Notes on Canonical Forms

John Beachy, Spring 2007

Throughout these notes V will denote a finite dimensional vector space over a field F and $T : V \to V$ will be a linear transformation. Our goal is to choose a basis for V in such a way that the corresponding matrix for T has as "simple" a form as possible. We will try to come as close to a diagonal matrix as possible.

The notes will follow the terminology of Curtis in *Linear Algebra: an introductory approach*, as much as possible. The order of some results may be changed, and some proofs will be different.

22. BASIC CONCEPTS

Eigenvalues and eigenvectors

Definition 23.10: *The linear transformation T is called* diagonalizable *if there exists a basis for V with respect to which the matrix for T is a diagonal matrix.*

Let's look at what it means for the matrix of T to be diagonal. Recall that we get the matrix by choosing a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, and then entering the coordinates of $T(\mathbf{u}_1)$ as the first column, the coordinates of $T(\mathbf{u}_2)$ as the second column, etc. The matrix is diagonal, with entries $\alpha_1, \alpha_2, \ldots, \alpha_n$, if and only if the chosen basis has the property that $T(\mathbf{u}_i) = \alpha_i \mathbf{u}_i$, for $1 \leq i \leq n$. This leads to the definition of an "eigenvalue" and its corresponding "eigenvectors".

Definition 22.7: *A scalar $\alpha \in F$ is called an* eigenvalue *of T if there exists a nonzero vector $\mathbf{v} \in V$ such that*

$$T(\mathbf{v}) = \alpha \mathbf{v} \ .$$

In this case, $\mathbf{v}$ is called an eigenvector *of T belonging to the eigenvalue α.*

As noted above, the significance of this definition is that there is a diagonal matrix that represents T if and only if there exists a basis for V that consists of eigenvectors of T. Note that the terms *characteristic value* and *eigenvalue* are used synonymously, as are the terms *characteristic vector* and *eigenvector*.

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors belonging to the eigenvalue α of T, then so is any nonzero vector $a_1\mathbf{v}_1 + a_2\mathbf{v}_2$, since

$$T(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) = a_1\alpha\mathbf{v}_1 + a_2\alpha\mathbf{v}_2 = \alpha(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) \ .$$

For each eigenvalue α of T, the set of eigenvectors that belong to it, together with the zero vector, forms a subspace.

In finding eigenvectors $\mathbf{v}$ with $T(\mathbf{v}) = \alpha\mathbf{v}$, it is often easier to use the equivalent condition $[T - \alpha 1_V](\mathbf{v}) = \mathbf{0}$, where 1_V is the identity mapping on V. Then the eigenvectors corresponding to an eigenvalue α are precisely the nonzero vectors in the null space $n(T - \alpha 1_V)$ (which is the same as the kernel of $T - \alpha 1_V$).

Definition: *If α is an eigenvalue of T, then the null space $n(T - \alpha 1_V)$ is called the* eigenspace *of α.*

For example, 0 is an eigenvalue of T if and only if T is not one-to-one. In this case the eigenspace of 0 is just the null space of T.

In searching for eigenvectors there is yet another point of view. If $T(\mathbf{v}) = \alpha\mathbf{v}$, then $T(\lambda\mathbf{v}) = \alpha\lambda\mathbf{v}$, showing that the vectors in the one dimensional subspace spanned by $\mathbf{v}$ are all mapped by T back into the same one dimensional subspace. For example, if $V = \mathbf{R}_3$, then to find eigenvectors of T we need to look for lines through the origin that are mapped by T back into themselves. This point of view leads to an important definition.

Definition 23.1: *A subspace W of V is called a T-*invariant subspace *if $T(\mathbf{w}) \in W$, for all $\mathbf{w} \in W$.*

With this terminology, a nonzero vector $\mathbf{v}$ is an eigenvector of T if and only if it belongs to a one-dimensional invariant subspace of V.

It is easy to check that the null space of T and the range of T are T-invariant subspaces. An interesting case occurs when T is a projection, with $T^2 = T$. Then we have shown that $V = n(T) \oplus T(V)$, and so V is a direct sum of T-invariant subspaces. It is important to know that if α is an eigenvalue of T, then the eigenspace of α is a T-invariant subspace. This follows directly from the fact that if $\mathbf{v}$ is an eigenvalue belonging to α, then so is $T(\mathbf{v})$, since $T(\mathbf{v}) = \alpha\mathbf{v}$.

We have been considering the eigenvectors belonging to a given eigenvalue. But how do we find the eigenvalues? From the definition of an eigenvalue, the scalar α is an eigenvalue if and only if $n(T - \alpha 1_V) \neq 0$. This allows us to conclude that α is an eigenvalue if and only if the determinant $D(T - \alpha 1_V)$ is zero. Then we can find the eigenvalues of T as the zeros of the polynomial equation $D(x1_V - T) = 0$. Thus we must choose a basis for V and then find the matrix A which represents T relative to the basis. Finally, we must solve the equation $D(xI - A) = 0$, where I is the identity matrix.

Definition 24.6: *If A is an $n \times n$ matrix, then the polynomial $f(x) = D(xI - A)$ is called the* characteristic polynomial *of A.*

We want to be able to talk about the characteristic polynomial of a linear transformation, independently of the basis chosen for its matrix representation. The next proposition shows that this is indeed possible (see page 205 of Curtis).

Proposition: *Similar matrices have the same characteristic polynomial.*

Proof. Let A, B be $n \times n$ matrices. If there exists an invertible $n \times n$ matrix S with $B = S^{-1}AS$, then

$$\begin{aligned} D(xI - B) &= D(xI - S^{-1}AS) \;=\; D(xS^{-1}S - S^{-1}AS) \\ &= D(S^{-1}xIS - S^{-1}AS) \;=\; D(S^{-1}(xI - A)S) \\ &= D(S^{-1})D(xI - A)D(S) \;=\; D(xI - A)\ . \end{aligned}$$

This completes the proof. □

Definition 24.6′: *Let A be a matrix of T with respect to some basis of V. Then the polynomial $f(x) = D(xI - A)$ is called the* characteristic polynomial *of T.*

We have already noted that α is an eigenvalue of T if and only if $D(T - \alpha 1_V) = 0$. We can state the final theorem of the section in a slightly different way than Curtis does. Its proof should be clear from the above discussion.

Theorem 22.10: *The eigenvalues of T are the zeros of its characteristic polynomial.*

In practice, there are many applications in which it is important to find the eigenvalues of a matrix. Using the characteristic polynomial of the matrix turns out to be impractical, and so various computational techniques have been developed. These are beyond the scope of this course, but you can get some idea about them from Kolman and Hill (the Math 240 book).

It turns out that we need to work not only with T but with combinations of powers of T. In particular, if we have any polynomial with coefficients in F, there is a corresponding linear transformation found by substituting T into the polynomial. The only problem is what to do with the constant terms, since the final result must be a linear transformation.

Example 1: If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then its characteristic polynomial is

$$f(x) = D(xI - A) = \begin{vmatrix} x-a & -b \\ -c & x-d \end{vmatrix} = (x-a)(x-d) - bc = x^2 - (a+d)x + (ad - bc)\ .$$

Note that $f(x)$ involves the trace and determinant of A. Curtis does this calculation in Example A, on page 187, and then writes out the characteristic polynomial of a 3×3 matrix in Example B, on page 188.

Definition 22.1: *Let* $f(x) = a_m x^m + \ldots + a_1 x + a_0$ *be a polynomial in* $F[x]$*. Then* $f(T)$ *denotes the linear transformation*

$$f(T) = a_m T^m + \ldots + a_1 T + a_0 1_V : V \to V \ ,$$

where 1_V *is the identity on* V*. Similarly, if* A *is an* $n \times n$ *matrix, then* $f(A)$ *denotes the* $n \times n$ *matrix*

$$f(A) = a_m A^m + \ldots + a_1 A + a_0 I \ ,$$

where I *is the* $n \times n$ *identity matrix.*

If $f(T)$ *is the zero transformation on* V*, we say that* T satisfies *the polynomial* $f(x)$*. Similarly, a matrix* A satisfies *the polynomial* $f(x)$ *if* $f(A)$ *is the zero matrix.*

Before proving the next result, we need to make an observation about the action of powers of T on eigenvectors. Suppose that $\mathbf{v} \in V$ and $T(\mathbf{v}) = \alpha\mathbf{v}$ for some scalar α. Then for any polynomial $a_n x^n + \ldots + a_1 x + a_0$ we have

$$\begin{aligned} [a_n T^n + \ldots + a_1 T + a_0 1_v](\mathbf{v}) &= a_n T^n(\mathbf{v}) + \ldots + a_1 T(\mathbf{v}) + a_0(\mathbf{v}) \\ &= a_n \alpha^n \mathbf{v} + \ldots + a_1 \alpha \mathbf{v} + a_0 \mathbf{v} \\ &= (a_n \alpha^n + \ldots + a_1 \alpha + a_0)\mathbf{v} \end{aligned}$$

or equivalently, $[f(T)](\mathbf{v}) = f(\alpha) \cdot \mathbf{v}$.

The proof of the next theorem is a slick one, using the above paragraph. It makes use of the functions f_i that turn up in the proof of the Lagrange interpolation formula. Note that Curtis has a proof that is easier to remember, and maybe easier to understand.

Theorem 22.8: *Let* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ *be eigenvectors belonging to the distinct eigenvalues* $\alpha_1, \alpha_2, \ldots, \alpha_r$ *of* T*. Then the vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ *are linearly independent.*

Proof. It is possible to choose polynomials $f_i(x)$ such that

$$f_i(\alpha_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

If $\sum_{j=1}^r \lambda_j \mathbf{v}_j = \mathbf{0}$, then multiplying by $f_i(T)$ gives

$$\mathbf{0} = [f_i(T)](\textstyle\sum_{j=1}^r \lambda_j \mathbf{v}_j) = \sum_{j=1}^r \lambda_j f_i(\alpha_j) \cdot \mathbf{v}_j = \lambda_i \mathbf{v}_i$$

which shows that $\lambda_i = 0$ for all i. Thus the given eigenvectors are linearly independent. □

Remember that diagonalizing a matrix is equivalent to finding a basis consisting of eigenvectors. We need to recall the procedure in Math 240 for attempting to diagonalize an $n \times n$ matrix A. We first find the eigenvalues of A by finding the zeros of the characteristic polynomial $f(x)$. These correspond to linear factors of $f(x)$, and if we cannot factor $f(x)$ into a product of linear factors, then we cannot possibly find enough linearly independent eigenvectors, so A is not diagonalizable.

If we find that there are enough eigenvalues, then for each eigenvalue α, we solve the system of equations $(A-\alpha I)\mathbf{x} = \mathbf{0}$ to find a basis for the corresponding eigenspace. The next problem is that some of the zeros may have multiplicity greater than 1, corresponding to linear factors of degree greater than 1. If α is not a repeated root, then the solution space is one dimensional and we can easily find a basis, which gives the eigenvector corresponding to α. If α *is* a repeated root, say of multiplicity m, then we need m linearly independent eigenvectors belonging to α. In order for this to occur, we need the nullity of $A - \alpha I$ to be equal to m.

I think that this review is enough to prove the following theorem. Just remember that the eigenspace of eigenvectors belonging to α_i is the null space $n(T - \alpha_i 1_V)$.

Theorem: *Let $\alpha_1, \ldots, \alpha_s$ be the distinct eigenvalues of T, and let $V_i = n(T - \alpha_i 1_v)$. The following are equivalent.*

(1) *T is diagonalizable;*

(2) *the characteristic polynomial of T is $f(x) = (x - \alpha_1)^{e_1} \cdots (x - \alpha_s)^{e_s}$, and $\dim(V_i) = e_i$, for all i;*

(3) $\dim(V_1) + \ldots + \dim(V_s) = \dim(V)$.

Proof. It is easy to see that (1) implies (2), since we can compute the characteristic polynomial using a diagonal matrix that represents T.

We have (2) implies (3) since the degree of $f(x)$ is equal to $\dim(V)$.

Finally, if (3) holds then we can choose a basis for each of the nonzero eigenspaces V_i, and then by Theorem 22.8, the union of these bases must be a basis for V. Since we have found a basis of eigenvectors, the corresponding matrix is diagonal, and thus T is diagonalizable. □

From Math 240, you are probably familiar with the following result (page 274 of Curtis). The proof requires a deeper analysis of matrices over the complex numbers, so we will almost certainly not have time to prove it.

Theorem 31.12: *Let V be a finite dimensional vector space over the field $\mathbf{R}$ of real numbers. If $T \in L(V,V)$ is symmetric, then there exists an orthonormal basis for V consisting of eigenvectors of T.*

The minimal polynomial for T

We have shown that $L(V,V)$ is a vector space, and since it is isomorphic to the vector space $M_n(F)$ of $n \times n$ matrices, it has dimension n^2. If we consider the set $I, T, \ldots, T^{n^2}$ in $L(V,V)$, we have $n^2 + 1$ elements, so they cannot be linearly independent. Thus there is some linear combination $a_0 I + a_1 T + \ldots + a_{n^2+1} T^{n^2}$ that equals the zero function. This is the proof given by Curtis that every $T \in L(V,V)$ satisfies some polynomial of degree $\leq n^2$. Knowing that there is *some* polynomial that T satisfies, we can find a polynomial of minimal degree that T satisfies, and then we can divide by its leading coefficient to obtain a monic polynomial. We will verify the uniqueness of this "minimal polynomial" in Theorem 22.3.

Definition 22.4: *The monic polynomial $m(x)$ of minimal degree such that $m(T) = 0$ is called the* minimal polynomial *of T.*

Theorem 22.3:

(a) *The minimal polynomial of T exists and is unique.*

(b) *If $m(x)$ is the minimal polynomial of T and $f(x)$ is any polynomial with $f(T) = 0$, then $m(x)$ is a factor of $f(x)$.*

Proof. Let $m(x)$ be a monic polynomial of minimal degree such that $m(T) = 0$. Suppose that $f(x)$ is any polynomial with $f(T) = 0$. The division algorithm holds for polynomials with coefficients in the field F, so it is possible to write $f(x) = q(x)m(x) + r(x)$, where either $r(x) = 0$ or $\deg(r(x)) < \deg(m(x))$. If $r(x)$ is nonzero, then we have $r(T) = f(T) - q(T)m(T) = 0$. We can divide $r(x)$ by its leading coefficient to obtain a monic polynomial satisfied by T, and then this contradicts the choice of $m(x)$ as a monic polynomial of minimal degree satisfied by T. We conclude that the remainder $r(x)$ is zero, so $f(x) = q(x)m(x)$ and $m(x)$ is thus a factor of $f(x)$.

If $h(x)$ is another monic polynomial of minimal degree with $h(T) = 0$, then by the preceding paragraph $m(x)$ is a factor of $h(x)$, and $h(x)$ is a factor of $m(x)$. Since both are monic polynomials, this forces $h(x) = m(x)$, showing that the minimal polynomial is unique. □

We need to investigate the relationship between the characteristic polynomial and the minimal polynomial of T. They are closely related, and the interplay between them will help to determine the best form for a matrix for T. We now return to eigenvalues and eigenvectors, showing that the minimal polynomial can also be used to find the eigenvalues of T.

Proposition: *The characteristic and minimal polynomials of a linear transformation have the same zeros (except for multiplicities).*

Proof. Let $m(x)$ be the minimal polynomial of T. We need to prove that $m(\alpha) = 0$ if and only if α is an eigenvalue of T.

First assume that $m(\alpha) = 0$. Then $m(x) = (x - \alpha)q(x)$ for some $q(x) \in F[x]$, since zeros correspond to linear factors. Since $\deg(q(x)) < \deg(m(x))$, we have $q(T) \neq 0$, because $m(x)$ has minimal degree among polynomials satisfied by T. Therefore there exists $\mathbf{u} \in V$ with $[q(T)](\mathbf{u}) \neq \mathbf{0}$. Letting $\mathbf{v} = [q(T)](\mathbf{u})$ gives

$$[T - \alpha 1_V](\mathbf{v}) = [(T - \alpha 1_V)q(T)](\mathbf{u}) = [m(T)](\mathbf{u}) = \mathbf{0}$$

and thus $\mathbf{v}$ is an eigenvector of T with eigenvalue α.

Conversely, assume that α is an eigenvalue of T, say $T(\mathbf{v}) = \alpha\mathbf{v}$ for some $\mathbf{v} \neq \mathbf{0}$. Then $m(\alpha) \cdot \mathbf{v} = [m(T)](\mathbf{v}) = \mathbf{0}$ implies that $m(\alpha) = 0$ since $\mathbf{v} \neq \mathbf{0}$. $\square$

A stronger statement than the above proposition is that the minimal polynomial of T is a factor of the characteristic polynomial of T. This follows from the Cayley-Hamilton theorem, which shows that any matrix satisfies its characteristic polynomial. This also gives an easier proof of part of the previous proposition, since the zeros of a polynomial $m(x)$ are automatically zeros of any multiple $m(x)q(x)$.

Theorem 24.8 [Cayley-Hamilton]: *If $f(x)$ is the characteristic polynomial of an $n \times n$ matrix A, then $f(A) = 0$.*

Proof. Let $f(x) = D(xI - A) = a_0 + a_1 x + \ldots + x^n$, and let $B(x)$ be the matrix which is the adjoint of $xI - A$. The entries of $B(x)$ are cofactors of $xI - A$, and so they are polynomials of degree $\leq n - 1$, which allows us to factor out powers of x to get $B(x) = B_0 + B_1 x + \ldots + B_{n-1}x^{n-1}$, for some $n \times n$ matrices $B_0, B_1, \ldots, B_{n-1}$. By definition of the adjoint, $(xI - A) \cdot B(x) = D(xI - A) \cdot I$, or

$$(xI - A)(B_0 + B_1 x + \ldots + B_{n-1}x^{n-1}) = (a_0 + a_1 x + \ldots + x^n) \cdot I$$

If we equate coefficients of x and then multiply by increasing powers of A we obtain

$$\begin{array}{rcl} -AB_0 = a_0 I & \text{and then} & -AB_0 = a_0 I \\ B_0 - AB_1 = a_1 I & & AB_0 - A^2 B_1 = a_1 A \\ \vdots & & \vdots \\ B_{i-1} - AB_i = a_i I & & A^i B_{i-1} - A^{i+1} B_i = a_{i-1} A^i \\ \vdots & & \vdots \\ B_{n-1} = I & & A^n B_{n-1} = A^n \end{array}$$

If we add the right hand columns, we get $0 = a_0 I + a_1 A + \ldots + A^n$, or simply $f(A) = 0$. $\square$

Curtis gives a special case of The Cayley-Hamilton theorem in Theorem 24.8, and then outlines a proof of the general case in Exercise 8 on page 226. The above proof is easier, so it makes sense to present it here, since we can use it immediately.

As a corollary of the Cayley-Hamilton theorem, we can now see that the minimal polynomial of T has degree $\leq n$, Note that in the next section Curtis characterizes diagonalizable matrices in terms of their minimal polynomials: A is diagonalizable if and only if its minimal polynomial has the form $m(x) = (x - \alpha_1) \cdots (x - \alpha_s)$, where $\alpha_i \neq \alpha_j$ for $i \neq j$. That is, A is diagonalizable if and only if its minimal polynomial is a product of linear factors, with no repetitions.

EXERCISES

1. Find the characteristic polynomial $f(x)$ of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}$. Show that $f(A) = 0$, and find the minimal polynomial of A.

   Hint: To check that $f(A) = 0$, it may be easiest to work with $f(x)$ in factored form. That way you can check for the minimal polynomial along the way to showing that $f(A) = 0$.

2. Find the eigenvalues and corresponding linearly independent eigenvectors of $\begin{bmatrix} 1 & 0 & 0 \\ -8 & 4 & -6 \\ 8 & 1 & 9 \end{bmatrix}$.

3. Show that if α is an eigenvalue of T, then α^k is an eigenvalue of T^k. How are the corresponding eigenvectors related?

4. We know that if A and B are similar $n \times n$ matrices, say $B = S^{-1}AS$, for an invertible matrix S, then A and B have the same eigenvalues. How are the corresponding eigenvectors related?

5. Show that if A and B are invertible $n \times n$ matrices, then AB^{-1} and $B^{-1}A$ have the same eigenvalues.

23. INVARIANT SUBSPACES

We have worked with the direct sum of two subspaces, but now we need to be able to talk about the direct sum of any finite number of subspaces.

Definition 23.3: *Let $V_1, \ldots, V_s$ be subspaces of V. Then V is said to be the* direct sum *of the subspaces V_i if each vector $\mathbf{v} \in V$ can be written uniquely as a sum $\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_s$, where $\mathbf{v}_i \in V_i$, for $1 \leq i \leq s$. In this case we use the notation $V = V_1 \oplus \cdots \oplus V_s$.*

In a situation reminiscent of the conditions for a basis, Curtis shows in Lemma 23.5 that the uniqueness part of the above definition is equivalent to the condition that the representation of $\mathbf{0}$ is unique. That is, uniqueness is equivalent to the statement that if $\mathbf{v}_1 + \ldots + \mathbf{v}_s = \mathbf{0}$, with $\mathbf{v}_i \in V_i$, then $\mathbf{v}_i = 0$ for all i.

The next proposition gives a useful way to think about a direct sum of subspaces. Given $V = V_1 \oplus \cdots \oplus V_s$, we can choose a basis for V that consists of basis vectors for the subspaces V_i. The converse is also true. Because a set of vectors is a basis for V if and only if each vector in V can be written uniquely as a linear combination of the vectors, the proof of the proposition should be clear.

Proposition: *Let $V_1, \ldots, V_s$ be subspaces of V, and suppose that $\mathcal{B}_i$ is a basis for V_i. Then $V = V_1 \oplus \cdots \oplus V_s$ if and only if $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_s$ is a basis for V.*

We recall that a subspace W of V is called T-invariant if $T(\mathbf{w}) \in W$ for all $\mathbf{w} \in W$. A special case occurs with eigenvectors: the vector $\mathbf{v}$ is an eigenvector of T if and only if the one-dimensional subspace it determines is a T-invariant subspace. Let's use these ideas in talking about diagonalization. Instead of saying that T can be diagonalized if and only if there is a basis for V consisting of eigenvectors of T, we can now say that T can be diagonalized if and only if V can be written as a direct sum of one-dimensional T-invariant subspaces. It is this statement that can be generalized.

Why should we try to look for a direct sum of T-invariant subspaces? Suppose that $V = V_1 \oplus \cdots \oplus V_s$. Choose a basis $\mathcal{B}_i$ for each subspace V_i, and put them together to form a basis for V. If each subspace V_i is T-invariant, then for any basis vector $\mathbf{u}_{ij} \in V_i$, the image $T(\mathbf{u}_{ij})$ can be written as a linear combination of the basis vectors for V_i, with 0's in every other component. The net result is that the matrix for T will have a block diagonal form. (On the top of page 203, Curtis says it is clear that the matrix of T will be in block diagonal form in this case.) The blocks represent the action of T on the subspaces V_i; because each V_i is T-invariant we can restrict $T : V \to V$ to $T_i : V_i \to V_i$. Partitioning the basis into T-invariant subsets is what leads to the block diagonal form. We can formalize this as follows.

Proposition: *Let $V_1, \ldots, V_s$ be T-invariant subspaces with $V = V_1 \oplus \cdots \oplus V_s$. If we choose a basis $\mathcal{B}_i$ for each V_i, then the matrix of T relative to the basis $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_s$ has block diagonal form* $\begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_s \end{bmatrix}$.

Our immediate goal is to use the factorization of the minimal polynomial of T to write V as a direct sum of T-invariant subspaces. This will give us a block diagonal matrix for T, and then we can analyze the blocks individually. First we need another characterization of direct sums of subspaces.

If $V = V_1 \oplus \cdots \oplus V_s$, let E_i be the projection of V onto V_i. That is, $E_i \in L(V, V)$, and V_i is the range of E_i. From the properties of a projection, we have $E_i^2 = E_i$, for all i. Because V is a direct sum of the subspaces, if we first project onto one component and then follow by projecting onto another component, we will always get $\mathbf{0}$. This shows that $E_iE_j = 0$ whenever $i \neq j$. Finally, every vector is the sum of its projections onto the components, so we have $E_1 + \ldots + E_s = 1_V$.

Example 2: Let $V = \mathbf{R}_3$, let V_1 be the x, y-plane, and let V_2 be the z-axis. Then $V = V_1 \oplus V_2$, and the matrices for the projections onto V_1 and V_2 are, respectively, $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $E_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. It is clear that $E_1 + E_2 = I$, $E_1^2 = E_1$, $E_2^2 = E_2$, $E_1E_2 = 0$, and $E_2E_1 = 0$.

Definition 23.8: *If $E \in L(V, V)$ and $E^2 = E$, then E is called a* projection mapping *or an* idempotent *linear transformation.*

Lemma 23.6: *If there exist $E_1, \ldots, E_s \in L(V, V)$ such that*

(a) $E_1 + \ldots + E_s = 1_V$ *and*

(b) $E_iE_j = 0$ *for* $i \neq j$,

then $E_i^2 = E_i$ for all i and $V = E_1V \oplus \cdots \oplus E_sV$.

Proof. Since $E_1 + \ldots + E_s = 1_V$ and $E_iE_j = 0$ if $i \neq j$, we have $E_i = E_i1_V = E_i(E_1 + \ldots + E_i + \ldots + E_s) = E_i^2$.

If $\mathbf{v} \in V$, then $\mathbf{v} = 1_V(\mathbf{v}) = [E_1 + \ldots + E_s](\mathbf{v}) = E_1(\mathbf{v}) + \ldots + E_s(\mathbf{v})$, which shows that $V = E_1V + \ldots + E_sV$. To show uniqueness, suppose that $\mathbf{0} = E_1\mathbf{v}_1 + \ldots + E_s\mathbf{v}_s$. Multiplying by E_i gives $\mathbf{0} = E_i^2\mathbf{v}_i = E_i\mathbf{v}_i$, since $E_iE_j = 0$ for $i \neq j$ and $E_i^2 = E_i$. Thus each term in $E_1\mathbf{v}_1 + \ldots + E_s\mathbf{v}_s$ is zero, and it follows that $V = E_1V \oplus \cdots \oplus E_sV$. □

We are now ready to prove the *primary decomposition theorem.* A nonconstant polynomial in $F[x]$ is called *irreducible* if it has no factors of strictly lower degree.

We begin with the minimal polynomial $m(x)$ of T, and factor it completely into irreducible factors, to get $m(x) = p_1(x)^{e_1} p_2(x)^{e_2} \cdots p_s(x)^{e_s}$. We then consider the linear transformations $p_i(T)^{e_i}$, and their null spaces $V_i = n(p_i(T)^{e_i})$. Each V_i is T-invariant, since if $\mathbf{w} \in V_i$, then $T(\mathbf{w}) \in V_i$ because $[p_i(T)^{e_i}](T(\mathbf{w})) = T(p_i(T)^{e_i}(\mathbf{w})) = T(\mathbf{0}) = \mathbf{0}$. (See Lemma 23.2 in Curtis.) Note that if $p_i(x)^{e_i}$ happens to be a linear factor $x - \alpha$, then $V_i = n(T - \alpha 1_V)$ is the eigenspace of α.

Theorem 23.9: *Let $m(x) = p_1(x)^{e_1} p_2(x)^{e_2} \cdots p_s(x)^{e_s}$ be the minimal polynomial of T, factored completely as a product of powers of distinct monic irreducible polynomials $p_i(x)$. For each i, let V_i be the null space $n(p_i(T)^{e_i})$.*

Then $V = V_1 \oplus \cdots \oplus V_s$, showing that V is a direct sum of T-invariant subspaces.

Proof. Let $q_i(x) = m(x)/p_i(x)^{e_i}$, so that $q_i(x)$ contains all but one of the irreducible factors of $m(x)$. Note that if $i \neq j$, then our definition guarantees that $q_i(x)q_j(x)$ has $m(x)$ as a factor. The polynomials $q_i(x)$, for $1 \leq i \leq s$, have no irreducible factor in common, and so since they are relatively prime there must exist polynomials $a_i(x)$ with $\sum_{i=1}^{s} a_i(x)q_i(x) = 1$.

Let $E_i = a_i(T)q_i(T)$. Then $\sum_{i=1}^{s} E_i = 1_V$, and $E_iE_j = 0$ if $i \neq j$ since $E_iE_j = m(T)q(T)$ for some polynomial $q(x)$. Thus $V = E_iV \oplus \cdots \oplus E_sV$ by Lemma 23.6.

It remains to show that $E_iV = V_i$, for all i. To prove that $E_iV \subseteq V_i$, let $\mathbf{v} \in E_iV$. Then $\mathbf{v} = [q_i(T)](\mathbf{w})$, for some $\mathbf{w} \in V$, and so $[p_i(T)^{e_i}](\mathbf{v}) = [p_i(T)^{e_i}](q_i(T)\mathbf{w}) = [p_i(T)^{e_i}q_i(T)](\mathbf{w}) = m(T)\mathbf{w} = \mathbf{0}$.

To prove that $V_i \subseteq E_iV$, let $\mathbf{v} \in V_i$. Then $[q_j(T)](\mathbf{v}) = \mathbf{0}$, since $q_j(T)$ has $p_i(T)^{e_i}$ as a factor. But $1_V = \sum_{j=1}^{s} E_j$, so $\mathbf{v} = [\sum_{j=1}^{s} E_j](\mathbf{v}) = E_i(\mathbf{v}) \in E_iV$. □

Theorem 23.11: *The linear transformation T is diagonalizable if and only if its minimal polynomial is a product of distinct linear factors.*

Proof. If T is diagonalizable, we can compute its minimal polynomial using a diagonal matrix, and then it is clear that we just need one linear factor for each of the distinct entries along the diagonal.

Conversely, let the minimal polynomial of T be $m(x) = (x-\alpha_1)(x-\alpha_2) \cdots (x-\alpha_s)$. Then V is a direct sum of the null spaces $n(T - \alpha_i 1_V)$, which shows that there exists a basis for V consisting of eigenvectors, and so T is diagonalizable.. □

Example 3: We can now show that Example 2 gives a model for all idempotent linear transformations, in the following sense. If $E^2 = E$, then E satisfies the polynomial $x^2 - x$, so its minimal polynomial is $x(x-1)$ or x or $x - 1$. It follows from Theorem 23.11 that E is diagonalizable, so it is possible to find a basis for V such that the corresponding matrix for E is diagonal, with 1's and/or 0's down the main diagonal, as in Example 2.

24. TRIANGULAR FORM

Definition: *The linear transformation T is called* triangulable *if there exists a basis for V such that the matrix of T relative to the basis is an upper triangular matrix.*

Lemma: *Let W be any proper T-invariant subspace of V. If the minimal polynomial of T is a product of linear factors, then there exists $\mathbf{v} \notin W$ such that $[T-\alpha 1_V](\mathbf{v}) \in W$ for some eigenvalue α of T.*

Proof. Let $m(x)$ be the minimal polynomial of T, and let $\mathbf{u} \in V$ be any vector which is not in W. Since $m(T) = 0$ we certainly have $[m(T)](\mathbf{u}) \in W$. Let $p(x)$ be a monic polynomial of minimal degree with $[p(T)](\mathbf{u}) \in W$. An argument similar to that for the minimal polynomial shows that $p(x)$ is a factor of every polynomial $f(x)$ with the property that $[f(T)](\mathbf{u}) \in W$. (See Lemma 25.4 in Curtis.) In particular, $p(x)$ is a factor of $m(x)$, and so by assumption it is a product of linear factors.

Suppose that $x - \alpha$ is a linear factor of $p(x)$, say $p(x) = (x-\alpha)q(x)$ for some $q(x) \in F[x]$. Note that α is a eigenvalue of T since it is a zero of $m(x)$. Then $\deg(q(x)) < \deg(p(x))$, and so $\mathbf{v} = [q(T)](\mathbf{u}) \notin W$, but

$$[T-\alpha 1_V](\mathbf{v}) = [T-\alpha 1_V](q(T)\mathbf{u}) = [(T-\alpha 1_V)q(T)](\mathbf{u}) = [p(T)](\mathbf{u})$$

and this element belongs to W. □

Theorem 24.1: *The linear transformation T is triangulable if and only if the minimal polynomial $m(x)$ of T is a product of linear factors.*

Proof. If T has a matrix that is upper triangular, then its characteristic polynomial is a product of linear factors, which guarantees that its minimal polynomial has the same property.

To prove the converse, choose a basis for V as follows. Let $\mathbf{u}_1$ be a eigenvector with eigenvalue α_1. Since the subspace W_1 spanned by $\mathbf{u}_1$ is T-invariant, we can apply the previous lemma to obtain a vector $\mathbf{u}_2 \notin W_1$ with $[T-\alpha_2 1_V](\mathbf{u}_2) \in W_1$, for some eigenvalue α_2 of T. (We are not claiming that α_2 is different from α_1.) Then $T(\mathbf{u}_2) = \alpha_2\mathbf{u}_2 + \mathbf{w}_1$, where $\mathbf{w}_1 \in W_1$. Next let W_2 be the subspace spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Since T maps the basis for W_2 back into W_2, the subspace W_2 is T-invariant and we can repeat the argument. Thus $T(\mathbf{u}_i) = \alpha_i\mathbf{u}_i + \mathbf{w}_{i-1}$, where $\mathbf{w}_{i-1}$ belongs to the subspace spanned by $\{\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}\}$. It follows that the matrix for T relative to the basis which has been chosen is an upper triangular matrix. □

If a matrix A is upper triangular, it is easy to find its characteristic polynomial. A similar statement can be made about matrices in block diagonal form. Since the determinant of a matrix in block diagonal form is the product of the determinants of the blocks, the characteristic polynomial of $\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_s \end{bmatrix}$ is the product of the characteristic polynomials of the blocks A_i. The corresponding statement for linear transformations is that if V is a direct sum $V = V_1 \oplus \cdots \oplus V_s$ of T-invariant subspaces, then we can write the characteristic polynomial of T as a product of the characteristic polynomials of the linear transformations $T_i : V_i \to V_i$ that are the restriction of T to V_i.

We now return to Theorem 23.9, and let T_i be the restriction of T to V_i. Let $m_i(x)$ be the minimal polynomial of T_i. Then $m_i(x)$ is a divisor of $p_i(x)^{e_i}$ since $p_i(T)^{e_i} = 0$ on V_i. On the other hand, $m_i(T)f_i(T) = 0$ since $f_i(T)V_j = 0$ for $i \neq j$ while $m_i(T)W_i = 0$. This implies that the minimal polynomial $m(x)$ of T is a divisor of $m_i(x)f_i(x)$, and then $p_i(x)^{e_i}$ must be a divisor of $m_i(x)$. It follows that the minimal polynomial of T_i is $m_i(x) = p_i(x)^{e_i}$.

EXERCISES

1. Let $A = [a_{ij}]$ be an $n \times n$ matrix with characteristic polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$. By substituting $x = 0$, show that $a_0 = \pm D(A)$, for the determinant $D(A)$ of A.

2. Let $A = [a_{ij}]$ be an $n \times n$ matrix with characteristic polynomial $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$. Use induction to show that $f(x) = (x - a_{11}) \cdots (x - a_{nn}) + q(x)$, where $q(x)$ has degree at most $n - 2$. Then show that $-a_{n-1}$ is equal to the trace $\mathrm{tr}(A)$ of A.

3. Let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \in F[x]$. The *companion matrix* of $f(x)$ is the matrix
$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$. Show that the characteristic polynomial of A is $f(x)$, and thus $f(A) = 0$ by the Cayley-Hamilton theorem.

   Hint: Find the determinant by expanding by cofactors along the last column.

25. JORDAN CANONICAL FORM

Definition 25.17: *The matrix A is said to be in* Jordan form *if it satisfies the following conditions.*

(i) *It must be in block diagonal form, where each block has a fixed scalar on the main diagonal and 1's or 0's on the superdiagonal. These blocks are called the primary blocks of A.*

(ii) *The scalars for different primary blocks must be distinct.*

(iii) *Each primary block must be made up of secondary blocks with a scalar on the diagonal and only 1's on the superdiagonal. These blocks must be in order of decreasing size (moving down the main diagonal).*

Example 4: Here are two matrices in Jordan form. The first has two primary blocks, while the second has three.

$$\left[\begin{array}{ccccccccc} 3 & 1 & 0 & & \cdot & & & & \\ 0 & 3 & 1 & & \cdot & & & & \\ 0 & 0 & 3 & & \cdot & & & & \\ & & & 3 & 1 & & & & \\ \cdot & \cdot & \cdot & 0 & 3 & & & & \\ & & & & & 2 & 1 & \cdot & \cdot \\ & & & & & 0 & 2 & & \\ & & & & & \cdot & & 2 & 1 \\ & & & & & \cdot & & 0 & 2 \end{array}\right] \qquad \left[\begin{array}{ccccccccccc} 3 & 1 & 0 & & \cdot & & & & & & \\ 0 & 3 & 1 & & \cdot & & & & & & \\ 0 & 0 & 3 & & \cdot & & & & & & \\ & & & 3 & \cdot & & & & & & \\ \cdot & \cdot & \cdot & \cdot & 3 & & & & & & \\ & & & & & 2 & 1 & 0 & 0 & & \\ & & & & & 0 & 2 & 1 & 0 & & \\ & & & & & 0 & 0 & 2 & 1 & & \\ & & & & & 0 & 0 & 0 & 2 & & \\ & & & & & & & & & 5 & \cdot \\ & & & & & & & & & \cdot & 5 \end{array}\right]$$

As a point of interest, the first matrix has characteristic polynomial $(x-3)^5(x-2)^4$ and minimal polynomial $(x-3)^3(x-2)^2$. The second matrix has characteristic polynomial $(x-3)^5(x-2)^4(x-5)^2$ and minimal polynomial $(x-3)^3(x-2)^4(x-5)$. Our general problem is to use a given matrix, together with its characteristic and minimal polynomials, to choose a basis which gives the Jordan form. This can be done if the characteristic polynomial is a product of linear factors, which happens, for example, over the field of complex numbers, where every polynomial can be completely factored into a product of linear terms.

Let's look at the type of basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ that we need to produce the matrix $\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$. We must have

$$T(\mathbf{u}_1) = 3\mathbf{u}_1, \quad T(\mathbf{u}_2) = 3\mathbf{u}_2 + \mathbf{u}_1, \quad T(\mathbf{u}_3) = 3\mathbf{u}_3 + \mathbf{u}_2\ .$$

We can rewrite this in the following way:

$$[T - 31_V](\mathbf{u}_1) = \mathbf{0}, \quad [T - 31_V](\mathbf{u}_2) = \mathbf{u}_1, \quad [T - 31_V](\mathbf{u}_3) = \mathbf{u}_2 \ .$$

This also implies that

$$[T - 31_V](\mathbf{u}_1) = \mathbf{0}, \quad [T - 31_V]^2(\mathbf{u}_2) = \mathbf{0}, \quad [T - 31_V]^3(\mathbf{u}_3) = \mathbf{0} \ .$$

In this case $\mathbf{u}_2$ is called a *generalized eigenvector* of degree 2, and $\mathbf{u}_3$ is called a *generalized eigenvector* of degree 3.

Definition: *Let $W_1 \subseteq W_2$ be subspaces of V. Let $\mathcal{B}$ be a basis for W_1. Any set $\mathcal{B}'$ of vectors such that $\mathcal{B} \cup \mathcal{B}'$ is a basis for W_2 is called a* complementary basis *for W_1 in W_2.*

Lemma: *Let $L : V \to V$ be a linear transformation. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a complementary basis for $n(L^m)$ in $n(L^{m+1})$. Then $\{L(\mathbf{u}_1), \ldots, L(\mathbf{u}_k)\}$ is part of a complementary basis for $n(L^{m-1})$ in $n(L^m)$.*

Proof. Since $\mathbf{u}_i \in n(L^{m+1})$, we have $L^{m+1}(\mathbf{u}_i) = \mathbf{0}$, so $L^m(L(\mathbf{u}_i)) = \mathbf{0}$, and therefore $L(\mathbf{u}_i) \in n(L^m)$. If $\sum_{i=1}^k c_i L(\mathbf{u}_i) \in n(L^{m-1})$, then

$$L^m \left(\sum_{i=1}^k c_i \mathbf{u}_i \right) = L^{m-1} \left(\sum_{i=1}^k c_i L(\mathbf{u}_i) \right) = 0 \ .$$

This implies that $\sum_{i=1}^k c_i \mathbf{u}_i \in n(L^m)$, and then $c_i = 0$ for all i since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a complementary basis for $n(L^m)$ in $n(L^{m+1})$. This shows that $L(\mathbf{u}_1), \ldots, L(\mathbf{u}_k)$ are linearly independent and also linearly independent of any vectors in $n(L^{m-1})$, so they are part of a complementary basis for $n(L^{m-1})$ in $n(L^m)$. $\square$

Theorem [Jordan canonical form]: *If the characteristic polynomial of T is a product of linear factors, then it is possible to find a matrix representation for T which has Jordan form.*

In particular, suppose that $f(x) = (x - \alpha_1)^{m_1} \cdots (x - \alpha_s)^{m_s}$ is the characteristic polynomial of T and suppose that $m(x) = (x - \alpha_1)^{e_1} \cdots (x - \alpha_s)^{e_s}$ is the minimal polynomial of T. Then the matrix for T has primary blocks of size $m_i \times m_i$ corresponding to each eigenvalue α_i, and in each primary block the largest secondary block has size $e_i \times e_i$.

Proof. By using Theorem 23.9 (the primary decomposition theorem) we can obtain a block diagonal matrix, and then we can deal with each block separately. This reduces the proof to the case in which the characteristic polynomial is $(x - \alpha)^m$ and the minimal polynomial is $(x - \alpha)^r$.

Choose a complementary basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ for $n((T - \alpha 1_V)^{r-1})$ in $n((T - \alpha 1_V)^r)$. By the preceding lemma, the vectors

$$[T - \alpha 1_V]^{r-1}(\mathbf{u}_1), \ldots, [T - \alpha 1_V](\mathbf{u}_1), \ \ \mathbf{u}_1,$$

$$[T - \alpha 1_V]^{r-1}(\mathbf{u}_2), \ldots, [T - \alpha 1_V](\mathbf{u}_2), \ \ \mathbf{u}_2, \ldots\ldots$$

$$[T - \alpha 1_V]^{r-1}(\mathbf{u}_k) \ldots, [T - \alpha 1_V](\mathbf{u}_k), \ \ \mathbf{u}_k,$$

are linearly independent. Starting with $[T - \alpha 1_V](\mathbf{u}_1), \ldots, [T - \alpha 1_V](\mathbf{u}_k)$ we find additional vectors $\mathbf{w}_1, \ldots, \mathbf{w}_t$ to complete a complementary basis for $n((T - \alpha 1_V)^{r-2})$ in $n((T - \alpha 1_V)^{r-1})$. Then we also include

$$[T - \alpha 1_V]^{r-2}(\mathbf{w}_1), \ldots, [T - \alpha 1_V](\mathbf{w}_1) \ ,$$

and so on. We then consider $n((T - \alpha 1_V)^{r-2})$, etc., and continue this procedure until we have found a basis for the generalized eigenspace $n((T - \alpha 1_V)^r$.

Finally we must compute the matrix of T with respect to the basis we have chosen. It is useful to write $T = \alpha 1_V + (T - \alpha 1_V)$. Then, for example,

$$T([T - \alpha 1_V]^j(\mathbf{u}_i)) = c[T - \alpha 1_V]^j(\mathbf{u}_i) + [T - \alpha 1_V]^{j+1}(\mathbf{u}_i)$$

and it follows rather easily that the basis we have chosen leads to a matrix in Jordan form. $\square$

We have given a direct proof that shows that all matrices over an algebraically closed field (such as $\mathbf{C}$) can be put into Jordan form. This proof does not make use of the more general rational canonical form that is developed by Curtis, which applies to all matrices.

Example 5 will illustrate the method of proof of the main theorem. Example 6 will give a more efficient computation of the Jordan canonical form.

Example 5: Let A be the following matrix.

$$\begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 3 & 1 & 0 \\ -1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Since A was chosen to have enough real eigenvalues, we can work in $\mathbf{R}_5$. The matrix $xI - A$ can be broken up into 2×2 and 1×1 blocks for which it is easy to find the determinant. The only work we need to do to find the characteristic polynomial $f(x)$ is to evaluate

$$\begin{vmatrix} x-3 & -1 \\ 1 & x-1 \end{vmatrix} = (x-3)(x-1)+1 = x^2-4x+4 = (x-2)^2 \ .$$

We conclude that $f(x) = (x-2)^5$. To find the minimal polynomial of A, since we know it must be a power of $(x-2)$, we just need to find the first power of $A-2I$ which gives the zero matrix.

$$(A-2I)^2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This tells us that the minimal polynomial is $(x-2)^2$. The next step is to find $n(A-2I)$, the eigenspace of 2. The generalized eigenspace $n(A-2I)^2$ is all of R^5.

We first give the block upper triangular form of the matrix obtained by choosing a basis for the invariant subspace consisting of all eigenvectors. Solving the system $(A-2I)x = 0$ gives $x_2 = -x_1$, $x_4 = -x_3$. We can choose a basis for the eigenspace in the usual way, by considering x_1, x_3, x_5 as independent variables. We then extend it to a basis for the whole space, by choosing two more linearly independent vectors. This gives us the following basis: $\mathbf{u}_1 = (1,-1,0,0,0)$, $\mathbf{u}_2 = (0,0,1,-1,0)$ $\mathbf{u}_3 = (0,0,0,0,1)$, $\mathbf{u}_4 = (1,0,0,0,0)$, $\mathbf{u}_5 = (0,0,1,0,0)$. Computation shows that with respect to this basis we get the following matrix.

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

To obtain the Jordan form of the matrix we must choose our basis more carefully. We begin with $\mathbf{u}_4$ and $\mathbf{u}_5$, since they are in the generalized eigenspace, but are linearly independent of the basis vectors for the eigenspace. By the lemma preceding the theorem, the vectors $(A-2I)\mathbf{u}_4$ and $(A-2I)\mathbf{u}_5$ are in the eigenspace and are linearly independent. To span the eigenspace we need one more linearly independent vector. It can be checked that $\mathbf{u}_3$ is linearly independent of the previously chosen vectors. To give us the desired form we choose $\mathbf{w}_1 = (A-2I)(\mathbf{u}_4)$, $\mathbf{w}_2 = \mathbf{u}_4$, $\mathbf{w}_3 = (A-2I)(\mathbf{u}_5)$, $\mathbf{w}_4 = \mathbf{u}_5$, $\mathbf{w}_5 = \mathbf{u}_3$.

To compute the matrix of A with respect to this basis, note that we have $A = 2I + (A-2I)$. Then

$$A(\mathbf{w}_1) = (2I + (A-2I))(A-2I)(\mathbf{u}_4)2(A-2I)(\mathbf{u}_4) + (A-2I)^2(\mathbf{u}_4) = 2\mathbf{w}_1$$

$$A(\mathbf{w}_2) = (2I + (A - 2I))(\mathbf{u}_4) = 2\mathbf{u}_4 + (A - 2I)(\mathbf{u}_4) = 2\mathbf{w}_2 + \mathbf{w}_1$$

since $(A - 2I)^2 = 0$. Similarly, $A(\mathbf{w}_3) = 2\mathbf{w}_3$ and $A(\mathbf{w}_4) = 2\mathbf{w}_4 + \mathbf{w}_3$. Finally, $A(\mathbf{w}_5) = 2\mathbf{w}_5$ since $\mathbf{w}_5$ is an eigenvector, and we have the following matrix, in Jordan form.

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

There had to be at least one 2×2 block (no larger) because the minimal polynomial had degree two. The second 2×2 block showed up because the dimension of the generalized eigenspace was 2 more than the dimension of the eigenspace. Fortunately we only had to compare the two subspaces, since $A - 2I$ was nilpotent of degree 2. In general, by comparing the dimensions of all of the generalized eigenspaces we could write down the Jordan canonical form without having to actually find the basis.

Example 6: Let A be the following matrix.

$$\begin{bmatrix} 3 & 2 & 1 & 1 & 0 & -3 & 1 \\ -1 & 2 & 0 & 0 & 1 & -1 & 1 \\ 1 & 1 & 3 & 0 & -6 & 5 & -1 \\ 0 & 2 & 1 & 4 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 5 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 5 \end{bmatrix}$$

When finding the characteristic polynomial, the matrix can be broken up into blocks. The 3×3 block in the lower right hand corner gives $(x - 3)(x - 5)^2$. For the 4×4 block in the upper left hand corner we have the following computation. (Add row 2 to row 3; subtract row 1 from row 4; add rows 3 and 4 to row 1; add row 2 to row 1.)

$$\begin{vmatrix} x-3 & -2 & -1 & -1 \\ 1 & x-2 & 0 & 0 \\ -1 & -1 & x-3 & 0 \\ 0 & -2 & -1 & x-4 \end{vmatrix} = (x-3) \begin{vmatrix} x-3 & -2 & -1 & -1 \\ 1 & x-2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -2 & -1 & x-4 \end{vmatrix}$$

$$= (x-3)^2 \begin{vmatrix} x-3 & -2 & -1 & -1 \\ 1 & x-2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{vmatrix} = (x-3)^2 \begin{vmatrix} x-4 & -1 & 0 & 0 \\ 1 & x-2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{vmatrix} = (x-3)^4$$

Thus the characteristic polynomial of A is $f(x) = (x - 3)^5(x - 5)^2$. Next we must compute the dimensions of the generalized eigenspaces for each of the eigenvalues 3 and 5. We first row-reduce $A - 3I$.

$$A - 3I = \begin{bmatrix} 0 & 2 & 1 & 1 & 0 & -3 & 1 \\ -1 & -1 & 0 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & -6 & 5 & -1 \\ 0 & 2 & 1 & 1 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We conclude that the eigenspace corresponding to 3 has dimension $7 - 5 = 2$. Next we compute $(A - 3I)^2$ and row-reduce it.

$$(A - 3I)^2 = \begin{bmatrix} -1 & 1 & 1 & 1 & 1 & -6 & 4 \\ 1 & -1 & -1 & -1 & -1 & 2 & 0 \\ -1 & 1 & 1 & 1 & -7 & 6 & 0 \\ -1 & 1 & 1 & 1 & 1 & -6 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus the generalized eigenspace of degree 2 of the eigenvalue 3 has dimension $7-3 = 4$. The dimension of the generalized eigenspace of degree 3 can be at most 5, since $(x - 3)$ has exponent 5 in the characteristic polynomial, and so it must be equal to 5. Nonetheless, we can easily do the computation which shows that $(A - 3I)^3$ has rank 2, and hence a nullity of 5.

$$(A - 3I)^3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -8 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -16 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & -8 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -8 & 8 & 0 \\ 0 & 0 & 0 & 0 & -8 & 0 & 8 \end{bmatrix}$$

To summarize, we can find two linearly independent eigenvectors, two linearly independent generalized eigenvectors of rank 2, and one linearly independent generalized eigenvector of rank 3. In the matrix for A, the primary block corresponding to the eigenvalue 3 will have one secondary 3×3 block, corresponding to the one generalized eigenvector of rank 3. Since there are two generalized eigenvectors of rank 2, one will go in the 3×3 block, and one will go into the next secondary block, which will therefore be 2×2.

Finally, we must compute the dimensions of the generalized eigenspaces for 5. Since the exponent of $(x - 5)$ in the characteristic polynomial is 2, the dimension will be 2 and we should reach this in either one step or two. In fact, by looking at the last three rows we can tell that the rank is no more than 5, so the nullity must be 2, giving 2 eigenvectors.

$$A - 5I = \begin{bmatrix} -2 & 2 & 1 & 1 & 0 & -3 & 1 \\ -1 & -3 & 0 & 0 & 1 & -1 & 1 \\ 1 & 1 & -2 & 0 & -6 & 5 & -1 \\ 0 & 2 & 1 & -1 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 \end{bmatrix}$$

We can now give the Jordan canonical form of the matrix A.

$$\begin{bmatrix} 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Example 7: Suppose that A is a matrix with characteristic polynomial $f(x) = (x-7)^6$ and minimal polynomial $p(x) = (x-7)^3$. Then A must be a 6×6 matrix, and the first secondary block must be a 3×3 block since the minimal polynomial has degree 3. We do not have enough information to determine A, but we have listed below all possible cases.

$$\begin{bmatrix} 7 & 1 & 0 & & & \\ 0 & 7 & 1 & & & \\ 0 & 0 & 7 & & & \\ & & & 7 & 1 & 0 \\ & & & 0 & 7 & 1 \\ & & & 0 & 0 & 7 \end{bmatrix} \quad \begin{bmatrix} 7 & 1 & 0 & & & \\ 0 & 7 & 1 & & & \\ 0 & 0 & 7 & & & \\ & & & 7 & 1 & \\ & & & 0 & 7 & \\ & & & & & 7 \end{bmatrix} \quad \begin{bmatrix} 7 & 1 & 0 & & & \\ 0 & 7 & 1 & & & \\ 0 & 0 & 7 & & & \\ & & & 7 & & \\ & & & & 7 & \\ & & & & & 7 \end{bmatrix}$$

Proposition: *Let A and B be $n \times n$ matrices over the field $\mathbf{C}$ of complex numbers. Then A and B are similar if and only if they have the same Jordan canonical form, up to the order of the eigenvalues.*

Proof. The assumption that the matrices have complex entries guarantees that their characteristic polynomials can be factored completely, and so they *can* be put into Jordan canonical form.

If A and B are similar, then they represent the same linear transformation T. The Jordan canonical form depends only on the linear transformation T, since it is constructed from the zeros of the characteristic polynomial of T and the dimensions of the corresponding generalized eigenspaces. However, we have no way to specify the order of the eigenvalues. Thus A and B have the same Jordan canonical form, except for the order of the eigenvalues.

Conversely, if A and B have the same Jordan canonical form, except for the order of the eigenvalues, then the matrices in Jordan canonical form are easily seen to be similar. Thus A and B are similar to the similar matrices. Since similarity is an equivalence relation, it follows that A and B are similar. □

EXERCISES

1. Find all possible Jordan canonical forms of each of the following matrices, over the field of complex numbers.

   (a) The matrix A has characteristic polynomial $(x-2)^4(x-10)^3$ and minimal polynomial $(x-2)^3(x-10)$.

   (b) The matrix B has rank 4, characteristic polynomial x^7, and minimal polynomial x^3.

2. Let A be a matrix over the field of real numbers. If the characteristic polynomial of A is $(x-2)(x-3)^6$ and the minimal polynomial of A is $(x-2)(x-3)^3$, list the possible Jordan canonical forms for A.

3. Find the characteristic polynomial and Jordan canonical form of these matrices.

   (a) $A = \begin{bmatrix} 3 & 0 & 1 & 1 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ (b) $B = \begin{bmatrix} 1 & 4 & 4 & 8 \\ -1 & 5 & 6 & 10 \\ 0 & 0 & 9 & 9 \\ 0 & 0 & -4 & -3 \end{bmatrix}$

4. Let a, b, c be nonzero real numbers, and let $A = \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix}$. Find the Jordan canonical form J_A of A, and find a matrix Q such that $Q^{-1}AQ = J_A$.

5. Let L be the linear transformation of $\mathbf{R}^3$ to itself given by the matrix

$$A = \frac{1}{25}\begin{bmatrix} 15 & 12 & -16 \\ -20 & 9 & -12 \\ 0 & 20 & 15 \end{bmatrix}$$

   (with respect to the standard basis of $\mathbf{R}^3$). You may assume that L is a rotation about the axis $\boldsymbol{a} = (-2, 1, 2)$ by an angle θ. Find $\cos(\theta)$ and a matrix in Jordan canonical form that is similar to A.