GENERALIZATIONS OF A WELL-KNOWN RESULT IN MATRIX THEORY

by R. C. THOMPSON

(Received 5 March, 1964)

Let A and C be $m \times m$ matrices and let B and D be $n \times n$ matrices, all with elements in a field F. Let A^T denote the transpose of A. A well-known theorem states that, if every $m \times m$ matrix X for which $AX = XA$ also satisfies $CX = XC$, then $C = \phi(A)$ for some polynomial $\phi(\lambda)$. In this note we establish the following simple generalizations.

THEOREM 1. *Let A and B have the same minimal polynomial $m(\lambda)$. If each $m \times n$ matrix X over F for which $AX = XB$ also satisfies $CX = XD$, then $C = \phi(A)$ and $D = \phi(B)$ for a polynomial $\phi(\lambda)$ over F.*

THEOREM 2. *Let $n = m$. If each symmetric $m \times m$ matrix X over F for which $AX = XA^T$ also satisfies $CX = XD$, then $D^T = C = \phi(A)$ for a polynomial $\phi(\lambda)$ over F.*

These results may be proved as easily as the classical result. It is possible to base the proofs on the Jordan canonical form under similarity by first extending F to a field K in which $m(\lambda)$ splits, next showing that the hypotheses will still be valid if X is permitted to have elements in K, then transforming A and B to their Jordan forms and hence establishing the existence of $\phi(\lambda)$ over K, and finally showing that $\phi(\lambda)$ may be taken to have coefficients in F. We give proofs based on the rational canonical form under similarity.

Let $f(\lambda) = f_r\lambda^r + \ldots + f_0$ and $g(\lambda) = g_s\lambda^s + \ldots + g_0$ be two monic ($f_r = g_s = 1$) nonconstant polynomials. Let $C(f(\lambda))$ denote the companion matrix of $f(\lambda)$; it is defined in [**1**, p. 148]. A persymmetric matrix is one constant along each diagonal perpendicular to the main diagonal. Let $\overline{C}$ be a matrix with r columns and let $\overline{D}$ be a matrix with s rows.

LEMMA. *Each $r \times s$ matrix X satisfying*

$$C(f(\lambda))X = XC(g(\lambda))^T \tag{1}$$

is persymmetric. If $f(\lambda)$ divides $g(\lambda)$ and if the first column of $\overline{C}X$ is zero for each X over F satisfying (1), *then $\overline{C} = 0$. If $g(\lambda)$ divides $f(\lambda)$ and if the first row of $X\overline{D}$ is zero for each X over F satisfying* (1), *then $\overline{D} = 0$.*

Proof. Let $X = (x_{ij})$. Comparing the (i, j) elements of the two sides of (1) for $i < r$ and $j < s$, we have $x_{i+1,\, j} = x_{i,\, j+1}$. Hence X is persymmetric and we may set $x_{ij} = x_{i+j-2}$ for $1 \leqq i \leqq r$ and $1 \leqq j \leqq s$. Let $g(\lambda)/f(\lambda) = h(\lambda) = h_t\lambda^t + \ldots + h_0$. For (1) to hold it is necessary and sufficient that

$$\sum_{\alpha=0}^{r} f_\alpha x_{\alpha+\beta} = 0 \quad \text{for all } \beta \geqq 0 \text{ and } \leqq s-2, \tag{2a}$$

$$\sum_{\alpha=0}^{s} g_\alpha x_{\alpha+\beta} = 0 \quad \text{for all } \beta \geqq 0 \text{ and } \leqq r-2, \tag{2b}$$

$$\sum_{\alpha=0}^{s-r-1} g_\alpha x_{\alpha+r-1} + \sum_{\alpha=0}^{r-1} (g_{\alpha+s-r} - f_\alpha) x_{\alpha+s-1} = 0. \tag{2c}$$

https://doi.org/10.1017/S2040618500035127 Published online by Cambridge University Press

If $r=1$ there are no equations (2b). Equation (2c) is an identity if $r=s$. The polynomial identity $g(\lambda)=f(\lambda)\,h(\lambda)$ implies relations among the g_i, f_i, h_i which in turn imply that

$$\sum_{\alpha=0}^{s} g_\alpha x_{\alpha+\beta} = \sum_{\mu=0}^{t} h_\mu \left(\sum_{\alpha=0}^{r} f_\alpha x_{\alpha+\mu+\beta} \right) \quad \text{for all } \beta \geqq 0 \text{ and } \leqq r-2.$$

Hence each equation (2b) is a linear combination of some of equations (2a). Since $f(\lambda)$ also divides $g(\lambda)-\lambda^{s-r}f(\lambda)$, equation (2$c$) is, in a like manner, a linear combination of some of equations (2a). Hence (1) will hold if and only if (2a) is satisfied. From the form of (2a) it follows that $x_0, x_1, \ldots, x_{r-1}$ are independent variables with $x_r, \ldots, x_{r+s-2}$ determined in terms of these independent variables. Then, if $\overline{C}_i$ is column i of $\overline{C}$, the first column of $\overline{C}X$ is $\overline{C}_1 x_0 + \ldots + \overline{C}_r x_{r-1}$, which can be zero for all choices of $x_0, \ldots, x_{r-1}$ in F only if $\overline{C}=0$. The proof of the other case is similar.

We now give together the proofs of Theorems 1 and 2. Let $f_1(\lambda), f_2(\lambda), \ldots, f_a(\lambda)$ denote the nontrivial invariant factors of $\lambda I_m - A$ and let $g_1(\lambda), g_2(\lambda), \ldots, g_b(\lambda)$ denote the nontrivial invariant factors of $\lambda I_n - B$. Each $f_i(\lambda)$ and each $g_i(\lambda)$ divides the common minimal polynomial $m(\lambda)=f_a(\lambda)=g_b(\lambda)$ of A and B.

In Theorem 2 we have $B=A^T$, $b=a$, and $g_i(\lambda)=f_i(\lambda)$ for all $i \leqq a$. Nonsingular matrices S and T exist over F such that

$$SAS^{-1} = C(f_1(\lambda)) \dotplus C(f_2(\lambda)) \dotplus \ldots \dotplus C(f_a(\lambda)) = A_1, \text{ say,}$$

and

$$TBT^{-1} = C(g_1(\lambda))^T \dotplus C(g_2(\lambda))^T \dotplus \ldots \dotplus C(g_b(\lambda))^T = B_1, \text{ say.}$$

Here $\dotplus$ denotes direct sum. For Theorem 2 take $T^{-1}=S^T$. Let $C_1=SCS^{-1}$, $X_1=SXT^{-1}$, $D_1=TDT^{-1}$. In Theorem 2, X_1 is symmetric if and only if X is. Then $AX=XB$, $AX=XA^T$, $CX=XD$ will hold if and only if, respectively, $A_1X_1=X_1B_1$, $A_1X_1=X_1A_1^T$, $C_1X_1=X_1D_1$. Partition the matrices C_1, X_1, D_1 into the forms $C_1=(C_{\alpha\beta})$, $X_1=(X_{\alpha\beta})$, $D_1=(D_{\alpha\beta})$, where $C_{\alpha\beta}$ is (degree $f_\alpha(\lambda)$) $\times$ (degree $f_\beta(\lambda)$), $X_{\alpha\beta}$ is (degree $f_\alpha(\lambda)$) $\times$ (degree $g_\beta(\lambda)$), $D_{\alpha\beta}$ is (degree $g_\alpha(\lambda)$) $\times$ (degree $g_\beta(\lambda)$).

For Theorem 1 set all $X_{\alpha\beta}$ equal to zero except for X_{ib} for one fixed $i \leqq a$. Then $A_1X_1=X_1B_1$ will hold if $C(f_i(\lambda))X_{ib}=X_{ib}C(g_b(\lambda))^T$. From $C_1X_1=X_1D_1$ follows $C_{ji}X_{ib}=0$ if $j \neq i$. Hence $C_{ji}=0$. Next set all $X_{\alpha\beta}$ equal to zero except for X_{aj} for one fixed $j \leqq b$. Then $A_1X_1=X_1B_1$ will hold if $C(f_a(\lambda))X_{aj}=X_{aj}C(g_j(\lambda))^T$. From $C_1X_1=X_1D_1$ follows $X_{aj}D_{ji}=0$ for $i \neq j$. Hence $D_{ji}=0$.

For Theorem 2 set all $X_{\alpha\beta}$ equal to zero except for X_{jj} for one fixed $j \leqq a$. Then $A_1X_1=X_1A_1^T$ will hold if $C(f_j(\lambda))X_{jj}=X_{jj}C(f_j(\lambda))^T$. So X_{jj} and therefore X_1 is symmetric. Then $C_1X_1=X_1D_1$ implies that $C_{ij}X_{jj}=0$ and $X_{jj}D_{ji}=0$ for $i \neq j$. Therefore $C_{ij}=0$ and $D_{ji}=0$. Thus, in both Theorems, C_1 and D_1 are block diagonal.

For the moment suppose that the first row of the last diagonal block C_{aa} of C_1 is zero. Set all $X_{\alpha\beta}$ equal to zero except for X_{ab}. Then $A_1X_1=X_1B_1$ or $A_1X_1=X_1A_1^T$ will hold if $C(f_a(\lambda))X_{ab}=X_{ab}C(g_b(\lambda))^T$. Thus X_1 is symmetric in the case $B=A^T$. From $C_1X_1=X_1D_1$ we get $C_{aa}X_{ab}=X_{ab}D_{bb}$. Since the first row of $C_{aa}X_{ab}$ is zero, the first row of $X_{ab}D_{bb}$ is zero also; hence $D_{bb}=0$, and then $C_{aa}=0$. Next set all $X_{\alpha\beta}$ equal to zero except for X_{ib} for one fixed $i < a$ and X_{aj} for one fixed $j < b$. For Theorem 2 take $i = j$ and $X_{ai} = X_{ia}^T$ so as to make X_1 symmetric. Then $A_1X_1 = X_1A_1^T$ will hold if $C(f_i(\lambda))X_{ib} = X_{ib}C(g_b(\lambda))^T$ and $A_1X_1 = X_1B_1$ will hold if also $C(f_a(\lambda))X_{aj} = X_{aj}C(g_j(\lambda))^T$. Then $C_1X_1 = X_1D_1$ yields $C_{ii}X_{ib} = 0$ and $X_{aj}D_{jj} = 0$ (because $C_{aa} = 0$ and $D_{bb} = 0$). Thus $C_{ii} = 0$ and $D_{jj} = 0$ and so both C_1 and D_1 are zero.

Now, for a fixed i with $0 \leqq i <$ degree $m(\lambda)$, the first row of $[C(f_a(\lambda))]^i$ is entirely zero except for a single 1 at column $i+1$. Thus it is possible to find a polynomial $\phi(\lambda)$ over F such that the first row of $C_{aa} - \phi(C(f_a(\lambda))$ is entirely zero. It follows from $A_1X_1 = X_1B_1$ that $A_1^kX_1 = X_1B_1^k$ for $k = 0, 1, 2, \ldots$; hence $\phi(A_1)X_1 = X_1\phi(B_1)$, and therefore $(C_1 - \phi(A_1))X_1 = X_1(D_1 - \phi(B_1))$. Since the first row of the last diagonal block of $C_1 - \phi(A_1)$ is zero, it now follows that $C_1 - \phi(A_1) = 0$ and $D_1 - \phi(B_1) = 0$. Thus $C = \phi(A)$ and $D = \phi(B)$ as required.

COROLLARY. *If each X over F for which $AX = XA$ also satisfies $CX = XD$ then $C = D = \phi(A)$. If A is symmetric and each symmetric X over F for which $AX = XA$ also satisfies $CX = XD$, then $C = D = \phi(A)$ and so $C = D$ is symmetric.*

The last statement follows because if $A = A^T$ then $AX = XA$ is the same as $AX = XA^T$ so $C = \phi(A)$ and $D = \phi(A^T) = \phi(A) = C$.

REFERENCE

1. S. Perlis, *The theory of matrices* (Cambridge, Mass., 1952).

THE UNIVERSITY OF BRITISH COLUMBIA
(Now at the University of California, Santa Barbara)