On a Classical Commutator Result

MARVIN MARCUS¹

&

ROBERT C. THOMPSON²

Communicated by J. TODD

1. Introduction. It is a classical result due to Frobenius [2; see also 1, 3] that if A and B are unitary matrices (both n -square) such that the eigenvalues of B cover an arc less than a semicircle on the boundary of the unit circle, then the conditions

$$(1) \qquad C = ABA^{-1}B^{-1}, \quad CA = AC$$

imply $C = I_n$, the n -square identity matrix. Recently the equations (1) were reexamined (for unitary A and B) by O. Taussky [4], H. Zassenhaus [8], and H. Wielandt [7]. In this paper we examine the equations (1) under the hypothesis that A and C are normal matrices. In Theorem 1 we present a result substantially more general than the classical result, for which the proof is extremely short. Theorem 2 contains another generalization of the classical result in which B is assumed to be normal. In Theorem 3 we derive explicit formulas for all solutions of (1). These formulas were first noticed by Wielandt [7]. We use these formulas to derive in Theorems 4, 5, 6 certain inequalities involving the multiplicity of the eigenvalue $\gamma = 1$ of C in (1). As Corollary 1 to Theorem 6 we derive an inequality which contains as a special case the known fact [1] that if (1) holds with A and B unitary then $\|I_n - C\| < 2$ implies $C = I_n$. (See Theorem 6 for the definition of $\|X\|$.)

2. Notation. The field of values of the matrix B , $F(B)$, is the set of all complex numbers x^*Bx where $x^*x = 1$. Here x is a column n -tuple, and $*$ denotes the complex conjugate transpose. It is well known that $F(B)$ is a bounded, closed, convex set in the complex plane. When B is normal, $F(B)$ is precisely the convex hull of the eigenvalues of B . We let $\Delta(B)$ denote the distance from the origin to the most distant point in $F(B)$. For normal B , $\Delta(B)$ is just the

---

^{1,2} The research of both authors was supported by U. S. Air Force Grant AFOSR 698-65.

spectral radius of B . For unitary B , $\Delta(B) = 1$. If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of B and $1 \leq k \leq n$, we let $\Gamma_k = \Gamma_k(B)$ denote the convex hull of all sums

$$(2) \quad k^{-1}(\lambda_{i_1} + \lambda_{i_2} + \dots + \lambda_{i_k}), \quad 1 \leq i_1 < i_2 < \dots < i_k \leq n.$$

We let (x, y) denote the inner product x^*y in unitary n -space. We let 0_{mn} denote the $m \times n$ matrix of zeros. If $m = n$ we write 0_n for 0_{nn} . The transpose of the matrix P is P^T .

3. Results.

Theorem 1. *Let A, B , and C satisfy (1). If A and C are normal and $0 \notin F(B)$, then $C = I_n$.*

Proof. A simultaneous unitary similarity of A, B, C leaves (1) unaffected. Because $AC = CA$, we can assume that both A and C are diagonal. Let $A = \text{diag}(\alpha_1, \dots, \alpha_n)$, $C = \text{diag}(\gamma_1, \dots, \gamma_n)$, $B = (b_{ij})$. Then (1) implies $CBA = AB$, hence $\gamma_i \alpha_i b_{ii} = \alpha_i b_{ii}$. Thus $b_{ii} \alpha_i (\gamma_i - 1) = 0$. Since $b_{ii} \notin F(B)$ and $0 \notin F(B)$, we get $b_{ii} \neq 0$. Thus $\gamma_i = 1$ for all i , and $C = I_n$.

In order to state Theorem 2, we require the following Lemma.

Lemma. *Let B be normal. Then $\Gamma_1 \supseteq \Gamma_2 \supseteq \dots \supseteq \Gamma_n$. If $\alpha \notin \Gamma_{r+1}$ and $(Bx_1, x_1) = (Bx_2, x_2) = \dots = (Bx_t, x_t) = \alpha$ for orthonormal vectors $x_1, \dots, x_t$ then $t \leq r$.*

Proof. To show $\Gamma_k \subseteq \Gamma_{k-1}$ it will suffice to show that the points (2) lie in Γ_{k-1} . This is clear from

$$\begin{aligned} k^{-1}(\lambda_{i_1} + \dots + \lambda_{i_k}) \\ = k^{-1} \left\{ \sum_{i=1}^k (\lambda_{i_1} + \dots + \lambda_{i_{t-1}} + \lambda_{i_{t+1}} + \dots + \lambda_{i_k}) (k-1)^{-1} \right\}. \end{aligned}$$

We give two proofs of the second assertion of the Lemma.

Let $e_1, \dots, e_n$ be an orthonormal set of eigenvectors of B belonging respectively, to the eigenvalues $\lambda_1, \dots, \lambda_n$. Let $x_1, \dots, x_n$ be orthonormal vectors. Then

$$(Bx_i, x_i) = \sum_{\alpha=1}^n \lambda_{\alpha} |(x_i, e_{\alpha})|^2,$$

and hence

$$\sum_{i=1}^k (Bx_i, x_i) = \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{i=1}^k |(x_i, e_{\alpha})|^2.$$

Let S be the doubly stochastic matrix $(|(x_i, e_{\alpha})|^2)_{1 \leq i, \alpha \leq n}$.

It is a standard result due to Birkhoff that S is a convex combination of permutation matrices. Hence let $S = \sum_{\sigma} \theta_{\sigma} P_{\sigma}$ be a convex combination of permutation matrices P_{σ} , where θ_{σ} are nonnegative numbers summing to one.

Let $e = (1, 1, \dots, 1, 0, \dots, 0)^T$ be a column n -tuple with the first k entries equal to 1, the rest zero. Let $\Lambda = (\lambda_1, \dots, \lambda_n)$. Then

$$\begin{aligned}\sum_{i=1}^k (Bx_i, x_i) &= \Lambda Se \\ &= \sum_{\sigma} \theta_{\sigma} \Lambda P_{\sigma} e.\end{aligned}$$

Since $\Lambda P_{\sigma} e$ is of the form

$$(2') \quad \lambda_{i_1} + \dots + \lambda_{i_k}, \quad 1 \leq i_1 < \dots < i_k \leq n,$$

this shows that $(Bx_1, x_1) + \dots + (Bx_k, x_k)$ lies in the convex hull of the numbers (2'). If $(Bx_1, x_1) = \dots = (Bx_k, x_k) = \alpha$, we immediately see that $\alpha \in \Gamma_k$. Hence $k \leq r$.

Another proof that $(Bx_1, x_1) + \dots + (Bx_k, x_k)$ lies in the convex hull of the numbers (2') is given in [6].

Theorem 2. Let A, B and C be normal and satisfy (1). Suppose $0 \notin \Gamma_{r+1}(B)$. Then $\gamma = 1$ is an eigenvalue of C of multiplicity at least $n - r$.

Proof. As in Theorem 1, we let $C = \text{diag}(\gamma_1, \dots, \gamma_n)$, $A = \text{diag}(\alpha_1, \dots, \alpha_n)$, $B = (b_{ij})$. Then we get $b_{ii}\alpha_i(\gamma_i - 1) = 0$. Since there exist n orthonormal vectors $u_1, \dots, u_n$ such that $b_{ii} = (Bu_i, u_i)$, it follows that if $b_{ii} = 0$ for k values of i , then $k \leq r$. Whenever $b_{ii} \neq 0$, $\gamma_i = 1$. Hence $\gamma_i = 1$ for at least $n - k$ values i . But $n - k \geq n - r$.

Theorem 3. Suppose C and A are normal and that (1) holds. After a simultaneous unitary similarity of A, B , and C , we may assume that

$$(3) \quad A = \tilde{\alpha}_1 I_{n_1} \oplus \tilde{\alpha}_2 I_{n_2} \oplus \dots \oplus \tilde{\alpha}_s I_{n_s}$$

with $\tilde{\alpha}_1, \dots, \tilde{\alpha}_s$ distinct, and that C is diagonal. Then

$$(4) \quad B = PW,$$

$$(5) \quad C = APA^{-1}P^{-1},$$

where P is a permutation matrix and where $W = W_1 \oplus \dots \oplus W_s$ is the direct sum of n_i -square matrices W_i , $1 \leq i \leq s$. Conversely if matrices W_i are arbitrary n_i -square and nonsingular, and if C, A , and B are given by (3), (4), (5) with P a permutation matrix, then (1) holds and C is diagonal.

Proof. Let $\tilde{\alpha}_1, \dots, \tilde{\alpha}_s$ be the distinct eigenvalues of A , where $\tilde{\alpha}_i$ has multiplicity n_i , $1 \leq i \leq s$. After a simultaneous unitary similarity of C and A , we may assume A is given by (3) and $C = \text{diag}(\gamma_1, \dots, \gamma_n)$. It follows from $A^{-1}C = BA^{-1}B^{-1}$ and the fact that both $A^{-1}C$ and A^{-1} are diagonal that $A^{-1}C = PA^{-1}P^T$ for some permutation matrix P . Then (5) holds. But then $PA^{-1}P^T = BA^{-1}B^{-1}$ so that $(P^TB)A = A(P^TB)$. Hence P^TB commutes with A . This forces $P^TB = W_1 \oplus \dots \oplus W_s$ where W_i is n_i -square. Hence $B = P(W_1 \oplus \dots \oplus W_s) = PW$ where $W = W_1 \oplus \dots \oplus W_s$.

Conversely, if A , B and C are given by (3), (4), and (5), it follows that $ABA^{-1}B^{-1} = APWA^{-1}W^{-1}P^T = APA^{-1}P^T$, since $WA^{-1}W^{-1} = A^{-1}$. This is because A^{-1} partitions into scalar segments on its main diagonal conformally with the partitioning of W . Moreover $CA = APA^{-1}P^TA = A^2PA^{-1}P^T = AC$, since $PA^{-1}P^T$ and A commute (because each is diagonal). Hence (1) holds.

It will somewhat simplify the next discussion if we impose some structure on P in (3), (4), and (5). Let $Q = Q_1 \oplus \cdots \oplus Q_s$, $R = R_1 \oplus \cdots \oplus R_s$ be permutation matrices where the partitioning is conformal with the partitioning of W . Then Q and R commute with A . Perform a similarity transformation of (3), (4), (5) by Q . Then we get (3), and

$$(4') \quad QBQ^T = QPRR^TWQ^T,$$

$$(5') \quad QCQ^T = A(QPR)A^{-1}(QPR)^T.$$

If $P = (P_{ii})$ where P_{ii} is n_i -square, then $QPR = (Q_iP_{ii}R_i)$. We can choose Q_i and R_i as n_i -square permutation matrices such that all the ones in P_{ii} are moved onto the initial segment of the main diagonal in $Q_iP_{ii}R_i$. That is, $Q_iP_{ii}R_i = I_{m_i} \oplus 0_{n_i-m_i}$ for some integer m_i , $0 \leq m_i \leq n_i$. Now change notation and denote QPR by P , R^TWQ^T by W , QCQ^T by C , QBQ^T by B . Then (3), (4), (5) hold and $P = (P_{ii})$ with

$$(6) \quad P_{ii} = I_{m_i} \oplus 0_{n_i-m_i}, \quad 1 \leq i \leq s.$$

From here on, P is assumed to be in this form. We still have W in the form $W = W_1 \oplus \cdots \oplus W_s$ where W_i is n_i -square.

Now, let $B = (B_{ii})$, conformally with the partitioning of W . Then

$$(7) \quad B_{ii} = P_{ii}W_i = \begin{bmatrix} W'_i \\ 0_{n_i-m_i, n_i} \end{bmatrix}$$

where W'_i consists of the first m_i rows of W_i .

Second proof of Theorem 1. If $0 \notin F(B)$, no diagonal element of B is zero. Hence by (7), $n_i = m_i$. Hence by (6), $P = I_n$. Since W commutes with A , $C = I_n$.

Second proof of Theorem 2. Here B can have at most r zero diagonal entries. Then, by (7), $\sum_i (n_i - m_i) \leq r$, hence $n - r \leq \sum_i m_i$. Let σ be a permutation of $1, 2, \dots, n$ such that the one in row i of P appears at position $(i, \sigma(i))$. Then, from (5) and $C = \text{diag}(\gamma_1, \dots, \gamma_n)$,

$$(8) \quad \gamma_i = \alpha_i \alpha_{\sigma(i)}^{-1}, \quad 1 \leq i \leq n.$$

Let J_t denote the interval $[n_1 + \cdots + n_{t-1} + 1, n_1 + \cdots + n_t]$, $1 \leq t \leq s$. If $i \in J_t$ and $\sigma(i) \neq i$, then, by (6), $\sigma(i) \notin J_t$. The number of integers $i \in J_t$ such that $\sigma(i) \notin J_t$ is $n_t - m_t$. Whenever $i \in J_t$ and $\sigma(i) \notin J_t$, (8) shows that

$\gamma_i \neq 1$. Hence the number of eigenvalues of C different from one is $\Sigma_i n_i - m_i = n - \Sigma_i m_i$. Thus the number of eigenvalues of C equal to one is $\Sigma_i m_i \geq n - r$. This completes the second proof of Theorem 2, and in fact proves Theorem 4.

Theorem 4. *Let (1) hold with C and A normal. After a simultaneous unitary similarity of C , A , and B , we may assume that (3), (4), (5), and (6) hold, where $P = (P_{ij})$, $B = (B_{ij})$, and $W = W_1 \oplus \cdots \oplus W_s$ are partitioned conformally with the partitioning in (3). Then the multiplicity of $\gamma = 1$ as an eigenvalue of C is $\Sigma_i m_i$.*

Theorem 5. *Let C and A be fixed normal matrices, and let B be a variable matrix such that (1) holds. Then the multiplicity of $\gamma = 1$ as an eigenvalue of C is*

$$(9) \quad \max_B |\text{trace } B| \Delta(B)^{-1}.$$

Moreover, it is possible to attain this maximum with a unitary B .

Proof. From (7), B can have at most $\Sigma_i m_i$ nonzero diagonal elements. Each diagonal element b_{ii} of B lies in $F(B)$, hence $|b_{ii}| \leq \Delta(B)$. Thus

$$|\text{trace } B| \leq \Sigma_i |b_{ii}| \leq \Delta(B) \Sigma_i m_i,$$

hence

$$\Sigma_i m_i \geq |\text{trace } B| \Delta(B)^{-1}.$$

Since, by Theorem 4, the multiplicity of $\gamma = 1$ as an eigenvalue of C is $\Sigma_i m_i$, $|\text{trace } B| \Delta(B)^{-1}$ does not exceed the multiplicity of $\gamma = 1$ as an eigenvalue of C . If we put $W = I_n$ (as we may since the W_i are arbitrary nonsingular blocks), we get $B = P$, hence $\text{trace } B = \Sigma_i m_i =$ multiplicity of eigenvalue $\gamma = 1$ of C . This B is unitary. This shows that (9) has the claimed property.

Theorem 6. *Let C , A , and B satisfy (1), with C and A normal and B unitary. Let V be any unitary matrix commuting with A . Let p be the number of eigenvalues of C different from one (including multiplicities). Let $\|X\|^2 = \text{trace } XX^*$. Then*

$$(10) \quad p \leq \frac{1}{2} \|V - B\|^2.$$

Moreover, for each fixed unitary B for which (1) holds, a unitary V commuting with A exists such that equality holds in (10); and for each fixed unitary V commuting with A , there exists a unitary B satisfying (1) such that equality holds in (10).

Proof. Let C , A , B , and W be as described in Theorem 3 and let $P = (P_{ij})$ with (6) and (7) in force. Any unitary V commuting with A has the form $V = V_1 \oplus \cdots \oplus V_s$ conformally with the partitioning of A . Then

$$\begin{aligned} \|V - B\|^2 &= \text{trace } (V - B)(V^* - B^*) \\ &= 2n - 2 \text{Re trace } VB^*. \end{aligned}$$

But $VB^* = VW^*P^T$, hence $\text{trace } VB^* = \sum_i \text{trace } V_i W_i^* P_{ii}$. Since $V_i W_i^*$ is unitary, we get

$$\begin{aligned} (11) \quad ||V - B||^2 &\geq 2n - 2\sum_i m_i \\ &= 2(n - \sum_i m_i) \\ &= 2p. \end{aligned}$$

Moreover, when $VW^* = I_n$, equality will hold in (11). This completes the proof.

Corollary 1. *Let C and A be normal. If the distance d from a unitary B satisfying (1) to the multiplicative group of unitary matrices commuting with A satisfies $d < 2$, then $C = I_n$.*

Proof. If $||V - B|| < 2$, then $p < 2$. Hence $p = 1$ or $p = 0$. But $p = 1$ is impossible since C has determinant one. Thus $p = 0$ and hence $C = I_n$.

Finally, in Corollary 2, we state the classical result (extended to normal matrices).

Corollary 2. *Let (1) hold with C , A , and B normal. If the eigenvalues of B lie in a cone in the complex plane with apex at the origin and apex angle strictly less than π , then $C = I_n$.*

Proof. The hypothesis implies $0 \notin F(B)$. Use Theorem 1.

REFERENCES

- [1] CURTIS, C. W. & REINER, I., *Representation theory of finite groups and associative algebras*, Interscience (New York), 1962, p. 259–260.
- [2] FROBENIUS, G., *Über den von L. Bieberbach gefundenen Beweis eines Satzes von C. Jordan*, Sitzber. Preuss. Akad. Wiss. (1911), p. 241–248.
- [3] SPEISER, A., *Die Theorie der Gruppen*, Dover (New York), 1954, p. 217.
- [4] TAUSKY, O., Commutators of unitary matrices that commute with one factor, *J. Math. and Mech.*, 10 (1961) 175–178.
- [5] TAUSKY, O., Matrices with zero trace, *Amer. Math. Monthly*, 69 (1962) 40–42.
- [6] THOMPSON, R. C., Principal submatrices of normal and Hermitian matrices, *Illinois J. Math.*, 10 (1966) 296–308.
- [7] WIELANDT, H., review of [4] in *Zentralblatt für Mathematik und ihre Grenzgebiete*, 96 (1962) 9.
- [8] ZASSENHAUS, H., A remark on a paper of O. Taussky, *J. Math. and Mech.*, 10 (1961) 179–181.

University of California
Santa Barbara, California

Date Communicated: MARCH 29, 1966