MATHEMATICS 217
NOTES

PART I – THE JORDAN CANONICAL FORM

The characteristic polynomial of an $n \times n$ matrix A is the polynomial $\chi_A(\lambda) = \det(\lambda I - A)$, a monic polynomial of degree n; a monic polynomial in the variable λ is just a polynomial with leading term λ^n. Note that similar matrices have the same characteristic polynomial, since $\det(\lambda I - C^{-1}AC) = \det C^{-1}(\lambda I - A)C = \det(\lambda I - A)$. It is possible to substitute the matrix A itself for the variable λ in any polynomial, when the coefficients are viewed as being scalars times the identity matrix.

Theorem 1 (Cayley-Hamilton Theorem) $\chi_A(A) = 0$.

Proof: See page 203 of Apostol.

It may be that $f(A) = 0$ for a polynomial $f(\lambda)$ of degree less than n, where A is an $n \times n$ matrix.

Theorem 2 *For any $n \times n$ matrix A there is a unique monic polynomial $m_A(\lambda)$ of minimal degree such that $m_A(A) = 0$; and $f(A) = 0$ for another polynomial $f(\lambda)$ if and only if $f(\lambda) = m_A(\lambda) \cdot q(\lambda)$ for some polynomial $q(\lambda)$.*

Proof: Let $m_A(\lambda)$ be a monic polynomial of minimal degree such that $m_A(A) = 0$; there is such a polynomial, since in particular $\chi_A(A) = 0$ by the Cayley-Hamilton Theorem. For any polynomial $f(\lambda)$ the customary division algorithm for polynomials shows that $f(\lambda) = m_A(\lambda)q(\lambda) + r(\lambda)$ for some polynomials $q(\lambda)$ and $r(\lambda)$, where the degree of $r(\lambda)$ is strictly less than the degree of $m_A(\lambda)$. If $f(A) = 0$ then $0 = f(A) = m_A(A)q(A) + r(A) = r(A)$, and since $r(\lambda)$ has lower degree than that of $m_A(\lambda)$ follows that $r(\lambda) = 0$. Consequently $f(\lambda) = m_A(\lambda)q(\lambda)$; and if $f(\lambda)$ is of the same degree as $m_A(\lambda)$ then $q(\lambda) = 1$.

The polynomial $m_A(\lambda)$ is called the *minimal polynomial* of the matrix A. Note that if $p(A) = 0$ for a polynomial $p(\lambda)$ then $p(C^{-1}AC) = C^{-1}p(A)C = 0$ for any nonsingular matrix C; hence similar matrices have the same minimal polynomial, and the characteristic and minimal polynomials of a linear transformation T thus can be defined to be the corresponding polynomials of any matrix representing T. As a consequence of the preceding theorem, the minimal polynomial $m_A(\lambda)$ divides the characteristic polynomial $\chi_A(\lambda)$ for any matrix A; that will be indicated by writing $m_A(\lambda)|\chi_A(\lambda)$.

EXAMPLES:

$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$: $\chi_A(x) = (x-2)^2,\ m_A(x) = (x-2)$

$B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$: $\chi_B(x) = (x-2)^2,\ m_B(x) = (x-2)^2$

If A and B are square matrices, $A \oplus B$ is defined to be the square matrix

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

and is called the *direct sum* of the matrices A and B.

Theorem 3 *If $C = A \oplus B$ where A and B are square matrices then*
(1) *the characteristic polynomials of these matrices satisfy*

$$\chi_C(\lambda) = \chi_A(\lambda) \cdot \chi_B(\lambda);$$

(2) *the minimal polynomial $m_C(\lambda)$ is the least common multiple of the minimal polynomials $m_A(\lambda)$ and $m_B(\lambda)$, that is, is the monic polynomial $m_C(\lambda)$ of lowest degree such that $m_A(\lambda)|m_C(\lambda)$ and $m_B(\lambda)|m_C(\lambda)$.*

Proof: The first result follows from the observation that

$$\det\Big((\lambda I - A) \oplus (\lambda I - B)\Big) = \det(\lambda I - A) \cdot \det(\lambda I - B).$$

For the second result, if $C = A \oplus B$ and $f(\lambda)$ is any polynomial then

$$\begin{aligned} f(C) = 0 &\iff f(A) = 0 \text{ and } f(B) = 0 \\ &\iff m_A(\lambda)|f(\lambda) \text{ and } m_B(\lambda)|f(\lambda). \end{aligned}$$

A vector space V is called a *direct sum* of vector subspaces $X_i \subset V$ for $1 \le i \le m$ if every vector $v \in V$ can be written uniquely as a sum $v = x_1 + x_2 + \cdots + x_m$ of vectors $x_i \in X_i$; this will be indicated by writing

$$V = X_1 \oplus X_2 \oplus \cdots \oplus X_m.$$

For example, if $e_1, \ldots, e_n$ is a basis for V and $L(e_i)$ is the set of all scalar multiples of the vector e_i (the span of e_i, a one-dimensional subspace of V), then $V = L(e_1) \oplus \cdots L(e_n)$. The necessary and sufficient condition that $V = X \oplus Y$ for a finite-dimensional vector space V is that $\dim V = \dim X + \dim Y$ and that $X \cap Y = 0$.

If $T : V \longrightarrow V$ is a linear mapping from a vector space V to itself, a subspace $W \subset V$ is T-*invariant* if $TW \subset W$. For example, the kernel and image of T are T-invariant; and any eigenspace $E(\lambda)$ of the mapping T is T-invariant.

Suppose that $T : V \longrightarrow V$ is a linear mapping from a vector space V to itself and that $V = X \oplus Y$ where X and Y are T-invariant subspaces of V. Since X is T-invariant, the restriction of T to the subspace X is a well defined linear mapping $T|X : X \longrightarrow X$, and similarly for the restriction $T|Y$. If $\{e_1, \dots, e_m\}$ is a basis for X, in terms of which the transformation $T|X : X \longrightarrow X$ is represented by a matrix A, and if $\{f_1, \dots, f_n\}$ is a basis for Y, in terms of which the transformation $T|Y : Y \longrightarrow Y$ is represented by a matrix B, then $\{e_1, \dots, e_m, f_1, \dots, f_n\}$ is a basis for V and in terms of this basis the matrix describing the linear transformation T is $A \oplus B$. Conversely for the linear transformation T defined by a matrix $A \oplus B$, where A is an $m \times m$ matrix and B is an $n \times n$ matrix, the subspaces X spanned by the basis vectors $e_1, \dots, e_m$ and Y spanned by the basis vectors $e_{m+1}, \dots, e_{m+n}$ are invariant subspaces, on which the action of T is represented by the matrices A and B, and $V = X \oplus Y$. Here e_i is the vector with entry 1 in the i-th place and entries 0 elsewhere, one of the standard basis vectors for the vector space $\mathbb{R}^{m+n}$.

If $T : V \longrightarrow V$ is a linear transformation from a vector space V to itself, a vector $v \in V$ is said to be a *cyclic vector* for T if the vectors $v, Tv, T^2v, \dots$ span the vector space V. If v is a cyclic vector for a transformation T on a finite-dimensional vector space V, there is an integer n such that the vectors $v, Tv, \dots, T^{n-1}v$ are linearly independent but the vectors $v, Tv, \dots, T^n v$ are linearly dependent; consequently there is an identity of the form

$$a_0 v + a_1 Tv + \cdots + a_{n-1} T^{n-1} v + T^n v = 0 \tag{1}$$

for some scalars $a_1, \dots, a_{n-1}$. It follows by induction that $T^m v$ is in the space spanned by the vectors $v, Tv, \dots, T^{n-1}v$ for every $m \geq n$ and hence that the vectors $e_i = T^{i-1}v$ for $1 \leq i \leq n$ are a basis for V. Note that $Te_i = e_{i+1}$ for $1 \leq i \leq n-1$ and that $Te_n = -a_0 e_1 - a_2 e_2 - \cdots - a_{n-1} e_{n-1}$, and therefore that the matrix A representing the linear mapping T in terms of this basis has the form

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & 1 & \cdots & 0 & -a_3 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

This matrix is called the *companion matrix* of the polynomial $p(\lambda) = a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$. Conversely if A is the companion matrix to a polynomial $p(\lambda)$ as above then $Ae_i = e_{i+1}$ for $1 \leq i \leq n-1$ and $Ae_n = -a_0 e_1 - a_2 e_2 - \cdots - a_{n-1} e_{n-1}$, where e_i are the standard basis vectors in $\mathbb{R}^n$; hence e_1 is a cyclic vector for the linear mapping defined by this matrix.

Theorem 4 *If A is the companion matrix of the polynomial $p(\lambda)$ then*

$$\chi_A(\lambda) = m_A(\lambda) = p(\lambda).$$

Proof: The identity (1) for the linear transformation $A = T$ and the cyclic vector $v = e_1$ shows that $p(A)e_1 = 0$; and then $0 = A^{i-1}p(A)e_1 = p(A)A^{i-1}e_1 = p(A)e_i$ for $1 \leq i \leq n$, so that $p(A) = 0$ since the vectors e_i are a basis. If $f(\lambda) = b_0 + b_1\lambda + \cdots + b_m\lambda^m$ is a polynomial of degree $m < n$ such that $f(A) = 0$ then similarly

$$\begin{aligned} 0 = f(A)e_1 &= b_0e_1 + b_1Ae_1 + \cdots + b_mA^me_1 \\ &= b_0e_1 + b_1e_2 + \cdots + b_me_{m+1}; \end{aligned}$$

since the vectors $e_1, e_2, \ldots, e_m$ are linearly independent for any $m < n$ it follows that $b_i = 0$ and hence that $f(\lambda) = 0$. Therefore there are no polynomials $f(\lambda)$ of degree $m < n$ for which $f(A) = 0$, so that $p(\lambda)$ must be the polynomial of least degree for which $p(\lambda) = 0$ and hence $p(\lambda) = m_A(\lambda)$. Since $m_A(\lambda)|\chi_A(\lambda)$ and $\chi_A(\lambda)$ also is of degree n it further follows that $p(\lambda) = \chi_A(\lambda)$.

There do not necessarily exist any cyclic vectors at all for a linear transformation T. The preceding theorem shows that if there is a cyclic vector for T then $m_T(\lambda) = \chi_T(\lambda)$; the converse is also true, but will not be needed here. Actually it is more useful to write a matrix A as a direct sum of companion matrices to the simplest possible polynomials; over the complex numbers these are just the polynomials of the form $p(\lambda) = (\lambda - a)^n$. In this case the companion matrix takes a particularly convenient form in terms of another basis. In place of the basis $e_i = T^{i-1}(v)$ for the cyclic vector v consider the vectors $f_i = (T - aI)^{n-i}v$. Note that $f_n = v = e_1,\ f_{n-1} = (T - aI)v = e_2 - ae_1,\ f_{n-2} = (T - aI)f_{n-1} = (T - aI)(e_2 - ae_1) = e_3 - 2ae_2 + a^2e_1$, and so on; so it is clear that the basis vectors e_i can be expressed as linear combinations of the vectors f_i and hence that $f_1, \ldots, f_n$ is also a basis for V. For this basis

$$\begin{aligned} (T - aI)f_1 &= (T - aI)(T - aI)^{n-1}v \\ &= (T - aI)^n v = p(T)v = 0 \end{aligned}$$

so that

$$Tf_1 = af_1 \tag{2}$$

and hence f_1 is an eigenvector for the eigenvalue a_1; and for $2 \leq i \leq n$

$$\begin{aligned} (T - aI)f_i &= (T - aI)(T - aI)^{n-i}v \\ &= (T - aI)^{n-i+1}v = f_{i-1} \end{aligned}$$

so that

$$Tf_i = f_{i-1} + af_i \quad \text{for} \quad 2 \leq i \leq n. \tag{3}$$

It follows from equations (2) and (3) that the matrix B representing the transformation T in terms of the basis $f_1, \ldots, f_n$ has the form

$$B(n; a) = \begin{pmatrix} a & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a & 1 & \cdots & 0 & 0 \\ & & & & \cdots & & \\ & & & & \cdots & & \\ 0 & 0 & 0 & 0 & \ldots & a & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & a \end{pmatrix}.$$

This matrix is called the *Jordan Block* of size n with eigenvalue a.

Theorem 5 *For the Jordan Block $B(n, a)$*

$$\chi_B(\lambda) = m_B(\lambda) = (\lambda - a)^n;$$

the matrix B has the single eigenvalue a, and the eigenspace $E(a)$ is one dimensional.

Proof: The Jordan block is similar to the companion matrix for the polynomial $p(\lambda) = (\lambda - a)^n$, since it arises from a change of basis, so it has the same characteristic and minimal polynomials as the companion matrix; hence $\chi_B(\lambda) = m_B(\lambda) = (\lambda - a)^n$ by the preceding theorem. The matrix $aI - B$ has entries 1 just above the main diagonal and 0 otherwise, so $(aI - B)v = 0$ only for scalar multiples of the vector

$$v = \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}.$$

A Jordan block of size $n > 1$ is thus not diagonalizable; but the Jordan blocks provide normal forms for arbitrary matrices.

Theorem 6 *For any linear transformation $T : V \longrightarrow V$ of a finite-dimensional complex vector space V to itself, there is a basis of V in terms of which T is represented by a matrix*

$$A = B(n_1; a_1) \oplus B(n_2; a_2) \oplus \cdots \oplus B(n_k; a_k)$$

where $B(n_i; a_i)$ are Jordan blocks.

Proof: The proof will be by induction on $n = \dim V$. When $n = 1$ the result is trivial, since any 1×1 matrix is itself a Jordan block. Next consider a linear transformation $T : V \longrightarrow V$ where $\dim V = n$ and assume that the result holds for all vector spaces W with $\dim W < n$.

First suppose that T is singular, so has a nontrivial kernel. Since $W = T(V) \subset V$ is a T-invariant subspace with $\dim W = m < n$ it follows from the inductive hypothesis that there is a basis of W for which $T|W$ is represented by a matrix

$$A' = B(n_1; a_1) \oplus B(n_2; a_2) \oplus \cdots \oplus B(n_k; a_k)$$

where $B(n_i; a_i)$ are Jordan blocks; this corresponds to a direct sum decomposition

$$W = X_1 \oplus X_2 \oplus \cdots \oplus X_k$$

where $X_i \subset W \subset V$ are T-invariant subspaces and $T|X_i : X_i \longrightarrow X_i$ is represented by the matrix $B(n_i; a_i)$. Each subspace X_i contains a single eigenvector v_i with eigenvalue a_i, by the preceding theorem. Let $K \subset V$ be the kernel of the linear transformation T, so $\dim K = n - m$, and suppose that $\dim(K \cap W) = r \leq n - m$. There are r linearly independent vectors $v_i \in K \cap W$ such that $T(v_i) = 0$, which are eigenvectors of $T|W$ with eigenvalues 0; so it can be supposed further that these are just the eigenvectors $v_i \in X_i$ for $1 \leq i \leq r$. Extend these vectors to a basis $v_1, \ldots, v_r, u_{r+1}, \ldots, u_{n-m}$ for the subspace K. The vectors $u_{r+1}, \ldots, u_{n-m}$ span a T-invariant subspace $U \subset V$ with $\dim U = n - m - r$. On each of the subspaces X_i for $1 \leq i \leq r$ the transformation T is represented by a Jordan block $B(n_i; 0)$ in terms of a basis $e_{i1}, e_{i2}, \ldots, e_{in_i}$; so $Te_{i1} = 0$ and $e_{i1} = v_i$ is one of the basis vectors of the kernel K, while $Te_{ij} = e_{i\,j-1}$ for $2 \leq i \leq n_i$. That clearly shows that the vectors $e_{i2}, \ldots, e_{i\,n_i-1}$ are in the image $W = T(V)$, which of course they must be since $X_i \subset W$; but $e_{i\,n_i} \in W = T(V)$ as well, so there must be a vector $f_i \in V$ for which $e_{i\,n_i} = T(f_i)$. The vector f_i is not contained in X_i, since no linear combination of the vectors in X_i can have as its image under T the vector $e_{i\,n_i}$; hence $X'_i = X_i \oplus L(f_i)$ is a linear subspace of V with $\dim X'_i = \dim X_i + 1$, and x'_i is also a T-invariant subspace of V. Note that in terms of the basis $e_{i1}, e_{i2}, \ldots, e_{in_i}, f_i$ for X'_i the restriction $T|X'_i : X'_i \longrightarrow X'_i$ is represented by a Jordan block $B(n_i + 1; 0)$. The proof of this part of the theorem will be concluded by showing that

$$V = X'_i \oplus \cdots \oplus X'_r \oplus X_{r+1} \cdots \oplus X_k \oplus L(u_{r+1}) \oplus \cdots \oplus L(u_{n-m}).$$

The dimension count is correct, since $W = X_1 \oplus \cdots \oplus X_k$ has dimension m, each X'_i increases the dimension by 1 for an additional r, and there are $n - m - r$ of the vectors u_i. It is hence enough just to show that there can be no nontrivial linear relation of the form

$$0 = \sum_{i=1}^{k} c_i x_i + \sum_{i=1}^{r} c' f_i + \sum_{i=1}^{n-m} c''_i u_i$$

for any vectors $x_i \in X_i$. Applying T to this relation yields

$$0 = \sum_{i=1}^{k} c_i T x_i + \sum_{i=1}^{r} c' e_{i\,n_i}$$

since $Tu_i = 0$, and since $V = X_1 \oplus \cdots \oplus X_k$ this identity can hold only when $c_i T x_i + c'_i e_{i\,n_i} = 0$ for each i; but $e_{i\,n_i}$ cannot be the image of any vector in X_i, so that $c_i = c'_i = 0$. Finally $\sum_{i=r+1}^{n-m} c'' u_i = 0$, and since the vectors u_i are linearly independent it follows that $c''_i = 0$ as well.

For the case in which T is nonsingular, the linear transformation T has an eigenvalue λ over the complex numbers so the transformation $T - \lambda I$ will be singular. The preceding part of the proof shows that $T - \lambda I$ can be represented by a matrix A that is a sum of Jordan blocks; and then T clearly is represented by the matrix $A + \lambda I$, which is easily seen also to be a sum of Jordan blocks. That concludes the proof.

An equivalent statement of the preceding theorem is that any matrix is similar over the complex numbers to a matrix that is a direct sum of Jordan blocks; this is often called the Jordan normal or Jordan canonical form for the matrix. These blocks actually are determined uniquely and characterize the matrix up to similarity.

Theorem 7 *Over the complex numbers, any matrix is similar to a unique direct sum of Jordan blocks.*

Proof: By the preceding theorem, over the complex numbers any matrix is similar to one of the form

$$A = B(n_1; a_1) \oplus B(n_2; a_2) \oplus \cdots \oplus B(n_k; a_k) \tag{4}$$

where $B(n_i; a_i)$ are Jordan blocks. The distinct eigenvalues of A are the distinct roots of the characteristic polynomial $\chi_A(\lambda)$, so are uniquely determined. Each Jordan block corresponds to a single eigenvector, so the number of distinct Jordan blocks for which $a_i = a$ for any particular eigenvalue a is also uniquely determined. It is thus only necessary to show that the sizes n_i of the Jordan blocks associated to an eigenvalue a are also uniquely determined.

Suppose therefore that $a_1 = \cdots = a_r = a$ but that $a_i \neq a$ for $i > r$, so a is an eigenvector for which $\dim E(a) = r$. For a Jordan block $B(n; a)$ the matrix $B(n; a) - aI$ is an $n \times n$ matrix with entries 1 above the main diagonal and zero elsewhere, so its kernel is $L(e_1)$; the matrix $(B(n; a) - aI)^2$ has entries 1 on the second line above the main diagonal and zero elsewhere, so its kernel is $L(e_1) \oplus L(e_2)$; and so on. Thus the dimension of the kernel of the linear mapping $(B(n; a) - aI)^j$ is j if $j \leq n$ but is n if $j \geq n$; and hence

$$\dim\ker(B(n;a) - aI)^j - \dim\ker(B(n;a) - aI)^{j-1} = \begin{cases} 1 & \text{if} \quad j \leq n \\ \\ 0 & \text{if} \quad j > n. \end{cases}$$

Now the kernel of $(A - aI)^j = (B(n_1; a_1) - aI)^j \oplus \cdots \oplus (B(ns_k; a_k) - aI)^j$ is the direct sum of the kernels of the separate blocks, so

$$\dim\ker A^j - \dim\ker A^{j-1} = \text{number of blocks} \;\; B(n_i; a_i) \text{ for which } n_i \geq j.$$

That determines the number of blocks $B(n;a)$ of any size n uniquely just in terms of the matrix A, and concludes the proof.

EXAMPLE

For the matrix

$$A = \begin{pmatrix} -5 & 9 & -4 \\ -4 & 8 & -4 \\ 1 & 0 & 0 \end{pmatrix}$$

it is a simple calculation to verify that $\det(\lambda I - A) = \lambda^3 - 3\lambda^2 + 4 = (\lambda+1)(\lambda-2)^2$; hence the eigenvalues are $-1, 2$ and the eigenvectors are solutions of the system of linear equations $(A - \lambda I)v = 0$. These systems can be solved by row reduction, so

$$\text{for } \lambda = -1: \left(\begin{array}{ccc|c} -4 & 9 & -4 & 0 \\ -4 & 9 & -4 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

$$\text{for } \lambda = 2: \left(\begin{array}{ccc|c} -7 & 9 & -4 & 0 \\ -4 & 6 & -4 & 0 \\ 1 & 0 & -2 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad v_2 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

There is only one eigenvector v_2 for the eigenvalue $\lambda = 2$, so there must be a Jordan block of the form $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$; thus in addition to the vector v_2 there must be another vector v_3 for which $Av_3 = 2v_3 + v_2$ or $(A - 2I)v_3 = v_2$. That vector also can be found by solving this last system of linear equations by row reduction, so

$$\left(\begin{array}{ccc|c} -7 & 9 & -4 & 2 \\ -4 & 6 & -4 & 2 \\ 1 & 0 & -2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -2 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

for any value x_3; in particular it is possible just to take $x_3 = 0$. With this choice the normal form for the matrix A is

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

in terms of the basis v_1, v_2, v_3; so for the matrix

$$C = (v_1\ v_2\ v_3) = \begin{pmatrix} -1 & 2 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

it follows that $AC = CB$ or $C^{-1}AC = B$.

The characteristic and minimal polynomials for any matrix can be read directly from the Jordan block normal form. For instance, for the matrix

$$A = B(2;1) \oplus B(3;1) \oplus B(4;2) \oplus B(5;3)$$

it follows that

$$\chi_A(\lambda) = (\lambda-1)^5(\lambda-2)^4(\lambda-3)^5, \quad m_A(\lambda) = (\lambda-1)^3(\lambda-2)^4(\lambda-3)^5.$$

Conversely if A is a 3×3 matrix for which $\chi_A(\lambda) = (\lambda-7)^2(\lambda-1)$ and $m_A(\lambda) = (\lambda-7)(\lambda-1)$ then A is similar to the matrix $B = B(1;1) \oplus B(1;7) \oplus B(1;7)$, explicitly

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

The normal form of a matrix is not always determined uniquely just by its characteristic and minimal polynomials, although there are only finitely many different possibilities up to order.

The exponentials of Jordan blocks are easy to calculate. A Jordan block has the form $B(n;a) = aI + N$ where the matrix N has entries 1 on the line above the main diagonal and 0 elsewhere. The matrix N^2 has entries 1 on the second line above the main diagonal and 0 elsewhere, N^3 has entries 1 on the third line above the main diagonal and 0 elsewhere, and so on. Since the matrices I and N commute

$$e^{B(n;a)t} = e^{aIt+Nt} = e^{at}e^{Nt} = e^{at}\begin{pmatrix} 1 & t & \frac{1}{2!}t^2 & \frac{1}{3!}t^3 & \cdots & \frac{1}{(n-1)!}t^{n-1} \\ 0 & 1 & t & \frac{1}{2!}t^2 & \cdots & \frac{1}{(n-2)!}t^{n-2} \\ 0 & 0 & 1 & t & \cdots & \frac{1}{(n-3)!}t^{n-3} \\ & & & & & \\ 0 & 0 & 0 & 0 & \cdots & t \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

PART 2 – INNER PRODUCT SPACES

In this part of these notes, suppose that V is a finite-dimensional vector space V over either the real or complex numbers, with an inner product (v, w).

Theorem 8 *For any linear transformation $T : V \longrightarrow V$ of a finite dimensional vector space V with an inner product there is a unique linear transformation $T^* : V \longrightarrow V$ such that $(T^*v, w) = (v, Tw)$ for any vectors $v, w \in V$.*

Proof: If $e_1, \ldots, e_n \in V$ is an orthonormal basis and if there is such a linear transformation T^* then

$$(T^*e_i, e_j) = (e_i, Te_j)$$

hence necessarily

$$T^*e_i = \sum_{j=1}^{n}(e_i, Te_j)e_j.$$

On the other hand this last formula does define a linear transformation T^* with the asserted properties on a basis for V, hence on all of V.

The linear transformation T^* is called the *adjoint* transformation to T.

Theorem 9 *The adjoint transformation has the following properties:*
(i) $(S + T)^* = S^* + T^*$
(ii) $(aT)^* = \overline{a}T^*$
(iii) $(ST)^* = T^*S^*$
(iv) $I^* = I$ *for the identity transformation* I
(v) $T^{**} = T$
(vi) $(Tv, w) = (v, T^*w)$ *for all* $v, w \in V$
(vii) *If T is represented by a matrix A in terms of an orthonormal basis for V then T^* is represented by the matrix $\overline{A^t}$ in terms of the same basis.*

Proof: The proof follows quite easily from the existence and uniqueness of the adjoint, the properties of the inner product, and the following observations, for any vectors $v, w \in V$.
(i) $(v, (S + T)w) = (v, Sw) + (v, Tw) = (S^*v, w) + (T^*v, w) = ((S^* + T^*)v, w)$
(ii) $(v, aTw) = \overline{a}(v, Tw) = \overline{a}(T^*v, w) = (\overline{a}T^*v, w)$
(iii) $(v, STw) = (S^*v, Tw) = (T^*S^*v, w)$
(iv) $(v, w) = (v, Iw) = (I^*v, w)$
(v) $(v, Tw) = (T^*v, w) = \overline{(w, T^*v)} = \overline{(T^{**}w, v)} = (v, T^{**}w)$
(vi) $(Tv, w) = \overline{(w, Tv)} = \overline{(T^*w, v)} = (v, T^*w)$
(vii) If $e_i \in V$ is an orthonormal basis then the matrix A representing T in terms of this basis has entries a_{ji} defined by $Te_i = \sum_j a_{ji}e_j$, and similarly the matrix A^* representing T^* has entries a^*_{ji} defined by $T^*e_i = \sum_j a^*_{ji}e_j$. Since the basis is orthonormal

$$a^*_{ji} = (T^*e_i, e_j) = (e_i, Te_j) = \overline{(Te_j, e_i)} = \overline{a}_{ij}$$

.

Theorem 10 *The following conditions on a linear transformation $T : V \longrightarrow V$ on a finite-dimensional real or complex vector space V with an inner product are equivalent:*
(i) $TT^* = T^*T = I$*;*
(ii) $(Tv, Tw) = (v, w)$ *for all* $v, w \in V$*;*
(iii) $||Tv|| = ||v||$ *for all* $v, w \in V$*.*

Proof:
(i)$\Longrightarrow$ (ii):
If (i) holds then $(Tv, Tw) = (v, T^*Tw) = (v, w)$ for all $v, w \in V$.
(ii)$\Longrightarrow$(i):
If (ii) holds then $(v, w) = (Tv, Tw) = (v, T^*Tw)$ for all $v, w \in V$, hence $T^*T = I$; and for a finite-dimensional vector space it follows from this that $TT^* = I$ as well.
(ii) $\Longrightarrow$(iii):
If (ii) holds then for the special case $w = v$ it follows that $||Tv||^2 = (Tv, Tv) = (v, v) = ||v||^2$, hence $||Tv|| = ||v||$ since both are positive.
(iii) $\Longrightarrow$ (ii):
Recall that $||v + w||^2 = (v + w, v + w) = (v, v) + (v, w) + (w, v) + (w, w) = ||v||^2 + 2\Re(v, w) + ||w||^2$. where $\Re z$ denotes the real part of a complex number z. If (iii) holds then $2\Re(Tv, Tw) = ||T(v + w)||^2 - ||Tv||^2 - ||Tw||^2 = ||v + w||^2 - ||v||^2 - ||w||^2|| = \Re(v, w)$, and that is condition (ii) in the real case. For the complex case note that $-\Im(Tv, Tw) = \Re i(Tv, Tw) = \Re(Tiv, Tw) = \Re(iv, w) = \Re i(v, w) = -\Im(v, w)$ where $\Im z$ denotes the imaginary part of a complex number z. The equalities for the real and imaginary parts separately show that (ii) holds.

TERMINOLOGY

A linear transformation T such that $T^* = T$ is called *self adjoint*; for real vector spaces a self-adjoint transformation is also called *symmetric*, while for complex vector spaces a self-adjoint transformation is also called *Hermitian.* A matrix representing a symmetric transformation in terms of an orthonormal basis is a symmetric matrix in the usual sense by Theorem 9(vii), while a matrix representing a Hermitian transformation satisfies $\overline{A^t} = A$. A linear transformation T such that $T^* = -T$ is called *skew-symmetric* in the real case and *skew-Hermitian* in the complex case. A linear transformation satisfying any of the equivalent conditions of Theorem 10 is called *orthogonal* in the real case and *unitary* in the complex case; the representative matrices in terms of an orthonormal basis satisfy the conditions $A\,A^t = A^t\,A = I$ in the real case and $A\,\overline{A^t} = \overline{A^t}\,A$ in the complex case. The conditions of Theorem 10 amount to the conditions that orthogonal or unitary linear transformations preserve distances and angles. A linear transformation T is called *normal* if $TT^* = T^*T$. This is a somewhat less intuitive notion, the significance of which will become apparent later. Note that any symmetric, Hermitian, skew-symmetric, skew-Hermitian, orthogonal or unitary transformation is normal; but there are also other normal linear transformations. The matrices of these special types of normal linear transformations have eigenvalues of special forms as well.

Theorem 11 *If a is an eigenvalue of a linear transformation $T : V \longrightarrow V$ then:*
(i) *a is real whenever $T = T^*$;*
(ii) *a is purely imaginary whenever $T = -T^*$;*
(iii) *$|a| = 1$ whenever $TT^* = I$.*

Proof: Suppose that $Tv = av$ for some vector $v \in V$ for which $||v|| = 1$.
(i) If $T = T^*$ then $a = (Tv, v) = (v, T^*v) = (v, Tv) = \overline{(Tv, v)} = \overline{a}$.
(ii) If $T^* = -T$ then $a = (Tv, v) = (v, T^*v) = -(v, Tv) = -\overline{(Tv, v)} = -\overline{a}$.
(iii) If $T^*T = I$ then $|a|^2 = (av, av) = (Tv, Tv) = (v, T^*Tv) = (v, v) = 1$.

Theorem 12 *If $T : V \longrightarrow V$ is a normal linear transformation then*
(i) $\ker T = \ker T^*$.
(ii) *If $Tv = av$ then $T^*v = \overline{a}v$.*
(iii) *If $Tv_1 = a_1v_1$, $Tv_2 = a_2v_2$ for some vectors $v_i \in V$ and distinct complex numbers a_1, a_2 then $(v_1, v_2) = 0$.*

Proof:
(i) If $Tv = 0$ then $0 = (Tv, Tv) = (v, T^*Tv) = (v, TT^*v) = (T^*v, T^*v)$ so that $T^*v = 0$.
(ii) Note that $T - aI$ is normal whenever T is normal, and that $(T - aI)^* = T^* - \overline{a}I$; then from (i) it follows that $(T - aI)v = 0$ if and only if $(T^* - \overline{a}I)v = 0$.
(iii) Since $a_1(v_1, v_2) = (a_1v_1, v_2) = (Tv_1, v_2) = (v_1, T^*v_2) = (v_2, \overline{a_2}v_2) = a_2(v_1, v_2)$ it follows that $(a_1 - a_2)(v_1, v_2) = 0$ so that $(v_1, v_2) = 0$ if $a_1 \neq a_2$.

It follows from the preceding theorem that a normal linear transformation T and its adjoint T^* have the same eigenvectors, although with complex conjugate eigenvalues; and that the eigenspaces for distinct eigenvalues are orthogonal.

Theorem 13 (Spectral Theorem) *If $T : V \longrightarrow V$ is a normal linear transformation on a finite-dimensional complex vector space V then V has an orthonormal basis of eigenvectors for T.*

Proof: The proof will be by induction on the dimension of the space V. If $\dim V = 1$ the result is trivial; so consider a space V with $\dim V = n$ and assume that the theorem holds for all spaces of lower dimension.

Over the complex numbers the transformation T will have an eigenvalue a, with eigenspace W. If $W = V$ then every vector in V is an eigenvector with eigenvalue a, so in particular every orthonormal basis consists of eigenvectors. Suppose therefore that $W \neq V$, and consider the orthogonal decomposition $V = W \oplus W^{\perp}$. The subspace W of course is T-invariant, as an eigenspace of T. The subspace $W^\perp$ is then T^*-invariant; indeed if $v \in W^\perp$ then for all $w \in W$ it follows that $(T^*v, w) = (v, Tw) = 0$, since $Tw \in W$, hence $T^*v \in W^\perp$ as well. The subspace W is also T^*-invariant, since it is the eigenspace of T^* for the eigenvalue $\overline{a}$ by Theorem 12(ii); and since $T^{**} = T$ it follows by the argument of the preceding sentence that the subspace $W^\perp$ is also T-invariant. The restrictions of T and T^* to the subspace $W^\perp$ also satisfy $(T^*v, w) = (v, Tw)$ for any $v, w \in W^\perp$, so that the restriction of T^* to $W^\perp$ is the adjoint to the restriction of T to that subspace; and therefore the restriction of T to that subspace is also normal. It follows from the induction hypothesis that there is an orthonormal basis for $W^\perp$ that consists of eigenvectors of T; and when that basis is extended to an orthonormal basis of V by adjoining enough additional vectors in W there results an orthonormal basis of V, which concludes the proof.

If there is an orthonormal basis for V consisting of eigenvectors for T then the basis also consists of eigenvectors for T^* by Theorem 12(ii), so both T and T^* are represented by diagonal matrices; and it follows that $TT^* = T^*T$, so that T must be normal. Thus the normality of T is both necessary and sufficient for there to exist an orthonormal basis of eigenvectors for T. The spectral theorem can be restated in terms of matrices as follows.

Corollary 14 *If A is a matrix such that $A\overline{A^t} = \overline{A^t}A$ then there is a unitary matrix U such that $U^*AU = D$ is a diagonal matrix.*

Proof: If A is viewed as the representation of a linear transformation T in terms of an orthonormal basis $e_1, \ldots, e_n$ then the hypothesis is that T is normal; hence there is an orthonormal basis $f_1, \ldots, f_n$ consisting of eigenvectors for T, so T is represented by a diagonal matrix in terms of this basis. Now the entries of the matrix A are $a_{ji} = (Te_i, e_j)$ and the entries of the matrix D are $d_{ji} = (Tf_i, f_j)$. Thus if $f_i = \sum_j u_{ki} e_k$ then

$$d_{ji} = (T(\sum_k u_{ki}e_k), \sum_l u_{lj}e_l) = \sum_{kl} u_{ki}\overline{u}_{lj}(Te_k, e_l) = \sum_{kl} \overline{u}_{lj} a_{lk} u_{ki}$$

or in matrix notation $D = U^*AU$. Since f_i and e_i are both orthonormal bases it follows further that

$$\delta^i_j = (f_i, f_j) = (\sum_k u_{ki}e_k, \sum_l u_{lj}e_j) = \sum_{kl} u_{ki}\overline{u}_{lj}(e_k, e_l) = \sum_k u_{ki}\overline{u}_{kj}$$

where δ^i_j are the entries of the identity matrix (also called Kronecker's symbol), or in matrix terms $I = U^t\,\overline{U}$; since the identity matrix is real this is the same as $I = UU^*$, so the matrix U is unitary.

The hypothesis that V is a complex vector space is essential, since linear transformations of real vector spaces need not have any eigenvectors at all. However for special real linear transformations the spectral theorem does hold.

Theorem 15 (Real Spectral Theorem) *If $T : V \longrightarrow V$ is a symmetric transformation of a real vector space V then V has an orthonormal basis of eigenvectors for T.*

Proof: If A is a matrix representing T in terms of an orthonornal basis for V then A is a real symmetric matrix. It can be viewed as a complex matrix, and as such it is Hermitian; so from the ordinary spectral theorem it follows that there is a unitary matrix U such that $U^*AU = D$ is diagonal. Now $D^* = U^*A^*U = U^*AU = D$ since A is real and symmetric, so the diagonal matrix D must be real; consequently A has all real eigenvalues. For real matrixes with all eigenvalues real the same argument as in the proof of the complex spectral theorem works, so the theorem holds in this case as well.

Corollary 16 *If A is a real symmetric matrix then there exists a real orthogonal matrix O such that $O^t\,A\,O = D$ is diagonal.*

Proof: The same argument as in the preceding corollary yields the desired result.