Georgia Southern University
Digital Commons@Georgia Southern

Legacy ETDs

Spring 1972

The Jordan Canonical Form

Richard A. Freeman

Follow this and additional works at: https://digitalcommons.georgiasouthern.edu/etd_legacy

Part of the Mathematics Commons

Recommended Citation
Freeman, Richard A., "The Jordan Canonical Form" (1972). *Legacy ETDs*. 1031.
https://digitalcommons.georgiasouthern.edu/etd_legacy/1031

This thesis (open access) is brought to you for free and open access by Digital Commons@Georgia Southern. It has been accepted for inclusion in Legacy ETDs by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.

724

THE JORDAN CANONICAL FORM

Richard A. Freeman

QA
263
F7

Georgia Southern College
Library

THE JORDAN CANONICAL FORM

by

Richard A. Freeman

A Thesis Submitted to the Faculty of the
Mathematics Department of Georgia Southern
College in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Mathematics

May, 1972

Approved by:

David R. Stone
Major Professor

C. Koziol
Member of Thesis Committee

C. L. Christmas
Member of Thesis Committee

Jack N. Averitt
Dean of the Graduate School

263
F7

ACKNOWLEDGMENT

I would like to express my sincere appreciation to Dr. David R. Stone for his generous advice and his helpful criticism which helped to formulate this paper.

RAF

TABLE OF CONTENTS

CHAPTER I 1

CHAPTER II 5

CHAPTER III 22

CHAPTER IV 34

BIBLIOGRAPHY 48

Chapter I

The study of matrices constitute a major part of any study of linear algebra. The fact that linear transformations determine matrices and matrices determine linear transformations (which will be brought out in more detail later) requires a thorough look at the relationships between matrices and linear maps.

The following are statements of the notation used in this and following chapters. Unless otherwise stated, all vector spaces will be denoted by U and V, and vectors by u, v, w, and z. F will denote a field and a, b, and c will denote field elements; F[x] will denote the ring of polynomials in one indeterminant with coefficients from F. Also, unless otherwise stated, polynomials in F[x] will be denoted by P, Q, and R and linear transformations of vector spaces by f, g and h. $\langle u_1, \ldots, u_n \rangle$ denotes the space generated by the vectors $u_1, \ldots, u_n$. All vector spaces mentioned are finite dimensional.

(1.1) Definition. Let f be a linear transformation from V onto U. If the vectors $v_1, \ldots, v_n$ are a basis of V and the vectors $u_1, \ldots, u_m$ are a basis of U, then the matrix determined by f relative to these bases is the matrix $A=(a_{ij})$ where $f(v_j)=\sum_{i=1}^{m} a_{ij}u_i$. Also an m by n matrix $B=(b_{ij})$ determines a transformation g defined by $g(v_j)=\sum_{i=1}^{m} b_{ij}u_i$, as long as the elements of the matrix B are from the field F.

(1.2 Definition. The set $L(V,V)$ is the set of all linear transformations f of V into V. Such an f is called an <u>endomorphism</u> of V.

If $f \in L(V,V)$ and A is the matrix determined by f relative to the basis $\{v_1, \ldots, v_n\}$ and B the matrix determined by f relative to another basis $\{u_1, \ldots, u_n\}$ of $V(A=(a_{ij})$ where $f(v_j)=\sum_{i=1}^{n} a_{ij}v_i$ and $B=(b_{ij})$ where $f(u_j)=\sum_{i=1}^{n} b_{ij}u_i)$, then there exists a matrix C such that $A=CBC^{-1}$. That is, A and B are called similar matrices. (Note: similarity of matrices is an equivalence relation.)

Many times, especially in application, it is necessary to find the simplest possible form of a matrix to work with. So, given an n by n matrix A with elements a field F, then A determines an endomorphism f of a vector V over F. It suffices to find an n by n matrix B similar to A whose form is as simple as possible. That is, it suffices to find a basis of V such that the matrix of f relative to this basis is in as simple a form as possible.

(1.3) Definition. An n by n matrix whose elements not on the main diagonal are zero is called a <u>diagonal</u> matrix.

Naturally, the simplest such matrix possible is a diagonal matrix. There is a relatively large class of endomorphisms that can be represented by diagonal matrices.

(1.4) Definition. Let V be a vector space over a field F and let $f \in L(V,V)$. If $v \in V$ such that $f(v) = av$, $a \in F$, then v is said to be an <u>eigenvector</u> of f and a is the <u>eigenvalue</u> corresponding to v.

(1.6) Theorem. Let f be in L(V,V). Then f can be represented by a diagonal matrix if and only if there exists a basis of V consisting of eigenvectors of f.

Proof: Suppose there exists a basis of eigenvectors $v_1, \ldots, v_n$ of f and the set $a_1, \ldots, a_n$ are the corresponding eigenvalues. Then $f(v_i) = a_i v_i$ so the matrix of f relative to this basis has the form:

$$\begin{bmatrix} a_1 & 0 & . & . & 0 \\ 0 & a_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & a_n \end{bmatrix}$$

Conversely, if f is represented by a diagonal matrix, then the vectors in that basis are eigenvectors.

Usually, one is not given a linear transformation directly. A matrix A is usually given representing f with respect to some unspecified basis. In this case, Theorem 1.6 should read: A matrix A is similar to a diagonal matrix if and only if there exists a basis of eigenvectors of the endomorphism determined by A.

Of course, there are many endomorphisms that cannot be represented by a diagonal matrix. In this case, given a matrix A it is

necessary to find a matrix similar to A that is in the simplest possible form. Chapters II and III of this paper represent two different derivations of this form, the first being based on a method used by J. Walter Nef [10, pp. 281-95] and the second on a method used by Charles W. Curtis [3, pp. 189-96]. Chapter IV of this paper dwells entirely on uses, applications, and examples of this form.

Chapter II

The development in this chapter of the Jordan canonical form is based on the development of this form by Walter Nef [10, pp. 281-95]. In this chapter, Definitions 2.1, 2.2, 2.4, as well as Theorem 2.3 are given as prelimary information.

(2.1) Definition. An <u>eigenvector</u> of an endomofphism f of a vector space V is any non-zero vector v in V which is mapped to a scalar multiple of itself. If $f(v) = av$, then a is referred to as the <u>eigenvalue</u> of f corresponding to the eigenvector v.

(2.2) Definition. If A is the matrix representation of an endomorphism f, then $\det(A - xI)$ is called the <u>characteristic polynomial</u> of f.

(2.3) Theorem. If A is the matrix representation of an endomorphism f, then a is an eigenvalue of f if and only if $\det(A - aI) = 0$.

(2.4) Definition. Let U and V be vector spaces and let f be a linear map of V into U. Then the <u>rank</u> of f is the dimension of the image of f.

Let V be a vector space and let $f \in L(V,V)$. If $P = \sum_{i=1}^{n} a_i x^i$ is an element of $F[x]$, note that $P_f = \sum a_i f^i$ is again an endomorphism on V and P_f commutes with f. Also, note that if P and Q are polynomials, then $(PQ)_f = P_f Q_f$ and $(P + Q)_f = P_f + Q_f$.

(2.5) Definition. A subspace H of V is said to be <u>f-invariant</u> if $f(H) \subseteq H$.

If H is f-invariant, then since $f^k(H) \subseteq H$ and $a_k f^k(H) \subseteq H$ for all $k > 0$, then H is also P_f-invariant for any polynomial P in $F[x]$.

Since the dimension of the vector space $L(V,V)$ is n^2 (where the dimension of V is n) and the endomorphisms $f, f^2, \ldots, f^k$ (where $k > n^2$) are dependent, then there is a polynomial P such that P_f is zero.

(2.6) Definition. The <u>minimal polynomial</u> of an endomorphism f is the unique monic polynomial m of the least positive degree such that $m_f = 0$.

If $u \in V$ and m is the minimal polynomial of f, then $m_f(u) = 0$, so there exists at least one polynomial P of positive degree such that $P_f(u) = 0$.

(2.7) Definition. The <u>f-annihilator</u> of a vector u in V is the unique monic polynomial h of the least positive degree such that $h_f(u) = 0$.

Notice that by the division alogorithm, if $k_f(u) = 0$, then h divides k. In particular, h divides the minimal polynomial m. Notice that the degree of h is greater than or equal to one.

(2.8) Theorem. If s is the degree of the f-annihilator h of u in V and $t \geq s$, then the vectors $u, f(u), \ldots, f^{s-1}(u)$ are linearly independent,

and the vectors $u, f(u), \ldots, f^{s-1}(v), f^{t}(u)$ are linearly dependent.

Proof: Suppose that $u, f(u), \ldots, f^{s-1}(u)$ are linearly dependent. Then there exists a polynomial k of degree less than s such that $k_f(u) = 0$ which contradicts the fact that h is the f-annihilator.

If $t = s$, then the theorem is true from the fact that $h_f(u)$ is nothing more than a linear combination of $u, f(u), \ldots, f^{s}(u)$.

Suppose the theorem is true for $t > s$, that is, $f^{t}(u)$ is an element of $\langle u, \ldots, f^{s-1}(u) \rangle$. Therefore, $f^{t+1}(u)$ is an element of $\langle f(u), \ldots, f^{s}(u) \rangle$, but $f^{s}(u)$ is an element of $\langle u, \ldots, f^{s-1}(u) \rangle$ so $f^{t+1}(u)$ is an element of $\langle u, \ldots, f^{s-1}(u) \rangle$. So the theorem is true by induction.

(2.9) Theorem. If P is a polynomial over F and f is an endomorphism of V, then ker P_f is an f-invariant subspace of V.

Proof: If u is in the ker P_f, then $P_f(u) = 0$, and hence $P_f[f(u)] = f[P_f(u)] = f(0) = 0$ which implies $f(u) \in$ ker P_f.

(2.10) Theorem. If P and Q are polynomials over F and P divides Q, then ker $P_f \subseteq$ ker Q_f.

Proof: If u is in the ker P_f, then $P_f(u) = 0$ and hence if $Q = JP$, $Q_f(u) = J_f P_f(u) = 0$ which implies that u is an element of ker Q_f.

(2.11) Theorem. If P is a proper divisor of Q and Q is a divisor of the minimal polynomial m of f, then ker P_f is a proper subset of ker Q_f.

Proof: Put $m = QJ$ and let $R = PJ$.

Since P is a proper divisor of Q, then R is a proper divisor of m

and α , $R < \deg m$, so $R_f \neq 0$. Therefore, there exists a v in V such that $R_f(v) \neq 0$. If $w = J_f(v)$, then $P_f(u) = P_f J_f(v) = (PJ)_f(v) = R_f(v) \neq 0$ and hence w is not in the kernel of P_f. On the other hand, $Q_f(w) = Q_f[J_f(v)] = (QJ)_f(v) = m_f(v) = 0$ which implies that $w \in \ker Q_f$. Hence, from Theorem 2.10, ker P_f is a proper subset of ker Q_f.

(2.12) Theorem. If J is the greatest common divisor of P and Q, then

$$\ker J_f = \ker P_f \cap \ker Q_f.$$

Proof: From Theorem 2.10, it follows that $\ker J_f \subseteq \ker P_f \cap \ker Q_f$. Suppose that $u \in \ker P_f \cap \ker Q_f$; that is, $P_f(u) = Q_f(u) = 0$. Since there exists a R and T such that $J = RP + TQ$, then $J_f(u) = (RP + TQ)_f(u) = R_f P_f(u) + T_f Q_f(u) = 0 + 0 = 0$ which implies that u is an element of $\ker J_f$.

(2.13) Theorem. If J is the least common multiple of P and Q, then $\ker J_f = \ker P_f + \ker Q_f$.

Proof: From Theorem 2.10, it follows that $\ker P_f \subseteq \ker J_f$ and $\ker Q_f \subseteq \ker J_f$. Therefore, $\ker P + \ker Q_f \subseteq \ker J_f$. Since J is the l.c.m. of P and Q, J can be written as $P_1 P$ and also as $Q_f Q$ where the g.c.d. of P_1 and Q_1 is 1. Hence there exists an R and T such that $P_1 R + Q_1 T = 1$. Suppose that $u \in \ker J_f$ and let $v = P_{1f} R_f(u)$. Then $P_f(v) = P_f[P_{1f} R_f(u)] = [P_f P_{1f} R_f](u) = [J_f R_f](u) = 0$ which implies that v is in the $\ker P_f$. Similarly, $w = Q_{1f} T_f(u)$ is in $\ker Q_f$. Further, $v + w = P_{1f} R_f(u) + Q_{1f} T_f(u) = 1_f(u) = u$. Therefore u is in the $\ker P_f + \ker Q_f$, so $\ker J_f = \ker P_f + \ker Q_f$.

(2.1 Corollary. If P and Q are relatively prime, then ker $(PQ)_f$ = ker $P_f \oplus$ ker Q_f.

(2.15) Definition. A subspace H of V is said to be <u>f-irreducible</u> if H is f-invariant and there is no decomposition $H = H_1 \oplus H_2$ of H into a direct sum of non-zero f-invariant subspaces.

(2.16) Theorem. Let v be a vector space and $f \in L(V,V)$, then V is equal to the direct sum of f-irreducible subspaces.

Proof: If n = 1, the theorem is obviously true.

Suppose the theorem is true for all vector spaces of dimension less than n. Let the dimension of V be n. If V is already f-irreducible, the theorem is true. If V is not f-irreducible, then there exists a decomposition $V = H_1 \oplus H_2$ where each H_i is f-invariant and non-zero. Then dim $H_i \leq n-1$, i = 1, 2. From the induction hypothesis, the theorem is true for H_1 and H_2 and hence true for V.

(2.17) Definition. A subspace H of V is said to be <u>f-cyclic</u> if there exists a u in H such that $H = \langle u, f(u), f^2(u), \ldots \rangle$. Then u is referred to as an <u>f-generator</u> of H and H is said to f-generated by u.

Notice that an f-cyclic subspace H of V is also f-invariant.

By Theorem 2.8, the vectors $\{u, f(u), \ldots, f^{s-1}(u)\}$ form a maximal independent subset of $\{u, f(u), \ldots\}$, where s is the degree of the f-annihilator of u. So, if s is the degree of the f-annihilator of u, then the set $\{u, f(u), \ldots, F^{s-1}(u)\}$ forms a basis of the f-cyclic subspace f-generated by u. From this, the following theorem is obtained.

(2.18) Theorem. If u is an f-generator of an f-cyclic subspace H of V, then the dimension of H is equal to the degree of the f-annihilator of u.

If L is an f-cyclic subspace f-generated by u, then for every vector $v \in L$, there exists a polynomial Q such that $v = Q_f(u)$. If h is the f-annihilator of u, then $h_f(v) = h_f\, Q_f(u) = h_f Q_f\,(u) = h_f(u)Q_f(u) = 0$. Hence, the f-annihilator of u is the same as the minimal polynomial of the restriction of f to the subspace.

(2.19) Theorem. Suppose that $f \in L(V,V)$ for a vector space V. If f has the minimal polynomial $m = P^s$ where P is an irreducible polynomial, then V is the direct sum of f-cyclic subspaces.

Proof. If $\dim V = 1$, then V itself is f-cyclic and the theorem is true. By induction, the theorem will be proved on the dimension n. Assume the theorem is true for all vectors spaces of dimension less than or equal to $n - 1$. Suppose the $\dim V = n$.

The f-annihilator h of a vector u in V is a divisor of the minimal polynomial m of f. Therefore, h must be of the form $h = P^k$, where k is less than or equal to s. Choose a u in V such that the corresponding k is as large as possible. Then $0 < k \leq s$.

Let L_o be the f-cyclic subspace in V f-generated by this vector u. Then by Theorem 2.18, $\dim L_o = k \deg P \geq 1$. Thus if $\bar{V}$ is the quotient space V/L_o, then $\dim \bar{V} < \dim V = n$.

Define an endomorphism $\bar{f}$ on $\bar{V}$ by $\bar{v} \to \bar{f}(\bar{v}) = \overline{f(v)}$. If $\bar{v} = \bar{w}$, then $v - w \in L_o$ and $f(v) - f(w) \in L_o$ which implies $\overline{f(v)} = \overline{f(w)}$. Now, for any polynomial D, $D_{\bar{f}}(\bar{v}) = \overline{D_f(v)}$, and in particular $m_{\bar{f}}(\bar{v}) = \overline{m_f(v)} = 0$ for

all $v \in V$. Hence the minimal polynomial $\bar{m}$ of $\bar{f}$ is a divisor of m and therefore $\bar{m} = P^t$ for some $t \leq s$.

Since dim $V < n$, then by the induction hypothesis, the theorem is true for $\bar{V}$ with the endomorphism $\bar{f}$. Thus there is a decomposition $\bar{V} = \bar{L}_1 \oplus \ldots \oplus \bar{L}_r$ of V into a direct sum of f-cyclic subspaces $\bar{L}_1, \ldots, \bar{L}_r$.

Let $\bar{u}_1$ be an $\bar{f}$-generator of $\bar{L}_1$ and let $v_1 \in \bar{u}_1 \subseteq V$. The f-annihilator of v_1 has the form P^j where $j \leq k \leq s$. Since it follows from $P_f{}^j(v_1) = 0$ that $P_{\bar{f}}{}^j(\bar{u}_1) = \overline{P_f{}^j(v_1)} = 0$, the $\bar{f}$-annihilator of $\bar{u}_1$ is a divisor of P^j and therefore has the form P^i where $i \leq j \leq k \leq s$.

From $P_f{}^j(v_1) = 0$ it follows that $P_f{}^k(v_1) = P_f^{k-i} P_f{}^i(v_1) = 0$.

Since $P_{\bar{f}}{}^i(\bar{u}_1) = 0$, $P_f{}^i(u_1) \in L_o$ and therefore there exists a polynomial Q such that $P_f{}^i(v_1) = Q_f(u)$. Therefore, P^k divides $P^{k-1}Q$, because P^k is the f-annihilator of u. Hence there exists a polynomial Q^* such that $Q = P^i Q^*$ and it follows that $P_f{}^i(v_1) = P_f{}^i Q^*(u)$, that is, there exists a $w = Q_f^*(u) \in L_o$ such that $P_f{}^i(v_1) = P_f{}^i(w)$.

Putting $z = v_1 - w \in \bar{u}_1 \subseteq V$ then $P_f{}^i(z) = 0$ so that the f-annihilator of z is a divisor of P^i. But by putting z in the place of v_1, it can also be determined that the f-annihilator of z is also a multiple of P^i. Therefore it is equal to P^i.

Now suppose that L_1 is an f-cyclic subspace of V f-generated by z. If the degree of P^i is denoted by y, then the set $\{z, f(z), \ldots, f^{y-1}(z)\}$ is a basis of L_1 and $\{\bar{z}, \bar{f}(\bar{z}), \ldots, \bar{f}^{y-1}(\bar{z})\}$ is a basis of $\bar{L}$. The canonical mapping $u \in L_1 \rightarrow \bar{u} \in \bar{L}$ is therefore an isomorphism of L_1 onto L_2.

If the same argument is carried out for $\bar{L}_2, \ldots, \bar{L}_r$ as for $\bar{L}_1$, then the f-cyclic subspaces $L_2, \ldots, L_r$ of V are obtained in addition to L_1.

Therefore, $V = L_0 \oplus \ldots \oplus L_r$.

(2.20) Theorem. Let $f \in L(V,V)$. Then V is the direct sum of f-cyclic subspaces which are f-irreducible. That is, $V = L_1 \oplus \ldots \oplus L_r$, where $L_1, \ldots, L_r$ are non-zero.

If m is the minimal polynomial of f, then the minimal polynomial of the restriction of to L_k is of the form $m_k = P_k^{s_k}$ where P_k is irreducible, $s_k \geq 1$ and $P_k^{s_k}$ divides m $(k = 1, 2, \ldots, r)$.

Proof: By theorem 2.16 there is a decomposition of the form $V = L_1 \oplus \ldots \oplus L_r$ in which each L_i is f-irreducible, $i = 1, \ldots, r$. The minimal polynomial m_k obviously divides m.

Note that $m_k \neq 1$ since $L_k \neq 0$.

Also, m_k is of the form $P_k^{s_k}$, for some irreducible polynomial P_k. For if this is not the case, m_k equals some PQ, for some relatively prime factors P and Q. Then by Corollary 2.14, $L_k = \ker (m_k)_f = \ker P_f + \ker Q_f$ and by Throrem 2.9 and 2.11, $\ker P_f$ and $\ker q_f$ are proper f-invariant subspaces of $\ker (m_k)_f = L_k$, so L_k is not f-irreducible.

Each $L_k (k = 1, \ldots, r)$ is the direct sum of f-cyclic subspaces by Theorem 2.19, but L_k is f-irreducible so L_k itself is f-cyclic.

(2.21) Definition. Two f-invariant subspaces L_1, L_2 are said to be <u>f-equivalent</u> if there exists an isomorphism g of L_1 onto L_2 which

commutes with f, that is, $g[f(u)] = f[g(u)]$ for every u in L_1. (It may be noted here that g need only be defined from L_1 onto L_2 since L_1 and L_2 are f-invariant.)

(2.22) Theorem. Let $f \in L(V,V)$. Two f-cyclic subspaces L_1 and L_2 of V are f-equivalent if and only if the restrictions of f to L_1 and L_2 have the same minimal polynomial.

Proof. Suppose that L_1 and L_2 are f-equivalent subspaces of V and g is an isomorphism of L_1 onto L_2 which commutes with f. Since g also commutes with powers of f, $P_f g = g P_f$ for any polynomial P. Let m_1 and m_2 be the minimal polynomials of the restrictions of f to L_1 and L_2 respectively. Then $(m_1)_f g(u) = g(m_1)_f(u) = 0$ for all u in L_1. Therefore, $(m_1)_f(v) = 0$ for all v in L_2 since g is an isomorphism. Hence, m_2 divides m_1. Using a similar argument, it can also be shown that m_1 divides m_2 which implies that $m_1 = m_2$.

(It may be noted that L_1 and L_2 need not be f-cyclic for this argument to work.)

Conversely, suppose that the restrictions of f to L_1 and L_2 have the same minimal polynomial m. Let u and v be the f-generators of the f-cyclic subspaces L_1 and L_2 respectively. Then for each vector w in L_1, w can be written in the form $P_f(u)$ for some polynomial P.

(Note: If H is an f-cyclic subspace f-generated by u and m is the minimal polynomial of f restricted to H, then for any polynomial P, $P_f(u) = 0$ if and only if m divides P. If $P_f(u) = 0$, suppose that m does not divide P. P does not divide m since if it did it would contradict the fact that m is the minimal polynomial of f restricted

to m since u is the f-generator of H. Therefore, P and m would be relatively prime. Then a contradiction is reached since the kernel of P_f restricted to H is a subset of ker m_f and by Corollary 2.14.)

For any two polynomials P and Q, $P_f(u) = Q_f(u)$ if and only if m divides P - Q. Likewise, $P_f(v) = Q_f(v)$ if and only if m divides P - Q. Therefore, $P_f(u) = Q_f(u)$ if and only if $P_f(v) = Q_f(v)$.

Define a mapping of g of L_1 to L_2 by $g[P_f(u)] = P_f(v)$. The map is well defined by the above paragraph. Also, $g[aP_f(u) + bQ_f(u)] = g[(aP + bQ)_f(u)] = (aP + bQ)_f(v) = aP_f(v) + bQ_f(v) = g[P_f(u)] + g[Q_f(u)]$. Hence g is linear.

Let $w \in L_1$ be such that $g(w) = 0$. By the nature of L_1, there exists a polynomial R such that $R_f(u) = w$. Then $0 = g(w) = g[R_f(u)] = R_f(v)$. But $R_f(v) = 0$ if and only if m divides R which implies that $R_f(u) = 0$. Also, pick a vector $z \in L_2$. Then $z = P_f(v)$ for a suitable polynomial P. But $P_f(u)$ is in L_1 and $g[P_f(u)] = z$. Hence g is an isomorphism of L_1 onto L_2.

Now, f(u) can be written the form $P_f(u)$ where P = x and u can be written in the form $Q_f(u)$ where Q = 1. Then $g[f(u)] = g[P_f(u)] = P_f(v) = f(v)$. Also, $f[g(u)] = fg[Q_f(u)] = f[Q_f(v) = f(v)$. Hence g commutes with f since u is the f-generator of L_1.

(2.23) **Theorem.** Let V be a vector space and let $f \in L(V,V)$. Suppose that $V = L_1 \oplus \ldots \oplus L_r$ where each L_i is non-zero, f-cyclic, and f-irreducible. Suppose in addition that

$$V = L_1{}^* \oplus \ldots \oplus L_s{}^*$$

is a decomposition of the same type. Then $r = s$ and by renumbering, L_k and L_k^* are f-equivalent ($k = 1, \ldots, r$).

Proof: Let $P_1^{s_1} P_2^{s_2} \ldots P_j^{s_j}$ be the minimal polynomial of f where $P_1, \ldots, P_j$ are distinct monic irreducible polynomials. By Theorem 2.20, the minimal polynomial m_k of f restricted to L_k has the form $m_k = P_i^s$ where $1 \leq i \leq j$ and $1 \leq s \leq s_i$. Using the previous theorem, this theorem can be proven by showing that for each pair (i,s), the numbers $n(i,s)$ of L_k with $m_k = P_i^s$ are uniquely determined by f. Following this procedure, the following is obtained.

$$n(i,s) = \frac{1}{\deg P_i} \left[\operatorname{rank}(P_i^{s-1})_f - 2\operatorname{rank}(P_i^s)_f + \operatorname{rank}(P_i^{s+1})_f\right]$$

[10, pp. 290-91].

Hence the numbers $n(i,s)$, $(1 \leq i \leq j;\ 1 \leq s \leq s_i)$, are uniquely determined by the irreducible factors P_i of the minimal polynomial of f, therefore by f itself.

Let $A = (a_{ij})$ be an $n \times n$ matrix whose elements are from a field F. If V is a vector space of dimension n with a basis $\{e_1, \ldots, e_n\}$ then an endomorphism f on V is defined by $f(e_j) = \sum_{i=1}^{n} a_{ij} e_i$.

Considering this endomorphism f, let $V = L_1 \oplus \ldots \oplus L_r$ be a decomposition of V where each L_i is non-zero, f-cyclic, and f-irreducible. If u_k is the f-generator of L_k and if q_k is the dimension of L_k, that is, the degree of the minimal polynomial m_k, then by Theorem 2.8, the vectors $u_k, f(u_k), \ldots, f^{q_k-1}(u_k)$ form a basis of L_k since m_k is also the f-annihilator of u_k. Then the set

$\{f^i(u_j) \mid j = 1, \ldots, r;\ i = 1, \ldots, q_k-1\}$

is a basis of V. Since $f[f^i(u_k)] = f^{i+1}(u_k)$, each basis vector is mapped to the next for $i = 1, \ldots, q_k-1$. And $f[f^{q_k-1}(u_k)] = f^{q_k}(u_k)$, which is a linear combination of $u_k, \ldots, f^{q_k-1}(u_k)$, say $f^{q_k}(u_k) = -a_{k,q_k-1}f^{q_k-1}(u_k)-\ldots-a_{k,1}f(u_k) - a_{k,o}u_k$.

Hence, $\sum_{i=1}^{q_k} a_{k,i}f^i(u_k) = 0$ where $a_{k,q_k} = 1$. Therefore, the part of the matrix of f corresponding to the basis vectors $u_k, \ldots, f^{q_k-1}(u_k)$ of the subspace L_k has the form

$$C_k = \begin{bmatrix} 0 & 0 & . & . & 0 & -a_{k,o} \\ 1 & 0 & . & . & 0 & -a_{k,1} \\ 0 & 1 & . & . & 0 & -a_{k,2} \\ . & . & & & . & . \\ . & . & & & . & . \\ 0 & 0 & . & . & 0 & -a_{k,q_k-2} \\ 0 & 0 & . & . & 1 & -a_{k,q_k-1} \end{bmatrix}$$

It follows that the matrix of f with respect to the basis $\{f^i(u_j) \mid j = 1, \ldots, r;\ i = 1, \ldots, q_k-1\}$ is of the form

$$C = \begin{bmatrix} C_1 & 0 & . & . & . & 0 \\ 0 & C_2 & . & . & . & 0 \\ . & . & & & & . \\ . & . & & & & . \\ 0 & 0 & . & . & . & C_r \end{bmatrix}$$

The blocks $C_1, C_2, \ldots, C_r$ are uniquely determined (by Theorem 2.23) by the original matrix A (up to order).

Recalling the previous information on similar matrices plus the fact that the two matrices A and C are relative to the same **linear** map f, then the following theorem is obtained.

(2.24) Theorem. Let A be an n x n matrix whose elements are from a field F. Then there exists an n x n matrix $C = SAS^{-1}$ similar to A and having the form shown above. Except for order, the blocks C_1, $C_2, \ldots, C_r$ are uniquely determined.

Note that in the theorem above the characteristic polynomial of C is $\det(C - xI) = \det(SAS^{-1} - xI) = \det[s(A - xI)S^{-1}] = \det(A = xI)$ which is identical to the characteristic polynomial of A.

(2.25) Theorem. Let $f \in L(V,V)$ where V is a vector space. Then the minimal polynomial of f divides the characteristic polynomial.

Proof: First the characteristic polynomial of the block C_k needs to be

obtained. This is simply the determinant of:

$$\begin{bmatrix} -x & 0 & . & . & -a_{k,o} \\ 1 & -x & . & . & -a_{k,1} \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & -a_{k,q_k-1} - x \end{bmatrix}$$

Expanding about the first row, the following is obtained: $(-1)^{q_k}(x^{q_k} + a_{k,q_k-1}x^{q_k-1} + \ldots + a_{k,1}x + a_{k,o})$. However, since the minimal polynomial of f restricted to L_k is $\sum_{i=1}^{q_k} a_{k,i}x^i$ (where $a_{k,q_k} = 1$), the characteristic polynomial of C_k is equal to (or the negative of) the minimal polynomial of the restriction of f to L_k.

Therefore, the theorem is true from the fact that the characteristic polynomial of f is equal to the product of those for each of the individual blocks C_k.

(2.26) Theorem. Let $f \in L(V,V)$. Let h be the characteristic polynomial of f. Then $h_f = 0$. (That is, every endomorphism satisfies its characteristic polynomial.)

Proof: From Theorem 2.25 and the fact that $m_f = 0$.

The above theorem is known as the **Cayley-Hamilton theorem. In**

this particular case the Caley-Hamilton theorem is easily proven by the use of the minimal polynomial and its characteristics. However, many authors use this theorem (verified by other means) to prove the existence of the minimal polynomial.

Suppose that the irreducible factors $p_j (o \leqq j \leqq q)$ are linear, that is $p_j = x - c_j$ and $m = (x - c_1)^{s_1} \ldots (x - c_q)^{s_q}$, so $(p_j)_f = f - c_j e (o \leqq j \leqq q)$; $m = (f - c_1 e)^{s_1} \ldots (f - c_q e)^{s_q}$ where e is the identity map on V. This will always be the case where F is an algebraically closed field.

The minimal polynomial m_k of the restriction of f to the subspace L_k then has the form $m_k = (x - c_k)^{t_k}$. (For simplicity, c_k is used in the place of c_{j_k}.) Hence dim $L_k = t_k$. If u_k is the f-generator of L_k, then the vectors u_k, $(f - c_k e)(u_k)$, ... , $(f - c_k e)^{t_k - 1}(u_k)$ are linearly independent (otherwise the degree of the f-annihilator of u_k would be less than t_k) and therefore form a basis of L_k. For $0 \leqq t \leqq t_k - 1$, $(f - c_k e)^t (u_k)$ is mapped by f onto $(f - c_k e + c_k e)$ $(f - c_k e)^t (u_k) = [(f - c_k e)(f - c_k e)^t + c_k (f - c_k e)^t](u_k) =$ $(f - c_k e)^{t+1}(u_k) + c_k (f - c_k e)^t (u_k)$. If $t = t_k - 1$, then the first term is zero. Hence the natrix of the restriction of f to L_k relative to the given basis is of the form:

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

Combining the bases of each L_k into a basis of V, then the matrix of f has the form:

$$J = \begin{bmatrix} J_1 & 0 & . & . & 0 \\ 0 & J_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & J_r \end{bmatrix}$$

The matrix J is known as the <u>Jordan canonical form</u> of A and the blocks J_k are called the <u>Jordan blocks</u> of A. It may be noted here that in the derivation of the Jordan canonical form, many authors have the ones on the super diagonal of the Jordan blocks. This change may

be due to the form of the matrix of the linear map or the order of the basis elements of each f-cyclic subspace used by the individual author.

The characteristic polynomial of the block J_k is $(x - c_k)^{t_k}$ and that of J is therefore $(x - c_1)^{t_1} \ldots (x - c_r)^{t_r}$. The diagonal elements $c_1, \ldots, c_r$ are therefore the eigenvalues of f and in fact each eigenvalue appears in J just as many times as its multiplicity as a root of the characteristic polynomial. Since the blocks J_k correspond to the subspaces L_k which are uniquely determined apart from f-equivalence, and since the restrictions of f to two f-equivalent subspaces have the same characteristic equation, then J is uniquely determined by A except for order of the blocks J_k. (Note: There are as many J_k's as there are L_k's.) This proves the following theorem.

(2.27) Theorem. Every square matrix A with elements from an algebraically closed field F is similar to a matrix $J = SAS^{-1}$ in the Jordan canonical form. The diagonal elements of the Jordan canonical form are the eigenvalues of A and each appears as often as its multiplicity as a root of the characteristic polynomial. The Jordan canonical form is uniquely determined by A except for the order of the Jordan blocks.

Chapter III

In his book Charles W. Curtis [5, pp. 158-62] uses a somewhat different approach to the Jordan canonical form than was studied in the previous chapter. In this chapter, the fact that a vector space can be broken down into the direct sum of non-zero f-cyclic subspaces is proved using the concept of the dual space. This approach is also used by Jimmie D. Gilbert [4, pp. 242-46] and John T. Moore [8, pp. 327-32].

(3.1) Definition. A linear transformation $F \in L(V,V)$ is called <u>diagonable</u> if there exists a basis of V consisting of eigenvectors of f. The matrix of f relative to this basis of eigenvectors is a diagonal matrix. That is, it has the form

$$\begin{bmatrix} a_1 & 0 & \cdot & \cdot & 0 \\ 0 & a_2 & \cdot & \cdot & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & 0 & \cdot & \cdot & a_n \end{bmatrix}$$

with zeros except in the (i,i) positions, $(1 \leq i \leq n)$.

Note that the matrix of f relative to any other basis of V is similar to the above matrix.

(3.2) Definition. Let $f \in L(V,V)$. If there exists an i such that

$f^i = 0$, then f is said to be a <u>nilpotent</u> linear transformation. The smallest such positive i is called the <u>index of nilpotency</u>.

Now a close study will be made of a nilpotent endomorphism f on V. Since f is nilpotent, there is a positive integer k such that $f^k = 0$. So f satisfies the polynomial x^k. Since the minimal polynomial m divides any annihilator of V then $m = x^t$, $t \leq k$.

(3.3) Lemma. Let W be an f-cyclic subspace of V. Then there exists a $w \in W$ and a positive integer k such that $f^k(w) = 0$ and such that the set $\{f^{k-1}(w), \ldots, f(w), w\}$ is a basis of W.

Proof: Since W is f-cyclic there exists a $w \in W$ such that the set $\{w, f(w), \ldots\}$ generates W. Since W is also finite dimensional there exists a k such that the set $\{f^{k-1}(w), \ldots, f(w), w\}$ is a basis of W. Since f is nilpotent on V then the restriction f_1 of f to W is nilpotent on W, and hence the minimal polynomial of f_1 has the form x^t where $t = k$. (This follows from the fact that w is the f-generator of W, that the degree of the f-annihilator of w is the dimension of W, and that the minimal polynomial of f_1 is equal to the f-annihilator of w.) Therefore, $f^k(w) = 0$.

Note that the matrix of f restricted to W relative to the basis $\{f^{k-1}(w), \ldots, f(w), w\}$ has the form

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

(3.4) Definition. Let V be a vector space over a field F. The <u>dual space</u> V^* of V is the vector space L(V,F). Elements of V^* are called <u>linear functionals</u> of V.

(3.5) Lemma. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for V. Then there exist linear functionals $\{f_1, f_2, \ldots, f_n\}$ in V^* such that for each $i, f_i(v_i) = 1$ and $f_i(v_j) = 0, i \neq j$. This set of linear functionals forms a basis of V^*.

Proof: The set of linear functionals exists since linear transformations may be defined which map basis vectors into arbitrary vectors in the image space.

Suppose that $a_1 f_1 + \ldots + a_n f_n = 0$. Then applying both sides to the vector v_i, the following relation is obtained: $a_1 f_1(v_i) + \ldots + a_i f_i(v_i) + \ldots + a_n f_n(v_i) = 0$ which implies that $0 = a_i \cdot 1 = a_i$. So the functionals $f_1, \ldots, f_n$ are linearly

independent.

Let $f \in V^*$ and suppose $f(v_i) = a_i$. Then $(a_1f_1 + \ldots + a_nf_n)(v_i) = a_if_i(v_i) = a_i$ for each $i = 1, \ldots, n$. So $f = a_1f_1 + \ldots + a_nf_n$.

Therefore the set $\{f_1, \ldots, f_n\}$ is a basis of V^*.

(3.6) Definition. The basis $\{f_1, \ldots, f_n\}$ described in the above lemma is called the <u>dual basis</u> of $\{v_1, \ldots, v_n\}$.

(3.7) **Lemma.** The function $v \to \tilde{v}$ is an isomorphism of V with V^{**}, where $\tilde{v} \in V^{**}$ is the linear functional on V^* defined by $\tilde{v}(g) = g(v)$.

Proof: As defined, $\tilde{v}$ is clearly in V^{**} and the function $v \to \tilde{v}$ is clearly linear. If $\tilde{v} = 0$, then $g(v) = 0$ for all $g \in V^*$, so $v = 0$. That is, the correspondence is one-to-one. Since $\dim V^{**} = \dim V^* = \dim V$, the correspondence is an isomorphism.

Notice that if $v_i, \ldots, v_n$ is a basis of V and $g_1, \ldots, g_n$ is a basis of V^*, then the basis of V^{**} dual to $g_1, \ldots, g_n$ is $\tilde{v}_1, \ldots, \tilde{v}_n$.

(3.8) Definition. If W is a subspace of V, let $W^\perp$ be the set of all linear functionals $g \in V^*$ such that $g(W) = 0$. Then $W^\perp$ is a subspace.

Note that $W^\perp$ is indeed a subspace since if $g_1, g_2 \in W^\perp$, $w \in W$, and $a \in F$, then $(ag_1 + g_2)(w) = ag_1(w) + g_2(w) = a_10 + 0 = 0 + 0 = 0$. Moreover, if $\dim W = k$ and $\dim V = n$, then $\dim W^\perp = n - k$. For if $v_1, \ldots, v_n$ is a basis of V, $v_1, \ldots, v_k$ is a basis of W, and if $g_1, \ldots, g_n$ is the basis of V^* dual to $v_1, \ldots, v_n$; then since $g_i(v_j) =$

$\delta_{i,j}$ (Kronecker delta), $g_{k+1}, \ldots, g_n$ is a basis of W

If S is a subspace of V^*, then $S^\perp$ is a subspace of V^{**}. However, under the identification described in Lemma 3.7, $S^\perp$ can be considered the set $\{v \in V \mid g(v) = 0 \text{ for all } g \in S\}$ and $\dim S^\perp = \dim V^{**} - \dim S = \dim V - \dim S = n - k$.

(3.9) Definition. Let V be a vector space over F and let $f \in L(V,V)$. Define $f^* \in K(V^*,V^*)$ as follows: for $g \in V^*$ let $(f^*g)(v) = g[f(v)]$, $v \in V$. The linear transformation f^* is called the <u>transpose</u> of f.

(Note that f^*g is linear since $(f^*g)(av_1 + bv_2) = g[f(av_1 + bv_2)] = g[af(v_1) + bf(v_2)] = af[f(v_1)] + bg[f(v_2)] = a(f^*g)(v_1) + b(f^*g)(v_2)$, for $v_1, v_2 \in V$, $g \in V^*$ and $a,b \in F$.

Also $f^* \in L(V^*,V^*)$ since for $g_1, g_2 \in V^*$, $[f^*(ag_1 + bg_2)](v) = (ag_1 + bg_2)[f(v)] = ag_1[f(v)] + bg_2[f(v)] = a(f^*g_1)(v) + b(f^*g_2)(v)$ for all $v \in V$ and $a,b \in F$.)

(3.10) Theorem. Let f be a nilpotent linear transformation on a vector space V. Then V can be expressed as a direct sum $V = V_1 \oplus \ldots \oplus V_r$ where each V_i is a non-zero f-cyclic subspace.

Proof: First, it will be shown that either V is f-irreducible or that it is the direct sum of f-irreducible subspaces. If $\dim V = 1$, then V itself is f-irreducible. Suppose that all vector spaces of dimension less than or equal to $n - 1$ are either f-irreducible or can be broken down into the direct sum of f-irreducible subspaces, and suppose that $\dim V = n$. Then either V is f-irreducible or it is not. If it is not

f-irreducible, then $V = H_1 \oplus H_2$ where $H_i \neq \{0\}$, $i = 1,2$. By the induction hypothesis, H_1 and H_2 are either f-irreducible or can be written as the direct sum of f-irreducible subspaces (since $\dim H_i \leq n - 1$, $i = 1,2$). Hence, either V is f-irreducible or it can be written as the direct sum of f-irreducible subspaces.

In order to show that if V is f-irreducible then it is f-cyclic, it suffices to show that if V is not f-cyclic then it is not f-irreducible. Let m_k be the minimal polynomial on V. Then there exists an f-cyclic subspace W of V with basis $\{w,, \ldots ,f^{k-1}(w)\}$. Otherwise, by Lemma 3.3, $f^{k-1} = 0$ on V, contrary to the assumption that m_k is the minimal polynomial of f. Since V is not f-cyclic, $V \neq W$.

Let f* be the transpose of f. Then for $v \in V$ and $h \in V^*$, $[(f^*)^i h](v) = h[(f)^i](v) = [h\ f^i](v) = [(f^i)^* h](v)$; that is, $(f^*)^i = (f^i)^*$. Also, for some g in V*, $(f^*)g \notin W^\perp$. Otherwise, $(f^*)^i g \in W^\perp$ for all $g \in V^*$. Then $0 = [(f^*)^{k-1} g](w) = g[f^{k-1}(w)]$ for every $g \in V^*$ which implies that $f^{k-1}(w) = 0$ which contradicts the fact that $\{w, \ldots ,f^{k-1}(w)\}$ is a basis of W. Therefore, $(f^*)^{k-1} g \notin W^\perp$ for some $g \in V^*$.

In relation to this g, suppose $a_0 g + a_1(f^*)g + \ldots + a_{k-1}\ (f^*)^{k-1} g \in W^\perp$ and suppose that $a_0 = \ldots = a_{i-1} = 0$ and $a_i \neq 0$. Then $a_i\ (f^*)^i g + \ldots a_{k-1}[(f^*)^{k-1} g] \in W^\perp$. Applying $(f^*)^{k-1-i}$ to this relation then $a_i[(f^*)^{k-1} g] \in W^\perp$ since $(f^*) \neq 0$. But this contradicts the fact that $(f^*)^{k-1} g \notin W^\perp$. Hence if $a_0 g + a_1[(f^*)g] + \ldots + a_{k-1}[(f^*)^{k-1} g] \in W^\perp$, then $a_0 = \ldots = a_{k-1} = 0$. Therefore, if S is the subspace of

V^* generated by $g, f^*g, \ldots, (f^k)^{k-1}g$, then $S \cap W^\perp = 0$ and $\dim S = k$. So $V^* = S \oplus W$.

Now consider the subspace $S^\perp$ of V. If $v \in S^\perp$ and $h \in S$, then $h[f(v)] = f^*[h(v)] = f^*(0) = 0$. So $S^\perp$ is f-invariant.

If $v \in W \cap S^\perp$, then since $V^* = W^\perp \oplus S$, $h(v) = 0$ for all $h \in V^*$. Therefore $v = 0$. Hence $W \cap S^\perp = 0$. Since $\dim S^\perp = n - k$, $\dim W = k$ and $W \cap S^\perp = 0$, then $V = W \oplus S^\perp$. But this contradicts the assumption that V is f-irreducible.

Therefore V is f-cyclic.

It is a well-known fact that a nilpotent matrix is similar to an upper triangular matrix. The following corollary says that this upper triangular matrix can be chosen in a very special form.

(3.11) Corollary. Let f be a nilpotent transformation on a vector space V. Then V has a basis such that the matrix of f relative to this basis has the form

$$\begin{bmatrix} A_1 & 0 & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & \cdot & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & 0 & \cdot & \cdot & A_r \end{bmatrix}$$

with zeros except in the diagonal blocks and each A_i of the form

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

Proof: Decompose V as in Theorem 3.8 such that $V = V_1 \oplus \ldots \oplus V_r$ where each V_i has the basis $\{f^{k-1}(v), \ldots , f(v), v\}$. Then the matrix of f restricted to V_i is A_i.

(3.12) Definition. Let $f \in L(V,V)$. The <u>null space</u> of f is the set $\{v \in V \mid f(v) = 0\}$.

(3.13) Corollary. Let $f \in L(V,V)$ and let the minimal polynomial of f be $m = (x - a_1)^{d_1} \ldots (x - a_s)^{d_s}$ where $a_1, \ldots , a_s$ are the eigenvalu of f. Then there exists a basis of V such that the matrix of f relative to this basis has the form

$$\begin{bmatrix} J_1 & 0 & . & . & 0 \\ 0 & J_2 & . & . & . \\ . & . & & & . \\ 0 & 0 & . & . & J_s \end{bmatrix}$$

with zero except in the diagonal blocks and each block having the form

$$\begin{bmatrix} a_i & 1 & 0 & . & . & 0 \\ 0 & a_i & 1 & . & . & 0 \\ . & . & . & & & . \\ . & . & . & & & . \\ 0 & 0 & 0 & . & . & 1 \\ 0 & 0 & 0 & . & . & a_i \end{bmatrix}$$

Proof: Let $m = P_1^{e_1} \ldots P_s^{e_s}$ be the minimal polynomial of f factored into distinct primes $P_i \in F[x]$ (that is $P_i = (x - a_i)$.)

Let $Q_i = \frac{m}{P_i^{e_i}}$, $1 \leq i \leq s$. Then the Q_i's are polynomials in $F[x]$ with no common prime factors. Hence, there exist polynomials R_i, $1 \leq i \leq s$, such that

$1 = Q_1R_1 + \ldots + Q_sR_s$ [3, pp. 169-70]. Substituting f,

$1 = (Q_1)_f(R_1)_f + \ldots + (Q_s)_f(R_s)_f$.

Now let $v \in V$, then

$v = (Q_1)_f(R_1)_f(v) + \ldots + (Q_s)_f(R_s)_f(v)$ and for each i,

$(P_i)_f^{e_i}(Q_i)_f(R_i)_f(v) = m_f(R_i)_f(v) = (R_i)_f m_f(v) = 0$. Therefore

$(Q_i)_f(R_i)_f(v) \in n[(P_i)_f^{e_i}]$, the null space of $(P_i)_f^{e_i}$. Thus

$v \in n[(P_1)_f^{e_i}] + \ldots + n[(P_s)_f^{e_s}]$.

Suppose that $v_1 + \ldots + v_s = 0$, $v_i \in n[(P_i)_f^{e_i}]$, $1 \leq i \leq s$.

Then there are polynomials $P^*_1Q^*$ such that

$1 = P^*P_1^{e_1} + Q^*P_2^{e_2} \ldots P_s^{e_s}$. Then

$1 = P_f^*(P_1)_f^{e_1} + Q^*(P_2)_f^{e_2} \ldots (P_s)_f^{e_s}(-v_2 - \ldots -v_s) = 0$.

A similar argument proves $v_2 = \ldots = v_s = 0$. Therefore, $V = n[(P_1)_f^{e_1}] \oplus \ldots \oplus n[(P_s)_f^{e_s}]$. [3, p.170].

Let $V_i = n[(f - a_iI)^{d_i}]$. On the space V_i, $f - a_iI$ is a nilpotent transformation. Apply Corollary 3.9 to $f - a_iI$ on the space V_i. There exists a basis of V_i such that the matrix A of $f - a_iI$ has the form:

$$\begin{bmatrix} A_1 & 0 & . & . & 0 \\ 0 & A_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & A_s \end{bmatrix}$$

where each block is in the form as in Corollary 3.9. If B is the matrix of f on the space V_i relative to this basis, then $A = B - a_iI$ and $B = a_iI + A$.

It follows that

$$B = \begin{bmatrix} B_1 & 0 & . & . & 0 \\ 0 & B_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & B_s \end{bmatrix}$$

where each B_i has the form

$$\begin{bmatrix} a_i & 1 & 0 & 0 & . & . & 0 \\ 0 & a_i & 1 & 0 & . & . & 0 \\ 0 & 0 & a_i & 1 & . & . & 0 \\ . & . & . & . & & & . \\ . & . & . & . & & & . \\ 0 & 0 & 0 & 0 & . & . & 1 \\ 0 & 0 & 0 & 0 & . & . & a_i \end{bmatrix}$$

Combining these bases for the various spaces $V_i = n[(f - a_i I)^{d_i}]$, then a basis of V is obtained such that the matrix of f relative to this basis has the required form.

(3.14) Theorem. If $f \in L(V,V)$, then there exists a basis of such

that the matrix B of f relative to this basis is in the Jordan canonical form.

Notice that in this form the ones are on the superdiagonal of each Jordan block. This results from the fact that in Lemma 2.5 the basis of the f-cyclis subspace was given as $\{f^{k-1}(w), f^{k-2}(w), \ldots, f(w), w\}$. The matrix of f restricted to this subspace relative to this basis therefore has ones on the superdiagonal. If the basic elements of the f-cyclic subspace had been taken in the reverse order, then the ones would have been on the subdiagonal as in Chapter II.

Chapter IV

Since any field can be extended to an algebraically closed field, then any square matrix with elements from a field F is similar to a matrix in the Jordan canonical form. From Chapter II there are as many blocks in the Jordan canonical form as there are independent eigenvectors. Also, each eigenvalue a_i appears on the diagonal as many times as its multiplicity as a root of the characteristic polynomial. The order of each block is equal to the dimension of the f-cyclic subspace L_i that determines it. (Note that f is the endomorphism determined by the matrix A.)

In Chapter III it was found that if the minimal polynomial of the linear transformation f has the form $(x - a_1)^{d_1} \ldots (x - a_s)^{d_s}$ where the $\{a_i\}$ are eigenvalues, then each a_i appears on the diagonal as many times as dimension of the null space $n(f - a_i I)^{d_i}$. The order of the largest block is d_i and there are as many blocks with a_i on the diagonal as there are independent eigenvectors of f belonging to the eigenvalue a_i.

(4.1) Definition. Let $f \in L(V,V)$ and let a_i be an eigenvalue of f. Then the number of times a_i appears as a root of the characteristic polynomial of f is called the <u>algebraic multiplicity</u> of a_i. If $S(a_i)$ is the set of all eigenvectors of f corresponding to a_i, then $S(a_i)$ is called the <u>eigenspace</u> associated with a_i and dim $S(a_i)$ is said to be the <u>geometric multiplicity</u> of a_i; that is, the geometric

multiplicity of a_i is the number of independent eigenvectors associated with a_i. Note that the geometric multiplicity of a_i is less than or equal to the algebraic multiplicity of a_i.

Condensing the information from Chapters II and III, the following is obtained.

(4.2) Theorem. Let A be a matrix with characteristic polynomial $P = (x - a_1)^{r_1} \ldots (x - a_k)^{r_k}$ and minimal polynomial $m = (x - a_1)^{s_1} \ldots (x - a_k)^{s_k}$, where $s_i \leq r_i$, then A is similar to a matrix J with submatrices of the form

$$\begin{bmatrix} a_i & 1 & 0 & . & . & 0 & 0 \\ 0 & a_i & 1 & . & . & 0 & 0 \\ 0 & 0 & a_i & . & . & 0 & 0 \\ . & . & . & & & . & . \\ . & . & . & & & . & . \\ 0 & 0 & 0 & . & . & a_i & 1 \\ 0 & 0 & 0 & . & . & 0 & a_i \end{bmatrix}$$

along the main diagonal. All other elements of J are zero. For each a_i there is at least one J_i of order s_i. All other blocks with a_i on the diagonal are of order less than or equal to s_i. The number of J_i corresponding to a_i is equal to the geometric multiplicity of a_i. The sum of the orders of all the blocks determined by a_i is the algebraic multiplicity of a_i. While the ordering of the J_i along the

diagonal is not unique, the number of J_i of each possible ordering is uniquely determined by A.

For example, an eigenvalue a of algebraic multiplicity 1 determines the one-by-one block [a] in the Jordan form. Hence, if a matrix is similar to a diagonal matrix, then its Jordan form is simply a diagonal matrix with the eigenvalues on the diagonal. Also, if a matrix is nilpotent, its Jordan form has zeros on the diagonal and ones on the superdiagonal.

Example 1. Suppose the characteristic and minimal polynomial of a transformation f are

$$P = (x - 2)^4(x - 3)^3 \text{ and}$$

$$m = (x - 2)^2(x - 3)^2 \text{ respectively.}$$

The Jordan form of f has either one of the two following forms:

$$\left[\begin{array}{cc|cc|cc|c} 2 & 1 & & & & & \\ 0 & 2 & & & & 0 & \\ \hline & & 2 & 1 & & & \\ 0 & & 0 & 2 & & & \\ \hline & & & & 3 & 1 & \\ & & & & 0 & 3 & \\ \hline & & & & & & 3 \end{array}\right],$$

$$\left[\begin{array}{cc|c|c|cc|c}
2 & 1 & & & & & \\
0 & 2 & & & & 0 & \\
\hline
 & & 2 & & & & \\
\hline
 & & & 2 & & & \\
\hline
 & 0 & & & 3 & 1 & \\
 & & & & 0 & 3 & \\
\hline
 & & & & & & 3
\end{array}\right]$$

where the matrix is of the first form if f has two independent eigenvectors belonging to the eigenvalue 2, or of the second form if f has three independent eigenvectors belonging to 2. (Note that there must be a 2 x 2 block corresponding to the eigenvalue 2 and a 2 x 2 block corresponding to the eigenvalue 3 since $(x - 2)^2$ and $(x - 3)^2$ are factors of the minimal polynomial.)

Example 2. Find all possible Jordan canonical forms of an endomorphism f whose characteristic polynomial is given by

$$P = (x - 2)^3(x - 5)^2.$$

Since $(x - 2)$ has three for an exponent in the characteristic polynomial, 2 will appear three times on the diagonal. Also, 5 will appear twice. The possibles forms are:

$$\left[\begin{array}{c|c|c|cc} 2 & & & & \\ \hline & 2 & & 0 & \\ \hline & & 2 & & \\ \hline & & & 5 & 1 \\ 0 & & & 0 & 5 \end{array}\right] \qquad \left[\begin{array}{cc|c|cc} 2 & 1 & & & \\ 0 & 2 & & 0 & \\ \hline & & 2 & & \\ \hline & 0 & & 5 & \\ \hline & & & & 5 \end{array}\right]$$

$$\left[\begin{array}{c|c|c|cc} 2 & & & & \\ \hline & 2 & & 0 & \\ \hline & & 2 & & \\ \hline & & & 5 & \\ \hline 0 & & & & 5 \end{array}\right] \qquad \left[\begin{array}{ccc|cc} 2 & 1 & 0 & & \\ 0 & 2 & 1 & & 0 \\ 0 & 0 & 2 & & \\ \hline & & & 5 & 1 \\ & 0 & & 0 & 5 \end{array}\right]$$

$$\left[\begin{array}{ccc|cc} 2 & 1 & 0 & & \\ 0 & 2 & 1 & & 0 \\ 0 & 0 & 2 & & \\ \hline & & & 5 & \\ & 0 & & & 5 \end{array}\right] \qquad \left[\begin{array}{cc|c|cc} 2 & 1 & & & \\ 0 & 2 & & & 0 \\ \hline & & 2 & & \\ \hline & 0 & & 5 & 1 \\ & & & 0 & 5 \end{array}\right]$$

The correct form, of course, depends on the endomorphism f. Of course, the eigenvalues of f are 2 and 5, but the number of independeny eigenvectors of each eigenvalue is unknown.

Clearly, as the size of the matrix and/or the number of eigenvalues increase, the problem of finding all possible Jordan forms becomes more difficult.

Example 3. Let A be the 3 x 3 matrix

$$\begin{bmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{bmatrix}$$

The characteristic polynomial of A is $(x - 2)^2(x + 1)$. Either this is the minimal polynomial, in which case A is similar to

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

or the minimal polynomial is $(x - 2)(x + 1)$, in which case A is similar to

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} .$$

A is similar to this diagonal matrix if and only if $(A - 2I)(A + I) = 0$ (since A must satisfy its minimal polynomial). But $(A - 2I)(A + I) =$

$$\begin{bmatrix} 0 & 0 & 0 \\ 3a & 0 & 0 \\ ac & 0 & 0 \end{bmatrix}$$

which is zero if and only if $a = 0$. So A is similar to a diagonal matrix if and only if $a = 0$.

Example 4. If A is the matrix

$$\begin{bmatrix} 2 & 2 & -1 \\ -1 & -1 & 1 \\ -1 & -2 & 2 \end{bmatrix}$$

then the characteristic polynomial of A is $(x - 1)^3$. By solving the equation $(A - I)X = 0$, where X is a 3 x 1 matrix, then it is found that the two independent eigenvectors are $X_1 = (1, 0, 1)$

and $X_2 = (0, 1, 2)$. Since the geometric multiplicity of 1 is 2 and the algebraic multiplicity of 1 is 3, then A is similar to the Jordan matrix

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

Example 5. Let

$$A = \begin{bmatrix} 5 & 4 & 3 \\ -1 & 0 & -3 \\ 1 & -2 & 1 \end{bmatrix}.$$

Then the characteristic polynomial of A is $(x + 2)(x - 4)^2$. Solving the equation $(A + 2I)X = 0$ and $(A - 4I)X = 0$, then the only two independent eigenvectors are $X_1 = (1,-1,-1)$ and $X_2 = (1, -1, 1)$. Solving the equation $(A - 4I)X_3 = X_2$, the vector $X_3 = (0, 1,-1) + aX_2$ in the null space $n(A - 4I)$ is obtained. If a is taken to be zero, then $Q^{-1}AQ = J$, where

$$Q = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$$

$$Q^{-1} = \frac{1}{12} \begin{bmatrix} 0 & -1 & -1 \\ 2 & 1 & 1 \\ 2 & 2 & 0 \end{bmatrix}$$

$$J = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

As was noted earlier, a square matrix A is similar to a matrix in the Jordan canonical form if the elements of A are from an algebraically closed field. Naturally, one such field is the complex field. Many times, however, one does not wish to work over the complex field. This problem may be overcome. If A is a complex matrix whose eigenvalues are real, then A is similar to a matrix whose entries are real, since the entries in the Jordan form of A are zero, one, and the eigenvalues of A.

Example 5. Let A be a complex square matrix. If P and Q are the characteristic polynomials of A and A^T (transpose of A) respectively, then $P = \det(A - xI) = \det (A - xI)^T = \det (A^T - xI^T) = \det (A^T - xI) = Q$. So A and A^T have the same characteristic polynomial. Also, it follows from the fact that $(A^T)^n = (A^n)^T$ and $(aA + bB)^T = a(A^T) + b(B^T)$, that A and A^T satisfy the same polynomials, so A and A^T have the same minimal polynomial. Therefore, the eigenvalues of A are the same as the eigenvalues of A^T, and these eigenvalues the same geometric multiplicity. But more can be said.

Suppose

$$\begin{bmatrix} a_i & 1 & . & . & 0 \\ 0 & a_i & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & a_i \end{bmatrix}$$

is an m x m Jordan block of the Jordan matrix similar to A. Then if P_1 is the matrix

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

then it is a straight calculation that $P_i J_i P_i^{-1} = J_i^T$. (Note also that P_i^{-1} is P_i.) If J is the Jordan matrix of A where

$$J = \begin{bmatrix} J_1 & 0 & . & . & 0 \\ 0 & J_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & J_r \end{bmatrix}$$

then $PJP^{-1} = J^T$ where

$$P = \begin{bmatrix} P_1 & 0 & . & . & 0 \\ 0 & P_2 & . & . & 0 \\ . & . & & & . \\ . & . & & & . \\ 0 & 0 & . & . & P_r \end{bmatrix}$$

Hence J is similar to its transpose. From the fact that similarity of matrices is an equivalence relation, then it follows that A is similar to A^T. As a consequence, the geometric multiplicity of each eigenvalue a_i of A is the same as the geometric multiplicity of the eigenvalue a_i of A^T.

Linear differential equations with constant coeffients provide a nice illustration of the use of the Jordan canonical form. Let $a_o, \ldots, a_{n-1}$ be complex numbers and let V be the space all n

times differentiable functions f on an interval of the real line which satisfy the differential equation

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1}f}{dx^{n-1}} + \ldots\ldots + a_1\frac{df}{dx} + a_0 f = 0.$$

Let D be the differential operator. Then V is invariant under D, because V is the null space P(D), where

$$P = x^n + a_{n-1}x^{n-1} + \ldots\ldots + a_1 x + a_0.$$

What is the Jordan canonical form of the differential operator on V?

Let $c_1, \ldots, c_k$ be the distinct complex roots of P, that is $P = (x - c_1)^{r_1}\ldots(x - c_k)^{r_k}$. Let $V_i = n(D - c_i I)^{r_i}$, that is, the set of solutions to the differential equation $(D - c_i I)^{r_i}_f = 0$. As was noted in Chapter III, $V = V_1 \oplus \ldots \oplus V_k$. Let N_i be the restriction of $D - c_i I$ to V_i. The Jordan canonical form for the operator D (on V) is then determined by the rational canonical forms for the nilpotent operators $N_1, \ldots, N_k$ on the spaces $V_1, \ldots, V_k$.

What must be determined now is the form (for the various c_i's) for the operator $N = D - cI$ on the space V_c which consists of the solutions of the equation $(D - cI)^r_f = 0$. (For simplicity, the subscripts are dropped while concerning the form pertaining to one of the roots of P.) The number of nilpotent blocks associated with c is the nullity of N, that is, the dimension of the eigenspace associated with the eigenvalue c. That dimension is one, because any function which satisfies the differential equation $D_f = cf$ is a

scalar multiple of the exponential function $h(x) = e^{cx}$. Therefore, the operator N (on V_c) has a cyclic vector. A good choice of this vector is $g = x^{r-1}h$:

$$g(x) = x^{r-1}e^{cx}.$$

This gives

$$Ng = (r - 1)x^{r-2}h$$

$$\vdots$$

$$N^{r-1}g = (r - 1)!h.$$

Thus, the Jordan canonical form of D (on V) is the direct sum of k elementary Jordan matrices, one for each root c_i.

The following is another example of the use of the Jordan canonical form. Any homogeneous differential equation with constant coefficients can be written in the form

$$z_1' = \frac{dz_1}{dt} = a_{11}z_1 + \dots + a_{1n}z_n$$

$$z_2' = \frac{dz_2}{dt} = a_{21}z_1 + \dots + a_{2n}z_n$$

$$\vdots$$

$$z_n' = \frac{dz_n}{dt} = a_{n1}z_1 + \dots + a_{nn}z_n$$

which can be written in matrix notation as

$$Z' = AZ \qquad (*)$$

$$\text{where } Z' = \begin{bmatrix} z_1' \\ \cdot \\ \cdot \\ \cdot \\ z_n' \end{bmatrix} \quad \text{and } Z = \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix}.$$

A is similar to a Jordan matrix, say $Q^{-1}AQ = J$. By the change of variables $Q^{-1}Z = Y$, the equation (*) becomes $Y' = JY$, an equivalent equation involving a much simpler matrix. This matrix can be solved, block-by-block, and the pieces can be put back together and the variable changed back to obtain a solution of the original differential equation.

The elementary nature of the Jordan canonical form of a matrix often simplifies proofs. For a more theoretical example than the above, the reader may see the proofs of several theorems in Russell Merris' paper "A Generalization of the Associated Transformation." in Linear Algebra and its Applications, vol. 4, no. 3, 1971, pp. 393-406.

[1]Dennis B. Ames, Fundamentals of Linear Algebra, Scranton, International Textbook Company, 1970.

[2]Earl A. Coddingon and Norman Levinson, Theory of Ordinary Differential Equations, New York, McGraw-Hill, 1955.

[3]Charles W. Curtis, Linear Algebra, an Introductory Approach, Boston, Allyn and Bacon, 1968.

[4]Jimmie D Gilbert, Elements of Linear Algebra, Scranton, International Textbook Company, 1970.

[5]Kenneth Hoffman and Ray Kunze, Linear Algebra, Englewood Cliffs, Prentice-Hall, 1971.

[6]Nathan Jacpbson, Lectures in Linear Algebra, Princeton, D. Van Nostrand Company, 1953.

[7]Serge Lange, Linear Algebra, New York, Addison-Wesley, 1968.

[8]John T. Moore, Elements of Linear Algebra and Matrix Theory, New York, McGraw-Hill, 1968.

[9]David C. Murdoch, Linear Algebra, New York, John Wiley and Sons, 1970.

[10]Walter Nef, Linear Algebra, trans., J. C. Ault, Berkshire, McGraw-Hill, 1966.

[11]Evar D. Nering, Linear Algebra and Matrix Theory, New York, John Wiley and Sons, 1963.

[12]Ben Noble, Applied Linear Algebra, Englewood Cliffs, Prentice-Hall, 1969.

[13]Malcom F. Smiley, Algebra of Matrices, Boston, Allyn and Bacon, 1965.

[14]R. R. Stoll and E. J. Wong, Linear Algebra, New York, Academic Press, 1968.

[15]Paul A. White, Linear Algebra, Belmont, Dickinson Publishing Company, 1966.

[16]Samuel F. Wolfenstien, Linear Algebra and Differential Equations, San Francisco, Holen-Day, 1969.