho(V) \geqq 1$ (B unstable) provided $\rho(U)\rho(V) < 1$ is still satisfied. This corresponds to the fact that X satisfies (4) if and only if

$$(A - \alpha I)X + X(B + \alpha I) = C.$$

Thus, with $\alpha > 0$ and A, B stable, we can arrange for $B + \alpha I$ to be unstable and still have the same solution matrix X. We refer to this as the "spectral shift" property of (4).

A simple a priori bound for X is easily found. Write $X = C_1 + UXV$ and, with any matrix norm, we have

$$\|X\| \leqq \|C_1\| + \|U\|\,\|X\|\,\|V\|.$$

If $\kappa = \|U\|\,\|V\| < 1$, we have

$$\|X\| \leqq (1 - \kappa)^{-1}\|C_1\|. \tag{11}$$

In this case (9) yields $X - X_n = U^nXV^n$, and combining this with (11) we obtain the error estimates (in any matrix norm)

$$\|X - X_n\| \leqq \frac{\kappa^n}{1 - \kappa}\|C_1\|. \tag{12}$$

COROLLARY (to Theorem 3). *If $\|\cdot\|$ denotes any matrix norm, we have $\kappa = \|U\|\,\|V\| < 1$, and we write*

$$X_n = \sum_{j=1}^{n} U^{j-1}XV^{j-1};$$

then condition (12) *obtains for $n = 1, 2, \cdots$.*

Smith [17] discusses a technique for accelerating the convergence of the series in numerical practice.

We should note that (4) can generally be solved without first transforming to the form (10). In [6], Jameson describes a finite recursive technique for direct computation of the solution of (4) in the nonsingular case. This involves the inversion of the matrix of order min (m, n).

6. Results from spectral theory. We shall next discuss Krein's analysis of (3). In order to use his techniques we need some of the results of the spectral theory of operators as developed in generality by Dunford and Schwarz [1], for example, and in the present context by Lancaster [8]. We summarize here some of these results.

If a matrix $A \in C_{n \times n}$ has distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_s$, define the *index* of $\lambda_k, k = 1, 2, \cdots, s$, to be the least positive integer, m_k, for which $\mathcal{N}(I\lambda_k - A)^{m_k} = \mathcal{N}(I\lambda_k - A)^{m_k+1}$. We then say that a complex-valued function f of a complex variable is defined on the spectrum of A if f and its first $m_k - 1$ derivatives exist at $\lambda_k, k = 1, 2, \cdots, s$. For such a function f the matrix $f(A)$ is well-defined and there exist *component matrices* Z_{kj}, $k = 1, 2, \cdots, s$, $j = 1, 2, \cdots, m_k$, depending only on A, which are linearly independent members of $C_{n \times n}$ and

$$f(A) = \sum_{k=1}^{s} \sum_{j=1}^{m_k} f_k^{(j-1)} Z_{kj}, \tag{13}$$

where $f_k^{(j-1)}$ is the $(j - 1)$st derivative of f evaluated at λ_k. The component matrices are polynomials in A (and hence commute with A) and may also be expressed as contour integrals

$$Z_{kj} = \frac{1}{(j - 1)!2\pi i} \int_{C_k} (z - \lambda_k)^{j-1} R(z, A)\, dz, \tag{14}$$

where C_k is a sufficiently small circle with center λ_k and $R(z, A)$ is the *resolvent* of A, $(Iz - A)^{-1}$. The reader will easily verify the *resolvent equation*

$$R(\lambda, A) - R(\mu, A) = (\mu - \lambda)R(\lambda, A)R(\mu, A). \tag{15}$$

The matrices Z_{kl} are particularly important. They are *idempotent*, $Z_{k1}^2 = Z_{k1}$, and have the property

$$\mathscr{R}(Z_{k1}) = \mathcal{N}(I\lambda_k - A)^{m_k}. \tag{16}$$

We also describe an idempotent matrix K as a *projection* onto $\mathscr{R}(K)$. Since $I = Z_{11} + Z_{21} + \cdots + Z_{s1}$ (deduce from (13)), it follows that (using the *direct sum* of subspaces)

$$C_n = \mathscr{R}(Z_{11}) \oplus \cdots \oplus \mathscr{R}(Z_{s1}) = \sum_{k=1}^{s} \oplus \mathscr{R}(Z_{k1}).$$

The matrices Z_{k1} are also *orthogonal*, in the sense that $Z_{j1}Z_{k1} = 0$ when $j \neq k$. Important special cases of (13) are:

$$A = \sum_{k=1}^{s} (\lambda_k Z_{k1} + Z_{k2}), \tag{17}$$

$$R(\lambda, A) = \sum_{k=1}^{s} \sum_{j=1}^{m_k} \frac{(j - 1)!}{(\lambda - \lambda_k)^j} Z_{kj}. \tag{18}$$

A *contour* in the complex plane will always mean a finite number of rectifiable Jordan curves oriented in the usual sense. If Γ is a contour, $\tilde{\Gamma}$ denotes the interior of Γ and $\hat{\Gamma} = \Gamma \cup \tilde{\Gamma}$, the union of Γ and $\tilde{\Gamma}$.

Now let Γ be a contour with no member of $\sigma(A)$ on Γ ($\Gamma \cap \sigma(A) = \emptyset$, the empty set). The operator

$$P = \frac{1}{2\pi i}\int_\Gamma R(z, A)\, dz$$

is a projection and depends only on the poles of $R(z, A)$ in $\tilde{\Gamma}$, that is, on the members of $\sigma(A)$ in $\tilde{\Gamma}$. If $\tilde{\Gamma} \cap \sigma(A) = \sigma_1$, we refer to P as the projection *defined by A and σ_1*. In fact, we have

$$P = \sum_k Z_{k1}, \qquad \mathscr{R}(P) = \sum_k \oplus\, \mathscr{R}(Z_{k1})$$

summed over those k for which $\lambda_k \in \sigma_1$. The operator $I - P$ is also a projection and $C_n = \mathscr{R}(P) + \mathscr{R}(I - P)$. Note also that $AP = PA$.

If f is continuous on Γ and regular in $\tilde{\Gamma}$, then we have, more generally,

$$Pf(A) = f(A)P = \frac{1}{2\pi i}\int_\Gamma f(z)R(z, A)\, dz. \tag{19}$$

Note that this integral is well-defined even though $f(A)$ itself may not be defined as a result of singularities in f at points of the spectrum of A outside Γ. To allow for such cases we may also write $Pf(A) = f_P(A)$.

Finally, we note that if P is any projection (idempotent) and $\mathbf{x} \in \mathscr{R}(P)$, then $P\mathbf{x} = \mathbf{x}$. To see this, observe that $\mathbf{x} \in \mathscr{R}(P)$ implies that there is a $\mathbf{y}$ with $\mathbf{x} = P\mathbf{y}$. Multiplying by P and putting $P^2 = P$, we get

$$P\mathbf{x} = P^2\mathbf{y} = P\mathbf{y} = \mathbf{x}.$$

7. The theorem of Krein. We first need a preliminary result.

LEMMA. *Let $A \in C_{n\times n}$, F_1, F_2 be regular functions in a neighborhood N of $\sigma(A)$ with values in $C_{l\times n}$, $C_{n\times m}$ respectively, and Γ a contour in the interior of N with $\sigma(A) \cap \Gamma = \emptyset$. Then*

$$\begin{aligned}&\left[\frac{1}{2\pi i}\int_\Gamma F_1(\lambda)R(\lambda, A)\, d\lambda\right]\left[\frac{1}{2\pi i}\int_\Gamma R(\lambda, A)F_2(\lambda)\, d\lambda\right] \\ &\quad = \frac{1}{2\pi i}\int_\Gamma F_1(\lambda)R(\lambda, A)F_2(\lambda)\, d\lambda.\end{aligned} \tag{20}$$

Proof. Choose a new contour Γ_1 with $\Gamma \subset \tilde{\Gamma}_1$, $\Gamma_1 \subset N$ and $\tilde{\Gamma} \cap \sigma(A) = \hat{\Gamma}_1 \cap \sigma(A)$. Then write

$$\int_\Gamma R(\lambda, A)F_2(\lambda)\, d\lambda = \int_{\Gamma_1} R(\mu, A)F_2(\mu)\, d\mu.$$

Using the resolvent equation (15) we manipulate the left-hand side of (20) as follows:

$$\left[\frac{1}{2\pi i}\int_{\Gamma} F_1(\lambda)R(\lambda, A)\,d\lambda\right]\left[\frac{1}{2\pi i}\int_{\Gamma_1} R(\mu, A)F_2(\mu)\,d\mu\right]$$

$$= -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma} F_1(\lambda)R(\lambda, A)R(\mu, A)F_2(\mu)\,d\lambda\,d\mu$$

$$= \frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma} F_1(\lambda)\frac{R(\lambda, A) - R(\mu, A)}{\lambda - \mu}F_2(\mu)\,d\lambda\,d\mu$$

$$= \frac{1}{4\pi^2}\int_{\Gamma} F_1(\lambda)R(\lambda, A)\left(\int_{\Gamma_1}\frac{F_2(\mu)}{\lambda - \mu}\,d\mu\right)d\lambda$$

$$- \frac{1}{4\pi^2}\int_{\Gamma_1}\left(\int_{\Gamma}\frac{F_1(\lambda)}{\lambda - \mu}\,d\lambda\right)R(\mu, A)F_2(\mu)\,d\mu.$$

From our definition of Γ_1 we see that $F_2(\mu)/(\lambda - \mu)$ is a regular function of μ in $\hat{\Gamma}_1$ save for a simple pole at $\mu = \lambda$, and that $F_1(\lambda)/(\lambda - \mu)$ is regular in $\hat{\Gamma}$. Hence

$$\int_{\Gamma_1}\frac{F_2(\mu)}{\lambda - \mu}\,d\mu = -2\pi i F_2(\lambda), \qquad \int_{\Gamma}\frac{F_1(\lambda)}{\lambda - \mu}\,d\lambda = 0$$

and the above expression reduces to

$$\frac{1}{2\pi i}\int_{\Gamma} F_1(\lambda)R(\lambda, A)F_2(\lambda)\,d\lambda$$

which is the right-hand side of (20).

We can now prove the main result concerning (3). We first recall the definition $\phi(x, y) = \sum_{j,k}\alpha_{jk}x^j y^k$.

THEOREM 4 (M. G. Krein). *Let $A \in C_{m\times m}$, $B \in C_{n\times n}$, $C \in C_{m\times n}$ and suppose that there exist $\sigma_1 \subseteq \sigma(A)$ and $\sigma_2 \subseteq \sigma(B)$ such that $\phi(\lambda, \mu) \neq 0$ for $\lambda \in \sigma_1$, $\mu \in \sigma_2$. If P_1, P_2 are the projections defined by A, σ_1 and B, σ_2 respectively, and if $P_1CP_2 = C$, then* (3) *has the solution*

$$X = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\frac{R(\lambda, A)CR(\mu, B)}{\phi(\lambda, \mu)}\,d\mu\,d\lambda, \tag{21}$$

where Γ_1, Γ_2 are contours containing and sufficiently close to σ_1, σ_2, respectively.

Proof. The proof is by verification. From (19) we have

$$A^jP_1 = \frac{1}{2\pi i}\int_{\Gamma_1}\lambda^j R(\lambda, A)\,d\lambda, \quad \text{and} \quad P_2B^k = \frac{1}{2\pi i}\int_{\Gamma}\mu^k R(\mu, B)\,d\mu.$$

Then, putting $C = P_1CP_2$ in (21), we have

$$A^jXB^k = A^jP_1\left[-\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\frac{R(\lambda, A)CR(\mu, B)}{\phi(\lambda, \mu)}\,d\mu\,d\lambda\right]P_2B^k$$

$$= \left[\frac{1}{2\pi i}\int_{\Gamma_1}\lambda^jR(\lambda, A)\,d\lambda\right]\left[\frac{1}{2\pi i}\int_{\Gamma_1}R(\lambda, A)C\left(\frac{1}{2\pi i}\int_{\Gamma_2}\frac{R(\mu, B)}{\phi(\lambda, \mu)}\,d\mu\right)d\lambda\right]$$
$$\cdot\left[\frac{1}{2\pi i}\int_{\Gamma_2}\mu^kR(\mu, B)\,d\mu\right].$$

Since $\phi(\lambda, \mu) \neq 0$ for $\lambda \in \sigma_1$, $\mu \in \sigma_2$, we can choose Γ_1, Γ_2 so close to σ_1, σ_2 that the integral in round brackets exists and is a regular function of λ on Γ_1. We may then apply the lemma to the first two square brackets and obtain

$$A^jXB^k = \left[\frac{1}{2\pi i}\int_{\Gamma_1}\lambda^jR(\lambda, A)C\left(\frac{1}{2\pi i}\int_{\Gamma_2}\frac{R(\mu, B)}{\phi(\lambda, \mu)}\,d\mu\right)d\lambda\right]\left[\frac{1}{2\pi i}\int_{\Gamma_2}\mu^kR(\mu, B)\,d\mu\right]$$

$$= \frac{1}{2\pi i}\int_{\Gamma_2}\left(\frac{1}{2\pi i}\int_{\Gamma_1}\frac{\lambda^jR(\lambda, A)}{\phi(\lambda, \mu)}\,d\lambda\right)CR(\mu, B)\mu^k\,d\mu,$$

having used the lemma a second time. Thus

$$A^jXB^k = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\frac{\lambda^j\mu^k}{\phi(\lambda, \mu)}R(\lambda, A)CR(\mu, B)\,d\mu\,d\lambda$$

and

$$\sum_{j,k}\alpha_{jk}A^jXB^k = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}R(\lambda, A)CR(\mu, B)\,d\mu\,d\lambda$$
$$= \left(\frac{1}{2\pi i}\int_{\Gamma_1}R(\lambda, A)\,d\lambda\right)C\left(\frac{1}{2\pi i}\int_{\Gamma_2}R(\mu, B)\,d\mu\right)$$
$$= P_1CP_2$$
$$= C.$$

Remark 1. The hypothesis $P_1CP_2 = C$ is equivalent to $(I - P_1)C = 0$ and $C(I - P_2) = 0$.

To see this we multiply $P_1CP_2 = C$ on the left by P_1 and on the right by P_2 to obtain $P_1C = C$ and $CP_2 = C$, whence $(I - P_1)C = 0$ and $C(I - P_2) = 0$. Conversely, putting $C = P_1C$ on the right of $C = CP_2$, we get $P_1CP_2 = C$.

Remark 2. The solution (21) also satisfies the hypothesis on C:

$$X = P_1XP_2. \tag{22}$$

Remark 3. If $\sigma_1 = \sigma(A)$, $\sigma_2 = \sigma(B)$, then $\phi(\lambda, \mu) \neq 0$ for $\lambda \in \sigma(A)$, $\mu \in \sigma(B)$, and X is the *unique* solution of (3). Note that in this case $P_1 = P_2 = I$.

In general X is the unique solution of the pair of equations (3) and (22).

Remark 4. Using the functional notation introduced at the end of § 6, we may write

$$\frac{1}{2\pi i}\int_{\Gamma_1}\frac{R(\lambda, A)}{\phi(\lambda, \mu)}\,d\lambda = \phi_{P_1}(A, I\mu)^{-1}.$$

Hence

$$X = \frac{1}{2\pi i}\int_{\Gamma_2}\phi_{P_1}(A, I\mu)^{-1}CR(\mu, B)\,d\mu. \tag{23}$$

Similarly, we may write

$$X = \frac{1}{2\pi i}\int_{\Gamma_1}R(\lambda, A)C\phi_{P_2}(I\lambda, B)^{-1}\,d\lambda. \tag{24}$$

8. The singular case. The most useful applications of Theorem 4 will doubtless be in the nonsingular case, when $\phi(\lambda, \mu) \neq 0$ for all $\lambda \in \sigma(A)$ and $\mu \in \sigma(B)$. However it also provides us with sufficient conditions for the existence of a solution in the singular case. Let us investigate this case a little more closely. Let τ_1 be the complement of σ_1 in $\sigma(A)$ and τ_2 be the complement of σ_2 in $\sigma(B)$. Assume that, in addition to the hypothesis of Theorem 4, for each $\lambda \in \tau_1$, $\phi(\lambda, \mu) = 0$ for some $\mu \in \tau_2$, and conversely. Then $Q_1 = I - P_1$ is the projection defined by A, τ_1 and $Q_2 = I - P_2$ is the projection defined by B, τ_2, and in Remark 1 above we have seen that if $Q_1C = CQ_2 = 0$, then X is a solution.

Now, if $\mathbf{x}$ is a right eigenvector of B with eigenvalue $\mu \in \tau_2$, then $\mathbf{x} \in \mathscr{R}(Q_2)$. (Note that $Q_2 = \sum Z_{k1}$ summed over those k with $\mu_k \in \tau_2$ and also (16).) Hence $Q_2\mathbf{x} = \mathbf{x}$ and $CQ_2 = 0$ implies that $C\mathbf{x} = \mathbf{0}$. Thus, $CQ = 0$ implies that $C\mathbf{x} = \mathbf{0}$ for all right eigenvectors $\mathbf{x}$ of the eigenvalues of B in τ_2. Similarly, it can be shown that $Q_1C = 0$ implies that $\mathbf{y}'C = \mathbf{0}'$ for all left eigenvectors $\mathbf{y}$ of the eigenvalues of A in τ_1.

Comparing this with Theorem 2 we see that either of the conditions $CQ_2 = 0$ or $Q_1C = 0$ alone is stronger than the necessary condition for existence given there.

9. The equation $AX + XB = C$. In this case (21) reduces to

$$X = -\frac{1}{4\pi^2}\int_{\Gamma_1}\int_{\Gamma_2}\frac{R(\lambda, A)CR(\mu, B)}{\lambda + \mu}\,d\mu\,d\lambda. \tag{25}$$

Using (23) and (24) we may also write

$$X = \frac{1}{2\pi i}\int_{\Gamma_2}(I\mu + A)_{P_1}^{-1}C(I\mu - B)^{-1}\,d\mu = \frac{1}{2\pi i}\int_{\Gamma_1}(I\lambda - A)^{-1}C(I\lambda + B)_{P_2}^{-1}\,d\lambda. \tag{26}$$

There is an interesting case in which X can be expressed as a real, improper integral. We follow the notation of the last theorem.

THEOREM 5 (M. G. Krein). *If* $\operatorname{Re}(\lambda + \mu) < 0$ *for all* $\lambda \in \sigma_1$ *and* $\mu \in \sigma_2$, *and if* $P_1CP_2 = C$, *then*

$$X = -\int_0^\infty e^{At}Ce^{Bt}\,dt \tag{27}$$

is a solution of $AX + XB = C$.

Proof. With the proper choice of Γ_1 and Γ_2 we will have $\mathrm{Re}(\lambda + \mu) < 0$ for all $\lambda \in \Gamma_1$, $\mu \in \Gamma_2$ and in (25) we may write

$$\frac{1}{\lambda + \mu} = -\int_0^\infty e^{(\lambda+\mu)t}\,dt.$$

We then have

$$X = \frac{1}{4\pi^2}\int_0^\infty \left(\int_{\Gamma_1} e^{\lambda t}R(\lambda, A)\,d\lambda\right) C\left(\int_{\Gamma_2} e^{\mu t}R(\mu, B)\,d\mu\right) dt.$$

Using (19) we obtain

$$X = -\int_0^\infty e^{At}P_1CP_2e^{Bt}\,dt = -\int_0^\infty e^{At}Ce^{Bt}\,dt.$$

Note, in particular, that for eigenvalues of A with $\mathrm{Re}(\lambda) < 0$ and of B with $\mathrm{Re}(\mu) < 0$ we have $\mathrm{Re}(\lambda + \mu) < 0$. Thus, if A and B are stable, then $P_1 = P_2 = I$ and (27) is then the unique solution.

Let us now return to the second integral expression of (26) and recall the constraints on Γ_1. We must have $\sigma_1 \subset \hat{\Gamma}_1$, no other points of $\sigma(A)$ in $\hat{\Gamma}_1$ and no members of $\sigma(-B)$ in $\hat{\Gamma}_1$. Bearing this in mind we ask to what extent Γ_1 can be continuously deformed without changing the value of the integral. We may suppose that we begin the deformation from a set of small circles with centers at the points of σ_1. However, noting that $P_1 = \sum_r Z_{r1}$ summed over those r for which $\lambda_r \in \sigma_1$, and using (18) together with the orthogonality of the component matrices, we can write for the integrand of (26):

$$(I\lambda - A)^{-1}P_1C(I\lambda + B)^{-1}_{P_2} = \left(\sum_r \sum_{j=1}^{m_r} \frac{(j-1)!}{(\lambda - \lambda_r)^j} Z_{rj}\right) C \cdot \left(\sum_s \sum_{k=1}^{m_s} \frac{(-1)^{k-1}(k-1)!}{(\lambda + \mu_s)^k} Y_{sk}\right)$$

and the s sum is over selected numbers of $\sigma(-B)$ *outside* $\hat{\Gamma}_1$ (that is, the members of σ_2). The matrices Y_{sk} are, of course, the component matrices of B.

Thus, the property $C = P_1CP_2$ implies that we may continue the deformation so that $\hat{\Gamma}_1$ contains other eigenvalues of A, provided they are not also eigenvalues of $-B$ in σ_2. By doing so, we introduce no further singularities into the integrand. We may also include some eigenvalues of $-B$ in $\hat{\Gamma}_1$, namely, the members of $\sigma(-B) - \sigma_2$.

Similar arguments may be applied to deformations of Γ_1 and Γ_2 in the general solution (21).

The study of solutions in the form (26) is simplified if there are simple geometrical boundaries in the complex plane separating σ_1 from $\sigma(-B)$ which can then be used as the contour Γ_1 (or part of it). For example, if σ_1 is contained in a circle and $\sigma(-B)$ is outside this circle (the circle may be supposed to have its center at the origin using the spectral shift property) we can use this circle for Γ_1, thus transforming to what may be a more tractable real integral.

As another example we develop a formal proof of Theorem 5. We shall need the following lemma (which looks like a Laplace transform).

LEMMA. *If* $\operatorname{Re}(\mu) < R(\lambda)$ *for all* $\mu \in \sigma(A)$, *then*

$$(I\lambda - A)^{-1} = \int_0^\infty e^{-(I\lambda - A)t}\,dt.$$

The proof is left to the reader.

Now suppose that $\max_{\mu\in\sigma(A)} \operatorname{Re}(\mu) < c_1$, $\max_{\mu\in\sigma(B)} \operatorname{Re}(\mu) = c_2$ and that $c_1 + c_2 < 0$. We consider two semicircular contours as indicated in the sketch. By choosing their common radius sufficiently large we can arrange for $\sigma(A)$,

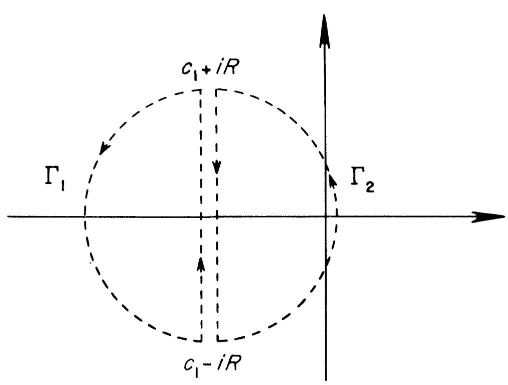

FIG. 1. *The complex λ-plane*

$\sigma(-B)$ to be inside the left and right contours respectively. Choose the left-hand contour for Γ_1, and name the right-hand contour Γ_2. It is a routine matter to show that in (26) the integral over the semicircle approaches zero as $R \to \infty$ and hence

$$X = \frac{1}{2\pi i}\int_{c_1 - i\infty}^{c_1 + i\infty} (I\lambda - A)^{-1}C(I\lambda + B)^{-1}\,d\lambda.$$

Using the lemma, we have

$$\begin{aligned} X &= \frac{1}{2\pi i}\int_{c_1 - i\infty}^{c_1 + i\infty}\left(\int_0^\infty e^{-(I\lambda - A)t}\,dt\right)C(I\lambda + B)^{-1}\,d\lambda \\ &= \frac{1}{2\pi i}\int_0^\infty e^{At}C\int_{c_1 - i\infty}^{c_1 + i\infty} e^{-\lambda t}(I\lambda + B)^{-1}\,d\lambda\,dt \\ &= -\frac{1}{2\pi i}\int_0^\infty e^{At}C\left[\int_{\Gamma_2} e^{-\lambda t}(I\lambda + B)^{-1}\,d\lambda\right]dt \\ &= -\int_0^\infty e^{At}Ce^{Bt}\,dt, \end{aligned}$$

and we have used (19) at the last step.

It is of interest to ask more generally when integrals of the kind found in (26) can be solutions of (4), taking a general contour Γ for the integration. The answer is provided in the following theorem.

THEOREM 6. *Let Γ be any contour with no points of $\sigma(A) \cup \sigma(-B)$ on Γ. Let $\tau_1 = \sigma(A) \cap \tilde{\Gamma}$ and $\tau_2 = \sigma(-B) \cap \tilde{\Gamma}$ and K_1, K_2 be the projections defined by A, τ_1 and $-B, \tau_2$ respectively. Let δ be any subset of $\sigma(-B)$ and J be the projection defined by $-B$, δ. Then*

$$X = \frac{1}{2\pi i}\int_\Gamma (I\lambda - A)^{-1}C(I\lambda + B)_J^{-1}\,d\lambda \tag{28}$$

is a solution of $AX + XB = C$ if and only if $K_1CJ = C$ and $CJ_0 = 0$, where $J_0 = K_2J$ (the projection defined by $-B$ and $\tau_2 \cap \delta$).

Proof. Note first that

$$A(I\lambda - A)^{-1} = \lambda(I\lambda - A)^{-1} - I, \qquad (I\lambda + B)_J^{-1}B = J - \lambda(I\lambda + B)_J^{-1},$$

and supposing that (28) is a solution we have

$$\begin{aligned}
AX + XB &= \frac{1}{2\pi i}\int_\Gamma A(I\lambda - A)^{-1}C(I\lambda + B)_J^{-1}\,d\lambda \\
&\quad + \frac{1}{2\pi i}\int_\Gamma (I\lambda - A)^{-1}C(I\lambda + B)_J^{-1}B\,d\lambda \\
&= \frac{1}{2\pi i}\int_\Gamma \{\lambda(I\lambda - A)^{-1} - I\}C(I\lambda + B)_J^{-1}\,d\lambda \\
&\quad + \frac{1}{2\pi i}\int (I\lambda - A)^{-1}C\{J - \lambda(I\lambda + B)_J^{-1}\}\,d\lambda \\
&= \frac{1}{2\pi i}C\int_\Gamma (I\lambda + B)_J^{-1}\,d\lambda + \frac{1}{2\pi i}\int_\Gamma (I\lambda - A)^{-1}\,d\lambda \cdot CJ \\
&= K_1CJ - CK_2J = C.
\end{aligned}$$

Since this argument is reversible we see that (28) is a solution if and only if $(K_1C - CK_2)J = C$. We prove the theorem by showing that this single condition is equivalent to $CJ_0 = 0$ *and* $K_1CJ = C$.

That these two conditions imply $(K_1C - CK_2)J = C$ is obvious. Conversely, premultiply the latter equation by K_1 and postmultiply by J and we see that $K_1CK_2J = 0$. By postmultiplying $(K_1C - CK_2)J = C$ by K_2J we find that $CK_2J = 0$. It follows immediately that $K_1CJ = C$ and the proof is complete.

It is easily seen that (28) itself also satisfies the conditions on C; namely, $XJ_0 = 0$ and $K_1XJ = X$.

We obtain the result of (26) when $\tilde{\Gamma}$ contains no eigenvalues of $-B$ so that $\tau_2 = \varnothing$ and we may take $J = P_2$. Then $K_2J = J_0 = 0$ and $K_1 = P_1$ so that the two conditions on C reduce to $P_1CP_2 = C$, the hypothesis of Theorem 4.

10. Solutions in terms of component matrices. As before we have $A \in C_{m \times m}$, $B \in C_{n \times n}$ and $C \in C_{m \times n}$. Let A, B have distinct eigenvalues $\lambda_1, \cdots, \lambda_s$ and $\mu_1, \cdots, \mu_t$ respectively, and suppose that $\lambda_1 = -\mu_1, \cdots, \lambda_r = -\mu_r$, with no other coincidences among the λ_j, $-\mu_k$. Let $B_\alpha = B + \alpha I$, $\alpha \neq 0$, and choose α small enough so that $\lambda_j \neq -(\mu_k + \alpha)$ for *all* j, k. There is then a unique $X_\alpha \in C_{m \times n}$ such that

$$AX_\alpha + X_\alpha B_\alpha = C.$$

We now choose a contour Γ with $\sigma(A) \subset \tilde{\Gamma}$ and $\sigma(B) \cap \tilde{\Gamma} = \emptyset$. Let Z_{kj}, $k = 1, 2, \cdots, s$; $j = 1, 2, \cdots, m_k$, be the component matrices for A, and with similar conventions, Y_{ih} are the component matrices for B (and hence B_α). Using (18) we find that, if μ_i has index $\bar{m}_i$, $i = 1, 2, \cdots, t$, then

$$(I\lambda + B_\alpha)^{-1} = \sum_{i=1}^{t} \sum_{h=1}^{\bar{m}_i} \frac{(-1)^{h-1}(h-1)!}{(\lambda + \mu_i + \alpha)^h} Y_{ih}. \tag{29}$$

Using this expansion in (28) we have

$$\begin{aligned} X_\alpha &= \frac{1}{2\pi i} \int_\Gamma (I\lambda - A)^{-1} C (I\lambda + B_\alpha)^{-1} \, d\lambda \\ &= \frac{1}{2\pi i} \int_\Gamma (I\lambda - A)^{-1} C \sum_{i,h} \frac{(-1)^{h-1}(h-1)}{(\lambda + \mu_i + \alpha)^h} Y_{ih} \, d\lambda. \end{aligned}$$

Using (19) we have

$$\frac{1}{2\pi i} \int_\Gamma (\lambda + \mu_i + \alpha)^{-h} (I\lambda - A)^{-1} \, d\lambda = (I(\mu_i + \alpha) + A)^{-h}$$

and so, using (13) for this function of A,

$$X_\alpha = \sum_{i,h} (-1)^{h-1}(h-1)!(I(\mu_i + \alpha) + A)^{-h} C Y_{ih}. \tag{30}$$

$$X_\alpha = \sum_{i,h} \sum_{k,j} \frac{(-1)^{h+j}(h+j-2)!}{(\mu_i + \lambda_k + \alpha)^{h+j-1}} Z_{kj} C Y_{ih}. \tag{31}$$

We see at once that we may run into trouble if we approach the limit $\alpha = 0$ in those terms for which $\mu_i + \lambda_k = 0$. We also see that, when $r = 0$, we may simply put $\alpha = 0$ in (31) and obtain the unique solution X, for the nonsingular case. We go on to investigate when solutions of the same kind obtain in the singular case, $r > 0$.

Now suppose that X_α is written out in powers of α. Thus

$$X_\alpha = \sum_{v=-\rho}^{\infty} X_v \alpha^v, \qquad X_{-\rho} \neq 0.$$

Then we see that

$$X_0 = \sum_{i=1}^{t} \sum_{h=1}^{\bar{m}_i} \sum_{k=1}^{s} {\sum_{j=1}^{m_k}}' \frac{(-1)^{h+j}(h+j-2)!}{(\mu_i + \lambda_k)^{h+j-1}} Z_{kj} C Y_{ih} \tag{32}$$

and the prime on the summation means that *in this sum we simply exclude* all those terms for which $\mu_i + \lambda_k = 0$. We also have

$$X_{-1} = \sum_{i,k=1}^{r} Z_{k1}CY_{i1}, \qquad X_{-2} = \sum_{i,k=1}^{r} (Z_{k2}CY_{i1} + Z_{k1}CY_{i2}). \tag{33}$$

Then from $AX_\alpha + X_\alpha B_\alpha = C$ we deduce that

$$\begin{aligned} AX_{-\rho} + X_{-\rho}B &= 0, \\ AX_{-\rho+1} + X_{-\rho+1}B + X_{-\rho} &= 0, \\ &\vdots \\ AX_{-1} + X_{-1}B + X_{-2} &= 0, \quad (34) \\ AX_0 + X_0B + X_{-1} &= C, \quad (35) \\ &\vdots \end{aligned}$$

Now substitute (33) into (34), write (using (17))

$$AZ_{k1} = \lambda_k Z_{k1} + Z_{k2}, \qquad Y_{i1}B = \mu_i Y_{i1} + Y_{i2} \tag{36}$$

and multiply the result from left and right by Z_{k1}, Y_{i1}, respectively, to obtain

$$(\lambda_k Z_{k1} + Z_{k2})CY_{i1} + Z_{k1}C(\mu_i Y_{i1} + Y_{i2}) = Z_{k2}CY_{i1} + Z_{k1}CY_{i2}.$$

This simplifies to

$$(\lambda_k + \mu_i)Z_{k1}CY_{i1} = 0.$$

Thus, if $\lambda_{jk} + \mu_i \neq 0$, $1 \leqq i, k \leqq r$, then $Z_{k1}CY_{i1} = 0$ and since

$$(j-1)!Z_{kj} = (A - \lambda_k I)^{j-1}Z_{k1}, \qquad (h-1)!Y_{ih} = Y_{i1}(B - \mu_i I)^{h-1},$$

we deduce that

$$Z_{kj}CY_{ih} = 0 \qquad \text{for } 1 \leqq i, k \leqq r, \quad i \neq k.$$

The expressions (33) therefore simplify and we have, in particular,

$$X_{-1} = \sum_{k=1}^{r} Z_{k1}CY_{k1}.$$

Substituting in (35) we obtain the following theorem.

THEOREM 7. *Let* $A \in C_{m\times m}$, $B \in C_{n\times n}$ *and* $\sigma(A) = \{\lambda_1, \cdots, \lambda_s\}$, $\sigma(B) = \{\mu_1, \cdots, \mu_t\}$. *Suppose that* $\lambda_j = \mu_k$ *if and only if* $1 \leqq j = k \leqq r$. *Then*

$$AX_0 + X_0B + \sum_{k=1}^{r} Z_{k1}CY_{k1} = C,$$

where X_0 *is given by* (32).

We can use this at once to improve on a result of Rosenblum [12].

THEOREM 8. *With the hypothesis of Theorem* 7 *we have*:

(i) *If* $Z_{k1}CY_{k1} = 0$ *for* $k = 1, 2, \cdots, r$, *then there exists a solution of* $AX + XB = C$ *and* X_0 *is such a solution.*

(ii) *If the common eigenvalues* $\lambda_1, \cdots, \lambda_r$ *have linear elementary divisors in both* A *and* $-B$ *and there exists a solution of* $AX + XB = C$, *then* $Z_{k1}CY_{k1} = 0$, $k = 1, 2, \cdots, r$.

Proof. Part (i) follows immediately from Theorem 7. For part (ii) we deduce from $AX + XB = C$ by multiplying from left and right by Z_{k1}, Y_{k1} respectively and using (36),

$$(\lambda_k Z_{k1} + Z_{k2})XY_{k1} + Z_{k1}X(-\lambda_k Y_{k1} + Y_{k2}) = Z_{k1}CY_{k1},$$

whence

$$Z_{k1}CY_{k1} = Z_{k2}XY_{k1} + Z_{k1}XY_{k2}.$$

If $\lambda_1, \cdots, \lambda_r$ have linear elementary divisors in both A and $-B$, then $Z_{k2} = Y_{k2} = 0$ for $k = 1, 2, \cdots, r$ and we deduce that $Z_{k1}CY_{k1} = 0$.

To see that $Z_{k1}CY_{k1} = 0$ is not necessary in every case, take

$$A = B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} c & c_1 \\ 0 & c \end{bmatrix} \neq 0.$$

Then,

$$X = \begin{bmatrix} c_1 & 0 \\ c & 0 \end{bmatrix}$$

is a solution of $AX + XB = C$ and $Z_{11}CY_{11} = C \neq 0$.

We obtain the following corollary as a special case of the theorem.

COROLLARY. *If* A *and* B *are simple matrices, then there exists a solution of* $AX + XB = C$ *if and only if* $Z_{k1}CY_{k1} = 0$ *for* $k = 1, 2, \cdots, r$.

Note that when A and B are simple the summation (32) reduces to

$$X_0 = \sum_{i=1}^{t} \sum_{k=1}^{s}{}' \frac{Z_{k1}CY_{i1}}{\mu_i + \lambda_k}.$$

In the case of a Lyapunov equation with A simple we have

$$X_0 = \sum_{i=1}^{s} \sum_{k=1}^{s}{}' \frac{Z_{k1}CZ_{i1}^*}{\lambda_k + \bar{\lambda}_i}.$$

To make a connection with preceding results let us apply Theorem 4 taking

$$P_1 = \sum_{u=r+1}^{s} Z_{u1}, \qquad P_2 = \sum_{v=r+1}^{t} Y_{v1}.$$

The hypothesis $P_1CP_2 = C$ gives

$$\sum_{u=r+1}^{s} Z_{u1}C \sum_{v=r+1}^{t} Y_{v1} = C$$

and multiplying from left and right by Z_{k1}, Y_{k1}, respectively, with $1 \leqq k \leqq r$ we obtain $Z_{k1}CY_{k1} = 0$ again.

Conversely, $Z_{k1}CY_{k1} = 0$ for $k = 1, 2, \cdots, r$ does *not* necessarily imply $P_1CP_2 = C$, so the hypotheses of Theorem 8 are weaker than those of Theorem 4. We can certainly use the solutions (26), obtained from Theorem 4, to obtain summations of the form (32), but the i, k summations will run from $r + 1$ to s and $r + 1$ to t, respectively, thus removing the need for the "primed" summation convention.

11. Solutions in terms of adjoint matrices. We first recall some definitions ([2] or [8]). The *reduced* adjoint of $I\lambda - A$ is the adjoint matrix, $\operatorname{adj}(I\lambda - A)$, divided by the greatest common divisor of its elements. The minimal polynomial of A is the characteristic polynomial of A divided by this same factor. Let $E(\lambda)$, $F(\lambda)$ be the reduced adjoints of $I\lambda - A, I\lambda - B$ respectively and let ψ, ϕ be the minimal polynomials of A, B. It is easily seen that we have

$$E(\lambda)(I\lambda - A) = (I\lambda - A)E(\lambda) = \psi(\lambda)I. \tag{37}$$

It should also be noted that if a matrix has all its eigenvalues distinct then the reduced adjoint coincides with the adjoint, and the characteristic and minimal polynomials coincide.

Now it is known (Gantmacher [2]) that the component matrices can be expressed in terms of the reduced adjoint and minimal polynomials. We have, for example,

$$Y_{ih} = \frac{1}{(h-1)!(m_i - h)!}\left[\frac{F(\lambda)}{\phi_i(\lambda)}\right]^{(m_i - h)}_{\lambda = \mu_i}, \tag{38}$$

where the index denotes a derivative, the subscript says the derivative is to be evaluated at μ_i, and $\phi_i(\lambda) = \phi(\lambda)/(\lambda - \mu_i)^{m_i}$.

Thus, all the results obtained in the last section in terms of component matrices can also be expressed in terms of adjoints and minimal polynomials. Let us illustrate in the relatively simple case in which either A or B is simple. Thus, if B is simple and A and $-B$ have no eigenvalues in common, we can use (30) to write the unique solution in the form

$$X = \sum_{i=1}^{t} (I\mu_i + A)^{-h} C Y_{i1}.$$

From (38) we have

$$Y_{i1} = \frac{F(\mu_i)}{\phi^{(1)}(\mu_i)},$$

and from (37),

$$(I\mu_i + A)^{-1} = \frac{E(-\mu_i)}{\psi(-\mu_i)}.$$

Hence, with B simple and a unique solution, we may write

$$X = \sum_{i=1}^{t} \frac{E(-\mu_i)CF(\mu_i)}{\psi(-\mu_i)\phi^{(1)}(\mu_i)}.$$

This generalizes considerably solutions given by Givens [3] in the case of Lyapunov's equation with no repeated eigenvalues in the coefficient matrix, A.

12. A priori bounds on solutions of $AX + XB = C$. We present bounds in cases related to the stability problem. The first result is due to E. Heinz [4].

LEMMA 1. *If* $a = \max \sigma((A + A^*)/2)$, *then in the spectral norm we have*

$$\|e^A\| \leqq e^a.$$

The reader is referred to the paper of Heinz for proof.

LEMMA 2. *If* $a = \max \sigma((A + A^*)/2)$, *then for all* $\lambda \in \sigma(A)$ *we have* $\operatorname{Re}(\lambda) \leqq a$.

Proof. Let $A\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{x}^*\mathbf{x} = 1$. Then $\mathbf{x}^*A^* = \mathbf{x}^*\bar{\lambda}$ and

$$\mathbf{x}^*\left(\frac{1}{2}(A + A^*)\right)\mathbf{x} = \frac{1}{2}(\mathbf{x}^*A\mathbf{x} + \mathbf{x}^*A^*\mathbf{x}) = \frac{1}{2}(\lambda + \bar{\lambda}) = \operatorname{Re}(\lambda).$$

We then obtain $\operatorname{Re}(\lambda) \leqq a$ from the extreme value of the Rayleigh quotient for $(A + A^*)/2$.

THEOREM 9 (E. Heinz). *If* $a = \max \sigma((A + A^*)/2)$, $b = \max \sigma((B + B^*)/2)$ *and* $a + b < 0$, *then* (27) *obtains and, in the spectral norm,*

$$\|X\| \leqq |a + b|^{-1}\|C\|.$$

Proof. First, it follows from Lemma 2 that for $\lambda \in \sigma(A)$, $\operatorname{Re}(\lambda) \leqq a$ and for $\mu \in \sigma(B)$, $\operatorname{Re}(\mu) \leqq b$. Hence, the hypothesis of Theorem 5 follows from $a + b < 0$ (and $P_1 = P_2 = I$). The integral of (27) is therefore the unique solution. We now have

$$\|X\| \leqq \int_0^\infty \|e^{At}\|\,\|e^{Bt}\|\,dt \cdot \|C\|$$

and from Lemma 1,

$$\int_0^\infty \|e^{At}\|\,\|e^{Bt}\|\,dt \leqq \int_0^\infty e^{(a+b)t}\,dt = \frac{1}{|a + b|}.$$

The bound of Heinz' theorem will coincide with a similar bound obtained from a theorem of Smith [17] when $a = b$. When $a \neq b$, Heinz' theorem is the stronger. As Smith points out this kind of estimate may find some application in estimating and improving approximations to the solution of $AX + XB = C$. Thus, suppose the approximation is $X + \Delta$. The discrepancy

$$D = A(X + \Delta) + (X + \Delta)B - C = A\Delta + \Delta B$$

is computed and the error Δ satisfies an equation involving the same operator as that for X itself. If enough is known about A and B, the theorem can be used to obtain an estimate for the norm of the error, $\|\Delta\|$, in terms of $\|D\|$.

We next present some bounds for the spectrum of the solution matrix X of the Lyapunov equation in the case that the coefficient matrix, A, is stable. In this case it is known (and will be proved in § 13) that if W is a positive definite matrix then the solution X of

$$AX + XA^* = -W \tag{39}$$

is also positive definite. For this result it is convenient to write, for *any* square matrix A,

$$m(A) = \min \operatorname{Re}(\lambda), \qquad M(A) = \max \operatorname{Re}(\lambda),$$

the minimum and maximum being taken over the spectrum of A. The following results can be found in a paper by Smith [16].

THEOREM 10. *If W is positive definite, and A is stable, then for the solution of* (39) *we have*

$$\frac{m(W)}{|m(A + A^*)|} \leqq m(X) \leqq \frac{M(W)}{2|m(A)|}, \tag{40}$$

$$\frac{m(W)}{2|M(A)|} \leqq M(X) \leqq \frac{M(W)}{|M(A + A^*)|}. \tag{41}$$

Proof. Let $\mathbf{x}^*A = \lambda\mathbf{x}^*$ with $\mathbf{x} \neq \mathbf{0}$. Then $A^*\mathbf{x} = \bar{\lambda}\mathbf{x}$ and multiplying (39) on left and right by $\mathbf{x}^*$, $\mathbf{x}$ respectively, we obtain

$$(\lambda + \bar{\lambda})\mathbf{x}^*X\mathbf{x} = -\mathbf{x}^*W\mathbf{x}.$$

Since W is positive definite, X is also positive definite and $\operatorname{Re}(\lambda) < 0$. We have

$$\mathbf{x}^*X\mathbf{x} = -\frac{\mathbf{x}^*W\mathbf{x}}{2\operatorname{Re}(\lambda)},$$

and using the extremal properties of the Rayleigh quotient we obtain

$$m(X) \leqq \frac{M(W)}{2|m(A)|} \quad \text{and} \quad M(X) \geqq \frac{m(W)}{2|M(A)|}$$

which gives one inequality for each of (40) and (41).

Now let $X\mathbf{y} = \mu\mathbf{y}$ with $\mathbf{y} \neq 0$. Then $\mathbf{y}^*X = \mu\mathbf{y}^*$, $\mu > 0$, and

$$\mu\mathbf{y}^*(A + A^*)\mathbf{y} = -\mathbf{y}^*W\mathbf{y}.$$

Thus, $\mathbf{y}^*(A + A^*)\mathbf{y} < 0$. Choosing $\mu = M(X)$ we obtain

$$M(X) \leqq \frac{M(W)}{|M(A + A^*)|},$$

and choosing $\mu = m(X)$ we get

$$m(X) \geqq \frac{m(W)}{|m(A + A)^*|}.$$

This completes the four inequalities (40) and (41).

13. A theorem on the inertia of a matrix. We wish finally to present Krein's proof of an important theorem due to Ostrowski and Schneider [10], although we simplify Krein's result here to the matrix case.

For any $M \in C_{n \times n}$ we denote the fact that M is positive (negative) definite by writing $M \gg 0$ ($\ll 0$).

Note first of all that if the Lyapunov equation

$$AX + XA^* = C,$$

with C Hermitian, has a solution then this may be supposed to be Hermitian.

For we also have

$$X^*A^* + AX^* = C,$$

and adding these equations we find that $(X + X^*)/2$ is also a solution and is Hermitian.

THEOREM 11. *There exists a Hermitian matrix W such that $AW + WA^* \ll 0$ if and only if A has no pure imaginary eigenvalues.*

Proof. Suppose first that A has no pure imaginary eigenvalues. We introduce projections P_+, P_- defined by A and the parts of $\sigma(A)$ in the right and left half-planes, respectively.

Suppose $H \gg 0$ and consider the equation

$$-AX - XA^* = P_+HP_+^*.$$

We apply Theorem 5, taking σ_1 as the set of eigenvalues of $-A$ in the left half-plane. Then $P_+ = P_1$, $\sigma_2 = \bar{\sigma}_1$ and $P_+^* = P_2$. Hence $\mathrm{Re}(\lambda + \mu) < 0$ for all $\lambda \in \sigma_1$, $\mu \in \sigma_2$ and we can write

$$X = -\int_0^\infty e^{-At}P_+HP_+^*e^{-A^*t}\,dt.$$

Similarly it is found that a solution of

$$AY + YA^* = -P_-HP_-^*$$

is

$$Y = \int_0^\infty e^{At}P_-HP_-^*e^{A^*t}\,dt.$$

If we define the Hermitian matrix $W = X + Y$, then adding the equations for X and Y we have $AW + WA^* = -H_1$, where $H_1 = P_+HP_+^* + P_-HP_-^*$. We have finished this part of the proof if we can prove that $H_1 \gg 0$. But this is the case, for if $\mathbf{x} \in C_n$ and $\mathbf{x} \neq \mathbf{0}$ we can write $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 = P_+^*\mathbf{x}$, $\mathbf{x}_2 = P_-^*\mathbf{x}$. Then

$$\mathbf{x}^*H_1\mathbf{x} = \mathbf{x}^*P_+HP_+^*\mathbf{x} + \mathbf{x}^*P_-HP_-^*\mathbf{x} = \mathbf{x}_1^*H\mathbf{x}_1 + \mathbf{x}_2^*H\mathbf{x}_2$$

and this is positive since $H \gg 0$ and $x \neq 0$ implies that $\mathbf{x}_1$ and $\mathbf{x}_2$ cannot be zero simultaneously.

Conversely, suppose there exists a Hermitian W such that $AW + WA^* = -2H$ and $H \gg 0$ and that $\mathbf{x}^*A = i\mu\mathbf{x}^*$ with μ real and $\mathbf{x} \neq 0$. Then $A^*\mathbf{x} = -i\mu\mathbf{x}$ and

$$i\mu\mathbf{x}^*W\mathbf{x} - i\mu\mathbf{x}^*W\mathbf{x} = -2\mathbf{x}^*H\mathbf{x}.$$

But this is impossible because the left-hand side is zero and the right-hand side is negative. So we have a contradiction, and A has no pure imaginary eigenvalue.

Ostrowski and Schneider define the *inertia* of a matrix M to be the triple of nonnegative integers (π, ν, δ), where π is the number of eigenvalues of M with positive real parts, there are ν with negative real parts, and δ eigenvalues are pure imaginary. We now complete the Ostrowski–Schneider result with the following theorem.

THEOREM 12. *A Hermitian solution W of $AW + WA^* = -H$, where $H \gg 0$, has the same inertia as $-A$.*

Proof. Let $T = f(A)$, where $f(z) = (z + a)(z - a)^{-1}$ and $a > 0$. This is the transformation used in § 3 and we see that our equation for W transforms to (see (6)):

$$W - TWT^* = \frac{1}{2a}(T - I)H(T^* - I).$$

Since W exists we have from Theorem 11 that A has no pure imaginary eigenvalues and so T has no eigenvalues on the unit circle. Thus $T - I$ is nonsingular and $H \gg 0$ implies that

$$TWT^* - W \ll 0.$$

Furthermore, $(zI - T)^{-1}$ exists for all z with $|z| = 1$. It is easily verified that

$$TWT^* - |z|^2 W = (zI - T)\{W - z(zI - T)^{-1}W - \bar{z}W(\bar{z}I - T^*)^{-1}\}(\bar{z}I - T^*),$$

and so on the circle $|z| = 1$ we have

$$W - z(zI - T)^{-1}W - \bar{z}W(\bar{z}I - T^*)^{-1} \ll 0. \tag{42}$$

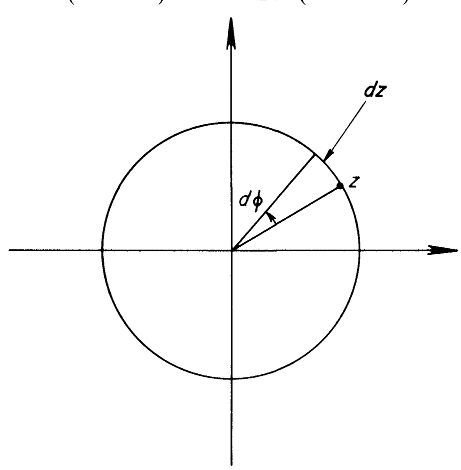

FIG. 2. *The unit circle in the complex z-plane*

On $|z| = 1$ we have

$$z\,d\phi = \frac{1}{i}\,dz, \qquad \bar{z}\,d\phi = -\frac{1}{i}\,d\bar{z},$$

so integrating (42) with respect to ϕ from 0 to 2π we obtain

$$2\pi W - \frac{1}{i}\oint (zI - T)^{-1}\,dz \cdot W + \frac{1}{i}W\oint (zI - T^*)^{-1}\,d\bar{z} \ll 0,$$

or

$$W - P_- W - WP_-^* \ll 0, \tag{43}$$

where P_- is the projection defined by

$$P_- = \frac{1}{2\pi i}\oint (zI - T)^{-1}\,dz.$$

Thus P_- is the projection defined by T and the part of $\sigma(T)$ inside the unit circle. But this part of $\sigma(T)$ is just the image of that part of the spectrum of A in the left half-plane. It follows that P_- is also the projection defined by A and the part of its spectrum with negative real part. We then deduce that $P_+ = I - P_-$ is the projection defined by A and the part of its spectrum in the right half-plane.

Now, if $\mathbf{x} \in \mathscr{R}(P_+^*)$, we have $P_-^*\mathbf{x} = \mathbf{0}$, and from (43) we obtain $\mathbf{x}^*W\mathbf{x} < 0$.

Then if $\mathbf{x} \in \mathscr{R}(P_-^*)$, we have $P_-^*\mathbf{x} = \mathbf{x}$, and this time (43) yields $\mathbf{x}^*W\mathbf{x} > 0$.

These results imply: (i) that the number of negative eigenvalues of W (counted according to multiplicities) is equal to the dimension of the range of P_+^* and (ii) the number of positive eigenvalues of W is the dimension of the range of P_-^*. From the above characterization of P_- and P_+ we see that the number of negative (positive) eigenvalues of W is equal to the number of eigenvalues of A with positive (negative) real parts. That is, the inertia of W coincides with the inertia of $-A$.

COROLLARY (Lyapunov theorem). *The matrix A is stable if and only if there is a positive definite W such that $AW + WA^* \ll 0$.*

Proof. If we are given that A is stable, then from Theorem 11 we deduce the existence of a Hermitian W with $AW + WA^* \ll 0$, and from Theorem 12 we deduce that W has the inertia of $-A$. That is, W is positive definite.

Conversely, if we are given the existence of $W \gg 0$ with $AW + WA^* \ll 0$, then we apply Theorem 12 at once and see that A must be stable.

REFERENCES

[1] N. DUNFORD AND J. T. SCHWARZ, *Linear Operators, Part I*, Interscience, New York, 1966.

[2] F. R. GANTMACHER, *The Theory of Matrices*, vol. 1, Chelsea, New York, 1960.

[3] W. GIVENS, *Elementary divisors and some properties of the Lyapunov mapping $X \rightarrow AX + XA^*$*, Argonne National Laboratory, Illinois, 1961.

[4] E. HEINZ, *Beiträge zur Störungstheorie der Spektralzerlegung*, Math. Ann., 123 (1951), pp. 415–438.

[5] A. S. HOUSEHOLDER, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

[6] A. JAMESON, *Solution of the equation $AX + XB = C$ by inversion of an $M \times M$ or $N \times N$ matrix*, SIAM J. Appl. Math., 16 (1968), pp. 1020–1023.

[7] M. G. KREIN, *Lectures on Stability Theory in the Solution of Differential Equations in a Banach Space*, Inst. of Math., Ukrainian Acad. Sci., 1964. (In Russian.)

[8] P. LANCASTER, *Theory of Matrices*, Academic Press, New York, 1969.

[9] E. C. MA, *A finite series solution of the matrix equation $AX - XB = C$*, SIAM J. Appl. Math., 14 (1966), pp. 490–495.

[10] C. C. MACDUFFEE, *The Theory of Matrices*, Chelsea, New York, 1956.

[11] A. OSTROWSKI AND H. SCHNEIDER, *Some theorems on the inertia of general matrices*, J. Math. Anal. Appl., 4 (1962), pp. 72–84.

[12] M. ROSENBLUM, *On the operator equation $BX - XA = Q$*, Duke Math. J., 23 (1956), pp. 263–270.

[13] ———, *The operator equation $BX - XA = Q$ with self-adjoint A and B*, Proc. Amer. Math. Soc., 20 (1969), pp. 115–120.

[14] W. E. ROTH, *The equations $AX - YB = C$ and $AX - XB = C$ in matrices*, Ibid., 3 (1952), pp. 392–396.

[15] D. E. RUTHERFORD, *On the solution of the matrix equation $AX + XB = C$*, Nederl. Akad. Wetensch. Proc. Ser. A, 35 (1932), pp. 53–59.

[16] R. A. SMITH, *Bounds for quadratic Lyapunov functions*, J. Math. Anal. Appl., 12 (1965), pp. 425–435.

[17] ———, *Matrix equations $XA + BX = C$*, SIAM J. Appl. Math., 16 (1968), pp. 198–201.

[18] A. TRAMPUS, *A canonical basis for the matrix transformation $X \rightarrow AXB$*, J. Math. Anal. Appl., 14 (1966), pp. 153–160.

[19] ———, *A canonical basis for the matrix transformation $X \rightarrow AX + XB$*, Ibid., 14 (1966), pp. 242–252.

[20] J. H. M. WEDDERBURN, *Note on the linear matrix equation*, Proc. Edinburgh Math. Soc., 22 (1904), pp. 49–53.