

A journey through decompositions of linear transformations

Anirbit

*Department of Theoretical Physics
Tata Institute of Fundamental Research,
Mumbai 400005, India*

(Dated: September 30, 2009)

The objective of this article is to indicate the arguments that go into understanding the basic kinds of decompositions of linear transformation. The core ideas have been pointed out with the expectation that the reader can fill in the detailed proofs but the two decomposition theorems mentioned next are proven in details.

Starting point is to understand the notion of diagonalizability of matrices of which 3 versions will be explained through matching of algebraic and geometric dimensions of eigenvalues, through existence of distinct roots of minimal polynomials and through families of projection operators.

The final goal is to understand the *Primary Decomposition Theorem (Rational Form)* and the *Cyclic Decomposition Theorem* and how the first alone gives the powerful *Jordan-Chevalley Decomposition* and the both together give the miraculous *Jordan Decomposition*.

En route we will need to understand the concepts of conductors and annihilators of vectors, cyclic vectors and cyclic vector spaces, minimal polynomials and characteristic polynomials, Cayley-Hamilton Theorem, Lagrange Polynomials and the idea of companion matrices.

I. MOTIVATION

Given a matrix representation of a linear transformation on a finite dimensional vector space one can ask the following natural questions:

- Can it be decided whether the matrix is diagonalizable *without* finding the eigen vectors?
- Can *any* matrix be block-diagonalized?
- Can *any* matrix be brought to a form where there are entries only along the diagonal and super/sub diagonal?

The answer to the first two questions is affirmative and over algebraically closed fields like \mathbb{C} the answer to the last question is also affirmative.

The point of this article is to indicate the key ideas that explain the above.

A. Decomposition of Linear Transformations

For the cause of Representation Theory it is important to understand the elementary ideas that go into the idea of decomposition of a linear transformation into transformations on smaller dimensional vector-spaces. It is desirable that a given linear transformation on an n -dimensional vector-space is writable as a "direct sum" of n linear transformations on one-dimensional subspaces of the original space. This is what is the idea behind "diagonalization". But we know that all linear transformations are not diagonalizable and then it becomes

necessary to understand which transformations are diagonalizable and when. Further if the matrix of the linear transformation is not diagonalizable it might still be reducible into a form that is "block-diagonal" and we need to see what is the simplest block-diagonal form to which a matrix can be reduced. Here I shall list out in a logical sequence the theorems which establish the above things and I shall indicate the basic idea behind them omitting the detailed proofs.

All vector-spaces in the following section are finite dimensional. Many of the concepts might not naturally extend for infinite dimensional vector-spaces.

3 elementary operations on vector spaces:

- Given two subspaces W_1 and W_2 of a vector space V we denote as $W_1 + W_2$ the **sum** of the two subspaces defined as the set $\{w_1 + w_2 | w_1 \in W_1 \text{ and } w_2 \in W_2\}$. The notion can be naturally extended to arbitrary number of subspaces.

One notes that $\dim(W_1) + \dim(W_2) = \dim(W_1 \cap W_2) + \dim(W_1 + W_2)$.

- Given k subspaces of V , say W_1, W_2, \dots, W_k this set of subspaces is called **independent** if for any set of k vectors one from each subspace (Say ω_i from $W_i, i \in \{1, 2, \dots, k\}$) the relation $w_1 + w_2 + w_3 + \dots + w_k = 0$ implies each $w_i = 0$
- Given k subspaces of V , say W_1, W_2, \dots, W_k , the subspace W defined as $W = W_1 + W_2 + \dots + W_k$ is said to be a **direct sum** of these k subspaces (denoted as $W = W_1 \oplus W_2 \oplus W_3 \oplus \dots \oplus W_k$ if these k subspaces are independent.

One notes that W could be equal to V and then one would say that the k subspaces $W_i, i \in \{1, 2, \dots, k\}$ of V form a **direct sum decomposition** of V . One would then denote it as $V = W_1 \oplus W_2 \oplus W_3 \oplus \dots \oplus W_k$. Then one can see that an ordered basis of V is obtained by concatenating together an ordered basis each from W_i .

I shall soon show a natural example of a set of subspaces of V connected to a given endomorphism on it which will always be independent but in general will NOT form a direct sum decomposition of the full space V . Such an example comes from the following crucial concept.

The following are equivalent ways of defining a **Characteristic Value** (also known as an “Eigen Value”) for T an endomorphism of a finite dimensional vector-space V :

- c is a characteristic value of T
- The operator $(T - cI)$ is singular (not invertible)
- $\det(T - cI) = 0$

Just by virtue of the above 3 equivalent definitions of an eigenvalue the following observations and concepts follow:

- Linear transformations related by a similarity transformation have the same characteristic values. Being similar is an equivalence relation on the set of all linear transformations and this splits the space into equivalence classes. And the characteristic value gives a multivalued function on each equivalence class. Such structures shall be ubiquitous in Representation Theory and are called **Class Functions**.
- The polynomial defined as $f(x) = \det(T - xI)$ is called the **Characteristic Polynomial** and its roots are precisely the characteristic values.
- If the field of the vector space is not closed the existence of characteristic values isn't guaranteed. One can easily see that the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ doesn't have any characteristic values in \mathbb{R} .
- If c is a characteristic value of an endomorphism A acting on the vector space V then an element $v \in V$ is called a **Characteristic Vector** (or an “Eigen Vector”) for c if $Av = cv$.
- An endomorphism of a vector space V is called **Diagonalizable** if there is a basis of V consisting of eigen vectors. It is easy to see that in that basis the matrix representation of the map will be diagonal.
- One can see that for a characteristic value c of an endomorphism of V , the subset of V consisting of characteristic vectors for c forms a vector subspace of V . The space spanned by all the characteristic

vectors of a given eigen value is called the **Characteristic Space** or **Eigen Space** for that characteristic value. This subspace can also be thought of as the **Null Space** of the operator $T - cI$ (where T is the linear map).

Dimension of the eigen subspace of an eigen value is called the **Geometric Dimension** of that eigen value and its multiplicity in the characteristic polynomial is called its **Algebraic Dimension**.

1. Diagonalizability when algebraic dimension equals geometric dimension

We will realize the criteria of diagonalizability in multiple ways, some of which are conceptually cleaner and some of which are computationally efficient.

Diagonalizability (Version 1)

Let T be an endomorphism of a finite dimensional vector space V and let $\{c_1, c_2, c_3, \dots, c_k\}$ be the set of characteristic values of T and W_i be the null space of $(T - c_iI)$. Then the following are equivalent:

- T is diagonalizable
- The characteristic polynomial of T is of the form

$$f(x) = (x - c_1)^{\dim(W_1)}(x - c_2)^{\dim(W_2)} \dots (x - c_k)^{\dim(W_k)}$$

. Or in other words that for each eigen value its **Algebraic Dimension = Geometric Dimension**.

- $\dim(V) = \dim(W_1) + \dim(W_2) + \dim(W_3) + \dots + \dim(W_k)$

The crucial ingredients that go into the above recipe are:

- Nullity of a diagonal matrix is equal to the number of 0s along its diagonal.
- One sees that the set of null spaces, one for each characteristic value of a given endomorphism of a finite dimensional vector space V are all linearly independent and hence if $W = W_1 + W_2 + \dots + W_k$ then $\dim(W) = \dim(W_1) + \dim(W_2) + \dots + \dim(W_k)$. But in general $\dim(W) < \dim(V)$ and hence in general the null-spaces don't give a direct sum decomposition. But when the third criteria above is true $\dim(W) = \dim(V)$ then the set of null-spaces of the distinct eigen values give a direct sum decomposition of the full space and hence diagonalizability.

An example to see the geometric dimension and algebraic dimension conflict

Consider 2 matrices

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

A simple calculation shows that both of them have the same characteristic polynomial

$$(x - 1)(x - 2)^2$$

. So both of them have eigenvalues 1 and 2.

Now let us look at these operators

$$A - I = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}, A - 2I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$

$$B - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}, B - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}$$

In what follows we shall use the following technique that geometric dimension of an eigen value a of an automorphism T is the nullity of the operator $T - aI$ which can be deduced from the *Rank - Nullity Theorem* which states that **Rank + Nullity of an automorphism = dimension of the vector-space on which it acts**. Further rank is usually easier to see either by just staring at the matrix hard enough or by solving a set of simultaneous equations.

One can observe the following from the above:

- Rank of $A - I = 2$. So for A geometric dimension of the eigen value 1 is 1 (same as its algebraic dimension).
- Rank of $A - 2I = 2$. So for A geometric dimension of the eigen value 2 is 1 (whereas algebraic dimension was 2).
- Rank of $B - I = 2$. So for B geometric dimension of the eigen value 1 is 1 (same as its algebraic dimension).
- Rank of $B - 2I = 2$. So for B geometric dimension of the eigen value 2 is 2 (same as its algebraic dimension).

Hence A is NOT diagonalizable whereas B is.

The similarity transformation by $\begin{bmatrix} 3 & 2 & 2 \\ -1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$ diagonalizes B .

From this example we conclude that whether or not a matrix is diagonalizable is a deeper story than just

the characteristic polynomial. To understand what really matters we have to look at some more subtle polynomials that capture this phenomenon.

Annihilating Polynomial for a linear transformation over a ring is a polynomial p such that $p(T) = 0$ where T is a matrix representation of the linear transformation.

The following concepts follow from the above construction:

- Since the vector-space is finite dimensional say of dimension n , then the space $L(V, V)$ is of dimension n^2 and hence there is an annihilating polynomials of degree $< n^2$. But stronger is true as follows.
- For a fixed linear transformation, set of its annihilating polynomials forms an ideal in the ring of polynomials. If in a ring there is Euclid's algorithm then one knows that there is a unique generator of the ideal which is the monic in the ideal of the lowest degree. This unique monic generator of the ideal of annihilating polynomials is called the **minimal polynomial**
- The characteristic polynomial and the minimal polynomial have exactly the same roots upto multiplicities. This immediately shows that if the linear transformation has all distinct eigen values then its minimal and characteristic polynomial are the same.
- The **Cayley-Hamilton Theorem (weak form)** states that the minimal polynomial divides the characteristic polynomial and hence a linear transformation satisfies its own characteristic polynomial.
- If W is an invariant subspace for a linear transformation T and T_W be the restriction of T to W . Then the characteristic polynomial and the minimal polynomial for T_W divides the characteristic and the minimal polynomial for T respectively.

2. Conductors and Triangulability

If W is an invariant subspace for the linear transformation of T on the vector space V and $v \in V$, then the **T -conductor of v into W** is the set of all polynomials g in the corresponding field such that $g(T)v \in W$.

As for annihilators, the set of conductors also form an ideal and it has a unique monic generator. (By abuse of terminology, henceforth we shall call this unique monic generator as the, “ T -conductor of v into W ” by) Since the minimal polynomial will take everything to 0, one notes that **every T -conductor for a linear transformation T divides its minimal polynomial**. Hence like above if the factorization of the minimal polynomial into irreducibles is known then that highly constraints the form of all the conductors.

The following concepts follow from the above construction:

- Let T be a linear operator acting on a finite dimensional vector space V with an invariant subspace W and let the minimal polynomial of T factorize completely into linear polynomials. Then there exists a $v \in V, v \notin W$ and a characteristic value c of T such that $(T - cI)v \in W$.

The above existence theorem is what shall crucially make the Cyclic Decomposition Theorem produce the Jordan Form.

- Call a linear transformation **Triangulable** if there is a basis in which the matrix is upper/lower triangular. A linear transformation on a finite dimensional vector space is triangulable iff its minimal polynomial is a product of linear polynomials with coefficients in the corresponding field. (Hence any linear transformation over an algebraically closed field like \mathbb{C} is triangulable.)

The first concept is almost a tautology! The central idea in the first concept is that the one can pick any arbitrary vector, say $\beta \in V \setminus W$ and c comes from one of the common factors between the minimal polynomial and the characteristic polynomial (which is guaranteed by the linear decomposition) and then $v = \frac{T\text{-conductor of } \beta}{x-c} \beta$ does the job. And by the minimality of the degree of the minimal polynomial, we have $v \notin W$. To get the second concept one just needs to iterate the first by starting with a W spanned by an eigen-vector of T . Let the eigen vector be v_1 , then one is guaranteed that there is a $v_2 \in V \setminus W$ s.t $(T - cI)v_2 = mv_1$ for some c and m and hence $Tv_2 = mv_1 + cv_2$. In the next iteration choose W as the subspace spanned by v_1 and v_2 . Continuing one generates the required ordered basis. The converse is trivial.

3. Diagonalizability when minimal polynomial has distinct roots

This leads us immediately to the computationally most efficient test of diagonalizability.

Diagonalizability (Version 2)

An endomorphism of a finite dimensional vector space is diagonalizable iff its minimal polynomial factorizes into a product of distinct linear factors with coefficients in the corresponding field.

Stated otherwise it means that if $\{c_1, c_2, \dots, c_k\}$ are the distinct eigenvalues of the linear operator T then T is diagonalizable iff as operators $(T - c_1I)(T - c_2I)\dots(T - c_kI) = 0$.

The central idea in the above proof is to show that if the characteristic vectors can't span the whole space then the minimal polynomial must have repeated roots. If c is an eigen value then there exists a vector α beyond the span of the characteristic vectors, say W , such that $(T - cI)\alpha \in W$. Since the minimal polynomial factors as $(x - c)q$, q being some polynomial then $q(T)\alpha \in W$ since $q(T)\alpha$ is an eigen vector of T with eigen value c . Then look at the polynomial $q(x) - q(c)$ and one can see that since c is a root of it, $(q(T) - q(c))\alpha \in W$ (since W is an invariant subspace of T) and hence $q(c)\alpha \in W$. But since $\alpha \notin W$ by definition, we have $q(c) = 0$ which is in contradiction to the assumption that the minimal polynomial has repeated roots. Hence proved. The converse is trivial.

4. Diagonalizability through projection operators

Projection operators and direct sum decompositions are intimately related to each other in the sense that giving one gives the other. This notion is made precise in the following way,

If $V = W_1 \oplus W_2 \oplus \dots \oplus W_k$, then there exists k linear operators E_1, E_2, \dots, E_k on V such that:

- Each E_i is a projection i.e $E_i^2 = E_i$
- $E_i E_j = 0$ if $i \neq j$
- $I = E_1 + E_2 + \dots + E_k$
- The range of E_i is W_i

Conversely if E_1, E_2, \dots, E_k are linear operators on V which satisfy these above mentioned conditions and if W_i is the range of E_i then $V = W_1 \oplus W_2 \oplus \dots \oplus W_k$.

The first part is trivially satisfied by choosing the E_i s as the projection operator to the W_i s. To prove the converse note that the condition $I = E_1 + E_2 + \dots + E_k$

ensures that $V = W_1 + W_2 + \dots + W_k$. This splits any vector $\alpha \in V$ as $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_k$ and one can then see that $E_j\alpha = \alpha_j$ by using the construction that for any i , $E_i\alpha = E_i\beta_i$ for some $\beta_i \in W_i$ using the surjectivity of E_i on W_i . This proves the uniqueness of the decomposition of α and hence the proof.

The above idea coupled with a concept we had earlier seen, that subspace of eigen-vectors of a given eigenvalue form an invariant subspace of the linear transformation and these subspaces for different eigen values are independent, gives us another way to characterize diagonalizability.

Diagonalizability (Version 3)

Let T be a linear operator on a finite dimensional vector space V . T is diagonalizable if there exists k distinct scalars c_1, c_2, \dots, c_k and k non-zero linear operators E_1, E_2, \dots, E_k such that

- $T = c_1E_1 + c_2E_2 + \dots + c_kE_k$
- $I = E_1 + E_2 + \dots + E_k$
- $E_iE_j = 0, i \neq j$

It would also imply that:

- c_1, c_2, \dots, c_k are the eigenvalues of T
- $E_i^2 = E_i$ (E_i is a projection)
- The range of E_i is the characteristic space for T associated with eigenvalue c_i

Conversely if T is diagonalizable then k non-zero linear operators E_1, E_2, \dots, E_k are guaranteed to exist which satisfy all the six criteria above.

The converse is easy to prove by choosing E_i to be the projection operator into the characteristic space of eigenvalue c_i and by showing that for any $v \in V$, $Tv = c_1E_1v + c_2E_2v + \dots + c_kE_kv$ using the idea that since E_i is the projection into an invariant space of T , it commutes with T i.e. $[T, E_i] = 0 \forall i$.

To prove the diagonalizability condition one notes that $TE_i = c_iE_i$ and hence the range of E_i is the nullspace of $(T - c_iI)$. Thus c_i s are the characteristic values of T and there is no other (which can be shown by using the fact that if there is one say c then $T = cI = \sum_i (c_i - c)E_i$ and the consequence follows by the linear independence of the known eigenspaces).

So we have seen that the range of all the E_i s is the space spanned by all the characteristic vectors of T and that is the whole space since $I = \sum_i E_i$. Hence diagonalizable.

Further we can see that for any $v \in V$, such that $Tv = c_iv$ then it implies $E_iv = 0$.

So the range of E_i is the nullspace of $T - c_iI$.

One can make the following important observations about the projection operators:

- If g is any polynomial over the field F and hence if T is a diagonalizable linear operator as above with E_i s being projection operators into its eigenspaces then

$$g(T) = g(c_1)E_1 + g(c_2)E_2 + \dots + g(c_k)E_k$$

- Thus if T is a diagonalizable operator then $g(T) = 0$ iff all $g(c_i) = 0$ and hence the minimal polynomial is $\prod_i^k (x - c_i)$.
- Conversely if $\prod_i^k (x - c_i)$ is the minimal polynomial for T then consider the set of k **Lagrange Polynomials** defined by these c_i s as:

$$p_j = \prod_{i \neq j} \frac{x - c_i}{c_j - c_i}$$

One then observes that $p_j(c_i) = \delta_{ij}$ and hence if T is diagonalizable then $p_j(T) = E_j$ and hence the projection operators to the characteristic spaces of a diagonalizable operator are polynomials in the operator. (given by the Lagrange Polynomials)

Conversely by defining $E_j = p_j(T)$ one can show that these E_j s satisfy all the criteria's of the third diagonalizability test and hence T is diagonalizable.

This gives an independent proof of the fact that a linear transformation is diagonalizable iff its minimal polynomial factorizes into distinct linear polynomials.

5. *Primary Decomposition Theorem*
and
Jordan-Chevalley Decomposition

We had earlier seen in the two matrices considered which had minimal polynomial being $(x-1)(x-2)^2$ the algebraic and the geometric dimension of the eigenvalues didn't match for the two matrices considered and the characteristic spaces for 1 and 2 were less than the full space. But in those examples one can see that the nullspaces of the operators $(T-2I)^2$ and $(T-I)$ could give a direct sum decomposition of the full space. This idea is captured in the following theorem:

Primary Decomposition Theorem *Let T be a linear operator on a finite dimensional vector space V over the field F and let p be the minimal polynomial for T where*

$$p = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$$

where the p_i are distinct irreducible monic polynomials over F and r_i are positive integers. Let W_i be the nullspace of the operator $p_i^{r_i}$. Then the following holds:

- $V = W_1 \oplus W_2 \oplus \dots \oplus W_k$
- Each W_i is invariant under T_i
- If T_i is the operator induced on W_i by T , then the minimal polynomial for T_i is $p_i^{r_i}$

The block diagonal form for any linear transformation that this theorem guarantees is called the **Rational Form**

This theorem gives a very powerful consequence which is crucial in representation theory. If E_i s denote the projection operators to the W_i s defined above and if $p_i = T - c_i I$ then one can show by construction that $I = E_1 + E_2 + \dots + E_k$ and we can observe the following:

- The operator D defined as $D = c_1 E_1 + c_2 E_2 + \dots + c_k E_k$ along with the E_i s and the c_i s satisfies all the criteria of the projection version of diagonalizability. Hence D is diagonalizable. In this context D is said to be semisimple.
- One can show that operator N defined as $N = T - D$ is nilpotent.
- N and D commute and this decomposition of T is unique (as can be shown by direct evaluation)
- Since we know that the projections are given by polynomials in the original linear transformation (specifically by Lagrange Polynomials), we conclude that the diagonalizable and the nilpotent parts of the operator are also given as polynomials in the operator.

This unique decomposition of any linear operator over an algebraically closed field as a sum of a pair of commuting nilpotent and semisimple operators (which are polynomials in the original operator) is called the **Jordan-Chevalley Decomposition**.

The way these operators are constructed is as follows. Consider the polynomials

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

. One can see that all the f_i s are relatively prime and hence their g.c.d is one and hence there exists polynomials g_i such that $\sum_i f_i g_i = 1$. One can then see that if E_i is defined as

$$E_i = f_i g_i$$

then E_i form a set of projection operators. ($p | E_i E_j$ and hence $E_i(T) E_j(T) v = 0, \forall v \in V$). Further by definition if $v \in \{\text{range of } E_i\}$ then $E_i v = v$ and $p_i(T)^{r_i} v = 0$ by substituting the former into the LHS of the later equation. Further for $i \neq j$, $p_i^{r_i} | E_j$ and hence $E_j v = 0$ if $p_i^{r_i} v = 0$. Since $\sum_i E_i = 1$ (by construction), we have $E_i v = v$. So the range of E_i is equal to the nullspace of $p_i^{r_i}$.

On an algebraically closed field, as a consequence of the Jordan Form it shall become obvious that the dimension of W_i (the nullspace of the operator $p_i^{r_i}$) is the algebraic dimension of the eigenvalue it corresponds to. Stated otherwise, the dimension of the nullspace of $(T - cI)^{(\text{its minimal polynomial multiplicity})}$ is the multiplicity of c in the characteristic polynomial.

Since $p_i(T)^{r_i} = 0$ on W_i we have $p_i(T_i)^{r_i} = 0$. So the minimal polynomial for T_i divides $p_i^{r_i}$. Further if g is any other annihilating polynomial of T_i then $g(T) f_i(T) = 0$ since after the direct sum decomposition is proven one sees that all the projections of a vector along W_j ($j \neq i$) is annihilated by the factor $p_j^{r_j}$ of f_i and the projection along W_i will be killed by g . So trivially $p | g(T) f_i(T)$ and so $p_i^{r_i} | g$. So the minimal polynomial of T_i divides $p_i^{r_i}$ and $p_i^{r_i}$ divides any annihilating polynomial of T_i . So $p_i^{r_i}$ is the minimal polynomial of T_i .

6. *Cyclic vectors, Cyclic subspaces
and
Companion Matrices*

Let T be a linear operator on the vector space V over the field F . In our aim to understand the decomposition of linear transformation further, the following constructions shall be of interest to us now:

- Given a $v \in V$, one shall then be interested in polynomials $g \in F[x]$ such that $g(T)v = 0$. This is an ideal in the ring of polynomials and it is non-zero since it contains the minimal polynomial.
- Given a $v \in V$, one shall then be interested in the smallest T -invariant subspace of V which contains v . This subspace is also the intersection of all T -invariant subspaces containing v . This space is also the space spanned by all vectors of the form $g(T)v, \forall g \in F[x]$.
- Given a T -invariant subspace, W of V one asks the question as to whether there is another T -invariant subspace W' such that $V = W \oplus W'$.

This leads us to define the following two concepts:

- If $v \in V$ then the **T-cyclic subspace generated by v** is the subspace of all vectors of the form $g(T)v, \forall g \in F[x]$ denoted as $Z(v, T)$.
- If $Z(v, T) = V$ then v is called a **cyclic vector for T**.
- If $v \in V$ then the **T-annihilator of v** is the ideal in the ring of polynomials $F[x]$ generated by $g \in F[x]$ such that $g(T)v = 0$. This ideal is denoted as $M(v, T)$. The unique monic generator of this ideal will also be called the *T-annihilator of v* (by abuse of notation!)
- A T -invariant subspace W of V is **T-admissible** if for every $v \in V$ and $f \in F[x]$ there exists a $w \in W$ such that $f(T)v = f(T)w$.

One notes the following:

- The T-cyclic subspace generated by 0 is 0.
- $Z(v, T)$ is 1-dimensional iff v is an eigen vector of T .
- For the identity operator every non-zero vector generates a cyclic subspace. So for $\dim(V) > 1$, identity operator has no cyclic vector.
- If a T -invariant subspace has a complementary T -invariant subspace then it is also T -admissible.

Let p_v denote the unique monic generator of $M(v, T)$. Then one notes the following:

- $\deg(p_v) = \dim(Z(v, T))$
- If $\deg(p_v) = k$ then the vectors $v, Tv, T^2v, \dots, T^{k-1}v$ forms a basis for $Z(v, T)$.
- The minimal polynomial for $T|_{Z(v, T)}$ is p_v .

The above is seen by noting that the remainder obtained by dividing any polynomial by p_v has to be of degree less than $\deg(p_v)$ and hence has to be a linear sum of $v, Tv, T^2v, \dots, T^{k-1}v$ which are linearly independent since any non-trivial relation between them will contradict the definition of p_v as the minimal polynomial.

The last assertion is established by noting that for any $g \in F[x]$, $p_v(T|_{Z(v, T)})g(T)v = 0$ and hence $p_v(T|_{Z(v, T)})$ has $Z(v, T)$ as its kernel. And by definition there cannot be any polynomial of lower degree annihilating $Z(v, T)$ since that will contradict the definition of p_v .

From the above assertions one can see that if there is a cyclic vector then the minimal polynomial and the characteristic polynomial are the same.

Let us for now look at a linear operator L on a vector space W of dimension k which has $w \in W$ as its cyclic vector. Then the set $\{w, Lw, L^2w, \dots, L^{k-1}w\}$ forms a basis for W . Consider the vectors defined as $w_i = L^{i-1}w$ then in the basis $\{w_1, w_2, \dots, w_k\}$ L looks like

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

where the monic polynomial p_w is $c_0 + c_1x + \dots + c_{k-1}x^{k-1}$. And this matrix is called the **Companion Matrix** for this monic polynomial.

One notes the following about the Companion Matrix

- If T is a linear operator on a vector space V then T has a cyclic vector iff there is a basis of V in which T looks like the companion matrix of its minimal polynomial.
- Minimal and the characteristic polynomial of a companion matrix is the same monic polynomial from which it came.

These companion matrices in the cyclic subspaces shall be the building blocks from which we shall try to build the full linear transformations.

7. The idea behind cyclic decomposition (Part 1)

The argument for Cyclic Decomposition Theorem needs synthesis of many concepts from linear algebra and hence we shall give the constituent ideas in instalments in an effort to make it more palatable.

Basically, if T is a linear operator on a finite dimensional vector space V then we can show that there are vectors $\{v_1, v_2, \dots, v_r\}$ such that

$$V = Z(v_1, T) \oplus Z(v_2, T) \oplus \dots \oplus Z(v_r, T)$$

Say one has found vectors $\{v_1, v_2, \dots, v_j\}$ inductively and the subspace

$$W_j = Z(v_1, T) \oplus Z(v_2, T) \oplus \dots \oplus Z(v_j, T)$$

is proper. One now needs to see that all this ensures that there is another vector $v_{j+1} \in V$ such that $W_j \cap Z(v_{j+1}, T) = \{0\}$.

Take a vector $v \in V, v \notin W_j$. Then let $f \in F[x]$ be the T -conductor of v into W_j and **IF** W_j is T -admissible then there is a $w \in W_j$ s.t $f(T)v = f(T)w$. Let $x = v - w$. Then since $v - x \in W_j$, and by definition W_j is T -invariant, $g(T)v \in W_j$ iff $g(T)w \in W_j$ for some $g \in F[x]$, i.e if the the T -conductor into W_j of both x and v match i.e if f is the T -conductor of x into W_j .

But $f(T)x = 0$, f being the T -conductor of x into W by the above argument and if $g(T)v \in W_j$ for any $g \in F[x]$, then by definition $f|g$ and hence $g(T)v \in W_j$ iff $g(T)v = 0$.

Since by definition $Z(v, T)$ is the space of all $g(T)v$ for arbitrary $g \in F[x]$, we see that

$$W_j \cap Z(v_{j+1}, T) = \{0\}$$

So the thing to ensure for the induction to continue is that W_j is T -admissible.

8. Cyclic Decomposition Theorem and Jordan Form

The idea in the above section of decomposition of a vector space into direct sum of cyclic subspaces is made precise in the following sense:

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 . Then there exists a set of vectors $\{v_1, v_2, \dots, v_r\}$ in V with respective T -annihilators p_1, p_2, \dots, p_r such that

- $V = W_0 \oplus Z(v_1, T) \oplus Z(v_2, T) \oplus Z(v_3, T) \dots \oplus Z(v_r, T)$
- $p_k | p_{k-1}, k \in \{2, 3, \dots, r\}$

The integer r and the annihilators p_1, p_2, \dots, p_r are uniquely determined by the above 2 conditions and the fact that $v_k \neq 0, \forall k$.

(Note that one can always choose W_0 as the null space since it is trivially always an admissible space further recall that earlier it had been shown that $\deg(p_i) = \dim(Z(v_i, T))$)

Some of the consequences of existence of the above cyclic decomposition are:

- Given a linear operator T on a vector space V , every T -admissible subspace of it has a complementary T -invariant subspace.

The direct sum of the cyclic spaces given by the above decomposition gives the required complementary space

- A linear transformation has a cyclic vector iff its minimal polynomial and the characteristic polynomial match.

One was shown earlier by the companion matrix. The other way is obvious since if minimal and characteristic polynomial match then the size of the first cyclic space will exhaust the dimension

- A priori, it may be the case that the minimal polynomial for the whole vector space is some polynomial but each specific vector has a smaller minimal polynomial (and the lcm of these smaller minimal polynomials is the minimal polynomial over the whole space).

But one sees that there is a vector whose minimal polynomial is the minimal polynomial over the whole space. From the next part of the explanation of this theorem it shall be clear that if one chooses W_0 as the null space then this vector is v_i

- **Cayley-Hamilton Theorem (strong form)** is a consequence of this theorem which states that if a linear operator T on a vector space V has minimal polynomial p and characteristic polynomial f then,

- $p|f$
- p and f have the same prime factors except for multiplicities.
- if the prime factorization of p is $p = \prod_1^k f_i^{k_i}$ then $f = \prod_1^k f_i^{d_i}$ where d_i is the nullity of $f_i(T)^{k_i}$ divided by $\text{deg}(f_i)$.

The central idea is that p_i is the minimal polynomial for $T|_{Z(v_i, T)}$ and since $Z(v_i, T)$ is a cyclic space p_i is also the characteristic polynomial for this restricted operator and hence by the block diagonal form given by the cyclic decomposition $f = \prod_1^r p_i$ and if W_0 is chosen as the null space then $p_1 = p$ and hence the first part is shown and by the fact that $p_i|p_{i-1}$ we get the next part. From the Primary Decomposition we know that if T_i is the null space of $f_i(T)^{k_i}$ then $f_i^{k_i}$ is the minimal polynomial for T_i . So by the fact just now proved that prime factors arising in the factorization of minimal and the characteristic polynomials are the same we see that the characteristic polynomial for T_i would be $f_i^{d_i}$ with some $d_i \geq r_i$. Since the degree of a characteristic polynomial is the dimensionality of the vector space we automatically have $d_i = \frac{\text{dim}(V_i)}{\text{deg}(f_i)}$. And further by the direct sum structure it follows that $f = \prod_1^k f_i^{d_i}$

- **The Jordan Form** is obtainable over an algebraically closed field by doing the Cyclic Decomposition of the induced operator in every subspace given by the Primary Decomposition.

Before understanding how the proof of the above big result works let us try to understand how the above leads to the Jordan Form.

Suppose the characteristic polynomial f of a linear operator T on a vector space V over an algebraically closed field F factors as:

$$f = \prod_1^k (x - c_i)^{d_i}$$

and hence the c_i s are its distinct eigenvalues and $d_i > 1 \forall i$ and hence its minimal polynomial p has to be of the form

$$p = \prod_1^k (x - c_i)^{r_i}$$

with $1 \leq r_i \leq d_i$. If W_i is the nullspace of $(T - c_i I)^{r_i}$ then the Primary Decomposition theorem tells us that

$V = \bigoplus_1^k W_i$ and the operator T_i induced on W_i has as its minimal polynomial $(x - c_i)^{r_i}$.

Hence the operator N_i on W_i defined as $N_i = T_i - c_i I$ is nilpotent. So now we want to do a Cyclic Decomposition of W_i with respect to the nilpotent operator $T_i - c_i I$ on it.

Hence we need to see how Cyclic Decomposition works for any nilpotent operator say N on some finite dimensional vector space V .

Cyclic Decomposition Theorem tells guarantees us the existence of r non-zero vectors $\{v_1, v_2, \dots, v_r\}$ with N -annihilators $\{p_1, p_2, \dots, p_r\}$ such that

$$V = \bigoplus_1^r Z(v_i, N)$$

and $p_{i+1}|p_i$. Since N is nilpotent its minimal polynomial is x^k for some $k \leq n$ and hence each $p_i = x^{k_i}$ where $k_1 = k$ and $k_r \geq 1$ and $k_{i+1} \leq k_i$.

Now the idea of **Companion Matrix** guarantees that there exists a basis in the subspace $Z(v_i, N)$ in which the induced operator is represented by a $k_i \times k_i$ matrix A_i of the form:

$$A_i = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

So the Cyclic decomposition theorem says that a nilpotent operator on the space V has an ordered basis in which

$$N = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & A_r \end{bmatrix}$$

where A_i is a $k_i \times k_i$ companion matrix of the type explained above.

Hence going back to T_i , since N_i can be written as above, we see that T_i can be written as a direct sum of matrices of the type:

$$\begin{bmatrix} c_i & 0 & \dots & 0 & 0 \\ 1 & c_i & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ & & & c_i & 0 \\ 0 & 0 & \dots & 1 & c_i \end{bmatrix}$$

The above kind of matrices are called **Elementary Jordan Block of eigen value c_i** .

By abuse of notation calling T the matrix representation of the operator T and similarly for T_i in the basis just generated we see that the final form looks like:

$$T = \begin{bmatrix} T_1 & 0 & \dots & 0 \\ 0 & T_2 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & T_k \end{bmatrix}$$

where

$$T_i = \begin{bmatrix} J_{i1} & 0 & \dots & 0 \\ 0 & J_{i2} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & J_{in_i} \end{bmatrix}$$

where J_{im} is the m^{th} elementary Jordan block for the eigen value c_i (corresponding to the block for T_i) and n_i being the number of cyclic spaces into which the nullspace of $(x - c_i I)^{r_i}$ splits.

The above is called the **Jordan Form** for the linear operator.

From the Jordan form three crucial observations immediately follow:

- One can see that the dimension of the null space of $(T - c_i I)^{r_i}$ is d_i i.e $\dim(W_i) = d_i$ (where d_i is the algebraic multiplicity of the eigenvalue c_i and r_i is its geometric multiplicity). Hence $\dim(T_i) = d_i$.
- n_i is the geometric multiplicity for the eigenvalue c_i since in the Cyclic Decomposition of N_i each cyclic subspace $Z(v_m, N_i)$ gives one vector i.e $N_i^{k_m}$ which is non-zero and is in the null space of the operator $N_i = T_i - c_i I$. Where $k_m = \dim(Z(v_m, N_i))$.

This again shows the diagonalizability criteria that a linear transformation is diagonalizable iff $n_i = d_i, \forall i$ where n_i and d_i have been just now argued to be the geometric and the algebraic multiplicities respectively of the eigenvalue c_i

- Dimension of the Elementary Jordan Block J_{i1} for all i is the multiplicity of the eigenvalue c_i in the minimal polynomial.

9. The idea behind cyclic decomposition (Part 2)

Starting with the T -admissible space W_0 one looks for a vector w_1 such that if $p_1 = s(w_1, W_0)$ (the conductor of w_1 into W_0) then $p_1 = \max_{w \in V} \text{degs}(w, W_0)$ and the corresponding w is taken as w_1 .

One continues this induction so that after k steps one has $W_k = W_0 + \sum_{i=1}^k Z(w_i, T)$ and polynomials p_1, p_2, \dots, p_k such that $w_k \in V, w_k \notin W_{k-1}$ and among all T -conductors into W_{k-1} it has the maximal degree conductor p_k .

So we see that if $w \in W$ and $f \in s(w, W_{k-1})$ then $fw = w_0 + \sum_{1 \leq i \leq k-1} g_i w_i$ where g_i are some polynomials and $w_i \in W_i$. One can then show that $f|g_i$ and $w_0 = fz_0$ for some $z_0 \in W_0$. (call this the "Divisibility Claim")

After k steps of the induction call the w of above as w_k and f as p_k and we have for some set of polynomials say h_i and with z_0 as above, the relation

$$p_k w_k = p_k z_0 + \sum_{1 \leq i \leq k-1} p_k h_i w_i$$

and define

$$v_k = w_k - z_0 - \sum_{1 \leq i \leq k-1} h_i w_i$$

Since $w_k - v_k \in W_{k-1}$ we have $s(v_k, W_{k-1}) = s(w_k, W_{k-1}) = p_k$ and since $p_k v_k = 0$ we have

$$W_{k-1} \cap Z(v_k, T) = \{0\}$$

And we have the construction that

$$W_k = W_0 \oplus Z(v_1, T) \oplus Z(v_2, T) \oplus \dots \oplus Z(v_k, T)$$

and we have the trivial relation

$$p_k v_k = 0 + p_1 v_1 + p_2 v_2 + \dots + p_{k-1} v_{k-1}$$

on which applying the Divisibility Claim we have that

$$p_i | p_{i-1}$$

After getting W_{k-1} one searches for the vector w_k in the rest of the space which has a conductor into W_{k-1} of maximal degree to construct $W_k = W_{k-1} + Z(v_k, T)$. And $\dim(W_k) > \dim(W_{k-1})$ and hence this induction will end after atmost $\dim(V)$ steps.

10. *The idea behind cyclic decomposition (Part 3)*

Let us now try to understand the basic idea behind how the divisibility claim works, now that we have a hang of how the cyclic decomposition works as to how to inductively search for the cyclic vectors which will exhaust the full space by their cyclic subspaces.

The argument is initially almost the same as above.

Let $g_i = fh_i + r_i$ and $r_i = 0$ or $\deg(r_i) < \deg(f)$ and define $z = w - \sum_1^{k-1} h_i w_i$. Since $z - w \in W_{k-1}$ we have $fz = w_0 + \sum_1^{k-1} r_i w_i$ and let j be the largest i for which $r_i \neq 0$. Since $W_{j-1} \subset W_{k-1}$ we have that there exists a polynomial g such that $p = gf$ where $p = s(z, W_{j-1})$ and $f = s(z, W_{k-1})$. Then we have

$$pz = gfsz = gr_j w_j + gw_0 + \sum_{1 \leq i \leq (j-1)} gr_i w_i$$

Since $pz \in W_{j-1}$ it implies that $gr_j w_j \in W_{j-1}$.

Now we remember that the degree of the monic generator of an ideal is the polynomial of the least degree in that ideal and also that the $p_i s$ were chosen to be the maximum degree polynomials among all the conductors into the respective $W_i s$ and hence combining these two we have the inequality

$$\deg(gr_j) \geq \deg(s(w_j, W_{j-1})) = p_j \geq \deg(s(z, W_{j-1})) = \deg(p)$$

and

$$\deg(p) = \deg(fg)$$

Hence we have $\deg(r_j) \geq \deg(f)$ which is absurd by the definition of j and hence we have that all the $r_i = 0$. This also shows that if

$$g_i = fh_i$$

and

$$z = \sum_1^{k-1} h_i w_i$$

then

$$w_0 = fz$$

but since W_0 is by definition an admissible space we have some z_0 such that $w_0 = fz_0$ and hence we have for any $w \in V$ and $f = s(w, W_{k-1})$,

$$f(T)w = f(T)(z_0 + \sum_1^{k-1} h_i w_i)$$

and hence it shows that **at every step of the induction each W_i is a T -admissible space!**

This completes the proof of the Cyclic Decomposition Theorem.

11. *The larger picture*

One notes that for the Cyclic Decomposition to work we crucially needed two things

- That the ideals of conductors are principal coupled with the non-existence of 0-divisors.
- The uniqueness coming from the existence of a notion of unique prime factorization.

The first property was coming from the fact that we were working in the polynomial ring where the Euclid's division algorithm gives the monic generators of the ideals. But in general we can carry over all that for any Principal Integral Domain (PID) (not necessarily a Euclidean Domain) which also are Unique Factorization Domains and since we never needed any specific property of vector spaces for this we can do the same Cyclic Decomposition on any module over a PID.

This gives us the "Structure theorem for modules over PID" and since any abelian group is a \mathbb{Z} -module where \mathbb{Z} is a PID we can have the "Structure Theorem for Abelian Groups" which is conventionally stated as

Every finitely generated abelian group is a direct sum of cyclic groups of prime power order and of a free abelian group.
or

If G is a finitely generated group then $V = L \oplus C_{d_1} \oplus C_{d_2} \oplus \dots \oplus C_{d_k}$ where C_i is a cyclic group of order i and $d_i > 1$ and $d_1 | d_2 | \dots | d_k$. One can identify any C_n to \mathbb{Z}_n .

II. ACKNOWLEDGEMENT

Interspersed in this document are influences arising from the few hundred emails exchanged over the last 4 years with Vipul (formerly my college-mate at CMI (Chennai Mathematical Institute) and currently a mathematics graduate student at UChicago) . A large part of the core content is from the algebra book by Hoffman and Kunze and there is the undeniable influence of the algebra books by Herstein and Artin on me over the last 4 years, Herstein being my initiation into the subject. Some of the contents in the second appendix of this article are from this breathtaking repository of group theory created by Vipul http://groupprops.subwiki.org/wiki/Main_Page