

# NOTES ON THE STRUCTURE OF LINEAR TRANSFORMATIONS

LANCE D. DRAGER

## 1. INTRODUCTION AND PRELIMINARIES

The goal of these notes is to give an accessible introduction to the structure theory of linear transformations. We want to develop enough of the theory for use in differential equations, with only basic linear algebra as a prerequisite.

In a following set of notes, we will show how these results can be applied to differential equations.

In the rest of the section, we will cover some topics that will be useful tools for the final results and which may be unfamiliar to the reader.

In order to deal with results over both the real and complex numbers, we will use the symbol  $\mathbb{K}$  to mean either  $\mathbb{R}$  or  $\mathbb{C}$ . Thus, a statement made in terms of  $\mathbb{K}$  is supposed to hold in both the real and complex cases.

We will *not* use boldface notation for vectors in these notes.

### 1.1. Direct Sum Decompositions.

1.1.1. *Basics.* Let  $V$  be a vectorspace over  $\mathbb{K}$ , and let  $S_1, S_2, \dots, S_k \subseteq V$  be subspaces. The sum of these subspaces, denoted  $S_1 + S_2 + \dots + S_k$  is the set of all vectors that can be obtained by taking a sum of the form  $s_1 + s_2 + \dots + s_k$ , where each  $s_j$  is in  $S_j$ ,  $j = 1, \dots, k$ . It is easy to check that this set of vectors is a subspace of  $V$ .

We say that **the sum**  $S_1 + \dots + S_k$  **is direct** if every vector  $v$  in  $S_1 + \dots + S_k$  has a *unique* representation  $v = s_1 + \dots + s_k$  where  $s_j \in S_j$ .

**Proposition 1.1.** *The sum  $S_1 + \dots + S_k$  is direct if and only if the following condition holds: If  $s_j \in S_j$  and  $s_1 + \dots + s_k = 0$ , then each  $s_j$  is 0.*

*Proof.* Since the  $S_j$ 's are subspaces,  $0 \in S_j$ . Thus,  $0 \in S_1 + \dots + S_k$ , since  $0 = 0 + \dots + 0$ , where  $0 \in S_1, \dots, 0 \in S_k$ . If the sum is direct, this must be the unique representation of zero as an element of the sum, so the condition in the Proposition holds.

On the other hand, suppose the condition in the Proposition holds, and that  $v \in S_1 + \dots + S_k$  has representations

$$\begin{aligned}v &= v_1 + \dots + v_k, & v_j &\in S_j \\v &= w_1 + \dots + w_k, & w_j &\in S_j.\end{aligned}$$

Subtracting these two equations gives

$$0 = (v_1 - w_1) + \dots + (v_k - w_k).$$

Since each  $S_j$  is a subspace,  $v_j - w_j \in S_j$ . By the hypothesis, we must have  $v_j - w_j = 0$  for  $j = 1, \dots, k$ . Thus, the representation of  $v$  is unique and the sum is direct.  $\square$

If the sum  $S_1 + \cdots + S_k$  is direct, it is usually denoted by  $S_1 \oplus \cdots \oplus S_k$  or

$$\bigoplus_{j=1}^k S_j.$$

**Exercise 1.2.** In the case of two subspaces  $S_1, S_2 \subseteq V$ , show that the sum  $S_1 + S_2$  is direct if and only if  $S_1 \cap S_2 = \{0\}$ .

**Proposition 1.3.** Suppose that  $V = V_1 \oplus \cdots \oplus V_k$  and that for each  $j$ , the vectors

$$(1.1) \quad v_1^j, v_2^j, \dots, v_{n_j}^j$$

are a basis of  $V_j$ . The collection of vectors

$$(1.2) \quad v_1^1, v_2^1, \dots, v_{n_1}^1, v_1^2, v_2^2, \dots, v_{n_2}^2, \dots, v_1^k, v_2^k, \dots, v_{n_k}^k$$

forms a basis of  $V$

*Proof.* If  $v \in V$ , we can write  $v$  uniquely as  $v = v_1 + v_2 + \cdots + v_k$  where  $v_j \in V_j$ . Since the vectors in (1.1) are a basis of  $V_j$ , we have

$$v_j = \sum_{\ell=1}^{n_j} c_\ell^j v_\ell^j$$

for some coefficients  $c_\ell^j \in \mathbb{K}$ . Thus, we have

$$v = \sum_{j=1}^k \sum_{\ell=1}^{n_j} c_\ell^j v_\ell^j.$$

This shows that the vectors in (1.2) span  $V$ .

To show that the vectors in (1.2) are linearly independent, suppose that we have

$$\sum_{j=1}^k \left[ \sum_{\ell=1}^{n_j} c_\ell^j v_\ell^j \right] = 0.$$

The sum in brackets is in  $V_j$ . Since the sum of the subspaces is direct, we must have

$$\sum_{\ell=1}^{n_j} c_\ell^j v_\ell^j = 0$$

for all  $j$ . Since the vectors in (1.1) are independent, we conclude that all of the scalars  $c_\ell^j$  must be zero. Thus, the vectors in (1.2) are independent.  $\square$

**Corollary 1.4.** If  $V = V_1 \oplus \cdots \oplus V_k$  then

$$\dim(V) = \dim(V_1) + \cdots + \dim(V_k).$$

1.1.2. *Projections.* To make the notation less cumbersome, we'll usually discuss the case of the direct sum of two subspaces when the generalization to more summands is clear.

Our next topic is the projections associated to a direct sum decomposition. Suppose that  $U = V \oplus W$ . Then each  $u \in U$  has a unique decomposition  $u = v + w$  where  $v \in V$  and  $w \in W$ . Thus, it makes sense to define mappings  $P: U \rightarrow U$  and  $Q: U \rightarrow U$  by

$$\begin{aligned} P(u) &= P(v + w) = v \\ Q(u) &= Q(v + w) = w, \end{aligned}$$

and it is easy to check that  $P$  and  $Q$  are linear. The image (i.e., range) of  $P$  is  $V$  and the image of  $Q$  is  $W$ . We have  $P(v) = P(v + 0) = v$  for  $v \in V$  and  $Q(w) = Q(0 + w) = w$  for  $w \in W$ . This implies that  $P^2 = P$  and  $Q^2 = Q$ . If  $u = v + w$ , then  $P(u) + Q(u) = P(v + w) + Q(v + w) = v + w = u$ . Thus,  $P + Q = I$ , where  $I$  stands for the identity transformation on  $U$ . Also note that  $P(Q(u + w)) = P(w) = 0$ , so  $PQ = 0$ . Similarly,  $QP = 0$ .

**Exercise 1.5.** Suppose that  $P: V \rightarrow V$  is a linear transformation such that  $P^2 = P$ . Show that  $V$  is the direct sum of the image of  $P$  and the nullspace of  $P$ .

1.1.3. *Direct Sum Decompositions of Linear Transformations.* Suppose that  $U = V \oplus W$  and that we are given linear transformations  $L_1: V \rightarrow V$  and  $L_2: W \rightarrow W$ . Then we can define a linear transformation  $L = L_1 \oplus L_2$  from  $U$  to  $U$  by  $L(v + w) = L_1(v) + L_2(w)$ . Notice that the subspaces  $V$  and  $W$  are invariant under  $L$ , i.e.,  $L(V) \subseteq V$  and  $L(W) \subseteq W$ .

Conversely, suppose that  $L: U \rightarrow U$  is a linear transformation that leaves the subspaces  $V$  and  $W$  invariant, i.e.,  $L(V) \subseteq V$  and  $L(W) \subseteq W$ . Then we can define a restricted linear transformation  $L_1: V \rightarrow V$  by  $L_1(v) = L(v)$  and a restricted linear transformation  $L_2: W \rightarrow W$  by  $L_2(w) = L(w)$ . Then we can write  $L = L_1 \oplus L_2$ . If we can understand the transformations  $L_1$  and  $L_2$ , we will get a good picture of  $L$ .

Lets examine this situation in terms of matrices. Suppose that  $\mathcal{V} = [v_1 \ \dots \ v_k]$  is an ordered basis of  $V$  and that  $\mathcal{W} = [w_1 \ \dots \ w_\ell]$  is an ordered basis of  $W$ . The matrix of  $L_1: V \rightarrow V$  with respect to  $\mathcal{V}$  will be a  $k \times k$  matrix, call it  $A$ . The matrix of  $L_2: W \rightarrow W$  with respect to  $\mathcal{W}$  will be an  $\ell \times \ell$  matrix  $B$ . Recall that the defining equations of these matrices are  $L_1(\mathcal{V}) = \mathcal{V}A$  and  $L_2(\mathcal{W}) = \mathcal{W}B$ .

As we saw above,  $\mathcal{U} = [v_1 \ \dots \ v_k \ w_1 \ \dots \ w_\ell]$  is an ordered basis of  $U = V \oplus W$ .

We want to determine the matrix of  $L = L_1 \oplus L_2$  with respect to  $\mathcal{U}$ . Lets call this matrix  $C$ . To find  $C$  we have to find the matrix so that  $L(\mathcal{U}) = \mathcal{U}C$ . But, we have

$$\begin{aligned} L(\mathcal{U}) &= [L(v_1) \ \dots \ L(v_k) \ L(w_1) \ \dots \ L(w_\ell)] \\ &= [L_1(v_1) \ \dots \ L_1(v_k) \ L_2(w_1) \ \dots \ L_2(w_\ell)] \end{aligned}$$

Consider the column of  $C$  corresponding to  $L_1(v_j)$ . Since  $L_1(v_j) \in V$ , it can be expressed as a linear combination of  $v_1, \dots, v_k$  and the coefficients are given by column  $j$  of  $A$ . Thus, we have

$$L(v_j) = L_1(v_j) = [v_1 \ \dots \ v_k \ w_1 \ \dots \ w_\ell] \begin{bmatrix} a_{1j} \\ \vdots \\ a_{kj} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Similarly, column  $k+p$  of  $C$  will give the expansion of  $L(w_p)$  with respect to  $\mathcal{U}$ . But  $L(w_p) = L_2(w_p)$  and  $L_2(w_p) \in W$ , so it can be expressed as a linear combination

of  $w_1, \dots, w_\ell$ , with the coefficients given by column  $p$  of  $B$ . Thus, we have

$$L(w_p) = L_2(w_p) = \begin{bmatrix} v_1 & \dots & v_k & w_1 & \dots & w_\ell \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b_{1p} \\ \vdots \\ b_{\ell p} \end{bmatrix}.$$

We conclude that the matrix  $C$  looks like

$$C = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2k} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & b_{11} & b_{12} & \dots & b_{1\ell} \\ 0 & 0 & \dots & 0 & b_{21} & b_{22} & \dots & b_{2\ell} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & b_{\ell 1} & b_{\ell 2} & \dots & b_{\ell \ell} \end{bmatrix}.$$

It is convenient to write a matrix like this in block form as

$$C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where  $A$  stands for the block of entries from  $A$ ,  $B$  stands for the block of entries from  $B$  and “0” stands for a block of zeroes. A matrix of this form is called **block diagonal**.

Thus, if  $V$  splits as a direct sum of subspaces that are invariant under  $L$ , then we can find a basis of  $V$  so the matrix of  $L$  is block diagonal. Similarly, if a linear transformation has a block diagonal form in some basis, then  $V$  splits as a sum of invariant subspaces.

**Exercise 1.6.** Suppose  $A_1$  and  $A_2$  are  $n \times n$  matrices and  $B_1$  and  $B_2$  are  $m \times m$  matrices. Verify the following multiplication rule for block diagonal matrices:

$$\begin{bmatrix} A_1 & 0 \\ 0 & B_1 \end{bmatrix} \begin{bmatrix} A_2 & 0 \\ 0 & B_2 \end{bmatrix} = \begin{bmatrix} A_1 A_2 & 0 \\ 0 & B_1 B_2 \end{bmatrix}.$$

1.1.4. *Determinants of Block Diagonal Matrices.* We want to prove the following formula for the determinant of a block diagonal matrix.

**Proposition 1.7.** *Let*

$$C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

*be a block diagonal matrix (with entries in  $\mathbb{K}$ ), and suppose that  $A$  and  $B$  are square matrices. Then*

$$\det(C) = \det(A) \det(B)$$

*Proof.* Suppose that  $A$  is  $n \times n$  and  $B$  is  $m \times m$ .

First consider the case of an  $n \times n$  elementary matrix  $E$ . Considering the three types of row operations, it is easy to see that

$$E' = \begin{bmatrix} E & 0 \\ 0 & I_m \end{bmatrix}$$

(where  $I_m$  is the  $m \times m$  identity matrix) is the matrix for the same type of elementary row operation as  $E$  and that  $\det(E') = \det(E)$ . Similarly, if  $F$  is an  $m \times m$  elementary matrix then

$$F' = \begin{bmatrix} I_n & 0 \\ 0 & F \end{bmatrix}$$

is an elementary matrix of the same type and  $\det(F') = \det(F)$ .

Now consider the matrix  $C$  in the Proposition. If  $A$  is noninvertible, there is a vector  $v \in \mathbb{C}^n$  such that  $v \neq 0$  but  $Av = 0$ . If let  $v' \in \mathbb{C}^{n+m}$  be the vector given in block form by

$$v' = \begin{bmatrix} v \\ 0 \end{bmatrix}$$

then  $Cv' = 0$ . Thus  $C$  is not invertible. Similarly, if  $B$  is not invertible,  $C$  is not invertible.

So, if one of  $A$  and  $B$  is not invertible, the equation  $\det(C) = \det(A)\det(B)$  is true because both sides are zero.

If both  $A$  and  $B$  are invertible, we start by writing  $A$  as a product of elementary matrices, say  $A = E_1 \cdots E_k$ . We know that  $\det(A) = \det(E_1) \cdots \det(E_k)$ . We then have

$$C = \begin{bmatrix} E_1 \cdots E_k & 0 \\ 0 & B \end{bmatrix} = \begin{bmatrix} E_1 & 0 \\ 0 & I_m \end{bmatrix} \cdots \begin{bmatrix} E_k & 0 \\ 0 & I_m \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & B \end{bmatrix}.$$

Similarly, we can write  $B = F_1 F_2 \cdots F_\ell$ . Then we have

$$\begin{aligned} C &= \begin{bmatrix} E_1 & 0 \\ 0 & I_m \end{bmatrix} \cdots \begin{bmatrix} E_k & 0 \\ 0 & I_m \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & F_1 \cdots F_\ell \end{bmatrix} \\ &= \begin{bmatrix} E_1 & 0 \\ 0 & I_m \end{bmatrix} \cdots \begin{bmatrix} E_k & 0 \\ 0 & I_m \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & F_1 \end{bmatrix} \cdots \begin{bmatrix} I_n & 0 \\ 0 & F_\ell \end{bmatrix}. \end{aligned}$$

From this equation and our observations above we conclude

$$\begin{aligned} \det(C) &= \det\left(\begin{bmatrix} E_1 & 0 \\ 0 & I_m \end{bmatrix}\right) \cdots \det\left(\begin{bmatrix} E_k & 0 \\ 0 & I_m \end{bmatrix}\right) \det\left(\begin{bmatrix} I_n & 0 \\ 0 & F_1 \end{bmatrix}\right) \cdots \det\left(\begin{bmatrix} I_n & 0 \\ 0 & F_\ell \end{bmatrix}\right) \\ &= [\det(E_1) \cdots \det(E_k)][\det(F_1) \cdots \det(F_\ell)] \\ &= \det(A)\det(B) \end{aligned}$$

□

**1.2. Algebra of Polynomials.** We need a few facts about the algebra of polynomials with complex coefficients. Some of these are standard college algebra, some may be new to the reader. Following mathematical practice we denote the set of all polynomials in the variable  $z$  with complex coefficients by  $\mathbb{C}[z]$ . Thus, a typical element  $p(z) \in \mathbb{C}[z]$  looks like

$$(1.3) \quad p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$$

for some  $n$ , where  $a_0, \dots, a_n \in \mathbb{C}$ . The exponent of the highest power of  $z$  that appears in  $p(z)$  (i.e., has a nonzero coefficient) is called the **degree** of  $p(z)$ , denoted by  $\deg(p(z))$ . The constant polynomials  $p(z) = a_0$  are in  $\mathbb{C}[z]$ . A **nonzero** constant polynomial has degree 0. By convention the degree of the zero constant polynomial is  $-\infty$ . The coefficient of the highest power of  $z$  in  $p(z)$  is called the **leading coefficient** of  $p(z)$ . If the leading coefficient of  $p(z)$  is 1, we say that  $p(z)$  is a **monic** polynomial.

The set  $\mathbb{C}[z]$  is closed under addition of polynomials and multiplication of polynomials. Since multiplication of polynomials includes multiplication by scalars,  $\mathbb{C}[z]$  is a complex vector space. With the degree conventions above, we have

$$\begin{aligned}\deg(p(z) + q(z)) &\leq \max(\deg(p(z)), \deg(q(z))) \\ \deg(p(z)q(z)) &= \deg(p(z)) + \deg(q(z)).\end{aligned}$$

1.2.1. *Division and Factoring.* In high school and college algebra, they teach you the division algorithm for polynomials. The results of this algorithm can be stated formally as follows.

**Theorem 1.8.** *Let  $p(z)$  be a polynomial and let  $d(z)$  be a nonzero polynomial (called the divisor). Then there are unique polynomials  $q(z)$  (the quotient) and  $r(z)$  (the remainder) such that*

$$p(z) = q(z)d(z) + r(z)$$

and  $\deg(r(z)) < \deg(d(z))$ .

If  $r(z) = 0$ , we have  $p(z) = q(z)d(z)$  and in this case we say that  $d(z)$  **divides**  $p(z)$ , which is written as  $d(z)|p(z)$  in symbols.

If  $p(z)$  is given by (1.3), we can plug a complex number  $\zeta$  in for  $z$  and get a complex number  $p(\zeta)$ . Thus,

$$p(\zeta) = a_n\zeta^n + a_{n-1}\zeta^{n-1} + \cdots + a_1\zeta + a_0 \in \mathbb{C}.$$

If  $p(\zeta) = 0$ , we say that  $\zeta$  is a **root** of  $p(z)$ .

**Proposition 1.9.** *Suppose that  $p(z) \in \mathbb{C}[z]$  and that  $\zeta$  is a root of  $p(z)$ . Then  $(z - \zeta)$  divides  $p(z)$ .*

*Proof.* Applying the division algorithm, we have

$$p(z) = q(z)(z - \zeta) + r(z).$$

Since  $\deg(z - \zeta) = 1$ , we must have  $\deg(r(z)) < 1$ , so  $r(z)$  must be a constant, call it  $c$ . The last equation then reads

$$p(z) = q(z)(z - \zeta) + c.$$

Plugging in  $\zeta$  for  $z$ , we see that  $0 = p(\zeta) = 0 + c$ , so  $c = 0$ . Thus, we have  $p(z) = q(z)(z - \zeta)$ , so  $(z - \zeta)|p(z)$ .  $\square$

The following theorem is called the Fundamental Theorem of Algebra. In particular, it asserts that every nonconstant polynomial with complex coefficients has at least one root; this is really a theorem of Analysis. The rest of the theorem follows from the last Proposition

**Theorem 1.10.** *Let  $p(z) \in \mathbb{C}[z]$  be a nonconstant polynomial, say*

$$p(z) = a_n z^n + \cdots + a_1 z + a_0, \quad a_n \neq 0.$$

*Then  $p(z)$  factors as*

$$(1.4) \quad p(z) = a_n (z - r_1)^{m_1} \cdots (z - r_k)^{m_k}$$

*where  $r_1, \dots, r_k$  are the roots of  $p(z)$ . The positive integer  $m_j$  is called the multiplicity of  $r_j$ , and we have  $m_1 + \cdots + m_k = n$ . The factorization (1.4) is unique, up to the order of the factors.*

If  $d(z)$  is another nonconstant polynomial of degree  $m$  (with leading coefficient  $d_m$ ) then  $d(z)$  divides  $p(z)$  if and only if

$$d(z) = d_m \prod_{j=1}^k (z - r_j)^{\ell_j}, \quad 0 \leq \ell_j \leq m_j$$

( $\ell_j = 0$  means, in effect, that that factor does not occur). In particular, if  $d(z)|p(z)$ , the roots of  $d(z)$  are among the roots of  $p(z)$  and the multiplicity of a root of  $d(z)$  is less than or equal to its multiplicity in  $p(z)$ .

### 1.2.2. Ideals and Minimal Polynomials.

**Definition 1.11.** A subset  $J$  of  $\mathbb{C}[z]$  is called an **ideal** if it satisfies the following conditions.

- (1)  $0 \in J$ .
- (2)  $J$  is closed under addition:  $p(z), q(z) \in J \implies (p(z) + q(z)) \in J$ .
- (3)  $J$  is absorbing under multiplication:  $p(z) \in \mathbb{C}[z], j(z) \in J \implies p(z)j(z) \in J$ .

Two trivial ideals are the whole set  $\mathbb{C}[z]$  and the subset  $\{0\}$ . For a less trivial example, consider the following.

**Example 1.12.** let  $\zeta \in \mathbb{C}$  be a fixed complex number and take

$$J = \{p(z) \in \mathbb{C}[z] \mid p(\zeta) = 0\}.$$

It should be easy to check that  $J$  is an ideal.

If  $g_1(z), g_2(z), \dots, g_k(z)$  are polynomials, **the ideal generated by  $g_1(z), g_2(z), \dots, g_k(z)$**  is the smallest ideal in  $\mathbb{C}[z]$  that contains the polynomials  $g_1(z), g_2(z), \dots, g_k(z)$ . This ideal is denoted  $(g_1(z), g_2(z), \dots, g_k(z))$ . It is easy to see that

$$(g_1(z), g_2(z), \dots, g_k(z)) = \{p_1(z)g_1(z) + p_2(z)g_2(z) + \dots + p_k(z)g_k(z) \mid p_j(z) \in \mathbb{C}[z], j = 1, \dots, k\}.$$

In particular, if we have a single polynomial  $g(z)$ , the ideal it generates is

$$(g(z)) = \{p(z)g(z) \mid p(z) \in \mathbb{C}[z]\},$$

i.e.,  $(g(z))$  is the set of all multiples of  $g(z)$  or, to put it another way,  $(g(z))$  is the set of all polynomials that are divisible by  $g(z)$ .

An important, and perhaps surprising, result is that *every* ideal in  $\mathbb{C}[z]$  is  $(g(z))$  for some polynomial  $g(z)$ . (The trivial ideals are  $(1) = \mathbb{C}[z]$  and  $(0) = \{0\}$ ).

**Exercise 1.13.** Verify that the ideal in Example 1.12 is  $(g(z))$  where  $g(z) = z - \zeta$ .

Here is the formal theorem, which introduces an important definition.

**Theorem 1.14.** Let  $J \subseteq \mathbb{C}[z]$  be an ideal. If  $J \neq (0)$ , there is a unique monic polynomial  $\mu(z)$  such that

$$(1.5) \quad \deg(\mu(z)) = \min\{\deg(p(z)) \mid p(z) \in J, p(z) \neq 0\}.$$

The polynomial  $\mu(z)$  generates  $J$ , i.e.  $J = (\mu(z))$ . Thus  $p(z)$  is in  $J$  if and only if  $p(z)$  is divisible by  $\mu(z)$ .

We will call  $\mu(z)$  the **minimal polynomial of  $J$** . If  $J \neq (1)$ , the minimal polynomial  $\mu(z)$  has degree greater than or equal to 1.

*Proof.* Since  $J \neq (0)$ , it contains nonzero polynomials. Thus,

$$\{\deg(p(z)) \mid p(z) \in J, p(z) \neq 0\}$$

is a nonempty set of nonnegative integers and we can choose a polynomial  $g(z) \in J$  such that

$$\deg(g(z)) = \min\{\deg(p(z)) \mid p(z) \in J, p(z) \neq 0\}.$$

If  $\deg(g(z)) = 0$ ,  $g(z)$  is a constant, say  $g(z) = c$ , so  $c \in J$ . But then every polynomial  $p(z)$  is in  $J$ , since  $p(z) = (p(z)/c)c$  and  $J$  is absorbing under multiplication. In this case  $J = (c) = (1)$ . The monic polynomial of minimal degree in  $J$  is  $\mu(z) = 1$ . For the rest of the proof, we can assume  $\deg(g(z)) \geq 1$ .

Suppose that the leading coefficient of  $g(z)$  is  $c$ . Since  $J$  is absorbing under multiplication  $\mu(z) = (1/c)g(z)$  is in  $J$ . Thus,  $\mu(z)$  is a monic polynomial in  $J$  such that (1.5) holds.

Certainly any multiple of  $\mu(z)$  is in  $J$ . We need to prove the converse, that every polynomial in  $J$  is a multiple of  $\mu(z)$ . Of course, the zero polynomial is a multiple of  $\mu(z)$ .

Suppose that  $p(z) \neq 0$  is in  $J$ . By the Division Algorithm, we have

$$p(z) = q(z)\mu(z) + r(z)$$

where  $\deg(r(z)) < \deg(\mu(z))$ . But  $r(z) = p(z) - q(z)\mu(z)$  is an element of  $J$ . We can't have  $r(z) \neq 0$ , because this would contradict (1.5). Thus,  $r(z) = 0$ , which means that  $p(z) = q(z)\mu(z)$ , so  $p(z)$  is a multiple of  $\mu(z)$ .

Finally, suppose that  $m(z)$  is a monic polynomial in  $J$  with of the same degree as  $\mu(z)$ . From our results  $m(z)$  is divisible by  $\mu(z)$ . Since the degrees are the same, the quotient must be a constant. Thus,  $m(z) = c\mu(z)$ . Comparing the leading coefficients on each side of this equation, we see that  $c$  must be 1, which means that  $m(z) = \mu(z)$ .  $\square$

**Exercise 1.15.** What is the minimal polynomial of the ideal in Example 1.12?

## 2. THE STRUCTURE OF A LINEAR TRANSFORMATION

Let  $V$  be a vector space over  $\mathbb{K}$  and let  $T: V \rightarrow V$  be a linear transformation. Recall that it makes sense to talk about the determinant of  $T$ . If we choose a basis of  $V$ , we can find the matrix of  $T$  with respect to this basis, call it  $A$ . If we choose a different basis, we will get a different matrix  $B$ . The two matrices are related by  $B = P^{-1}AP$ , where  $P$  is a change of basis matrix. But then

$$\det(B) = \det(P^{-1})\det(A)\det(P) = \frac{1}{\det(P)}\det(A)\det(P) = \det(A).$$

Thus, the determinant of the matrix of  $T$  is independent of which basis we use and the common value is called the determinant of  $T$ .

We will use  $I$  to stand for both an identity matrix and an identity linear transformation.

Since the determinant of a linear transformation is well defined, we can consider the function  $\chi_T(z) = \det(T - zI)$ . If we choose a basis and  $T$  has the matrix  $A$  in this basis,  $T - zI$  will have the matrix  $A - zI$ . Thus,  $\chi_T(z) = \det(A - zI)$ . The right-hand side of this equation is the characteristic polynomial of the matrix  $A$ . Thus,  $\chi_T(z)$  is a polynomial, called the characteristic polynomial of  $T$ . It is equal to the characteristic polynomial  $\chi_A(z)$  of any matrix representation  $A$  of  $T$ .

If

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0$$

is a polynomial with complex coefficients and  $A$  is a matrix, we can substitute  $A$  for  $z$  in  $p(z)$  to get another matrix:

$$p(A) = a_n A^n + a_{n-1} A^{n-1} + \dots + a_1 A + a_0 I.$$

Similarly, we can plug a linear transformation  $T$  into  $p(z)$  to get a linear transformation

$$p(T) = a_n T^n + a_{n-1} T^{n-1} + \dots + a_1 T + a_0 I.$$

The algebraic operations on matrices are chosen to correspond to the algebraic operations on linear transformations, so if  $T$  has matrix representation  $A$  with respect to some basis, the matrix representation of  $p(T)$  will be  $p(A)$ .

Given a linear transformation  $T$  (or a matrix) we can define a set of polynomials  $J_T$  by

$$J_T = \{p(z) \in \mathbb{C}[z] \mid p(T) = 0\}.$$

A little thought shows that  $J_T$  is an ideal in  $\mathbb{C}[z]$ . The minimal polynomial of this ideal is called the minimal polynomial of  $T$  and is denoted by  $\mu_T(z)$ . Thus,  $p(T) = 0$  if and only if  $\mu_T(z) \mid p(z)$ .

If  $v \in V$  is a nonzero vector, we can define a set of polynomials by

$$J_v = \{p(z) \in \mathbb{C}[z] \mid p(T)v = 0\}.$$

Again  $J_v$  is an ideal. The minimal polynomial of this ideal, denoted  $\mu_v(z)$  (when  $T$  is understood) is called the minimal polynomial of  $v$  with respect to  $T$ . Since  $\mu_T(T)v = 0$ , we have  $\mu_v(z) \mid \mu_T(z)$ .

**Proposition 2.1.** *The roots of the minimal polynomial of  $T$  are precisely the eigenvalues of  $T$ .*

*Proof.* Suppose that  $\lambda$  is an eigenvalue of  $T$ . Then there is a nonzero vector  $v$  so that  $(T - \lambda I)v = 0$ . Thus,  $\mu_v(z) = z - \lambda$ . We conclude that  $(z - \lambda) \mid \mu_T(z)$ , so  $\lambda$  is a root of  $\mu_T(z)$ .

To do the other direction, factor  $\mu_T(z)$  as

$$\mu_T(z) = \prod_{j=1}^{\ell} (z - r_j)^{t_j}$$

where the  $r_j$ 's are the roots of the minimal polynomial. Fix a root  $r_k$  and define a polynomial  $q(z)$  by

$$q(z) = (z - r_k)^{t_k - 1} \prod_{\substack{j=1 \\ j \neq k}}^{\ell} (z - r_j)^{t_j}.$$

We then have  $q(z) \neq 0$  and  $\deg(q(z)) < \deg(\mu_T(z))$ . Thus,  $q(T) \neq 0$ . We can find a vector  $v$  such that  $q(T)v \neq 0$ . But then  $(T - r_k)q(T)v = \mu_T(T)v = 0$ . Thus,  $q(T)v$  is an eigenvector of  $T$  with eigenvalue  $r_k$ . This shows that all of the roots of  $\mu_T(z)$  are eigenvalues.  $\square$

**2.1. Upper Triangularization and the Cayley-Hamilton Theorem.** In this subsection, we will prove the following two theorems together.

**Theorem 2.2.** [Upper Triangularization Theorem] *Let  $V$  be a complex vector space and let  $T: V \rightarrow V$  be a linear transformation. Then there is an ordered basis of  $V$  such that the matrix representation of  $T$  with respect to this basis is upper triangular.*

**Theorem 2.3.** [Cayley-Hamilton Theorem] *Every linear transformation (and matrix) satisfies its characteristic polynomial, i.e.  $\chi_T(T) = 0$ , where  $\chi_T(z) = \det(T - zI)$ .*

To prove these theorems, we start with a definition. Suppose that  $T: V \rightarrow V$ , where  $V$  is a complex vector space of dimension  $n$ . A **fan** for  $T$  is a sequence  $\{V_j\}_{j=1}^n$  such that

$$V_1 \subseteq V_2 \subseteq \dots \subseteq V_{n-1} \subseteq V_n = V,$$

where  $\dim(V_j) = j$  and  $T(V_j) \subseteq V_j$ .

Our first step is the following lemma.

**Lemma 2.4.** *If  $T$  is a linear transformation on a complex vector space  $V$ , there exists a fan for  $T$ .*

*Proof of Lemma.* The proof is by induction on the dimension  $n$  of  $V$ . If  $V$  has dimension one, the lemma is trivially true: the fan is  $\{V\}$ .

Suppose now that the Lemma is true for vector spaces of dimension  $n - 1$ , and suppose that  $T: V \rightarrow V$  where  $V$  has dimension  $n$ .

Since we are working over the complex numbers,  $T$  has an eigenvalue  $\lambda$  with eigenvector  $v_1$  (the characteristic polynomial has at least one complex root). Let  $V_1 = \text{span}(v_1)$ .

Now choose a subspace  $W \subseteq V$  so that  $V = V_1 \oplus W$ . This can be done as follows. The collection of vectors  $v_1$  is linearly independent and so can be completed to a basis of  $V$ , say  $v_1, w_1, \dots, w_{n-1}$ . Let  $W = \text{span}(w_1, \dots, w_{n-1})$ . Observe that  $W$  has dimension  $n - 1$ .

The direct sum decomposition  $V = V_1 \oplus W$  means that every vector  $v \in V$  can be written uniquely in the form  $v = \zeta v_1 + w$  for a scalar  $\zeta$  and a vector  $w \in W$ . Let  $P$  be the projection onto  $V_1$  and let  $Q$  be the projection onto  $W$ . Thus,  $P$  and  $Q$  are defined by

$$P(\zeta v_1 + w) = \zeta v_1$$

$$Q(\zeta v_1 + w) = w.$$

Since  $QT(V) \subseteq W$ , we certainly have  $QT(W) \subseteq W$ . Thus, the restriction of  $QT$  to  $W$  is a linear transformation on  $W$ , which has dimension  $n - 1$ . By the induction hypothesis there is a fan for this linear transformation. Thus, we can find subspaces  $\{W_j\}_{j=1}^{n-1}$  such that  $\dim(W_j) = j$ ,

$$W_1 \subseteq W_2 \subseteq \dots \subseteq W_{n-1} = W$$

and  $QT(W_j) \subseteq W_j$ .

Now define subspaces  $V_1, V_2, \dots, V_n$  of  $V$  by

$$(2.1) \quad \begin{aligned} V_1 &= V_1 \text{ (already defined)} \\ V_2 &= V_1 + W_1 \\ V_3 &= V_1 + W_2 \\ &\vdots \\ V_n &= V_1 + W_{n-1}. \end{aligned}$$

We claim these subspaces form a fan for  $T$ . Obviously we have  $V_1 \subseteq \dots \subseteq V_n = V$ . Since  $V = V_1 \oplus W$ , it is clear that the sums in (2.1) are direct. Thus, we have  $\dim(V_j) = j$ . It remains to show that the subspaces  $V_j$  are invariant under  $T$ . Clearly  $V_1$  is invariant under  $T$ , since  $v_1$  is an eigenvector. Consider the case of  $V_j$ ,  $j > 1$ . A typical element  $v$  of  $V_j$  looks like  $v = \alpha v_1 + w$  where  $w \in W_{j-1}$ . We can write  $T(v) = PT(v) + QT(v)$ . Of course,  $PT(v) \in V_1$ , so  $PT(v) = \beta v_1$  for some scalar  $\beta$ . For  $QT(v)$ , we have

$$\begin{aligned} QT(v) &= QT(\alpha v_1 + w) \\ &= Q(\alpha T(v_1) + T(w)) \\ &= Q(\alpha \lambda v_1 + T(w)) \\ &= 0 + QT(w) \\ &= QT(w) \in W_{j-1} \end{aligned}$$

since  $W_{j-1}$  is invariant under  $QT$ . Thus,  $T(v) = \beta v_1 + w'$ , where  $w' \in W_{j-1}$ , so  $T(v) \in V_j$ .  $\square$

We are now ready to show that there is a basis so that the matrix representation of  $T$  with respect to this basis is upper triangular. To see this, choose a fan

$$V_1 \subseteq V_2 \subseteq \dots \subseteq V_n = V$$

for  $T$ . Since  $V_1$  is of dimension one, we can choose a vector  $v_1$  that spans  $V_1$ . Since  $V_2$  is two dimensional, we must have  $V_1 \subsetneq V_2$ . Choose a vector  $v_2 \in V_2 \setminus V_1$ . We claim that the vectors  $v_1, v_2$  are independent. To see this, suppose that  $c_1 v_1 + c_2 v_2 = 0$ . If  $c_2 = 0$ , we can conclude  $c_1 = 0$ , since  $v_1 \neq 0$ . But if  $c_2 \neq 0$ , we would have  $v_2 = (-c_1/c_2)v_1 \in V_1$ , which contradicts our choice of  $v_2$ . Thus,  $v_1$  and  $v_2$  are independent. Since  $V_2$  is two dimensional,  $v_1, v_2$  form a basis of  $V_2$ .

We proceed inductively in this way. Suppose we have chosen vectors  $v_1, v_2, \dots, v_k$  such that  $v_1, v_2, \dots, v_j$  is a basis of  $V_j$  for  $j = 1, \dots, k$ . Since  $V_{k+1}$  has dimension  $k+1$ , we must have  $V_k \subsetneq V_{k+1}$ , so we can choose a vector  $v_{k+1} \in V_{k+1} \setminus V_k$ . We claim that the vectors  $v_1, \dots, v_k, v_{k+1}$  are independent. To see this, suppose we have a relation

$$c_1 v_1 + c_2 v_2 + \dots + c_k v_k + c_{k+1} v_{k+1} = 0.$$

If  $c_{k+1} = 0$ , we can conclude the rest of the  $c_j$ 's are zero, since the vectors  $v_1, \dots, v_k$  are independent. But, if  $c_{k+1} \neq 0$ , we would have

$$v_{k+1} = (-c_1/c_{k+1})v_1 + \dots + (-c_k/c_{k+1})v_k \in V_k,$$

which contradicts our choice of  $v_{k+1}$ . Thus, the vectors  $v_1, \dots, v_k, v_{k+1}$  are independent and so must be a basis of  $V_{k+1}$ .

With this procedure, we have constructed a basis  $v_1, \dots, v_n$  such that  $v_1, \dots, v_j$  is a basis for  $V_j$ ,  $j = 1, \dots, n$ . Let  $A$  be the matrix of  $T$  with respect to this basis.

Column  $j$  of  $A$  consists of the coefficients that tell us how to express  $T(v_j)$  in terms of the basis vectors. But  $v_j \in V_j$  and  $V_j$  is invariant under  $T$ , so  $T(v_j) \in V_j = \text{span}(v_1, \dots, v_j)$ . Thus we must have

$$(2.2) \quad T(v_j) = \sum_{k=1}^j v_k a_{kj},$$

so the entries  $a_{kj}$  are zero for  $k > j$ . In other words, the matrix  $A$  looks like

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

and so is upper triangular.

Since  $A$  is upper triangular, the eigenvalues are  $a_{11}, \dots, a_{nn}$  (repeated with their multiplicities). The characteristic polynomial of  $A$ , which is the same as the characteristic polynomial of  $T$ , is evidently

$$\chi(z) = (a_{11} - z)(a_{22} - z) \cdots (a_{nn} - z).$$

We want to prove that  $\chi(T) = 0$ . To do this, let

$$\chi_k(z) = \prod_{j=1}^k (a_{jj} - z)$$

We will show inductively that  $\chi_k(T)V_k = 0$ . The case  $k = n$  will then give us  $\chi(T)V = \chi_n(T)v_n = 0$ .

We have  $\chi_1(T)V_1 = (a_{11} - T)V_1 = 0$ , since  $V_1 = \text{span}(v_1)$  and  $T(v_1) = a_{11}v_1$ .

For the induction step, suppose we know  $\chi_{k-1}(T)V_{k-1} = 0$  and consider  $\chi_k(T)V_k$ . Since  $v_1, \dots, v_{k-1}, v_k$  are a basis of  $V_k$ , it will suffice to show that  $\chi_k(T)$  sends each of these basis vectors to zero. We have

$$\chi_k(T)v_j = (a_{kk}I - T)\chi_{k-1}(T)v_j = 0, \quad j = 1, \dots, k-1$$

by the induction hypothesis. From (2.2) we have

$$T(v_k) = a_{kk}v_k + \sum_{j=1}^{k-1} v_j a_{jk}$$

and so

$$(a_{kk}I - T)v_k = w \in V_{k-1},$$

so

$$\chi_k(T)v_k = \chi_{k-1}(T)(a_{kk}I - T)v_k = \chi_{k-1}(T)w = 0,$$

by the induction hypothesis. This completes the proof of both our theorems.

**2.2. The Structure of a Linear Transformation.** In this subsection,  $V$  will be a complex vector space of dimension  $n$  and  $T: V \rightarrow V$  will be a linear transformation. Let  $\lambda_1, \dots, \lambda_r$  be the distinct eigenvalues of  $T$ .

The characteristic polynomial of  $T$  is  $\chi_T(z) = \det(T - zI)$ , which has degree  $n$ . Let  $a_n$  be the leading coefficient of  $\chi_T(z)$  (which is  $(-1)^n$ ). Then the characteristic polynomial factors as

$$(2.3) \quad \chi_T(z) = a_n \prod_{j=1}^r (z - \lambda_j)^{m_j}$$

and we have

$$m_1 + \dots + m_r = n.$$

Let  $E_j$  denote the eigenspace corresponding to  $\lambda_j$ , so

$$E_j = N(T - \lambda_j I).$$

We define the **generalized eigenspace  $G_j$  corresponding to  $\lambda_j$**  to be the set of all vectors  $v \in V$  such that  $(T - \lambda_j I)^p v = 0$  for some power  $p$ . In this case,  $(T - \lambda_j I)^q v = 0$  for  $q > p$ .

Suppose that  $v \neq 0$  and  $p$  is the smallest power so that  $(T - \lambda_j I)^p v = 0$ . Then the minimal polynomial of  $v$  must be  $\mu_v(z) = (z - \lambda_j)^p$ . The minimal polynomial of  $v$  divides the minimal polynomial of  $T$ , which in turn divides the characteristic polynomial of  $T$  (2.3). Thus, we must have  $p \leq m_j \leq n$ . In summary, if  $(T - \lambda_j I)^p v = 0$  for any power  $p$ , we must have  $(T - \lambda_j I)^n v = 0$ . Thus, the generalized eigenspace  $G_j$  can be described as

$$G_j = N((T - \lambda_j I)^n).$$

Since  $T$  commutes with any polynomial in  $T$ , we have  $T(G_j) \subseteq G_j$ . Of course,  $E_j \subseteq G_j$ .

**Proposition 2.5.** *We have*

$$V = G_1 \oplus \dots \oplus G_r.$$

Before proving the Proposition, we prove a Lemma.

**Lemma 2.6.** *For each  $k$ , define*

$$P_k(z) = \prod_{\substack{j=1 \\ j \neq k}}^r (z - \lambda_j)^{m_j}.$$

*Then,*

- (1)  $P_k(T)G_\ell = 0$  for  $\ell \neq k$ .
- (2) If  $v \in G_k$  and  $v \neq 0$ , then  $P_k(T)v \neq 0$ .

*Proof of lemma.* The first part of the Lemma follows because  $P_k(T)$  contains the factor  $(T - \lambda_\ell)^{m_\ell}$ , which annihilates  $G_\ell$ , as we saw above (you can apply the factors in any order).

For the second part, suppose  $v \in G_k$  and  $v \neq 0$ . By definition of  $G_k$ ,  $(T - \lambda_k I)^p(v) = 0$  for some power  $p$ . Assume  $p$  is the least such integer. Then the minimal polynomial of  $v$  must be  $(z - \lambda_k)^p$ . If  $P_k(T)v = 0$ , the minimal polynomial of  $v$  would divide  $P_k(z)$ . But this is not the case, so  $P_k(T)v \neq 0$ .  $\square$

*Proof of Proposition.* Let  $G = G_1 + \cdots + G_r$ . We first want to prove that this sum is direct. So, suppose we have

$$v_1 + v_2 + \cdots + v_r = 0, \quad v_j \in G_j.$$

Fix  $k$  and apply  $P_k(T)$  to both sides of this equation. By the Lemma, the result is

$$P_k(T)v_k = 0,$$

and, by the lemma, this implies that  $v_k = 0$ . Since  $k$  was arbitrary, we conclude that  $v_1 = v_2 = \cdots = v_r = 0$ , so the sum is direct.

Next, we want to prove that  $V = G$ . Suppose, for a contradiction, that  $G \subsetneq V$ . Then, we can find a vector  $v_0 \in V \setminus G$  such that

$$\deg(\mu_{v_0}(z)) = \min\{\deg(\mu_v(z)) \mid v \in V \setminus G\}.$$

We can factor  $\mu_{v_0}$  as

$$\mu_{v_0}(z) = \prod_{j=1}^r (z - \lambda_j)^{h_j}, \quad 0 \leq h_j \leq m_j.$$

Consider first the case where two of the exponents are nonzero, say  $h_k > 0$  and  $h_\ell > 0$ , where  $k \neq \ell$ .

Let  $Q_k(z)$  be the polynomial

$$Q_k(z) = (z - \lambda_k)^{h_k - 1} \prod_{\substack{j=1 \\ j \neq k}}^r (z - \lambda_j)^{h_j}.$$

First note that  $Q_k(T)v_0 \neq 0$ , since  $Q_k(z)$  is a nonconstant polynomial of degree less than the degree of the minimal polynomial  $\mu_{v_0}(z)$  of  $v_0$ . We also have

$$Q_k(T)[(T - \lambda_k I)v_0] = \mu_{v_0}(T)v_0 = 0$$

but this means that  $(T - \lambda_k)v_0 \in G$ , since its minimal polynomial has degree strictly less than the degree of  $\mu_{v_0}(z)$ , and  $v_0$  was chosen to minimize the degree of the minimal polynomial among all vectors not in  $G$ . If we set  $(T - \lambda_k I)v_0 = g_1 \in G$ , we have  $Tv_0 = \lambda_k v_0 + g_1$ .

We can apply exactly the same argument with the index  $\ell$ , to conclude  $T(v_0) = \lambda_\ell v_0 + g_2$ , where  $g_2 \in G$ . Subtracting the equations

$$\begin{aligned} T(v_0) &= \lambda_k v_0 + g_1 \\ T(v_0) &= \lambda_\ell v_0 + g_2 \end{aligned}$$

gives

$$0 = (\lambda_k - \lambda_\ell)v_0 + g_1 - g_2.$$

Since  $\lambda_k \neq \lambda_\ell$ , we can solve this equation to give

$$v_0 = \frac{1}{\lambda_k - \lambda_\ell}(g_1 - g_2) \in G.$$

But this contradicts the choice of  $v_0$ !

The remaining case would be the case where only one exponent in  $\mu_{v_0}(z)$  is not zero, say  $h_k > 0$ . But then  $\mu_{v_0} = (z - \lambda_k)^{h_k}$ , so  $v_0$  is annihilated by a power of  $(T - \lambda_k I)$ . But this is exactly the definition of  $v_0 \in G_k$ , which again contradicts the choice of  $v_0$ .

Thus, our assumption  $G \subsetneq V$  leads to a contradiction, so we conclude  $G = V$ .  $\square$

Next, we need a little more machinery. If  $W$  is a vector space over  $\mathbb{K}$  and  $N: W \rightarrow W$ , we say that  $N$  is **nilpotent** if  $N^p = 0$  for some power  $p$ . If  $p$  is the least such integer, so  $N^p = 0$  but  $N^{p-1} \neq 0$ , we say that  $p$  is **the degree of nilpotency of  $N$** . In this case, the minimal polynomial of  $N$  is  $\mu_N(z) = z^p$ . Thus,  $N$  has exactly one eigenvalue, which is 0.

Observe that the upper triangularization theorem says we can find a basis of  $W$  so that the matrix, call it  $A$ , of  $N$  with respect to this basis is upper triangular. But the diagonal entries of  $A$  are the eigenvalues of  $N$ , and so must all be zero. We can say that  $A$  is **strictly upper triangular**.

**Exercise 2.7.** If  $A$  is a strictly upper triangular matrix square matrix, show that  $A$  is nilpotent, i.e.,  $A^p = 0$  for some  $p$ . Hint: what does  $A$  do to the standard basis vectors?

**Exercise 2.8.** Let  $N_1, N_2: W \rightarrow W$  be nilpotent transformations which commute, i.e.,  $N_1N_2 = N_2N_1$ . Show that  $N_1 + N_2$  is nilpotent. Hint: you can expand  $(N_1 + N_2)^p$  by the binomial theorem, i.e.,

$$(N_1 + N_2)^p = \sum_{k=0}^p \binom{p}{k} N_1^k N_2^{p-k}.$$

**Proposition 2.9.** *The dimension of  $G_j$  is  $m_j$ . If  $T_j: G_j \rightarrow G_j$  is the linear transformation induced by  $T$ , the characteristic polynomial of  $T_j$  is  $(\lambda_j - z)^{m_j}$ .*

*Proof.* Temporally, denote the dimension of  $G_j$  by  $d_j$ . By definition,  $N_j = T_j - \lambda_j I$  is nilpotent on  $G_j$ . We can find a basis of  $G_j$  so that the matrix  $A_j$  of  $N_j$  is strictly upper triangular. Thus, the matrix of  $T_j$  with respect to this basis is  $B_j = \lambda_j I + A_j$ , which is upper triangular and has size  $d_j \times d_j$ . The diagonal entries are all  $\lambda_j$ . Thus, the characteristic polynomial of  $B_j$  is  $(\lambda_j - z)^{d_j}$ . If we do this on each  $G_j$  and put the resulting basis together into a basis for  $V$ , we see that the matrix of  $T$  with respect to this basis will have the block diagonal form

$$\begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_r \end{bmatrix}$$

with all other entries zero. Thus, the characteristic polynomial of  $T$  is the product of the characteristic polynomials of the blocks, so

$$\chi_T(z) = \prod_{j=1}^r (\lambda_j - z)^{d_j}.$$

But the factorization of a polynomial into linear factors is unique, so, comparing with (2.3), we must have  $d_j = m_j$ .  $\square$

**Proposition 2.10.** *The transformation  $T$  is diagonalizable if and only if  $E_j = G_j$  for  $j = 1, \dots, r$ .*

*The transformation  $T$  is diagonalizable if and only if the minimal polynomial of  $T$  is*

$$\mu_T(z) = \prod_{j=1}^r (z - \lambda_j),$$

*i.e., the eigenvalues all have multiplicity one as roots of the minimal polynomial.*

*Proof.* We know that  $V = G_1 \oplus \cdots \oplus G_r$  and  $E_j \subseteq G_j$ . If  $E_j = G_j$  for all  $j$ , then  $V = E_1 \oplus \cdots \oplus E_r$ . If we choose a basis for each  $E_j$ , we get a basis of  $V$  consisting of eigenvectors, so  $T$  is diagonalizable.

Conversely, suppose  $T$  is diagonalizable. Then choosing a basis of each eigenspace gives us a basis of eigenvectors, so we must have

$$(2.4) \quad \dim(V) = \dim(E_1) + \cdots + \dim(E_r).$$

We know that

$$(2.5) \quad \dim(V) = \dim(G_1) + \cdots + \dim(G_r).$$

Subtracting (2.4) from (2.5) yields

$$(2.6) \quad (\dim(G_1) - \dim(E_1)) + \cdots + (\dim(G_r) - \dim(E_r)) = 0.$$

Since  $E_j \subseteq G_j$ , we have  $\dim(G_j) - \dim(E_j) \geq 0$ , so (2.6) implies that  $\dim(E_j) = \dim(G_j)$  for all  $j$ , and this implies that  $E_j = G_j$  for all  $j$ .

For the second part of the proof, suppose first that  $T$  is diagonalizable. Then  $E_j = G_j$  for all  $j$ . But then  $(T - \lambda_j I)G_j = 0$ . Since  $V$  is the direct sum of the  $G_j$ 's, we see that  $p(T) = 0$ , where

$$p(z) = \prod_{j=1}^r (z - \lambda_j).$$

Since all of the eigenvalues must be roots of  $\mu_T(z)$ , we see that  $p(T)$  has the minimal possible degree for a polynomial that annihilates  $T$ . Thus,  $\mu_T(z) = p(z)$ .

Conversely, suppose that

$$\mu_T(z) = \prod_{j=1}^r (z - \lambda_j).$$

Suppose that  $v \in G_j$ . Then  $(T - \lambda_j I)^p v = 0$  for some power  $p$ . If we take  $p$  to be as small as possible, then the minimal polynomial  $\mu_v(z)$  must be  $(z - \lambda_j)^p$ . But,  $\mu_v(z)$  divides  $\mu_T(z)$ , so we must have  $p = 1$ . But this means that  $v$  is an eigenvector. We conclude that  $G_j = E_j$  for all  $j$ , so  $T$  is diagonalizable.  $\square$

**Proposition 2.11.** *There are unique transformations  $S$  and  $N$  such that the following three conditions hold:*

- (1)  $T = S + N$ .
- (2)  $SN = NS$
- (3)  $S$  is diagonalizable and  $N$  is nilpotent.

*Proof.* Since  $V$  is the direct sum of the generalized eigenspaces and the generalized eigenspaces are invariant under  $T$ , we can write  $T = T_1 \oplus \cdots \oplus T_r$  where  $T_k: G_k \rightarrow G_k$  is the transformation induced by  $T$ .

By the definition of  $G_k$ ,  $T_k - \lambda_k I$  is nilpotent. If we set  $S_k = \lambda_k I$  on  $G_k$  and  $N_k = T_k - \lambda_k I$ , we have  $T_k = S_k + N_k$  and  $S_k N_k = N_k S_k$ .

Now set  $S = S_1 \oplus \cdots \oplus S_r$  and  $N = N_1 \oplus \cdots \oplus N_r$ . It is easy to see that  $T = S + N$  and  $SN = NS$ . Clearly,  $G_k$  is the  $\lambda_k$ -eigenspace of  $S$ , so  $V$  is the direct sum of the eigenspaces of  $S$ . Thus,  $S$  is diagonalizable. To see that  $N$  is nilpotent, note that  $N^p = N_1^p \oplus \cdots \oplus N_r^p$ , so if we let  $p$  be a common multiple of the degrees of nilpotency of the  $N_j$ 's, we will have  $N^p = 0$ .

To prove the decomposition  $T = S + N$  is unique, suppose that  $S'$  and  $N'$  are transformations which satisfy the three conditions of the Proposition. It is clear

that  $S'$  and  $N'$  commute with  $T$ , and hence leave each generalized eigenspace  $G_j$  invariant. Thus, we have  $S' = S'_1 \oplus \cdots \oplus S'_r$  and  $N' = N'_1 \oplus \cdots \oplus N'_r$  and it will suffice to show that  $S'_k = S_k$  and  $N'_k = N_k$ .

On  $G_k$ ,  $S_k = \lambda_k I$ , so clearly  $S_k S'_k = S'_k S_k$ . From this, we can deduce that  $N'_k = T_k - S'_k$  and  $N_k = T_k - S_k$  commute also. We have  $T_k = S_k + N_k = S'_k + N'_k$ , so  $S'_k - S_k = N_k - N'_k$ . Exercise 2.8 shows that  $N_k - N'_k$  is nilpotent, and so  $S'_k - S_k = S'_k - \lambda_k I$  is nilpotent, say  $(S'_k - \lambda_k I)^q = 0$ , where  $q$  is the least such power.

Let  $v$  be a nonzero vector in  $G_k$ . Then  $(S' - \lambda_k I)^q v = (S'_k - \lambda_k I)^q v = 0$ . Thus, the minimal polynomial of  $v$  with respect to  $S'$  must be  $(z - \lambda_k)^t$  for some  $t$  with  $1 \leq t \leq q$ . But  $\mu_v(z)$  must divide  $\mu_{S'}(z)$ , the minimal polynomial of  $S'$ . Since  $S'$  is diagonalizable, all the factors of  $\mu_{S'}(z)$  have multiplicity one. Thus, we must have  $t = 1$ , which means that  $(S'_k - \lambda_k I)v = (S' - \lambda_k I)v = 0$ . Since  $v$  was arbitrary, we must have  $S'_k = \lambda_k I = S_k$  and hence  $N'_k = N_k$  on  $G_k$ . □

**2.3. Jordan Canonical Form (Optional).** This section may be regarded as optional. The material is not really necessary for our treatment of differential equations. On the other hand, this is a standard topic in linear algebra and we're almost there, so it would be a shame not to include it.

Let  $T: V \rightarrow V$  be a linear transformation on a complex vector space. According to the last subsection,  $T$  decomposes uniquely as  $T = S + N$  where  $SN = NS$ ,  $S$  is diagonalizable and  $N$  is nilpotent.

Our first goal is to get a canonical form for nilpotent transformations. In other words, if  $N: V \rightarrow V$  is nilpotent, we will show that we can choose a basis so that the matrix of  $N$  has a particular form.

We will outline an algorithm for finding this basis. The idea of the algorithm is pretty easy, but the notation to cover the general case would be a real mess of indices. What we'll do is demonstrate the basics of the algorithm using specific numbers for some of the dimensions involved, and hope the general case will be obvious from our example.

So, suppose that  $N: V \rightarrow V$  is nilpotent. To be specific, suppose that  $N^3 = 0$  and  $N^2 \neq 0$ . We then have the following chain of increasing subspaces

$$N^2(V) \subseteq N(V) \subseteq V.$$

We will construct our basis by working up the chain, starting with  $N^2(V)$ .

To start with, choose any basis of  $N^2(V)$ . For definiteness, let's say we need two basis vectors, call them  $a_1$  and  $b_1$ . Since everything in  $N^2(V)$  gets sent to 0 by  $N$ , we have  $Na_1 = 0$  and  $Nb_1 = 0$ .

In the next step, we construct a basis for  $N(V)$ . Since  $N$  maps  $N(V)$  onto  $N^2(V)$ , we can find vectors  $a_2$  and  $b_2$  in  $N(V)$  such that  $Na_2 = a_1$  and  $Nb_2 = b_1$ . Define

$$K_1 = \{v \in N(V) \mid Nv = 0\}$$

Since  $a_1, b_1 \in N^2(V) \subseteq N(V)$  and  $Na_1 = 0$ ,  $Nb_1 = 0$ , we have  $a_1, b_1 \in K_1$  and we know that  $a_1$  and  $b_1$  are linearly independent. Thus, we can complete the list  $a_1, b_1$  to a basis of  $K_1$ . For definiteness, suppose we need two more basis vectors, so we get a basis  $a_1, b_1, c_1, d_1$  of  $K_1$ .

The situation thus far can be summarized in the following table:

$$(2.7) \quad \begin{array}{c|cccc} V & & & & \\ N(V) & a_2 & b_2 & c_1 & d_1 \\ N^2(V) & a_1 & b_1 & & \end{array}$$

Here each vector is in the subspace to the left, and hence in all the subspaces above that one. Each vector maps to the one below it under  $N$ . If there is nothing below the vector, it maps to zero under  $N$ .

We claim that the vectors in (2.7) are linearly independent. To see this, suppose that we have a linear relation

$$(2.8) \quad \alpha_1 a_1 + \beta_1 b_1 + \alpha_2 a_2 + \beta_2 b_2 + \gamma_1 c_1 + \delta_1 d_1 = 0.$$

If we apply  $N$  to both and we are left with

$$\alpha_2 a_1 + \beta_2 b_1 = 0.$$

Since  $a_1$  and  $b_1$  are independent, we must have  $\alpha_2 = \beta_2 = 0$ . Plugging this into (2.8), we have

$$(2.9) \quad \alpha_1 a_1 + \beta_1 b_1 + \gamma_1 c_1 + \delta_1 d_1 = 0.$$

But,  $a_1, b_1, c_1, d_1$  are a basis of  $K_1$ , so all the coefficients in (2.9) must be zero. Thus, all the coefficients in (2.8) are zero, so these vectors are linearly independent.

We next claim that the vectors in (2.7) are a basis of  $N(V)$ . By the table, the dimension of  $N^2(V)$  is two and the dimension of  $K_1$  is 4. The transformation  $N_1: N(V) \rightarrow N^2(V)$  induced by  $N$  has rank (dimension of the range) equal to  $\dim(N^2(V)) = 2$ , since it maps onto  $N^2(V)$ . Its nullspace is  $K_1$ , which has dimension 4. The rank theorem for  $N_1$  then says  $\dim(N(V)) = \text{rank}(N_1) + \dim(\text{nullspace}(N_1)) = 2 + 4 = 6$ . Thus,  $N(V)$  has dimension 6, so our six linearly independent vectors in  $N(V)$  must form a basis of  $N(V)$ .

For the next step, we construct a basis of  $V$ . Since  $N$  maps  $V$  onto  $N(V)$ , we can find  $a_3, b_3, c_2, d_2$  so that  $Na_3 = a_2$ ,  $Nb_3 = b_2$ ,  $Nc_2 = c_1$  and  $Nd_2 = d_1$ .

Now define

$$K_0 = \{v \in V \mid Nv = 0\}.$$

We know that the vectors  $a_1, b_1, c_1, d_1$  are in  $K_0$  and are independent, so we can complete this list to a basis of  $K_0$ . For definiteness, say we need one more vector  $e_1$  for a basis of  $K_0$  (nothing to do with the standard basis vector  $e_1$ ). We can summarize the situation in the following table:

$$(2.10) \quad \begin{array}{c|ccccc} V & a_3 & b_3 & c_2 & d_2 & e_1 \\ N(V) & a_2 & b_2 & c_1 & d_1 & \\ N^2(V) & a_1 & b_1 & & & \end{array}$$

We now claim that the vectors in (2.10) are independent. To see this, suppose that we have a linear relation

$$(2.11) \quad \alpha_1 a_1 + \beta_1 b_1 + \alpha_2 a_2 + \beta_2 b_2 + \gamma_1 c_1 + \delta_1 d_1 + \alpha_3 a_3 + \beta_3 b_3 + \gamma_2 c_2 + \delta_2 d_2 + \varepsilon_1 e_1 = 0$$

If we apply  $N$  to both sides of this relation, we get

$$(2.12) \quad \alpha_2 a_1 + \beta_2 b_1 + \alpha_3 a_2 + \beta_3 b_2 + \gamma_2 c_1 + \delta_2 d_1 = 0.$$

But, the vectors in table (2.7) were independent, so all of the coefficients in (2.12) must be zero. Plugging this into (2.11) gives

$$(2.13) \quad \alpha_1 a_1 + \beta_1 b_1 + \gamma_1 c_1 + \delta_1 d_1 + \varepsilon_1 e_1 = 0.$$

But the vectors in this relation were chosen to be a basis of  $K_0$ , so all of the coefficients in (2.13) must be zero. Thus, all of the coefficients in (2.11) are zero, so the vectors are independent.

Finally, we claim that the vectors in table (2.10) are a basis of  $V$ . The transformation  $N$  maps  $V$  onto  $N(V)$ , which has dimension 6. Thus,  $\text{rank}(N) = 6$ . The nullspace of  $N$  is  $K_0$ , which has dimension 5. Thus, the rank theorem for  $N$  says that  $V$  must have dimension 11, which is the number of vectors in table (2.10). Thus, the vectors in table (2.10) are a basis of  $V$ .

Now that we have constructed the basis of  $V$ , let  $A$  be the subspace of  $V$  spanned by  $a_1, a_2$  and  $a_3$ , with ordered basis  $\mathcal{A} = [a_1 \ a_2 \ a_3]$ , let  $B$  be the subspace spanned by  $b_1, b_2$  and  $b_3$  with ordered basis  $\mathcal{B} = [b_1 \ b_2 \ b_3]$ , let  $C$  be the subspace spanned by  $c_1$  and  $c_2$ , with ordered basis  $\mathcal{C} = [c_1 \ c_2]$ , and so forth. We then have the direct sum decomposition

$$V = A \oplus B \oplus C \oplus D \oplus E.$$

To find the matrix on  $N$  on  $A$ , call it  $B_1$ , we have to find the matrix  $B_1$  that fills in the equation  $N(\mathcal{A}) = \mathcal{A}B_1$ . But,

$$N(\mathcal{A}) = [N(a_1) \ N(a_2) \ N(a_3)] = [0 \ a_1 \ a_2]$$

and we have

$$[0 \ a_1 \ a_2] = [a_1 \ a_2 \ a_3] \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

from which we conclude that

$$B_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Similarly, the matrix  $B_2$  of  $N$  on  $B$  is

$$B_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix  $B_3$  of  $N$  on  $C$  is

$$B_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

and the matrix  $B_4$  of  $N$  on  $D$  is the same as  $B_3$ . Finally, the matrix of  $N$  on the one dimensional space  $E$  is simply  $[0]$ .

The matrix of  $N$  on all of  $V$  will then have block diagonal form with  $B_1, B_2, B_3, B_4$  and  $B_5$  down the diagonal. Thus, the matrix of  $N$  with respect to the basis

we have constructed is

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

with the blocks on the diagonal indicated by boxes.

We introduce the following terminology for these blocks. The  $n \times n$  **nilpotent block** is the  $n \times n$  matrix with ones on the superdiagonal and zeros elsewhere. For example, the  $1 \times 1$  nilpotent block is just  $[0]$ . The  $2 \times 2$  nilpotent block is

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

The  $4 \times 4$  nilpotent block is

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

Let us say that a nilpotent matrix is in **nilpotent canonical form** if it is block diagonal, where each of the blocks on the diagonal is a nilpotent block.

The algorithm above can be formulated as follows.

**Theorem 2.12.** *If  $V$  is a vector space over  $\mathbb{K}$  and  $N: V \rightarrow V$  is a nilpotent transformation, there is a basis of  $V$  so that the matrix of  $N$  with respect to this basis is in nilpotent canonical form. The nilpotent canonical form of a transformation is unique, up to the order of the nilpotent blocks along the diagonal.*

If  $A$  is a nilpotent matrix, we can consider the nilpotent transformation defined by multiplication by  $A$ . The nilpotent canonical form of this transformation is called the nilpotent canonical form of the matrix. The following is a fairly easy corollary of the theorem above.

**Corollary 2.13.** *Two nilpotent matrices are similar if and only if they have the same nilpotent canonical form.*

Finally, we can apply these results on nilpotent matrices to get the Jordan Canonical Form.

Let  $V$  be a complex vector space of dimension  $n$ , let  $T: V \rightarrow V$  be linear and let  $\lambda_1, \dots, \lambda_r$  be the distinct eigenvalues of  $T$ . Then,  $V$  is the direct sum of the generalized eigenspaces  $G_j$ , which are invariant under  $T$ . Thus,  $T = T_1 \oplus \dots \oplus T_r$ , where  $T_j: G_j \rightarrow G_j$  is the induced transformation

Since  $G_j$  is the nullspace of  $(T - \lambda_j I)^n$ , the transformation  $N_j: G_j \rightarrow G_j$  defined by  $N_j = T_j - \lambda_j I$  is nilpotent. Hence, we can find a basis of  $G_j$  that puts  $N_j$  into nilpotent canonical form. Thus, the matrix  $M_j$  of  $N_j$  with respect to this basis will be block diagonal with nilpotent blocks on the diagonal. But then the matrix with respect to this basis of  $T_j = \lambda_j I + N_j$  will be  $A_j = \lambda_j I + M_j$ . This matrix will have  $\lambda_j$ 's on the main diagonal. Thus, it will be block diagonal with blocks that are a nilpotent block with  $\lambda_j$ 's added down the diagonal.

A  $k \times k$  **Jordan Block** with eigenvalue  $\lambda$  is a  $k \times k$  matrix with  $\lambda$ 's on the diagonal and 1's on the superdiagonal. Thus, the  $1 \times 1$  block with eigenvalue  $\lambda$  is just  $[\lambda]$ , the  $2 \times 2$  block with eigenvalue  $\lambda$  is

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

the  $4 \times 4$  block is

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

and so forth.

We can find a basis of  $G_j$  in which the matrix of  $T_j$  is block diagonal with Jordan blocks with eigenvalue  $\lambda_j$  down the diagonal. If we choose such a basis for each  $G_j$  and put them together into a basis for  $V$ , the matrix of  $T$  will be a block diagonal matrix with Jordan blocks down the diagonal. This matrix is said to be the **Jordan Canonical Form** of  $T$ .

Thus, we say that a matrix is in Jordan Canonical Form if it is block diagonal with Jordan blocks down the diagonal.

We've now proved the following theorem.

**Theorem 2.14.** *If  $V$  is a complex vector space and  $T$  is a linear transformation, there is a basis of  $V$  so that the matrix of  $T$  with respect to this basis is in Jordan Canonical Form. The Jordan Canonical Form of  $T$  is unique up to the ordering of the Jordan blocks.*

*Two matrices are similar if and only if they have the same Jordan Canonical Form.*

As exercises, we'll give some matrices to work on, with machine assistance, of course! First, some computational observations in the form of exercises.

**Exercise 2.15.** Let  $V$  be a vector space over  $\mathbb{K}$  and let  $A$  be an  $n \times n$  matrix, which defines a linear transformation  $T: \mathbb{K}^n \rightarrow \mathbb{K}^n$  by  $T(v) = Av$ . Suppose that  $S$  is a subspace of  $\mathbb{K}^n$  with basis  $v_1, \dots, v_k$ . Define

$$K = \{v \in S \mid T(v) = 0\}.$$

How would you find a basis of  $K$ ? Hint: what is the matrix of the linear transformation  $S \rightarrow \mathbb{K}^n$  induced by  $T$ ?

**Exercise 2.16.** Here is a general theorem. Let  $V$  be a vector space over  $\mathbb{K}$  of dimension  $n$ . Suppose that  $v_1, \dots, v_n$  is a basis of  $V$  and that  $a_1, \dots, a_k$  is a linearly independent set. Then the linearly independent set can be completed to a basis by adding vectors from  $v_1, \dots, v_n$ . In other words, we can choose vectors  $v_{j_1}, \dots, v_{j_{n-k}}$  from among  $v_1, \dots, v_n$  so that  $a_1, \dots, a_k, v_{j_1}, \dots, v_{j_{n-k}}$  is a basis.

Describe an algorithm for finding the vectors  $v_{j_1}, \dots, v_{j_{n-k}}$  explicitly in the case  $V = \mathbb{K}^n$ .

**Exercise 2.17.** The following matrix  $N$  is nilpotent. Find a matrix  $P$  so that  $P^{-1}NP$  is in nilpotent canonical form.

$$N = \begin{bmatrix} 7 & -51 & -6 & 56 & 137 & -46 & 90 & 36 & 152 & -41 & 14 & 12 & 9 & -62 \\ 1 & -7 & -87 & 84 & 56 & -8 & -192 & 52 & 21 & -24 & -106 & 45 & 30 & -139 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & -27 & 18 & 0 & -99 & 9 & 0 & -9 & -54 & 0 & 0 & 0 \\ 14 & -110 & 222 & -301 & 271 & -139 & 208 & -20 & 312 & -68 & 27 & -117 & -60 & 278 \\ -4 & 32 & -66 & 97 & -79 & 39 & -45 & 6 & -90 & 20 & -1 & 36 & 18 & -87 \\ 0 & 0 & 0 & 3 & -2 & 0 & 11 & -1 & 0 & 1 & 6 & 0 & 0 & 0 \\ 19 & -149 & 282 & -416 & 397 & -195 & 127 & -7 & 424 & -106 & -51 & -147 & -75 & 340 \\ 0 & 0 & -30 & 27 & 12 & 0 & -69 & 16 & 0 & -6 & -36 & 15 & 10 & -45 \\ 47 & -369 & 726 & -1005 & 930 & -470 & 590 & -51 & 1048 & -238 & 30 & -381 & -195 & 897 \\ 0 & 0 & 1 & 6 & -4 & 0 & 22 & -2 & 0 & 2 & 12 & 0 & 0 & 0 \\ 78 & -618 & 1287 & -1719 & 1462 & -768 & 1372 & -153 & 1746 & -353 & 276 & -687 & -350 & 1653 \\ -19 & 149 & -318 & 452 & -385 & 195 & -196 & 25 & -424 & 100 & 15 & 165 & 87 & -394 \\ 26 & -206 & 419 & -592 & 508 & -256 & 343 & -38 & 582 & -128 & 30 & -225 & -114 & 539 \end{bmatrix}$$

**Exercise 2.18.** In each part, find a matrix  $P$  so that  $P^{-1}AP$  is in Jordan Canonical Form.

(1)

$$A = \begin{bmatrix} -4 & 4 & 2 \\ -10 & 9 & 3 \\ 6 & -5 & 1 \end{bmatrix}$$

(2)

$$A = \begin{bmatrix} 40 & -12 & 0 & 11 & -12 & -4 & -1 & 5 & -1 \\ 748 & -1273 & 120 & 205 & -126 & -415 & 351 & 55 & -27 \\ -2654 & 4924 & -463 & -677 & 474 & 1607 & -1341 & -204 & 178 \\ -200 & 291 & -26 & -52 & 42 & 95 & -74 & -18 & 10 \\ 409 & -418 & 33 & 110 & -100 & -137 & 90 & 42 & -19 \\ -3303 & 4643 & -416 & -903 & 658 & 1516 & -1200 & -277 & 122 \\ -137 & -966 & 114 & -53 & 123 & -313 & 347 & -44 & -21 \\ 300 & -1162 & 121 & 69 & -5 & -378 & 357 & 9 & -34 \\ 376 & -342 & 26 & 104 & -92 & -112 & 72 & 37 & -10 \end{bmatrix}$$