Canonical forms

Shrivathsa Pandelu

November 5, 2020

Contents

1 Introduction **1**

2 Characteristic values **2**

3 Annhilating polynomials **3**

4 Invariant subspaces **4**

5 Simultaneous Triangulation; Simultaneous Diagonalization **6**

6 Direct sums **7**
6.1 Direct sum decompositions 7
6.2 Invariant direct sums 8

7 The Primary decomposition theorem **10**

8 Cyclic subspaces and Annihilators **12**

9 Cyclic Decomposition and the Rational form **12**
9.1 Motivation 12
9.2 The real deal 14
9.3 Corollaries and applications 16

10 Jordan Form **18**

11 Computing invariant factors **19**

12 Summary; Semi-Simple operators **22**

13 Reference **25**

1 Introduction

The goal of this article is to give an explanation and derivation of the rational canonical form and Jordan form. Much of the material is borrowed from the book Linear Algebra by Kenneth Hoffman and Ray Kunze.

We shall first consider a finite dimensional vector space V over a field F and a linear operator T on V. Often we want to compute properties of T such as its rank or null space. For example, knowing the null space of differential equations (which can be phrased as linear operators) tell us how big the solution space is.

We study T by studying its parts. We try to "break" V into parts $W_1, \ldots, W_k$ and study what T does to each parts. Here we face our first hurdle. We better have $T(W_i) \subseteq W_i$ as otherwise it is hard to see what T does to W_i and we might have to chase its images and coalesce them into one big subspace. Note that we must be able to put all the W_i back together to V. We define

Definition 1. *A subspace W of V is said to be T invariant if $T(W) \subseteq W$.*

Our goal is to find out whether, and if so when, we can decompose V into invariant subspaces such that T is simple on each such subspace. The word simple here refers to those where computations are easy. Diagonal operators are very simple, scalar multiplication is simpler still. It is not clear that we should be able to do so, however in the finite dimensional case, things are particularly convenient.

So one of the first things we will be doing is to consider simple invariant subspaces, spaces spanned by one vector or spaces where T acts like scalar multiplication. Then through a series of theorems and lemmas, show that a finite dimensional space can indeed be broken into invariant subspaces and that once broken T must behave in certain nice ways.

2 Characteristic values

The simplest non trivial subspaces are those spanned by single vectors. We define

Definition 2. *A characteristic value of T is a scalar $c \in F$ for which there is a non zero vector $\alpha \in V$ such that $T\alpha = c\alpha$. α is called a characteristic vector associated with the characteristic value c and the set of all such α together with 0 is called the characteristic space associated with c.*

This definition applies to infinite dimensional vector spaces as well. When V is finite dimensional, we have the following

Theorem 1. *The following are equivalent*

i) c is a characteristic value of T.
ii) The operator $(T - cI)$ is singular.
iii) det(T-cI) = 0.

Treating matrices over F as operators we can define the characteristic values of matrices. For matrices too, we have condition iii), so the characteristic values of a matrix A over F are the roots of the polynomial $det(xI - A)$.

Definition 3. *The polynomial $det(xI - A)$ is called the characteristic polynomial of the matrix A.*

Observe that similar matrices have the same characteristic polynomials as

$$det(xI - P^{-1}AP) = det(P^{-1}(xI - A)P)$$

where P is an invertible matrix over F. Therefore, we may unambiguously define the characteristic polynomial of an operator T as that of its matrix with respect to some basis.

It is easy to see that the operator T is diagonal if and only if there is a basis consisting of eigenvectors. In such a case we say that T is diagonalizable. When T is diagonalizable, we may choose a basis of V such that T has the matrix form

$$\begin{bmatrix} c_1 I_1 & & & \\ & c_2 I_2 & & \\ & & \ddots & \\ & & & c_k I_k \end{bmatrix}$$

where the scalars c_i are distinct and I_i is a $d_i \times d_i$ identity matrix, $d_i \geq 1$. It follows that the characteristic polynomial of T is then $(x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$. So, if the charateristic polynomial doesn't split into linear factors, then T cannot be diagonal over F, but may be over some extension in which case the dimension of V may change. Observe that d_i is the algebraic multiplicity of the eigenvalue c_i.

Lemma 1. *Let $c_1, \ldots, c_k$ be distinct eigenvalues of T and W_i be the eigenspace of c_i. If $W = W_1 + W_2 + \cdots + W_k$, then*

$$dim\ W = dim\ W_1 + \cdots + dim\ W_k.$$

In fact, if $\mathcal{B}_i$ is a basis for W_i, then $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ is a basis for W.

Proof. All we need to show that eigenvectors of different eigenvalues are independent. Assume $\beta_1, \ldots, \beta_k$ are eigenvectors with eigenvalues $c_1, \ldots, c_k$ all different. Suppose we have scalars $a_1, \ldots, a_k$ such that

$$a_1\beta_1 + \cdots + a_k\beta_k = 0.$$

By a repeated application of T, we obtain an invertible Vandermonde matrix with the column vector $\begin{bmatrix} a_1 & \ldots & a_k \end{bmatrix}^t$ as a solution which would mean each $a_i = 0$. Alternatively, one could use lagrange polynomials to show that each $a_i = 0$. For example if f_1 is a polynomial such that $f_1(c_1) = 1$ and $f_1(c_j) = 0, j \geq 2$, then $f_1(a_1\beta_1 + \cdots + a_k\beta_k) = a_1\beta_1 = 0 \Rightarrow a_1 = 0$. □

Theorem 2. *Let $c_1, \ldots, c_k$ be the eigenvalues of T and W_i the null space of $(T - c_iI)$. The following are equivalent.*

i) T is diagonalizable.
ii) The characteristic polynomial for T is

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$$

and $dim\ W_i = d_i, i = 1, \ldots, k$.
iii) $dim\ W_1 + \cdots + dim\ W_k = dim\ V$.

Proof. Clearly (i) implies (ii). Since the degree of the characteristic polynomial is $n = \dim V$, (ii) implies (iii). By the previous lemma, if (iii) holds, we must have $W = W_1 + \cdots + W_k = V$, therefore (iii) implies (i). □

3 Annhilating polynomials

Given the operator T, the polynomials in T provide us some useful informations about T. For example, the null space of $T - cI$ is the eigenspace of c.

Given a polynomial $f \in F[x]$, we can define the operator $f(T)$ on V. We have $(f + g)(T) = f(T) + g(T), (fg)(T) = f(g(T))$. Note that products become compositions as that is the operation in $L(V, V)$. This gives us an F algebra homomorphism from $F[x]$ to $L(V, V)$ mapping x to T. The kernel of this homomorphism is generated by a (monic) polynomial since $F[x]$ is a PID.

When V of dimensional n, $L(V, V)$ is of dimension n^2, so the operators $I, T, \ldots, T^{n^2}$ are not linearly independent hence the kernel is not zero. The same applies to matrices.

From an algebro-geometric perspective, we are looking at which affine varieties if T a part of in the affine plane $\mathbb{A}^{n^2}$. Usually we need to consider polynomials in n^2 variables, however here we have an action of $F[x]$ on the space of linear operators that in turn gives us n^2 linear equations (the polynomials indicating the entries of $f(T)$) and we are looking at the zero sets of the ideals spanned by such linear equations.

Definition 4. *The minimal polynomial for T is the unique monic generator of the ideal of polynomials over F that annihilate T.*

Let A be a matrix over F and F' a field containing F, then A is a matrix over F' as well. Let m, m' be the minimal polynomial of A over F, F' respectively, then clearly $m'|m$.

If $m' = a_0 + a_1x + \cdots + a_{t-1}x^{t-1} + x^t$, then the coefficients $a_0, \ldots, a_{t-1}$ satisfy some linear equations over F (for A is a matrix over F), hence those same equations have a solution over F which means that there is a polynomial over F of degree t annihilating A. This means that the degree of m' cannot be strictly less than that of m which means that $m' = m$. Therefore the minimal polynomial depends only on A and not on the base field.

Theorem 3. *The characteristic polynomial and minimal polynomial of T have the same roots.*

Proof. Let p be the minimal polynomial for T, f the characteristic polynomial and c a scalar.

If $p(c) = 0$, then $p = (x - c)q$ where q is a polynomial of smaller degree than p. By definition of p, we must have a vector β such that $q(T)\beta \neq 0$, so $p(T)\beta = (T - cI)q(T)\beta = 0$ Thus, $q(T)\beta$ is an eigenvector with eigenvalue c. Thus c is a root of f.

If c is an eigenvalue, then there is a non zero $\alpha \in V$ such that $T\alpha = c\alpha \Rightarrow p(T)\alpha = p(c)\alpha = 0 \Rightarrow p(c) = 0$ since $\alpha \neq 0$. Thus, p, f have the same roots. □

Theorem 4. ***(Caley-Hamilton)*** *Let T be a linear operator on a finite dimensional vector space. If f is the characteristic polynomial for T, then $f(T) = 0$; in other words, the minimal polynomial divides the characteristic polynomial for T.*

Proof. Let K be he ring of polynomials in T, $\{\alpha_1, \ldots, \alpha_n\}$ a basis for V and A the matrix of T with respect to this basis. Then

$$T\alpha_i = \sum_{j=1}^{j=n} A_{ji}\alpha_i$$

which can be written as

$$\sum_{j=1}^{j=n} (\delta_{ij}T - A_{ji}I)\alpha_j = 0, 1 \leq i \leq n.$$

Now comes the really clever part, we set $B_{ij} = \delta_{ij}T - A_{ji}I$ to create a "matrix" B over K treating B_{ij} as its entries. Observe that B is the matrix $xI - A$ "evaluated" at T. Also note that $\det B = f(T)$. Now, by the way we defined B, we have $B[\alpha_1 \ldots \alpha_n]^t = 0$.

The last equation is an abuse of notation as the "multiplication" is not really defined, it is a shorthand for writing long summations. However, we may think of it as the space of $n \times n$ matrices over K acting on V^n by the usual rules of matrix multiplication.

Continuing with the notational abuse, we may write

$$(adjB)B[\alpha_1 \ldots \alpha_n]^t = 0 \Rightarrow (\det B)I[\alpha_1 \ldots \alpha_n]^t = 0 \Rightarrow f(T)\alpha_i = 0, 1 \leq i \leq n.$$

Therefore $f(T)$ annihilates V and the minimal polynomial divides the characteristic polynomial. □

4 Invariant subspaces

If V is a vector space over F and T a linear operator, a subspace W is said to be invariant under T is $T(W) \subseteq W$. We denote the restriction of T to W by T_W Extending a basis of the invariant subspace W to one of V, we obtain a basis for V with respect to which T has a matrix of the form

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

where B is the matrix of T_W in the basis of W that we started with. We then have

Lemma 2. *The characteristic and minimal polynomials of T_W divide those of T.*

Proof. The proof is quite simple. □

A particularly nice invariant subspace is the one that is spanned by all eigenvectors. If $c_1, \ldots, c_k$ are the eigenvalues of T and W_i is the eigenspace corresponding to the scalar c_i, then $W = W_1 + \cdots + W_k$ is invariant with $\dim W = \dim W_1 + \cdots + \dim W_k$. The characteristic polynomial of T_W is $(x - c_1)^{e_1} \ldots (x - c_k)^{e_k}$ where $e_i = \dim W_i, 1 \leq i \leq k$. All of this makes Theorem 2 more transparent. Our strategy is to get large invariant subspaces and see how T behaves in them. In the previous sections we discussed characteristic vectors and characteristic spaces. These were invariant, spanned by one vector or a few vectors with the same eigenvalue. We would like to "grow" small invariant subspaces into larger ones.

Definition 5. *Let W be an invariant subspace for T and let α be a vector in V. The T-conductor of α into W is the set $S_T(\alpha; W)$, which consists of all polynomials g over F such that $g(T)\alpha \in W$.*

An important consequence of W being invariant is the following.

Lemma 3. *If W is invariant under T, then W is invariant under every polynomial in T. Thus, for each $\alpha, S(\alpha; W)$ is an ideal of the polynomial algebra $F[x]$.*

Once we know that the conductor is an ideal, we can talk about its monic generator. The unique monic generator of $S(\alpha; W)$ for an invariant subspace W is also called the T-conductor of α into W. Note that the T-conductor of α into W is the annihilator of α in V/W. We need W to be invariant so that T can induce a linear operator on V/W. Also not that every T-conductor divides the minimal polynomial for T.

Lemma 4. *Let V be a finite dimensional vector space over the field F. Let T be a linear operator on V such that the minimal polynomial for T is a product of linear factors*

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}, c_i \in F.$$

Let W be a proper invariant subspace. There exists a vector $\alpha \in V$ such that

(a) $\alpha \notin W$;
(b) $(T - cI)\alpha \in W$ for some characteristic value c of T.

Proof. Let $\beta \notin W$, then the T-conductor of β into W is of the form $f = (x - c_1)^{e_1} \dots (x - c_k)^{e_k}$ since it divides p. Since β is not in W, f is not a constant, so some $e_j > 0$ and $f = (x - c_j)g$. By definition of the conductor, $\alpha = g(T)\beta \notin W$. Also, $f(T)\beta = (T - c_j I)\alpha \in W$. Thus, α is the required vector. □

Theorem 5. *Let V be a finite dimensional vector space over the field F and T a linear operator on V. Then T is triangulable if and only if the minimal polynomial for T is a product of linear factors.*

Proof. Suppose the minimal polynomial is a product of linear factors, say

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}, c_i \in F.$$

Start with $W_0 = \{0\}$, by the lemma above there is a vector α_1 such that $T\alpha = c_{11}\alpha$ for some eigenvalue c_{11}. Take W_1 to be the space spanned by α_1, then there is a vector α_2 such that $T\alpha_2 = c_{12}\alpha_1 + c_{22}\alpha_2$. Take W_2 to be the space spanned by α_1, α_2. Repeat this procedure. Since $\alpha_i \notin W_i$, at each stage we have a linearly independent set of vectors $\{\alpha_1, \dots \alpha_i\}$. Further, $T\alpha_i$ is a linear combination of $\alpha_1, \dots, \alpha_i$. So, we obtain a basis (once the subspaces W_i become V, the lemma is not applicable) with respect to which T is upper triangular.

Conversely, if T is triangular, then it is clear that the characteristic, hence minimal (by Cayley-Hamilton theorem), polynomial is a factor of linear polynomials over F. □

As a corollary, every operator on a vector space over algebraically closed fields is triangulable.

Theorem 6. *Let V be a finite dimensional vector space over the field F and let T be a linear operator on V. Then T is diagonalizable if and only if the minimal polynomial for T has the form*

$$p = (x - c_1) \dots (x - c_k)$$

where $c_1, \dots, c_k$ are distinct elements of F.

Proof. If T is diagonalizable, we have a basis $\mathcal{B}$ consisting of eigenvectors and $p(T)$ annihilates each basis element and therefore p must be the minimal polynomial as the characteristic polynomial and minimal polynomials have the same roots.

Conversely, suppose the minimal polynomial is as given. Let W be the subspace spanned by all eigenvectors and that $W \neq V$ for if $W = V$, then T is diagonalizable. If W_i is the eigenspace of c_i,

then $W = W_1 + \cdots + W_k$. By Lemma 4, there is a vector $\alpha \notin W$ such that $\beta = (T - c_j I)\alpha \in W$. Now $p = (x - c_j)q$ for some polynomial q and

$$0 = p(T)\alpha = (T - c_j I)q(T)\alpha \Rightarrow Tq(T)\alpha = c_j q(T)\alpha$$

so $q(T)\alpha$ is an eigenvector and is in W. Now $q(T)\alpha \in W, (T - c_j I)\alpha \in W$ and $\alpha \notin W$. Therefore the T-conductor of α into W is not a constant and hence must be the polynomial $x - c_j$. Also, the T-conductor must divide $q(x)$ which means that $q(c_j) = 0$ contradicting the fact that p does not have any repeated roots. Thus $W = V$ and T is diagonalizable. □

As an aside, observe that a triangulable operator on a finite dimensional vector space satisfies its characteristic polynomial and that every operator over an algebraically closed field is triangulable. Let A be an $n \times n$ matrix over a field F and K be the algebraic closure of F, then we can act A on K^n and A is similar to a triangulable matrix. Thus, A satisfies its characteristic polynomial over K which is the same as that over F. This is an independent proof of the Cayley-Hamilton theorem.

5 Simultaneous Triangulation; Simultaneous Diagonalization

When can we say that the sums and differences of operators are triangulable or diagonalizable? More specifically, given a family $\mathcal{F}$ of linear operators on a vector space V, can we simultaneously triangulate or diagonalize all operators in $\mathcal{F}$, i.e., find one basis $\mathcal{B}$ which does the job for every operator in $\mathcal{F}$?

If we are to simultaneously diagonalize operators, then those operators should commute for diagonal operators do commute. This is no restriction for simultaneous triangulation, however we shall see that requiring them to commute is a sufficient condition for simultaneous triangulation. The subspace W is said to be invariant under $\mathcal{F}$ if it is invariant under each operator in $\mathcal{F}$. To find a basis that triangulates each operator, we need to have an extension of the previous theorems to a family of operators.

Lemma 5. *Let $\mathcal{F}$ be a commuting family of triangulable linear operators on V. Let W be a proper subspace of V which is invariant under $\mathcal{F}$. There exists a vector α in V such that*

(a) α is not in W;
(b) for each T in $\mathcal{F}$, the vector $T\alpha$ is in the subspace spanned by α and W.

Proof. We can assume that $\mathcal{F}$ is finite by replacing it with a basis for the linear span of $\mathcal{F}$. This is possible because the space of all linear operators on V is finite dimensional and an α satisfying (a), (b) for the basis, satisfies (a), (b) for every linear combination of the basis elements. So, take $\mathcal{F} = \{T_1, \ldots, T_r\}$ where $T_1, \ldots, T_r$ are linearly independent.

Since T_1 is triangulable, we have a vector β_1 not in W such that $(T_1 - cI)\beta_1 \in W$ for some c by the theorem for a single operator. Let W_1 be the set of all vectors β such that $(T_1 - cI)\beta \in W$. Observe that W_1 is a vector space strictly larger than W. Moreover, since the operators $T_1, \ldots, T_r$ commute and W is invariant under $\mathcal{F}$, W_1 is also invariant under $\mathcal{F}$.

Restrict $T_2, \ldots, T_r$ to W_1 and now we have fewer operators to worry about because by definition of W_1, for each $\beta \in W_1, (T_1 - cI)\beta \in W$ so any α that works for $T_2, \ldots, T_r$ in W_1 works for $T_1, \ldots, T_r$.

Now, we repeat this process. At each stage, we obtain a subspace $W \subsetneq W_{i+1} \subseteq W_i$. Since there are finitely many operators, this process terminates and we arrive at an α satisfying (a), (b). At each stage we need T_i to be triangulable for the process to continue. □

Theorem 7. *Let V be a finite dimensional vector space over the field F. Let $\mathcal{F}$ be a commuting family of triangulable linear operators on V. There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented by a triangular matrix in that basis.*

Proof. The proof is similar to the one given for a single linear operator. □

Corollary 1. *Let $\mathcal{F}$ be a commuting family of $n \times n$ matrices over an algebraically closed field F. There exists a non-singular matrix P with entries in F such that $P^{-1}AP$ is upper-triangular for every matrix A in $\mathcal{F}$*

Theorem 8. *Let $\mathcal{F}$ be a commuting family of diagonalizable linear operators on the finite dimensional vector space V. There exists an ordered basis for V such that every operator in $\mathcal{F}$ is represented in that basis by a diagonal matrix.*

Proof. In the case of a single linear operator, we took the space spanned by all eigenvalues and showed that it was all of V. But this cannot be applied here, as there are different sets of eigenvalues. We proceed by induction.

If $\dim V = 1$, there is nothing to prove. Assume that the theorem holds for all vector spaces of dimension less than n, and let V be n dimensional. If all T in $\mathcal{F}$ are scalar multiples of I, we are done. Else, choose T which is not a scalar multiple of I and let $c_1, \ldots, c_k$ be its distinct eigenvalues, and set W_i to be the null space of $T - c_i I, 1 \leq i \leq k$.

For a fixed i, W_i is invariant under every operator commuting with T. Let $\mathcal{F}_i$ be the family of operators obtained from $\mathcal{F}$ by restriction to W_i. Since each operator was diagonalizable, their restrictions are diagonalizable and since T is not a multiple of I, $\dim W_i < \dim V$. By induction hypothesis, ther is a basis $\mathcal{B}_i$ for W_i that simultaneously diagonalizes each operator in $\mathcal{F}_i$.

Set $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. This is a basis for V because T is diagonalizable, further in this basis, it is clear that all operators in $\mathcal{F}$ are diagonal. □

6 Direct sums

6.1 Direct sum decompositions

In this section we talk about how to "glue" back our subspaces into V.

Definition 6. *Let $W_1, \ldots, W_k$ be subspaces of the vector space V. We say that $W_1, \ldots, W_k$ are independent if*

$$\alpha_1 + \cdots + \alpha_k = 0, \alpha_i \in W_i$$

implies that each α_i is 0.

Lemma 6. *Let V be a finite dimensional vector space. Let $W_1, \ldots, W_k$ be subspaces of V and let $W = W_1 + \cdots + W_k$. The following are equivalent.*

(a) $W_1, \ldots, W_k$ are independent.
(b) For each $j, 2 \leq j \leq k$, we have

$$W_j \cap (W_1 + \cdots + W_{j-1}) = \{0\}.$$

(c) If $\mathcal{B}_i$ is an ordered basis for $W_i, 1 \leq i \leq k$, then the sequence $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ is an ordered basis for W.

If any of the conditions of the lemma above hold, we say that $W = W_1 + \cdots + W_k$ is the direct sum of $W_1, \ldots, W_k$ and we write $W = W_1 \oplus \cdots \oplus W_k$.

We wanted to first isolate the effects of a linear operator T on V and that required us to study invariant subspaces. Having found invariant subspaces, we would then like to take T back V and this requires us to glue back its isolated effects. If the invariant subspaces are not independent, then that means that they have non trivial intersection and this makes it hard to choose consistent bases or to study the effects of T on V from its effects on the subspaces. Having independence gives a nice block structure to T.

Consider for example the transformation on $\mathbb{R}^3$ given by the matrix

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

and take W_1 to be the $x-y$ plane and W_2 to be the $y-z$ plane. Both are invariant and have y axis as their intersection. Because of dependence, T doesn't have a nice block structure, although its effects on W_1, W_2 are particularly simple, one is identity and the other is projection.

Moreover, calculations involving dimensions of nullspaces, eigenspaces can be done separately in each invariant subspaces and then added together because the independence condition prevents certain subspaces (the intersections) from being multiply counted.

Definition 7. *If V is a vector space, a projection of V is an operator E on V satisfying $E^2 = E$.*

Suppose E is a projection. Let R be its range and N its null space.

1. The vector β is in the range if and only if $E\beta = \beta$.
2. By 1, $R \cap N = \{0\}$ and given $\alpha \in V, \alpha - E\alpha \in N$, so $\alpha = E\alpha + (\alpha - E\alpha)$. Therefore $V = R \oplus N$.

Also note that if we can write $V = R \oplus N$ then there is a unique projection which has range R and null space N. This projection is called the projection on R along N.

Theorem 9. *If $V = W_1 \oplus \cdots \oplus W_k$, then there exists k linear operators $E_1, \ldots, E_k$ on V such that*

(i) each E_i is a projection ($E_i^2 = E_i$);
(ii) $E_i E_j = 0$, if $i \neq j$;
(iii) $I = E_1 + \cdots + E_k$;
(iv) the range of E_i is W_i.

Conversely, if $E_1, \ldots, E_k$ are k linear operators on V which satisfy conditions (i), (ii), and (iii), and if we let W_i be the range of E_i, then $V = W_1 \oplus \cdots \oplus W_k$.

Proof. Given $V = W_1 \oplus \cdots \oplus W_k$, we let E_i be defined by $E_i\alpha = \alpha_i$ where $\alpha = \alpha_1 + \cdots + \alpha_k$ with $\alpha_i \in W_i$. By independence, E_i is well defined. It is clear that E_i is a linear operator. Further, the range of E_i is W_i and its null space is $W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_k$. Also, $E_i^2 = E_i$ and $E_i E_j = 0$ for $i \neq j$. Further any $\alpha \in V$ is given by $\alpha = \alpha_1 + \cdots + \alpha_k = E_1\alpha + \cdots + E_k\alpha$. Therefore, $E_1, \ldots, E_k$ satisfy the given conditions.

Conversely, suppose such $E_1, \ldots, E_k$ exist. We let W_i be the range of $E_i, 1 \leq i \leq k$. Then by $I = E_1 + \cdots + E_k$, we have $V = W_1 + \cdots + W_k$. Further, given $\alpha = \alpha_1 + \cdots + \alpha_k$ with $\alpha_i \in W_i$, we have $E_i\alpha = \alpha_i$ for $1 \leq i \leq k$, which means that $\alpha = 0$ if and only if each $\alpha_i = 0$. Thus $W_1, \ldots, W_k$ are independent. Hence, $V = W_1 \oplus \cdots \oplus W_k$. □

6.2 Invariant direct sums

Theorem 10. *Let T be a linear operator on $V = W_1 \oplus \cdots \oplus W_k$ and $E_1, \ldots, E_k$ be the associated projections. Then each W_i is invariant under T if and only if $TE_i = E_iT$ for every i.*

Proof. Suppose T commutes with E_i, then for $\alpha_i \in W_i$, we have $E_iT\alpha_i = TE_i\alpha_i = T\alpha_i \Rightarrow T\alpha_i \in W_i$ and hence, W_i is invariant.

Conversely, if each W_i is T invariant, then given $\alpha = \alpha_1 + \cdots + \alpha_k \in V$ with $\alpha_i \in W_i, 1 \leq i \leq k$, $T\alpha = T\alpha_1 + \cdots + T\alpha_k$. Because each W_i is invariant, $T\alpha_i \in W_i, 1 \leq i \leq k$, so

$$TE_i\alpha = T\alpha_i = E_iT\alpha \Rightarrow TE_i = E_iT, 1 \leq i \leq k$$

as required. □

Remark. The theorem above doesn't need V to be finite dimensional.

Theorem 11. *Let T be a linear operator on a finite dimensional vector space V. If T is diagonalizable and if $c_1, \ldots, c_k$ are the distinct eigenvalues of T, then there exist linear operators $E_1, \ldots, E_k$ on V such that*

(i) $T = c_1E_1 + \cdots + c_kE_k$;
(ii) $I = E_1 + \cdots + E_k$;

(iii) $E_iE_j = 0, i \neq j$;
(iv) $E_i^2 = E_i$;
(v) the range of E_i *is the eigenspace for* T *associated with* c_i.

Conversely, if there exist k *distinct scalars* $c_1, \ldots, c_k$ *and* k *non-zero linear operators* $E_1, \ldots, E_k$ *which satisfy conditions (i), (ii), and (iii), then* T *is diagonalizable,* $c_1, \ldots, c_k$ *are the distinct eigenvalues of* Tm *and conditions (iv) and (v) are also satisfied.*

Proof. Suppose T is diagonalizable with eigenvalues $c_1, \ldots, c_k$. Let W_i be the eigenspace corresponding to c_i, then $V = W_1 \oplus \cdots \oplus W_k$. Let $E_1, \ldots, E_k$ be the associated projections, then (ii), (iii), (iv), and (v) are satisfied.

To verify (i), take $\alpha = E_1\alpha + \cdots + E_k\alpha \in V$, then

$$T\alpha = c_1E_1\alpha + \cdots + c_kE_k\alpha$$

as $E_i\alpha \in W_i$. This is true for every vector $\alpha \in V$, therefore $T = c_1E_1 + \cdots + c_kE_k$.

Conversely, suppose we are given operators $E_1, \ldots, E_k$ on V and scalars $c_1, \ldots, c_k$ satisfying (i), (ii), and (iii). Multiplying (ii) by E_i, we get $E_i^2 = E_i, 1 \leq i \leq k$. Thus, (iv) is satisfied.

Multiplying (i) by E_i from the right, $TE_i = c_iE_i$, so c_i is a characteristic value and the range of E_i is contained in the corresponding eigenspace. Now, if c is an eigenvalue, then

$$T - cI = (c_1 - c)E_1 + \cdots + (c_k - c)E_k.$$

If α is an eigenvector for c, then $E_i\alpha \neq 0$ for some i and we have $E_i(T-cI)\alpha = (c_i-c)E_i\alpha = 0 \Rightarrow c_i = c$. Thus, $c_1, \ldots, c_k$ are the eigenvalues for T. Now, since the range of E_i is contained in the eigenspace corresponding to c_i and $I = E_1 + \cdots + E_k$, we have a basis of V consisting of eigenvectors and hence, T is diagonalizable.

All that is left is to show that the eigenspace corresponding to c_i is exactly the range of E_i, but this is clear because if $(T - c_iI)\alpha = (c_1 - c_i)E_1\alpha + \cdots + (c_k - c_i)E_k\alpha = 0$, then by multiplying by E_j for $j \neq i$, we have $(c_j - c_i)E_j\alpha = 0$ and since $c_j \neq c_i$, we must have $E_j\alpha = 0$ for $j \neq i$. Thus $\alpha = E_i\alpha$ is in the range of E_i proving the theorem. □

One of the more interesting features of the decomposition given in the theorem above is that any polynomial $p(T)$ of T is easily given as $p(c_1)E_1 + \cdots + p(c_k)E_k$. In particular, taking one of the Lagrange polynomials that is 1 at some c_i and 0 at all others, we have $p(T) = E_i$, i.e., each E_i is a polynomial in T. Now we give an independent proof of Theorem 6.

Observe that when T is diagonalizable, the polynomial $p(x) = (x - c_1)\ldots(x - c_k)$ annihilates T and since has all the roots of the characteristic polynomial of T must be the minimal polynomial. Conversely, if $p(x) = (x - c_1)\ldots(x - c_k)$ is the minimal polynomial of T, then taking p_i to be the Lagrange polynomial given by

$$p_i(x) = \prod_{j \neq i} \frac{(x - c_j)}{(c_i - c_j)}$$

we have for any polynomial g of degree $\leq k - 1$,

$$g(x) = g(c_1)p_1 + \cdots + g(c_k)p_k.$$

If $k = 1$, then T is a scalar multiple of identity and is diagonalizable, so assume $k \geq 2$, then taking $g = 1, x$, we have

$$1 = p_1 + \cdots + p_k$$

$$x = c_1p_1 + \cdots + c_kp_k.$$

Observe that for $i \neq j, p_ip_j = 0$. Taking $E_i = p_i(T)$, the operators $E_1, \ldots, E_k$ and the scalars $c_1, \ldots, c_k$ satisfy conditions (i), (ii), and (iii) of the theorem above. Note that $E_i \neq 0$ since p is the minimal polynomial for T and each p_i has smaller degree. Thus, T is diagonalizable.

7 The Primary decomposition theorem

An eigenspace of an operator T is the null space of a polynomial in T of the form $x - c$. When the minimal polynomial factorises into linear parts, we can talk of eigenspaces and triagularizing/diagonalising T. We now consider general factorisations of the minimal polynomial. Observe that if p is a polynomial, the null space of $p(T)$ is T-invariant. The decomposition below is called primary decomposition because it comes from decomposing the minimal polynomial into its primary components which are powers of prime polynomials.

Theorem 12. ***(Primary Decomposition Theorem)*** *Let T be a linear operator on the finite dimensional vector space V over the field F. Let p be the minimal polynomial for T,*

$$p = p_1^{r_1} \dots p_k^{r_k}$$

where the p_i are distinct irreducible monic polynomials over F and the r_i are positive integers. Let W_i be the null space of $p_i(T)^{r_i}, i = 1, \dots, k$. Then

(i) $V = W_1 \oplus \dots \oplus W_k$;
(ii) each W_i is invariant under T;
(iii) if T_i is the operator induced on W_i by T, then the minimal polynomial for T_i is $p_i^{r_i}$.

Proof. Firstly, observe that said nullspaces are not empty because $p/p_i^{r_i}$ is not the minimal polynomial, hence is not zero at some $\alpha \in V$, but $p(\alpha) = 0$.

For $i \neq j$, if $\alpha \in W_i \cap W_j$, then $p_i^{r_i}(T)\alpha = p_j^{r_j}(T)\alpha = 0$. Since p_i, p_j are two different irreducibles, their gcd is 1, so there are polynomials g_i, g_j such that

$$g_i p_i^{r_i} + g_j p_j^{r_j} = 1 \Rightarrow (g_i(T)p_i(T)^{r_i} + g_j(T)p_j(T)^{r_j})\alpha = \alpha,$$

but by assumption $\alpha \in W_i \cap W_j$, so $\alpha = 0$. Thus, $W_1, \dots, W_k$ are pairwise independent.

Now, if $\alpha_1 + \dots + \alpha_k = 0$ with $\alpha_i \in W_i$, then evaluating $p_1(T)^{r_1}$, we have $\beta_2 + \dots + \beta_k = 0$ where $\beta_j = p_1(T)^{r_1}\alpha_j \in W_j, j \geq 2$. Now, we proceed by induction on k, with two subspaces, we have already proved that they are independent. Using this manipulation, we drop k by 1 and so each $\beta_j = 0, j \geq 2$ and by the $k = 2$ case we must have each $\alpha_j = 0, j \geq 2$. This gives $\alpha_1 = 0$ as well. Therefore the subspaces we obtain are really independent.

Now we need to show that these pieces fit together properly. To do this, one might conisder trying to show that the ranks add up to the right number, but there is no neat direct way to do this. So, we consider the use of projections. Now, let

$$f_i = \frac{p}{p_i^{r_i}}.$$

Since $p(T)$ annihilates V, the range of $f_i(T)$ is contained in W_i. Further for $i \neq j, f_i f_j(T) = 0$. Thus, f_i projects V, in some way, onto W_i, but we want something more. We must have the projections add up to 1. To do this, we observe that $f_1, \dots, f_k$ are coprime, so there are polynomials $g_1, \dots, g_k$ such that $g_1 f_1 + \dots + g_k f_k = 1$. The range of $g_i f_i(T)$ is still contained in W_i, and we still have $g_i f_i(T).g_j f_j(T) = 0, i \neq j$. Using the last equation, we have $(g_i f_i(T))^2 = g_i f_i(T)$. So, we set

$$E_i = g_i f_i(T).$$

Then,

$$\begin{aligned} &E_1 + \dots + E_k = I \\ &E_i^2 = E_i, 1 \leq i \leq k \\ &E_i E_j = 0, 1 \leq i \neq j \leq k. \end{aligned}$$

This shows that $W_1, \dots, W_k$ are independent. All that is left to be shown is that the range of E_i is exactly W_i. Let $\alpha \in W_i$, since $p_i^{r_i} \mid f_j, j \neq i$, we have $E_j\alpha = 0, j \neq i$. Because $E_1 + \dots + E_k = I$, we have $E_i\alpha = \alpha$, thus the range of E_i is precisely W_i.

Finally, we look at the restriction T_i of T to W_i, it satisfies $p_i^{r_i}$, thus the minimal polynomial of T_i divides $p_i^{r_i}$. Let g be any polynomial such that $g(T_i) = 0$, then $g(T)f_i(T) = 0$ for any α not in W_i is annihilated by $f_i(T)$ and those in W_i are annihilated by $g(T)$. Therefore $p \mid gf_i \Rightarrow p_i^{r_i} \mid g$. Therefore the minimal polynomial for T_i is $p_i^{r_i}$. This completes the proof. □

Observe that we don't really need V to be finite dimensional for the theorem to hold nor do we need, for (i), (ii), that p be the minimal polynomial. It suffices that T satisfy some polynomial for (i), (ii) to hold.

Corollary 2. *If $E_1, \ldots, E_k$ are the projections associated with the primary decomposition of T, then each E_i is a polynomial in T, and accordingly if a linear operator U commutes with T then U commutes with each of E_i, i.e., each subspace W_i is invariant under U.*

So, we have arrived at an interesting result. We can decompose V into invariant subspaces on which T acts in a much simpler way, or at least we hope so, and we know how to put these subspaces together to get V. Well done!

Now, we have a look at a particular consequence of the primary decomposition theorem, the case where the minimal polynomial factors into linear polynomials. In this case we had shown that T is triangulable, next we look at its primary decomposition.

Let $p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}$ be the minimal polynomial. Let $E_1, \ldots, E_k$ be the projections associated with the primary decomposition for T. We set $D = c_1E_1 + \cdots + c_kE_k$, then by Theorem 9, D is diagonalizable. We set $N = T - D$.

$$T = TE_1 + \cdots + TE_k$$

$$D = c_1E_1 + \cdots + c_kE_k$$

so

$$N = (T - c_1I)E_1 + \cdots + (T - c_kI)E_k$$

and $N^r = 0$ for $r \geq r_1, \ldots, r_k$ because the range of E_i is in the nullspace of $(T - c_iI)^{r_i}$. Therefore N is nilpotent.

Theorem 13. *Let T be a linear operator on the finite dimensional vector space V over the field F. Suppose that the minimal polynomial for T decomposes over F into a product of linear polynomials. Then there is a diagonal operator D on V and a nilpotent operator N on V such that*

(i) $T = D + N$,
(ii) $DN = ND$.

The diagonal operator D and nilpotent operator N are uniquely determined by (i) and (ii) and each of them is a polynomial in T.

Proof. By the primary decomposition theorem, we have seen that we can obtain D, N which are polynomials in T satisfying (i) and (ii). Now suppose we also have $T = D' + N'$ where D' is diagonalizable and N' is nilpotent and $D'N' = N'D'$.

Since D', N' commute with each other and $T = D' + N'$, they must both commute with T and hence with any polynomial in T. In particular, D', N' commute with both D, N. So, we have

$$D - D' = N' - N.$$

Since N, N' are both nilpotent, their difference $N' - N$ is also nilpotent. Now, since D, D' commute and are both diagonalizable, they are simulatneously diagonalizable, and hence $D - D'$ is diagonalizable and nilpotent and therefore must be 0. Therefore $D = D', N = N'$. □

So, if we are working over an algebraically closed field, then every operator can be decomposed uniquely as in the theorem above and the study of operators reduces to that of nilpotent operators for we have a fair grip over diagonalizable operators.

8 Cyclic subspaces and Annihilators

We want to find invariant subspaces. One way was to look at the null spaces of certain polynomials. However, there is another way where we grow invariant subspaces - a bottom up approach in some sense. We start with a non zero vector $\alpha \in V$ and chase all its images. Eventually we end up with an invariant subspace. This subspace can also be obtained as the intersection of all invariant subspaces containing α.

Definition 8. *If α is any vector in V, the T cyclic subspace generated by α is the subspace $Z(\alpha;T)$ of all vectors of the form $g(T)\alpha, g \in F[x]$. If $Z(\alpha;T) = V$, then α is called a cyclic vector for T.*

An obvious basis for $Z(\alpha;T)$ is the set $\{\alpha, T\alpha, \dots\}$.

Definition 9. *If α is any vector in V, the T annihilator of α is the ideal $M(\alpha;T)$ in $F[x]$ consisting of all polynomials g over F such that $g(T)\alpha = 0$. The unique monic generator p_α of this ideal is also called the T annihilator of α.*

Theorem 14. *Let α be any non-zero vector in V and let p_α be the T annihilator of α.*

(i) The degree of p_α is equal to the dimesion of the cyclic subspace $Z(\alpha;T)$.
(ii) If the degree of p_α is k, then the vectors $\alpha, T\alpha, \dots, T^{k-1}\alpha$ form a basis for $Z(\alpha;T)$.
(iii) If U is the linear operator on $Z(\alpha;T)$ induced by T, then the minimal polynomial for U is p_α.

Let us first look at what happens to T when it has a cyclic vector. Let T operate on V with cyclic vector α, then $\alpha, T\alpha, \dots, T^{n-1}\alpha$ is a basis for V. Let $p = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} + x^n$ be the characteristic polynomial for T. Then with respect to the basis given by α, T has the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & 0 & \dots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -c_{n-1} \end{bmatrix}.$$

This matrix is called the companion matrix of p. Because α is a cyclic vector, the minimal polynomial for T must be the characteristic polynomial.

Theorem 15. *If U is a linear operator on the finite dimensional space W, then U has a cyclic vector if and only if there is some ordered basis for W in which U is represented by the companion matrix of the minimal polynomial for U.*

Proof. We have seen above that if U has a cyclic vector then there is a basis of W such that U is represented by the companion matrix of its minimal polynomial in that basis. Conversely if there is such a basis $\{\alpha_1, \dots, \alpha_n\}$, then clearly α_1 is a cyclic vector for U. □

Corollary 3. *If A is the companion matrix of a monic polynomial p, then p is both the minimal and the characteristic polynomial of A.*

Proof. Let U be the operator on F^n represented by A in the standard basis. Then e_1 is a cyclic vector for U and p is clearly the minimal polynomial for U. Next either by direct computation or using Cayley-Hamilton theorem we conclude that p is also the characteristic polynomial for U. □

9 Cyclic Decomposition and the Rational form

9.1 Motivation

To decompose V into invariant subspaces where T acts in a nice way, the primary decomposition is in some sense a top-down approach and gives a way to decompose V. Now we look at the rational form of a matrix/operator which provides a different way to decompose V.

We try to decompose V into cyclic subspaces. So how does one go about this task? We could start with $\alpha_1 \neq 0$ and look at its cyclic subspace. This is invariant and hopefully the first piece of the jigsaw puzzle that is V. To get another piece take an $\alpha_2 \notin Z(\alpha_1;T)$ and look at its cyclic subspace. It too is invariant, but need not be independent of $Z(\alpha_1;T)$. This is our first hurdle. If the two cyclic subspaces intersect, there is a polynomial f such that $f(T)\alpha_2 \in Z(\alpha_1;T)$, naturally f is a multiple of the T conductor of α_2 into $Z(\alpha_1;T)$. If we want them to be independent, then we want $f(T)\alpha_2 = 0$. From here it is easy to see that f should be the minimal polynomial of α_2. So, we want an α_2 whose minimal polynomial and conductor are the same. Note that in the reasoning above, we could have replaces $Z(\alpha_1;T)$ with any subspace of V.

Starting with our arbitrary α_2, suppose g is the conductor of α_2 into $Z(\alpha_1;T)$, now if we could find a $\beta \in Z(\alpha_1;T)$ such that $g(T)\alpha_2 = g(T)\beta$, then $g(T)(\alpha_2 - \beta) = 0$, and the minimal polynomial of $\alpha_2 - \beta$ is the same as its conductor into $Z(\alpha_1;T)$.

Here we have a simple result: if W is an invariant subspace and $\alpha \in V, \beta \in W$, then the conductor of $\alpha - \beta$ into W is the same as that of α.

In general, we have $g(T)\alpha_2 = f(T)\alpha_1$. We now want is that $g \mid f$ for writing $f(T)\alpha_1 = g(T)\beta$ for some $\beta \in Z(\alpha_1;T)$ is the same as saying $g \mid f$. So, if we write $f = qg + r$, then we have

$$r(T)\alpha_1 = g(T)(q(T)\alpha_1 - \alpha_2).$$

Set $\beta = q(T)\alpha_1 - \alpha_2$ (different from the previous one). Let p_1, p_2 be the minmal polynomials of α_1, βrespectively and let h_1 be the conductor of α_1 into $Z(\beta;T)$. Keep in mind that g is the conductor of β into $Z(\alpha_1;T)$. We then have polynomials f_1, f_2, g_1 such that

$$p_1 = h_1 f_1, r = h_1 f_2, p_2 = g g_1.$$

Then we have

$$g_1 r(T)\alpha_1 = p_2(T)\beta = 0.$$

So, we have

$$p_1 \mid g_1 r$$

which, since $p_2 \neq 0$, we have $g_1 \neq 0$, so as long as r is non zero, means that

$$\deg\, p_1 \leq \deg\, g_1 + \deg\, r.$$

By the division algorithm, we have $\deg\, r < \deg\, g$, so

$$\deg\, p_1 \leq \deg\, g_1 + \deg\, r < \deg\, g_1 + \deg\, g = \deg\, p_2.$$

Here we have our breakthrough. If we impose the condition that α_1 have a minimal polynomial of maximum degree, then we would have arrived at a contradiction and therefore g must divide f and we can arrive at a suitable α_2 whose cyclic subspace is independent of that of α_1. Phew!

This gives us an algorithm to find independent cyclic subspaces, namely at each stage we try to look at the element having conductor of maximum degree into the subspace we have already built.

Observe that requiring α_1 to have a minimal polynomial of maximum degree is same as saying that it have a conductor into $\{0\}$ of maximum degree. This seems to show that the "seed" of our building procedure is the zero subspace, which is always invariant. So, in what follows we will try to generalise this in the sense that we start with an arbitrary invariant subspace W_0, so that if we want to we can keep W_0 as it is in our decomposition procedure.

However, W_0 cannot be arbitrary, we need it t have an invariant complement, i.e., a W_1 such that $V = W_0 \oplus W_1$. Another thing is that whenever $g(T)\alpha \in W_0$, we need a $\beta \in W_0$ such that $g(T)\alpha = g(T)\beta$.

Conversely, if W_0 has a invariant complement W_1, then $V = W_0 \oplus W_1$ and given $\alpha = \alpha_0 + \alpha_1 \in V$, for any polynomial $g, g(T)\alpha = g(T)\alpha_0 + g(T)\alpha_1$ and if this is in W_0, then by independence, $g(T)\alpha_1 = 0$ and we have an $\alpha_0 \in W_0$ such that $g(T)\alpha = g(T)\alpha_0$.

Definition 10. *Let T be a linear operator on a vector space V and let W be a subspace of V. We say that W is T-admissible if*

(i) W is invariant under T;
(ii) if $f(T)\beta$ is in W, there exists a vector γ in W such that $f(T)\beta = f(T)\gamma$.

9.2 The real deal

Theorem 16. *(**Cyclic Decomposition Theorem**). Let T be a linear operator on a finite dimensional vector space V and let W_0 be a proper T-admissible subspace of V. There exist non-zero vectors $\alpha_1, \ldots, \alpha_r$ in V with respective T annihilators $p_1, \ldots, p_r$ such that*

1. $V = W_0 \oplus Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$;
2. p_k *divides* $p_{k-1}, k = 2, \ldots, r$.

Furthermore, the integer r and the annihilators $p_1, \ldots, p_r$ are uniquely determined by (i), (ii), and the fact that no α_k is 0.

Recall: Suppose W_0 is T-admissible, and let g be the conductor of β into W_0 for some β and say $g(T)\gamma = g(T)\beta, \gamma \in W_0$. Since $\gamma \in W_0$, the conductor of $\beta - \gamma$ which divides its annihilator is g, hence g is also the minimal polynomial of $\beta - \gamma$. Therefore, $W_0, Z(\beta - \gamma; T)$ are independent.

Proof. Step 1. We first find invariant subspaces and later make them independent. Starting with W_0, as mentioned earlier, it is plausible that all our α_k come from searching for vectors with maximum degree conductors.

Since V is finite, assuming we have an invariant W, we have the following inequality, where $s(\beta; W)$ is the T-conductor (the polynomial) of β into W:

$$0 \leq \max_{\alpha \in V} \deg\,(s(\alpha; W)) \leq \dim\, V.$$

Starting with W_0 and choose β_1 for which $s(\beta_1; W_0)$ has maximum degree. Since W_0 is invariant, $W_1 = W_0 + Z(\beta_1; T)$ is also invariant. If $W_0 = V$, then every such $\beta_1 \in W_0$, else W_1 is strictly larger than W_0. So, we keep repeating this procedure and obtain vectors $\beta_1, \ldots, \beta_r$ such that

1. $V = W_0 + Z(\beta_1; T) + \cdots + Z(\beta_r; T)$;
2. if $1 \leq k \leq r$ and
$$W_k = W_0 + Z(\beta_1; T) + \cdots + Z(\beta_k; T)$$
then the conductor $p_k = s(\beta_k; W_{k-1})$ has maximum degree among all T-conductors into the subspace W_{k-1}, i.e., for every k
$$\deg\,(p_k) = \max_{\alpha \in V} \deg\,(s(\alpha; W_{k-1}))$$

This procedure terminates because V is finite dimensional. Also this step does not require W_0 to be T admissible, invariance is sufficient.

Step 2. Continuing with the notation from Step 1, fix a $1 \leq k \leq r$ and let $\beta \in V$ with $f = s(\beta; W_{k-1})$, then if
$$f(T)\beta = \beta_0 + \sum_{i=1}^{k-1} g_i(T)\beta_i, \beta_0 \in W_0$$
we claim that $f \mid g_i$ for every i as before and that $\beta_0 = f(T)\gamma_0$ for some $\gamma_0 \in W_0$.

If $k = 1$, then this is just the definition of W_0 being T-admissible, so take $k > 1$ and let $g_i = fq_i + r_i$ by the division algorithm with $0 \leq \deg\,(r_i) < \deg\,(f)$. Set
$$\gamma = \beta - \sum_{i=1}^{k-1} q_i(T)\beta_i$$
then the conductor of γ into W_{k-1} is the same as that of β which is f and $f(T)\gamma = \beta_0 + \sum_{i=1}^{k-1} r_i(T)\beta_i$. Assume some $r_i \neq 0$ and let j be the largest index such that $r_j \neq 0$.

Let p be the T conductor of γ into W_{j-1}. Why $j - 1$? Recall that in the previous subsection we used the annihilator, which is the conductor into the zero subspace, the subspace at the previous level. We are using the fact that at some point a choice was made between γ and β_j based on their conductor into W_{j-1} and β_j was chosen.

Since W_k contains W_{j-1}, it follows that $f \mid p$, so we write $p = fh$. Then (when $j = 1$, the last term is 0)

$$p(T)\gamma = fh(T)\gamma = r_j h(T)\beta_j + h(T)\beta_0 + \sum_{i=1}^{j-1} r_i h(T)\beta_i.$$

Since $p(T)\gamma \in W_{j-1}$ we get

$$r_j h(T)\beta_j \in W_{j-1} \Rightarrow p_j \mid r_j h \Rightarrow \deg(p_j) \leq \deg(r_j h)$$

where p_j is the conductor of β_j into W_{j-1}. By the choice of β_j,

$$\deg(p) = \deg(fh) \leq \deg(p_j)$$

and by the condition on r_j, we have

$$\deg(r_j h) < \deg(fh)$$

and we have a contradiction.

Hence every $r_j = 0$ and writing $g_i = fq_i$, we have $\beta_0 = f(T)\alpha$ for some $\alpha \in V$. Because W_0 is T admissible, it follows that $\beta_0 = f(T)\gamma_0$ for some $\gamma_0 \in W_0$.

Step 3. In this step we will do the fine tuning of the vectors we found. What we proved in Step 2 is that at each stage the subpaces we get using $\beta_1, \ldots, \beta_k$ are actually T admissible.

Starting with $\beta_1, \ldots, \beta_r$, for $1 \leq k \leq r$, by applying Step 2, we obtain

$$p_k\beta_k = p_k\gamma_0 + \sum_{i=1}^{k-1} p_k h_i \beta_i$$

where γ_0 is in W_0 and $h_1, \ldots, h_{k-1}$ are polynomials. Let

$$\alpha_k = \beta_k - \gamma_0 - \sum_{i=1}^{k-1} h_i\beta_i.$$

Since $\beta_k - \alpha_k$ is in W_{k-1}, we have $s(\alpha_k; W_{k-1}) = s(\beta_k; W_{k-1}) = p_k$ and since $p_k\alpha_k = 0$, we also have $W_{k-1} \cap Z(\alpha_k; T) = \{0\}$.

Thus we have

$$W_k = W_0 \oplus Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_k; T)$$

because it can be checked that we can retrieve β_k from α_k. Further the conductor of α_k into W_{k-1} and its annihilator is p_k which satisfies the maximality conditions. Further we have the trivial relation

$$p_k\alpha_k = 0 + p_1\alpha_1 + \cdots + p_{k-1}\alpha_{k-1}$$

which would imply, by Step 2, that p_k divides each p_i with $i < k$. Note that even though we could add $p_{k+1}\alpha_{k+1}$ to this sum, we cannot apply step 2 to conclude that $p_k | p_{k+1}$ because we know that p_k has a higher degree. In the proof of step 2, we knew p_j had a higher degree by the construction, here we miss one inequality. Fantastic.

Step 4. This is the final step regarding uniqueness. This is expected because at each step, the choice of the p_j was quite determined.

Suppose $\gamma_1, \ldots, \gamma_s$ are vectors in V with annihilators $g_1, \ldots, g_s$ such that g_i divides g_j for $i > j$ and $V = W_0 \oplus Z(\gamma_1; T) \oplus \cdots \oplus Z(\gamma_s; T)$. We will show that $r = s$ and each $g_i = p_i$. Since $V = W_0 + \oplus Z(\gamma_1; T) \oplus \cdots \oplus Z(\gamma_s; T)$, we write

$$\alpha_1 = \gamma_0 + h_1(T)\gamma_1 + \cdots + h_s(T)\gamma_s$$

and upon applying $g_1(T)$, since each $g_i \mid g_1$, we have $g_1(T)\alpha_1 = g_1(T)\gamma_0$ which means that $p_1 \mid g_1$. Similarly, $g_1 \mid p_1$ and since both are monic, they must be equal. Since they are equal, the dimensions of $Z(\alpha_1; T)$ and $Z(\gamma_1; T)$ are equal. Now,

$$V = W_0 \oplus Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T) = W_0 \oplus Z(\gamma_1; T) \oplus \cdots \oplus Z(\gamma_s; T)$$

applying p_2 throughout gives

$$p_2V = p_2W_0 \oplus Z(p_2\alpha_1; T)$$
$$p_2V = p_2W_0 \oplus Z(p_2\gamma_1; T) \oplus \cdots \oplus Z(\gamma_s; T)$$

It is clear to see that the dimensions of $Z(p_2\alpha_1; T)$ and $Z(p_2\gamma_1; T)$ are equal because the annihilators of α_1, γ_1 are the same. So, by comparing dimensions, we must have $p_2\gamma_2 = 0$ i.e., $g_2 \mid p_2$. Note that this immediately gives $g_i \mid p_2$ for $i \geq 3$. By a similar calculation, we also get $p_2 \mid g_2$. Therefore, they are equal and α_2, γ_2 have the same annihilators.

Continuing similarly, we conclude that $r = s$ and $g_i = p_i$ for $1 \leq i \leq r$. □

9.3 Corollaries and applications

Corollary 4. *If T is a linear operator on a finite dimensional vector space, then every T admissible subspace has a complementary subspace which is also invariant under T.*

Proof. If $W_0 = V$, there is nothing to prove. Else, we get vectors $\alpha_1, \ldots, \alpha_r$ such that the subspace $Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$ is an invariant complement of W_0. □

Corollary 5. *Let T be a linear operator on a finite dimensional vector space V.*

(a) There exists a vector α in V such that the T-annihilator of α is the minimal polynomial for T.
(b) T has a cyclic vector if and only if its characteristic and minimal polynomials are identical.

Proof. If $V = \{0\}$, then there is nothing to prove. If $V \neq \{0\}$, then let $V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$ where the annihilators $p_1, \ldots, p_r$ are such that p_{k+1} divides $p_k, 1 \leq i \leq k$. Then it follows that the minimal polynomial is p_1 and α_1 is a vector whose annihilator is the minimal polynomial proving (a).

We have already seen that if there is a cyclic vector, then the minimal and characteristic polynomials are the same. If the two polynomials are the same, then get hold of an α as in (a), then by comparing dimensions, it follows that $V = Z(\alpha_1; T)$. □

Theorem 17. *(**Generalized Cayley-Hamilton Theorem**) Let T be a linear operator on a finite dimensional vector space V. Let p and f be the minimal and characteristic polynomials for T respectively.*

(i) p divides f.
(ii) p and f have the same prime factors, except for multiplicities.
(iii) If

$$p = f_1^{r_1} \ldots f_k^{r_k}$$

is the factorization of p, then

$$f = f_1^{d_1} \ldots f_k^{d_k}$$

where d_i is the nullity of $f_i(T)^{r_i}$ divided by the degree of f_i.

Proof. If V is the zero space, there is nothing to show, else we can write $V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_k; T)$ with $p_1, \ldots, p_k$ being the annihilators of $\alpha_1, \ldots, \alpha_k$ respectively. As noted in a corollary above, $p_1 = p$ is the minimal polynomial. Now if U_i denotes the restriction of T to $Z(\alpha_i; T)$, then U_i has a cyclic vector and has minimal polynomial p_i. Therefore, the charactersitc polynomial of U_i is also p_i. With a suitable basis, T has a block diagonal matrix and it is clear that the charactersitic polynomial should be $f = p_1 \ldots p_k$. This proves (i).

As for (ii), since $p_k \mid p_{k-1} \mid \cdots \mid p_1$, it follows that if g is a prime factor of p_1, then g is also the prime factor of f and if g is a prime factor of $f = p_1 \ldots p_k$, then g must divide some p_j and hence divides p_1. Therefore, p, f have the same prime factors.

Now, let $p = f_1^{r_1} \ldots f_k^{r_k}$ be the prime factorization of the minimal polynomial. By the primary decomposition theorem, if we let V_i be the nullspace of $f_i(T)^{r_i}$, then

$$V = V_1 \oplus \cdots \oplus V_r$$

and $f_i^{r_i}$ is the minimal polynomial of the operator T_i obtained by the restriction of T to V_i. By part (ii) proved above, the characteristic polynomial of V_i is $f_i^{d_i}$ where $d_i = \dim V_i$. Clearly, $d_i = \dim V_i/\deg f_i$ and $\dim V_i = \text{nullity } f_i(T)^{r_i}$. Since T is the direct sum of the operators $T_1, \ldots, T_k$, it follows that

$$f = f_1^{d_1} \ldots f_k^{d_k}. \qquad \square$$

Corollary 6. *If T is a nilpotent linear operator on a vector space of dimension n, then the characteristic polynomial of T is x^n.*

There is always an analogous decomposition theorem for matrices. In the case of cyclic decomposition, we can get the matrix of T into a block diagonal form where each block is a companion matrix for dome polynomial. The divisibility conditions are an added bonus. Such a matrix is said to be in its rational form.

Theorem 18. *Let F be a field and let B be an $n \times n$ matrix over F. Then B is similar over the field F to one and only one rational form.*

Proof. Let T be the linear operator on F^n that is represented by B in the standard basis (note that there is only one such T). Then, there is a basis in which T is represented by a matrix A in rational form. Then B is similar to A. Suppose B is similar to another matrix C which is in rational form. This means that that there is some ordered basis for F^n in which the operator T is represented by the matrix C. If C is the direct sum of companion matrices C_i of monic polynomials $g_1, \ldots, g_s$ and $g_s \mid g_{s-1} \mid \cdots \mid g_1$, then it is clear that we will have vectors $\beta_1, \ldots, \beta_s$ with annihilators $g_1, \ldots, g_s$ such that

$$V = Z(\beta_1; T) \oplus \cdots \oplus Z(\beta_s; T).$$

By uniqueness of the cyclic decomposition, we must have $C = A$. $\square$

The polynomials $p_1, \ldots, p_r$ are called the **invariant factors** for the matrix B.

We will compute the cyclic decomposition of a diagonalizable operator T on a vector space V of dimension n. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T and $V_1, \ldots, V_k$ be the corresponding eigenspaces. Then

$$V = V_1 \oplus \cdots \oplus V_k.$$

Let $\alpha \in V$. We will first have a look at $Z(\alpha; T)$. Let

$$\alpha = \beta_1 + \cdots + \beta_k$$

with $\beta_i \in V_i$. Then for any polynomial f, we have

$$f(T)\alpha = f(c_1)\beta_1 + \cdots + f(c_k)\beta_k.$$

By using Lagrange polynomials, we can see that each β_i is in $Z(\alpha; T)$. It is also clear that $Z(\alpha; T)$ is in the span of $\beta_1, \ldots, \beta_k$. So,

$$Z(\alpha; T) = \langle \beta_1, \ldots, \beta_k \rangle.$$

Next we need the annihilator of α. So, if $f(T)\alpha = 0$, then $f(c_i) = 0$ whenever $\beta_i \neq 0$. Therefore the annihilator of α is the product

$$\prod_{\beta_i \neq 0} (x - c_i).$$

Now, let $\beta_i = \{\beta_1^i, \ldots, \beta_{d_i}^i\}$ be an ordered basis for V_i and let $r = \max_i d_i$. We define vectors $\alpha_1, \ldots, \alpha_r$ by

$$\alpha_j = \sum_{d_i \geq j} \beta_j^i, 1 \leq j \leq r.$$

The cyclic subspace spanned by α_j is the subspace spanned by β_j^i, as i runs over those indices for which $d_i \geq j$. The T annihilator of α_j is

$$p_j = \prod_{d_i \geq j} (x - c_i).$$

We have

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$$

because β_j^i belongs to one and only one of the subspaces $Z(\alpha_1; T), \ldots, Z(\alpha_r; T)$ and $\beta = (\beta_1, \ldots, \beta_k)$ is a basis for V. Divisibility condition also follows from the definition fo $p_1, \ldots, p_k$.

10 Jordan Form

Let N be any nilpotent operator. Why are we studying nilpotent operators? We have shown earlier that if the characteristic polynomial or minimal polynomial of an operator T splits into linear factors, then T can be written as the sum of a diagonal operator D and a nilpotent operator N which are both polynomials in T and commute with each other and we understand D quite well.

Let N be a nilpotent operator on an n dimensional vector space V with characteristic polynomial x^n and minimal polynomial x^k for some k. Using the cyclic decomposition theorem, we have

$$V = Z(\alpha_1; T) \oplus \cdots \oplus Z(\alpha_r; T)$$

with the annihilator of α_i being x^{k_i} for some k_i. Further we must have

$$k = k_1 \geq k_2 \geq \cdots \geq k_r \geq 1.$$

The matrix of N restricted to $Z(\alpha_i; T)$ is the companion matrix of x^{k_i} which is a $k_i \times k_i$ matrix of the form

$$\begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

So, N will be similar to a direct sum of such matrices and not that there is a gap where two blocks meet. From this matrix representation, it is clear that the nullspace of N is spanned by $N^{k_i-1}\alpha_i$.

Alternatively, if some $\alpha = f_1\alpha_1 + \cdots + f_r\alpha_r$ is in the nullspace of N, then we must have $N(f_i\alpha_i) = 0 \Rightarrow xf_i\alpha_i = 0 \Rightarrow x^{k_i} \mid xf_i$. Since $\deg f_i < k_i$, we must have $f_i = cx^{k_i-1}$ for some constant c and therefore, α is in the span of $N^{k_1-1}\alpha_1, \ldots, N^{k_r-1}\alpha_r$.

Observe that instead of the characteristic polynomial being x^n, we could have worked with $(x-c)^n$ for some constant c. In this case, we take $N = T - cI$ to be the nilpotent matrix. Then T is the sum of cI and N. Since cI is similar to only itself, ot follows that there is a basis where T is represented by a matrix of the form above except with c on the diagonal.

Take a linear operator T on V whose minimal polynomial p splits into linear factors, say

$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}.$$

If W_i is the nullspace of $(T - c_iI)^{r_i}$, then by the primary decomposition theorem

$$V = W_1 \oplus \cdots \oplus W_k$$

and the minimal polynomial of T restricted to W_i is $(x - c_i)^{r_i}$. Now it is clear that by taking suitable bases for W_i and adjoining them, we get a basis for V in which T is represented by a matrix of the form

$$\begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & A_k \end{bmatrix}$$

where each A_k is a direct sum of elementary Jordan matrices with characteristic value c_i of decreasing size. An elementary Jordan matrix with characteristic value c is a matrix with c on the diagonal and 1s below the diagonal. A matrix as above is said to be in **Jordan form**.

Now we show that the Jordan form is unique upto the order in which we write the characteristic values. Suppose T is represented by a matrix of the form A above in some basis, then we must show that upto rearrangements of c_i, A is unique. If A_i is a $d_i \times d_i$ matrix, then it is clear that the multiplicity of c_i in the characteristic polynomials is d_i. So the numbers $d_1, \ldots, d_k$ are unique upto rearrangements and so are the scalars $c_1, \ldots, c_k$. Since A is the direct sum of A_i, it follows that

$$V = W_1 \oplus \cdots \oplus W_k$$

each of which are invariant under T. Note that W_i must be the nullspace of $(T - c_i I)^n$ where $n = \dim V$ (in fact it is sufficient to take $(T - c_i I)^{d_i}$) for $(A_i - c_i I)$ is nilpotent and for $i \neq j$, $A_j - c_i I$ is non-singular. So W_i are unique, hence A_i, given by the rational form of $T_i - c_i I$, is also unique.

Some observations:

1. The entries of A not on or below the diagonlar are 0. The diagonal contains k distinct values $c_1, \ldots, c_k$, c_i appearing d_i times where d_i is the multiplicity of c_i in the characteristic polynomial, i.e., $d_i = \dim W_i$.
2. For each i, the matrix A_i is the direct sum of some n_i elementary Jordan matrices where n_i is the dimension of the eigenspace associated with c_i. For, n_i is the number of elementary nilpotent blocks in the rational form for $T_i - c_i I$ and is thus equal to the dimension of the nullspace of $T - c_i I$. In particular T is diagonalizable if and only if $d_i = n_i$ for each i.
3. For each i, the first elementary Jordan block of A_i is of dimension $r_i \times r_i$ where r_i is the multiplicity of c_i in the minimal polynomial of T. This follows from the fact that the minimal polynomial for T_i is going to be that factor of the minimal polynomial of T that concerns to c_i and the fact that in the rational form, the the first block has the same dimensions as the degree of the minimal polynomial.

So, as you can see, the Jordan form allows us to extract a lot of information.Of course, there is an analogous result for matrices. Further there is a unique Jordan form for every matrix over an algebraically closed field.

11 Computing invariant factors

Having proved the existence of invariant factors and particular types of decompositions, we would like to know how to actually compute them. We start with a companion matrix. Let A be the companion matrix of a polynomial

$$p = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0.$$

Here e_1 is a cyclic vector under the action of A on F^n and has annihilator p, hence p will be the minimal polynomial of A. By Cayley-Hamilton theorem, p is also the characteristic polynomial of A. We can also prove this by a direct computation.

$$xI - A = \begin{bmatrix} x & 0 & 0 & \ldots & 0 & c_0 \\ -1 & x & 0 & \ldots & 0 & c_1 \\ 0 & -1 & x & \ldots & 0 & c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & x & c_{n-2} \\ 0 & 0 & 0 & \ldots & -1 & x + c_{n-1} \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & p \\ -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 0 \end{bmatrix}$$

Add x times the last row to row $n-1$. Then add x times the new row $n-1$ to row $n-2$. Repeat this until you have p as the last entry in row 1. Then using the columns, clear the other entries of the last column. Finally, rearranging the columns and multiplying by -1s, it is clear to see that we get a diagonal matrix with entries $p, 1, \ldots, 1$. During all these operations, the determinant doesn't change, so the characteristic polynomial of A is indeed p.

The goal is to systematically do such row and column operations on $xI - A$ for any matrix A to arrive at a diagonal matrix with the invariant factors on the diagonal. Let us for once clarify the rules of these operations. They should be such that the determinant remains the same upto a scalar multiplication, so for any matrix A over $F[x]$ we are allowed to

1. multiply any row by scalars in F;
2. add a polynomial multiple of one row to another;
3. interchange two rows.

Note that we can multiply by scalars, but not by non constant polynomials because at each stage we should be able to reverse the operation. The operations listed above are actually elementary row operations with the restriction that multiplication be done by scalars only.

Lemma 7. *Let M be a matrix in $F[x]^{n\times n}$ which has some non zero entry in its first column, and let p be the greatest common divisor of the entries in column 1 of M. Then M is row equivalent to a matrix N which has $\begin{bmatrix} p & 0 & \dots & 0 \end{bmatrix}^t$ as its first column.*

Proof. This is an application of the Euclidean algorithm for finding the gcd of polynomials. At each stage, make the smallest degree polynomial of the first column monic, then subtract appropriate multiples of the same from other rows to reduce the degree of each entry and repeat. In the end, swap rows so that p comes first. □

Theorem 19. *Let P be an $m \times m$ matrix with entries from $F[x]$. The following are equivalent.*

(i) P is invertible.
(ii) The determinant of P is a non-zero scalar polynomial.
(iii) P is row equivalent to the $m \times m$ identity matrix.
(iv) P is a product of elementary matrices.

Proof. (i)⇒ (ii) because if P is invertible, then the determinant, a polynomial, is invertible, hence must be a non-zero scalar.

(ii) ⇒ (iii) If the determinant is a non-zero scalar polynomial, then the gcd of entries of each column must be one. Using the previous lemma, we can find a row equivalent upper triangular matrix with 1s on the diagonal. Using even more row operations, we can make the off diagonal entries 0. So, A is row equivalent to the identity matrix.

(iii) ⇒ (iv) Since each row operation is invertible, by inverting these operations, we can go from identity matrix to A, so A is a product of elementary matrices.

(iv) ⇒ (i) is obvious. □

For major simplifications, it is useful to consider column operations as well. These are similar to the row operations except they are done on the columns instead of the rows.

Definition 11. *The matrix N is equivalent to the matrix M if we can pass from M to N by means of a sequence of operations*

$$M = M_0 \to M_1 \to \cdots \to M_k = N$$

each of which is an elementary row operation or an elementary column operation.

Since each row and column operation is obtained by some matrix multiplication, we have the obvious

Theorem 20. *Let M and N be $m \times n$ matrices over $F[x]$. Then N is equivalent to M if and only if*

$$N = PMQ$$

where P is an invertible matrix in $F[x]^{m\times m}$ and Q is an invertible matrix in $F[x]^{n\times n}$.

Theorem 21. *Let A be an $n \times n$ matrix with entries in the field F, and let $p_1, \dots, p_r$ be the invariant factors for A. The matrix $xI - A$ is equivalent to the $n \times n$ diagonal matrix with diagonal entries $p_1, \dots, p_k, 1, \dots, 1$.*

Proof. There is an invertible matrix P such that $P^{-1}AP$ is in the rational form

$$\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{bmatrix}$$

and $P^{-1}(xI - A)P = xI - P^{-1}AP$. Since each A_i is a companion matrix there are row and column operators that bring each A_i to a diagonal matrix with entries $p_i, 1, \ldots, 1$. Since the blocks are placed diagonally, row operations done one one A_i don't affect other A_js, so we can do all of them independently. It is now clear that the theorem is true. □

Definition 12. *Let N be a matrix in $F[x]^{m\times n}$. We say that N is in (Smith) normal form if*

(a) every entry off the main diagonal of N is 0;
(b) on the main diagonal of N there appear (in order) polynomials $f_1, \ldots, f_l$ such that f_k divides $f_{k+1}, 1 \leq k \leq l + 1$.

In the definition, $l = \min(m, n)$ and the main diagonal entries are $N_{kk}, k = 1, \ldots, l$.

Note that the previous theorem says that every square matrix is equivalent to a matrix in Smith normal form because the polynomials $p_1, \ldots, p_k$ can be arranged to follow the divisibility conditions. Next we show that this is true for any rectangular matrix and give an algorithm to do so.

Theorem 22. *Let M be an $m \times n$ matrix with entries from $F[x]$. Then M is equivalent to a matrix N which is in normal form.*

Proof. If $M = 0$, there is nothing to prove, so assume $M \neq 0$. Now, the idea is to use a similar procedure as earlier in order to make the first entry the gcd of all entries. So, first we find a column of M having an entry of the least degree. Bring that column to the first row. Using the previous algorithm, which I will call algorithm 1, find the equivalent matrix

$$\begin{bmatrix} p & a & \ldots & b \\ 0 & c & \ldots & d \\ \vdots & \vdots & & \vdots \\ 0 & e & \ldots & f \end{bmatrix}$$

If the other entries of the first column are not zero, using a similar algorithm for columns, which I will call algorithm 2, make the first row have only one non zero entry which appears in the first position as below

$$\begin{bmatrix} q & 0 & \ldots & 0 \\ a' & c' & \ldots & d' \\ \vdots & \vdots & & \vdots \\ b' & e' & \ldots & f' \end{bmatrix}$$

We may have altered the first column, however the degree of the first entry has decreased. In fact, it has to strictly decrease if at the end of these two steps, the first row and first column don't have zero entries besides q. This is because the gcd of the entries in the first row and column would not be q and would be strictly smaller. So, this procedure will eventually stop to result in an equivalent matrix of the form

$$\begin{bmatrix} g & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & S & \\ 0 & & & \end{bmatrix}$$

If g does not divide every entry of S, then find an entry not divisible by f and add that row to the first row and repeat this procedure, which I will call algorithm 3. Algorithm 3 is a repeated

application of the other two algorithms. This time the degree of g will strictly decrease too. So, we eventually arrive at an equivalent matrix of the form

$$\begin{bmatrix} f_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & R & \\ 0 & & & \end{bmatrix}$$

where f_1 divides every entry of R. Now, we can do the same algorithm on R and the effects are isolated to R, that is to say f_1 is unaffected by these operations on R. So, we can conclude that a normal form exists. □

Next we want to show that this form is unique. It should be quite clear that since gcd is unique, and all the entries of the normal form were gcds of specific entries, that the normal form is unique. We observe that row and column operations don't change the determinant of any square submatrix and neither do they change the gcds of the entries of a matrix (all row and column operations are invertible). The first entry we get in the normal form divides every other entry, so it is the gcd of all 1×1 submatrices. The second entry, with the notation used above would divide all entries of R. This however, is equivalent to saying that it divides all determinants of 2×2 submatrices because row and column operations don't change these determinants as shown below. We define

Definition 13. *Let M be an $m \times n$ matrix with entries in $F[x]$. If $1 \leq k \leq \min(m, n)$, we define $\delta_k(M)$ to be the greatest common divisor of the determinants of all $k \times k$ submatrices of M.*

Theorem 23. *If M and N are equivalent $m \times n$ matrices with entries in $F[x]$, then*

$$\delta_k(M) = \delta_k(N), 1 \leq k \leq \min(m, n).$$

Proof. It is enough to show this for a single row operation. We have three types of operations. With a little thinking, it should be clear that none of these operations change the gcd of determinants of $k \times k$ submatrices. There may be multiplication by scalars of some determinants, but in the end we are looking for monic gcds. □

Corollary 7. *Each matrix M in $F[x]^{n \times n}$ is equivalent to precisely one matrix N which is in normal form. The polynomials $f_1, \dots, f_l$ which occur on the main diagonal of N are*

$$f_k = \frac{\delta_k(M)}{\delta_{k-1}(M)}, 1 \leq k \leq \min(m, n)$$

where for convenience, we define $\delta_0(M) = 1$.

Proof. In a normal form N with diagonal entries $f_1, \dots, f_l$, it is easy to see that $\delta_k(N) = f_1 \dots f_k$. □

Now, for an $n \times n$ matrix A, the normal form of $xI - A$ has diagonal entries $1, \dots, 1, p_r, \dots, p_1$. The last corollary tells us what $p_1, \dots, p_r$ are in terms of submatrices of $xI - A$. The number $n - r$ is the largest k such that $\delta_k(xI - A) = 1$. The minimal polynomial p_1 s the characteristic polynomial divided by the gcd of the determinants of all $(n-1) \times (n-1)$ submatrices of $xI - A$.

12 Summary; Semi-Simple operators

We started with a linear operator T on a finite dimensional vector space V and wanted to decompose V in such a way that T acted in a simple manner on each of these parts. This lead us to the notion of invariant subspaces and the simplest invariant subspaces are the eigenspaces. We found a way to find eigenvalues and eigenvectors which lead us to look at annihilators of T and the notion of minimal polynomials. After that we found out relations between the minimal and characteristic polynomials and the Cayley-Hamilton theorem.

Next, we looked at invariant subspaces, how T behaves when restricted to invariant subspaces and the conductors. Later, we discussed direct sums which offer a way to glue subspaces back to V.

Then comes a findamental result, the primary decomposition theorem. Our first step towards our goal of decomposing V. On the way we saw projection operators. After that we looked at the cyclic decomposition theorem, which, in some sense, is a bottom up approach contrasting the top down approach of the primary decomposition theorem.

All of these results culminated in the Jordan rational form which finds a basis such that T is very elementary on V with respect to that basis. Given an operator, one can compute the Jordan form if one has sufficient time and patience.

One other result that was obtained with the primary decomposition theorem was that we can write T as the sum of a diagonal and a nilpotent operator. This however required that the underlying field be algebraically closed. Here we define an operator that serves the role of a digonal operator.

Definition 14. *Let V be a finite dimensional vector space over the field F, and let T be a linear operator on V. We say that T is semi-simple if every invariant subspace has a complementary T-invariant subspace.*

The definition is inspired by a general notion of semisimplicity. Let R be a ring (not necessarily commutative, or with unity), and M be an R-module. It is said to be simple if M has no nontrivial submodules and M is semisimple if it can be written as an arbitrary direct sum of simple submodules. The ring R is said to be semisimple if every R-module is semisimple. It is a theorem that M is semisimple if every submodule has an invariant complement.

Lemma 8. *Let T be a linear operator on a finite dimensional vector space V and $V = W_1 \oplus \cdots \oplus W_k$ be the primary decomposition for T. In other words, if p is the minimal polynomial for T and $p = p_1^{r_1} \dots p_k^{r_k}$ is the prime factorization of p, then W_j is the null space of $p_j^{r_j}$. Let W be any subspace of V which is invariant under T. Then*

$$W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$$

Proof. Recall that if $E_1, \dots, E_k$ are the projections associated with the decomposition above, then there are polynomials $h_1, \dots, h_k$ such that $h_i(T) = E_k, 1 \leq i \leq k$. Now, if $\alpha = \alpha_1 + \cdots + \alpha_k \in W$ with each $\alpha_i \in W_i$, then $E_i\alpha \in W$ because it is invariant. So, $\alpha_i \in W_i$. Therefore, $\alpha \in (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$. It follows that $W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$. □

Lemma 9. *Let T be a linear operator on V, and suppose that the minimal polynomial for T is irreducible over F. Then T is semi-simple.*

Proof. Let W be a subspace of V which is invariant under T. As seen in the cyclic decomposition theorem, it is sufficient to show that W is T admissible. Take a $\beta \in V$ and f a polynomial such that $f(T)\beta \in W$, we need an $\alpha \in W$ such that $f(T)\alpha = f(T)\beta$. If $f(T)\beta = 0$, take $\alpha = 0 \in W$, else $f(T)\beta \neq 0$, so p does not divide f. Therefore, f, p are coprime because p is irreducible and there exist polynomials g, h such that $fg + ph = 1$.

Then, $f(T)g(T) = I$ because $p(T) = 0$. So, $\beta = f(T)g(T)\beta = g(T)f(T)\beta$. Since $f(T)\beta \in W$ and W is invariant, we have $\beta \in W$. Taking $\alpha = \beta, W$ is T-admissible. □

Theorem 24. *Let T be a linear operator on the finite dimensional vector space V. A necessary and sufficient condition that T be semi-simple is that the minimal polynomial p for T be of the form $p = p_1 \dots p_k$ where $p_1, \dots, p_k$ are distinct irreducible polynomials over the scalar field F.*

Proof. Assume that T is semi-simple and that an irreducible (non constant) factor g repeated in the factorization of p, say $p = g^2h$. Take W to be the null space of $g(T)$. Clearly, W is T invariant. Since gh is not the minimal polynomial, there is an $\alpha \in V$ such that $\beta = gh(T)\alpha \neq 0$. Clearly, $\beta \in W$. However, there is no $\gamma \in W$ such that $gh(\gamma) = \beta$ for $gh(\gamma) = hg(\gamma) = 0 \neq \beta$. Therefore, W doesn't have a T invariant complement contradicting the assumption on T.

Conversely, assume that p is square-free and has the factorization stated above. Let $V = W_1 \oplus \cdots \oplus W_k$ be the primary decompostition of V, where W_j is the null space of $p_j(T)$ and p_j is the

minimal polynomial of $T_j = T|_{W_j}$. Now, $W \cap W_j$ is invariant and by the lemma above, there is a V_j such that $W_j = (W \cap W_j) \oplus V_j$. By the previous lemma we have $W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k)$. Now using the decomposition of each W_j and rearranging the factors of V, it is clear that W has a T invariant complement. □

Corollary 8. *If T is a linear operator on a finite dimensional vector space V over an algebraically closed field, then T is semi-simple if and only if it is diagonalizable.*

Lemma 10. ***(Taylor's Formula)*** *Let F be a field of characteristic zero and let g, h be polynomials over F. If f is any polynomial over F with $\deg f \leq n$, then*

$$f(g) = f(h) + f^{(1)}(h)(g-h) + \frac{f^{(2)}(h)}{2!}(g-h)^2 + \cdots + \frac{f^{(n)}(h)}{n!}(g-h)^n$$

where $f^{(k)}$ denotes the kth formal derivative of f and $f(g)$ is just the composition.

Proof. Since derivative and composition are linear, it is sufficient to prove this for $f = x^k$ for every $k \geq 0$. Observe that the expression given is just $g^k = [h + (g-h)]^k$. Thus the lemma is true. □

Next we give some equivalent conditions for f to be square free.

Lemma 11. *Let F be a field of characteristic zero, let f be a polynomial over F, and f' be the derivative of f. The following are equivalent:*

(a) f is the product of distinct irreducible polynomials over F.
(b) f, f' are coprime.
(c) f has no repeated root.

Proof. The proof is quite standard and left to the reader. The important point to note is that since the characteristic is zero, the derivative of a non constant polynomials cannot be zero. □

Theorem 25. *Let F be a field of characteristic zero, let V be a finite dimensional vector space over F, and let T be a linear operator on V. Let $\mathcal{B}$ be an ordered basis for V and let A be the matrix of T in this ordered basis. Then T is semi-simple if and only if the matrix A is similar to a diagonal matrix over the algebraic closure of F.*

Proof. Let p be the minimal polynomial of T. Then T is semi-simple if and only if p is square free. This is true if and only if p is a product of distinct linear factors over the algebraic closure of F because two distinct irreducibles cannot share a root and because F is of characteristic zero, each irreducible factor also has no repeated root. The latter condition is true if and only if A is similar to a diagonal matrix over the algebraic closure of F. □

Theorem 26. *Let F be a field of characteristic zero, let V be a finite dimensional vector space over F, and let T be a linear operator on V. There is a semi-simple operator S on V and a nilpotent operator N on V such that*

(i) $T = S + N$;
(ii) $SN = NS$.

Furthermore, the semi-simple S and nilpotent N satisfying (i) and (ii) are unique, and each is a polynomial in T.

Proof. Let $p_1^{r_1} \dots p_k^{r_k}$ be the prime factorization of the minimal polynomial for T. We know that the minimal polynomial for S should be square free, the first thing to expect is that the S we obtain would have $f = p_1 \dots p_k$ as its minimal polynomial. Next N should be a polynomial, say g, in T whose power should be divisible by p. Note that if r is the maximum of $r_1, \dots, r_k$, then f^r is divisible by p and no lesser power of f is divisible by p. We have $S = T - g(T)$ and we want g such that $f(T - g(T)) = 0$ i.e., $f(x - g(x))$ is divisible by p. Further we want $g(T)$ to be nilpotent, then p divides some power of g. Since each p_i is irreducible, we see that f should divide g.

Let us look at some example. If T was semi-simple, then $g(x) = 0, p = f$ and $f(x - g(x))$ is divisible by f. Next we take $g = cf$, since $f \mid g$, we will take g to be a scalar multiple of f. Then,

$$f(x - cf) = f(x) - cff'(x) + f^2 b$$

where b is some polynomial. Since p should divide this, and $f \neq p$, we want atleast f^2 to divide $f(x - cf)$. Since f, f' are coprime, it is clear that c can be chose in such a way that $1 - cf'$ is divisible by f but c will not be a polynomial because the degree of f' is strictly smaller than that of f.

This is where things get slightly trickier. We want g which is divisible by f and some power is divisible by p. What we do is write g as a sum of powers of f with polynomial coefficients. Since g is divisible by f, this is always possible. So, we write $g = \sum g_i f^i$ and we want

$$f(x - \sum g_i f^i)$$

to be divisible by p or by f^r.

We will find a sequence of polynomials $g_0, g_1, \ldots$ such that

$$f(x - \sum_{i=0}^{n} g_i f^i)$$

is divisible by f^{n+1} for $n = 0, 1, \ldots$. For $n = 0$, we can take $g_0 = 0$. Having found $g_0, \ldots, g_{n-1}$ set

$$h = x - \sum_{i=0}^{n-1} g_i f^i.$$

Now we want g_n such that

$$f(h - g_n f^n) = f(h) - g_n f^n f'(h) + f^{n+1} b$$

is divisible by f^{n+1}. Here b is some polynomial. Since $f(h) = qf^n$ by assumption, all we need to do is to make sure that $q - g_n f'(h)$ is divisible by f. Now f, f' are coprime, so there are polynomials a, b such that $cf + df' = 1$. Substituting h, we see that $c(h)f(h) + d(h)f'(h) = 1$, so

$$qc(h)f(h) + qd(h)f'(h) = q.$$

Since $f(h)$ is divisible by $f, g_n = qd(h)$ does the job.

Now, take $n = r - 1$, then there are polynomials $g_0, \ldots, g_{r-1}$ such that

$$f(x - \sum_{i=0}^{r-1} g_i f^i)$$

is divisible by f^r. Take $g = \sum_{i=0}^{r-1} g_i f^i$ and set $N = g(T)$. Since g is divisible by f, N is nilpotent. Further, if we set $S = T - N$, then $f(S) = 0$, so the minimal polynomial of S divides f, hence is square free and therefore, S is semi-simple.

Thus, we have written T as $S + N$ where S is semi-simple and N is nilpotent. Further, $SN = NS$. Now, go to the algebraic closure of F. There, S becomes diagonal and N stays nilpotent. We had previously shown that $T = D + N$ with D diagonal and N nilpotent and $DN = ND$ is a unique decomposition. It follows that S, N are unique. □

13 Reference

Linear Algebra - Kenneth Hoffman, Ray Kunze