SOME GENERALIZATIONS OF COMMUTATIVITY FOR LINEAR TRANSFORMATIONS ON A FINITE DIMENSIONAL VECTOR SPACE

Thesis by
Fergus John Gaines

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

California Institute of Technology
Pasadena, California
1966
(Submitted April 4, 1966)

ACKNOWLEDGMENTS

I wish to thank my thesis advisor, Dr. Olga Taussky Todd, for suggesting much of the area of research discussed in this thesis. Her advice and encouragement have been invaluable to me.

My special thanks go to Dr. E. A. Bender for much helpful criticism and advice, especially in connection with Chapter IV. Dr. E. C. Dade also assisted with some ideas in the earlier stages of my research.

Finally, I am indebted to the California Institute of Technology and to the National Science Foundation for financial support in the form of teaching and research assistantships.

ABSTRACT

Let $\mathcal{L}$ be the algebra of all linear transformations on an n-dimensional vector space V over a field $\mathcal{F}$ and let $A, B \in \mathcal{L}$. Let $A_{i+1} = A_i B - BA_i$, $i = 0,1,2,\ldots$, with $A = A_0$. Let $f_k(A,B;\delta) = A_{2K+1} - \delta_1 A_{2K-1} + \delta_2 A_{2K-3} - \ldots + (-1)^K \delta_K A_1$ where $\delta = (\delta_1, \delta_2, \ldots, \delta_K)$, δ_i belong to $\mathcal{F}$ and $K = k(k-1)/2$. Taussky and Wielandt [Proc. Amer. Math. Soc., 13(1962), 732-735] showed that $f_n(A,B;\delta) = 0$ if δ_i is the i^{th} elementary symmetric function of $(\beta_r - \beta_s)^2$, $1 \leq r < s \leq n$, $i = 1, 2, \ldots, N$, with $N = n(n-1)/2$, where β_r are the characteristic roots of B. In this thesis we discuss relations involving $f_k(X,Y;\delta)$ where $X, Y \in \mathcal{L}$ and $1 \leq k < n$. We show: 1. If $\mathcal{F}$ is infinite and if for each $X \in \mathcal{L}$ there exists δ so that $f_k(A,X;\delta) = 0$ where $1 \leq k < n$, then A is a scalar transformation. 2. If $\mathcal{F}$ is algebraically closed, a necessary and sufficient condition that there exists a basis of V with respect to which the matrices of A and B are both in block upper triangular form, where the blocks on the diagonals are either one- or two-dimensional, is that certain products $X_1 X_2 \ldots X_r$ belong to the radical of the algebra generated by A and B over $\mathcal{F}$, where X_i has the form $f_2(A,P(A,B);\delta)$, for all polynomials $P(x,y)$. We partially generalize this to the case where the blocks have dimensions $\leq k$. 3. If A and B generate $\mathcal{L}$, if the characteristic of $\mathcal{F}$ does not divide n and if there exists δ so that $f_k(A,B;\delta) = 0$, for some k with $1 \leq k < n$, then the characteristic roots of B belong to the splitting field of $g_k(w;\delta) = w^{2K+1} - \delta_1 w^{2K-1} + \delta_2 w^{2K-3} - \ldots \ldots + (-1)^K \delta_K w$ over $\mathcal{F}$. We use this result to prove a theorem involving a generalized form of property L [cf. Motzkin and Taussky, Trans. Amer. Math. Soc., 73(1952), 108-114]. 4. Also we give mild generalizations of results of McCoy [Amer. Math. Soc. Bull., 42(1936), 592-600] and Drazin [Proc. London Math. Soc., 1(1951), 222-231].

TABLE OF CONTENTS

INTRODUCTION

Let $\mathcal{L}$ be the algebra of all linear transformations on an n-dimensional vector space V over a field $\mathcal{F}$ and let $A,B \varepsilon \mathcal{L}$. It is well-known [cf. Jacobson, 4 pp. 120-121] that there exists a basis of V with respect to which the matrices of A and B are in block upper triangular form where corresponding blocks are of the same dimensions, the blocks on the diagonal are square and corresponding diagonal blocks cannot be reduced further by a simultaneous similarity. Most of this thesis is devoted to the problem of determining how certain properties of A and B are reflected in properties of the diagonal blocks. In [1] McCoy showed that, if $\mathcal{F}$ is algebraically closed, all the diagonal blocks are one-dimensional if and only if $P(A,B)(AB - BA)$ is nilpotent for every polynomial $P(x,y)$ in the non-commuting variables x and y with coefficients in $\mathcal{F}$. In Chapter I we prove a generalization of McCoy's theorem when the field $\mathcal{F}$ is quite arbitrary. We also generalize a theorem of Drazin which is related to McCoy's result.

To assist with the investigation of the diagonal blocks we introduce what we call "KTW commutator expressions" $f_k(A,B;\mathfrak{d})$. Let $A_{i+1} = A_iB - BA_i$, $i = 0,1,2,\ldots$, with $A_0 = A$. Let k be an integer, $1 \leq k \leq n$, and $K = k(k-1)/2$, then define

$$f_k(A,B;\mathfrak{d}) = A_{2K+1} - \mathfrak{d}_1A_{2K-1} + \mathfrak{d}_2A_{2K-3} - \cdots + (-1)^K\mathfrak{d}_KA_1$$

where $\mathfrak{d} = (\mathfrak{d}_1,\mathfrak{d}_2,\ldots,\mathfrak{d}_K)$ with $\mathfrak{d}_i \varepsilon \mathcal{F}$, $i = 1,2,\ldots,K$. We note $f_1(A,B;\mathfrak{d}) = AB - BA$. The Kato-Taussky-Wielandt (KTW) commutator relation [cf. 12] then says that $f_n(A,B;\mathfrak{d}) = 0$ if $\mathfrak{d}_i$ is the i^{th} elementary symmetric function of $(\beta_r - \beta_s)^2$, $1 \leq r < s \leq n$, where β_r are the characteristic roots of B. In Chapter II we prove that, if $\mathcal{F}$ is infinite and if for each $X \in \mathcal{L}$ there exists $\mathfrak{d}$ so that, $f_k(A,X;\mathfrak{d}) = 0$ for some k with $1 \leqslant k < n$, then A is a scalar transformation. We also show that if $\mathfrak{F}$ is algebraically closed then there exists a basis of V with respect to which the matrices of A and B are in block upper triangular form, where the diagonal blocks have dimensions $\leqslant k$, if, for each polynomial $P(x,y)$, there exists $\mathfrak{d}$ so that $f_k(A,P(A,B);\mathfrak{d})$ belongs to the radical $\mathcal{J}$ of the algebra generated by A and B over $\mathfrak{F}$.

In Chapter III we characterize those A,B for which there exists a basis of V with respect to which the matrices have block upper triangular forms where the diagonal blocks are one- or two-dimensional. The necessary and sufficient condition we give is, essentially, that certain products $X_1X_2 \ldots X_r$ belong to $\mathcal{J}$, where X_i has the form $f_2(A,P(A,B);\mathfrak{d})$, for every polynomial $P(x,y)$.

Now $f_1(A,B;\mathfrak{d}) = AB - BA$ and the Kato-Taussky-Wielandt commutator relation says that there exists $\mathfrak{d}$ so that $f_n(A,B;\mathfrak{d}) = 0$. So in Chapter IV we examine what happens if there exists $\mathfrak{d}$ so that $f_k(A,B;\mathfrak{d}) = 0$ for some k with $1 \leqslant k < n$. We show that if A and B generate $\mathcal{L}$, if the characteristic of $\mathfrak{F}$ does not divide n and if there exists $\mathfrak{d}$ so that $f_k(A,B;\mathfrak{d}) = 0$, for some k with $1 \leqslant k < n$, then the characteristic roots β_r of B belong to the splitting field of $g_k(w;\mathfrak{d})$ over $\mathfrak{F}$, where

$$g_k(w;\mathfrak{d}) = w^{2K+1} - \mathfrak{d}_1 w^{2K-1} + \ldots + (-1)^K \mathfrak{d}_K w.$$

Moreover if $k = 2$ and B has at least two distinct characteristic roots then there exists an ordering $\beta_1, \beta_2, \ldots, \beta_r$ of the distinct characteristic roots of B so that $\beta_1 - \beta_2 = \beta_2 - \beta_3 = \ldots = \beta_{r-1} - \beta_r$ satisfies $g_2(w;\mathfrak{d}) = 0$. Now $f_1(A,B;\mathfrak{d}) = 0$ means $AB = BA$ and this in turn implies A and B have property P, which means A and B have property L [cf. 16]. Property L demands that $xA + yB$ have all its characteristic roots of the form $x\alpha + y\beta$, for all $x,y \in \mathfrak{F}$.

We employ the main theorem of Chapter IV to prove a result which says that if A and B satisfy a certain KTW commutator equation then they have a generalized property L.

The results which are, perhaps, of most interest are Theorems 2.10, 2.11, 2.13, 3.1, 4.2, 4.9 and Corollary 4.6.

Note: throughout this thesis the symbol ■ denotes the end of a proof.

CHAPTER I

ON THEOREMS OF McCOY AND DRAZIN.

In this chapter we prove some thearems which generalize results of McCoy and Drazin on matrix commutators. Rational methods are used except in part of Theorem 1.9 and in Lemma 1.12.

Let A and B be linear transformations on an n-dimensional vector space V over a field $\mathfrak{F}$. To set the stage we introduce the so-called "P-property" and state the main theorem about it.

1.1 <u>Definition</u> [cf. McCoy, 1] . If $\mathfrak{F}$ is algebraically closed then A and B have <u>property P</u> if and only if there exist orderings $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$ of the characteristic roots of A and B, respectively, so that the characteristic roots of P(A,B) are $P(\alpha_i, \beta_i)$, $i = 1,2,\ldots,n$, for every polynomial P(x,y) in the non-commuting variables x and y, with coefficients in $\mathfrak{F}$.

Unless otherwise stated, "polynomial P(x,y)" in this thesis will always mean a polynomial in the non-commuting variables x and y with coefficients in $\mathfrak{F}$.

1.2 <u>Theorem</u> [McCoy, 1] . Assume $\mathfrak{F}$ is algebraically closed. Then the following three conditions are equivalent.

(i) A and B have property P.

(ii) P(A,B)(AB - BA) is nilpotent for every polynomial P(x,y).

(iii) There exists a basis of V with respect to which the matrices of A and B are in upper triangular form, i.e. all the elements below the main diagonals are zeros.

We aim to prove a generalization (Theorem 1.9) of this theorem where we do not assume that the field $\mathcal{F}$ is algebraically closed.

We shall need the well-known concept given in the following definition.

1.3 Definition [cf. Herstein, 2, p.4]. Let $\mathcal{R}$ be a ring, M an irreducible left $\mathcal{R}$-module, $O(M) = \{r \in \mathcal{R} : rM = (0)\}$. Then the (Jacobson) radical of $\mathcal{R}$ is $\bigcap O(M)$, where this intersection is taken over all irreducible $\mathcal{R}$-modules M.

The following result is well-known.

1.4 Theorem [cf. McCoy, 3, pp. 113, 120]. The radical of $\mathcal{R}$ contains every nil left (or right) ideal of $\mathcal{R}$. If $\mathcal{R}$ satisfies the descending chain condition on left ideals, then the radical of $\mathcal{R}$ is nilpotent.

In particular, this theorem applies if $\mathcal{R}$ is a finite-dimensional algebra over a field.

Now let $\mathcal{R}$ be the algebra of polynomials P(A,B) in A and B (including I, the identity transformation) and let $\mathcal{J}$ be the radical of $\mathcal{R}$. We denote AB - BA by [A,B].

1.5 Definition. A and B have property Q if and only if $[A,B] \in \mathcal{J}$.

1.6 Lemma [cf. McCoy, 1]. Let $C \in \mathcal{R}$. Then $C \in \mathcal{J}$ if and only if P(A,B)C is nilpotent for every polynomial P(x,y).

Proof. If $C \in \mathcal{J}$ then $P(A,B)C \in \mathcal{J}$ for every polynomial P(x,y). Hence P(A,B)C is nilpotent, by Theorem 1.4.

Conversely, if P(A,B)C is nilpotent for every polynomial P(x,y), then the left ideal $\mathcal{R}\cdot C$ is nil and hence is contained in $\mathcal{J}$, by Theorem 1.4. Thus $C \in \mathcal{J}$ since $I \in \mathcal{R}$. ∎

We note that essentially the same result was proved in McCoy[1] in a somewhat different manner.

1.7 <u>Lemma</u>. If A and B have property Q and V is irreducible as a left $\mathcal{R}$-module, then $AB = BA$.
<u>Proof</u>. $[A,B] \varepsilon \mathcal{J}$ and hence acts as 0 on all irreducible $\mathcal{R}$-modules. Hence $AB = BA$. ∎

The following definition is a natural analog of (iii) of Theorem 1.2.

1.8 <u>Definition</u>. A and B have <u>property T</u> if and only if there exists a basis of V with respect to which the matrices of A and B have the following block forms

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1t} \\ 0 & A_{22} & \cdots & A_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & A_{tt} \end{bmatrix} \text{ and } \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ 0 & B_{22} & \cdots & B_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & B_{tt} \end{bmatrix}$$

respectively, where A_{ii} and B_{ii} are square blocks of dimension n_i, $i = 1,2,\ldots,t$, and there are zeros below the main block diagonals. Also $A_{ii}B_{ii} = B_{ii}A_{ii}$ and the minimum polynomials of A_{ii} and B_{ii} are irreducible over $\mathcal{F}$, for $i = 1,2,\ldots,t$.

We prove now the promised generalization of Theorem 1.2.

1.9 <u>Theorem.</u> The following three conditions are equivalent.
(i) A and B have property Q.
(ii) A and B have property T.
(iii) There exist orderings $\alpha_1,\alpha_2,\ldots,\alpha_n$ and $\beta_1,\beta_2,\ldots,\beta_n$ of the characteristic roots (which are contained in some extension of $\mathcal{F}$) of A and B, respectively, so that P(A,B)

has characteristic roots $P(\alpha_i, \beta_i)$, $i = 1,2,\ldots,n$, for every polynomial $P(x,y)$.

Proof. We shall prove that (i)$\Rightarrow$(ii)$\Rightarrow$(iii)$\Rightarrow$(i). Let A and B have property Q. Let

$$V = V_1 \supset V_2 \supset \cdots \supset V_{t+1} = (0)$$

be an $\mathcal{R}$-composition series for V. In the usual manner [cf. Jacobson, 4, p. 120] choose a basis for V so that the matrices of A and B have the same block forms as the matrices in 1.8, where A_{ii} (resp. B_{ii}) is the matrix of A (resp. B) restricted to the quotient space V_i/V_{i+1}, with $i = 1,2,\ldots,t$. Now V_i/V_{i+1} is irreducible as an $\mathcal{R}$-module. Since A and B have property Q, the matrices A_{ii} and B_{ii} have property Q for $i = 1,2,\ldots,t$. Hence $A_{ii}B_{ii} = B_{ii}A_{ii}$ by Lemma 1.7.

Now the minimum polynomial of A_{ii} (resp. B_{ii}) is irreducible over $\mathcal{F}$, for each $i = 1,2,\ldots,t$. Suppose the minimum polynomial $p(x)$ of A_{ii} is reducible for some i. Then $p(x) = q(x)r(x)$, where $q(x)$ and $r(x)$ are non-constant polynomials with coefficients in $\mathcal{F}$. Let $W = V_i/V_{i+1}$. Let $W_1 = \{w \varepsilon W: r(A_{ii})w = 0\}$. Now W_1 is clearly a subspace of W and also $A_{ii}W_1 \subset W_1$. Now $B_{ii}W_1 \subset W_1$ since, if $r(A_{ii})w = 0$ then

$$0 = Br(A_{ii})w = r(A_{ii})Bw$$

because $A_{ii}B_{ii} = B_{ii}A_{ii}$. Thus W_1 is an $\mathcal{R}$-submodule of W. But W is irreducible. Hence $W_1 = (0)$ or $W_1 = W$. Now $W_1 \neq W$, since, then, $p(x)$ would not be the minimum polynomial of A_{ii}. Hence $W_1 = (0)$. But then $r(A_{ii})$ is non-singular and, since $0 = q(A_{ii})r(A_{ii})$, we have $q(A_{ii}) = 0$. This contradicts the fact that $p(x)$ is the minimum polynomial of A_{ii}. Hence $p(x)$ must be irreducible. In a similar fashion it can be shown that the minimum polynomial of B_{ii} is irreducible. (We note that the discussion of this paragraph is contained in Jacobson [4, p. 133]). Thus A and B have property T, so (i) $\Rightarrow$ (ii).

Now suppose A and B have property T. We have $A_{ii}B_{ii} = B_{ii}A_{ii}$, $i = 1,2,\ldots,t$. Let $\mathcal{G}$ be the algebraic closure of $\mathcal{F}$. Let $\mathcal{R}_i$ be the algebra of all polynomials $P(A_{ii}, B_{ii})$ where $P(x,y)$ is a polynomial in the non-commuting variables x and y with coefficients in $\mathcal{G}$. Then $[A_{ii}, B_{ii}]$ belongs to the radical of $\mathcal{R}_i$. Now consider A_{ii} and B_{ii} as matrices with elements in $\mathcal{G}$ and apply the fact that (i) $\Rightarrow$ (ii) in the present theorem. Hence there exists a non-singular matrix U_i with elements in $\mathcal{G}$ so that $U_i^{-1}A_{ii}U_i$ and $U_i^{-1}B_{ii}U_i$ are both in upper triangular form (i.e. there are only zeros below the main diagonals), $i = 1,2,\ldots,t$. So, if A_1 and B_1 denote the matrices in 1.8, there exists a non-singular nxn matrix U with elements in $\mathcal{G}$ so that $U^{-1}A_1U$ and $U^{-1}B_1U$ are both in upper triangular form. The matrix $U^{-1}A_1U$ (resp. $U^{-1}B_1U$) has the characteristic roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ (resp. $\beta_1, \beta_2, \ldots, \beta_n$) of A (resp. B) on the main diagonal. Thus $P(A,B)$ has characteristic roots $P(\alpha_i, \beta_i)$, $i = 1,2,\ldots,n$ for every polynomial $P(x,y)$ in the non-commuting variables x and y with coefficients in $\mathcal{G}$. Then (iii) follows as an immediate consequence.

We now show (iii) $\Rightarrow$ (i). Assume (iii). Then $P(A,B)[A,B]$ has only 0 as a characteristic root, and hence is nilpotent, for every polynomial $P(x,y)$. Thus A and B have property Q by Lemma 1.6. ∎

1.10 <u>Corollary</u> [Jacobson, 4, p.133] . If $AB = BA$, then A and B have property T.

1.11 <u>Corollary</u> Let A and B be nxn matrices with elements in $\mathcal{F}$. Then each of the statements (i), (ii) and (iii) in Theorem 1.9 is equivalent to the fact that A and B have property P, considered as matrices with elements in

$\mathcal{G}$, where $\mathcal{G}$ is the algebraic closure of $\mathcal{F}$.

Proof. If there are orderings $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$ of the characteristic roots of A and B, respectively, so that P(A,B) has characteristic roots $P(\alpha_i, \beta_i)$, $i = 1,2,\ldots,n$, for each polynomial P(x,y) in the non-commuting variables x and y with coefficients in $\mathcal{G}$, then this statement is clearly true for those P(x,y) with coefficients in $\mathcal{F}$. Hence A and B satisfy (iii) of Theorem 1.9.

The converse is contained in the proof given above that (ii) $\Rightarrow$ (iii) in Theorem 1.9. ∎

We remark that if $\mathcal{F}$ is algebraically closed in Theorem 1.9, we get Theorem 1.2.

The following lemma gives an example of a pair of linear transformations with property Q.

1.12 Lemma. Let A and B be linear transformations on an n-dimensional vector space V over a field $\mathcal{F}$, where the characteristic of $\mathcal{F}$ does not divide n. If each polynomial P(A,B) has only one characteristic root then A and B have property Q.

Proof. All the characteristic roots of [A,B] are equal - equal c, say. Thus 0 = trace [A,B] = nc. Hence c = 0. Now P(A,B)[A,B] has all its characteristic roots equal, for each polynomial P(A,B). But 0 is a characteristic root, since [A,B] is nilpotent. Hence P(A,B) [A,B] is nilpotent and thus A and B have property Q. ∎

McCoy's formulation of property Q is that P(A,B)[A,B] is nilpotent for each polynomial P(A,B). We raise the question here whether it is possible to assume P(A,B)[A,B] nilpotent for a smaller class of polynomials P(A,B) and still get property Q. Williamson [5] has shown that if $\mathcal{F}$ is the complex numbers and if A is non-derogatory (i.e.

if the minimum polynomial of A equals its characteristic polynomial) and if we assume $h(A)[A,B]$ nilpotent for certain polynomials $h(A)$ of degree $\leqslant n - 2$, then A and B have property Q.

We should at least like to bound the degrees of the $P(A,B)$. If $\mathcal{F}$ is algebraically closed, then a careful examination of the proof of Drazin et al. [6] of McCoy's theorem (Theorem 1.2) gives the result that if $P(A,B)[A,B]$ is nilpotent for each polynomial $P(A,B)$ of degree $\leqslant$ $n(a + b - 1) - 2$ where a and b are the degrees of the minimum polynomials of A and B, respectively, then A and B have property P. This bound seems much too large. We prove the following result for some small values of n and conjecture its validity in general.

1.13 <u>Theorem</u>. Let A and B be linear transformations on V of dimension $n (n \geqslant 2)$ over a field $\mathcal{F}$ so that $P(A,B)[A,B]$ is nilpotent for each polynomial $P(x,y)$ of degree $\leqslant n - 2$ then for $n = 2,3$ the transformations A and B have property Q.

<u>Proof</u>. <u>Case 1</u>. $n = 2$. We assume $[A,B]$ is nilpotent. Hence either $[A,B] = 0$ or the rational canonical form matrix of $[A,B]$ is $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. If $[A,B] = 0$, the result is true. Otherwise replace A and B by matrices, again called A and B, so that $[A,B] = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. There should be no confusion caused by the new meanings for A and B. We have $A[A,B] = [A,AB]$ and also $B[A,B] = [BA,B]$. Hence $A[A,B]$ and $B[A,B]$ are nilpotent (since each matrix has its determinant and its trace equal to zero). If $X = (x_{ij})$ then

$$X[A,B] = \begin{bmatrix} 0 & x_{11} \\ 0 & x_{21} \end{bmatrix}$$

and, if this matrix is nilpotent, $x_{21} = 0$. Hence if $A = (a_{ij})$ and $B = (b_{ij})$ we get $a_{21} = b_{21} = 0$ and thus the matrices A and B are in upper triangular form. Hence A and B have property Q.

Case 2. We assume that $(xA + yB + zI)[A,B]$ is nilpotent for all $x,y,z \in \mathcal{F}$ and V is three-dimensional over $\mathcal{F}$. As above assume that A and B are matrices so that $[A,B]$ is in rational canonical form. Hence

$$[A,B] = 0 \quad \text{or} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

If $[A,B] = 0$, the result is true. Assume

$$[A,B] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If $X = (x_{ij})$, then

$$X[A,B] = \begin{bmatrix} 0 & x_{11} & x_{12} \\ 0 & x_{21} & x_{22} \\ 0 & x_{31} & x_{32} \end{bmatrix}.$$

Now let $X = xA + yB + zI$, where $A = (a_{ij})$ and $B = (b_{ij})$. Then $X[A,B]$ is nilpotent, and thus $\begin{bmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}$ is nilpotent. Thus we have

$$\begin{bmatrix} xa_{21} + yb_{21} & xa_{22} + yb_{22} + z \\ xa_{31} + yb_{31} & xa_{32} + yb_{32} \end{bmatrix}$$

nilpotent. Thus both the trace and the determinant are zero. Hence $a_{32} = -a_{21}$ and $b_{32} = -b_{21}$. Also

$$-(xa_{21} + yb_{21})^2 - (xa_{22} + yb_{22} + z)(xa_{31} + yb_{31}) = 0.$$

This gives $a_{31} = b_{31} = 0$ and $a_{21} = b_{21} = 0$. Thus A and B are both in upper triangular form, and property Q follows immediately.

If $[A,B] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then

$$X[A,B] = \begin{bmatrix} 0 & x_{11} & 0 \\ 0 & x_{21} & 0 \\ 0 & x_{31} & 0 \end{bmatrix}$$

where $X = (x_{ij}) = xA + yB + zI$ with $A = (a_{ij})$ and $B = (b_{ij})$. Then $X[A,B]$ nilpotent gives $x_{21} = 0$, which leads to $a_{21} = 0$ and $b_{21} = 0$. If both a_{31} and b_{31} are zero then A and B are in block upper triangular form, and we can apply Case 1 to the 2x2 blocks to prove the result.

So assume $a_{31} \neq 0$. Then let $b_{31} = ha_{31}$. The (2,1) element of $[A,B]$ is $ha_{31}a_{23} - b_{23}a_{31}$ and this is zero. Hence $b_{23} = ha_{23}$, since $a_{31} \neq 0$. Thus

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & ha_{23} \\ ha_{31} & b_{32} & b_{33} \end{bmatrix}.$$

If $a_{23} = 0$, then, by means of a permutation similarity, simultaneously interchange the first and second rows and the first and second columns of A and B. The matrices obtained are in block lower triangular form and, once again, we can apply Case 1 to the 2x2 blocks to prove the result.

So assume $a_{23} \neq 0$. The (1,1) element of $[A,B]$ gives

$$a_{11}b_{11} + ha_{13}a_{31} - b_{11}a_{11} - b_{13}a_{31} = 0.$$

Hence $b_{13} = ha_{13}$, since $a_{31} \neq 0$. The (2,2) element of $[A,B]$ gives

$$a_{22}b_{22} + a_{23}b_{32} - b_{22}a_{22} - ha_{23}a_{32} = 0$$

and hence $b_{32} = ha_{32}$, since $a_{23} \neq 0$. The (3,1) element gives

$$a_{31}b_{11} + ha_{33}a_{31} - ha_{31}a_{11} - b_{33}a_{31} = 0$$

and hence $b_{11} - b_{33} = h(a_{11} - a_{33})$, since $a_{31} \neq 0$. The (2,3) element gives

$$ha_{22}a_{23} + a_{23}b_{33} - b_{22}a_{23} - ha_{23}a_{33} = 0$$

and hence $b_{22} - b_{33} = h(a_{22} - a_{33})$. The (1,3) element gives

$$ha_{11}a_{13} + ha_{12}a_{23} + a_{13}b_{33} - b_{11}a_{13} - b_{12}a_{23} - ha_{13}a_{33} = 0$$

which is the same as

$$[(b_{33} - b_{11}) - h(a_{33} - a_{11})]a_{13} + a_{23}(ha_{12} - b_{12}) = 0$$

which gives $b_{12} = ha_{12}$ on applying the equation derived from the (3,1) element, and using the fact that $a_{23} \neq 0$.
Then the (1,2) element of $[A,B]$ is

$$ha_{11}a_{12} + a_{12}b_{22} + ha_{13}a_{32} - b_{11}a_{12} - ha_{12}a_{22} - ha_{13}a_{32}$$

which equals $a_{12}[h(a_{11} - a_{22}) - (b_{11} - b_{22})]$. This last expression equals

$$a_{12}[h(a_{11} - a_{33}) - (b_{11} - b_{33}) + h(a_{33} - a_{22}) \\ - (b_{33} - b_{22})]$$

which is zero, by the identities derived from the (3,1) and (2,3) elements. But the (1,2) element of $[A,B]$ is 1. This contradiction shows that at least one of a_{31} and a_{23} is zero and thus the result is true, by the arguments given above for these cases. ∎

Drazin [7] has proved a generalization of McCoy's theorem (Theorem 1.2) when the field $\mathfrak{F}$ is algebraically closed of characteristic zero. We shall now generalize Drazin's results to an arbitrary field $\mathfrak{F}$ (insofar as they do generalize). Our proofs, of course, are completely different.

1.14 Definition.

$$\mathcal{C}_0 = \{A,B\}, \quad \mathcal{C}_1 = \{[A,B]\},$$

$$\mathcal{C}_{k+1} = \bigcup \{[A,C_k], [B,C_k]\} \text{ for } k \geqslant 1,$$

where the union is over all linear transformations C_k in $\mathcal{C}_k$.

We remark that there are at most 2^{k-1} linear transformations in $\mathcal{C}_k$, $k \geqslant 1$.

1.15 Definition. [Drazin, 7]. A and B have property Q_k for some $k \geqslant 1$ if and only if $P(A,B)C_k$ is nilpotent for every C_k in $\mathcal{C}_k$ and for all polynomials $P(x,y)$.

It is clear, by Lemma 1.6, that A and B have property Q_k if and only if $\mathcal{C}_k \subset \mathcal{J}$. Property Q_1 is thus property Q.

1.16 Definition. [Drazin, 7]. A and B are quasi k-commutative for some $k \geqslant 1$ if and only if $\mathcal{C}_k = \{0\}$.

We shall need the following lemma which is due to Jacobson.

1.17 Lemma. [8] Let $\mathfrak{A}$ be an associative algebra over a field $\mathfrak{F}$ and let $x, y \in \mathfrak{A}$. Assume that x commutes with $[x,y]$ and that x is algebraic over $\mathfrak{F}$ with minimum polynomial of degree r. Then if $\mathfrak{F}$ has characteristic 0 or $p > r$, the commutator $[x,y]$ is nilpotent.

Proof. Jacobson stated this theorem for characteristic 0 and there is a slight error in his proof (a fact which was pointed out by Drazin [unpublished]). We present a repaired version of Jacobson's proof. Let $m(z)$ be the minimum polynomial of x over $\mathfrak{F}$. Then $m(x) = 0$. The mapping $x \to x' = [x,y]$ is a derivation. Hence $m'(x)x' = 0$. Assume that

$$m^{(k)}(x)(x')^{2^k-1} = 0, \text{ for some } k \geqslant 1.$$

Hence

$$m^{(k+1)}(x)(x')^{2^k} + m^{(k)}(x)((x')^{2^k-1})' = 0.$$

Multiply on the left by $(x')^{2^k-1}$ and use the fact that $[x,x'] = 0$. Hence

$$m^{(k+1)}(x)(x')^{2^{k+1}-1} = 0.$$

Thus, by induction,

$$m^{(k)}(x)(x')^{2^k-1} = 0 \text{ for } k = 1,2,\ldots \ .$$

Now $m^{(r)}(x) = r!$. Hence

$$r!(x')^{2^r-1} = 0 \text{ and } (x')^{2^r-1} = 0,$$

since the characteristic of $\mathfrak{F}$ does not divide $r!$. ∎

We apply this lemma in proving the following result.

1.18 <u>Lemma</u>. Let $\mathfrak{F}$ have characteristic 0 or $p > n = \dim. V$. Suppose A and B are quasi k-commutative for some $k \geqslant 1$ and that V is irreducible as an $\mathcal{R}$-module, then AB = BA.

Proof. The proof is by induction on k. Assume $\mathcal{C}_k = \{0\}$. If k = 1, then AB = BA and the result is proved. Assume the lemma is true for some $k \geq 1$. Suppose $\mathcal{C}_{k+1} = \{0\}$, $k \geq 1$. Then $AC_k = C_kA$ and $BC_k = C_kB$ for all C_k in $\mathcal{C}_k$. Now each C_k in $\mathcal{C}_k$ is of the form $[A,C_{k-1}]$ or $[B,C_{k-1}]$ for some C_{k-1} in $\mathcal{C}_{k-1}$. Suppose $X = [A, C_{k-1}] \neq 0$ for some C_{k-1} in $\mathcal{C}_{k-1}$. Then X is nilpotent by Lemma 1.17. We see that $V_1 = XV$ is an invariant subspace of V for both A and B, since A and B commute with X. Hence either $V_1 = V$ or $V_1 = (0)$, since V is an irreducible $\mathcal{R}$-module. But $V_1 \neq (0)$, since $X \neq 0$. Hence $V_1 = V$ and thus X is nonsingular. But X is nilpotent and this gives a contradiction. Thus we must have $[A,C_{k-1}] = 0$. By similar reasoning $[B,C_{k-1}] = 0$. Thus $\mathcal{C}_k = \{0\}$ and, by the induction hypothesis, this means AB = BA. ∎

The theorem we now prove is our generalization of Drazin's main theorem[7].

1.19 Theorem. If $\mathcal{F}$ has characteristic 0 or $p > n$, then A and B have property Q_k for some $k \geq 1$ if and only if they have property T.

Proof. It is clear (as in the proof of Theorem 1.9) that, if A and B have property T, we can find a basis of V so that the matrix of each $P(A,B)C_k$, with respect to that basis, has zero blocks on and below the main block diagonal. Hence $P(A,B)C_k$ is nilpotent for all polynomials $P(x,y)$ and for each C_k in $\mathcal{C}_k$. Thus A and B have property Q_k, $k = 1,2,\ldots$.

Conversely, let A and B have property Q_k for some $k \geq 1$. Use the composition series argument of Theorem 1.9. Then A_{ii} and B_{ii} of that theorem are quasi k-commutative here, and, since V_i/V_{i+1} is an irreducible $\mathcal{R}$-module, we actually have $A_{ii}B_{ii} = B_{ii}A_{ii}$, $i = 1,2,\ldots,t$, by Lemma 1.18.

The minimum polynomial of A_{ii} (resp. B_{ii}) is irreducible as in Theorem 1.9, $i = 1,2,\ldots,t$. Hence A and B have property T. ∎

1.20 Corollary. If $\mathcal{F}$ has characteristic 0 or $p > n$ then the properties Q_k are equivalent, $k = 1,2,\ldots$.

1.21 Corollary. If $\mathcal{F}$ has characteristic 0 or $p > n$, and if A and B are quasi k-commutative for some $k \geqslant 1$, then they have property Q.

Finally we shall give a counter-example to Theorem 1.19, when the conditions on the characteristic of $\mathcal{F}$ are not satisfied. In order to construct the example we shall need the following well-known result.

1.22 Theorem. [Shoda, 9; Albert and Muckenhoupt, 10]. Let Z be an nxn matrix with elements in a field $\mathcal{F}$. Then there exist nxn matrices X and Y, with elements in $\mathcal{F}$, so that $Z = [X,Y]$, if and only if trace $Z = 0$.

1.23 Example. Let $\mathcal{F}$ have characteristic $p > 0$ and let I be the pxp identity matrix. Then trace $I = 0$ and hence there exist pxp matrices A and B with elements in $\mathcal{F}$ so that $I = [A,B]$, by Theorem 1.22. Then A and B are quasi 2-commutative, but they do not have property Q, since e.g. $[A,B]$ is not nilpotent.

For $n > p$ we construct a counter-example by "filling out" the above matrices with zeros.

CHAPTER II

THE KATO-TAUSSKY-WIELANDT COMMUTATOR RELATION

In the first chapter we considered a generalization of matrix commutativity (property Q) and extended known results by considering an arbitrary field $\mathcal{F}$. From here on we shall seek to generalize property Q itself, and thus to generalize the notion of commutativity. The kind of generalizations we want are stated as Problem 2.1.

So let A and B be linear transformations on an n-dimensional vector space V over a field $\mathcal{F}$. Again let $\mathcal{R}$ be the algebra generated by A and B and $\mathcal{J}$ the radical of $\mathcal{R}$. By applying the composition series argument cited in Theorem 1.9 to V as an $\mathcal{R}$-module we get the matrices

$$(*) \quad \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1t} \\ 0 & A_{22} & \cdots & A_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & A_{tt} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ 0 & B_{22} & \cdots & B_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & B_{tt} \end{bmatrix}$$

of A and B with respect to a suitable basis of V where A_{ii} and B_{ii} (which are square blocks of dimension n_i) cannot be reduced further by a simultaneous similarity , $i = 1,2,\ldots,t$. When $\mathcal{F}$ is algebraically closed McCoy's theorem (Theorem 1.2) characterizes those A,B for which $n_i = 1$, $i = 1,2,\ldots$. Actually, for McCoy's theorem to hold, it is only necessary to assume that $\mathcal{F}$ contains the characteristic roots of A and B. We now raise the following general question.

2.1 <u>Problem.</u> How can property Q be generalized so as to characterize those A and B for which the matrices in (*)

have diagonal blocks A_{ii} and B_{ii} of specific dimensions, $i = 1,2,\ldots,t$?

In this chapter and the next we partially answer this question (cf. Theorems 2.13 and 3.1).

If we wish to have A_{11} and B_{11}, say, of dimension k, then somehow we must introduce a relation satisfied identically by kxk matrices. This leads us to a commutator relation which was proved by Kato and Taussky [11] in the two-dimensional case, and by Taussky and Wielandt [12] in the general case.

Let A and B be linear transformations on V. Let $A_{i+1} = [A_i, B]$ for $i = 0,1,2,\ldots$, where $A_o = A$. Let $K = k(k-1)/2$, where k is a positive integer and let $\delta_1, \delta_2, \ldots, \delta_K$ be K arbitrary elements in $\mathcal{F}$.

2.2 Definition.

$$f_k(A,B;\delta) = A_{2K+1} - \delta_1 A_{2K-1} + \delta_2 A_{2K-3} - \cdots + (-1)^K \delta_K A_1$$

$$\text{where } \delta = (\delta_1, \ldots, \delta_K).$$

2.3 Theorem [11,12].

$$f_n(A,B;\delta) = 0$$

if δ_i, $i = 1,2,\ldots,N$, are the elementary symmetric functions of $(\beta_r - \beta_s)^2$, $1 \le r < s \le n$, where β_r are the characteristic roots of B, $i = 1,2,\ldots,n$, and $N = n(n-1)/2$.

Note.

$$\delta_1 = \sum_{1 \le r < s \le n} (\beta_r - \beta_s)^2, \quad \delta_2 = \sum (\beta_r - \beta_s)^2 (\beta_t - \beta_u)^2 \text{ etc.},$$

where the second sum is taken over $1 \le r < s \le n$, $1 \le t < u \le n$ and $(r,s) < (t,u)$, and this last ordering is lexicographic.

We shall call expressions of the type $f_k(A,B;\delta)$

"Kato-Taussky-Wielandt commutator expressions" or "KTW commutator expressions", for short. Theorem 2.3 is then the <u>Kato-Taussky-Wielandt commutator relation</u>.

We remark that Theorem 2.3 is independent of A. We can express the result in a somewhat different form. Let T_B be the linear transformation on the space of all linear transformations on V defined by $T_B(X) = [X,B]$, then Theorem 2.3 becomes

2.4 <u>Theorem</u> [Taussky and Wielandt, 12; Khan, 13].

$$T_B \prod_{r<s}(T_B^2 - (\beta_r - \beta_s)^2 I) = 0.$$

We do not need to assume that $(\beta_r - \beta_s)^2 \varepsilon \mathcal{F}$, $1 \leq r < s \leq n$, in either Theorem 2.3 or Theorem 2.4.

We note that $f_1(A,B;\mathfrak{d}) = [A,B]$ and thus, if there exist elements $\mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_K$ in $\mathcal{F}$ so that $f_k(A,B;\mathfrak{d}) = 0$ for some k with $1 \leq k < n$, we might suppose this to be a "good" generalization of commutativity, in that it might give a partial answer to the question raised in 2.1. However the condition turns out to be a bit too weak for this. This is borne out by Example 2.6.

2.5 <u>Definition</u>. If $X_1, X_2, \ldots, X_k$ are linear transformations on V then

$$[X_1, X_2, \ldots, X_k] = [[X_1, X_2, \ldots, X_{k-1}], X_k]$$

where, as usual, $[X_1, X_2] = X_1X_2 - X_2X_1$.

2.6 <u>Example</u>. Let $\mathcal{F}$ have characteristic $\neq 2$ or 3 and assume $\sqrt{3}\ \varepsilon\, \mathcal{F}$. Let A and B be matrices,

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and } B = \begin{bmatrix} x & x & x \\ x & -(1+x)/2 & (1-x)/2 \\ x & (1-x)/2 & -(1+x)/2 \end{bmatrix}$$

where $x = 1/\sqrt{3}$. Now

$$f_2(A,B;\mathfrak{d}) = [A,B,B,B] - \mathfrak{d}_1[A,B].$$

We have $A^2 = B^2 = I$ and this gives

$$[A,B,B,B] = 4[A,B].$$

Hence $f_2(A,B;\mathfrak{d}) = 0$, where $\mathfrak{d}_1 = 4$. Incidentally, we also have $f_2(B,A;\mathfrak{d}) = 0$, where $\mathfrak{d}_1 = 4$. But even with both $f_2(A,B;\mathfrak{d})$ and $f_2(B,A;\mathfrak{d})$ zero, we cannot transform A and B to the form (*) by a simultaneous similarity, (even if we extend the field $\mathfrak{F}$) where the diagonal blocks A_{ii} and B_{ii} have dimensions ≤ 2. This comes from the fact that A and B have neither a row characteristic vector nor a column characteristic vector in common. Since A and B are symmetric, it is only necessary to verify this last statement for row vectors. The characteristic vectors of A are $(a,0,-a)$ where $a \in \mathfrak{F}$, corresponding to the characteristic value -1; and (a,b,a) where $a,b \in \mathfrak{F}$, corresponding to the characteristic value $+1$. Clearly B cannot have $(a,0,-a)$ as a characteristic vector. Suppose B has a characteristic vector of the form (a,b,a). Hence $(a,b,a)B = \pm 1(a,b,a)$. Thus

$$(x(2a + b),\ (a - b)(1 + x)/2,\ -(a - b)(1 - x)/2) = \pm(a,b,a).$$

Hence

$$x(2a + b) = \pm a \text{ or } b = -a(2x \mp 1)/x.$$

Also

$$(a - b)(1 + x) = \pm 2b \text{ or } b = a(1 + x)/(1 + x \pm 2).$$

But since $x = 1/\sqrt{3}$, these equations contradict each other unless $a = b = 0$. Hence A and B have no characteristic vectors in common. ∎

We shall, in Chapter IV, discuss a result concerning linear transformations A and B which satisfy $f_k(A,B;\mathfrak{d}) = 0$ for some k with $1 \leq k < n$. But the result in this chapter (Theorem 2.13) which sheds some light on Problem 2.1 involves KTW expressions of the type $f_k(A,P(A,B);\mathfrak{d})$, where $P(x,y)$ is a polynomial in x and y.

Before we can prove any results about KTW commutator expressions we need some preliminary results. The following well-known theorem will be of use to us.

2.7 Theorem [cf. Jacobson, 14, p.112]. Let $\mathfrak{F}$ be an infinite field and $g(x_1,x_2,\ldots,x_r)$ a non-zero polynomial in the polynomial domain $\mathfrak{F}[x_1,x_2,\ldots,x_r]$, where the x_i are algebraically independent, then there exist elements $c_1,c_2,\ldots,c_r$ in $\mathfrak{F}$ so that $g(c_1,c_2,\ldots,c_r) \neq 0$.

We need this to prove the next lemma.

2.8 Lemma. Let $\mathfrak{F}$ be an infinite field and n a positive integer, then it is possible to choose $x_1,x_2,\ldots,x_n$ in $\mathfrak{F}$ so that the $n(n-1)/2$ elements $(x_i - x_j)^2$, with $1 \leq i < j \leq n$, are all distinct and non-zero.

Proof. Let

$$g(y_1,y_2,\ldots,y_n) = \prod\{(y_i - y_j)^2 - (y_r - y_s)^2\}.$$

where the product is taken over $1 \leq i < j \leq n$ and $1 \leq r < s \leq n$ with $(i,j) \neq (r,s)$. Since $g(y_1,y_2,\ldots,y_n) \neq 0$, by Theorem 2.7 there exist $x_1,x_2,\ldots,x_n \in \mathfrak{F}$ so that $g(x_1,x_2,\ldots,x_n) \neq 0$. This proves the result. ∎

2.9 Lemma. Let $\mathfrak{F}$ be any field and X an nxn matrix with elements in $\mathfrak{F}$ so that, for every non-singular matrix U with elements in $\mathfrak{F}$, the matrix $U^{-1}XU$ is diagonal, then X is a scalar matrix.

Proof. Let $X = \operatorname{diag}(x_1, x_2, \ldots, x_n)$. Let $U_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Let $U = U_1$ if $n = 2$ and $U = U_1 \oplus I_{n-2}$ otherwise, where I_{n-2} is the $(n-2)x(n-2)$ identity matrix. Then

$$U^{-1}XU = \begin{bmatrix} x_1 & x_1-x_2 & 0\ldots0 \\ 0 & x_2 & 0\ldots0 \\ \cdot & \cdot & \cdot \quad \cdot \\ \cdot & \cdot & \cdot \quad \cdot \\ 0 & 0 & 0\ldots x_n \end{bmatrix}$$

and, since this matrix must be diagonal, we get $x_1 = x_2$. By a simultaneous permutation of rows and columns (using a permutation similarity) we can replace x_2 by any x_i for $i \geqslant 3$. Hence $x_1 = x_i$, $i = 2,3,\ldots$ and thus $X = x_1 I$. ∎

The result we prove next will be applied to prove a theorem about KTW commutator expressions, but it is of some interest in itself.

2.10 Theorem. Let $\mathfrak{F}$ be an infinite field and X an nxn matrix with elements in $\mathfrak{F}$. If X is not a scalar matrix, then there exists a non-singular matrix U with elements in $\mathfrak{F}$ so that $U^{-1}XU$ has none of its elements zero.

Proof. Let $X = (x_{ij})$ be the given matrix. We shall subject X to a succession of similarity transformations to put it in the required form. To avoid cumbersome notation, after each similarity we shall still refer to the new matrix as $X = (x_{ij})$. Since X is not a scalar matrix, by Lemma 2.8 we may assume it has at least one off-diagonal element which is non-zero (transform X by a similarity transformation, if necessary). By a simultaneous permutation of rows and columns (using a permutation similarity) put this element in the (1,2) place of X. Thus we may assume $X = (x_{ij})$ with $x_{12} \neq 0$.

Let $S = I + yE_{ij}$, with $i \neq j$, where I is the nxn identity matrix, E_{ij} has 1 in the (i,j) place and zeros elsewhere and $y \varepsilon \mathfrak{F}$. Then $S^{-1} = I - yE_{ij}$ and hence

$$S^{-1}XS = X + yXE_{ij} - yE_{ij}X - y^2E_{ij}XE_{ij}.$$

Thus the effect of the similarity $S^{-1}XS$ is to add y times the i^{th} column of X to the j^{th} column, subtract y times the j^{th} row from the i^{th} row and change the element x_{ij} to $x_{ij} + y(x_{ii} - x_{jj}) - y^2x_{ji}$. Call such a similarity transformation an <u>elementary similarity</u>.

Now consider the matrix X with $x_{12} \neq 0$. By means of an elementary similarity add y_i times the second column of X to the i^{th} column, for each $i \neq 2$, where $y_i = 0$ if $x_{1i} \neq 0$ and $y_i = 1$ if $x_{1i} = 0$. We have possibly changed some rows of X, but, in any case, we now have a new matrix $X = (x_{ij})$ with $x_{11}, x_{12}, \ldots, x_{1n}$ all non-zero.

We shall prove, by induction, that we can transform X by a similarity so that the first n - 1 rows of the matrix obtained contain no zeros. If n = 2, we have already proved this statement. If $n > 2$, let us assume that we have succeeded in transforming X by a similarity so that the first k rows contain no zeros, where $1 \leq k < n - 1$. So now we assume $X = (x_{ij})$ where the first k rows contain no zeros. If $x_{k+1,k+2} \neq 0$ we can proceed as in the next paragraph. Otherwise, by means of an elementary similarity subtract y times the k^{th} row from the $(k+1)^{st}$ row. This adds y times the $(k+1)^{st}$ column to the k^{th} column and changes the value of the element in the (k+1,k) place. Let

$$g(y) = y\prod_{i=1}^{k}(x_{ik} + yx_{i,k+1}).$$

Since $g(y) \not\equiv 0$, by Theorem 2.7 we can choose $y \varepsilon \mathcal{F}$ so that $g(y)$ is not zero (because $\mathcal{F}$ is an infinite field). We thus get a matrix $X = (x_{ij})$ whose first k rows contain no zeros and for which $x_{k+1,k+2} \neq 0$.

For each $i \neq k + 2$ add y_i times the $(k + 2)^{nd}$ column to the i^{th}, by means of an elementary similarity. This operation subtracts y_i times the i^{th} row from the $(k + 2)^{nd}$ row and changes the value of the element in the $(k + 2,i)$ place. So for each $i \neq k + 2$ let

$$g(y_i) = \prod_{j=1}^{k+1}(x_{ji} + y_i x_{j,k+2}).$$

Since $g(y_i) \not\equiv 0$, choose $y_i \varepsilon \mathcal{F}$ so that $g(y_i) \neq 0$. Thus, by induction, we have shown that the given matrix can be transformed by a similarity transformation so that the first $n - 1$ rows of the matrix obtained contain no zeros.

So now we have a matrix $X = (x_{ij})$ with zeros, perhaps, only in the last row. By means of an elementary similarity subtract y times the $(n - 1)^{st}$ row from the n^{th} row. This adds y times the n^{th} column to the $(n - 1)^{st}$ column and changes $x_{n,n-1}$ to $x_{n,n-1}+y(x_{nn}-x_{n-1,n-1})-y^2x_{n-1,n}$. Let

$$g(y) = [x_{n,n-1} + y(x_{nn} - x_{n-1,n-1}) - y^2x_{n-1,n}]\cdot$$
$$\prod_{i=1}^{n}{}'(x_{ni} - yx_{n-1,i})\prod_{i=1}^{n-1}(x_{in-1} + yx_{in})$$

where the prime means that the term containing $i = n - 1$ is omitted. Since $g(y) \not\equiv 0$ we can choose $y \varepsilon \mathcal{F}$ so that $g(y) \neq 0$. Thus we have obtained an $X = (x_{ij})$ where none of the elements is zero, and this proves the theorem. ∎

We shall now prove our first result involving KTW commutator expressions. It supports the idea that we can use the vanishing of a KTW commutator expression as a generalization of commutativity. The theorem we prove generalizes the fact that the center of the algebra of linear transformations on a finite-dimensional vector space over a field is the scalar transformations.

2.11 Theorem. Let V be an n-dimensional vector space over an infinite field $\mathfrak{F}$, A a linear transformation on V and k an integer with $1 \leq k < n$. Suppose that for each linear transformation X on V there exist elements $\delta_1, \delta_2, \ldots, \delta_K$ in $\mathfrak{F}$, where $K = k(k-1)/2$, so that

$$f_k(A, X; \delta) = 0,$$

then A is a scalar transformation.

Proof. Let B be the matrix of A with respect to some basis of V. Then for each nxn matrix Y with elements in $\mathfrak{F}$ there exist elements $\delta_1, \delta_2, \ldots, \delta_K$ in $\mathfrak{F}$ so that $f_k(B, Y; \delta) = 0$. Let $Y = \mathrm{diag}(y_1, y_2, \ldots, y_n)$ where the $y_i \in \mathfrak{F}$ are such that $(y_i - y_j)2$ are distinct and non-zero for $1 \leq i < j \leq n$ (by Lemma 2.8). If $B = (b_{ij})$, we get

$$b_{ij} y_{ji} (y^K - \delta_1 y^{K-1} + \ldots + (-1)^K \delta_K) = 0$$

where $y = y_{ji}^2$ and $y_{ji} = y_j - y_i$. Since there are $N = n(n-1)/2$ distinct non-zero values for y_{ji}^2 with $1 \leq i < j \leq n$ and $N > K$, we get $b_{ij} = 0$ for some i, j. Thus the matrix of A with respect to any basis of V has at least one element zero. Hence A is a scalar transformation, by Theorem 2.10. $\blacksquare$

We use the following known result in proving one of the main theorems of this chapter.

2.12 Theorem ["Burnside's theorem", cf.4, p.276]. If $\mathcal{R}$ is an irreducible algebra of linear transformations on a finite-dimensional vector space V over an algebraically closed field, then $\mathcal{R}$ is the complete algebra of linear transformations on V.

The next theorem is a generalization of part of McCoy's theorem on property Q (Theorem 1.2) and it gives some information about a solution to Problem 2.1.

2.13 Theorem. Let A and B be linear transformations on an n-dimensional vector space V over an algebraically closed field $\mathcal{F}$. Let k be an integer with $1 \leq k < n$. Suppose that for each polynomial $P(x,y)$, there exist $\delta_1, \delta_2, \ldots, \delta_K$ in $\mathcal{F}$, where $K = k(k-1)/2$, so that

$$f_k(A, P(A,B); \delta) \ \varepsilon \ \mathcal{J}$$

then there exists a basis of V with respect to which the matrices of A and B have the forms (*), where $f_k(A_{ii}, P(A_{ii}, B_{ii}); \delta) = 0$ and $\dim. A_{ii} = \dim. B_{ii} \leq k$.

Proof. By the usual composition series argument we can find a basis of V so that the matrices of A and B have the forms (*) where A_{ii} and B_{ii} cannot be reduced any further by a simultaneous similarity and $f_k(A_{ii}, P(A_{ii}, B_{ii}); \delta) = 0$ with $i = 1,2,\ldots,t$. We claim that $\dim. A_{ii} = \dim. B_{ii} \leq k$ for $i = 1,2,\ldots,t$. For suppose $\dim. A_{ii} = n_i > k$ for some i. Since A_{ii} and B_{ii} cannot be reduced by the same similarity, the algebra $\mathcal{R}_i$ of polynomials in A_{ii} and B_{ii} is irreducible and hence by "Burnside's theorem" (Theorem 2.12), $\mathcal{R}_i$ is the complete algebra of $n_i x n_i$ matrices with elements in $\mathcal{F}$. Hence, given any $n_i x n_i$ matrix X with elements in $\mathcal{F}$, we can find K elements $\delta_1, \delta_2, \ldots$

$\ldots,\delta_K$ in $\mathfrak{F}$ so that $f_k(A_{ii},X;\delta) = 0$. Since $k<n_i$ this means A_{ii} is a scalar matrix, by Theorem 2.10. But then B_{ii} can be reduced to upper triangular form by a similarity (since $\mathfrak{F}$ is algebraically closed) and this leaves A_{ii} unchanged. This contradicts the fact that A_{ii} and B_{ii} cannot be reduced by the same similarity transformation. Thus $n_i \leqslant k$ and this completes the proof of the theorem. ∎

Remark. This theorem does not characterize those A,B for which the matrices in (*) have dim.A_{ii}(resp. dim.B_{ii}) $\leqslant k$. The exact conditions on A and B for this characterization would appear to be quite complicated. However, in Chapter III we give necessary and sufficient conditions on A and B that the matrices in (*) have dim.A_{ii} = dim.B_{ii} $\leqslant 2$.

We close this chapter with some remarks on the Kato-Taussky-Wielandt commutator relation. When V is two-dimensional the relation is

$$[A,B,B,B] - (\beta_1 - \beta_2)^2[A,B] = 0$$

or

$$T_B(T_B^2 - (\beta_1 - \beta_2)^2 I) = 0$$

where β_1 and β_2 are the characteristic roots of B and T_B is the linear transformation defined on the space of all linear transformations on V by $T_B(X) = [X,B]$. The Kato-Taussky-Wielandt commutator relation is not the most general commutator relation between A and B in the two-dimensional case. We prove the following generalization.

2.14 Theorem. Let A and B be linear transformations on a two-dimensional vector space V over a field $\mathfrak{F}$ and let the linear transformation T_X be defined by $T_X(Y) = [Y,X]$

where X and Y are linear transformations on V. Then

$$T_B T_A T_B - xT_B = 0$$

where x = trace Atrace B + 2detA + 2detB - 2det(A + B)#.

Proof. Replace A and B by matrices, again called A and B. We let $A = (a_{ij})$ and $B = (b_{ij})$. Then

$$T_A = \begin{bmatrix} 0 & -a_{12} & a_{21} & 0 \\ -a_{21} & \alpha_{12} & 0 & a_{21} \\ a_{12} & 0 & \alpha_{21} & -a_{12} \\ 0 & a_{12} & -a_{21} & 0 \end{bmatrix}, \quad T_B = \begin{bmatrix} 0 & -b_{12} & b_{21} & 0 \\ -b_{21} & \beta_{12} & 0 & b_{21} \\ b_{12} & 0 & \beta_{21} & -b_{12} \\ 0 & b_{12} & -b_{21} & 0 \end{bmatrix}$$

where $\alpha_{ij} = a_{ii} - a_{jj}$ and $\beta_{ij} = b_{ii} - b_{jj}$. We are using the fact that , if X is any square matrix, then $T_X = X^t \otimes I - I \otimes X$, where "t" means transpose and "$\otimes$" is the tensor product sign [cf. 13]. Let I be the 4x4 identity matrix, then

$$T_A T_B - xI =$$

The notation "detX" means "the determinant of the linear transformation X".

$$\begin{bmatrix} a_{12}b_{21} + a_{21}b_{12} - x & -a_{12}\beta_{12} & a_{21}\beta_{21} & -a_{12}b_{21} - a_{21}b_{12} \\ -\alpha_{12}b_{21} & 2a_{21}b_{12} + \alpha_{12}\beta_{12} - x & -2a_{21}b_{21} & \alpha_{12}b_{21} \\ \alpha_{21}b_{12} & -2a_{12}b_{12} & 2a_{12}b_{21} + \alpha_{21}\beta_{21} - x & -\alpha_{21}b_{12} \\ -a_{12}b_{21} - a_{21}b_{12} & a_{12}\beta_{12} & -a_{21}\beta_{21} & a_{12}b_{21} + a_{21}b_{12} - x \end{bmatrix}$$

Let $y = 2a_{21}b_{12} + 2a_{12}b_{21} + \alpha_{12}\beta_{12} - x$. Then

$$T_BT_AT_B - xT_B = \begin{bmatrix} 0 & -b_{12}y & b_{21}y & 0 \\ -b_{21}y & \beta_{12}y & 0 & b_{21}y \\ b_{12}y & 0 & \beta_{21}y & -b_{12}y \\ 0 & b_{12}y & -b_{21}y & 0 \end{bmatrix}$$

where, in computing, we use the facts that $\alpha_{12} = -\alpha_{21}$ and $\beta_{12} = -\beta_{21}$. Thus $y = 0$ gives

$$T_BT_AT_B - xT_B = 0.$$

This happens if

$$x = 2a_{21}b_{12} + 2a_{12}b_{21} + \alpha_{12}\beta_{12}.$$

Now

$$\begin{aligned} & \text{traceAtraceB} + 2\det A + 2\det B - 2\det(A+B) \\ = \; & (a_{11} + a_{22})(b_{11} + b_{22}) + 2(a_{11}a_{22} - a_{21}a_{12}) + \\ & +2(b_{11}b_{22} - b_{21}b_{12}) - 2(a_{11}a_{22} - a_{21}a_{12}) + \\ & -2(b_{11}b_{22} - b_{21}b_{12}) - 2(a_{11}b_{22} - a_{21}b_{12}) + \\ & -2(b_{11}a_{22} - b_{21}a_{12}) \\ = \; & (a_{11} - a_{22})(b_{11} - b_{22}) + 2a_{21}b_{12} + 2a_{12}b_{21} \\ = \; & x. \end{aligned}$$

This completes the proof of the theorem. ∎

2.15 <u>Corollary</u>. Let A, B and C be linear transformations on a two-dimensional vector space over a field. Then

$$[C,B,A,B] - x[C,B] = 0$$

where $x = \text{traceAtraceB} + 2\det A + 2\det B - 2\det(A + B)$.

The Kato-Taussky-Wielandt commutator relation for the two-dimensional case follows from this corollary on putting $A = B$.

We note that we do not use the generalized form of the Kato-Taussky-Wielandt commutator relation (Theorem 2.14) in this thesis. We also remark that we have not succeeded in generalizing Theorem 2.14 when the dimension of V is greater than 2.

CHAPTER III

THE TWO-DIMENSIONAL BLOCKS

In Theorem 2.13 we gave a condition on A and B sufficient to guarantee that $\dim.A_{ii} = \dim.B_{ii} \leq k$ in (*). However this does not characterize matrices of the form (*) with $\dim.A_{ii} \leq k$, $i = 1,2,\ldots,t$. In this chapter we characterize those linear transformations A and B on V for which $\dim.A_{ii} = \dim B_{ii} \leq 2$ in (*) for $i = 1,2,\ldots,t$. The characterization is in terms of two-dimensional KTW commutator expressions.

Again we have linear transformations A and B on an n-dimensional vector space V over a field $\mathcal{F}$. $\mathcal{R}$ is the algebra generated by A and B over $\mathcal{F}$ and $\mathcal{J}$ is the radical of $\mathcal{R}$.

We have the following main theorem.

3.1 <u>Theorem</u>. Let $\mathcal{F}$ be algebraically closed. Then the following statements are equivalent.

(a) For each polynomial $P(x,y)$ there exist an integer $r = r(A,B,P(A,B))$ and distinct elements $h_1, h_2, \ldots, h_r \in \mathcal{F}$ so that

$$X_{\pi(1)} X_{\pi(2)} \cdots X_{\pi(r)} \in \mathcal{J}$$

for every permutation $\pi(1), \pi(2), \ldots, \pi(r)$ of $1,2,\ldots,r$, where

$$X_s = [A,P,P,P] - h_s[A,P],$$

$s = 1,2,\ldots,r$ and $P = P(A,B)$.

(b) There exists a basis of V with respect to which the matrices of A and B have the forms

$$(*)\quad \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1t} \\ 0 & A_{22} & \cdots & A_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & A_{tt} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ 0 & B_{22} & \cdots & B_{2t} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & B_{tt} \end{bmatrix},$$

respectively, where A_{ii} and B_{ii} are either 1x1 or 2x2 matrices which cannot be reduced further by a simultaneous similarity, $i = 1,2,\ldots,t$.

Proof. Assume (b) holds. Let A' and B' denote the matrices in (b) (since we do not refer to transposes in this theorem, the notation is unambiguous). The block diagonal of $P(A',B')$ has blocks $P(A_{ii},B_{ii})$, $i = 1,2,\ldots,t$. Let A_j, B_j be the 2x2 blocks on the diagonals of A', B', respectively, $j = 1,2,\ldots,q$. Let x_{1j} and x_{2j} be the characteristic roots of $P(A_j,B_j)$. Form $(x_{1j} - x_{2j})^2$, $j = 1,2,\ldots,q$ and let $h_1, h_2,\ldots,h_r$ be the distinct elements among these. If we let $P_j = P(A_j,B_j)$, then we have

$$[A_j,P_j,P_j,P_j] - (x_{1j} - x_{2j})^2[A_j,P_j] = 0.$$

If we form $X'_s = [A',P',P',P'] - h_s[A',P']$ for $s = 1,2,\ldots,r$, where $P' = P(A',B')$, we see that $X'_{\pi(1)}X'_{\pi(2)}\cdots X'_{\pi(r)}$ has zero blocks on and below the main diagonal, for each permutation $\pi(1),\pi(2),\ldots,\pi(r)$ of $1,2,\ldots,r$. This is also true of $Q(A',B')X'_{\pi(1)}X'_{\pi(2)}\cdots X'_{\pi(r)}$ for all polynomials $Q(x,y)$. Thus $Q(A,B)X_{\pi(1)}X_{\pi(2)}\cdots X_{\pi(r)}$ is nilpotent for all $Q(A,B)$ and hence $X_{\pi(1)}X_{\pi(2)}\cdots X_{\pi(r)} \in \mathcal{J}$ by Lemma 1.6. Hence (a) holds.

It is clear, from this part of the proof, why we must take a product of X_s's, instead of a single one.

Conversely, let (a) hold. By the usual argument, there exists a basis of V with respect to which the matrices of A and B have the forms (*), where A_{ii} and B_{ii} cannot be reduced any further by a simultaneous similarity,

$i = 1,2,\ldots,t$. Let A_{ii} and B_{ii} be $n_i x n_i$ matrices. We wish to show that $n_i \leqslant 2$.

So assume, for some i, that $n_i > 2$. Under this hypothesis we shall show that A_{ii} is actually a scalar matrix. As in Theorem 2.13 this means A_{ii} and B_{ii} can be reduced by a simultaneous similarity, which is a contradiction. Hence $n_i \leqslant 2$ for $i = 1,2,\ldots,t$.

We proceed to show that $n_i > 2$ implies A_{ii} is a scalar matrix. For simplicity, let $A_{ii} = C$, $B_{ii} = D$ and $P = P(C,D)$. Then, if $Y_s = [C,P,P,P] - h_s[C,P]$, we have $Y_{\pi(1)}Y_{\pi(2)}\cdots Y_{\pi(r)} = 0$ for all permutations $\pi(1),\pi(2),\ldots,\pi(r)$ of $1,2,\ldots,r$.

Since C and D cannot be reduced by the same similarity transformation, the algebra of all polynomials in C and D with coefficients in $\mathcal{F}$ is irreducible. Hence, by "Burnside's theorem" (Theorem 2.12), this algebra is the complete algebra of $n_i x n_i$ matrices with elements in $\mathcal{F}$. Thus, if X is any $n_i x n_i$ matrix with elements in $\mathcal{F}$, for some integer r (depending on X) there exist distinct elements $h_1,h_2,\ldots,h_r \in \mathcal{F}$ so that, if $X_s = [C,X,X,X] - h_s[C,X]$, then $X_{\pi(1)}X_{\pi(2)}\cdots X_{\pi(r)} = 0$ for all permutations of $1,2,\ldots,r$.

Without loss of generality, assume C is in Jordan canonical form.

$$C = \sum_{j=1}^{m}\oplus C_j \text{ where } C_j = \alpha_j I_j + E_j, \quad j = 1,2,\ldots,m,$$

where I_j is an identity matrix and E_j is a matrix of the same dimension as I_j, with 1's on the superdiagonal and zeros elsewhere, i.e. the C_j are Jordan blocks. We wish to show C is a scalar matrix. This we do in three stages.

<u>Stage 1</u>. Each Jordan block C_j has dimension $\leqslant 2$.

<u>Proof</u>. Suppose some block, C_1 say, has dimension $\geqslant 3$. Then

$$C_1 = \begin{bmatrix} \alpha_1 & 1 & 0 & \ldots & 0 \\ 0 & \alpha_1 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \ldots & \alpha_1 \end{bmatrix}.$$

We may ignore the diagonal, since we shall be taking commutators. Thus we consider E_1. Let

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

i.e. the leading 3x3 diagonal block of E_1. Since we may consider any $n_i x n_i$ matrix X in the expression $[C,X,X,X] - h_s[C,X]$, we shall now consider only those matrices X which have arbitrary 3x3 blocks Z in the place corresponding to the block E and zeros elsewhere. Thus we may restrict ourselves to 3x3 matrices. Let

$$Z = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

then

$$[E,Z] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix},$$

$$[E,Z,Z] = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}$$

and

$$[E,Z,Z,Z] = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

Hence

$$Z_s = [E,Z,Z,Z] - h_s[E,Z] = \begin{bmatrix} -h_s & 0 & 0 \\ 1 & 1 & 1-h_s \\ 0 & -1 & -(1-h_s) \end{bmatrix}$$

which is in block lower triangular form. We have $Z_{\pi(1)}Z_{\pi(2)}\cdots Z_{\pi(r)} = 0$, for each permutation of $1,2,\ldots,r$. Hence some h_s is zero, say $h_1 = 0$. Let

$$H_s = \begin{bmatrix} 1 & 1-h_s \\ -1 & -(1-h_s) \end{bmatrix},$$

then $H_uH_v = h_uH_v$, $1 \leqslant u,v \leqslant r$. Hence

$$\begin{aligned} H_rH_{r-1}\cdots H_1 &= H_1 \text{ if } r = 1, \\ &= \left(\prod_{s=2}^{r} h_s\right)H_1, \text{ otherwise.} \end{aligned}$$

Thus $H_rH_{r-1}\cdots H_1 \neq 0$ since $H_1 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ and $\prod_{s=2}^{r} h_s \neq 0$, $r \geqslant 2$. This is a contradiction. Hence $\dim.C_j \leqslant 2$, $j = 1,2,\ldots,m$.

<u>Stage 2</u>. C is a diagonal matrix.
<u>Proof</u>. Suppose for some j that $\dim.C_j = 2$. Let $C_j = \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}$. By simultaneous permutations of rows and columns, if necessary, we may assume C_j is not the leading block on the diagonal of C (since $\dim.C = n_i \geqslant 3$). Thus

$$C_{j-1} \oplus C_j = \begin{bmatrix} \beta & 0 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix} \text{ or } \begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \alpha & 1 \\ 0 & 0 & 0 & \alpha \end{bmatrix}$$

since $\dim . C_l \leq 2$, $l = 1,2,\ldots,m$. Let

$$G = \begin{bmatrix} \beta & 0 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix}.$$

If $C_{j-1} \oplus C_j$ is 4x4, it equals

$$\begin{bmatrix} \beta & 1 & 0 & 0 \\ 0 & & & \\ 0 & & G & \\ 0 & & & \end{bmatrix}$$

which is in block upper triangular form. We shall be multiplying C by matrices which have zeros everywhere except in the place corresponding to G, so again we restrict ourselves to 3x3 matrices. Let

$$Z = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & x & 0 \end{bmatrix}$$

where x is chosen in $\mathcal{F}$ so that $x \neq 0$ and $x \neq \beta - \alpha$. Then

$$[G,Z] = \begin{bmatrix} 0 & 0 & \beta-\alpha \\ 0 & x & 0 \\ 0 & 0 & -x \end{bmatrix}, \quad [G,Z,Z] = \begin{bmatrix} 0 & x(\beta-\alpha) & x-(\beta-\alpha) \\ 0 & 0 & 0 \\ 0 & -2x^2 & 0 \end{bmatrix}$$

and

$$[G,Z,Z,Z] = \begin{bmatrix} 0 & 3x^2-2x(\beta-\alpha) & (\beta-\alpha)-x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Hence

$$Z_s = [G,Z,Z,Z] - h_s[G,Z] = \begin{bmatrix} 0 & 3x^2-2(\beta-\alpha)x & \beta-\alpha-x-h_s(\beta-\alpha) \\ 0 & -h_s x & 0 \\ 0 & 0 & h_s x \end{bmatrix}.$$

Now $Z_{\pi(1)}Z_{\pi(2)}\cdots Z_{\pi(r)} = 0$ and thus $(\prod_{s=1}^{r} h_s)x^r = 0$. This means some h_s is zero, say $h_1 = 0$. Thus

$$Z_1 = \begin{bmatrix} 0 & 3x^2-2(\beta-\alpha)x & \beta-\alpha-x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The $(1,3)$ element of $Z_1Z_2\ldots Z_r$ is $(\beta-\alpha-x)cx^{r-1}$, where $c =$ $= 1$ if $r = 1$ and $c = h_2h_3\ldots h_r \neq 0$, otherwise. But this gives a contradiction, since $x \neq 0$ and $x \neq \beta-\alpha$. Hence $C = \mathrm{diag}(\alpha_1,\alpha_2,\ldots,\alpha_{n_i})$.

Stage 3. C is a scalar matrix.
Proof. Since we may put any three of $\alpha_1,\alpha_2,\ldots,\alpha_{n_i}$ in the first three places on the diagonal of C (by simultaneous row and column permutations), we may as well assume that $C = \mathrm{diag}(\alpha_1,\alpha_2,\alpha_3)$. Let

$$Z = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

We have, on letting $\alpha_i - \alpha_j = \alpha_{ij}$,

$$[C,Z] = \begin{bmatrix} 0 & 0 & \alpha_{13} \\ 0 & 0 & \alpha_{23} \\ \alpha_{31} & 0 & 0 \end{bmatrix}, \qquad [C,Z,Z] = \begin{bmatrix} 2\alpha_{13} & 0 & \alpha_{31} \\ \alpha_{23}+\alpha_{13} & 0 & \alpha_{32} \\ \alpha_{31} & 0 & 2\alpha_{31} \end{bmatrix},$$

and

$$[C,Z,Z,Z] = \begin{bmatrix} 0 & 0 & 5\alpha_{13} \\ \alpha_{12} & 0 & 2\alpha_{23}+3\alpha_{13} \\ 5\alpha_{31} & 0 & 0 \end{bmatrix}.$$

Hence

$$Z_s = [C,Z,Z,Z] - h_s[C,Z] = \begin{bmatrix} 0 & 0 & (5-h_s)\alpha_{13} \\ \alpha_{12} & 0 & (2-h_s)\alpha_{23}+3\alpha_{13} \\ (5-h_s)\alpha_{31} & 0 & 0 \end{bmatrix}.$$

The product of an even number of Z_s's has the form $\begin{bmatrix} * & 0 & 0 \\ * & 0 & * \\ 0 & 0 & * \end{bmatrix}$ and the product of an odd number has the form $\begin{bmatrix} 0 & 0 & * \\ * & 0 & * \\ * & 0 & 0 \end{bmatrix}$,

where $*$ denotes an element not necessarily equal to zero. The (1,3) element (if r is odd) or the (1,1) element (if r is even) of the product $Z_1Z_2\ldots Z_r$ gives $c\alpha_{13}^r = 0$, where $c = \prod_{s=1}^{r}(5-h_s)$. If no $h_s = 5$ we get $\alpha_1 = \alpha_3$. Otherwise h_1 (say) $= 5$. Thus

$$Z_1 = \begin{bmatrix} 0 & 0 & 0 \\ \alpha_{12} & 0 & 3\alpha_{12} \\ 0 & 0 & 0 \end{bmatrix}.$$

The (2,1) (if r is odd) or (2,3) (if r is even) element of $Z_1Z_2\ldots Z_r$ gives $c\alpha_{12}\alpha_{13}^{r-1} = 0$ where $c = 1$ if $r = 1$ and $c = \prod_{s=2}^{r}(5-h_s)$ otherwise. Hence $\alpha_1 = \alpha_2$ or $\alpha_1 = \alpha_3$. Now this last statement is true whether or not some $h_s = 5$. If $\alpha_1 = \alpha_3$, put α_2 in the (1,1) place (by a simultaneous row and column permutation) and repeat the argument. Thus $\alpha_2 = \alpha_1$ or $\alpha_2 = \alpha_3$ i.e. $\alpha_1 = \alpha_2 = \alpha_3$. Similarly $\alpha_1 = \alpha_2$ implies $\alpha_1 = \alpha_2 = \alpha_3$. Hence C is a scalar matrix i.e. A_{ii} is a scalar matrix.

Thus $n_i \leq 2$ and the result is proved. ∎

We make some remarks about the above theorem. Note, in the course of the proof, we did not use <u>all</u> the products $X_{\pi(1)}X_{\pi(2)}\cdots X_{\pi(r)}$. It seems likely that it should be possible to consider even fewer of these products than were used in the course of the proof; we do not have an example which contradicts this surmise.

It should be noted that the proof that (a) $\Longrightarrow$ (b) does not guarantee that the constants $h_1, h_2, \ldots, h_r$ are of the form $(x_1 - x_2)^2$ where x_1, x_2 are characteristic roots of $P(A,B)$. It is clear, for example, that if A and B have property Q, then the constants $h_1, h_2, \ldots, h_r$ may be quite arbitrary.

Finally, it seems likely that a similar theorem should characterize those matrices (*) with dim.$A_{ii} \leq k$, if we replace the two-dimensional KTW commutator expressions by k-dimensional ones. But we have made no progress with this problem.

CHAPTER IV

THE RELATION $f_k(A,B;\delta) = 0$, WHEN $k < n$

Example 2.6 indicates that the relation $f_k(A,B;\delta) = 0$, with $1 \leq k < n$ is not a "good" generalization of commutativity in that it does not help us to solve Problem 2.1. However, it seems reasonable to ask what the relation $f_k(A,B;\delta) = 0$ does imply. We might hope to prove that if $\delta = (\delta_1, \delta_2, \ldots, \delta_K)$ then $\delta_1, \delta_2, \ldots, \delta_K$ are elementary symmetric functions of some of $(\beta_r - \beta_s)^2$, $1 \leq r, s \leq n$, where β_r are the characteristic roots of B. When $k = 2$ we do get a result like this (cf. 4.6)but, in general, the result we obtain is not quite this strong (cf. Theorem 4.2).

We begin the investigation with a definition.

4.1 <u>Definition</u>. Let $f_k(X,Y;\delta)$ be a KTW commutator expression, with $\delta = (\delta_1, \delta_2, \ldots, \delta_K)$ where $K = k(k-1)/2$. Then

$$g_k(w;\delta) = w^{2K+1} - \delta_1 w^{2K-1} + \delta_2 w^{2K-3} - \ldots + (-1)^K \delta_K w.$$

The main theorem in this chapter is

4.2 <u>Theorem</u>. Let $\mathcal{L}$ be the algebra of linear transformation on an n-dimensional vector space V over a field $\mathcal{F}$ and let $A, B \in \mathcal{L}$. Suppose A and B generate $\mathcal{L}$ i.e. every linear transformation in $\mathcal{L}$ has the form $P(A,B)$ where $P(x,y)$ is a polynomial. If the characteristic of $\mathcal{F}$ does not divide n and if there exist $\delta_1, \delta_2, \ldots, \delta_K$ in $\mathcal{F}$ so that

$$f_k(A,B;\delta) = 0$$

where $1 \leq k < n$ and $\delta = (\delta_1, \delta_2, \ldots \delta_K)$, then the characteristic roots of B belong to the splitting field of $g_k(w;\delta)$ over $\mathcal{F}$.

We shall prove this theorem presently, but first we introduce some well-known facts from the Theory of Graphs which we shall employ in the course of the proof. A general reference for the graph theoretical material we use is Varga [15].

4.3 Definition. Let X be an nxn matrix with elements in a field $\mathfrak{F}$, then X is P-irreducible if it cannot be transformed by a permutation similarity to the form

$$\begin{bmatrix} Y & Z \\ 0 & W \end{bmatrix}$$

where Y and W are square matrices.

Note. The "P" in "P-irreducible" stands for "permutation". Our "P-irreducible" is the same as the "irreducible" or "indecomposable" of the Perron-Frobenius theory of non-negative matrices [cf. 15].

We associate a directed graph G(X) with an nxn matrix $X = (x_{ij})$ in the following way. G(X) consists of vertices numbered 1,2,...,n and there is an edge from i to j, i.e. $i \rightarrow j$, if and only if $x_{ij} \neq 0$.

The following well-known theorem is stated in Varga [15]. We include a proof since there does not seem to be one in the literature.

4.4 Theorem. Let X be an nxn matrix with elements in a field $\mathfrak{F}$. Then X is P-irreducible if and only if G(X) is strongly connected.

Proof. If P is a permutation matrix then $G(P^{-1}XP)$ is obtained by relabeling the vertices of G(X). Suppose G(X) is strongly connected. Then X is P-irreducible. For suppose otherwise; then there exists a permutation matrix P so that

$$
(\#) \quad P^{-1}XP = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1r} & x_{1,r+1} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2r} & x_{2,r+1} & \cdots & x_{2n} \\
\cdot & \cdot & & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & & \cdot \\
x_{r1} & x_{r2} & \cdots & x_{rr} & x_{r,r+1} & \cdots & x_{rn} \\
0 & 0 & \cdots & 0 & x_{r+1,r+1} & \cdots & x_{r+1,n} \\
\cdot & \cdot & & \cdot & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \cdot & & \cdot \\
0 & 0 & \cdots & 0 & x_{n,r+1} & \cdots & x_{nn}
\end{bmatrix}
$$

where $r < n$. Now $G(P^{-1}XP)$ is not strongly connected since there does not exist a path from $r+1$ to 1. This contradiction shows X is P-irreducible.

Conversely, let X be P-irreducible. Then $G(X)$ is strongly connected. Suppose it is not. Then there exists a vertex i to which at least one other vertex is not connected. By renumbering the vertices we get $i = 1$. Let $2,3,\ldots,r$ be the vertices which are connected to 1 by some path in $G(X)$ (again by renumbering). It is clear that $r \geq 2$, since we cannot have $x_{j1} = 0$, $j = 2,3,\ldots,n$, because this would contradict the fact that X is P-irreducible. It is also clear that $r < n$, since we have assumed $G(X)$ is not strongly connected. These renumberings of the vertices of $G(X)$ correspond to a permutation similarity $P^{-1}XP$ of X. We claim that $P^{-1}XP$ has the form (#) above, because there does not exist a path in $G(P^{-1}XP)$ from j to i where $r < j \leq n$ and $1 \leq i < r$. Since, if j can be connected to i, then j can be connected to 1, because i can be connected to 1. This is a contradiction of the fact that $2,3,\ldots,r$ are the only vertices that can be connected to 1 by a path in $G(X)$. Hence X is P-reducible (i.e. not P-irreducible). This contradicts the initial assumption. Hence $G(X)$ is strongly connected. ∎

4.5 <u>Proof of Theorem 4.2</u>. Firstly, we dispose of two easy cases. If $k = 1$, then $AB = BA$ and since A and B generate $\mathcal{L}$ this means $n = 1$ and the theorem is then obvious. The second easy case is when B has only one characteristic root β. Then trace$B = n\beta \in \mathcal{F}$. Hence $\beta \in \mathcal{F}$, since the characteristic of $\mathcal{F}$ does not divide n. So the result is trivial in this case also.

So assume that there exist $\mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_K$ in $\mathcal{F}$ so that $f_k(A,B;\mathfrak{d}) = 0$ where $1 < k < n$ and where B has at least two distinct characteristic roots. Replace A and B by matrices, again called A and B and extend $\mathcal{F}$ to a field $\mathcal{G}$ which contains the characteristic roots of B. Let U be a non-singular matrix with elements in $\mathcal{G}$ so that $N = U^{-1}BU$ is the Jordan canonical form of B with $N = \sum_{i=1}^{r} \oplus B_i$ where B_i is a direct sum of Jordan blocks all of which have the same characteristic root β_i and $\beta_i \neq \beta_j$ when $i \neq j$. Let $A_1 = U^{-1}AU$, then

$$f_k(A_1, N; \mathfrak{d}) = 0.$$

Let $A_1 = (A_{ij})$ be the partition of A_1 corresponding to that of $N = \sum_{i=1}^{r} \oplus B_i$. We shall prove the following statements.

1. If $A_{ij} \neq 0$, then $\beta_i - \beta_j$ satisfies the equation $g_k(w;\mathfrak{d}) = 0$.

2. $A_1 = (A_{ij})$ is P-irreducible as a block matrix.

If we assume Statements 1 and 2 we can complete the proof of the theorem in short order. For let $G(A_1)$ be the graph of A_1 considered as a block matrix, i.e. $i \longrightarrow j$ if and only if $A_{ij} \neq 0$. Then Statement 2 and a modification of Theorem 4.4 (for block matrices) imply that $G(A_1)$ is strongly connected. Thus, if β_i, β_j are distinct characteristic roots of B, there exists a sequence $i, i_1, i_2, \ldots$ $\ldots, i_u, j$ so that $\beta_i - \beta_{i_1}, \beta_{i_1} - \beta_{i_2}, \ldots, \beta_{i_u} - \beta_j$ satisfy $g_k(w;\mathfrak{d}) = 0$. Let $\mathcal{K}$ be the splitting field of $g_k(w;\mathfrak{d})$ over $\mathfrak{F}$. Thus

$$\beta_i - \beta_j = (\beta_i - \beta_{i_1}) + (\beta_{i_1} - \beta_{i_2}) + \ldots + (\beta_{i_u} - \beta_j) \varepsilon \mathcal{K}.$$

Let n_j be the multiplicity of β_j as a characteristic root of B, $j = 1, 2, \ldots, r$. Then

$$\sum_{j=1}^{r} n_j \beta_i - \sum_{j=1}^{r} n_j \beta_j \varepsilon \mathcal{K}.$$

Hence $n\beta_i - \text{trace} B \varepsilon \mathcal{K}$. Thus $\beta_i \varepsilon \mathcal{K}$, $i = 1, 2, \ldots, r$, since the characteristic of $\mathcal{K}$ does not divide n.

It remains to prove Statements 1 and 2 to complete the proof of the theorem.

<u>Proof of Statement 1</u>. We employ the relation $f_k(A_1, N;\mathfrak{d}) =$ $= 0$. Suppose $A_{ij} \neq 0$. Let a_{st} be the "first" non-zero element of A_{ij} in the following sense: if the lower left-hand corner element of A_{ij} is non-zero, let this be a_{st}; otherwise let a_{st} be a non-zero element of A_{ij} so that $a_{uv} = 0$ if $u \geqslant s$ and $v \leqslant t$ and $(u,v) \neq (s,t)$ where, of course, we only consider those elements a_{uv} in A_{ij}. Thus

$$A_{ij} = \begin{bmatrix} & * & & & * \\ 0 & \ldots & 0 & a_{st} & \\ 0 & \ldots & 0 & 0 & \\ \cdot & & \cdot & \cdot & * \\ \cdot & & \cdot & \cdot & \\ 0 & \ldots & 0 & 0 & \end{bmatrix}$$

where * denotes elements which make up the rest of A_{ij}.

If a_{st} is the (a,b) element of A_{ij}, we shall calculate the (a,b) element of the (i,j) block of $f_k(A_1,N;\eth)$. To simplify calculations assume that B_j has characteristic root 0 and B_i has characteristic root $\beta_{ij} = \beta_i - \beta_j$ (Subtract $\beta_j I$ from N. Since we take commutators, this operation does not affect the end result of the calculations.). The matrix $f_k(A_1,N;\eth)$ is a linear combination of matrices of the type $[A_1,N,N,\ldots,N]$. The (i,j) block of $[A_1,N]$ is $A_{ij}B_j - B_iA_{ij}$. The (i,j) block of $[A_1,N,N,\ldots,N]$ only involves A_{ij}, B_i and B_j; it consists of a linear combination of matrices of the type $B_i^cA_{ij}B_j^d$, where $c + d$ is the number of times the commutator operation is applied in $[A_1,N,N,\ldots,N]$. The (a,b) element of $B_i^cA_{ij}B_j^d$ is obtained by multiplying the b^{th} column of $A_{ij}B_j^d$ by the a^{th} row of B_i^c. Those elements in the b^{th} column of $A_{ij}B_j^d$ from the a^{th} element down are all that matter here. But these elements are zeros, except when $d = 0$, since B_j has zeros on and below the main diagonal. Thus the (a,b) element of $[A_1,N,N,\ldots,N]$, where the commutator operation is performed m times, is $(-1)^m\beta_{ij}^m a_{st}$. Since the monomials in $g_k(w;\eth)$ are of odd degree, the equation $f_k(A_1,N;\eth) = 0$ gives $g_k(\beta_{ij};\eth)a_{st} = 0$ and hence $g_k(\beta_{ij};\eth) = 0$ since $a_{st} \neq 0$. Thus we have shown that if $A_{ij} \neq 0$, then $\beta_i - \beta_j$ satisfies the equation $g_k(w;\eth) = 0$.

<u>Proof of Statement 2</u>. We now show that $A_1 = (A_{ij})$ is P-irreducible as a block matrix. For suppose there exists a block permutation matrix P, partitioned conformally with $A_1 = (A_{ij})$, so that $P^{-1}A_1P$ has the form

$$\begin{bmatrix} A_{11} & \cdots & A_{1m} & A_{1,m+1} & \cdots & A_{1r} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ A_{m1} & \cdots & A_{mm} & A_{m,m+1} & \cdots & A_{mr} \\ 0 & \cdots & 0 & A_{m+1,m+1} & \cdots & A_{m+1,r} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ 0 & \cdots & 0 & A_{r,m+1} & \cdots & A_{rr} \end{bmatrix}$$

where $m < r$, then A_1 and N may be reduced by a simultaneous similarity, since the block permutation matrix P simply permutes the blocks on the diagonal of N. Thus the algebra of matrices of the form $P(A_1,N)$ is reducible, where $P(x,y)$ is a polynomial in the non-commuting variables x and y with coefficients in $\mathcal{G}$. But the matrices A and B generate the complete algebra of nxn matrices with coefficients in $\mathfrak{F}$. Hence A_1 and N generate the complete algebra of nxn matrices with coefficients in $\mathcal{G}$ which is irreducible. This contradicts the assumption that $A_1 = (A_{ij})$ is not P-irreducible as a block matrix. Thus A_1 is P-irreducible as a block matrix. $\blacksquare$

4.6 <u>Corollary</u>. Let A and B satisfy the conditions of Theorem 4.2 with $k = 2$ and let B have at least two distinct characteristic roots. Then there exists an ordering $\beta_1, \beta_2, \ldots, \beta_r$ of the distinct characteristic roots of B so that $\beta_1 - \beta_2 = \beta_2 - \beta_3 = \cdots = \beta_{r-1} - \beta_r$ satisfies $w^2 - \mathfrak{d}_1 = 0$.

<u>Proof</u>. Consider the matrix $A_1 = (A_{ij})$ defined in 4.5. We have

$$f_2(A_1,N;\mathfrak{d}) = 0 \quad \text{where } \mathfrak{d} = (\mathfrak{d}_1).$$

We claim that A_1 cannot have more than two off-diagonal blocks in each row or column which are non-zero. Suppose, for example, that A_{ij}, A_{ik} and $A_{i\ell}$ are non-zero off-diagonal blocks, where j, k and ℓ are distinct. Then $\beta_i - \beta_j$, $\beta_i - \beta_k$ and $\beta_i - \beta_\ell$ all satisfy $x^2 - \delta_1 = 0$, by Statement 1 of 4.5. Hence at least two of β_j, β_k, β_ℓ must be equal, and this contradicts the fact that the β's are distinct. We also note that if A_{ij} and A_{ik} (resp. A_{ji} and A_{ki}) are non-zero off-diagonal blocks, where $j \neq k$, then an argument similar to the one just given shows that if A_{si} (resp. A_{is}) is a non-zero off-diagonal block then $s = j$ or $s = k$.

Let $G(A_1)$ be the graph of A_1 considered as a block matrix. We shall write $i \sim j$ if $i \longrightarrow j$ or $j \longrightarrow i$. The discussion of the last paragraph shows that if $i \sim j$ and $i \sim k$, where i, j, k are distinct, then $i \sim l$ implies that l must be either j or k. We claim that by renumbering the vertices of $G(A_1)$ we get the subgraph

$$\underset{1}{\bullet} \text{———} \underset{2}{\bullet} \text{———} \underset{3}{\bullet} \;\cdots\; \underset{r-1}{\bullet} \text{———} \underset{r}{\bullet}$$

where $i \sim i+1$, for $i = 1, 2, \ldots, r-1$. For let

$$\mu = \underset{1}{\bullet} \text{———} \underset{2}{\bullet} \text{———} \underset{3}{\bullet} \;\cdots\; \underset{s-1}{\bullet} \text{———} \underset{s}{\bullet}$$

be a maximal "path" in $G(A_1)$ (on renumbering vertices) where $i \sim i+1$, for $i = 1, 2, \ldots, s-1$ and suppose $s \neq r$. If j is a vertex $\notin \mu$, then neither $j \sim 1$ nor $j \sim s$ can hold, since μ is maximal. Since $G(A_1)$ is strongly connected there exists an internal vertex $i \,\varepsilon\, \mu$ and a vertex $j \notin \mu$ so that $i \sim j$. But $i \sim i+1$ and $i \sim i-1$, and since $j \neq i-1$ and $j \neq i+1$, this is a contradiction. Thus $G(A_1)$ contains the required subgraph. Hence, on renumbering the distinct characteristic roots β_i of B, we see that $\beta_1 - \beta_2$, $\beta_2 - \beta_3$,

$\ldots, \beta_{r-1} - \beta_r$ satisfy $w^2 - b_1 = 0$. Now we must have

$$\beta_i - \beta_{i+1} = \beta_{i+1} - \beta_{i+2}, \quad i = 1,2,\ldots,$$

since $\beta_i - \beta_{i+1} = \beta_{i+2} - \beta_{i+1}$ implies that $\beta_i = \beta_{i+2}$. This is impossible, since the β_i's are distinct. ∎

4.7 Remarks on Theorem 4.2. Example 2.6 illustrates Theorem 4.2 in a trivial fashion. Later on in this chapter we shall give a non-trivial example (Example 4.12) which illustrates Theorem 4.2, Corollary 4.6 and a theorem we have yet to state (Theorem 4.9). We note that Theorem 4.2 says that if β is a characteristic root of B, then it satisfies an equation of degree at most $k(k-1)!$ over $\mathfrak{F}$. At first sight this statement does not look too promising, but if k is "small" compared to n it says something about the reducibility of the characteristic polynomial of B over $\mathfrak{F}$. We shall return to this fact in a moment when we discuss a generalization of the so-called "L-property". We note one more fact about Theorem 4.2: it is, that the condition on the characteristic of $\mathfrak{F}$ is necessary. We give a counterexample to the theorem later (Example 4.13) where this condition is not satisfied.

Assume for the moment that the characteristic roots of A and B belong to $\mathfrak{F}$. We recall that A and B have property P if there exist orderings $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n$ of the characteristic roots of A and B respectively so that $P(A,B)$ has characteristic roots $P(\alpha_i, \beta_i)$, $i = 1,2,\ldots,n$, for each polynomial $P(z,w)$. Now a weak form of property P is property L (a term due to Kac) which demands that property P only hold for linear polynomials $P(z,w) = xz + yw$, where $x,y \in \mathfrak{F}$ [Motzkin and Taussky, 16]. Property L then says that the characteristic polynomial $p(x,y,z)$ of $xA + yB$ splits into linear factors, i.e.

$$p(x,y,z) = \prod_{i=1}^{n}(x\alpha_i + y\beta_i - z)$$

where α_i and $\beta_i \varepsilon \mathcal{F}$, $i = 1,2,\ldots,n$. We shall use Theorem 4.2 to prove a result about a pair of linear transformations A and B which implies they have a property very much like that of property L (instead of only linear factors in $p(x,y,z)$ we get both linear and quadratic factors). To prove the result we need a preliminary lemma. We denote by $\mathcal{F}[x,y]$ the integral domain of polynomials $P(x,y)$ and by $\mathcal{F}(x,y)$ the quotient field of this integral domain. (We assume, of course, that the indeterminates commute, in contradiction to the convention on page 4).

4.8 Lemma. Let $p(x,y,z)$ be a homogeneous polynomial in x,y and z with coefficients in a field $\mathcal{F}$. Suppose

$$p(x,y,z) = \sum_{i=1}^{r} p_i^{k_i}$$

where each p_i is an irreducible polynomial in z over $\mathcal{F}(x,y)$, then each p_i is a homogeneous polynomial in x,y and z with coefficients in $\mathcal{F}$.

Proof. $\mathcal{F}[x,y]$ is a Gaussian domain (unique factorization domain) [cf. Jacobson 14, p.126]. Since a Gaussian domain is integrally closed [ibid. p.184], the coefficients of the powers of z in p_i must be polynomials in x and y.

Suppose p_i is not homogeneous in x,y and z. Define $M(q)$ (resp. $m(q)$) to be the maximum (resp. minimum) degree of the monomials in a polynomial q. Now $M(p_i) > m(p_i)$ and M (resp. m) has the property that $M(qr) = M(q) + M(r)$ ($m(qr) = m(q) + m(r)$), for polynomials q and r. Hence $M(p) > m(p)$ where $p = p(x,y,z)$. But this is false. Hence $M(p_i) = m(p_i)$ and thus p_i is homogeneous.

4.9 Theorem. Let $\mathcal{L}$ be the algebra of linear transformation on an n-dimensional vector space V over an infinite field $\mathfrak{F}$ whose characteristic does not divide n. Assume A and B $\varepsilon \mathcal{L}$ are such that A and B generate $\mathcal{L}$ and suppose that for each x,y $\varepsilon \mathfrak{F}$ there exists $\mathfrak{d}_1 \varepsilon \mathfrak{F}$ so that

$$f_2(A, xA + yB; \mathfrak{d}) = 0$$

where $\mathfrak{d} = (\mathfrak{d}_1)$. Then the characteristic polynomial p(x,y,z) of xA + yB splits into linear and quadratic homogeneous factors in x,y and z with coefficients in $\mathfrak{F}$.

Proof. We may assume that $n \geqslant 3$, since the result is trivial otherwise. Replace A and B by matrices. Let xA + yB = X. Then

$$f_2(A,X;\mathfrak{d}) = [A,X,X,X] - \mathfrak{d}_1 [A,X].$$

If $[A,X] = 0$ then AB = BA and this implies n = 1. So the result is trivial in this case. So assume $[A,X] \neq 0$. Then the equation $f_2(A,X;\mathfrak{d}) = 0$ says that $\mathfrak{d}_1$ is a rational function of x and y. Replace x and y by two algebraically independent indeterminates, again called x and y, respectively. Then the equation $f_2(A,X;\mathfrak{d}) = 0$ still holds, by Theorem 2.7, since $\mathfrak{F}$ is infinite. A and X clearly generate the algebra of nxn matrices with elements in $\mathfrak{F}(x,y)$, so we may apply Theorem 4.2. Now $g_2(w;\mathfrak{d}) = w^3 - \mathfrak{d}_1 w$. So each characteristic root of X = xA + yB satisfies an equation of degree at most 2 over $\mathfrak{F}(x,y)$. The theorem then follows on applying Lemma 4.8.

4.10 Corollary. Let A and B be linear transformations on an n-dimensional vector space V over an algebraically closed field $\mathfrak{F}$ whose characteristic is either zero or greater than n. If, for each x,y $\varepsilon \mathfrak{F}$, there exists $\mathfrak{d}_1 \varepsilon \mathfrak{F}$ so that

$$f_2(A, xA + yB; \mathfrak{d}) \in \mathcal{J}$$

where $\mathcal{J}$ is the radical of the algebra generated by A and B over $\mathcal{F}$, then the characteristic polynomial $p(x,y,z)$ of $xA + yB$ splits into linear and quadratic homogeneous factors in x,y and z with coefficients in $\mathcal{F}$.

Proof. Apply the usual composition series argument to V to get matrices for A and B in block upper triangular form. "Burnside's theorem" (Theorem 2.12) shows that simultaneously irreducible blocks A_{ii} and B_{ii} (cf. (*) on p.18) generate a complete matrix algebra, and the condition on the characteristic guarantees that it does not divide the dimension of any diagonal block. The result follows immediately on applying Theorem 4.9 to the diagonal blocks.

The above results do not guarantee that $p(x,y,z)$ actually has a linear factor, but part of the next result does.

4.11 Theorem. Let A and B be nxn matrices with elements in a field $\mathcal{F}$ whose characteristic is either zero or greater than n. Suppose A and B generate the complete algebra of nxn matrices and let x,y be algebraically independent indeterminates. If there exists $\mathfrak{d}_1 \in \mathcal{F}(x,y)$ so that

$$f_2(A, xA + yB; \mathfrak{d}) = 0$$

where $\mathfrak{d} = (\mathfrak{d}_1)$, then the characteristic polynomial $p(x,y,z)$ of $xA + yB$ splits into linear and quadratic homogeneous factors in x,y and z with coefficients in $\mathcal{F}$. Moreover if $xA + yB$ has an odd number of distinct characteristic roots then $p(x,y,z)$ has at least one linear factor.

Proof. The first part of this theorem is almost a repetition of Theorem 4.9 and is proved in the same manner (Note: the field $\mathcal{F}$ need not be infinite here). So assume

$xA + yB$ has r distinct characteristic roots, where r is odd. If $r = 1$, the result is trivial. So assume $r \geqslant 3$. By Corollary 4.6 there exists an ordering $z_1, z_2, \ldots, z_r$ of the distinct characteristic roots of $xA + yB$ so that $z_1 - z_2 = z_2 - z_3 = \ldots = z_{r-1} - z_r$ satisfies $w^2 - \delta_1 = 0$. Now the irreducible factors of $p(x,y,z)$ are separable, since the characteristic of $\mathfrak{F}$ is either zero or greater than n. Hence

$$\sum_{i=1}^{r} z_i \in \mathfrak{F}(x,y).$$

Now

$$z_i - z_{i+1} = \sqrt{\delta_1}, \quad i = 1,2,\ldots,r-1.$$

Hence

$$\sum_{i=1}^{r} z_i = r\{z_r + ((r-1)/2)\sqrt{\delta_1}\}.$$

Thus $z_r + ((r-1)/2)\sqrt{\delta_1} \in \mathfrak{F}(x,y)$. Now $(r-1)/2 = s$ is an integer, since r is odd and $z_r + ((r-1)/2)\sqrt{\delta_1} = z_{r-s}$. Hence $z_{r-s} \in \mathfrak{F}(x,y)$. By Lemma 4.8 $z_{r-s} = x\alpha + y\beta$ where $\alpha, \beta \in \mathfrak{F}$. ∎

Finally we give two examples to illustrate the results of this chapter.

4.12 <u>Example</u>. This example illustrates Theorems 4.9 and 4.11 (and <u>a fortiori</u> Theorem 4.2 and Corollary 4.6). Let $\mathfrak{F}$ be a field whose characteristic $\neq$ 2 or 3. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

If E_{ij} is the 3x3 matrix with 1 in the (i,j) place and zeros elsewhere, then

$$E_{33} = (1/2)(A^2 - 3A + 2I), \quad E_{22} = -A^2 + 4A - 3I,$$
$$E_{11} = I - E_{22} - E_{33}, \quad E_{12} = E_{11}(B - I),$$
$$E_{32} = E_{33}(B - I), \quad E_{21} = (B - I)E_{11},$$
$$E_{23} = (B - I)E_{33}, \quad E_{13} = E_{12}E_{23}, \quad E_{31} = E_{32}E_{21}.$$

Hence A and B generate the complete algebra of 3x3 matrices with elements in $\mathcal{F}$. Let x,y be two algebraically independent indeterminates over $\mathcal{F}$ and let $X = xA + yB$. Clearly A and X generate the algebra of all 3x3 matrices with elements in $\mathcal{F}(x,y)$. We have

$$X = xA + yB = \begin{bmatrix} x+y & y & 0 \\ y & 2x+y & y \\ 0 & y & 3x+y \end{bmatrix},$$

$$[A,X] = \begin{bmatrix} 0 & -y & 0 \\ y & 0 & -y \\ 0 & y & 0 \end{bmatrix}, \qquad [A,X,X] = \begin{bmatrix} -2y^2 & -xy & 0 \\ -xy & 0 & -xy \\ 0 & -xy & 2y^2 \end{bmatrix}$$

and

$$[A,X,X,X] = \begin{bmatrix} 0 & -2y^3-x^2y & 0 \\ 2y^3+x^2y & 0 & -2y^3-x^2y \\ 0 & 2y^3+x^2y & 0 \end{bmatrix}$$

$$= (x^2 + 2y^2)[A,X].$$

Hence

$$f_2(A, xA + yB; (x^2+2y^2)) = 0.$$

$$p(x,y,z) = \det(X - zI) = \begin{vmatrix} x+y-z & y & 0 \\ y & 2x+y-z & y \\ 0 & y & 3x+y-z \end{vmatrix}.$$

Expand the determinant by the second row to give

$$p(x,y,z) = (2x + y - z)(z^2 - (4x + 2y)z + 3x^2 + 4xy - y^2)$$

(cf. Theorem 4.9). Hence the characteristic roots of X are

$$z_2 = 2x + y,$$

$$z_1 = 2x + y + \sqrt{x^2 + 2y^2}, \quad z_3 = 2x + y - \sqrt{x^2 + 2y^2}.$$

Clearly

$$z_1 - z_2 = z_2 - z_3 = \sqrt{x^2 + 2y^2}$$

(cf. Corollary 4.6). We see that

$$z_1, z_2, z_3 \in \mathfrak{F}(x, y, \sqrt{x^2 + 2y^2})$$

(cf. Theorem 4.2). Since $X = xA + yB$ has an odd number of distinct characteristic roots, $p(x,y,z)$ has a linear factor (cf. Theorem 4.11).

4.13 Example. The example we give here is a counter-example to Theorems 4.2 and 4.9 when the condition on the characteristic of $\mathfrak{F}$ is not satisfied. Let $\mathfrak{F}$ have characteristic 3 and let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Then

$$E_{13} = A^2, \quad E_{11} = E_{13}B^2 - E_{13}B,$$

$$E_{12} = E_{13}B^2 + E_{13}B, \quad E_{23} = A - E_{12},$$

$$E_{33} = BE_{13} - E_{23}, \quad E_{22} = I - E_{11} - E_{33},$$

$$E_{32} = BE_{12} - E_{22}, \quad E_{21} = E_{11} - B^2E_{11} - BE_{11},$$

$$E_{31} = B^2E_{11} - BE_{11} - E_{11}.$$

Hence A and B generate the complete algebra of 3x3 matrices with elements in $\mathfrak{F}$. As usual let x,y be indeterminates and $X = xA + yB$. Then A and X generate the complete algebra of 3x3 matrices with elements in $\mathfrak{F}(x,y)$. We have

$$X = \begin{bmatrix} 0 & x-y & -y \\ y & 0 & x+y \\ y & -y & 0 \end{bmatrix}, \quad [A,X] = \begin{bmatrix} y & 0 & -y \\ y & y & 0 \\ 0 & -y & y \end{bmatrix},$$

and

$$[A,X,X] = \begin{bmatrix} -xy & 0 & 0 \\ 0 & -xy & 0 \\ 0 & 0 & -xy \end{bmatrix}.$$

Hence

$$[A,X,X,X] = 0 = 0[A,X].$$

Therefore

$$f_2(A, xA + yB; \mathfrak{d}) = 0, \text{ where } \mathfrak{d} = (0).$$

Now

$$g_2(w;\mathfrak{d}) = w^3$$

and

$$p(x,y,z) = \det(xA + yB - zI) = \begin{vmatrix} -z & x-y & -y \\ y & -z & x+y \\ y & -y & -z \end{vmatrix}$$

$$= x^2y - z^3.$$

Now

$$x^2y - z^3 = ((x^2y)^{1/3} - z)^3$$

and hence the characteristic roots of $xA + yB$ do not belong to the splitting field of $g_2(w;\mathfrak{d})$ over $\mathfrak{F}(x,y)$ (cf. Theorem 4.2). Also $x^2y - z^3$ has no linear factors over

$\mathcal{F}(x,y)$ (cf. Theorem 4.9).

BIBLIOGRAPHY

1 N. H. McCoy, On the characteristic roots of matric polynomials, Amer. Math. Soc. Bull., 42(1936), 592-600.

2 I. N. Herstein, Theory of Rings, University of Chicago Mathematics Lecture Notes, Chicago 1961.

3 N. H. McCoy, The Theory of Rings, Macmillan, New York, 1964.

4 N. Jacobson, Lectures in Abstract Algebra, Volume II, Linear Algebra, Van Nostrand, Princeton, 1953.

5 J. Williamson, The simultaneous reduction of two matrices to triangle form, Amer. J. Math., 57(1935), 281-293.

6 M. P. Drazin, J. W. Dungey and K. W. Gruenberg, Some theorems on commutative matrices, J. London Math. Soc., 26(1951), 221-228.

7 M. P. Drazin, Some generalizations of matrix commutativity, Proc. London Math. Soc., 1(1951), 222-231.

8 N. Jacobson, Rational methods in the theory of Lie algebras, Ann. of Math., 4(1935), 875-881.

9 K. Shoda, Einige Sätze über Matrizen, Jap. J. Math., 13(1936), 361-365.

10 A. Albert and B. Muckenhoupt, On matrices of trace zero, Michigan Math. J., 4(1957), 1-3.

11 T. Kato and O. Taussky, Commutators of A and A*, J. Wash. Acad. Sci., 46(1956), 38-40.

12 O. Taussky and H. Wielandt, Linear relations between higher additive commutators, Amer. Math. Soc. Proc., 13(1962), 732-735.

13 N. A. Khan, Linear relations between higher matrix commutators, Canadian J. Math., 16(1964), 315-320.

14 N. Jacobson, Lectures in Abstract Algebra, Volume I, Basic Concepts, Van Nostrand, Princeton, 1951.

15 R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.

16 T. S. Motzkin and O. Taussky, Pairs of matrices with property L, Trans. Amer. Math. Soc., 73(1952), 108-114.