

UNITARY BLOCK DIAGONALIZATION AND THE CHARACTERISTIC POLYNOMIAL
OF A PENCIL GENERATED BY HERMITIAN MATRICES

Thesis by
Helene Marian Shapiro

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1979

(Submitted May 21, 1979)

ACKNOWLEDGEMENTS

I am grateful to Professor Olga Taussky Todd, my thesis advisor, for her generous guidance and encouragement, and for devoting so much of her time and interest to my work. I am indebted to her for many ideas and suggestions, and for steering my interests in appropriate directions. It has been a privilege to work with her. I would also like to thank Professor Steven Kleiman for several helpful discussions, Dr. Robert Guralnick for his illuminating comments and suggestions, and Robert Calderbank for checking the examples.

Thanks are due to the mathematics department of Caltech for general academic support and interest, to my fellow graduate students, and to the secretaries in the math office for creating a pleasant atmosphere. I also thank them for allowing me to use the office typewriters to prepare this thesis.

I am grateful to the National Science Foundation and the California Institute of Technology for financial support.

To my parents, friends, and teachers, I would like to express my appreciation for everything they have done and been for me, and I would like to take this opportunity to thank the people of the Kenyon College mathematics department.

ABSTRACT

We consider the problem of simultaneously putting a set of square matrices into the same block upper triangular form with a similarity transformation, and obtain a result linking the size of the largest block to polynomial identities. This generalizes McCoy's theorem, which gives necessary and sufficient conditions for a set of matrices over an algebraically closed field to be simultaneously similar to upper triangular matrices.

A theorem of Specht states that when the algebra \mathcal{A} generated by a set of complex matrices satisfies the condition $\mathcal{A} = \mathcal{A}^*$, where $*$ denotes conjugate transpose, the algebra can be simultaneously, unitarily block diagonalized if and only if it can be simultaneously, block upper triangularized. Applying this to a single complex matrix A , we see that A can be unitarily block diagonalized if and only if A and A^* can be simultaneously block triangularized. Let $A = H + iK$, where H and K are Hermitian. Then A can be unitarily block diagonalized if and only if H and K can be simultaneously block upper triangularized.

Hence, to study the problem of unitarily, block diagonalizing the matrix A , we consider the pair of Hermitian matrices H and K . We study the pencil $xH + yK$ and the characteristic polynomial of this pencil, $f(x,y,z) = \det(zI - xH - yK)$. Motzkin and Taussky proved that H and K commute if and only if $f(x,y,z)$ factors into linear factors. If H and K can be simultaneously block diagonalized, then $f(x,y,z)$ splits into factors corresponding to the blocks, but examples show that the converse is not true. However, we prove that if $f(x,y,a)$ has a repeated linear factor of high enough multiplicity, then H and

H and K have a common eigenvector. We also show that if $f(x,y,z)$ is a power of a quadratic polynomial, then H and K are simultaneously, unitarily similar to block diagonal matrices, each of which is a sum of identical 2×2 blocks. This is a sharper version of a result of Kippenhahn.

Motzkin and Taussky studied the algebraic curve of degree n whose equation is $f(x,y,z) = 0$. Murnaghan and Kippenhahn independently showed that the curve whose equation in line coordinates is $f(x,y,z) = 0$ determines the numerical range of the matrix A . These geometric interpretations yield some information about $A = H + iK$ when $f(x,y,z)$ has only a linear factor and a quadratic factor.

Kippenhahn conjectured that if $f(x,y,z)$ has a repeated factor, then A is unitarily similar to a matrix which is block diagonal. We verify this conjecture for $n \leq 5$.

CONTENTS

Acknowledgements.	ii
Abstract.	iii
Introduction.	1
Chapter I. Simultaneous Block Triangularization of Sets of Matrices and Unitary Block Diagonalization of Sets of Matrices.	
1. Simultaneous triangularization of sets of matrices.	4
2. Block triangularization.	6
3. The main result of Chapter I--Theorem 1.	9
4. Polynomial identities.	13
5. Unitary block triangularization and block diagonalization.	15
6. Some applications of Specht's theorem B and Theorem 2.	19
Chapter II. Pencils of Matrices, the Characteristic Polynomial and Property L.	
1. Definition of the matrix pencil $xA + yB$ and the characteristic polynomial $f(x, y, z)$	23
2. The pencil $xH + yK$ where $A = H + iK$	24
3. Property L.	31
4. A generalization of property L.	33

Chapter III. Linear Factors of the Characteristic Polynomial.

1. Linear factors of $\det(zI - xA - yB)$ and the eigenvalues of $A^{-1}B$ 40
2. A pair of $n \times n$ Hermitian matrices, H and K , which do not have a common eigenvector but such that $\det(zI - xH - yK)$ does have a linear factor. 41
3. A theorem about linear factors of high multiplicity. 43
4. Examples related to Theorem 4. 50

Chapter IV. A Theorem About an Algebra Generated by a Pair of

- Matrices Which Satisfy Polynomials of Degree Two. 55

Chapter V. The Minimal Polynomial of a Pencil.

1. Definition of the minimal polynomial of a pencil and its relation to the characteristic polynomial. 58
2. The minimal polynomial of $xH + yK$ 59
3. The case where the minimal polynomial has degree two. 61

Chapter VI. Geometric Interpretations of $f(x, y, z) = 0$.

1. Geometric Background. 66
2. The method of Motzkin and Taussky--the characteristic curve. 68
3. The case $\det(zI - xH - yK) = (z - \alpha x - \beta y)^r [g(x, y, z)]^s$, where $g(x, y, z)$ has degree two. 69
4. Line coordinates, dual curves and foci. 72
5. The numerical range of a matrix. 74
6. The relation of $f(x, y, z) = \det(zI - xH - yK)$ to the numerical range of $A = H + iK$ 75

Chapter VII. A Conjecture of Kippenhahn.

1. Statement of the conjecture.	79
2. Reduction to the case $f(x, y, z) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2$	81
3. Two lemmas.	86
4. The proof that A can be unitarily block diagonalized.	92
5. Theorem 4 as a consequence of Kippenhahn's conjecture.	106
Appendix	108
References.	117

INTRODUCTION

Any square matrix with entries from a field F is similar, via a nonsingular matrix over F , to an upper triangular matrix, provided F contains the eigenvalues of A . However, given a finite set of $n \times n$ matrices, Ω , it is not always possible to put them all in triangular form with the same similarity. Such sets are characterized by a theorem of McCoy [24, 9]. If Ω cannot be simultaneously upper triangularized, it may still be possible to put the matrices in Ω into the same block triangular form with the same similarity. Gaines and R. C. Thompson [15] used commutator relations to study such matrices. We study such sets of matrices in Chapter 1; Theorem 1 links the size of the largest block in the finest possible block triangular form for Ω to polynomial identities. McCoy's theorem follows as a special case of Theorem 1.

Now let the field F be \mathbb{C} , the field of complex numbers. If the algebra \mathcal{A} generated by Ω satisfies the condition $\mathcal{A} = \mathcal{A}^*$, then a theorem of Specht [40] tells us \mathcal{A} can be unitarily block diagonalized, with diagonal blocks of sizes n_1, \dots, n_t if and only if \mathcal{A} can be simultaneously block triangularized, with blocks of sizes n_1, \dots, n_t . Applying this to a single matrix A , we see that A is unitarily similar to a block diagonal matrix if and only if A and A^* are simultaneously similar to block triangular matrices. In particular, A is normal if and only if A and A^* are simultaneously similar to triangular matrices. Let $A = H + iK$, where H and K are Hermitian. The matrix A is normal if and only if $HK = KH$, and A is unitarily similar to a block diagonal matrix if and only if

H and K are simultaneously similar to block triangular matrices. This may be viewed as a generalization of normal matrices.

Thus, studying the problem of unitarily block diagonalizing the matrix A is equivalent to considering the problem of simultaneously block triangularizing the pair of Hermitian matrices H and K . This amounts to finding subspaces which are invariant under both H and K . We consider the matrix pencil $xH + yK$ and the characteristic polynomial of this pencil $f(x,y,z) = \det(zI - xH - yK)$ in order to study the pair H and K . Theorem 3 shows that the polynomial $f(x,y,z)$ determines $A = H + iK$ up to unitary similarity when A is 2×2 , but examples show this does not hold for larger matrices.

Motzkin and Taussky [27, 28] studied pairs of matrices A and B with property L and the characteristic curve $\det(zI - xA - yB) = 0$. Among other things, they proved that if H and K are Hermitian, $HK = KH$ if and only if the polynomial $\det(zI - xH - yK)$ factors into linear factors. Thus, $A = H + iK$ is normal if and only if $\det(zI - xH - yK)$ factors into linear factors. Each linear factor corresponds to an eigenvalue of A . We try to generalize this to matrices which are unitarily similar to block diagonal matrices and find that if A is unitarily similar to a block diagonal matrix, then $f(x,y,z)$ splits into factors which correspond to the blocks, but examples show the converse is not true. In particular, the presence of a linear factor in $f(x,y,z)$ does not guarantee the existence of a common eigenvector for H and K . However, Theorem 4 shows that if the linear factor appears to a high enough multiplicity, then H and K must have a common eigenvector. Kippenhahn showed that if $f(x,y,z)$ is a power of a quadratic polynomial, then A is unitarily

similar to a block diagonal matrix. We give a new proof of this.

The equation $f(x,y,z) = 0$ may be viewed as an algebraic curve of degree n --this technique is due to Motzkin and Taussky [28]. Matrices in the pencil then correspond to lines through the origin; the intersection points of these lines with the curve $f(x,y,z) = 0$ correspond to eigenvalues. Using this approach, we study the situation where $f(x,y,z)$ has a linear and a quadratic factor.

Viewed as an equation in line coordinates, $f(x,y,z) = 0$ is the equation of an algebraic curve of class n . Kippenhahn [21] and Murnaghan [30] showed this curve determines the numerical range of A . We use this fact to give another treatment of the case where $f(x,y,z)$ has a linear and a quadratic factor.

Finally, we consider a conjecture of Kippenhahn. The conjecture is that if $f(x,y,z)$ has a repeated factor, then A is unitarily similar to a block diagonal matrix. Using previous theorems, we show the conjecture holds for $n \leq 4$. We then verify the conjecture for $n = 5$ with a lengthy computation. Finally, we note that Theorem 4 would be a consequence of the conjecture.

I. Simultaneous Block Triangularization of Sets of Matrices and
Unitary Block Diagonalization of Sets of Matrices.

Section 1. Simultaneous triangularization of sets of matrices.

An $n \times n$ matrix A over an algebraically closed field F can always be put in upper triangular form via a similarity transformation-- i.e. there exists a nonsingular matrix S such that $S^{-1}AS$ is upper triangular. However, given a finite set, A_1, \dots, A_m of $n \times n$ matrices, it is not always possible to simultaneously upper triangularize A_1, \dots, A_m with the same similarity. Frobenius [11] showed that if the A_i 's commute pairwise, then they can be simultaneously upper triangularized.

Theorem (Frobenius) [11]. Let A_1, \dots, A_m be a set of $n \times n$ matrices over an algebraically closed field F . If $A_i A_j = A_j A_i$ for all $i, j = 1, \dots, m$, then there is a nonsingular matrix S such that $S^{-1}A_i S$ is upper triangular for $i = 1, \dots, m$.

The following example shows that the converse of Frobenius's theorem is not true.

Example 1. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

Then A and B are both upper triangular, but

$$AB = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad BA = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{so} \quad AB \neq BA.$$

The following theorem of McCoy characterizes sets of matrices which can be simultaneously upper triangularized.

Theorem (McCoy) [9, 24]. If A_1, \dots, A_m is a given set of $n \times n$ matrices over a field F containing all of the eigenvalues of A_1, \dots, A_m then the following are equivalent.

(I) For every scalar polynomial $P(x_1, \dots, x_m)$ in the non-commutative variables x_1, \dots, x_m , each of the matrices $P(A_1, \dots, A_m)(A_i A_j - A_j A_i)$ for $i, j = 1, \dots, m$ is nilpotent.

(II) There exists a nonsingular matrix S such that $S^{-1} A_i S$ is upper triangular for $i = 1, \dots, m$.

(III) There exists an ordering of the eigenvalues $\alpha_k^{(i)}$ of each A_i such that the eigenvalues of any scalar polynomial function $P(A_1, \dots, A_m)$ of A_1, \dots, A_m are $P(\alpha_k^{(1)}, \dots, \alpha_k^{(m)})$ for $k = 1, \dots, n$.

The proofs that (II) implies (III) and (III) implies (I) are immediate; the significance of the theorem is the assertion that (I) implies (II). Note that condition (I) says precisely that

$[A_i, A_j] = A_i A_j - A_j A_i$, the commutator of A_i and A_j , lies in the radical of the algebra generated by A_1, \dots, A_m for $i, j = 1, \dots, m$.

A set of matrices satisfying (I), (II), or (III) is said to have property P.

Section 2. Block triangularization.

In this section we examine sets of matrices which can be simultaneously put into the same block upper triangular form via a similarity transformation.

Gaines and Thompson [14,15] proved a generalization of McCoy's theorem which deals with the case where the field F is arbitrary and need not contain the eigenvalues of the matrices. They replace condition (II) of McCoy's theorem by:

(II)' There exists a nonsingular $n \times n$ matrix S over F such that each $S^{-1}A_i S$ is block triangular,

$$S^{-1}A_i S = \begin{pmatrix} A_{11}^{(i)} & A_{12}^{(i)} & \dots & A_{1k}^{(i)} \\ 0 & A_{22}^{(i)} & \dots & A_{2k}^{(i)} \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A_{kk}^{(i)} \end{pmatrix} \quad i = 1, \dots, m.$$

where, for each fixed t , $1 \leq t \leq k$, the matrices

$A_{tt}^{(1)}, A_{tt}^{(2)}, \dots, A_{tt}^{(m)}$ are commutative.

In this section, we study an arbitrary set of square matrices over an algebraically closed field and transform them into the finest possible block upper triangular form via a simultaneous similarity. Gaines [12, 14] used commutator relations to study such sets of matrices and to obtain information about the sizes of the blocks. We will prove a result which links the sizes of the blocks to polynomial identities. McCoy's theorem will be a special case of this result.

First we review some basic theory and introduce some notation.

Let V be an n -dimensional vector space over a field F and let Ω be a set of linear transformations of V . Let \mathcal{A} be the algebra of linear transformations generated by the set Ω over F . In the usual manner [17, Chapter IV] let \mathcal{A} act on V and regard V as a left \mathcal{A} -module. A subspace U of V is called an invariant subspace of V , relative to the set Ω , provided $A(U) \subseteq U$ for all A in Ω . Note that U is invariant relative to Ω if and only if U is invariant relative to \mathcal{A} .

Definition [17, p. 120]. A sequence of invariant subspaces

$$0 \subset V_1 \subset V_2 \subset \dots \subset V_t = V$$

is called a composition series for V relative to Ω , provided there exists no invariant subspace U such that $V_i \subset U \subset V_{i+1}$ for $i = 1, \dots, t-1$.

The set Ω induces a set of linear transformations on the factor spaces V_i/V_{i-1} ; since there is no invariant subspace U such that $V_{i-1} \subset U \subset V_i$, the factor space V_i/V_{i-1} is irreducible with respect to Ω for $i = 1, \dots, t$.

The irreducible spaces $V_1, V_2/V_1, \dots, V_t/V_{t-1}$ are called the composition factors of the series. Since V is a finite dimensional vector space,

a composition series for V relative to Ω exists. (If there are no proper Ω -invariant subspaces of V , then $0 \subset V$ is the only composition series for V .) Now suppose $0 \subset W_1 \subset \dots \subset W_s = V$ and $0 \subset V_1 \subset \dots \subset V_t = V$ are two composition series. The Jordan-Hölder theorem states that $s = t$ and the composition factors

$W_1, W_2/W_1, \dots, W_s/W_{s-1}$ are isomorphic to the composition factors

$V_1, V_2/V_1, \dots, V_t/V_{t-1}$ in some order. Let n_i denote the dimension

of V_i/V_{i-1} . The set of integers $\{n_1, \dots, n_t\}$ lists the dimensions of the composition factors and hence is uniquely determined by V and Ω . Note $n_1 + n_2 + \dots + n_t = n$. If $0 \subset V$ is the only composition series, $t = 1$ and $n_1 = n$.

Let $\{\alpha_{11}, \dots, \alpha_{1 n_1}, \alpha_{21}, \dots, \alpha_{2 n_2}, \dots, \alpha_{t1}, \dots, \alpha_{t n_t}\}$ be a basis for V such that $\{\alpha_{i1}, \dots, \alpha_{i n_i}\}$ is a basis for V_i , for $i = 1, \dots, t$. Represent the linear transformations in Ω as $n \times n$ matrices with elements in F , relative to the basis $\{\alpha_{11}, \dots, \alpha_{t n_t}\}$. Since each V_i is invariant under Ω , these matrices assume the following block upper triangular form

$$A = \begin{pmatrix} A_{11} & & & * \\ 0 & A_{22} & & \\ \vdots & & \ddots & \\ 0 & 0 & \dots & A_{tt} \end{pmatrix} .$$

All elements below the diagonal blocks A_{11}, \dots, A_{tt} are zero. The i 'th diagonal block, A_{ii} , is $n_i \times n_i$ and is the matrix of the transformation induced by A on the factor space V_i/V_{i-1} .

Definition. An $n \times n$ matrix A is $BT(n_1, \dots, n_t)$ provided A is block upper triangular with t diagonal blocks of sizes n_1, \dots, n_t , where $t \geq 1$ and $n_1 + n_2 + \dots + n_t = n$.

The set Ω will be called $BT(n_1, \dots, n_t)$ provided all matrices in Ω are $BT(n_1, \dots, n_t)$. Note that Ω is $BT(n_1, \dots, n_t)$ if and only if the algebra \mathcal{A} is $BT(n_1, \dots, n_t)$.

Definition. The set Ω is $BT^*(n_1, \dots, n_t)$ provided Ω is $BT(n_1, \dots, n_t)$ and the integers n_1, \dots, n_t are the dimensions of the composition factors of V relative to Ω .

Note that Ω is $BT^*(n_1, \dots, n_t)$ if and only if \mathcal{O} is $BT^*(n_1, \dots, n_t)$. If n_1, \dots, n_t are the dimensions of the composition factors of V relative to Ω , and Ω is $BT(m_1, \dots, m_s)$, then $s \leq t$ and $m_i \geq n_{\sigma(i)}$ for some permutation σ of $1, \dots, t$.

Section 3. The main result of Chapter I--Theorem 1.

Let Ω be a set of $n \times n$ matrices. From the discussion in I.2, it is clear that there is a nonsingular matrix S such that $S^{-1}\Omega S = \{S^{-1}AS \mid A \in \Omega\}$ is $BT^*(n_1, \dots, n_t)$. We wish to investigate the problem of determining conditions on Ω which force the diagonal blocks to be a certain size--in particular, if $m = \max\{n_1, \dots, n_t\}$, what conditions on Ω guarantee $m \leq k$, for some positive integer k . McCoy's theorem deals with the case $m = 1$. We will generalize condition (I) of McCoy's theorem to obtain a characterization of sets of matrices for which $m = k$, where k is a positive integer.

We first discuss some facts about polynomials of matrices. Let $P(x_1, \dots, x_r)$ be a polynomial in the non-commuting variables x_1, \dots, x_r with coefficients in F . Suppose A_1, \dots, A_r are $n \times n$ matrices which are $BT(n_1, \dots, n_t)$. Then $P(A_1, \dots, A_r)$ is also $BT(n_1, \dots, n_t)$ and if A_{ij} is the j 'th diagonal block of A_i , $i = 1, \dots, r$, then $P(A_{1j}, \dots, A_{rj})$ is the j 'th diagonal block of $P(A_1, \dots, A_r)$.

Suppose P vanishes on all r -tuples of $k \times k$ matrices over F . Let $h \leq k$ and let B_1, \dots, B_r be an r -tuple of $h \times h$ matrices over F . For $i = 1, \dots, r$ let \bar{B}_i denote the $k \times k$ matrix with B_i in the upper left hand corner and zeroes in the remaining $k-h$ rows and columns.

$$\bar{B}_i = \begin{array}{c} h \\ k-h \end{array} \left[\begin{array}{c|c} \overbrace{B_i}^h & \overbrace{0}^{k-h} \\ \hline 0 & 0 \end{array} \right]$$

Since $P(\bar{B}_1, \dots, \bar{B}_r) = 0$, $P(B_1, \dots, B_r) = 0$. Thus P vanishes on all r -tuples of $h \times h$ matrices when $h \leq k$. Suppose P does not vanish on all r -tuples of $k+1 \times k+1$ matrices. Then a similar argument shows that P does not vanish on all r -tuples of $n \times n$ matrices whenever $n \geq k+1$.

The following theorem of Burnside will be needed in the proof of Theorem 1.

Theorem (Burnside) [3, 17(p. 276)]. If U is a non-zero, irreducible algebra of linear transformations of a finite dimensional vector space over an algebraically closed field, then U is the complete algebra of linear transformations.

We now state and prove the main theorem of this chapter.

Theorem 1. Let $Q(x_1, \dots, x_r)$ be a polynomial in the non-commuting variables x_1, \dots, x_r with coefficients in an algebraically closed field F . Suppose the equation $Q(x_1, \dots, x_r) = 0$ is satisfied by every r -tuple of $k \times k$ matrices over F , but there exists an r -tuple of $k+1 \times k+1$ matrices over F which does not satisfy the equation.

Let \mathcal{A} be an algebra of $n \times n$ matrices over F which is $BT^*(n_1, \dots, n_t)$. Let $m = \max \{n_1, \dots, n_t\}$. Then the following two conditions are equivalent.

$$(1) \quad m \leq k.$$

$$(2) \quad Q(A_1, \dots, A_r) \text{ is in the radical of } \mathcal{A} \text{ for every } r\text{-tuple } (A_1, \dots, A_r) \text{ of matrices of } \mathcal{A}.$$

Proof. Assume $m \leq k$ and let (A_1, \dots, A_r) be any r -tuple of matrices of \mathcal{A} . Denote the j 'th diagonal block of A_i by A_{ij} , for $i = 1, \dots, r$ and $j = 1, \dots, t$. The matrix $Q(A_1, \dots, A_r)$ is $BT(n_1, \dots, n_t)$ with $Q(A_{1j}, \dots, A_{rj})$ as the j 'th diagonal block. Since $m \leq k$, the hypothesis on Q implies $Q(A_{1j}, \dots, A_{rj}) = 0$ for $j = 1, \dots, t$. Thus all of the diagonal blocks of $Q(A_1, \dots, A_r)$ are blocks of zeroes. Now let C be any element of \mathcal{A} . The matrix $C Q(A_1, \dots, A_r)$ is again $BT(n_1, \dots, n_t)$ and the diagonal blocks of $C Q(A_1, \dots, A_r)$ are the products of the corresponding diagonal blocks of C and $Q(A_1, \dots, A_r)$. Hence all of the diagonal blocks of $C Q(A_1, \dots, A_r)$ are blocks of zeroes. This shows that $C Q(A_1, \dots, A_r)$ is nilpotent for all C in \mathcal{A} . Therefore, $Q(A_1, \dots, A_r)$ is in the radical of \mathcal{A} .

Conversely, suppose condition (2) holds. We have $m = \max \{n_1, \dots, n_t\} = n_i$ for some i . Let

$$\mathcal{A}_i = \{A_{ii} \mid A_{ii} \text{ is the } i\text{'th diagonal block of some } A \in \mathcal{A}\}.$$

If $m = 1$, we certainly have $m \leq k$. Hence, we may assume $m > 1$ and $\mathcal{A}_i \neq 0$. Since \mathcal{A} is $BT^*(n_1, \dots, n_t)$, the algebra \mathcal{A}_i is an irreducible algebra of $m \times m$ matrices over F . The theorem of Burnside implies \mathcal{A}_i is the complete matrix algebra, $M_m(F)$, of $m \times m$ matrices over F . Since $Q(A_1, \dots, A_r)$ is in the radical of \mathcal{A}

for every r -tuple (A_1, \dots, A_r) of \mathcal{A} , the $m \times m$ matrix $Q(B_1, \dots, B_r)$ must be in the radical of \mathcal{A}_i for every r -tuple (B_1, \dots, B_r) of matrices of \mathcal{A}_i . But $\mathcal{A} = M_m(F)$ which has a zero radical. Therefore, $Q(B_1, \dots, B_r) = 0$ and Q vanishes on $\mathcal{A}_i = M_m(F)$. By our hypothesis on Q , $m \leq k$. ■

We now show McCoy's theorem follows from Theorem 1. We will prove that condition (II) of McCoy's theorem follows from condition (I) of McCoy's theorem; the implications that (II) implies (III) and (III) implies (II) are straightforward. Let $\Omega = \{A_1, \dots, A_p\}$ be a set of $n \times n$ matrices over an algebraically closed field F and let \mathcal{A} be the algebra generated by Ω over F . Assume condition (I) of McCoy's theorem holds--i.e. that $[A_i, A_j] = A_i A_j - A_j A_i$ is in the radical of \mathcal{A} for $i, j = 1, \dots, p$. We apply Theorem 1 with $Q(x_1, x_2) = x_1 x_2 - x_2 x_1$ and $k = 1$.

We now need to show that condition (2) of Theorem 1 holds--i.e. that $[C, D]$ is in the radical of \mathcal{A} for any pair (C, D) of matrices of \mathcal{A} . Since the radical is an ideal and $x_1 x_2 - x_2 x_1 = [x_1, x_2]$ is linear in each of its variables, we may assume C and D are monomials in A_1, \dots, A_p . By repeatedly applying the commutator identity

$$[XY, Z] = [X, Y] Z + Y [X, Z] + [Y, Z] X + Z [Y, X]$$

and using the fact that the radical of \mathcal{A} is a two sided ideal, we may assume $C = A_i^r$ and $D = A_j^s$ where r and s are non-negative integers. Finally, we use induction on r and s and the identity

$$[X^n, Y] = X [X^{n-1}, Y] + [X, Y] X^{n-1}$$

to show that if $[A_i, A_j]$ is in the radical of \mathcal{A} for $i, j = 1, \dots, p$,

then $[C, D]$ is in the radical of \mathcal{A} . Hence, condition (I) of McCoy's theorem implies that condition (2) of Theorem 1 holds. Therefore, $m = 1$ and A_1, \dots, A_p can be simultaneously upper triangularized.

Theorem 1 immediately implies the following corollary.

Corollary 1. Let $Q(x_1, \dots, x_r)$ be a polynomial in r non-commuting variables with coefficients in the algebraically closed field F . Suppose the equation $Q(x_1, \dots, x_r) = 0$ is satisfied by every r -tuple of $k \times k$ matrices over F , but there exists an r -tuple of $k+1 \times k+1$ matrices over F which does not satisfy the equation. Let \mathcal{A} be an algebra of $n \times n$ matrices over F and suppose \mathcal{A} is $BT^*(n_1, \dots, n_t)$. Then if $Q(A_1, \dots, A_r) = 0$ holds for every r -tuple (A_1, \dots, A_r) of matrices of \mathcal{A} , $m = \max \{n_1, \dots, n_t\} \leq k$.

Setting $k = 1$ and $Q(x_1, x_2) = x_1x_2 - x_2x_1$ in Corollary 1 yields Frobenius's theorem.

Section 4. Polynomial identities.

The hypothesis of Theorem 1 calls for a polynomial Q in r non-commuting variables which vanishes on the space of all $k \times k$ matrices over F but which does not vanish for all r -tuples of $k+1 \times k+1$ matrices. In this section we discuss some known examples of such polynomials.

Definition. The standard polynomial in m variables is:

$$S_m(x_1, \dots, x_m) = \sum^{\pm} x_{\sigma(1)} x_{\sigma(2)} \cdot \dots \cdot x_{\sigma(m)}$$

where the sum is over all permutations σ of the integers $1, \dots, m$ and the coefficient of the term $x_{\sigma(1)} x_{\sigma(2)} \dots x_{\sigma(m)}$ is $+1$ if σ is an even permutation and -1 if σ is an odd permutation.

The equation $S_m(x_1, \dots, x_m) = 0$ is called a standard identity. Amitsur and Levitzki [1, pp. 450-455] showed that the standard identity in $2k$ variables is satisfied by all $2k$ -tuples of $k \times k$ matrices. Furthermore, Levitzki [23, pp. 336-338] showed that the minimal degree of a polynomial identity for the complete matrix algebra of $k \times k$ matrices is at least $2k$. Hence, $S_{2k} = 0$ is a polynomial identity of minimal degree for $M_k(F)$ and the polynomial $S_{2k}(x_1, \dots, x_{2k})$ satisfies the hypothesis of Theorem 1.

The polynomial $P(X, Y) = [X, Y]^2$ is a central polynomial for $M_2(F)$ --i.e. if X and Y are 2×2 matrices, then $[X, Y]^2$ is always a scalar matrix. Formanek [10] and Razmyslov [34] constructed central polynomials for the algebra of $k \times k$ matrices over a field, for each positive integer k . The Formanek polynomial $F_k(X, Y_1, \dots, Y_k)$ is non-vanishing on $M_k(F)$ and is not a central polynomial for $M_{k+1}(F)$. Hence the polynomial

$$G_k(Z, X, Y_1, \dots, Y_k) = [Z, F_k(X, Y_1, \dots, Y_k)]$$

satisfies the hypothesis of Theorem 1.

Now let A, B_1, \dots, B_{k+1} be a $k+2$ -tuple of $k \times k$ matrices. Let $\bar{A}, \bar{B}_1, \dots, \bar{B}_{k+1}$ be the $k+1 \times k+1$ matrices obtained by adding a row and column of zeroes to each of A, B_1, \dots, B_{k+1} . The matrix $F_{k+1}(\bar{A}, \bar{B}_1, \dots, \bar{B}_{k+1})$ must be a scalar matrix. Since $F_{k+1}(\bar{A}, \bar{B}_1, \dots, \bar{B}_{k+1})$

is a block diagonal matrix with $F_{k+1}(A, B_1, \dots, B_{k+1})$ in the first block (which has size $k \times k$) and a 1×1 zero matrix in the second block, $F_{k+1}(\bar{A}, \bar{B}_1, \dots, \bar{B}_{k+1}) = 0$. Hence, $F_{k+1}(A, B_1, \dots, B_{k+1}) = 0$ and F_{k+1} vanishes on $M_k(F)$. Since F_{k+1} does not vanish on $M_{k+1}(F)$, the polynomial F_{k+1} satisfies the hypothesis of Theorem 1.

Section 5. Unitary block triangularization and block diagonalization.

We now let F be \mathbb{C} , the field of complex numbers and let V denote an n -dimensional vector space over \mathbb{C} . The following theorem is due to Specht [40]. We state the result a bit differently here and give a different proof.

Theorem A (Specht) [40, p. 210]. Let Ω be a nonempty set of complex $n \times n$ matrices. Then there is a nonsingular matrix S such that $S^{-1}\Omega S$ is $BT(n_1, \dots, n_t)$ if and only if there is a unitary matrix U such that $U^{-1}\Omega U = U^*\Omega U$ is $BT(n_1, \dots, n_t)$.

Proof. Suppose there is a nonsingular matrix S such that $S^{-1}\Omega S$ is $BT(n_1, \dots, n_t)$. Then V has a basis

$$\{\alpha_{11}, \alpha_{12}, \dots, \alpha_{1 n_1}, \alpha_{21}, \dots, \alpha_{2 n_2}, \dots, \alpha_{t1}, \dots, \alpha_{t n_t}\}$$

such that the subspace $V_i = \langle \alpha_{11}, \dots, \alpha_{1 n_1}, \dots, \alpha_{i1}, \dots, \alpha_{i n_i} \rangle$

is invariant under Ω for $i = 1, \dots, t$. Thus, we have a chain of subspaces $0 \subset V_1 \subset V_2 \subset \dots \subset V_t = V$ such that each V_i is Ω -invariant and the dimension of V_i/V_{i-1} is n_i . Use the Gram-Schmidt process to construct an orthonormal basis of V ,

$$\{\beta_{11}, \dots, \beta_{1 n_1}, \beta_{21}, \dots, \beta_{2 n_2}, \dots, \beta_{t1}, \dots, \beta_{t n_t}\}$$

such that for each $i = 1, \dots, t$, the set

$$\{\beta_{11}, \dots, \beta_{1n_1}, \dots, \beta_{i1}, \dots, \beta_{in_i}\}$$

is an orthonormal basis for V_i . Let U be the change of basis matrix from the standard basis of V to the β -basis. Since the β -basis is orthonormal, U is a unitary matrix, and since each V_i is Ω -invariant, $U^* \Omega U$ is $BT(n_1, \dots, n_t)$. ■

Thus, if we can simultaneously put a set Ω of complex matrices into the form $BT(n_1, \dots, n_t)$ by a similarity transformation, then we can do this with a unitary transformation.

Suppose now that the matrices in Ω can be simultaneously put into the same block diagonal form with a similarity transformation. The following example shows that it is not always possible to do this with a unitary similarity.

Example 2. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$.

The vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ are a pair of linearly independent

eigenvectors for A and B . If $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then

$$S^{-1}AS = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad S^{-1}BS = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

However, since neither A nor B can be unitarily diagonalized, they clearly cannot be simultaneously, unitarily diagonalized.

Proof. Assume there is a nonsingular matrix S such that $S^{-1} \mathcal{A} S$ is $BT(n_1, \dots, n_t)$. By Theorem A (Specht) there is a unitary matrix U such that $U^* \mathcal{A} U$ is $BT(n_1, \dots, n_t)$. Then $\mathcal{A} = \mathcal{A}^*$ implies $(U^* \mathcal{A} U)^* = U^* \mathcal{A}^* U = U^* \mathcal{A} U$. Since the transpose conjugate of a block upper triangular matrix is block lower triangular, the only nonzero elements of the matrices of $U^* \mathcal{A} U$ are in the diagonal blocks. ■

Corollary 2. An $n \times n$ complex matrix A is unitarily similar to a block diagonal matrix with blocks of sizes n_1, \dots, n_t if and only if A and A^* are simultaneously similar to block upper triangular matrices with blocks of sizes n_1, \dots, n_t .

Proof. Let \mathcal{A} be the algebra generated by A and A^* . Then $\mathcal{A} = \mathcal{A}^*$ and Theorem B applies. ■

Note that for the case $n_1 = n_2 = \dots = n_t = 1$, the corollary says that A is normal if and only if A and A^* are simultaneously similar to upper triangular matrices.

We now apply Specht's theorem B to Theorem 1 (Section I.3) to obtain the following result.

Theorem 2. Let Ω be a nonempty set of complex $n \times n$ matrices and let \mathcal{A} be the algebra generated by Ω over \mathbb{C} . Assume $\mathcal{A} = \mathcal{A}^*$. Let $Q(x_1, \dots, x_r)$ be a polynomial satisfying the hypothesis of Theorem 1. Then the following are equivalent.

- (1) There is a unitary matrix U such that $U^* \Omega U$ is $D(n_1, \dots, n_t)$ with $\max \{n_1, \dots, n_t\} \leq k$ (where k is as in Theorem 1).
- (2) $Q(A_1, \dots, A_r) = 0$ for all $A_1, \dots, A_r \in \mathcal{A}$.

This may be regarded as a generalization of the fact that a set Ω of normal matrices can be simultaneously, unitarily diagonalized if and only if the matrices in Ω commute [9, pp. 224-225].

Section 6. Some applications of Specht's theorem B and Theorem 2.

McCoy [25] introduced the notion of quasi-commutative matrices. Roth [36] and Drazin [8] generalized this concept and studied k -commutative matrices.

Definition (Drazin) [8, p. 223]. Let Γ_0 be a set of $n \times n$ matrices A_1, \dots, A_m . Define sets Γ_k ($k = 1, 2, \dots$) inductively as follows.

$$\Gamma_0 = \{A_1, \dots, A_m\}$$

$$\Gamma_1 = \{ [A_i, A_j] \mid i, j = 1, \dots, m \}$$

⋮

$$\Gamma_k = \{ [A_i, C_{k-1}] \mid C_{k-1} \in \Gamma_{k-1} \text{ and } i = 1, \dots, m \}$$

Definition (Drazin) [8, p. 223]. The set $\Gamma_0 = \{A_1, \dots, A_m\}$ will be said to be quasi-commuting of the k 'th order for some $k \geq 1$ provided

$$\Gamma_k = 0.$$

Note that $\Gamma_1 = 0$ means that the set is commutative. McCoy's property of quasi-commuting [25] is included under $\Gamma_2 = 0$. McCoy proved that if $\Gamma_2 = 0$, then the matrices A_1, \dots, A_m have property P [25]. Drazin extended this result. In addition to other theorems about quasi-commutativity, he proved the following result.

Theorem (Drazin) [8]. Let F be algebraically closed and let A_1, \dots, A_m be $n \times n$ matrices over F . If A_1, \dots, A_m are quasi-commuting of the k 'th order, for some $k \geq 1$, then A_1, \dots, A_m have property P .

We now apply Drazin's theorem to Corollary 2 (Section I.5). This shows that A is normal if and only if A and A^* are quasi-commuting of the k 'th order for some $k \geq 1$.

Now let A_1, \dots, A_m be a set of normal matrices. Since A_i is normal, A_i^* is a polynomial in A_i . Hence the algebra, \mathcal{A} , generated by A_1, \dots, A_m satisfies $\mathcal{A} = \mathcal{A}^*$. Using Drazin's theorem and Specht's theorem B, we see that the normal matrices A_1, \dots, A_m commute if and only if they are quasi-commuting of order k for some $k \geq 1$.

We now use Theorem 2 (Section I.5) to give an alternative proof of a theorem of Watters [50].

Theorem (Watters) [50]. If S is a set of $n \times n$ normal matrices and \mathcal{A} is the algebra generated by S over the complex numbers, the matrices in S are simultaneously, unitarily similar to block diagonal matrices with blocks of size 1×1 or 2×2 if and only if

$$[A, B]^2 P = P [A, B]^2$$

holds for all A in \mathcal{A} and B and P in S .

To obtain this result from Theorem 2, let

$$Q(x_1, x_2, x_3) = [x_1, x_2]^2 x_3 - x_3 [x_1, x_2]^2.$$

This polynomial vanishes for all triples of 2×2 matrices, but not for all triples of $n \times n$ matrices whenever $n \geq 3$. Hence,

$Q(x_1, x_2, x_3)$ satisfies the hypothesis of Theorem 1 with $k = 2$.

Assume the second condition of Watter's theorem holds--i.e. that $Q(A, B, P) = 0$ for all A in \mathcal{A} and B and P in S . Since $[A, B]^2$ commutes with every element of \mathcal{A} if and only if it commutes with every element of S , we clearly have $Q(A, B, Z) = 0$ for all A and Z in \mathcal{A} and B in S .

Since $\mathcal{A} = \mathcal{A}^*$, Specht's theorem B implies that \mathcal{A} is unitarily similar to an algebra of matrices which is $D(n_1, \dots, n_t)$ where the numbers n_1, \dots, n_t are the dimensions of the composition factors of the algebra \mathcal{A} . We consider only the i 'th diagonal block and hence may assume $\mathcal{A} = M_m(\mathbb{C})$, where $m = n_i$. We need to show $m \leq 2$.

Suppose $m \geq 3$ and $B = (b_{ij}) \in S$ is not a scalar matrix. If $b_{ij} \neq 0$ for some $i \neq j$, then $[E_{ji}, B]^2$ is not scalar, where E_{ji} is the matrix with a one in the j, i position and zeroes elsewhere. Thus, $[E_{ji}, B]^2$ does not commute with every $Z \in \mathcal{A}$. If B is diagonal with $b_{ii} \neq b_{jj}$, let $A = E_{ii} + E_{ij} + E_{ji} + E_{jj}$. Then $[A, B]^2$ is not scalar and hence does not commute with every $Z \in \mathcal{A}$.¹ Hence $Q(A, B, Z) = 0$ for all A and Z in $M_m(\mathbb{C})$ and B in S implies $m \leq 2$.

Finally, we consider a single $n \times n$ matrix A of complex numbers. Write $A = H + iK$, where H and K are Hermitian. Then $H = \frac{A + A^*}{2}$ and $K = \frac{A - A^*}{2i}$. The matrix A is normal if and only if $AA^* = A^*A$. Note that $AA^* = A^*A$ if and only if $HK = KH$. Thus, by Specht's theorem B, the matrix A is normal if and only if

¹This argument was suggested by the referee of [39].

H and K can be simultaneously upper triangularized.

Now suppose A is unitarily similar to a matrix which is $D(n_1, \dots, n_t)$. This may be viewed as a generalization of the concept of normality. Again applying Specht's theorem B, we see that A is unitarily similar to a matrix which is $D(n_1, \dots, n_t)$ if and only if the pair A and A^* are simultaneously similar to matrices which are $BT(n_1, \dots, n_t)$. This occurs if and only if H and K are simultaneously similar to matrices which are $BT(n_1, \dots, n_t)$. Thus, we may study these generalized normal matrices by studying pairs of Hermitian matrices with common invariant subspaces. This idea will be used in the remaining chapters.

II. Pencils of Matrices, the Characteristic Polynomial and Property L.

Section 1. Definition of the matrix pencil $xA + yB$ and the characteristic polynomial $f(x,y,z)$.

For the remainder of this work, A and B will denote $n \times n$ complex matrices of complex numbers. As described at the end of I.6, we will consider $A = H + iK$, where H and K are Hermitian. The letters H and K will always designate $n \times n$ complex, Hermitian matrices.

Definition. The pencil generated by A and B is

$$\{ xA + yB \mid x, y \in \mathbb{C} \}.$$

Definition. The characteristic polynomial of the pencil $xA + yB$ is

$$f(x,y,z) = \det(zI - xA - yB).$$

Note that $f(x,y,z)$ is a homogeneous polynomial of degree n in the variables x , y , and z , with complex coefficients. Fixing values of x and y , say $x = x_0$ and $y = y_0$, yields

$$f(x_0, y_0, z) = \det(zI - x_0A - y_0B),$$

the usual characteristic polynomial of the matrix $x_0A + y_0B$.

Applying a simultaneous similarity to A and B does not change the characteristic polynomial, for

$$\det(zI - xA - yB) = \det(zI - xS^{-1}AS - yS^{-1}BS).$$

Section 2. The pencil $xH + yK$ where $A = H + iK$.

We will use the pencil $xH + yK$ and the characteristic polynomial $f(x,y,z) = \det(zI - xH - yK)$ to study the matrix $A = H + iK$, where $H = \frac{A + A^*}{2}$ and $K = \frac{A - A^*}{2i}$ are Hermitian. In this section we prove that the characteristic polynomial has real coefficients, and that it uniquely determines A up to unitary similarity when A is 2×2 .

Proposition 1. Let H and K be $n \times n$ Hermitian matrices. Then the coefficients of the polynomial $f(x,y,z) = \det(zI - xH - yK)$ are real.

Proof. If U is a unitary matrix, then U^*HU and U^*KU are again Hermitian and $\det(zI - xH - yK) = \det(zI - xU^*HU - yU^*KU)$. Since we can unitarily diagonalize H , we may, without loss of generality, assume $H = \text{diag}(h_1, h_2, \dots, h_n)$. Letting $K = (k_{ij})$, we have

$$zI - xH - yK = \begin{pmatrix} z-h_1x-k_{11}y & -k_{12}y & \dots & -k_{1n}y \\ -\bar{k}_{12}y & z-h_2x-k_{22}y & \dots & -k_{2n}y \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ -\bar{k}_{1n}y & -\bar{k}_{2n}y & \dots & z-h_nx-k_{nn}y \end{pmatrix}.$$

Consider a term, $a_{ijk}x^i y^j z^k$ of the polynomial $f(x,y,z)$. In expanding $\det(zI - xH - yK)$, such a term arises by taking z from exactly k diagonal terms and x from exactly i diagonal terms. Thus, a_{ijk} is a sum of numbers, each of which is the product of i eigenvalues of H with the determinant of a $j \times j$ principal minor of K . Since H is Hermitian, it has real eigenvalues. Since K is Hermitian, any

principal minor of K is also Hermitian, and hence has a real determinant. Therefore, a_{ijk} is a real number. ■

Theorem 3. Let $A = H_1 + iK_1$ and $B = H_2 + iK_2$, where A and B are 2×2 complex matrices and $H_1, K_1, H_2,$ and K_2 are Hermitian. Suppose $\det(zI - xH_1 - yK_1) = \det(zI - xH_2 - yK_2)$. Then there exists a unitary matrix U such that $U^*AU = B$.

Proof. We may unitarily put A into upper triangular form, so without loss of generality, we may assume

$$A = \begin{pmatrix} \lambda_1 & 2a \\ 0 & \lambda_2 \end{pmatrix}$$

where λ_1 and λ_2 are the eigenvalues of A and $a \in \mathbb{C}$.

Write $\lambda_1 = r_1 + is_1$ and $\lambda_2 = r_2 + is_2$, where $r_1, s_1, r_2,$ and s_2 are real numbers. Then,

$$H_1 = \frac{A + A^*}{2} = \begin{pmatrix} r_1 & a \\ \bar{a} & s_1 \end{pmatrix} \quad \text{and} \quad K_1 = \frac{A - A^*}{2i} = \begin{pmatrix} s_1 & -ia \\ ia & s_2 \end{pmatrix}$$

$$\begin{aligned} \det(zI - xH_1 - yK_1) &= \det \begin{pmatrix} z - r_1x - s_1y & -a(x-iy) \\ -\bar{a}(x+iy) & z - r_2x - s_2y \end{pmatrix} \\ &= z^2 - [(r_1 + r_2)x + (s_1 + s_2)y]z + (r_1x + s_1y)(r_2x + s_2y) - |a|^2(x^2 + y^2). \end{aligned}$$

Setting $x = 1$ and $y = i$ in the equation

$$\det(zI - xH_1 - yK_1) = \det(zI - xH_2 - yK_2)$$

yields $\det(zI - A) = \det(zI - B)$. This shows that A and B have

the same eigenvalues. Applying a unitary similarity to B , we may assume

$$B = \begin{pmatrix} \lambda_1 & 2b \\ 0 & \lambda_2 \end{pmatrix} = H_2 + iK_2.$$

As before, we have

$$\begin{aligned} \det(zI - xH_2 - yK_2) &= z^2 - [(r_1 + r_2)x + (s_1 + s_2)y] \\ &\quad + (r_1x + s_1y)(r_2x + s_2y) - |b|^2(x^2 + y^2). \end{aligned}$$

Since $\det(zI - xH_1 - yK_1) = \det(zI - xH_2 - yK_2)$, we must have $|a| = |b|$. Hence, $b = a e^{i\theta}$ for some $0 \leq \theta \leq 2\pi$.

Let $U = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$. Then U is unitary and

$$U^* A U = \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \begin{pmatrix} \lambda_1 & a e^{i\theta} \\ 0 & \lambda_2 e^{i\theta} \end{pmatrix} = \begin{pmatrix} \lambda_1 & a e^{i\theta} \\ 0 & \lambda_2 \end{pmatrix} = B.$$

Hence, A and B are unitarily similar. ■

Theorem 3 shows that if A is 2×2 , then it is uniquely determined, up to unitary similarity, by the polynomial $\det(zI - xH - yK)$. The following examples show that this is not true for larger matrices.

Example 3. Let $A = \begin{pmatrix} 0 & i & i \\ i & a+i & i \\ i & i & -a+i \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a+i & i\sqrt{3} \\ 0 & i\sqrt{3} & -a+i \end{pmatrix}$

where a is a nonzero real number.

$$A = H_1 + iK_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -a \end{pmatrix} + i \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$B = H_2 + iK_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -a \end{pmatrix} + i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \sqrt{3} \\ 0 & \sqrt{3} & 1 \end{pmatrix}$$

The eigenvectors of H_1 are the standard basis vectors,

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and scalar multiples of these three vectors.}$$

Since none of these are eigenvectors of K_1 , the matrices H_1 and K_1 do not have a common eigenvector. Hence A is not unitarily similar to a matrix which is $D(1,2)$. However, B is $D(1,2)$. Hence A is not unitarily similar to B .

We now compute $\det(zI - xH_1 - yK_1)$, expanding by cofactors of the elements of the first column.

$$\begin{aligned} \det(zI - xH_1 - yK_1) &= \begin{vmatrix} z & -y & -y \\ -y & z-ax-y & -y \\ -y & -y & z+ax-y \end{vmatrix} \\ &= z [(z - ax - y)(z + ax - y) - y^2] + y [(-y)(z + ax - y) - y^2] \\ &\quad - y [y^2 + y(z - ax - y)] \\ &= z [(z - ax - y)(z + ax - y) - y^2] + y [-2y^2 - y(2z - 2y)] \end{aligned}$$

$$= z [(z - ax - y)(z + ax - y) - y^2] - 2y^2z$$

$$= z [(z - ax - y)(z + ax - y) - 3y^2].$$

We now compute $\det(zI - xH_2 - yK_2)$.

$$\det(zI - xH_2 - yK_2) = \begin{pmatrix} z & 0 & 0 \\ 0 & z-ax-y & -\sqrt{3}y \\ 0 & -\sqrt{3}y & z+ax-y \end{pmatrix}$$

$$= z [(z - ax - y)(z + ax - y) - 3y^2].$$

Thus, $\det(zI - xH_1 - yK_1) = \det(zI - xH_2 - yK_2)$, but A is not unitarily similar to B.

Note that in Example 3, the characteristic polynomial factors into a linear factor and a quadratic factor. In the next example, the characteristic polynomial is irreducible.

Example 4.

$$\text{Let } A = \begin{pmatrix} 0 & \sqrt{\frac{1}{2}} & 2 \\ 0 & 0 & \sqrt{\frac{1}{2}} \\ 0 & 0 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$A = H_1 + iK_1 = \begin{pmatrix} 0 & \frac{1}{2}\sqrt{\frac{1}{2}} & 1 \\ \frac{1}{2}\sqrt{\frac{1}{2}} & 0 & \frac{1}{2}\sqrt{\frac{1}{2}} \\ 1 & \frac{1}{2}\sqrt{\frac{1}{2}} & 0 \end{pmatrix} + i \begin{pmatrix} 0 & -\frac{i}{2}\sqrt{\frac{1}{2}} & -1 \\ \frac{i}{2}\sqrt{\frac{1}{2}} & 0 & -\frac{i}{2}\sqrt{\frac{1}{2}} \\ i & \frac{i}{2}\sqrt{\frac{1}{2}} & 0 \end{pmatrix}.$$

$$B = H_2 + iK_2 = \begin{pmatrix} 0 & \frac{1}{2}\sqrt{\frac{1}{2}} & \frac{1}{2}\sqrt{\frac{1}{2}} \\ \frac{1}{2}\sqrt{\frac{1}{2}} & 0 & 1 \\ \frac{1}{2}\sqrt{\frac{1}{2}} & 1 & 0 \end{pmatrix} + i \begin{pmatrix} 0 & -\frac{i}{2}\sqrt{\frac{1}{2}} & -\frac{i}{2}\sqrt{\frac{1}{2}} \\ \frac{i}{2}\sqrt{\frac{1}{2}} & 0 & -i \\ \frac{i}{2}\sqrt{\frac{1}{2}} & i & 0 \end{pmatrix}.$$

$$\det(zI - xH_1 - yK_1) = \det \begin{pmatrix} z & -\frac{1}{2}\sqrt{\frac{1}{2}}(x-iy) & -(x-iy) \\ -\frac{1}{2}\sqrt{\frac{1}{2}}(x+iy) & z & -\frac{1}{2}\sqrt{\frac{1}{2}}(x-iy) \\ -(x+iy) & -\frac{1}{2}\sqrt{\frac{1}{2}}(x+iy) & z \end{pmatrix}$$

$$= z^3 + \left[-\frac{1}{8}(x^2 + y^2) - (x^2 + y^2) - \frac{1}{8}(x^2 + y^2) \right] z \\ - \frac{1}{8}(x - iy)(x^2 + y^2) - \frac{1}{8}(x + iy)(x^2 + y^2)$$

$$= z^3 - \frac{5}{4}(x^2 + y^2)z - \frac{1}{4}x(x^2 + y^2).$$

$$\det(zI - xH_2 - yK_2) = \det \begin{pmatrix} z & -\frac{1}{2}\sqrt{\frac{1}{2}}(x-iy) & -\frac{1}{2}\sqrt{\frac{1}{2}}(x-iy) \\ -\frac{1}{2}\sqrt{\frac{1}{2}}(x+iy) & z & -(x-iy) \\ -\frac{1}{2}\sqrt{\frac{1}{2}}(x+iy) & -(x+iy) & z \end{pmatrix}$$

$$= z^3 + \left[-\frac{1}{8}(x^2 + y^2) - \frac{1}{8}(x^2 + y^2) - (x^2 + y^2) \right] z \\ - \frac{1}{8}(x - iy)(x^2 + y^2) - \frac{1}{8}(x + iy)(x^2 + y^2)$$

$$= z^3 - \frac{5}{4}(x^2 + y^2)z - \frac{1}{4}x(x^2 + y^2) = \det(zI - xH_1 - yK_1).$$

We now show the polynomial $f(x, y, z) = z^3 - \frac{5}{4}(x^2 + y^2)z - \frac{1}{4}x(x^2 + y^2)$

is irreducible. Suppose $f(x,y,z)$ is reducible. Then it must have a linear factor $(z - \alpha x - \beta y)$, since the degree of $f(x,y,z)$ is three. Setting $x = 1$ and $y = i$, we see that $\alpha + i\beta$ is an eigenvalue of A . Since 0 is the only eigenvalue of A , $\alpha + i\beta = 0$. Setting $x = 1$ and $y = 0$ shows that α is an eigenvalue of H_1 and hence is real. Similarly, β is an eigenvalue of K_1 and hence is real. Thus, $\alpha + i\beta = 0$ implies $\alpha = \beta = 0$ and the linear factor must be z . Since $z^3 - \frac{5}{4}(x^2 + y^2) - \frac{1}{4}x(x^2 + y^2)$ is clearly not divisible by z , the polynomial is irreducible.

Finally, we show A is not unitarily similar to B . Suppose there is a unitary matrix U such that $U^*AU = B$. Then $U^*A^*U = B^*$. Hence, if $W(A, A^*)$ is any word in A and A^* ,

$$U^* [W(A, A^*)] U = W(U^*AU, U^*A^*U) = W(B, B^*).$$

Thus, we would have $\text{tr}(W(A, A^*)) = \text{tr}(W(B, B^*))$ [41]. We compute $(A^*)^2 A^2$ and $(B^*)^2 B^2$ and show they have different traces.

$$A^* = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{\frac{1}{2}} & 0 & 0 \\ 2 & \sqrt{\frac{1}{2}} & 0 \end{pmatrix} \quad \text{and} \quad (A^*)^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix}.$$

$$(A^*)^2 A^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

$$B^* = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{\frac{1}{2}} & 0 & 0 \\ \sqrt{\frac{1}{2}} & 2 & 0 \end{pmatrix} \quad \text{and} \quad (B^*)^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2\sqrt{\frac{1}{2}} & 0 & 0 \end{pmatrix}.$$

$$(B^*)^2 B^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2\sqrt{\frac{1}{2}} & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 2\sqrt{\frac{1}{2}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Since $(A^*)^2 A^2$ has trace $\frac{1}{4}$ and $(B^*)^2 B^2$ has trace 2, the matrix A is not unitarily similar to B .

Thus, we see the polynomial $\det(zI - xH - yK)$ uniquely determines $A = H + iK$, up to unitary similarity, if $n \leq 2$, but not, in general, if A is larger than 2×2 .

Section 3. Property L.

Recall from I.1 that A and B are said to have property P provided there exists an ordering $\alpha_1, \dots, \alpha_n$ of the eigenvalues of A and β_1, \dots, β_n of the eigenvalues of B such that the matrix $Q(A, B)$ has eigenvalues $Q(\alpha_i, \beta_i)$, $i = 1, \dots, n$ for any polynomial Q . M. Kac suggested that matrices for which $xA + yB$ has eigenvalues $x\alpha_i + y\beta_i$, for all values of x and y , be studied. Motzkin and Taussky [27, 28] were the first to study such matrices.

Definition. If there is an ordering, $\alpha_1, \dots, \alpha_n$, of the eigenvalues of A and, β_1, \dots, β_n , of the eigenvalues of B such that $xA + yB$ has eigenvalues $x\alpha_i + y\beta_i$ for all values of x and y , then A and B are said to have property L .

Notice that property P implies property L . However, pairs of matrices which have property L need not have property P [27, p. 112].

Thus, in general, no two of the statements

- (1) A and B commute,
- (2) A and B have property P ,
- (3) A and B have property L ,

are equivalent, but (1) implies (2) and (2) implies (3).

For pairs of Hermitian matrices, however, the three properties are equivalent [27].

Theorem (Motzkin, Taussky) [27]. If H and K are Hermitian and have property L , then $HK = KH$.

Now suppose A has eigenvalues $\alpha_1, \dots, \alpha_n$ and B has eigenvalues β_1, \dots, β_n . If A and B have property L , then $xA + yB$ has eigenvalues $x\alpha_i + y\beta_i$ for all values of x and y . Hence

$$f(x, y, z) = \det(zI - xA - yB) = \prod_{i=1}^n (z - \alpha_i x - \beta_i y)$$

Conversely, if $f(x, y, z)$ factors into n linear factors of the form $(z - \alpha_i x - \beta_i y)$, $i = 1, \dots, n$, then A and B must have property L . Furthermore, $\alpha_1, \dots, \alpha_n$ are the eigenvalues of A , β_1, \dots, β_n are the eigenvalues of B and $x\alpha_i + y\beta_i$, for $i = 1, \dots, n$ are the eigenvalues of $xA + yB$. So A and B have property L if and

only if $f(x,y,z)$ factors into n linear factors [28].

Recall that $A = H + iK$ is normal if and only if $HK = KH$.

By the Motzkin-Taussky theorem, and the remark above, we see that A is normal if and only if $f(x,y,z) = \det(zI - xH - yK)$ factors into linear factors. Now if A is normal, there is a unitary matrix U such that $U^*AU = \text{diag}(\alpha_1, \dots, \alpha_n)$, where $\alpha_1, \dots, \alpha_n$ are the eigenvalues of A . Write $\alpha_j = h_j + ik_j$, where h_j and k_j are real, $j = 1, \dots, n$.

$$\begin{aligned} U^*AU &= \text{diag}(\alpha_1, \dots, \alpha_n) = \text{diag}(h_1, \dots, h_n) + i \text{diag}(k_1, \dots, k_n) \\ &= U^*HU + i U^*KU. \end{aligned}$$

Note that U^*HU and U^*KU are again Hermitian, and that H has eigenvalues h_1, \dots, h_n and K has eigenvalues k_1, \dots, k_n . Thus, we have

$$\det(zI - xH - yK) = \prod_{j=1}^n (z - h_jx - k_jy)$$

so each linear factor corresponds to a diagonal element of U^*AU .

Section 4. A generalization of property L--factors of the characteristic polynomial.¹

We saw above that if A is unitarily similar to a diagonal matrix, then the n diagonal elements correspond to the n linear factors of $\det(zI - xH - yK)$. Suppose now that A is unitarily similar to a matrix

¹Gaines [12] considered a generalized L-property and proved that if A and B satisfy a certain commutator relation, then the characteristic polynomial of $xA + yB$ splits into linear and quadratic homogeneous factors. Taussky [46] has also studied a weak L-property.

to the blocks in the matrices. In II.3 we used a theorem of Motzkin and Taussky to show that if all the factors are linear, the converse also holds. However, the following example shows that in general, the converse does not hold when there are non-linear factors.

Example 5. Let $H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -a \end{pmatrix}$ where a is a nonzero

real number, and let $K = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

The matrices H and K are Hermitian and we showed in Example 3 (II.2, pp. 26-28) that

$$\det(zI - xH - yK) = z [(z - ax - y)(z + ax - y) - 3y^2],$$

but H and K have no common eigenvector. Hence, $f(x,y,z)$ factors into a linear factor and a quadratic factor, but H and K are not simultaneously, unitarily similar to matrices which are $D(1, 2)$.

In their second paper on matrices with property L [28] Motzkin and Taussky proved the following theorem.

Theorem (Motzkin, Taussky) [28, p. 395]. Let $xA + yB$ be a pencil in which all matrices are diagonalizable. Then A and B can be diagonalized by the same similarity and therefore they commute.

This result does not generalize to block diagonalization. The

following example exhibits a pair A, B of 3×3 matrices such that every matrix in the pencil is similar to a matrix which is $D(1, 2)$, but A and B cannot be simultaneously block diagonalized.

Example 6.

$$\text{Let } A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The matrix A is diagonal. The matrix B has two distinct eigenvalues, namely 0 and 2 , and hence is similar to a matrix which is $D(1, 2)$.

$$xA + yB = \begin{pmatrix} 0 & 0 & 0 \\ y & 2x+y & y \\ y & y & x+y \end{pmatrix}$$

We show that for each fixed x_0 and y_0 , the matrix $x_0A + y_0B$ is similar to a matrix which is $D(1, 2)$.

Consider the three column vectors $\begin{pmatrix} y \\ y \end{pmatrix}$, $\begin{pmatrix} 2x + y \\ y \end{pmatrix}$, and $\begin{pmatrix} y \\ x + y \end{pmatrix}$.

Case 1.

$$\det \begin{pmatrix} 2x_0 + y_0 & y_0 \\ y_0 & x_0 + y_0 \end{pmatrix} \neq 0.$$

Then the vectors $\begin{pmatrix} 2x_0 + y_0 \\ y_0 \end{pmatrix}$ and $\begin{pmatrix} y_0 \\ x_0 + y_0 \end{pmatrix}$ are linearly independent.

Hence there exist scalars r and s such that

$$\begin{pmatrix} y_0 \\ y_0 \end{pmatrix} = r \begin{pmatrix} 2x_0 + y_0 \\ y_0 \end{pmatrix} + s \begin{pmatrix} y_0 \\ x_0 + y_0 \end{pmatrix}.$$

$$\text{Let } S = \begin{pmatrix} 1 & 0 & 0 \\ -r & 1 & 0 \\ -s & 0 & 1 \end{pmatrix}. \quad \text{Then } S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ r & 1 & 0 \\ s & 0 & 1 \end{pmatrix}.$$

$$\begin{aligned} S^{-1}(x_0A + y_0B)S &= \begin{pmatrix} 1 & 0 & 0 \\ r & 1 & 0 \\ s & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2x_0 + y_0 & y_0 \\ 0 & y_0 & x_0 + y_0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2x_0 + y_0 & y_0 \\ 0 & y_0 & x_0 + y_0 \end{pmatrix}. \end{aligned}$$

Hence $x_0A + y_0B$ is similar to a matrix which is $D(1, 2)$.

Case 2. $\det \begin{pmatrix} 2x_0 + y_0 & y_0 \\ y_0 & x_0 + y_0 \end{pmatrix} = 0.$

$$\text{Then } (2x_0 + y_0)(x_0 + y_0) - y_0^2 = 2x_0^2 + 3x_0y_0 = 0.$$

Thus $x_0(2x_0 + 3y_0) = 0$. If $x_0 = 0$, then $x_0A + y_0B = y_0B$, which is similar to a matrix which is $D(1, 2)$. If $x_0 \neq 0$, then $2x_0 + 3y_0 = 0$, so $y_0 = -\frac{2}{3}x_0$. Then $x_0A + y_0B = x_0(A - \frac{2}{3}B) = \frac{x_0}{3}(3A - 2B)$.

$$3A - 2B = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 4 & -2 \\ -2 & -2 & 1 \end{pmatrix}$$

$\text{Det}(\lambda I - (3A - 2B)) = \lambda^3 - 5\lambda^2 = \lambda^2(\lambda - 5)$. Since $3A - 2B$ has two distinct eigenvalues, it is similar to a matrix which is $D(1, 2)$.

Thus, for every choice of x_0 and y_0 , the matrix $x_0A + y_0B$ is similar to a matrix which is $D(1,2)$.

Furthermore, the characteristic polynomial of $xA + yB$ factors into a linear factor and a quadratic factor.

$$f(x,y,z) = \det(zI - xA - yB) = \det \begin{pmatrix} z & 0 & 0 \\ -y & z-2x-y & -y \\ -y & -y & z-x-y \end{pmatrix}.$$

$$= z [(z - 2x - y)(z - x - y) - y^2] = z [z^2 - (3x + 2y)z + (2x^2 + 3xy)].$$

However, we now show that A and B are not simultaneously similar to matrices which are $D(1, 2)$. Suppose there were a nonsingular matrix S such that $S^{-1}AS$ and $S^{-1}BS$ were both $D(1, 2)$. Then $[A, B]^2$ would have to commute with A . (See page 14.) We compute $[A, B]^2$ and show that it does not commute with A .

$$[A, B] = AB - BA = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

$$[A, B]^2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ -2 & 0 & -1 \end{pmatrix}.$$

$$A [A, B]^2 = \begin{pmatrix} 0 & 0 & 0 \\ 2 & -2 & 0 \\ -2 & 0 & -1 \end{pmatrix} \quad \text{and} \quad [A, B]^2 A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Hence A and B cannot be simultaneously similar to matrices which are $D(1, 2)$. Thus, the theorem of Motzkin and Taussky which states that if

every matrix in the pencil $xA + yB$ is diagonalizable, then A and B commute, does not extend to cover block diagonalization.

III. Linear Factors of the Characteristic Polynomial.

Section 1. Linear factors of $\det(zI - xA - yB)$ and the eigenvalues of $A^{-1}B$.

If A and B have property L , they do not necessarily have property P . However, if A is nonsingular, one can say something about the eigenvalues of $A^{-1}B$.

Theorem (Motzkin, Taussky) [27, 48]. Let A and B have property L and suppose A is nonsingular. Let $\alpha_1, \dots, \alpha_n$ be the eigenvalues of A and β_1, \dots, β_n be the eigenvalues of B . Then the eigenvalues of $A^{-1}B$ are $\alpha_i^{-1}\beta_i$, for $i = 1, \dots, n$.

We employ the same argument used by Motzkin and Taussky to show that if $\det(zI - xA - yB)$ has a linear factor $(z - \alpha x - \beta y)$, and if A is nonsingular, $\alpha^{-1}\beta$ is an eigenvalue of $A^{-1}B$. Suppose $\det(zI - xA - yB) = (z - \alpha x - \beta y) g(x, y, z)$ where $g(x, y, z)$ has degree $n-1$. Then $\alpha x_0 + \beta y_0$ is an eigenvalue of $x_0A + y_0B$ for any pair of complex numbers x_0 and y_0 . Setting $x_0 = \alpha^{-1}\beta$ and $y_0 = -1$, we see that 0 is an eigenvalue of $\alpha^{-1}\beta A - B$. Hence $0 = \det(\alpha^{-1}\beta A - B) = (\det A)(\det(\alpha^{-1}\beta I - A^{-1}B))$. Thus, $\alpha^{-1}\beta$ is an eigenvalue of $A^{-1}B$.

Section 2. A pair of $n \times n$ Hermitian matrices, H and K , which do not have a common eigenvector but such that $\det(zI - xH - yK)$ has a linear factor.

Suppose $f(x,y,z) = \det(zI - xH - yK) = (z - \alpha x - \beta y) g(x,y,z)$, where $g(x,y,z)$ has degree $n-1$. Example 5 (II.4, p.35) showed that H and K need not have a common eigenvector. The matrices in that example were 3×3 . The following example shows how to construct a pair of $n \times n$ Hermitian matrices which do not have a common eigenvector, but such that $f(x,y,z)$ does have a linear factor.

Example 7.

Let $H = \begin{pmatrix} 0 & & & & \\ & \alpha_2 & & & \\ & & \alpha_3 & & \\ & & & \ddots & \\ & & & & \alpha_n \\ & & & & & 0 \end{pmatrix}$, where $\alpha_2, \dots, \alpha_n$ are distinct real numbers.

$$\text{Let } K = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

H and K are $n \times n$ Hermitian matrices. Consider $f(x,y,z)$. This polynomial is divisible by z if and only if $\det(xH + yK) = 0$. We will show how to choose $\alpha_2, \dots, \alpha_n$ so that $\det(xH + yK) = 0$.

$$xH + yK = \begin{pmatrix} 0 & y & y & \cdot & \cdot & \cdot & y \\ y & \alpha_2 x + y & y & \cdot & \cdot & \cdot & y \\ y & y & \alpha_3 x + y & \cdot & \cdot & \cdot & y \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ y & y & y & \cdot & \cdot & \cdot & \alpha_n x + y \end{pmatrix}.$$

$\text{Det}(xH + yK)$ is a homogeneous polynomial of degree n in x and y .

$$\det(xH + yK) = r_0 y^n + r_1 x y^{n-1} + \dots + r_{n-1} x^{n-1} y + r_n x^n.$$

We compute the coefficients r_0, r_1, \dots, r_n . Clearly, $r_n = r_{n-1} = 0$.

To get a term of the form $c x^{n-2} y^2$, we must take exactly $n-2$ nonzero diagonal elements. Hence,

$$r_{n-2} = (-1)^{\sum_{i=2}^n} \alpha_2 \alpha_3 \dots \alpha_{i-1} \alpha_{i+1} \dots \alpha_n$$

Now to get a term of the form $c x^{n-3} y^3$, we must take exactly $n-3$ nonzero diagonal elements. Thus

$$r_{n-3} = \det \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \times (\text{a sum of products of } n-3 \text{ distinct } \alpha_i \text{'s})$$

Since $\det \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 0$, the coefficient $r_{n-3} = 0$.

A similar argument shows $r_0 = r_1 = \dots = r_{n-4} = 0$.

Hence $\det(xH + yK) = r_{n-2}x^{n-2}y^2$ and $\det(xH + yK) = 0$ if and only if

$$(*) \quad r_{n-2} = (-1) \sum_{i=2}^n \alpha_2 \alpha_3 \cdots \alpha_{i-1} \alpha_{i+1} \cdots \alpha_n = 0.$$

Thus, if we choose $\alpha_2, \dots, \alpha_n$ to satisfy equation (*), then $\det(zI - xH - yK)$ will be divisible by z . If, in addition, $0, \alpha_2, \dots, \alpha_n$ are distinct, then the eigenvectors of H are all scalar multiples of the standard basis vectors. Since none of these is an eigenvector of K , the matrices H and K have no common eigenvector. Thus, if it is possible to find distinct, nonzero real numbers $\alpha_2, \dots, \alpha_n$ which satisfy equation (*), then one can construct Hermitian matrices H and K such that $\det(zI - xH - yK)$ is divisible by z , but H and K have no common eigenvector. Note that if $\alpha_2 = a$ and $\alpha_3 = -a$ we obtain the matrices in Example 5 (II.4, p. 35).

Section 3. A theorem about linear factors of high multiplicity.

As we saw above, the presence of a linear factor in the characteristic polynomial of the pencil $xH + yK$ is not a sufficient condition for H and K to have a common eigenvector. However, if the linear factor occurs with a high enough multiplicity, then we can infer the existence of a common eigenvector.

Before stating the main theorem of this chapter, we prove a lemma which will be needed in the proof.

Lemma 1. Suppose $f(x, y, z) = \det(zI - xH - yK) = (z - \alpha x - \beta y)^r g(x, y, z)$, where $z - \alpha x - \beta y$ does not divide $g(x, y, z)$. Then there exist real numbers a and b such that $\alpha a + \beta b$ is an eigenvalue of the

matrix $aH + bK$ of multiplicity exactly r . (Equivalently, $aH + bK - (\alpha a + \beta b)I$ has rank $n-r$.)

Proof. Since $f(x,y,z) = (z - \alpha x - \beta y)^r g(x,y,z)$, the number $\alpha x_0 + \beta y_0$ is an eigenvalue of multiplicity at least r for all x_0 and y_0 . Suppose, by contradiction, that $\alpha a + \beta b$ is an eigenvalue of $aH + bK$ of multiplicity greater than r for all real a and b . Then $(z - \alpha a - \beta b)^{r+1}$ divides the polynomial $f(a,b,z)$ for all real values of a and b . This is possible if and only if $(z - \alpha x - \beta y)^{r+1}$ divides $f(x,y,z)$, which contradicts the assumption that $z - \alpha x - \beta y$ does not divide $g(x,y,z)$. Hence, for some real a and b , the number $\alpha a + \beta b$ is an eigenvalue of $aH + bK$ of multiplicity exactly r . Since $aH + bK$ is Hermitian, this is equivalent to saying $aH + bK - (\alpha a + \beta b)I$ has rank $n-r$. ■

Theorem 4. Let H and K be $n \times n$ Hermitian matrices. Suppose $f(x,y,z) = \det(zI - xH - yK) = (z - \alpha x - \beta y)^r g(x,y,z)$, where $z - \alpha x - \beta y$ does not divide $g(x,y,z)$. Then if $r > \frac{n}{3}$, the matrices H and K have a common eigenvector corresponding to the eigenvalues α of H and β of K .

Proof. We first show that without loss of generality, we may assume $H - \alpha I$ has rank $n-r$ (i.e. α is an eigenvalue of H of multiplicity exactly r) and that $\alpha = \beta = 0$. By Lemma 1, there exist real numbers a and b such that $\alpha a + \beta b$ is an eigenvalue of $aH + bK$ of multiplicity exactly r . If we choose real numbers c and d so

that $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$, then the matrices $aH + bK$ and $cH + dK$

generate the same pencil as H and K . Since a, b, c and d are real, $aH + bK$ and $cH + dK$ are Hermitian. If we replace H and K by $aH + bK$ and $cH + dK$, the polynomial $f(x, y, z)$ undergoes a linear change of variable and hence has the same type of factorization. Hence, we may replace H and K by $aH + bK$ and $cH + dK$ and thus assume that $H - \alpha I$ has rank $n-r$.

Next, we replace H by $H - \alpha I$.

$$\begin{aligned} \det(zI - x(H - \alpha I) - yK) &= \det((z + \alpha x)I - xH - yK) \\ &= (z + \alpha x - \alpha x - \beta y)^r g(x, y, z + \alpha x) = (z - \beta y)^r g(x, y, z + \alpha x) \end{aligned}$$

Since $z - \alpha x - \beta y$ does not divide $g(x, y, z)$, we see $z - \beta y$ does not divide $g(x, y, z + \alpha x)$. Furthermore, H and $H - \alpha I$ have the same eigenvectors. Thus, without loss of generality, we may replace H by $H - \alpha I$ and thus assume $\alpha = 0$. Similarly, we may assume $\beta = 0$. Thus, we may assume $f(x, y, z) = z^r g(x, y, z)$, where z does not divide $g(x, y, z)$ and H has rank $n-r$. Zero is then an eigenvalue of H of multiplicity exactly r , and an eigenvalue of K of multiplicity at least r .

We now diagonalize H with a unitary similarity, U , and apply the same similarity to K . The matrices $U^* H U$ and $U^* K U$ are again Hermitian, so we may assume $H = \text{diag}(0, 0, \dots, 0, h_{r+1}, \dots, h_n)$ where $h_{r+1} h_{r+2} \dots h_n \neq 0$. Partition the matrix K into the following block form.

$$K = \left(\begin{array}{c|c} \overbrace{\begin{matrix} K_{11} \end{matrix}}^r & \overbrace{\begin{matrix} K_{12} \end{matrix}}^{n-r} \\ \hline \overbrace{\begin{matrix} K_{12}^* \end{matrix}}^{n-r} & \overbrace{\begin{matrix} K_{22} \end{matrix}}^{n-r} \end{array} \right) \begin{matrix} r \\ n-r \end{matrix}$$

where K_{11} is $r \times r$, K_{12} is $r \times (n-r)$ and K_{22} is $(n-r) \times (n-r)$.

Let $K_{22} = (k_{ij})$ where $i, j = r+1, \dots, n$. Since K is Hermitian, K_{22} is also Hermitian and $k_{ji} = \bar{k}_{ij}$.

We first show that $K_{11} = 0$.

$$xH + K = \left(\begin{array}{c|ccc} K_{11} & & & \\ \hline & xh_{r+1} + k_{r+1,r+1} & \dots & k_{r+1,n} \\ K_{12}^* & \vdots & & \vdots \\ & \bar{k}_{r+1,n} & \dots & xh_n + k_{nn} \end{array} \right)$$

$\det(zI - xH - K) = f(x, 1, z) = z^r g(x, 1, z)$. The coefficient of z^i in $\det(zI - xH - K)$ is $(-1)^{n-i}$ times the sum of all the principal $n-i \times n-i$ minors of the matrix $xH + K$. Since the coefficient of z^i is zero for $i = 0, 1, \dots, r-1$, the sum of all the principal $n-i \times n-i$ minors of the matrix $xH + K$ is zero for each i , for $i = 0, 1, \dots, r-1$.

For a fixed i , the sum of all the principal $n-i \times n-i$ minors of $xH + K$ is a polynomial in x of degree $n-r$, for each $i = 0, \dots, r-1$. The coefficient of x^{n-r} in this polynomial is

$$(h_{r+1} h_{r+2} \dots h_n) \times (\text{the sum of the principal } r-i \times r-i \text{ minors of the } r \times r \text{ matrix } K_{11}).$$

Since $h_{r+1} \dots h_n \neq 0$, the sum of all the principal $r-i \times r-i$ minors of K_{11} is equal to zero, for each $i = 0, \dots, r-1$. Thus the characteristic polynomial of K_{11} is $\lambda^r = 0$ and K_{11} is nilpotent. Since K_{11} is Hermitian, $K_{11} = 0$. We now have

$$H = \left(\begin{array}{c|c} \begin{matrix} 0 & 0 \\ \hline 0 & h_{r+1} \\ & \vdots \\ & h_n \end{matrix} & \begin{matrix} \\ \\ \end{matrix} \\ \hline \end{array} \right) \begin{matrix} \left. \vphantom{\begin{matrix} 0 & 0 \\ \hline 0 & h_{r+1} \\ & \vdots \\ & h_n \end{matrix}} \right\} r \\ \left. \vphantom{\begin{matrix} 0 & 0 \\ \hline 0 & h_{r+1} \\ & \vdots \\ & h_n \end{matrix}} \right\} n-r \end{matrix} \quad \text{and} \quad K = \left(\begin{array}{c|c} \begin{matrix} 0 & \\ \hline K_{12}^* & K_{22} \end{matrix} & \begin{matrix} \\ \\ \end{matrix} \\ \hline \end{array} \right) \begin{matrix} \left. \vphantom{\begin{matrix} 0 & \\ \hline K_{12}^* & K_{22} \end{matrix}} \right\} r \\ \left. \vphantom{\begin{matrix} 0 & \\ \hline K_{12}^* & K_{22} \end{matrix}} \right\} n-r \end{matrix}$$

A column vector $\xi = (x_1, \dots, x_n)^T$ is in the null space of H if and only if $x_{r+1} = x_{r+2} = \dots = x_n = 0$. We will show that if $r > \frac{n}{3}$, it is possible to find a nonzero vector $\xi = (x_1, \dots, x_r, 0, \dots, 0)^T$ which is in the null space of K . This will complete the proof.

$$\text{We have} \quad K \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ K_{12}^* \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \left. \vphantom{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ K_{12}^* \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} \\ \vdots \\ 0 \end{pmatrix}} \right\} r \\ \left. \vphantom{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ K_{12}^* \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} \\ \vdots \\ 0 \end{pmatrix}} \right\} n-r \end{matrix}$$

The matrix K_{12}^* is $(n-r) \times r$. If $r > \frac{n}{2}$, then the null space of H , which we denote by $\eta(H)$, has dimension greater than $\frac{n}{2}$. Also, $\eta(K)$ has dimension greater than $\frac{n}{2}$. Hence $\eta(H) \cap \eta(K) \neq 0$ and H and K have a common eigenvector with eigenvalue zero.

It remains to consider the case $r < \frac{n}{2}$. Then $r < n-r$ so $\text{rank}(K_{12}^*) \leq r$. If $\text{rank}(K_{12}^*) < r$, then there exists a nonzero

$$\text{solution to the system} \quad K_{12}^* \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and we are done.}$$

We now show K_{12}^* cannot have rank r . If K_{12}^* has rank r , then K_{12} also has rank r . Hence the r rows of K_{12} are linearly independent. The full matrix K has rank at most $n-r$, since 0 is an eigenvalue of multiplicity at least r . The first r rows of K are linearly independent, so these r rows, together with some choice of $n-2r$ rows from the $n-r \times r$ matrix $\begin{pmatrix} K_{12}^* & | & K_{22} \end{pmatrix}$ generate the row space of K . Relabel the rows so that the first $n-2r$ rows of $\begin{pmatrix} K_{12}^* & | & K_{22} \end{pmatrix}$ satisfy this condition. Write

$$\begin{pmatrix} K_{12}^* & | & K_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{r+1} & | & \beta_{r+1} \\ \vdots & & \vdots \\ \alpha_{n-r} & | & \beta_{n-r} \\ \hline \gamma_1 & | & \beta_{n-r+1} \\ \vdots & & \vdots \\ \gamma_r & | & \beta_n \end{pmatrix} \begin{matrix} \left. \vphantom{\begin{matrix} \alpha_{r+1} \\ \vdots \\ \alpha_{n-r} \end{matrix}} \right\} n-2r \\ \left. \vphantom{\begin{matrix} \beta_{n-r+1} \\ \vdots \\ \beta_n \end{matrix}} \right\} r \end{matrix}$$

where $\alpha_{r+1}, \dots, \alpha_{n-r}$ and $\gamma_1, \dots, \gamma_r$ are row vectors of length r and $\beta_{r+1}, \dots, \beta_n$ are row vectors of length $n-r$. Then each of the

r rows $\begin{pmatrix} \gamma_1 & \beta_{n-r+1} \\ \gamma_2 & \beta_{n-r+2} \\ \vdots & \vdots \\ \gamma_r & \beta_n \end{pmatrix}$ is expressible as a linear combination

of the first $n-r$ rows of the matrix K . Since K_{11} is the $r \times r$ zero matrix, each of the last r rows, $\gamma_1, \dots, \gamma_r$, of the matrix K_{12}^* is a linear combination of the first $n-2r$ rows of K_{12}^* . Hence,

the rank of K_{12}^* is at most $n-2r$. But we assumed K_{12}^* had rank r .

Thus, $r \leq n-2r$ and $r \leq \frac{n}{3}$, contrary to the hypothesis $r > \frac{n}{3}$.

Hence K_{12}^* cannot have rank r . ■

Remark. Repeated application of Theorem 4 yields a sharper result on the number of linearly independent, common eigenvectors of H and K . If H and K satisfy the hypotheses of the theorem, then they have a common eigenvector, corresponding to the eigenvalue α of H and β of K . Since H and K are Hermitian, Specht's theorem B (I.5, p. 14) implies that there is a unitary matrix U such that

$$U^* H U = \left(\begin{array}{c|cccc} \alpha & 0 & 0 & 0 & 0 \\ \hline 0 & & & & \\ \vdots & & H_1 & & \\ \vdots & & & & \\ 0 & & & & \end{array} \right) \quad \text{and} \quad U^* K U = \left(\begin{array}{c|cccc} \beta & 0 & 0 & 0 & 0 \\ \hline 0 & & & & \\ \vdots & & K_1 & & \\ \vdots & & & & \\ 0 & & & & \end{array} \right),$$

where H_1 and K_1 are $(n-1) \times (n-1)$ Hermitian matrices. Now

$$\det(zI - xH - yK) = \det(zI - xU^* H U - yU^* K U) = (z - \alpha x - \beta y) \det(zI - xH_1 - yK_1).$$

Hence $\det(zI - xH_1 - yK_1) = (z - \alpha x - \beta y)^{r-1} g(x, y, z)$ where $z - \alpha x - \beta y$

does not divide $g(x, y, z)$. If $r-1 > \frac{n-1}{3}$, we may apply Theorem 4 to

H_1 and K_1 and obtain a common eigenvector of H_1 and K_1 with eigenvalue α corresponding to H_1 and β corresponding to K_1 .

The matrices H and K will thus have two linearly independent common eigenvectors corresponding to the eigenvalue α of H and the eigenvalue β of K . Continuing in this manner, we see that

if $r-i > \frac{n-i}{3}$, H and K will have $i+1$ linearly independent common eigenvectors. ■

Section 4. Examples related to Theorem 4.

The following examples illustrate Theorem 4, show that the inequality $r > \frac{n}{3}$ is the best possible, and show that the theorem does not hold for pairs of matrices which are not Hermitian.

Example 8. This example illustrates Theorem 4.

$$\text{Let } H = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & s & 0 \\ 0 & 0 & 0 & -s \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 0 & 0 & b & b \\ 0 & 0 & ab & ab \\ b & ab & c & 0 \\ b & ab & 0 & -c \end{pmatrix}$$

where $a, b, c,$ and s are nonzero real numbers.

$$\begin{aligned} \det(zI - xH - yK) &= \det \begin{pmatrix} z & 0 & -by & -by \\ 0 & z & -aby & -aby \\ -by & -aby & z-sx-cy & 0 \\ -by & -aby & 0 & z+sx+cy \end{pmatrix} \\ &= z^4 + z^2 [-(sx + cy)^2 - a^2 b^2 y^2 - b^2 y^2 - a^2 b^2 y^2 - b^2 y^2] \\ &\quad + z [a^2 b^2 y^2 (sx + cy) - a^2 b^2 y^2 (sx + cy) + b^2 y^2 (sx + cy) - b^2 y^2 (sx + cy)] \\ &= z^2 [z^2 - (sx + cy)^2 - 2a^2 b^2 y^2 - 2b^2 y^2]. \end{aligned}$$

Thus H and K satisfy the hypotheses of Theorem 4, with $r = 2$ and $n = 4$.

Note that $\begin{pmatrix} a \\ -1 \\ 0 \\ 0 \end{pmatrix}$ is a common eigenvector of H and K with eigenvalue 0.

Example 9. Let $H = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -a \end{pmatrix}$ where a is a nonzero real number.

Let $K = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. These matrices were used in Example 3 (pp. 26-28) and in Example 5 (p. 35).

We showed there that $\det(zI - xH - yK) = z [(z - ax - y)(z + ax - y) - 3y^2]$ and that H and K have no common eigenvector. Note that in this example, $n = 3$, and $r = 1 = \frac{n}{3}$ and the conclusion of Theorem 4 does not hold. Thus, this example shows that the inequality $r > \frac{n}{3}$ is the best possible.

Example 10. We use the previous example to construct a pair of 6×6 Hermitian matrices for which $r = 2 = \frac{6}{3}$ and the conclusion of Theorem 4 does not hold.

Let $H = \begin{pmatrix} H_1 & 0 \\ 0 & H_1 \end{pmatrix}$ and $K = \begin{pmatrix} K_1 & 0 \\ 0 & K_1 \end{pmatrix}$ where H_1 and

K_1 are the 3×3 matrices used in Example 9. Then

$$\begin{aligned} \det(zI - xH - yK) &= [\det(zI - xH_1 - yK_1)]^2 \\ &= z^2 [(z - ax - y)(z + ax - y) - 3y^2]^2. \end{aligned}$$

Thus $r = 2$ and $n = 6$, so $r = \frac{n}{3}$.

We now show H and K have no common eigenvector. Suppose α is a common eigenvector of H and K . Write

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad \text{where } \alpha_1 \text{ and } \alpha_2 \text{ are } 3 \times 1 \text{ column vectors.}$$

$$\text{Then } H(\alpha) = \begin{pmatrix} H_1\alpha_1 \\ H_1\alpha_2 \end{pmatrix} \quad \text{and} \quad K(\alpha) = \begin{pmatrix} K_1\alpha_1 \\ K_1\alpha_2 \end{pmatrix}. \quad \text{Hence } \alpha_1 \text{ and } \alpha_2$$

must be common eigenvectors of H_1 and K_1 . Since H_1 and K_1 have no common eigenvectors, H and K cannot have a common eigenvector.

Thus, $r = \frac{n}{3}$ and the conclusion of Theorem 4 does not hold.

Example 11. The construction used in Example 10 can be generalized to produce a pair of $3m \times 3m$ matrices H and K such that

$$\det(zI - xH - yK) = z^m [(z - ax - y)(z + ax - y) - 3y^2]^m.$$

Simply let H be the direct sum of m copies of the 3×3 matrix

$$H_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -a \end{pmatrix} \quad \text{and let } K \text{ be the direct sum of } m \text{ copies}$$

$$\text{of the } 3 \times 3 \text{ matrix } K_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad \text{Then the same}$$

argument used in Example 10 shows that H and K have no common eigenvector, although $r = m = \frac{3m}{3} = \frac{n}{3}$.

We now give some examples which show the condition that H and K be Hermitian is needed.

Example 12.

$$\text{Let } A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & \beta & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

where β is a complex number. The matrices A and B are not Hermitian.

$$\begin{aligned} \det(zI - xA - yB) &= \det \begin{pmatrix} z & -x & 0 \\ -y & z - \beta y & -x \\ 0 & y & z \end{pmatrix} \\ &= z^2(z - \beta y) - xyz + xyz = z^2(z - \beta y). \end{aligned}$$

Thus we have $r = 2$ and $n = 3$ so $r > \frac{n}{3}$. The matrix A has

exactly one linearly independent eigenvector, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. This is not

an eigenvector of B . Hence A and B have no common eigenvector and the conclusion of Theorem 4 does not hold.

Example 13.

$$\text{Let } A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

$$\text{Let } B = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

$$\det(zI - xA - yB) = \det \begin{pmatrix} z-x-y & -y & 0 & -y \\ -y & z+x-2y & -y & 0 \\ 0 & 0 & z & 0 \\ 0 & 0 & 0 & z \end{pmatrix}$$

$$= z^2 \det \begin{pmatrix} z-x-y & -y \\ -y & z+x-2y \end{pmatrix} = z^2 [(z-x-y)(z+x-2y) - y^2].$$

So $n = 4$ and $r = 2 > \frac{n}{3} = \frac{4}{3}$. The eigenvectors of A are

$(1 \ 0 \ 0 \ 0)^T$ with eigenvalue 1, $(0 \ 1 \ 0 \ 0)^T$ with eigenvalue -1,

and $\{(0 \ 0 \ a \ b) \mid a, b \text{ are complex numbers, not both zero}\}$ with eigenvalue zero.

$$\text{Now } B \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad B \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \quad \text{while} \quad B \begin{pmatrix} 0 \\ 0 \\ a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a \\ 0 \\ 0 \end{pmatrix}$$

so no eigenvector of A is an eigenvector of B . Note that both A and B can be diagonalized, so Theorem 4 does not hold for a pair of diagonalizable matrices.

IV. A Theorem About an Algebra Generated by a Pair of Matrices
Which Satisfy Polynomials of Degree Two.

In this chapter we prove a theorem which will be used later to determine the structure of H and K when $f(x,y,z)$ is a power of a quadratic polynomial.

Theorem 5. Let C and D be $n \times n$ matrices with elements in a field F , where F does not have characteristic two. Suppose there are polynomials of degree two, $f(x)$ and $g(x)$, with coefficients in F , such that $f(C) = 0$ and $g(D) = 0$. Let $\mathcal{A} = \mathcal{A}(C, D)$ be the algebra generated by C, D , and I over F . Then the dimension of \mathcal{A} , considered as a vector space over F , is at most $2n$.

Proof. Let $f(x) = x^2 + \alpha x + \beta$ and $g(x) = x^2 + \gamma x + \delta$, where α, β, γ , and $\delta \in F$. Since $f(C) = 0$, we have $C^2 + \alpha C + \beta I = 0$. Hence $(C + \frac{1}{2} \alpha I)^2 + (\beta - \frac{1}{4} \alpha^2) I = 0$ and $(C + \frac{1}{2} \alpha I)^2 = (\frac{1}{4} \alpha^2 - \beta) I$. Similarly, $(D + \frac{1}{2} \gamma I)^2 = (\frac{1}{4} \gamma^2 - \delta) I$. Since $\mathcal{A}(C, D) = \mathcal{A}(C + \frac{1}{2} \alpha I, D + \frac{1}{2} \gamma I)$, we may, without loss of generality, assume that $C^2 = cI$ and $D^2 = dI$ for some elements c and d of F . Now let $E = C + D$ and $G = C - D$. Since $C = \frac{E + G}{2}$ and $D = \frac{E - G}{2}$, the matrices E and G also generate the algebra \mathcal{A} . We have

$$E^2 = C^2 + CD + DC + D^2 \quad \text{and} \quad G^2 = C^2 - CD - DC + D^2.$$

Hence, $E^2 + G^2 = 2(C^2 + D^2) = 2(c + d)I$. Since $C^2 = \frac{E^2 + EG + GE + G^2}{4}$,
 $cI = \left(\frac{c+d}{2}\right)I + \frac{EG + GE}{4}$. Thus, $E, G,$ and I generate the algebra
 $\mathcal{A} = \mathcal{A}(C, D)$ and satisfy the relations

$$(1) \quad G^2 = 2(c + d)I - E^2$$

$$(2) \quad GE = 2(c - d)I - EG$$

The set of all words in $E, G,$ and I is a linear spanning set for \mathcal{A} , considered as a vector space over F . Relation (2) shows that any word in E and G may be expressed as a linear combination of words of the form $E^r G^s$, where r and s are nonnegative integers. Relation (1) shows that $E^r G^s$ can be expressed as a linear combination of words of the form E^k and $E^k G$, where k is a nonnegative integer.

Since E is an $n \times n$ matrix, the Cayley-Hamilton theorem insures that E^k can be expressed as a linear combination of $I, E, E^2, \dots, E^{n-1}$ for any nonnegative integer k . Hence, the set

$$\{I, E, E^2, \dots, E^{n-1}, G, EG, E^2G, \dots, E^{n-1}G\}$$

is a linear spanning set for \mathcal{A} and the dimension of \mathcal{A} as a vector space over F is at most $2n$. ■

Corollary 3. Let C and D satisfy the hypotheses of Theorem 5 and assume further that F is algebraically closed. Then \mathcal{A} is simultaneously similar to an algebra which is $BT^*(n_1, \dots, n_t)$ and each $n_i \leq 2$.

Proof. We may assume \mathcal{A} is already $BT^*(n_1, \dots, n_t)$. Let C_i and D_i denote the i 'th diagonal blocks of C and D , respectively. Let

\mathcal{A}_i be the algebra of all $n_i \times n_i$ matrices which occur as the i 'th diagonal block of some matrix in \mathcal{A} . Note that C_i and D_i generate \mathcal{A}_i , as an algebra over F . Since C and D satisfy the hypotheses of Theorem 5, so do C_i and D_i . Hence, the dimension of \mathcal{A}_i as a vector space over F is at most $2n_i$.

Since \mathcal{A} is $BT^*(n_1, \dots, n_t)$, the algebra \mathcal{A}_i is an irreducible matrix algebra of $n_i \times n_i$ matrices over F . The field F is algebraically closed, so we may use the theorem of Burnside (I.3, p.10) to deduce that $\mathcal{A}_i = M_{n_i}(F)$. Hence \mathcal{A}_i has dimension n_i^2 as a vector space over F . Therefore $n_i^2 \leq 2n_i$ and $n_i \leq 2$. ■

Corollary 4. Let A be an $n \times n$ complex matrix and suppose the minimal polynomial of A has degree two. Then there is a unitary matrix U such that U^*AU is block diagonal with blocks of size one or two.

Proof. Let \mathcal{A} be the algebra generated by A , A^* and I over \mathbb{C} . Since the minimal polynomial of A^* has the same degree as the minimal polynomial of A , the matrices A and A^* satisfy the hypothesis of Theorem 5. Hence, by Corollary 3, the algebra \mathcal{A} is similar to an algebra which is $BT^*(n_1, \dots, n_t)$ where $n_i \leq 2$ for $i = 1, \dots, t$. Since $\mathcal{A} = \mathcal{A}^*$, Specht's theorem B (I.5, p.17) implies that there is a unitary matrix U such that $U^*\mathcal{A}U$ is $D(n_1, \dots, n_t)$. Hence U^*AU is block diagonal with blocks of size one or two. ■

V. The Minimal Polynomial of a Pencil.

Section 1. Definition of the minimal polynomial of a pencil and its relation to the characteristic polynomial.

Kippenhahn [21, pp. 210-212] considered the minimal polynomial of the pencil $xH + yK$ and made a conjecture which will be discussed in Chapter VII. In this chapter, we review Kippenhahn's work, sometimes using a different point of view.

Definition. Let $m(x,y,z)$ be a homogeneous polynomial of degree k with complex coefficients. The polynomial $m(x,y,z)$ is called the minimal polynomial of the pencil $xA + yB$ provided it satisfies the following three conditions.

- (1) $m(x, y, xA + yB) = 0$.
- (2) If $g(x,y,z)$ is homogeneous of degree j and $g(x, y, xA + yB) = 0$, then $j \geq k$.
- (3) The coefficient of z^k in $m(x,y,z)$ is 1.

One may view $f(x,y,z) = \det(zI - xA - yB)$ as the usual characteristic polynomial of the matrix $xA + yB$ considered as a single $n \times n$ matrix with entries in the polynomial ring $\mathbb{C}[x, y]$. The minimal polynomial, $m(x,y,z)$ may then be considered as the minimal polynomial of the matrix $xA + yB$ with entries in $\mathbb{C}[x, y]$. Now factor $f(x,y,z)$ into irreducible factors.

$$f(x, y, z) = [\pi_1(x, y, z)]^{r_1} [\pi_2(x, y, z)]^{r_2} \dots [\pi_j(x, y, z)]^{r_j}$$

where $\pi_1, \pi_2, \dots, \pi_j$ are distinct irreducible polynomials and r_1, r_2, \dots, r_j are positive integers. Since $f(x, y, z)$ is homogeneous, so are the polynomials π_1, \dots, π_j [13, p. 125]. Since $m(x, y, z)$ and $f(x, y, z)$ are the minimal and characteristic polynomials, respectively, of the $n \times n$ matrix $xA + yB$ with entries in $\mathbb{C}[x, y]$, it follows that

$$m(x, y, z) = [\pi_1(x, y, z)]^{s_1} [\pi_2(x, y, z)]^{s_2} \dots [\pi_j(x, y, z)]^{s_j}$$

where $1 \leq s_i \leq r_i$ for $i = 1, \dots, j$.

Section 2. The minimal polynomial of $xH + yK$.

We saw above that $m(x, y, z) = \pi_1^{s_1} \dots \pi_j^{s_j}$. For the pencil $xH + yK$, where H and K are Hermitian, we have $s_1 = s_2 = \dots = s_j = 1$. Kippenhahn stated this fact but did not give the details of the proof.

Proposition 2. Let $f(x, y, z) = \det(zI - xH - yK) = \pi_1^{r_1} \dots \pi_j^{r_j}$ and let $m(x, y, z)$ be the minimal polynomial of $xH + yK$. Then $m(x, y, z) = \pi_1 \pi_2 \dots \pi_j$.

Proof.¹ Let $m_0(x, y, z) = \pi_1 \dots \pi_j$. Then $m_0(x, y, xH + yK)$ is a polynomial expression of the matrix $xH + yK$ and hence is an $n \times n$ matrix of polynomials in the variables x and y . Let

$$m_0(x, y, xH + yK) = \left(g_{ij}(x, y) \right)_{i, j = 1, \dots, n}$$

¹Thanks are due to Professor Stephen Kleiman for help with this proof.

If x_0 and y_0 are real numbers, the matrix $x_0 H + y_0 K$ is Hermitian and hence is diagonalizable. Thus $m_0(x_0, y_0, x_0 H + y_0 K)$ is the zero matrix for all real numbers x_0 and y_0 . This implies $g_{ij}(x_0, y_0) = 0$ for all real numbers x_0 and y_0 and $i, j = 1, \dots, n$. Hence $g_{ij}(x, y)$ must be the zero polynomial for all $i, j = 1, \dots, n$ and we have $m_0(x, y, xH + yK) = 0$. Therefore $m_0(x, y, z) = \pi_1 \dots \pi_j = m(x, y, z)$, the minimal polynomial of $xH + yK$. ■

The following example shows that the conclusion of Proposition 2 need not hold if the matrices are not Hermitian.

Example 14.

$$\text{Let } A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and } B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & \beta & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

where β is any complex number.

$$\det(zI - xA - yB) = \det \begin{pmatrix} z & -x & 0 \\ -y & z - \beta y & -x \\ 0 & y & z \end{pmatrix}$$

$$= z^2(z - \beta y) - xyz + xyz = z^2(z - \beta y) = f(x, y, z).$$

We now show that $(xA + yB)(xA + yB - \beta yI) \neq 0$ and hence that the minimal polynomial is equal to $f(x, y, z) = z^2(z - \beta y)$.

$$(xA + yB)(xA + yB - \beta yI) = \begin{pmatrix} 0 & x & 0 \\ y & \beta y & x \\ 0 & -y & 0 \end{pmatrix} \begin{pmatrix} -\beta y & x & 0 \\ y & 0 & x \\ 0 & -y & -\beta y \end{pmatrix}$$

$$= \begin{pmatrix} xy & 0 & x^2 \\ 0 & 0 & 0 \\ -y^2 & 0 & -xy \end{pmatrix} \neq 0.$$

Section 3. The case where the minimal polynomial has degree two.

We now study the situation where the minimal polynomial of $xH + yK$ has degree one or two. Note that if the minimal polynomial has degree one, then H and K are both scalar matrices, so this case is trivial. If the minimal polynomial has degree two, but factors into linear factors, then $f(x, y, z)$ also factors into linear factors. Hence H and K have property L, and by the theorem of Motzkin and Taussky (II.3, p. 32) H and K commute. It remains to consider the case where $m(x, y, z)$ is an irreducible quadratic.

Kippenhahn [21, pp. 211-212] proved that if the minimal polynomial has degree one or two, and $A = H + iK$ is not reducible by a unitary transformation, then A must be of size 1×1 or 2×2 . Kippenhahn's proof uses ideas which are similar to the ones used in the proof of Theorem 5 (IV). We use Theorem 5 to prove this result and then apply Theorem 3 (II.2, p.25) to obtain a more detailed result.

Theorem 6. Let $A = H + iK$ be a complex $n \times n$ matrix, where $n = 2m$. Suppose $f(x, y, z) = \det(zI - xH - yK) = [g(x, y, z)]^m$ where $g(x, y, z)$ is an irreducible polynomial of degree two. Then A is unitarily similar to a block diagonal matrix D which is the direct sum of m copies of a single 2×2 matrix.

Proof. By Proposition 2 (V.2, p. 59), $g(x,y,z)$ is the minimal polynomial of the pencil $xH + yK$. Setting $x = 1$ and $y = i$, we see that $g(1, i, A) = 0$. Since $g(1, i, z)$ is a polynomial of degree two in z , Corollary 4 (IV, p. 57) implies that A is unitarily similar to a block diagonal matrix with blocks of size 2×2 or 1×1 . Since $f(x,y,z) = [g(x,y,z)]^m$ and $g(x,y,z)$ is irreducible, all of the blocks must be 2×2 . Thus, A is unitarily similar to a matrix of the form

$$\begin{pmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \\ & & & A_m \end{pmatrix}$$

where each A_j is 2×2 . Let $A_j = H_j + iK_j$, for $j = 1, \dots, m$ and

$$\text{set } H = \begin{pmatrix} H_1 & & \\ & H_2 & \\ & & \ddots \\ & & & H_m \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} K_1 & & \\ & K_2 & \\ & & \ddots \\ & & & K_m \end{pmatrix}.$$

$$\text{Then } f(x,y,z) = \prod_{j=1}^m \det(zI - xH_j - yK_j) = [g(x,y,z)]^m.$$

Since $g(x,y,z)$ is irreducible, $\det(zI - xH_j - yK_j) = g(x,y,z)$

for all $j = 1, \dots, m$. By Theorem 3 (II.2, p. 25), all of the A_j 's are unitarily similar. Hence there exist 2×2 unitary matrices

U_2, U_3, \dots, U_m such that $U_j^* A_j U_j = A_1$ for $j = 2, \dots, m$. If we set

$$U = \left(\begin{array}{cc|c} 1 & 0 & \\ \hline 0 & 1 & \\ & & U_2 \\ & & \\ & & \\ & & \\ & & U_m \end{array} \right)$$

then $U^* \left(\begin{array}{c} A_1 \\ \\ A_2 \\ \\ \\ \\ A_m \end{array} \right) U = \left(\begin{array}{c} A_1 \\ \\ A_1 \\ \\ \\ \\ A_1 \end{array} \right) = D$

and A is unitarily similar to a matrix which is the direct sum of m identical 2×2 blocks. ■

Remark: Theorem 3 was used in the proof of Theorem 6 to show that all of the 2×2 blocks A_j must be unitarily similar. Since 3×3 matrices with the same characteristic polynomial $f(x,y,z)$ need not be unitarily similar (see Examples 3 and 4 of II.2, pp. 26-31) the conclusion of Theorem 6 need not hold if $g(x,y,z)$ is a cubic polynomial.

Kippenhahn [21, pp. 205-206] gives an example of an $n \times n$ matrix $A = H + iK$ which is not similar to a block diagonal matrix, but such that $\det(zI - xH - yK)$ factors into m quadratic factors when $n = 2m$ is even, and factors into m quadratic factors and one linear factor when $n = 2m + 1$ is odd. This shows that the splitting

of A into 2×2 blocks in Theorem 6 is due to the fact that the single quadratic factor $g(x,y,z)$ is repeated m times, and not merely to the fact that $f(x,y,z)$ splits into quadratic factors. We illustrate this for $n = 4 = 2(2)$ with the 4×4 case of Kippenhahn's example.

Example 15 (Kippenhahn) [21, pp. 205-206].

$$\text{Let } A = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 \end{pmatrix} = H + iK.$$

$$H = \begin{pmatrix} -1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & -1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & -1 \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 0 & -\frac{i}{2} & 0 & 0 \\ \frac{i}{2} & 0 & -\frac{i}{2} & 0 \\ 0 & \frac{i}{2} & 0 & -\frac{i}{2} \\ 0 & 0 & \frac{i}{2} & 0 \end{pmatrix}.$$

$$\det(zI - xH - yK) = \begin{vmatrix} z+x & \frac{1}{2}(x+iy) & 0 & 0 \\ \frac{1}{2}(x-iy) & z+x & \frac{1}{2}(x+iy) & 0 \\ 0 & \frac{1}{2}(x-iy) & z+x & \frac{1}{2}(x+iy) \\ 0 & 0 & \frac{1}{2}(x-iy) & z+x \end{vmatrix}$$

Expanding this by using cofactors of elements in the first column, we get

$$\begin{aligned} f(x,y,z) &= (z+x)[(z+x)^3 - \frac{1}{2}(x^2+y^2)(z+x)] \\ &\quad - \frac{1}{4}(x^2+y^2)[(z+x)^2 - \frac{1}{4}(x^2+y^2)] \\ &= (z+x)^4 - \frac{3}{4}(x^2+y^2)(z+x)^2 + \frac{1}{16}(x^2+y^2)^2 \end{aligned}$$

$$= \left[(z + x)^2 - \left(\frac{3 + \sqrt{5}}{8} \right) (x^2 + y^2) \right] \left[(z + x)^2 - \left(\frac{3 - \sqrt{5}}{8} \right) (x^2 + y^2) \right].$$

Hence $f(x,y,z)$ factors into two distinct quadratic factors, but the matrix A is not similar to a matrix which is block diagonal.

VI. Geometric Interpretations of $f(x,y,z) = 0$.

In this chapter we discuss two ways of viewing the characteristic polynomial $f(x,y,z) = \det(zI - xH - yK)$ geometrically. Motzkin and Taussky [28] studied the algebraic curve in the projective x,y,z -plane defined by the equation $\det(zI - xA - yB) = 0$. Kippenhahn [21] showed that the equation $\det(zI + xH + yK) = 0$, considered as an equation in line coordinates, defines an algebraic curve such that the closed convex hull of this curve is the numerical range (field of values) of the matrix $A = H + iK$. Murnaghan [30] used a different approach to derive this curve and also showed that its closed, convex hull is the numerical range of the matrix A .

Section 1. Geometric background.

In this section we review some terminology and facts from geometry [5, 33, 37, 49].

We will work over the field of complex numbers. A point in nonhomogeneous point coordinates is an ordered pair of complex numbers (x, y) . If x and y are real numbers, (x, y) is called a real point. A point in homogeneous point coordinates is an ordered triple of complex numbers, (x, y, z) , not all zero. If ρ is any nonzero complex number, then (x, y, z) and $(\rho x, \rho y, \rho z)$ denote the same point. Thus, while a point in homogeneous point coordinates is given by three numbers, $x, y,$ and $z,$ there only two independent

ratios, $\frac{x}{z}$ and $\frac{y}{z}$. One identifies the point given by homogeneous point coordinates (x, y, z) with the point $(\frac{x}{z}, \frac{y}{z})$ given in nonhomogeneous point coordinates. The point (x, y) becomes $(x, y, 1)$ in homogeneous coordinates. Any point in homogeneous coordinates whose third coordinate is zero (i.e. with $z = 0$) is a point at infinity.

The set of all points satisfying a homogeneous equation of degree one, $lx + my + nz = 0$, where l, m , and n are fixed complex numbers is a line. If $g(x, y, z)$ is a homogeneous polynomial of degree n , then the set of all points satisfying the equation $g(x, y, z) = 0$ is an algebraic curve of degree n . A curve of degree two is a conic.

Now let $lx + my + nz = 0$ be a fixed line and let $g(x, y, z) = 0$ be an algebraic curve of degree n . Setting $z = 1$ yields the nonhomogeneous equations

$$lx + my + n = 0$$

$$g(x, y, 1) = 0.$$

Assume $m \neq 0$. Then $y = \frac{-n - lx}{m}$ and the solutions to

$g(x, \frac{-n - lx}{m}, 1) = 0$ are the x -coordinates of the points common

to both the line and the curve. Since g has degree n , there are

precisely n solutions to the equation $g(x, \frac{-n - lx}{m}, 1) = 0$, where

multiple roots are counted according to their multiplicities. Thus,

a line intersects an algebraic curve of degree n in n points,

counted according to their multiplicities.

Section 2. The method of Motzkin and Taussky--the characteristic curve.

The geometric treatment discussed in this section is due to Motzkin and Taussky [28]. They used the algebraic curve C , determined by the equation $f(x,y,z) = \det(zI - xA - yB) = 0$, to study the pencil $xA + yB$.¹ The curve C is called the characteristic curve of the pencil $xA + yB$. Note that C is an algebraic curve of degree n and, when A and B have property L, C splits into n lines.

Since $\det(zI) \neq 0$, the point $P = (0, 0, 1)$ is not on C . The line $y_0x - x_0y = 0$ through P meets C in n points, (x_0, y_0, z_i) , $i = 1, \dots, n$. Since $\det(z_i - x_0A - y_0B) = 0$, the z_i 's are the n eigenvalues of the matrix $x_0A + y_0B$. A tangent of C , at a point T of C , is defined as a line having an intersection multiplicity $m > 1$ at T . The point T is called a point of contact.² Thus, the line $y_0x - x_0y$ is tangent to C if and only if the matrix $x_0A + y_0B$ has a multiple eigenvalue. To every multiple eigenvalue there corresponds a point of contact such that the intersection multiplicity there equals the multiplicity of the eigenvalue.

¹Motzkin and Taussky dealt with the case where A and B are $n \times n$ matrices over a field F ; we shall restrict ourselves to complex matrices.

²For a more complete discussion of this definition and its relation to other definitions of tangent, see Motzkin and Taussky [28, p. 390].

Section 3. The case $\det(zI - xH - yK) = (z - \alpha x - \beta y)^r [g(x,y,z)]^s$,
where $g(x,y,z)$ has degree two.

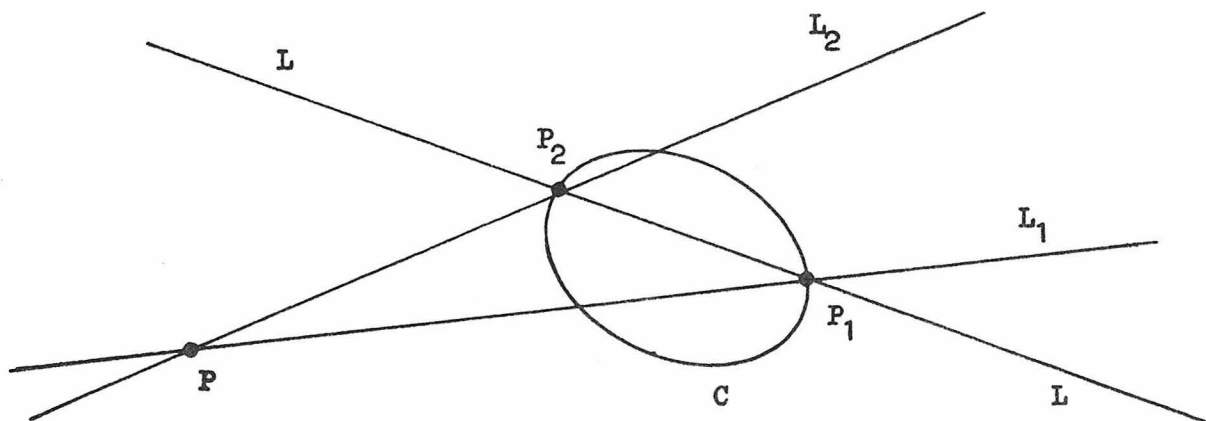
We now return to the pencil $xH + yK$. If x_0 and y_0 are real numbers, the matrix $x_0H + y_0K$ is Hermitian and hence has real eigenvalues. Thus, the line $y_0x - x_0y = 0$ intersects the curve $f(x,y,z) = 0$ in the n real points (x_0, y_0, z_i) , $i = 1, \dots, n$ where z_1, \dots, z_n are the eigenvalues of $x_0H + y_0K$.

Suppose $f(x,y,z) = \det(zI - xH - yK) = (z - \alpha x - \beta y)^r [g(x,y,z)]^s$, where r and s are positive integers and $g(x,y,z) = 0$ is an irreducible curve of degree two. Thus $g(x,y,z) = 0$ is an irreducible conic. The line $z - \alpha x - \beta y = 0$ intersects the conic in two points which may or may not be real points. If these points are real and distinct, then $f(x,y,z)$ determines the structure of $A = H + iK$ up to unitary similarity.

Theorem 7. Let $f(x,y,z) = \det(zI - xH - yK) = (z - \alpha x - \beta y)^r [g(x,y,z)]^s$, where r and s are positive integers and $g(x,y,z) = 0$ is an irreducible conic. If the line $z - \alpha x - \beta y = 0$ intersects the conic $g(x,y,z) = 0$ in two distinct real points, then the matrix $A = H + iK$ is unitarily similar to a block diagonal matrix which is the direct sum of s identical 2×2 blocks and r identical 1×1 blocks.

Proof. Let L denote the line $z - \alpha x - \beta y = 0$ and let C denote the conic $g(x,y,z) = 0$. The line L intersects C in two distinct real points, $P_1 = (a_1, b_1, c_1)$ and $P_2 = (a_2, b_2, c_2)$. Let L_1 be the line through $P = (0, 0, 1)$ and P_1 , and let L_2 be the line through

P and P_2 . Since P is not a point of L , but P_1 and P_2 are points of L , the lines L_1 and L_2 must be distinct. The equation of L_1 is $b_1x - a_1y = 0$ and the equation of L_2 is $b_2x - a_2y = 0$.



$$P = (0, 0, 1)$$

$$C : g(x, y, z) = 0$$

$$L : z - \alpha x - \beta y = 0$$

$$P_1 = (a_1, b_1, c_1)$$

$$L_1 : b_1x - a_1y = 0$$

$$P_2 = (a_2, b_2, c_2)$$

$$L_2 : b_2x - a_2y = 0$$

Let $H_1 = a_1H + b_1K$ and $K_1 = a_2H + b_2K$. Since $a_1, b_1, a_2,$ and b_2 are real numbers, H_1 and K_1 are Hermitian matrices. Since

L_1 and L_2 are distinct lines, $\det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \neq 0$, and H_1

and K_1 also generate the pencil $xH + yK$. Also, the algebra, $\mathcal{A}(H, K)$, generated by $I, H,$ and K over \mathbb{C} is the same as the algebra, $\mathcal{A}(H_1, K_1)$, generated by $I, H_1,$ and K_1 over \mathbb{C} .

We now apply the techniques of Motzkin and Taussky. The line L_1 intersects the curve $f(x, y, z) = 0$ in n points. Since

$f(x,y,z) = (z - \alpha x - \beta y)^r [g(x,y,z)]^s$ and $g(x,y,z)$ has degree two, there are at most three distinct intersection points. One lies on the line L and the other two are on the conic C . However, L_1 intersects the curve $f(x,y,z) = 0$ in the point P_1 , which lies on both L and C . Hence L_1 intersects $f(x,y,z) = 0$ in at most two distinct points. Thus the matrix $H_1 = a_1 H + b_1 K$ has at most two distinct eigenvalues. A similar argument using the line L_2 shows that $K_1 = a_2 H + b_2 K$ has at most two distinct eigenvalues. Since H_1 and K_1 are Hermitian, this implies that the minimal polynomials of H_1 and K_1 are of degree at most two. The matrices H_1 and K_1 then satisfy the hypothesis of Theorem 5 (IV, p. 55) and the dimension of the algebra $\mathcal{A}(H_1, K_1)$ is at most $2n$.

Now $\mathcal{A}(H_1, K_1) = \mathcal{A}(H, K) = \mathcal{A}$ satisfies the condition $\mathcal{A} = \mathcal{A}^*$ for H and K are Hermitian. By Corollary 3 (IV, p. 56) and Specht's theorem B (I.5, p. 17), there is a unitary matrix U such that $U^* \mathcal{A} U$ is $D(n_1, \dots, n_t)$, where for each $i = 1, \dots, t$, we have $n_i \leq 2$. Using Theorem 3 (II.2, p. 25), the fact that $f(x,y,z)$ is the product of $(z - \alpha x - \beta y)^r$ and $[g(x,y,z)]^s$, and Theorem 6 (V.3, p. 61) we see that A is unitarily similar to a block diagonal matrix of the form

$$(\alpha + i\beta) I_r \oplus \underbrace{B \oplus \dots \oplus B}_{s \text{ times}},$$

where I_r is the $r \times r$ identity matrix, and $B = H' + iK'$ is a 2×2 matrix such that $\det(zI - xH' - yK') = g(x,y,z)$. ■

Example 16.

$$\text{Let } A = \begin{pmatrix} 0 & i & i \\ i & a+i & i \\ i & i & -a+i \end{pmatrix}. \text{ The matrix } A \text{ was}$$

used in Example 3 (II.2, pp. 26-28) and in Example 5 (II.4, p. 35).

It was shown there that A is not unitarily similar to a matrix which is $D(1, 2)$, but $f(x, y, z) = z [(z - ax - y)(z + ax - y) - 3y^2]$.

To find the intersection points of the line $z = 0$ and the conic $(z - ax - y)(z + ax - y) - 3y^2 = 0$, we set $z = 0$ in the equation of the conic. This yields

$$a^2 x^2 + 2y^2 = 0.$$

$$(ax + i\sqrt{2} y)(ax - i\sqrt{2} y) = 0.$$

Since we cannot have $x = y = z = 0$, the intersection points of the line and the conic are not real points.

Section 4. Line coordinates, dual curves and foci.

To discuss the connection between the polynomial $f(x, y, z) = \det(zI - xH - yK)$ and the numerical range of the matrix $A = H + iK$, we need the concepts of line coordinates and dual curves [5, 33, 37, 49].

The line $lx + my + ny = 0$ is determined by the triple $[l, m, n]$ of complex numbers. A line in homogeneous line coordinates is defined to be a triple of complex numbers $[l, m, n]$, not all zero. The triples $[l, m, n]$ and $[\rho l, \rho m, \rho n]$ denote the same line for any nonzero constant ρ . The point $Q = (x_0, y_0, z_0)$ is on the line $[l, m, n]$ if and only if $lx_0 + my_0 + nz_0 = 0$. Thus, a line

$[l, m, n]$ satisfies the equation of the point Q if and only if the line $[l, m, n]$ goes through Q . The homogeneous equation of degree one

$$lx_0 + my_0 + nz_0 = 0$$

in the line coordinates $l, m,$ and n is the equation of the point $Q = (x_0, y_0, z_0)$.

If $g(x,y,z)$ is a homogeneous polynomial of degree n , the equation $g(x,y,z) = 0$ may be viewed as an equation in line coordinates. The set of lines $[l, m, n]$ such that $g(l,m,n) = 0$ can be considered a set of lines in the plane which form an envelope of a curve. Thus, $g(x,y,z) = 0$ is the equation satisfied by all lines which are tangent to that curve, where tangent is now used in the classical sense.

Thus, the equation $g(x,y,z) = 0$ can describe two curves, C_1 and C_2 , where C_1 is the curve obtained by viewing $g(x,y,z) = 0$ as an equation in point coordinates, and C_2 arises by considering the equation to be in line coordinates. The curve C_2 is called the dual curve of C_1 . The degree, n , of the polynomial $g(x,y,z)$ is the number of points in which a fixed line intersects C_1 , and is also the number of tangents to C_2 from a fixed point. The number n is called the class of C_2 .

A fixed algebraic curve, C , may be described by two different equations. Thus $g_1(x,y,z) = 0$ may be the equation of C in point coordinates and $g_2(x,y,z) = 0$ may be the equation of C in line coordinates. The degree of g_1 is the degree of the curve C , and the degree of the polynomial g_2 is the class of C .

Finally, we will need the definition of a focus of an algebraic curve.

Definition. Let C be an algebraic curve and let F be a point not equal to $(1, i, 0)$ or $(1, -i, 0)$.¹ Let L_1 be the line through F and $(1, i, 0)$ and let L_2 be the line through F and $(1, -i, 0)$. If L_1 and L_2 are tangent to C , at points other than $(1, i, 0)$ and $(1, -i, 0)$, the point F is called a focus of C .

In general, a curve of class m has m^2 foci, counted according to proper multiplicities. A curve of class m with real coefficients has m real foci, counted according to proper multiplicities, and $m^2 - m$ foci which are not real.

Section 5. The numerical range of a matrix.

The algebraic curve whose equation in line coordinates is $f(x, y, z) = \det(zI - xH - yK) = 0$ determines the numerical range of the matrix $A = H + iK$. In this section we review some known facts about the numerical range. In Section 6 we will return to the equation $f(x, y, z) = 0$.

Definition. The set of all complex numbers w^*Aw , where w ranges over all $n \times 1$ column vectors of length one (i.e. $w^*w = 1$) is called the numerical range of A and is denoted $F(A)$.

$$F(A) = \{w^*Aw \mid w = (w_1, \dots, w_n)^T, w_i \in \mathbb{C}, \text{ and } \sum_{i=1}^n |w_i|^2 = 1\}.$$

¹The points $(1, i, 0)$ and $(1, -i, 0)$ are called the circular points at infinity.

The numerical range is also called the field of values.

With the usual representation of complex numbers $a + ib$ as points (a,b) in the real plane, one may regard $F(A)$ as a subset of the real plane. Toeplitz [47] and Hausdorff [16] showed $F(A)$ is a compact, convex set. If A and B are unitarily equivalent, then $F(A) = F(B)$. The eigenvalues of A are contained in $F(A)$. Since $F(A)$ is convex, this implies that $F(A)$ contains the closed, convex hull of the eigenvalues of A . If A is normal, then $F(A)$ is equal to the closed, convex hull of the eigenvalues of A . The converse holds for $n \leq 4$, but is not true for $n \geq 5$ [29]. If an eigenvalue, $\alpha + i\beta$ is on the boundary of $F(A)$, then A is unitarily similar to a matrix of the form

$$\left(\begin{array}{c|cccc} \alpha + i\beta & 0 & . & . & 0 \\ \hline 0 & & & & \\ . & & & & \\ . & & A_1 & & \\ . & & & & \\ 0 & & & & \end{array} \right)$$

where A_1 is of size $n-1 \times n-1$ [6, 19, 21].

Section 6. The relation of $f(x,y,z) = \det(zI - xH - yK)$ to the numerical range of $A = H + iK$.

Kippenhahn [21] showed that the numerical range of A is the closed, convex hull of the set of real points of the algebraic curve whose equation in line coordinates is $\det(zI + xH + yK) = 0$. Thus, viewed as an equation in line coordinates, $f(x,y,z) = \det(zI - xH - yK) = 0$ is the equation of a curve such that the closed, convex hull of its real

part is $F(-A) = -F(A)$.

For the rest of this section, equations in x , y , and z are regarded as equations in line coordinates, unless otherwise specified.

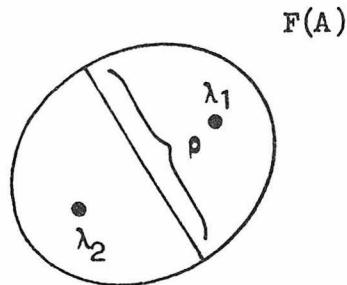
The algebraic curve $\det(zI + xH + yK) = 0$ has class n and has real coefficients. Hence it has n real foci. Kippenhahn [21] and Murnaghan [30] independently showed that these n real foci correspond to the eigenvalues of the matrix $A = H + iK$ in the following manner. If the eigenvalues of A are $a_1 + ib_1, \dots, a_n + ib_n$, where a_1, \dots, a_n and b_1, \dots, b_n are real, then the real foci of $\det(zI + xH + yK) = 0$ are the points $(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)$.

The dual curve of a conic is again a conic, so a curve has degree two if and only if it has class two. Hence, if A is 2×2 , the equation $\det(zI + xH + yK) = 0$ defines a conic. This equation has real coefficients by Proposition 1 (II.2, p. 24), so the real part of the curve is a hyperbola, parabola, or an ellipse (if $\det(zI + xH + yK)$ factors into two linear factors, we have a degenerate ellipse consisting of a pair of points). Since $F(A)$ is bounded, the real part of $\det(zI + xH + yK)$ must be an ellipse and $F(A)$ consists of the ellipse and its interior. The foci of the ellipse are the eigenvalues of A . Let U be a unitary matrix such that U^*AU is upper triangular. Since $F(A) = F(U^*AU)$, we may assume

$A = \begin{pmatrix} \lambda_1 & \rho \\ 0 & \lambda_2 \end{pmatrix}$, where λ_1 and λ_2 are the eigenvalues of A . We can

also assume ρ is a nonnegative real number since this can be achieved with a unitary similarity. Let $\lambda_1 = r_1 + is_1$ and $\lambda_2 = r_2 + is_2$, where r_1, s_1, r_2 , and s_2 are real numbers. Murnaghan [30] showed $F(A)$ is bounded by the ellipse with foci (r_1, s_1) and (r_2, s_2) and with minor axis of length ρ .

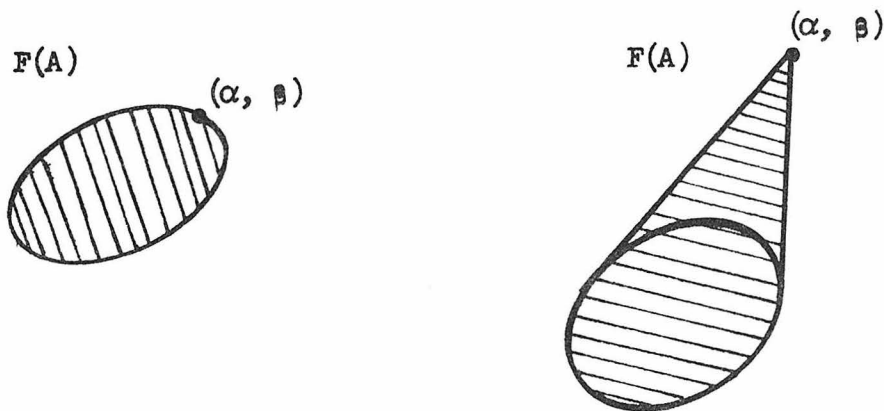
Since $F(A)$ uniquely determines the numbers λ_1, λ_2 and ρ , the polynomial $\det(zI + xH + yK)$ uniquely determines the matrix A up to unitary similarity. This gives an alternate proof of Theorem 3 (II.2, p. 25).



$$\lambda_1 = r_1 + is_1 \rightarrow (r_1, s_1)$$

$$\lambda_2 = r_2 + is_2 \rightarrow (r_2, s_2)$$

Suppose $\det(zI + xH + yK) = (z - \alpha x - \beta y)^r [g(x, y, z)]^s$ where $g(x, y, z)$ is irreducible of degree two. (This is the situation considered in Theorem 7, (VI.3, p. 69).) The curve then consists of the point (α, β) and the ellipse $g(x, y, z) = 0$. If the point (α, β) is on the ellipse, or is exterior to the ellipse, then it will be on the boundary of $F(A)$.



Hence, A is unitarily similar to a matrix of the form

$$\left(\begin{array}{c|cccc} \alpha + i\beta & 0 & . & . & 0 \\ \hline 0 & & & & \\ \cdot & & & & \\ \cdot & & & A_1 & \\ \cdot & & & & \\ 0 & & & & \end{array} \right)$$

where A_1 is $n-1 \times n-1$. Applying the same argument to A_1 and repeating the process r times, we see that A must be unitarily similar to a matrix of the form $(\alpha + i\beta) I_r \oplus A_{2s}$ where I_r is the $r \times r$ identity matrix and A_{2s} is $2s \times 2s$. Using the fact that $\det(zI + xH + yK) = (z - \alpha x - \beta y)^r [g(x,y,z)]^s$ and Theorem 6 (V.3, p. 61) we see that A_{2s} is unitarily similar to a block diagonal matrix which is a direct sum of s copies of a 2×2 matrix $B = H' + iK'$, where $\det(zI + xH' + yK') = g(x,y,z)$.

Note that saying the point (α, β) lies outside the ellipse is equivalent to saying there are two distinct, real tangent lines from (α, β) to the ellipse. If we now consider $(z - \alpha x - \beta y)^r [g(x,y,z)]^s = 0$ as the equation of an algebraic curve in point coordinates, this is equivalent to saying the line $z - \alpha x - \beta y = 0$ intersects the conic $g(x,y,z) = 0$ in two distinct real points. This yields another proof of Theorem 7 (VI.3, p. 69).

Finally, note that if the point (α, β) lies in the interior of the ellipse, the curve alone does not determine whether H and K have a common eigenvector. Example 3 (II.2, pp. 26-28) illustrates this.

VII. A Conjecture of Kippenhahn.

Section 1. Statement of the conjecture.

We now consider the conjecture of Kippenhahn [21] mentioned at the beginning of Chapter V. Recall that if $f(x,y,z) = \det(zI - xH - yK) = [\pi_1(x,y,z)]^{r_1} \dots [\pi_j(x,y,z)]^{r_j}$ where π_1, \dots, π_j are the distinct irreducible factors of $f(x,y,z)$ and r_1, \dots, r_j are positive integers, then $m(x,y,z) = \pi_1 \dots \pi_j$ is the minimal polynomial of $xH + yK$ (Proposition 2, V.2, p. 59). Hence, the degree of $m(x,y,z)$ is less than the degree of $f(x,y,z)$ if and only if $f(x,y,z)$ has a repeated factor. Kippenhahn made the following conjecture.

Conjecture (Kippenhahn) [21, p. 212]. If the degree of $m(x,y,z)$ is less than the degree of $f(x,y,z)$, then H and K have a common, non-trivial, invariant subspace and thus $A = H + iK$ is unitarily similar to a block diagonal matrix.

Kippenhahn proved that this conjecture holds if $m(x,y,z)$ has degree one or two (Theorem 6, V.3, p. 61). In this chapter we will show the conjecture holds if $n \leq 5$, where n is the degree of $f(x,y,z)$ and the size of the matrix A . We also show that Theorem 4 (III.3, p. 44) would be a consequence of this conjecture.

We first use Theorem 4 and Theorem 6 to show the conjecture holds for $n \leq 4$. If $n = 2$ and $m(x,y,z)$ has degree one, then

$m(x,y,z) = z - \alpha x - \beta y$ where α and β are real numbers. Hence,

$$H = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}. \quad \text{Thus the matrix}$$

$$A = H + iK = \begin{pmatrix} \alpha + i\beta & 0 \\ 0 & \alpha + i\beta \end{pmatrix} \quad \text{is scalar.}$$

If $n = 3$ and $m(x,y,z)$ has degree less than three, the characteristic polynomial $f(x,y,z)$ must have a repeated linear factor. By Theorem 4, H and K have a common eigenvector and A is unitarily similar to a matrix which is $D(1,2)$.

If $n = 4$ and $m(x,y,z)$ has degree less than four, either $f(x,y,z)$ has a repeated linear factor or is the square of an irreducible quadratic. In the former case, Theorem 4 applies and H and K have a common eigenvector. The latter case is covered by Theorem 6 which tells us A is unitarily similar to a matrix which is $D(2, 2)$.

We now examine the situation when $n = 5$. If $m(x,y,z)$ has degree less than five, the polynomial $f(x,y,z)$ can have a repeated linear factor or a repeated quadratic factor. If $f(x,y,z)$ has a repeated linear factor, then Theorem 4 applies and H and K have a common eigenvector. Otherwise, we must have

$$f(x,y,z) = (z - \alpha x - \beta y) [g(x,y,z)]^2$$

where $g(x,y,z)$ is irreducible of degree two. The next three sections are devoted to proving the following result.

Theorem 8. Let $A = H + iK$ be a 5×5 complex matrix. Suppose $f(x, y, z) = \det(zI - xH - yK) = (z - \alpha x - \beta y) [g(x, y, z)]^2$ where $g(x, y, z)$ is an irreducible polynomial of degree two. Then A is unitarily similar to a block diagonal matrix which is $D(3, 2)$ or $D(2, 2, 1)$.

Remark. If the point (α, β) lies outside, or on the conic whose equation in line coordinates is $g(x, y, z) = 0$, then Theorem 8 is a special case of the situation dealt with in Section VI.6 (pp. 77-78). In this case, A is unitarily similar to a matrix which is $D(2, 2, 1)$.

Section 2. Reduction to the case $f(x, y, z) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2$.

We first show that without loss of generality, we may assume $f(x, y, z) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2$. We use the methods of Motzkin and Taussky [28] discussed in Section VI.2.

Assume $f(x, y, z) = (z - \alpha x - \beta y) [g(x, y, z)]^2$. We will show that we may assume $g(x, y, z) = z^2 - x^2 - y^2$. Since $g(x, y, z) = 0$ is an irreducible conic, there are two distinct lines, L_1 and L_2 , through the point $P = (0, 0, 1)$ which are tangent to $g(x, y, z) = 0$. Let $\delta_1 x - \eta_1 y = 0$ be the equation of L_1 and let $\delta_2 x - \eta_2 y = 0$ be the equation of L_2 , where $\delta_1, \eta_1, \delta_2$ and η_2 are complex numbers. Then each of the matrices $B_1 = \eta_1 H + \delta_1 K$ and $B_2 = \eta_2 H + \delta_2 K$ has an eigenvalue of multiplicity four. Since L_1 and L_2 are distinct

lines through P , we have $\det \begin{pmatrix} \eta_1 & \delta_1 \\ \delta_2 & \eta_2 \end{pmatrix} \neq 0$. Hence B_1 and B_2

also generate the pencil $xH + yK$.

First consider the following special case. Suppose there exist real numbers r and s such that $\delta_1 = r\eta_1$ and $\delta_2 = s\eta_2$. Then $B_1 = \eta_1(H + rK)$ and $B_2 = \eta_2(H + sK)$. Since $H + rK$ and $H + sK$ are Hermitian, B_1 and B_2 can be diagonalized. Then since B_1 and B_2 each have at most two distinct eigenvalues, B_1 and B_2 satisfy polynomials of degree two. Hence, by Corollary 3 (IV, p. 56) the algebra generated by B_1 and B_2 , denoted $\mathcal{A}(B_1, B_2)$, is similar to an algebra which is $BT^*(2, 2, 1)$. Since H and K are elements of $\mathcal{A}(B_1, B_2)$, the matrices H and K must be simultaneously similar to matrices which are $BT(2, 2, 1)$. By Specht's theorem B (I.5, p. 17) the Hermitian matrices H and K are then simultaneously, unitarily similar to matrices which are $D(2, 2, 1)$ and hence A is unitarily similar to a matrix which is $D(2, 2, 1)$.

We now assume that $\delta_1 = r\eta_1$ and $\delta_2 = s\eta_2$ do not both hold. We will assume $\delta_1 \neq r\eta_1$ for any real number r . Write $\eta_1 = a + ib$ and $\delta_1 = c + id$ where $a, b, c,$ and d are real numbers. Now if

$\eta_1 = 0$, then from $\det \begin{pmatrix} \eta_1 & \delta_1 \\ \eta_2 & \delta_2 \end{pmatrix} \neq 0$ we know $\eta_2 \neq 0$. If

$\delta_2 = s\eta_2$ holds and $\eta_1 = 0$, then $B_1 = \delta_1 K$ and $B_2 = \eta_2(H + sK)$

are both diagonalizable and the argument used for the special case applies.

So if $\eta_1 = 0$, we may assume $\delta_2 \neq s\eta_2$ for any real number s and $\eta_2 \neq 0$. For convenience of notation, we will assume $\delta_1 \neq r\eta_1$ for any real number r and $\eta_1 \neq 0$. Then (a, b) and (c, d) are

linearly independent and $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$.

Now $B_1 = (a + ib)H + (c + id)K = (aH + cK) + i(bH + dK)$.

Since $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$, the Hermitian matrices $H' = aH + cK$

and $K' = bH + dK$ also generate the pencil $xH + yK$. Also,

$$\begin{aligned} \det(zI - xH' - yK') &= \det(zI - x(aH + cK) - y(bH + dK)) \\ &= \det(zI - (ax + by)H - (cx + dy)K). \end{aligned}$$

Hence, the polynomial $\det(zI - xH' - yK')$ is obtained from $\det(zI - xH - yK)$ by making the linear change of variable

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \\ z \end{pmatrix}$$

Therefore, $\det(zI - xH' - yK')$ also factors into a linear factor and a repeated quadratic. Thus, without loss of generality, we may replace H by H' , K by K' , and $A = H + iK$ by $B_1 = H' + iK'$. For convenience of notation, we will still use the letters H , K and A . Thus, we may assume that A has an eigenvalue λ of multiplicity four.

Now write $\lambda = r + is$ where r and s are real numbers.

$$A - \lambda I = (H - rI) + i(K - sI).$$

Note that $\det(zI - x(H - rI) - y(K - sI)) = \det((z + rx + sy)I - xH - yK)$ is obtained from $\det(zI - xH - yK)$ by making the linear change of variable

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ r & s & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

So we may replace H with $H - rI$, K with $K - sI$, and A with $A - \lambda I$.

Thus, we may assume that zero is an eigenvalue of A of multiplicity four. Applying a unitary similarity to upper triangularize A , we may assume

$$A = \begin{pmatrix} \alpha + i\beta & 2a_{12} & 2a_{13} & 2a_{14} & 2a_{15} \\ 0 & 0 & 2a_{23} & 2a_{24} & 2a_{25} \\ 0 & 0 & 0 & 2a_{34} & 2a_{35} \\ 0 & 0 & 0 & 0 & 2a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

$$xH + yK = \begin{pmatrix} \alpha x + \beta y & a_{12}(x-iy) & a_{13}(x-iy) & a_{14}(x-iy) & a_{15}(x-iy) \\ \bar{a}_{12}(x+iy) & 0 & a_{23}(x-iy) & a_{24}(x-iy) & a_{25}(x-iy) \\ \bar{a}_{13}(x+iy) & \bar{a}_{23}(x+iy) & 0 & a_{34}(x-iy) & a_{35}(x-iy) \\ \bar{a}_{14}(x+iy) & \bar{a}_{24}(x+iy) & \bar{a}_{34}(x+iy) & 0 & a_{45}(x-iy) \\ \bar{a}_{15}(x+iy) & \bar{a}_{25}(x+iy) & \bar{a}_{35}(x+iy) & \bar{a}_{45}(x+iy) & 0 \end{pmatrix} .$$

Let $g(x,y,z) = z^2 - (rx + sy)z - (tx^2 + uxy + vy^2)$, where $r, s, t, u,$ and v are constants. The coefficient of z^4 in the expansion of $f(x,y,z) = (z - \alpha x - \beta y)[g(x,y,z)]^2$ is then $-\alpha x - \beta y - 2(rx + sy)$. However, the coefficient of z^4 in the expansion of $\det(zI - xH - yK)$ must be $(-1)(\text{trace}(xH + yK))$ which is $-\alpha x - \beta y$. Therefore, $rx + sy = 0$, which yields $r = s = 0$. Thus, $g(x,y,z) = z^2 - (tx^2 + uxy + vy^2)$.

The coefficient of z^0 in $f(x,y,z)$ is then $-(\alpha x + \beta y)(tx^2 + uxy + vy^2)^2$. Thus we have

$$(\alpha x + \beta y)(tx^2 + uxy + vy^2)^2 = \det(xH + yK).$$

Inspection of the matrix $xH + yK$ shows that every term in the expansion of $\det(xH + yK)$ contains the product $(x - iy)(x + iy) = x^2 + y^2$. Therefore, $x^2 + y^2$ must divide $(\alpha x + \beta y)(tx^2 + uxy + vy^2)$. Since α and β are real, $tx^2 + uxy + vy^2$ must be divisible by $x^2 + y^2$. But then $u = 0$ and $t = v$, so $g(x,y,z) = z^2 - t(x^2 + y^2)$ and $f(x,y,z) = (z - \alpha x - \beta y)(z^2 - t(x^2 + y^2))^2$.

The eigenvalues of H are then $\alpha, \sqrt{t}, \sqrt{t}, -\sqrt{t}, -\sqrt{t}$ and the eigenvalues of K are $\beta, \sqrt{t}, \sqrt{t}, -\sqrt{t}, -\sqrt{t}$. Since H and K are Hermitian, they have real eigenvalues, so t is a nonnegative real number. We may replace H and K with $\frac{H}{\sqrt{t}}$ and $\frac{K}{\sqrt{t}}$.

(Note $t \neq 0$ as the conic is irreducible.) Hence, we may assume

$$t = 1 \text{ and } f(x,y,z) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2.$$

Section 3. Two lemmas.

In this section we prove two lemmas which will be used to transform the matrix A into a form which is easier to deal with. In particular, the lemmas will be used to show A is unitarily similar to a matrix with many zero entries. Lemma 2 is similar to a lemma concerning 3×3 matrices which appears in a paper by Percy [32, p. 426].

Lemma 2.

$$\text{Let } B = \begin{pmatrix} 0 & b_{12} & b_{13} & b_{14} \\ 0 & 0 & b_{23} & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then if $b_{12}b_{23}b_{34} = 0$, B is unitarily similar to a matrix of the

$$\text{form } \begin{pmatrix} 0 & 0 & c_{13} & c_{14} \\ 0 & 0 & 0 & c_{24} \\ 0 & 0 & 0 & c_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Proof. Since $b_{12}b_{23}b_{34} = 0$, we have $b_{12} = 0$, $b_{23} = 0$, or $b_{34} = 0$.

Suppose first that $b_{12} = 0$. There exist real numbers φ and ψ such

that $0 \leq \varphi \leq 2\pi$, $0 \leq \psi \leq 2\pi$ and $e^{i\varphi} b_{13} = |b_{13}|$ and $e^{i\psi} b_{23} = |b_{23}|$.

Now if $b_{23} \neq 0$, choose $0 \leq \theta \leq 2\pi$ such that

$$|b_{23}| \sin \theta + |b_{13}| \cos \theta = 0.$$

$$\text{Let } U = \begin{pmatrix} e^{i\varphi} \sin \theta & -e^{i\psi} \cos \theta & 0 & 0 \\ e^{i\varphi} \cos \theta & e^{i\psi} \sin \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad \text{The matrix } U \text{ is unitary.}$$

$$\text{Then } U^* = U^{-1} = \begin{pmatrix} e^{-i\varphi} \sin \theta & e^{-i\varphi} \cos \theta & 0 & 0 \\ -e^{-i\psi} \cos \theta & e^{-i\psi} \sin \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$$UBU^* = \begin{pmatrix} 0 & 0 & b_{13}^* & b_{14}^* \\ 0 & 0 & 0 & b_{24}^* \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad U^* = \begin{pmatrix} 0 & 0 & b_{13}^* & b_{14}^* \\ 0 & 0 & 0 & b_{24}^* \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which has the desired form.

If $b_{34} = 0$, and $b_{23} \neq 0$, choose real numbers φ , ψ , and θ so that $e^{i\psi} b_{23} \sin \theta + e^{i\varphi} b_{24} \cos \theta = 0$.

$$\text{Let } U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{-i\psi} \sin \theta & e^{-i\varphi} \cos \theta \\ 0 & 0 & -e^{-i\psi} \cos \theta & e^{-i\varphi} \sin \theta \end{pmatrix}. \quad \text{The matrix } U \text{ is unitary.}$$

$$U^* = U^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\psi} \sin \theta & -e^{i\psi} \cos \theta \\ 0 & 0 & e^{i\varphi} \cos \theta & e^{i\varphi} \sin \theta \end{pmatrix}.$$

$$UBU^* = U \begin{pmatrix} 0 & b_{12} & b_{13}^* & b_{14}^* \\ 0 & 0 & 0 & b_{24}^* \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & b_{12} & b_{13}^* & b_{14}^* \\ 0 & 0 & 0 & b_{24}^* \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So it remains to show that if $b_{23} = 0$, we can unitarily transform B to the desired form. If $b_{23} = 0$ and $b_{12} \neq 0$, then choose real numbers φ , ψ , and θ , all between 0 and 2π so that $b_{12} e^{i\psi} \sin \theta + b_{13} e^{i\varphi} \cos \theta = 0$.

$$\text{Let } U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\psi} \sin \theta & e^{-i\varphi} \cos \theta & 0 \\ 0 & -e^{-i\psi} \cos \theta & e^{-i\varphi} \sin \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad \text{The matrix } U \text{ is unitary.}$$

$$\text{Then } U^* = U^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\psi} \sin \theta & -e^{i\psi} \cos \theta & 0 \\ 0 & e^{i\varphi} \cos \theta & e^{i\varphi} \sin \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$$UBU^* = U \begin{pmatrix} 0 & 0 & b_{13}^* & b_{14} \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & b_{13}^* & b_{14} \\ 0 & 0 & 0 & b_{24}^* \\ 0 & 0 & 0 & b_{34}^* \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which has the desired form. This concludes the proof of Lemma 2. ■

Lemma 2 is now used to prove Lemma 3.

Lemma 3. Let B be a 5×5 complex matrix of the form

$$\begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ 0 & 0 & b_{23} & b_{24} & b_{25} \\ 0 & 0 & 0 & b_{34} & b_{35} \\ 0 & 0 & 0 & 0 & b_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and suppose B has rank at most three. Then B is unitarily similar to a matrix of the form

$$\begin{pmatrix} b_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ 0 & 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & 0 & 0 & c_{35} \\ 0 & 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Proof. We consider two cases: $b_{11} = 0$ and $b_{11} \neq 0$.

Case 1. $b_{11} = 0$. Since the rank of B is at most three,

one of the entries on the superdiagonal must be zero. If $b_{12}b_{23}b_{34} = 0$, then by Lemma 2 there exists a 4×4 unitary matrix U such that

$$U \begin{pmatrix} 0 & b_{12} & b_{13} & b_{14} \\ 0 & 0 & b_{23} & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} U^* \text{ has the form } \begin{pmatrix} 0 & 0 & c_{13} & c_{14} \\ 0 & 0 & 0 & c_{24} \\ 0 & 0 & 0 & c_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let V be the 5×5 unitary matrix

$$\begin{pmatrix} & & & & & 0 \\ & & & & & 0 \\ & & & & & 0 \\ & & & & & 0 \\ & & & & & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then VBV^* has the form

$$\begin{pmatrix} 0 & 0 & c_{13} & c_{14} & c_{15} \\ 0 & 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & 0 & c_{34} & c_{35} \\ 0 & 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We now apply Lemma 2 to the 4×4 matrix

$$\begin{pmatrix} 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & c_{34} & c_{35} \\ 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus, we let U be a 4×4 unitary matrix such that

$$U \begin{pmatrix} 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & c_{34} & c_{35} \\ 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 \end{pmatrix} U^* = \begin{pmatrix} 0 & 0 & d_{24} & d_{25} \\ 0 & 0 & 0 & d_{35} \\ 0 & 0 & 0 & d_{45} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then if $V = \left(\begin{array}{c|cccc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & & & & \\ 0 & & & & \\ 0 & & & & \\ 0 & & & & \end{array} \right)$, the matrix

$$V \begin{pmatrix} 0 & 0 & c_{13} & c_{14} & c_{15} \\ 0 & 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & 0 & c_{34} & c_{35} \\ 0 & 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} V^* \quad \text{has the form}$$

$$\begin{pmatrix} 0 & d_{12} & d_{13} & d_{14} & d_{15} \\ 0 & 0 & 0 & d_{24} & d_{25} \\ 0 & 0 & 0 & 0 & d_{35} \\ 0 & 0 & 0 & 0 & d_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{which is the desired form.}$$

Case 2. $b_{11} \neq 0$. Since the rank of B is at most three, not all of b_{23} , b_{34} and b_{45} can be nonzero. Hence $b_{23}b_{34}b_{45} = 0$. Applying Lemma 2, we see that there is a 4×4 unitary matrix U such that

$$U \begin{pmatrix} 0 & b_{23} & b_{24} & b_{25} \\ 0 & 0 & b_{34} & b_{35} \\ 0 & 0 & 0 & b_{45} \\ 0 & 0 & 0 & 0 \end{pmatrix} U^* \text{ has the form } \begin{pmatrix} 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & 0 & c_{35} \\ 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Hence, letting V be the 5×5 unitary matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & & & & \\ 0 & & & & \\ 0 & & & & \\ 0 & & & & \end{pmatrix} \begin{matrix} \\ \\ \\ \\ U \end{matrix}$$

we have VBV^* of the form

$$\begin{pmatrix} b_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ 0 & 0 & 0 & c_{24} & c_{25} \\ 0 & 0 & 0 & 0 & c_{35} \\ 0 & 0 & 0 & 0 & c_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This concludes the proof of Lemma 3. ■

Section 4. The proof that A can be unitarily block diagonalized.

In Section 3 of this chapter, it was shown that to establish Theorem 8, it suffices to prove the following statement.

Theorem 8'. Let $A = H + iK$ be a 5×5 complex matrix. Suppose

$$f(x, y, z) = \det(zI - xH - yK) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2.$$

Then A is unitarily similar to a block diagonal matrix which is $D(3, 2)$ or $D(2, 2, 1)$.

Proof. Since $A = H + iK$ and $f(x,y,z) = (z - \alpha x - \beta y)(z^2 - x^2 - y^2)^2$, we have $\det(zI - A) = (z - (\alpha + i\beta)) z^4$ and A has eigenvalues $\alpha + i\beta, 0, 0, 0, 0$. Applying a unitary transformation to upper triangularize A , we may assume

$$A = \begin{pmatrix} \alpha + i\beta & 2a_{12} & 2a_{13} & 2a_{14} & 2a_{15} \\ 0 & 0 & 2a_{23} & 2a_{24} & 2a_{25} \\ 0 & 0 & 0 & 2a_{34} & 2a_{35} \\ 0 & 0 & 0 & 0 & 2a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

The minimal polynomial of the pencil $xH + yK$ is

$(z - \alpha x - \beta y)(z^2 - x^2 - y^2)$ by Proposition 2 (V.2, p. 59). Hence A satisfies the equation $(z - (\alpha + i\beta)) z^2 = 0$. This implies that

A has at least two linearly independent eigenvectors with eigenvalue zero. Hence A has rank at most 3. By Lemma 3, we may assume

$$a_{23} = a_{34} = 0.$$

$$A = \begin{pmatrix} \alpha + i\beta & 2a_{12} & 2a_{13} & 2a_{14} & 2a_{15} \\ 0 & 0 & 0 & 2a_{24} & 2a_{25} \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 2a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

$$xH + yK = \begin{pmatrix} \alpha x + \beta y & a_{12}(x-iy) & a_{13}(x-iy) & a_{14}(x-iy) & a_{15}(x-iy) \\ \bar{a}_{12}(x+iy) & 0 & 0 & a_{24}(x-iy) & a_{25}(x-iy) \\ \bar{a}_{13}(x+iy) & 0 & 0 & 0 & a_{35}(x-iy) \\ \bar{a}_{14}(x+iy) & \bar{a}_{24}(x+iy) & 0 & 0 & a_{45}(x-iy) \\ \bar{a}_{15}(x+iy) & \bar{a}_{25}(x+iy) & \bar{a}_{35}(x+iy) & \bar{a}_{45}(x+iy) & 0 \end{pmatrix}.$$

Since the minimal polynomial of $xH + yK$ is $(z - \alpha x - \beta y)(z^2 - x^2 - y^2)$, the matrix $xH + yK + \left(\sqrt{x^2 + y^2}\right) I$ has rank at most three for all values of x and y . Hence, as functions of x and y , the determinants of all the 4×4 minors of $xH + yK + \left(\sqrt{x^2 + y^2}\right) I$ must be identically equal to zero. We use this fact to obtain a set of equations in the a_{ij} 's. The details of this computation are rather lengthy; we have placed them in an appendix at the end of this chapter. The equations resulting from this computation are listed on the next page.

- (1) $|a_{13}|^2 + |a_{14}|^2 + |a_{12}|^2 + |a_{24}|^2 - |a_{13}|^2 |a_{24}|^2 = 1$
- (2) $\alpha (1 - |a_{24}|^2) + a_{12} a_{24} \bar{a}_{14} + \bar{a}_{12} \bar{a}_{24} a_{14} = 0$
- (3) $\beta (1 - |a_{24}|^2) + i(\bar{a}_{12} \bar{a}_{24} a_{14} - a_{12} a_{24} \bar{a}_{14}) = 0$
- (4) $|a_{25}|^2 + |a_{35}|^2 + |a_{45}|^2 + |a_{24}|^2 - |a_{35}|^2 |a_{24}|^2 = 1$
- (5) $a_{24} a_{45} \bar{a}_{25} = 0$
- (6) $a_{13} a_{35} (1 - |a_{24}|^2) + a_{14} a_{45} + a_{12} a_{25} = 0$
- (7) $a_{14} a_{25} \bar{a}_{24} + a_{15} (1 - |a_{24}|^2) = 0$
- (8) $a_{12} a_{24} a_{45} = 0$
- (9) $a_{13} a_{24} a_{35} \bar{a}_{14} - |a_{13}|^2 a_{24} a_{45} + a_{24} a_{45} + \bar{a}_{12} a_{15} - a_{25} \alpha = 0$
- (10) $i(a_{13} a_{24} a_{35} \bar{a}_{14} - |a_{13}|^2 a_{24} a_{45} + a_{24} a_{45} - \bar{a}_{12} a_{15}) + a_{25} \beta = 0$
- (11) $|a_{13}|^2 a_{25} - a_{13} \bar{a}_{12} a_{35} + |a_{14}|^2 a_{25} - \bar{a}_{14} a_{24} a_{15} - \bar{a}_{12} a_{14} a_{45} - a_{25} + a_{24} a_{45} \alpha = 0$
- (12) $i(|a_{13}|^2 a_{25} - a_{13} \bar{a}_{12} a_{35} + |a_{14}|^2 a_{25} - \bar{a}_{14} a_{24} a_{15} - a_{12} a_{14} a_{45} - a_{25}) + a_{24} a_{45} \beta = 0$
- (13) $a_{24} a_{45} (\alpha + i\beta) = 0$
- (14) $|a_{12}|^2 + |a_{13}|^2 + |a_{14}|^2 + |a_{15}|^2 + |a_{24}|^2 + |a_{25}|^2 + |a_{35}|^2 + |a_{45}|^2 = 2$
- (15) $-|a_{15}|^2 + |a_{24}|^2 (1 - |a_{13}|^2 - |a_{35}|^2) = 0$
- (16) $\bar{a}_{12} \bar{a}_{24} + \bar{a}_{13} \bar{a}_{35} \bar{a}_{24} a_{25} - \bar{a}_{12} \bar{a}_{24} |a_{35}|^2 = 0$
- (17) $\bar{a}_{15} a_{45} = 0$
- (18) $\bar{a}_{14} |a_{25}|^2 - \bar{a}_{15} \bar{a}_{24} a_{25} - \bar{a}_{14} - \bar{a}_{25} a_{45} \bar{a}_{12} + \bar{a}_{14} |a_{35}|^2 - \bar{a}_{13} \bar{a}_{35} a_{45} = 0$

We complete the proof by considering the following five cases.

Case 1. $|a_{24}|^2 = 1.$

Case 2. $a_{24} = 0.$

Case 3. $|a_{24}|^2 \neq 1, a_{24} \neq 0,$ and $a_{25} = a_{45} = 0.$

Case 4. $|a_{24}|^2 \neq 1, a_{24} \neq 0,$ and $a_{45} \neq 0.$

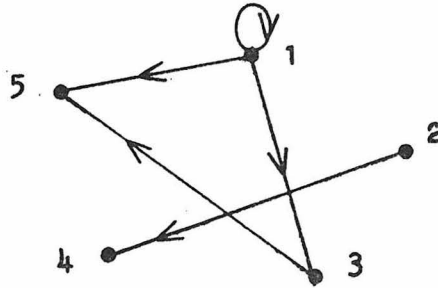
Case 5. $|a_{24}|^2 \neq 1, a_{24} \neq 0, a_{45} = 0,$ and $a_{25} \neq 0.$

Case 1. $|a_{24}|^2 = 1.$

Then (1) implies $a_{12} = a_{14} = 0$ and (4) implies $a_{25} = a_{45} = 0.$

$$\text{Thus } A = \begin{pmatrix} \alpha + i\beta & 0 & 2a_{13} & 0 & 2a_{15} \\ 0 & 0 & 0 & 2a_{24} & 0 \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The graph of A is



The graph of A shows that there is a permutation matrix P such that EAP^T is $D(3, 2).$

Case 2. $a_{24} = 0$.

Then (2) and (3) imply that $\alpha = \beta = 0$. From (7) we have $a_{15} = 0$ and (6) implies $a_{13}a_{35} + a_{14}a_{45} + a_{12}a_{25} = 0$. Thus, we have

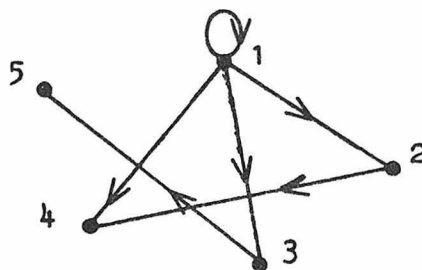
$$A = \begin{pmatrix} 0 & 2a_{12} & 2a_{13} & 2a_{14} & 0 \\ 0 & 0 & 0 & 0 & 2a_{25} \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 2a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $A^2 = 0$, by Corollary 4 (IV, p. 57) A is unitarily similar to a matrix which is $D(2, 2, 1)$.

Case 3. $a_{24} \neq 0$, $|a_{24}| \neq 1$, and $a_{25} = a_{45} = 0$.

From (7) we have $a_{15} = 0$. Equation (6) implies $a_{13}a_{35} = 0$.

$$A = \begin{pmatrix} \alpha + i\beta & 2a_{12} & 2a_{13} & 2a_{14} & 0 \\ 0 & 0 & 0 & 2a_{24} & 0 \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$



The graph of A is

Since $a_{13}a_{35} = 0$, either $a_{13} = 0$ or $a_{35} = 0$. Removing the edge $[1, 3]$ disconnects the graph into a component with three vertices and a component with two vertices. In this case, there is a permutation matrix P such that PAP^T is $D(3, 2)$. Removing the edge $[3, 5]$ disconnects the graph into a component with four vertices and a component with one vertex. Hence there is a permutation matrix P such that PAP^T is $D(4, 1)$. Since the 4×4 block corresponds to the polynomial $(z^2 - x^2 - y^2)^2$, A is unitarily similar to a matrix which is $D(2, 2, 1)$.

Case 4. $|a_{24}|^2 \neq 1$, $a_{24} \neq 0$, and $a_{45} \neq 0$.

Equation (13) implies $\alpha = \beta = 0$. From (5) we have $a_{25} = 0$, while (7) implies $a_{15} = 0$ and (8) yields $a_{12} = 0$. Thus,

$$A = \begin{pmatrix} 0 & 0 & 2a_{13} & 2a_{14} & 0 \\ 0 & 0 & 0 & 2a_{24} & 0 \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 2a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $a_{24} \neq 0$, equation (15) implies $|a_{13}|^2 + |a_{35}|^2 = 1$. Since $a_{12} = a_{15} = a_{25} = 0$, equation (14) then tells us

$$|a_{14}|^2 + |a_{24}|^2 + |a_{45}|^2 = 1.$$

$$\text{Let } \eta_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \text{ Let } \eta_2 = \frac{1}{2} (A \eta_1) = \begin{pmatrix} 0 \\ 0 \\ a_{35} \\ a_{45} \\ 0 \end{pmatrix} \text{ and let}$$

$$\eta_3 = \frac{1}{2} (A \eta_2) = \begin{pmatrix} a_{13}a_{35} + a_{14}a_{45} \\ a_{24}a_{45} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \text{ Let } W \text{ be the subspace}$$

spanned by the three vectors η_1 , η_2 , and η_3 . Note that $A(W) \subset W$.

We now show that $A^*(W) \subset W$.

$$A^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2\bar{a}_{13} & 0 & 0 & 0 & 0 \\ 2\bar{a}_{14} & 2\bar{a}_{24} & 0 & 0 & 0 \\ 0 & 0 & 2\bar{a}_{35} & 2\bar{a}_{45} & 0 \end{pmatrix}$$

$$A^*(\eta_1) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \frac{A^*(\eta_2)}{2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ |a_{35}|^2 + |a_{45}|^2 \end{pmatrix} \text{ and}$$

$$\frac{A^*(\eta_3)}{2} = \begin{pmatrix} 0 \\ 0 \\ |a_{13}|^2 a_{35} + \bar{a}_{13} a_{14} a_{45} \\ \bar{a}_{14} a_{13} a_{35} + |a_{14}|^2 a_{45} + |a_{24}|^2 a_{45} \\ 0 \end{pmatrix}.$$

It is clear that $A^*(\eta_1) \in W$ and $A^*(\eta_2) \in W$. We now show that $A^*(\eta_3) \in W$. To do this, we must show $A^*(\eta_3)$ is a scalar multiple of η_2 . To do this, it suffices to show

$$a_{45}[|a_{13}|^2 a_{35} + \bar{a}_{13} a_{14} a_{45}] = a_{35}[\bar{a}_{14} a_{13} a_{35} + |a_{14}|^2 a_{45} + |a_{24}|^2 a_{45}].$$

First consider the left hand side. From equation (9) we have

$$a_{13} a_{24} a_{35} \bar{a}_{14} - |a_{13}|^2 a_{24} a_{45} + a_{24} a_{45} = 0.$$

Since $a_{24} \neq 0$, we may divide this by a_{24} to obtain

$$a_{45}(1 - |a_{13}|^2) + a_{13} a_{35} \bar{a}_{14} = 0.$$

Using $a_{25} = 0$ in equation (6) we obtain $a_{14} a_{45} = a_{13} a_{35} (|a_{24}|^2 - 1)$.

$$\begin{aligned} \text{Thus } a_{45}|a_{13}|^2 a_{35} + \bar{a}_{13} a_{14} a_{45} a_{45} &= a_{45}|a_{13}|^2 a_{35} + \bar{a}_{13} (a_{13} a_{35} (|a_{24}|^2 - 1)) a_{45} \\ &= |a_{13}|^2 a_{35} a_{45} (1 + |a_{24}|^2 - 1) = |a_{13}|^2 |a_{24}|^2 a_{35} a_{45}. \end{aligned}$$

Now we work on the right hand side.

$$\begin{aligned} \bar{a}_{14} a_{13} a_{35}^2 + |a_{14}|^2 a_{35} a_{45} + |a_{24}|^2 a_{35} a_{45} &= a_{35} a_{45} (|a_{13}|^2 - 1) + \\ + a_{35} a_{45} (|a_{14}|^2 + |a_{24}|^2) &= a_{35} a_{45} (|a_{13}|^2 - 1 + |a_{14}|^2 + |a_{24}|^2). \end{aligned}$$

Now from equation (1) we have $|a_{13}|^2 + |a_{14}|^2 + |a_{24}|^2 - 1 = |a_{13}|^2 |a_{24}|^2$,

since $a_{12} = 0$. Hence, the right hand side is also equal to

$|a_{13}|^2 |a_{24}|^2 a_{35} a_{45}$. Thus $A^*(\eta_3)$ is a scalar multiple of η_2 .

Thus, we have shown the subspace W spanned by the vectors

η_1, η_2 , and η_3 is invariant under both A and A^* . This implies

that W^\perp is also A invariant and A is unitarily similar to a

block diagonal matrix.

Case 5. $|a_{24}|^2 \neq 1$, $a_{24} \neq 0$, $a_{45} = 0$, but $a_{25} \neq 0$.

$$A = \begin{pmatrix} \alpha + i\beta & 2a_{12} & 2a_{13} & 2a_{14} & 2a_{15} \\ 0 & 0 & 0 & 2a_{24} & 2a_{25} \\ 0 & 0 & 0 & 0 & 2a_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

$$A^* = \begin{pmatrix} \alpha - i\beta & 0 & 0 & 0 & 0 \\ 2\bar{a}_{12} & 0 & 0 & 0 & 0 \\ 2\bar{a}_{13} & 0 & 0 & 0 & 0 \\ 2\bar{a}_{14} & 2\bar{a}_{24} & 0 & 0 & 0 \\ 2\bar{a}_{15} & 2\bar{a}_{25} & 2\bar{a}_{35} & 0 & 0 \end{pmatrix}.$$

$$\text{Let } \eta_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \bar{a}_{14} \\ \bar{a}_{15} \end{pmatrix} \quad \text{and let } \eta_2 = \frac{1}{2} (A \eta_1) = \begin{pmatrix} |a_{14}|^2 + |a_{15}|^2 \\ \bar{a}_{14} a_{24} + \bar{a}_{15} a_{25} \\ \bar{a}_{15} a_{35} \\ 0 \\ 0 \end{pmatrix}.$$

$$\text{Let } \eta_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and let } W \text{ be the subspace spanned by the vectors}$$

η_1 , η_2 , and η_3 . Then W is invariant under A . We now compute $A^*(\eta_1)$, $A^*(\eta_2)$, and $A^*(\eta_3)$ and show that W is also invariant under the matrix A^* . First, we have that $A^*(\eta_1)$ is the zero vector.

$$\frac{1}{2} A^*(\eta_2) = \begin{pmatrix} \frac{1}{2} (\alpha - i\beta) (|a_{14}|^2 + |a_{15}|^2) \\ \bar{a}_{12} (|a_{14}|^2 + |a_{15}|^2) \\ \bar{a}_{13} (|a_{14}|^2 + |a_{15}|^2) \\ \bar{a}_{14} (|a_{14}|^2 + |a_{15}|^2) + \bar{a}_{14} |a_{24}|^2 + \bar{a}_{15} a_{25} \bar{a}_{24} \\ \bar{a}_{15} (|a_{14}|^2 + |a_{15}|^2) + \bar{a}_{14} a_{24} a_{25} + \bar{a}_{15} |a_{25}|^2 + \bar{a}_{15} |a_{35}|^2 \end{pmatrix}.$$

$$\frac{1}{2} A^*(\eta_3) = \begin{pmatrix} \frac{1}{2} (\alpha - i\beta) \\ \bar{a}_{12} \\ \bar{a}_{13} \\ \bar{a}_{14} \\ \bar{a}_{15} \end{pmatrix}.$$

Since $a_{45} = 0$, equation (18) implies

$$\bar{a}_{14}(|a_{25}|^2 + |a_{35}|^2 - 1) = \bar{a}_{15}\bar{a}_{24}a_{25}.$$

$$\begin{aligned} \text{Thus } \bar{a}_{14}(|a_{14}|^2 + |a_{15}|^2) + \bar{a}_{14}|a_{24}|^2 + \bar{a}_{15}\bar{a}_{24}a_{25} &= \\ \bar{a}_{14}(|a_{14}|^2 + |a_{15}|^2 + |a_{24}|^2 + |a_{25}|^2 + |a_{35}|^2 - 1). \end{aligned}$$

$$\text{From equation (7) } \bar{a}_{14}\bar{a}_{25}a_{24} = \bar{a}_{15}(|a_{24}|^2 - 1).$$

$$\begin{aligned} \text{Thus } \bar{a}_{15}(|a_{14}|^2 + |a_{15}|^2) + \bar{a}_{14}a_{24}\bar{a}_{25} + \bar{a}_{15}|a_{25}|^2 + \bar{a}_{15}|a_{35}|^2 \\ = \bar{a}_{15}(|a_{14}|^2 + |a_{15}|^2 + |a_{24}|^2 + |a_{25}|^2 + |a_{35}|^2 - 1), \end{aligned}$$

and the 2×1 column vector formed from the last two entries of the vector $\frac{1}{2}A^*(\eta_2)$ is a scalar multiple of the 2×1 column vector

$$\begin{pmatrix} \bar{a}_{14} \\ \bar{a}_{15} \end{pmatrix}. \text{ Since } \eta_3 \text{ is the vector } (1 \ 0 \ 0 \ 0 \ 0)^T, \text{ the first}$$

entry of any vector in W can have any value. Thus, the vectors

$A^*(\eta_2)$ and $A^*(\eta_3)$ will be linear combinations of the vectors

η_1 , η_2 , and η_3 provided $\begin{pmatrix} \bar{a}_{12} \\ \bar{a}_{13} \end{pmatrix}$ is a scalar multiple of

$$\begin{pmatrix} \bar{a}_{14}a_{24} + \bar{a}_{15}a_{25} \\ \bar{a}_{15}a_{35} \end{pmatrix}. \text{ We will show } \det \begin{pmatrix} \bar{a}_{12} & \bar{a}_{14}a_{24} + \bar{a}_{15}a_{25} \\ \bar{a}_{13} & \bar{a}_{15}a_{35} \end{pmatrix} = 0.$$

This will complete the argument that the subspace W is invariant

under both A and A^* .

$$\text{From (7) we have } \bar{a}_{15} = \frac{\bar{a}_{14}\bar{a}_{25}a_{24}}{|a_{24}|^2 - 1}.$$

In equation (16) we divide by a_{24} and obtain

$$\bar{a}_{12}(1 - |a_{35}|^2) + \bar{a}_{13}\bar{a}_{35}a_{25} = 0.$$

If $|a_{35}|^2 = 1$, then (4) implies $a_{25} = 0$, and in this case we are assuming $a_{25} \neq 0$. Hence $|a_{35}|^2 - 1 \neq 0$ and $\bar{a}_{12} = \frac{\bar{a}_{13}\bar{a}_{35}a_{25}}{|a_{35}|^2 - 1}$.

$$\text{Thus } \bar{a}_{12}\bar{a}_{15}a_{35} = \frac{(\bar{a}_{13}\bar{a}_{35}a_{25})(\bar{a}_{14}\bar{a}_{25}a_{24})a_{35}}{(|a_{35}|^2 - 1)(|a_{24}|^2 - 1)}.$$

Now $(|a_{35}|^2 - 1)(|a_{24}|^2 - 1) = |a_{35}|^2|a_{24}|^2 - |a_{24}|^2 - |a_{35}|^2 + 1$, which by (4) is equal to $|a_{25}|^2$, since in this case we have $a_{45} = 0$.

$$\text{Hence } \bar{a}_{12}\bar{a}_{15}a_{35} = \bar{a}_{13}\bar{a}_{14}a_{24}|a_{35}|^2.$$

$$\begin{aligned} \text{Now } \bar{a}_{13}\bar{a}_{14}a_{24} + \bar{a}_{13}\bar{a}_{15}a_{25} &= \bar{a}_{13}\bar{a}_{14}a_{24} + \frac{\bar{a}_{13}(\bar{a}_{14}\bar{a}_{25}a_{24})a_{25}}{|a_{24}|^2 - 1} \\ &= \bar{a}_{13}\bar{a}_{14}a_{24} \left(\frac{|a_{24}|^2 - 1 + |a_{25}|^2}{|a_{24}|^2 - 1} \right). \end{aligned}$$

Since $a_{45} = 0$, equation (4)

implies $|a_{24}|^2 + |a_{25}|^2 - 1 = |a_{35}|^2(|a_{24}|^2 - 1)$. This shows that

$$\bar{a}_{13}\bar{a}_{14}a_{24} + \bar{a}_{13}\bar{a}_{15}a_{25} = \bar{a}_{13}\bar{a}_{14}a_{24}|a_{35}|^2 = \bar{a}_{12}\bar{a}_{15}a_{35}. \text{ Hence,}$$

$$\det \begin{pmatrix} \bar{a}_{14}a_{24} + \bar{a}_{15}a_{25} & \bar{a}_{12} \\ \bar{a}_{15}a_{35} & \bar{a}_{13} \end{pmatrix} = 0.$$

Thus, the subspace W is invariant under A and A^* . Hence, both W and W^\perp are invariant under A , and A is unitarily similar to a block diagonal matrix. ■

Section 5. Theorem 4 as a consequence of Kippenhahn's conjecture.

In this section we show that Theorem 4 would be a consequence of Kippenhahn's conjecture. First, we review the statement of the conjecture. Let $f(x,y,z) = \det(zI - xH - yK) = \pi_1^{r_1} \pi_2^{r_2} \dots \pi_j^{r_j}$, where π_1, \dots, π_j are distinct, irreducible polynomials and r_1, \dots, r_j are positive integers. Then the minimal polynomial of $xH + yK$ is $m(x,y,z) = \pi_1 \dots \pi_j$. Kippenhahn's conjecture is that if the degree of $m(x,y,z)$ is less than n , (equivalently, $f(x,y,z)$ has a repeated factor), then $A = H + iK$ is unitarily similar to a block diagonal matrix.

Suppose the conjecture is true. Since the characteristic polynomial of a block diagonal matrix $xH + yK$ is the product of the characteristic polynomials of the blocks, repeated application of the conjecture shows that H and K are simultaneously similar to block diagonal matrices, H' and K' , such that the characteristic polynomials of the diagonal blocks of $xH' + yK'$ have no repeated factors. Thus, we may assume

$$xH + yK = \begin{pmatrix} xH_{11} + yK_{11} & & & \\ & xH_{22} + yK_{22} & & \\ & & \dots & \\ & & & xH_{rr} + yK_{rr} \end{pmatrix}$$

where $\det(zI - xH_{11} - yK_{11}) = \pi_1 \dots \pi_j$ and $\det(zI - xH_{ii} - yK_{ii})$ divides $\det(zI - xH_{i-1, i-1} - yK_{i-1, i-1})$ for $i = 2, 3, \dots, r$.

Now suppose π_1 is linear. The polynomial π_1 divides $\det(zI - xH_{ii} - yK_{ii})$ for $i = 1, \dots, r_1$. Consider what happens when $r_1 > \frac{n}{3}$. The first r_1 blocks cannot all be of size 3×3 or greater. Hence, one of the first r_1 blocks is of size 1×1 or 2×2 . Since π_1 divides $\det(zI - xH_{ii} - yK_{ii})$ for $i = 1, \dots, r_1$, this 1×1 or 2×2 block comes from Hermitian matrices, H_{ii} and K_{ii} , with property L. Thus, H_{ii} and K_{ii} can be simultaneously, unitarily diagonalized and H and K have a common eigenvector. Thus, Theorem 4 follows from the Kippenhahn conjecture.

More generally, suppose π_1 has degree d and that $r_1 > \frac{n}{d+1}$. Then not all of the first r_1 blocks can have size greater than d . Hence at least one of the first r_1 blocks is size $d \times d$ and has characteristic polynomial π_1 . Thus, H and K have a common d -dimensional, invariant subspace.

APPENDIX

In this appendix we derive the eighteen equations listed on page 95. This is done by computing 4×4 minors of $xH + yK + \left(\sqrt{x^2 + y^2}\right) I$ and setting the coefficients of the polynomials obtained equal to zero.

$$xH + yK + \left(\sqrt{x^2 + y^2}\right) I =$$

$$\begin{pmatrix} \alpha x + \beta y + \sqrt{x^2 + y^2} & a_{12}(x - iy) & a_{13}(x - iy) & a_{14}(x - iy) & a_{15}(x - iy) \\ \bar{a}_{12}(x + iy) & \sqrt{x^2 + y^2} & 0 & a_{24}(x - iy) & a_{25}(x - iy) \\ \bar{a}_{13}(x + iy) & 0 & \sqrt{x^2 + y^2} & 0 & a_{35}(x - iy) \\ \bar{a}_{14}(x + iy) & \bar{a}_{24}(x + iy) & 0 & \sqrt{x^2 + y^2} & a_{45}(x - iy) \\ \bar{a}_{15}(x + iy) & \bar{a}_{25}(x + iy) & \bar{a}_{35}(x + iy) & \bar{a}_{45}(x + iy) & \sqrt{x^2 + y^2} \end{pmatrix}$$

We first compute the determinant of the minor formed by rows and columns 1, 2, 3, and 4, expanding by cofactors of the entries in column three.

$$a_{13}(x - iy) \det \begin{pmatrix} \bar{a}_{12}(x + iy) & \sqrt{x^2 + y^2} & a_{24}(x - iy) \\ \bar{a}_{13}(x + iy) & 0 & 0 \\ \bar{a}_{14}(x + iy) & \bar{a}_{24}(x + iy) & \sqrt{x^2 + y^2} \end{pmatrix}$$

$$+ \sqrt{x^2 + y^2} \det \begin{pmatrix} \alpha x + \beta y + \sqrt{x^2 + y^2} & a_{12}(x - iy) & a_{14}(x - iy) \\ \bar{a}_{12}(x + iy) & \sqrt{x^2 + y^2} & a_{24}(x - iy) \\ \bar{a}_{14}(x + iy) & \bar{a}_{24}(x + iy) & \sqrt{x^2 + y^2} \end{pmatrix}$$

$$\begin{aligned}
&= -|a_{13}|^2(x^2 + y^2)[(x^2 + y^2) - |a_{24}|^2(x^2 + y^2)] \\
&\quad + \sqrt{x^2 + y^2}[(\alpha x + \beta y + \sqrt{x^2 + y^2})(x^2 + y^2) + a_{12}a_{24}\bar{a}_{14}(x - iy)(x^2 + y^2) \\
&\quad\quad + \bar{a}_{12}\bar{a}_{24}a_{14}(x + iy)(x^2 + y^2) - |a_{12}|^2(x^2 + y^2)\sqrt{x^2 + y^2} \\
&\quad\quad - |a_{14}|^2(x^2 + y^2)\sqrt{x^2 + y^2} - |a_{24}|^2(x^2 + y^2)(\alpha x + \beta y + \sqrt{x^2 + y^2})].
\end{aligned}$$

Removing the common factor $(x^2 + y^2)$, we have

$$\begin{aligned}
&[-|a_{13}|^2 + |a_{13}|^2|a_{24}|^2 + 1 - |a_{14}|^2 - |a_{12}|^2 - |a_{24}|^2](x^2 + y^2) \\
&\quad + [\alpha x + \beta y + a_{12}a_{24}\bar{a}_{14}(x - iy) + \bar{a}_{12}\bar{a}_{24}a_{14}(x + iy) \\
&\quad\quad - |a_{24}|^2(\alpha x + \beta y)]\sqrt{x^2 + y^2} = 0.
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
(1) \quad &|a_{13}|^2 + |a_{14}|^2 + |a_{12}|^2 + |a_{24}|^2 - |a_{13}|^2|a_{24}|^2 = 1 \\
(2) \quad &\alpha(1 - |a_{24}|^2) + a_{12}a_{24}\bar{a}_{14} + \bar{a}_{12}\bar{a}_{24}a_{14} = 0 \\
(3) \quad &\beta(1 - |a_{24}|^2) + i(\bar{a}_{12}\bar{a}_{24}a_{14} - a_{12}a_{24}\bar{a}_{14}) = 0.
\end{aligned}$$

Next, we find the determinant of the minor formed from rows and columns 2, 3, 4, and 5, again expanding by cofactors of the entries in column three.

$$\begin{aligned}
& \sqrt{x^2 + y^2} \det \begin{pmatrix} \sqrt{x^2 + y^2} & a_{24}(x - iy) & a_{25}(x - iy) \\ \bar{a}_{24}(x + iy) & \sqrt{x^2 + y^2} & a_{45}(x - iy) \\ \bar{a}_{25}(x + iy) & \bar{a}_{45}(x + iy) & \sqrt{x^2 + y^2} \end{pmatrix} \\
& + \bar{a}_{35}(x + iy) \det \begin{pmatrix} \sqrt{x^2 + y^2} & a_{24}(x - iy) & a_{25}(x - iy) \\ 0 & 0 & a_{35}(x - iy) \\ \bar{a}_{24}(x + iy) & \sqrt{x^2 + y^2} & a_{45}(x - iy) \end{pmatrix} \\
& = (x^2 + y^2)^2 + a_{24}a_{45}\bar{a}_{25}(x - iy)(x^2 + y^2)\sqrt{x^2 + y^2} \\
& \quad + \bar{a}_{24}\bar{a}_{45}a_{25}(x + iy)(x^2 + y^2)\sqrt{x^2 + y^2} \\
& \quad - (|a_{25}|^2 + |a_{24}|^2 + |a_{45}|^2)(x^2 + y^2)^2 \\
& \quad - |a_{35}|^2(1 - |a_{24}|^2)(x^2 + y^2)^2 \\
& = (1 - |a_{24}|^2 - |a_{25}|^2 - |a_{45}|^2 - |a_{35}|^2 + |a_{24}|^2|a_{35}|^2)(x^2 + y^2)^2 \\
& \quad + [a_{24}a_{45}\bar{a}_{25}(x - iy) + \bar{a}_{24}\bar{a}_{45}a_{25}(x + iy)](x^2 + y^2)\sqrt{x^2 + y^2} = 0.
\end{aligned}$$

Hence, we have

$$(4) \quad |a_{25}|^2 + |a_{35}|^2 + |a_{45}|^2 + |a_{24}|^2 - |a_{24}|^2|a_{35}|^2 = 1$$

$$(5) \quad a_{24}a_{45}\bar{a}_{25} = 0.$$

Now we compute the determinant of the minor formed by rows 1,2,3, and 4 and columns 2, 3, 4, and 5. Expanding by cofactors of the entries in column three, we have

$$\begin{aligned}
 & -a_{13}(x-iy) \det \begin{pmatrix} \sqrt{x^2+y^2} & a_{24}(x-iy) & a_{25}(x-iy) \\ 0 & 0 & a_{35}(x-iy) \\ \bar{a}_{24}(x+iy) & \sqrt{x^2+y^2} & a_{45}(x-iy) \end{pmatrix} \\
 & -\sqrt{x^2+y^2} \det \begin{pmatrix} a_{12}(x-iy) & a_{14}(x-iy) & a_{15}(x-iy) \\ \sqrt{x^2+y^2} & a_{24}(x-iy) & a_{25}(x-iy) \\ \bar{a}_{24}(x+iy) & \sqrt{x^2+y^2} & a_{45}(x-iy) \end{pmatrix} \\
 = & a_{13}a_{35}(1 - |a_{24}|^2)(x^2 + y^2)(x - iy)^2 \\
 & -\sqrt{x^2 + y^2} [a_{12}a_{24}a_{45}(x - iy)^3 + a_{14}a_{25}\bar{a}_{24}(x - iy)^2(x + iy) \\
 & + a_{15}(x - iy)(x^2 + y^2) - a_{15}|a_{24}|^2(x - iy)^2(x + iy) \\
 & - a_{14}a_{45}(x - iy)^2\sqrt{x^2 + y^2} - a_{12}a_{25}(x - iy)^2\sqrt{x^2 + y^2}].
 \end{aligned}$$

Removing the common factor $(x - iy)$, we have

$$\begin{aligned}
 & [a_{13}a_{35}(1 - |a_{24}|^2) + a_{14}a_{45} + a_{12}a_{25}](x - iy)(x^2 + y^2) \\
 & - [a_{12}a_{24}a_{45}(x - iy) + a_{14}a_{25}\bar{a}_{24}(x + iy) + a_{15}(x + iy) \\
 & - a_{15}|a_{24}|^2(x + iy)](x - iy)\sqrt{x^2 + y^2} = 0.
 \end{aligned}$$

This yields

$$(6) \quad a_{13}a_{35}(1 - |a_{24}|^2) + a_{14}a_{45} + a_{12}a_{25} = 0.$$

We also have $a_{12}a_{24}a_{45} + a_{14}a_{25}\bar{a}_{24} + a_{15} - a_{15}|a_{24}|^2 = 0$

and $-a_{12}a_{24}a_{45} + a_{14}a_{25}\bar{a}_{24} + a_{15} - a_{15}|a_{24}|^2 = 0.$

Adding these two equations together yields

$$(7) \quad a_{14}a_{25}\bar{a}_{24} + a_{15}(1 - |a_{24}|^2) = 0$$

$$(8) \quad a_{12}a_{24}a_{45} = 0.$$

The 4×4 minor obtained by removing row five and column two is

$$\det \begin{pmatrix} \alpha x + \beta y + \sqrt{x^2 + y^2} & a_{13}(x - iy) & a_{14}(x - iy) & a_{15}(x - iy) \\ \bar{a}_{12}(x + iy) & 0 & a_{24}(x - iy) & a_{25}(x - iy) \\ \bar{a}_{13}(x + iy) & \sqrt{x^2 + y^2} & 0 & a_{35}(x - iy) \\ \bar{a}_{14}(x + iy) & 0 & \sqrt{x^2 + y^2} & a_{45}(x - iy) \end{pmatrix}.$$

Expanding by cofactors of the entries in the second column, this equals

$$- a_{13}(x - iy) \begin{pmatrix} \bar{a}_{12}(x + iy) & a_{24}(x - iy) & a_{25}(x - iy) \\ \bar{a}_{13}(x + iy) & 0 & a_{35}(x - iy) \\ \bar{a}_{14}(x + iy) & \sqrt{x^2 + y^2} & a_{45}(x - iy) \end{pmatrix}$$

$$- \sqrt{x^2 + y^2} \begin{pmatrix} \alpha x + \beta y + \sqrt{x^2 + y^2} & a_{14}(x - iy) & a_{15}(x - iy) \\ \bar{a}_{12}(x + iy) & a_{24}(x - iy) & a_{25}(x - iy) \\ a_{14}(x + iy) & \sqrt{x^2 + y^2} & a_{45}(x - iy) \end{pmatrix}$$

$$\begin{aligned}
= & -a_{13}(x-iy)[a_{24}a_{35}\bar{a}_{14}(x-iy)^2(x+iy) + \bar{a}_{13}a_{25}(x^2+y^2)\sqrt{x^2+y^2} \\
& - \bar{a}_{13}a_{24}a_{45}(x-iy)^2(x+iy) - \bar{a}_{12}a_{35}(x^2+y^2)\sqrt{x^2+y^2}] \\
& - \sqrt{x^2+y^2}[a_{24}a_{45}(\alpha x + \beta y + \sqrt{x^2+y^2})(x-iy)^2 + |a_{14}|^2a_{25}(x-iy)^2(x+iy) \\
& + \bar{a}_{12}a_{15}(x^2+y^2)\sqrt{x^2+y^2} - \bar{a}_{14}a_{24}a_{15}(x-iy)^2(x+iy) \\
& - \bar{a}_{12}a_{14}a_{45}(x-iy)^2(x+iy) \\
& - a_{25}(\alpha x + \beta y + \sqrt{x^2+y^2})(x-iy)\sqrt{x^2+y^2}] .
\end{aligned}$$

Removing the common factor $(x-iy)$ leaves

$$\begin{aligned}
& [-a_{13}a_{24}a_{35}\bar{a}_{14}(x-iy) + |a_{13}|^2a_{24}a_{45}(x-iy) - a_{24}a_{45}(x-iy) \\
& \quad - \bar{a}_{12}a_{15}(x+iy) + a_{25}(\alpha x + \beta y)] (x^2+y^2) \\
& + [-|a_{13}|^2a_{25}(x+iy) + a_{13}\bar{a}_{12}a_{35}(x+iy) - a_{24}a_{45}(\alpha x + \beta y) - |a_{14}|^2a_{25}(x+iy) \\
& \quad + a_{14}a_{24}a_{15}(x+iy) + a_{12}a_{14}a_{45}(x+iy) + a_{25}(x+iy)] (x-iy)\sqrt{x^2+y^2} = 0.
\end{aligned}$$

From this we get four equations.

$$(9) \quad a_{13}a_{24}a_{35}\bar{a}_{14} - |a_{13}|^2a_{24}a_{45} + a_{24}a_{45} + \bar{a}_{12}a_{15} + a_{25}\alpha = 0$$

$$(10) \quad i(a_{13}a_{24}a_{35}\bar{a}_{14} - |a_{13}|^2a_{24}a_{45} + a_{24}a_{45} - \bar{a}_{12}a_{15}) + a_{25}\beta = 0$$

$$\begin{aligned}
(11) \quad & |a_{13}|^2a_{25} - a_{13}\bar{a}_{12}a_{35} + a_{24}a_{45}\alpha + |a_{14}|^2a_{25} \\
& - \bar{a}_{14}a_{24}a_{15} - \bar{a}_{12}a_{14}a_{45} - a_{25} = 0
\end{aligned}$$

$$\begin{aligned}
(12) \quad & i(|a_{13}|^2a_{25} - a_{13}\bar{a}_{12}a_{35} + |a_{14}|^2a_{25} - \bar{a}_{14}a_{24}a_{15} \\
& - \bar{a}_{12}a_{14}a_{45} - a_{25}) + a_{24}a_{45}\beta = 0
\end{aligned}$$

By adding (11) to i (12) we get

$$(13) \quad a_{24}a_{45}(\alpha + i\beta) = 0.$$

$$\begin{aligned} \text{Now, we are assuming } \det(zI - xH - yK) &= (z - (\alpha x + \beta y))(z^2 - (x^2 + y^2))^2 \\ &= z^5 - (\alpha x + \beta y)z^4 - 2(x^2 + y^2)z^3 + 2(\alpha x + \beta y)(x^2 + y^2)z^2 \\ &\quad + (x^2 + y^2)^2z - (\alpha x + \beta y)(x^2 + y^2)^2. \end{aligned}$$

Since the coefficient of z^3 in the expansion of $\det(zI - xH - yK)$ is the sum of the determinants of all the 2×2 principal minors of the matrix $xH + yK$, we must have

$$(14) \quad |a_{12}|^2 + |a_{13}|^2 + |a_{14}|^2 + |a_{15}|^2 + |a_{24}|^2 + |a_{25}|^2 \\ + |a_{35}|^2 + |a_{45}|^2 = 0.$$

Adding equations (1) and (4) and subtracting (14) then shows

$$(15) \quad |a_{24}|^2(1 - |a_{13}|^2 - |a_{35}|^2) - |a_{15}|^2 = 0.$$

The minor obtained by removing row one and column four is

$$\det \begin{pmatrix} \bar{a}_{12}(x+iy) & \sqrt{x^2+y^2} & 0 & a_{25}(x-iy) \\ \bar{a}_{13}(x+iy) & 0 & \sqrt{x^2+y^2} & a_{35}(x-iy) \\ \bar{a}_{14}(x+iy) & \bar{a}_{24}(x+iy) & 0 & a_{45}(x-iy) \\ \bar{a}_{15}(x+iy) & \bar{a}_{25}(x+iy) & \bar{a}_{35}(x+iy) & \sqrt{x^2+y^2} \end{pmatrix}.$$

Expanding by cofactors of the entries in column four, we have

$$\begin{aligned}
& -\sqrt{x^2+y^2} \det \begin{pmatrix} \bar{a}_{12}(x+iy) & \sqrt{x^2+y^2} & a_{25}(x-iy) \\ \bar{a}_{14}(x+iy) & \bar{a}_{24}(x+iy) & a_{45}(x-iy) \\ \bar{a}_{15}(x+iy) & \bar{a}_{25}(x+iy) & \sqrt{x^2+y^2} \end{pmatrix} \\
& - \bar{a}_{35}(x+iy) \det \begin{pmatrix} \bar{a}_{12}(x+iy) & \sqrt{x^2+y^2} & a_{25}(x-iy) \\ \bar{a}_{13}(x+iy) & 0 & a_{35}(x-iy) \\ \bar{a}_{14}(x+iy) & \bar{a}_{24}(x+iy) & a_{45}(x-iy) \end{pmatrix} \\
& = -\sqrt{x^2+y^2} [\bar{a}_{12}\bar{a}_{24}(x+iy)^2\sqrt{x^2+y^2} + \bar{a}_{15}a_{45}(x^2+y^2)\sqrt{x^2+y^2} \\
& \quad + \bar{a}_{14}\bar{a}_{25}a_{25}(x+iy)^2(x-iy) - \bar{a}_{15}\bar{a}_{24}a_{25}(x+iy)^2(x-iy) \\
& \quad - \bar{a}_{14}(x+iy)(x^2+y^2) - \bar{a}_{12}\bar{a}_{25}a_{45}(x+iy)^2(x-iy)] \\
& - \bar{a}_{35}(x+iy) [\bar{a}_{14}a_{35}(x^2+y^2)\sqrt{x^2+y^2} + \bar{a}_{13}\bar{a}_{24}a_{25}(x+iy)^2(x-iy) \\
& \quad - \bar{a}_{13}a_{45}(x^2+y^2)\sqrt{x^2+y^2} - \bar{a}_{12}\bar{a}_{24}a_{35}(x+iy)^2(x-iy)].
\end{aligned}$$

Removing the common factor of $(x+iy)$, we get

$$\begin{aligned}
& [-\bar{a}_{12}\bar{a}_{24}(x+iy) - \bar{a}_{15}a_{45}(x-iy) - \bar{a}_{13}\bar{a}_{24}\bar{a}_{35}a_{25}(x+iy) \\
& \quad + \bar{a}_{12}\bar{a}_{24}|a_{35}|^2(x+iy)](x^2+y^2) \\
& + [-\bar{a}_{14}a_{25} + \bar{a}_{15}\bar{a}_{24}a_{25} + \bar{a}_{14} + \bar{a}_{12}\bar{a}_{25}a_{45} - |a_{35}|^2\bar{a}_{14} \\
& \quad + \bar{a}_{13}\bar{a}_{35}a_{45}](x^2+y^2)\sqrt{x^2+y^2} \\
& = 0.
\end{aligned}$$

This yields

$$(a) \quad \bar{a}_{12}\bar{a}_{24} + \bar{a}_{15}a_{45} + \bar{a}_{13}\bar{a}_{35}\bar{a}_{24}a_{25} - \bar{a}_{12}\bar{a}_{24}|a_{35}|^2 = 0$$

$$(b) \quad \bar{a}_{12}\bar{a}_{24} - \bar{a}_{15}a_{45} + \bar{a}_{13}\bar{a}_{24}\bar{a}_{35}a_{25} - \bar{a}_{12}\bar{a}_{24}|a_{35}|^2 = 0.$$

From (a) + (b) and (a) - (b) we obtain

$$(16) \quad \bar{a}_{12}\bar{a}_{45} + \bar{a}_{13}\bar{a}_{24}\bar{a}_{35}a_{25} - \bar{a}_{12}\bar{a}_{24}|a_{35}|^2 = 0$$

$$(17) \quad \bar{a}_{15}a_{45} = 0.$$

Finally, we have

$$(18) \quad \bar{a}_{14}|a_{25}|^2 - \bar{a}_{15}\bar{a}_{24}a_{25} - \bar{a}_{14} - \bar{a}_{25}a_{45}\bar{a}_{12} + \bar{a}_{14}|a_{35}|^2 - \bar{a}_{13}\bar{a}_{35}a_{45} = 0.$$

REFERENCES

1. S. A. Amitsur, J. Levitzki, Minimal identities for algebras, Proc. Amer. Math. Soc., vol. 1, 1950, pp. 449-463.
2. G. P. Barker, L. Q. Eifler, T. P. Kezlan, A non-commutative spectral theorem, Linear Algebra Appl., vol. 20, 1978, pp. 95-100.
3. W. Burnside, On the condition of reducibility of any group of linear substitutions, Proc. Lond. Math. Soc., (2), vol. 3, 1905, pp. 430-434.
4. P. M. Cohn, Algebra, Vol 2, John Wiley and Sons, New York, 1977.
5. J. L. Coolidge, A Treatise on Algebraic Plane Curves, Oxford University Press, Oxford, 1931.
6. W. F. Donoghue, On the numerical range of a bounded operator, Michigan Math. J., vol. 4, 1957, pp. 261-263.
7. M. P. Drazin, On diagonalizable and normal matrices, Quart. J. Math., Oxford Ser. (2), vol. 2, 1951, pp. 189-198.
8. M. P. Drazin, Some generalizations of matrix commutativity, Proc. London Math. Soc., vol. 1, 1951, pp. 222-231.
9. M. P. Drazin, J. W. Dungey, K. W. Gruenberg, Some theorems on commutative matrices, J. London Math. Soc., vol 26, 1951, pp. 221-228.
10. E. Formanek, Central polynomials for matrix rings, J. Algebra, vol. 23, 1972, pp. 129-131.
11. G. Frobenius, Über vertauschbare Matrizen, Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin, 1896, pp. 601-614.
12. F. J. Gaines, Kato-Taussky-Wielandt commutator relations, Linear Algebra Appl., vol. 1, 1968, pp. 127-138.
13. F. J. Gaines, Kato-Taussky-Wielandt commutator relations and characteristic curves, Pacific J. of Math., vol. 61, 1975, pp. 121-128.
14. F. J. Gaines, Some generalizations of commutativity for linear transformations of a finite dimensional vector space, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1966.
15. F. J. Gaines, R. C. Thompson, Sets of nearly triangular matrices, Duke Math J., vol. 35, 1968, pp. 441-453.
16. F. Hausdorff, Der Wertvorrat einer Bilinearform, Math. Z., vol. 3, 1919, pp. 314-316.

17. N. Jacobson, Lectures in Abstract Algebra, Vol. II, Linear Algebra, D. Van Nostrand Company, Inc., New York, 1953.
18. N. Jacobson, PI-Algebras, An Introduction, Springer-Verlag, New York, 1975.
19. C. R. Johnson, Normality and the numerical range, Linear Algebra Appl., vol. 15, 1976, pp. 89-94.
20. T. Kato, O. Taussky, Commutators of A and A^* , J. Wash. Acad. Sci., vol. 46, 1956, pp. 38-40.
21. R. Kippenhahn, Über den Wertevorrat einer Matrix, Math. Nachrichten, vol. 6, 1951-52, pp. 193-228.
22. T. J. Laffey, Simultaneous quasi-diagonalization of complex matrices, Linear Algebra Appl. vol. 16, 1977, pp. 189-201.
23. J. Levitzki, A theorem on polynomial identities, Proc. Amer. Math. Soc., vol. 1, 1950, pp. 334-341.
24. N. H. McCoy, On the characteristic roots of matrix polynomials, Bull. Amer. Math. Soc., vol. 42, 1936, pp. 592-600.
25. N. H. McCoy, Quasi-commutative matrices, Trans. Amer. Math. Soc., vol. 36, 1934, pp. 327-340.
26. T. Motzkin, On $L(S)$ -tuples and l -pairs of matrices, Pacific J. Math., vol. 40, 1972, pp. 639-645.
27. T. S. Motzkin, O. Taussky, Pairs of matrices with property L , Trans. Amer. Math. Soc., vol. 73, 1952, pp. 108-114.
28. T. S. Motzkin, O. Taussky, Pairs of matrices with property L , II, Trans. Amer. Math. Soc., vol. 80, 1955, pp. 387-401.
29. B. N. Moyls, M. D. Marcus, Field convexity of a square matrix, Proc. Amer. Math. Soc., vol. 6, 1955, pp. 981-983.
30. F. D. Murnaghan, On the field of values of a square matrix, Proc. Nat. Acad. of Science U.S.A., vol. 18, 1932, pp. 246-248.
31. C. Pearcy, A complete set of unitary invariants for operators generating finite W^* -algebras of type I, Pacific J. Math., vol. 12, 1962, pp. 1405-1416.
32. C. Pearcy, A complete set of unitary invariants for 3×3 complex matrices, Trans. Amer. Math. Soc., vol. 104, 1962, pp. 425-429.
33. E. Primrose, Plane Algebraic Curves, Macmillan and Company, Limited, London, 1955.

34. Ju. P. Razmyslov, Trace identities of full matrix algebras over a field of characteristic zero, *Math. USSR Izvestiya*, vol. 8, 1974, pp. 727-760.
35. S. Rosset, A new proof of the Amitsur-Levitzki identity, *Israel J. Math.*, vol. 23, 1976, pp. 187-188.
36. W. E. Roth, On k -commutative matrices, *Trans. Amer. Math. Soc.*, vol. 39, 1936, pp. 483-495.
37. G. Salmon, *A Treatise on the Higher Plane Curves*, Hodges, Foster, and Figgis, Dublin, 1879.
38. H. Shapiro, The characteristic polynomial of a pencil generated by a pair of Hermitian matrices, (Abstract) *Notices Amer. Math. Soc.*, vol. 26 (1), 1979, p. A-49.
39. H. Shapiro, Simultaneous block triangularization and block diagonalization of sets of matrices, *Linear Algebra Appl.*, to appear.
40. W. Specht, Zur Theorie der Gruppen linearer Substitutionen, II, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 49, 1939, pp. 207-215.
41. W. Specht, Zur Theorie der Matrizen, II, *Jber. Deutsche Math.-Vereinig.*, vol. 50, 1940, pp. 19-23.
42. O. Taussky, A characterization of property L, *Journal de Mathématiques Pures et Appliquées*, 1961, pp. 291-293.
43. O. Taussky, Commutativity in finite matrices, *Amer. Math. Mon.*, vol. 64, 1957, pp. 227-235.
44. O. Taussky, Some results concerning the transition from the L- to the P- property for pairs of finite matrices, I, *J. Algebra*, vol. 20, 1972, pp. 271-283.
45. O. Taussky, Some results concerning the transition from the L- to the P- property for pairs of finite matrices, II, *Linear and Multilinear Alg.*, vol. 2, 1974, pp. 195-202.
46. O. Taussky, A weak property L for pairs of matrices, *Math. Z.* vol. 71, 1959, pp. 463-465.
47. O. Toeplitz, Das algebraische Analogon zu einem Satze von Fejer, *Math. Z.*, vol. 2, 1918, pp. 187-197.
48. D. B. Wales, H. J. Zassenhaus, On L-groups, *Mathematische Annalen*, vol. 198, 1972, pp. 1-12.
49. R. J. Walker, *Algebraic Curves*, Princeton University Press, Princeton, 1950.

50. J. P. Watters, Simultaneous quasi-diagonalization of matrices, *Linear Algebra Appl.*, vol. 9, 1974, pp. 103-117.
51. N. Wiegmann, Normal matrices with property L, *Proc. Amer. Math. Soc.*, vol. 4, 1953, pp. 35-36.
52. H. Wielandt, On the eigenvalues of $A+B$ and AB , *J. Res. Nat. Bur. Standards, Section B, 77B*, 1953, pp. 61-63.
53. H. Wielandt, Pairs of normal matrices with property L, *J. Res. Nat. Bur. Standards*, vol. 51, 1953, pp. 89-90.