

Advanced Linear Algebra

"...can there be no sorrow, be no pain!?"

November 26, 2022

Abstract

These are the notes of Linear Algebra 2 that ran during Autumn of 2022 at IISER Kolkata. There were two parts of the course. The first covered rational and Jordan forms and the second (quite quickly, I must add) covered the basic results in the theory of inner product spaces. One can quickly summarize the whole course in two sentences as follows. In the part 1, we wish to classify all linear operators on a finite dimensional vector space upto similarity (Theorem 3.2.9 & Theorem 3.3.4). In the part 2, we wish to see how diagonalizability of an operator over a finite dimensional IPS is characterized by the inner product structure (Theorem 6.0.6) and the fact that operators on a \mathbb{C} -IPS behaves quite a lot like complex numbers (Theorem 7.2.1). Part 1 remains quite detailed, except few results, but the same cannot be said for Part 2, which very often cites outside sources and gives proofs which, at times, can become very sketchy.

Contents

I	Classification of linear operators on a finite dimensional vector space upto similarity	3
1	Overview of basic linear algebra	4
1.1	Linear equations	4
1.2	Something about vector spaces	5
1.3	Rows-spaces and rank	6
2	Elementary canonical forms	7
2.1	Annihilating polynomials	10
2.2	Invariant subspaces	13
2.3	Characterization of diagonal operators in terms of minimal polynomial	14
2.4	Characterization of triangulable operators in terms of minimal polynomial	17
2.5	Simultaneous diagonalizability and triangulability	19
2.6	Direct sum decomposition	22
2.6.1	Projection operators	22
2.7	Invariant direct sum decomposition	23
2.8	Primary decomposition	26

3	Rational & Jordan forms	29
3.1	Cyclic subspaces and cyclic vectors	29
3.1.1	Companion matrix of a monic polynomial	31
3.2	Cyclic decomposition theorem	32
3.2.1	Corollaries of cyclic decomposition	33
3.2.2	Cyclic decomposition in terms of matrices	35
3.3	Jordan form	37
II	Inner product spaces	41
4	Inner products, orthonormal basis, functionals, adjoints and uniqueness	42
4.1	Linear functionals	43
4.2	Adjoint operators	43
4.2.1	Properties of adjoints on \mathbb{C} -IPS	44
4.3	Normal & unitary operators - 1	45
5	Best approximation and self-adjoint operators	46
5.1	Best approximation	46
6	Spectral theorem for normal operators	48
6.1	Unitary operators - 2	51
7	Polar decomposition	52
7.1	Square root of positive operators	52
7.2	Polar decomposition of complex operators	53

Part I

Classification of linear operators on a finite dimensional vector space upto similarity

Lecture # 1, 08/08/2022

This is a course on advanced linear algebra. In one sentence, the principal aim of everything that we do here will be to "classify all linear operators on a vector space". Of-course this is intractable, but in certain cases there is a very rich theory. We have already seen one example of this in the case of diagonalizable operators, we shall see some other types of similar nice operators.

Of-course we are assuming background in basic algebra, but we will try to give a rather terse look at the basic linear algebra and all the theorems therein, just as a refresher.

1 Overview of basic linear algebra

Whatever we will do in the following will be a terse overview of basic linear algebra but with all of its spectacular theorems present. Keeping in mind the general philosophy of linear algebra as to characterize all linear maps, we will see one such attempt that you have learned in the previous course.

1.1 Linear equations

For a vector space V over field F , a linear operator is a linear map $T : V \rightarrow V$. One thing that we humans like to do is to ask *what all elements of a space satisfies some collection of conditions?* When the conditions are *equational* in nature and the space is, say \mathbb{R}^2 , then we get subspaces of \mathbb{R}^2 , which may or may not be linear (depending on conditions) in the sense that sum of two elements may not be in that subspace. But now suppose that the defining conditions are linear, then we get groovy, for we can write those conditions as very simple looking equations of the form $a_1x_1 + \dots + a_nx_n = 0$, where $a_i \in F$; of-course there may be more than one such equational conditions. Now suppose we are working in a finite dimensional real vector space V , which of-course is thus isomorphic to $\mathbb{R}^{\dim V}$. We want to do the same thing and want to find all those subspaces of V which satisfy some equational linear conditions. Now, we can collect all the coefficients of linear conditions into a nice matrix A of shape equal to *number of total conditions* \times *number of total variables*. Of-course, number of total variables will be $n := \dim V$. So if total number of conditions are m , then A is an $m \times n$ matrix. Now any element of $V \cong \mathbb{R}^{\dim V}$ is simply written as $x = (x_1, x_2, \dots, x_n)$, so if x satisfies all the aforementioned linear conditions, then $Ax = 0$. Of-course, any linear transformation $T : V \rightarrow W$ has a corresponding matrix representation given by images of canonical basis vectors; the matrix corresponding to T is $A := [Te_1, \dots, Te_n]$. In particular

$$\{\text{All linear transformations } V \rightarrow W\} \cong \{\text{All matrices of shape } \dim W \times \dim V\}.$$

Finally, remember that there are three elementary row operations that one can perform on a matrix A of shape $m \times n$. 1) multiply a row of A by $c \in F$, 2) Take a row A_i and replace it with $A_i + cA_j$ for some other row A_j and $c \in F$ and 3) swap rows A_i and A_j with each other.

Let us now bombard you with results that you know already:

Theorem 1.1.1. Let V be a vector space over F . Let A and B be an $m \times n$ matrix:

1. If A and B are row-equivalent, then the systems $AX = 0$ and $BX = 0$ have same solutions.
2. Every matrix A of shape $m \times n$ is row-equivalent to a row-reduced echelon matrix.
3. ★ If A is an $m \times n$ matrix such that $m < n$, then there exists a non-trivial solution of $AX = 0$.

4. ★ If A is a matrix of shape $n \times n$, then A is row-equivalent to I_n if and only if $AX = 0$ has only trivial solution.
5. If A is an $n \times n$ matrix, then TFAE:
 - (a) A is invertible,
 - (b) row-echelon form of A is identity,
 - (c) A is a product of elementary matrices.
6. ★ If A is $n \times n$ matrix, the TFAE:
 - (a) A is invertible,
 - (b) (*injectivity*) $AX = 0$ has only trivial solution $X = 0$,
 - (c) (*surjectivity*) for all $Y \in \mathbb{R}^{\dim V} = V$, there exists $X \in \mathbb{R}^{\dim V} = V$ such that $AX = Y$.

That is, the matrix $A_{n \times n}$ is invertible iff the linear transformation $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is injective iff $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is surjective.

Proof. Let us only sketch the main ideas of the proof here:

1. Since $A = E_1 \dots E_k B$, so $AX = 0$ if and only if $E_1 \dots E_k BX = 0$ if and only if $BX = 0$.
2. You know that every matrix is row-equivalent to row-reduced matrix. Then getting to echelon form of that row-reduced matrix is just a matter of few more elementary row operations.
3. Let R be the row-echelon form of A . Then $R = E_1 \dots E_k A$. So R and A are row-equivalent. By 1. above, $RX = 0$ and $AX = 0$ have same set of solutions. Now, since $m < n$, so matrix R will have a non-zero row, and thus a non-zero solution, since in row-echelon form, the solutions are determined only by the pivot variables, rest all are assigned values as per choice. In particular, the dimension of the kernel of A is equal to the number of pivot variables which is equal to the number of non-zero rows of R .
4. $L \implies R$ is clear, but for $R \implies L$, we first observe that the row-echelon form of R also has to have trivial solution for $RX = 0$. Now, assume that R is not identity, then since R is an $n \times n$ matrix, then there will be a zero row of R , which, by virtue of 3. above tells us that there exists a non-trivial solution of $RX = 0$, a contradiction.
5. Since $A = E_1 \dots E_k R$, so A is invertible if and only if R is invertible if and only if R is identity; the last one follows from the fact that a row-echelon matrix will be invertible if and only if it is identity (think about it, how can R be invertible? only when all rows are non-zero).
6. a implies b follows from 5. For b implies c, we proceed as follows. Define a new matrix \tilde{A} of shape $n \times (n + 1)$ where extra column at end is the vector Y . Similarly define \tilde{X} to be the vector of length $n + 1$ which is obtained by appending the last entry of X by -1. Then, since \tilde{A} has more rows than columns, a use of 3. above tells us that $\tilde{A}\tilde{X} = 0$ has a non-trivial solution. But $\tilde{A}\tilde{X} = AX - Y$. Note that if \tilde{X} is non-trivial, then it clearly shouldn't be all zeros and then -1 because if that is a solution for $\tilde{A}\tilde{X} = 0$, then we will get that Y is zero, which we are not assuming in this case. Finally, c implies a follows from the observation that if A is not invertible, then AX always will have an element of the vector as 0 (follows from R having a zero row if A not invertible). This contradicts the surjectivity and hence A is invertible.

□

1.2 Something about vector spaces

So that's about the basics of matrix manipulation that you may already know. Now let me assume you know finite-dimensional vector spaces well and good. I will give one result, which I think not

a lot of you may know (well I certainly didn't before writing these notes !):

Lemma 1.2.1. Let V be an F -vector space and $W_1, W_2 \subseteq V$ be two finite-dimensional subspaces of V . Then, we have the following relation between the dimensions:

$$\dim W_1 + \dim W_2 = \dim(W_1 \cap W_2) + \dim(W_1 + W_2).$$

Proof. The idea is to simply divide the basis vectors of W_1 and W_2 into three parts: first one which is purely only in W_1 , other which is in $W_1 \cap W_2$ and is a basis for it and the third which is purely only in W_2 . Denote the cardinality of these three sets as j, k and l respectively. Clearly, $W_1 + W_2$ has a basis constituted by the sets of all first, second and third, that is, $\dim(W_1 + W_2) = j + k + l = j + k + k + l - k = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$. This is what we needed. \square

1.3 Rows-spaces and rank

Let $A_{m \times n}$ be a matrix over F . Then, one can view A as a linear transformation:

$$A : F^n \rightarrow F^m.$$

The row-space of A is simply the subspace of F^n generated by the m rows of A :

$$\text{Row}(A) := \left\{ \sum_{i=1}^m c_i A_i \mid c_i \in F \forall i = 1, \dots, m \right\}.$$

There are some results to keep in mind:

Proposition 1.3.1. We have the following results.

1. If $A_{m \times n}$ and $B_{m \times n}$ are two matrices of same shape such that they are row-equivalent, then A and B have same row-spaces:

$$A = EB, E_{m \times m} \text{ is invertible} \implies \text{Row}(A) = \text{Row}(B) \subseteq F^n.$$

2. If $A_{m \times n}$ is a matrix over F and R is the row-echelon form of A , then the non-zero rows of R forms a basis for $\text{Row}(A) \subseteq F^n$.
3. Let $W \subseteq F^n$ be a subspace such that $\dim W \leq m$. Then there exists a unique row-echelon matrix $R_{m \times n}$ such that $\text{Row}(R) = W$.
4. A diagonalizable matrix is always similar to a unique diagonal matrix, upto permutations of the diagonal entry.

Proof. As usual, these are brief sketches, not fully dressed proofs:

1. $A = EB$ implies rows of A are linear combination of rows of B , so $\text{Row}(A) \subseteq \text{Row}(B)$. Since E is invertible, we can thus write $(E)^{-1}A = B$.
2. Linear independence of non-zero rows of R is obvious by the very structure of the rows in R . We need only see whether they span the whole of $\text{Row}(A)$. Since we have $A = E_1 \dots E_k R$, that tells us that any row of A is a linear combination of rows of R , so if we take any $x \in \text{Row}(A)$, then $x = \sum_{i=1}^m c_i A_i$, but since A_i is itself a linear combination of rows of R , so we are done.

3. There exists such a matrix which is easy to see; simply take a basis of W which will consist of less than m vectors, and construct a matrix $A_{m \times n}$ whose rows are these vectors and rest all 0. Now, row-echelon form of this matrix will have row space equal to W by 1. However, to show uniqueness, we need to work a bit. Take any row-echelon matrix $R_{m \times n}$ such that its row space is W . Let the first r rows of it be non-zero. Then for any $x \in W \subseteq F^n$, we can write $x = \sum_{i=1}^r c_i R_i$. Now, we need to capture how exactly is it that $x \in W$ remains independent of choice of R_i s? Well, we see that the $x = \sum_{i=1}^r c_i R_i$ as we have taken above will be such that component will be exactly c_i . Now if $x = (x_1, \dots, x_n)$, then $x = \sum_{i=1}^r x_{k_i} R_i$ where k_i is the pivot index of i^{th} row of R . So, to determine an element x of W , where we are given a basis of W using R ($\text{Row}(R) = W$), we need only specify the coordinates of x at $k_1, \dots, k_r \leq n$ which are the pivot points of R . So, to conclude, take any $x \in W$, then x has first non-zero coordinate at index k_i for some $i = 1, \dots, r$, which is easy to see as explained above. But the same is true for the standard basis
4. Trivial.

□

Lecture # 2, 12/08/2022

2 Elementary canonical forms

One of our main aims in this course is the following: *Given a finite dimensional vector space V and a linear operator T on V , how can we write T in the simplest form possible, or what properties of T makes it easy representable by matrices?* Since we can represent linear operators on a finite dimensional vector space by finite matrices given an ordered choice of a basis, so the question reduces to: *For a linear operator T on a finite dimensional vector space V , does there exists a choice of basis vectors \mathcal{B} in V so to make the matrix $[T]_{\mathcal{B}}$ easy to analyze, maybe perhaps even a diagonal matrix or something close to it?*

As we shall soon see, one of the closest ally in our this task will be the eigenvalues of the operator T . We first recall the following basic theorem:

Theorem 2.0.1. Let T be a linear operator on a finite dimensional F -vector space V . Then, TFAE:

1. $\lambda \in F$ is an eigenvalue of T ,
2. the linear operator $T - \lambda I : V \rightarrow V$ is non-invertible,
3. the $\det(T - \lambda I) = 0$.

Since $\lambda \in F$ is an eigenvalue, so $Tv = \lambda v$ for some non-zero $v \in V$. Hence $(T - \lambda I)v = 0$, that is, $0 \neq v \in \text{Ker}(T - \lambda I)$. Hence $\text{Ker}(T - \lambda I)$ is non-trivial and by Theorem 1.1.1, 6, $T - \lambda I$ has to be invertible. By the same theorem, $\det(T - \lambda I) = 0$.

Motivated from above, we form the **characteristic polynomial** as $\det(A - xI) = 0$, which will be of degree $n = \dim V$ where A is $n \times n$ matrix. Now as can easily be seen, characteristic polynomial of T is independent of the basis chosen, because if \mathcal{B} and \mathcal{B}' are two bases for V , then we know that $[T]_{\mathcal{B}} = P^{-1}[T]_{\mathcal{B}'}P$ for some invertible matrix P . So we are good on the well-definedness side of the things about characteristic polynomials. You should be clear where do you want your eigenvalues from; an eigenvalue is a scalar $\lambda \in F$ so that there exists $v \in V$ (which is a F -vector space) such

that $Tv = \lambda v$. If there are no such λ , then there is no-eigenvalue of T . Equivalently, if there are no roots of characteristic polynomial IN F , then there are no eigenvalues of $T : V \rightarrow V$ where V is a fixed to be an F -vector space.

The **eigenspace** of T for eigenvalue λ is defined to be the following subspace of V :

$$\text{Ker}(T - \lambda I) \subset V.$$

In other words, the eigenspace of T for eigenvalue λ is the set of all eigenvectors of λ . If we are given that V is finite dimensional and $A = [T]_{\mathcal{B}}$, then we can analyze the operator $A - \lambda I$ in order to find, say dimension of eigenspace of λ (which you will do by using rank-nullity and by finding the rank of $A - \lambda I$) and what not.

We next study **diagonalizable operators**. An operator $T : V \rightarrow V$ on a finite dimensional F -vector space is said to be diagonalizable if there exists an ordered basis \mathcal{B} of V such that each vector $b \in \mathcal{B}$ is an eigenvector of T . The name comes from the fact that $[T]_{\mathcal{B}}$ will look as follows:

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c_n \end{bmatrix}$$

Let T be a diagonalizable operator on finite-dimensional vector space V . There is a subtle relation between dimension of the eigenspace of $A = [T]_{\mathcal{B}}$ corresponding to eigenvalue λ (geometric multiplicity) and the multiplicity of eigenvalue λ as the root of characteristic polynomial (algebraic multiplicity). In particular,

Lemma 2.0.2. A linear operator T is diagonalizable if and only if for each eigenvalue $\lambda \in F$ of T , the algebraic multiplicity of λ is equal to geometric multiplicity of λ . That is,

$\forall \lambda$ eigenvalue of T , multiplicity of λ in $\Phi(x)$, the characteristic polynomial of $T = \dim \text{Ker}(T - \lambda I)$.

Proof. (L \implies R) Since T is diagonalizable, therefore there exists a basis \mathcal{B} of V such that each vector $b \in \mathcal{B}$ is an eigenvector of T . Now, we know that $A = [T]_{\mathcal{B}}$ will look like as above, where we have to assume that c_i is repeated d_i many times. So characteristic polynomial $\Phi(x)$ of T will look like $\Phi(x) = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$. So if we take $\lambda = c_1$ WLOG, then algebraic multiplicity is equal to d_1 . Next, we will have that $A - c_1 I$ looks like a diagonal matrix whose first d_1 entries are zero and rest diagonal entries are $c_i - c_1$. Clearly, $\text{Ker}(A - c_1 I)$ will have dimension d_1 .

(R \implies L) Let λ_i be an eigenvalue of T . Then, the eigenspace $E_i \subseteq V$ corresponding to λ_i will have some basis, say $\{b_j^i\}_{j=1}^{d_i}$ where d_i is also the algebraic multiplicity of λ_i in $\Phi(x)$. It is because of this latter viewpoint that we can write $\sum_{i=1}^k d_i = n$. Now, consider the collection $\mathcal{B} = \{b_j^1\}_{j=1}^{d_1} \cup \{b_j^2\}_{j=1}^{d_2} \cup \dots \cup \{b_j^k\}_{j=1}^{d_k}$, which may consist of less than n vectors, but since we know that two eigenspaces always have trivial intersection, therefore \mathcal{B} will have n vectors. Now, we need to show that each element in \mathcal{B} is independent with each other. For that, suppose:

$$\sum_{b \in \mathcal{B}} c_b b = 0$$

where $c_b \in F$. Since each $\{b_j^i\}_{j=1}^{d_i}$ is a basis of eigenspace corresponding to λ_i , therefore the part of the above sum corresponding to $\sum_{j=1}^{d_i} c_{ij}b_j^i$ is an eigenvector of λ_i , for each $i = 1, \dots, k$. Therefore the above equation reduces to $v_1 + \dots + v_k = 0$ where each v_i is an eigenvector of λ_i . Since any two eigenvectors from different eigenspaces are necessarily independent, therefore each $v_i = 0$ and thus each c_b is zero by independence of the basis of eigenspace of λ_i .

I think its good to give a proof of the fact that vectors from distinct eigenspaces are necessarily independent. So suppose $v_1 + \dots + v_k = 0$ where v_i is in eigenspace corresponding to eigenvalue λ_i . Then we know that for each $f(x) \in F[x]$, we have

$$\begin{aligned} 0 &= f(T)0 \\ &= f(T)(v_1 + \dots + v_k) \\ &= f(c_1)v_1 + \dots + f(c_k)v_k. \end{aligned}$$

So if we simply consider $f(x) \in F[x]$ to be such that $f(x) = (x - c_2) \dots (x - c_k)$, then by above we get $0 = (c_1 - c_2)v_1 \iff v_1 = 0$ as $c_1 \neq c_2$. Doing same for other v_i s does the job. \square

Lemma 2.0.3. Let $T : V \rightarrow V$ be a linear operator on a finite-dimensional F -vector space. If every vector $v \in V$ is an eigenvector of T , then T is diagonalizable and eigenspaces are independent of each other.

Proof. Since any $v \in V$ is an eigenvector, therefore if we take any ordered basis $\mathcal{B} = \{v_1, \dots, v_n\}$ of V , then $v = c_1v_1 + \dots + c_nv_n$. Applying T on it, $Tv = c_1Tv_1 + \dots + c_nTv_n = c_1\lambda_1v_1 + \dots + c_n\lambda_nv_n$. Since $Tv = \lambda v$, therefore $\lambda(c_1v_1 + \dots + c_nv_n) = c_1\lambda_1v_1 + \dots + c_n\lambda_nv_n$. Thus, by independence of vectors v_i , we get $c_i = \lambda^{-1}c_i\lambda_i \iff \lambda = \lambda_i$ for those i for which $c_i \neq 0$ in $v = c_1v_1 + \dots + c_nv_n$. So the fact that each $v \in V$ is an eigenvector tells us that any basis \mathcal{B} of V has to consist of an element of an eigenspace from each eigenspace of V for T . Thus T is diagonalizable as any basis of V will be an eigenvector and T each eigenspace is not linearly dependent on other eigenspaces. \square

The main theorem is the following:

Theorem 2.0.4. Let T be a linear operator on a finite dimensional F -vector space V . Then TFAE:

1. T is a diagonalizable operator,
2. the characteristic polynomial of T is

$$\Phi(x) = (x - \lambda_1)^{d_1} \dots (x - \lambda_k)^{d_k}$$

where $\lambda_1, \dots, \lambda_k \in F$ are eigenvalues of T and $d_i = \dim \text{Ker}(T - \lambda_i I)$,

3. we have

$$\dim V = \sum_{i=1}^k \dim \text{Ker}(T - \lambda_i I),$$

4. for each eigenvalue λ_i of T , algebraic multiplicity of λ_i is equal to geometric multiplicity of λ_i .

Proof. (1 \implies 2) This is how we proved the L \implies R in Lemma 2.0.2 above.

(2 \implies 3) Since the characteristic polynomial of T is product of linear factors as given, therefore $d_1 + \dots + d_k = n = \dim V$ and since $d_i = \dim \text{Ker}(T - \lambda_i I)$.

(3 \implies 1 \iff 4) Since $\text{Ker}(T - \lambda_i I)$ spans V by dimension argument, we get a basis of V which consists of eigenvectors of T . \square

2.1 Annihilating polynomials

Ok, we need to be a bit quick here. We have that $\text{Hom}_{\text{Vect}_k^{\text{fin}}}(V, V) = L(V)$ is a ring under pointwise addition and composition. It is moreover an F -algebra under scalar multiplication. Let $T : V \rightarrow V$ be a linear operator on a finite dimensional F -vector space, so an element of $L(V)$. Then, we have an **annihilator ideal corresponding to T** of algebra $F[x]$ given by:

$$\text{Ann}(T) := \{f(x) \in F[x] \mid f(T) = 0 \in L(V)\}.$$

That is, we consider all the polynomials $f(x) \in F[x]$ such that the operator that we get after applying f to T will give you the zero 0 operator. Note that $L(V)$ doesn't obtain a $F[x]$ -module structure via the above mapping because $(T+U)^k \neq T^k + U^k$ for $k \in \mathbb{N}$. So the only thing we have going for us is that each element of $L(V)$ induces an ideal in $F[x]$:

$$\begin{aligned} L(V) &\longrightarrow \text{Ideal}(F[x]) \\ T &\longmapsto \text{Ann}(T). \end{aligned}$$

That's nice, we'll be using this annihilator ideal quite frequently.

Let us first see that $\text{Ann}(T)$ is always non-trivial in finite dimensional case:

Lemma 2.1.1. Let $T \in L(V)$ where V is a finite-dimensional F -vector space. Then, $\text{Ann}(T) \leq F[x]$ contains a non-zero element.

Proof. The proof uses the usual dimension argument. Since $\dim V = n$, therefore $\dim L(V) = n^2$ as an F -vector space, where one basis might be given as all matrices of the form $\{(a_{ij})\}_{i,j=1,\dots,n}$ where $a_{ij} = 1$ and rest zero in one particular matrix. So $L(V)$ is also a finite-dimensional F -vector space. Thus, if we take the operator $T \in L(V)$, then for the collection of elements $\{I, T, T^2, \dots, T^{n^2}\}$ in $L(V)$, we obtain a dependence relation as in any finite dimensional vector space, a collection of elements of size more than the dimension necessarily has to be dependent. Thus, we have:

$$T^n + c_{n-1}T^{n-1} + \dots + c_1T + c_0 = 0$$

in $L(V)$. What this tells us is that the corresponding polynomial $f(x) = x^n + c_{n-1}x^{n-1} + \dots + c_1x + c_0$ annihilates T , that is $f(T) = 0$. Thus, $f(x) \in \text{Ann}(T) \leq F[x]$ is a non-zero element of the ideal. \square

Ok so we know that if $T \in L(V)$, then for any $p(x) \in F[x]$, we get another linear operator $p(T) \in L(V)$. A natural question that now arises is that how does $p(T)$ act on a vector $v \in V$. Well in general we can only say that if $p(x) = \sum_{i=0}^n c_i x^i$, then $p(T)v = \sum_{i=0}^n c_i T^i(v)$, where T^i is of-course i -fold composition of T with itself. But this takes a much simpler form when we consider $v \in V$ to be an eigenvector of T :

Lemma 2.1.2. Let $T \in L(V)$, $\dim V = n$ and $v \in V$ an eigenvector of T corresponding to an eigenvalue $c \in F$. Then, for all $p(x) \in F[x]$,

$$p(T)v = p(c)v.$$

Proof. It is quite easy. \square

An important observation about $F[x]$ is that it is a PID. So this means that $\text{Ann}(T)$ is a principal ideal. Define that minimum degree monic polynomial $p(x) \in F[x]$ such that $\langle p(x) \rangle = \text{Ann}(T)$ to be the **minimal polynomial** of operator T .

There is a very important relationship between the minimal polynomial $p(x)$ and the characteristic polynomial $\Phi(x)$ for a given linear operator $T \in L(V)$.

Theorem 2.1.3. Let $T \in L(V)$ where V is a finite-dimensional F -vector space and denote $p(x)$ to be the minimal and $\Phi(x)$ to be the characteristic polynomial of T . Then,

1. $p(x)$ and $\Phi(x)$ has same roots,
2. the multiplicity of root c of $p(x) \leq$ the multiplicity of root c of $\Phi(x)$, for all roots of c of $p(x)$ (or of $\Phi(x)$, equivalently).

Proof. To show that $p(x)$ and $\Phi(x)$ have same roots, we need to show that $p(c) = 0$ if and only if c is an eigenvalue of T (it is necessary to show if and only if, because only then can we say that both of them will have identical roots). So suppose $p(c) = 0$ for some $c \in F$, then we can write $p(x) = (x - c)q(x)$ for $q(x) \in F[x]$. In order to show that c is an eigenvalue of T , we need only show that $\text{Ker}(T - cI) \subseteq V$ is non-trivial. We already have such a form in $p(x) = (x - c)q(x)$, well atleast the feeling is there. Ok, so we observe that $p(T) = (T - cI)q(T)$. Somehow we need an element $v \in V$ such that $p(T)v = 0$ but $q(T)v \neq 0$. Is such an element there? Now comes into the picture the minimality of polynomial $p(x)$ wrt to certain property. Since $\deg p(x) > \deg q(x)$, therefore $q(x)$ cannot possibly be in $\text{Ann}(T)$. Thus, $q(T) \neq 0$ in $L(V)$. Hence, $\exists v \in V$ such that $q(T)v \neq 0$. But since $p(T) = 0$ as $p(x) \in \text{Ann}(T)$, therefore $p(T)v = 0$ for the same v as above. So we can write $0 = p(T)v = (T - cI)q(T)v = (T - cI)w$, where $w = q(T)v \neq 0$. Hence $c \in F$ is an eigenvalue of T . Conversely, if c is an eigenvalue of T , then $\exists v \in V$ such that $(T - cI)v = 0$. Then by Lemma 2.1.2, we obtain that $p(T)v = p(c)v$. But since $p(x) \in \text{Ann}(T)$, so $p(T)v = 0$, irrespective of what special property $v \in V$ satisfies. Thus, $p(c)v = 0$. Now, two cases can happen, either $v = 0$ or $p(c) = 0$. The former cannot happen because v is an eigenvector, which are by definition non-zero. Hence $p(c) = 0$.

So $p(x)$ and $\Phi(x)$ have exactly the same roots. But how can we talk about multiplicities of those roots? Well, we'll soon see that $\Phi(x)$ is also in $\text{Ann}(T)$, thus, $p(x) | \Phi(x)$. Hence, the multiplicities of roots of $p(x)$ is equal to or lower than those of $\Phi(x)$. \square

We next observe that we can find the minimal polynomial of a diagonalizable operator by knowing its distinct eigenvalues. We cannot do the same by the above Theorem 2.1.3 because it only tells us that the multiplicities of its roots will be lower than that of $\Phi(x)$, but may not be always 1.

Lemma 2.1.4. ¹ Let $T \in L(V)$ be a diagonalizable operator with distinct eigenvalues c_1, \dots, c_k . Then the minimal polynomial of T , denoted $p(x)$ is given by:

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

Proof. Since T is diagonalizable, therefore $V = \bigoplus_{i=1}^k \text{Ker}(T - c_i I)$. Suppose $p(x)$ is the minimal polynomial of T . Then by Theorem 2.1.3, we clearly have that the roots of $p(x)$ are exactly all the

¹The second part of this proof is different than what was taught.

eigenvalues c_i , but possibly with multiplicities more than 1. So we reduce to showing that for any eigenvalue c , the polynomial $q(x) \in F[x]$ in the factorization

$$p(x) = (x - c)q(x),$$

where we have $\deg q(x) < \deg p(x)$, cannot have c as a root, that is, we need to show that $q(c) \neq 0$. So assume to the contrary that $q(c) = 0$. Since c is an eigenvalue of T , therefore $q(T)v = q(c)v = 0$ for an eigenvector $v \in V$ of eigenvalue $c \in F$. So $q(T)v = 0$ in V . Since $Tv = cv$, therefore $v \in \text{Ker}(T - cI)$. Since $v \in \text{Ker}(T - cI)$ was arbitrarily chosen, therefore $q(T)v = 0$ for all $v \in \text{Ker}(T - cI) \subseteq V$. Since $q(T) \in L(V)$, therefore $\text{Ker}(q(T)) \supseteq \text{Ker}(T - cI)$. Therefore $\dim \text{Ker}(q(T)) \geq \dim \text{Ker}(T - cI) \geq 1$. Therefore $\dim \text{Ker}(q(T)) \geq 1$. Thus, $\dim \text{Image}(q(T)) \leq n - 1$, where $n = \dim V$. Thus, $q(T)$ is not surjective which is equivalent to saying that $q(T)$ is not injective in this case because V is a finite-dimensional F -vector space. But since T is diagonalizable, therefore we can also write T in the matrix form as $A_T = P^{-1}\Lambda P$ where P is an invertible $n \times n$ matrix and Λ is a diagonal matrix whose entries are all the eigenvalues of T . Since the minimal/characteristic polynomials of both T and A_T are same, therefore we first observe that:

$$\begin{aligned} q(A_T) &= q(P^{-1}\Lambda P) \\ &= P^{-1}q(\Lambda)P \\ &= P^{-1}\Lambda'P. \end{aligned}$$

Thus, $q(A_T)$ and hence $q(T)$ is itself diagonalizable. Since diagonalizable operators are invertible, which is equivalent to saying that they are surjective, therefore $\dim \text{Image}(q(T)) = n$, a contradiction to the above. Hence, $q(c) \neq 0$. We are done. \square

Remark 2.1.5. (*On polynomials preserving niceness of operators*) One would have noticed in the above proof that we explicitly used the trivial observation that if T is diagonalizable, then so is $p(T)$ where $p(x) \in F[x]$ is any polynomial. This one example of the general feature that polynomials have wrt operators, that they preserves some of the nice features that T may have. Another one of such phenomenon is the following fact:

similar matrices/operators have same minimal and characteristic polynomials.

One should keep an eye out to find more such properties about operators on f.d. vector spaces which are preserved by polynomials, as and when they are introduced.

Ok so we finally state the Cayley-Hamilton theorem in this language of annihilators so that you can make sense of last line of the proof of Theorem 2.1.3 and then we'll move on to something more fascinating:

Theorem 2.1.6. Let $T \in L(V)$ where V is a finite-dimensional F -vector space. Denote $\Phi(x) \in F[x]$ to be the characteristic polynomial of T . Then,

$$\Phi(x) \in \text{Ann}(T). \tag{1}$$

2.2 Invariant subspaces

Let $T \in L(V)$ be a linear operator on an F -vector space. A **T -invariant subspace** $W \subseteq V$ is defined to be that subspace of V such that $T(W) \subseteq W$. Some examples are as follows:

- For any $T \in L(V)$, the subspaces $\{0\}$, V are T -invariant.
- For the F -vector space $F[x]$ and derivative linear operator in $L(F[x])$, the following subspace is invariant:

$$W_n = \{f(x) \in F[x] \mid \deg f \leq n\} \cup \{0\}.$$

Given two commuting linear operators on a vector space, we automatically obtain two more invariant subspaces for each of the linear operators in question:

Lemma 2.2.1. Let $T, U \in L(V)$ where V is an F -vector space. If $TU = UT$, then $\text{Image}(U)$ and $\text{Ker}(U)$ are invariant T -invariant subspaces.

Proof. There are no surprises here. □

As we observed in Remark 2.1.5, something similar happens between the commuting relation of $f(T)$ and T . In particular,

Lemma 2.2.2. Let $T \in L(V)$ where V is an F -vector space and let $f(x) \in F[x]$. Then $f(T)$ and T commutes with each other.

Proof. Nothing to show here. □

A corollary of Lemmas 2.2.1 and 2.2.2 is the following:

Corollary 2.2.3. Let $T \in L(V)$ and let $c \in F$ be an eigenvalue of T , then $\text{Ker}(T - cI)$ and $\text{Image}(T - cI)$ are T -invariant subspaces of V .

Proof. Lemma 2.2.2 tells us that if we consider $f(x) = x - c$, then $f(T) = T - cI$ and T commutes with each other. By Lemma 2.2.1, we obtain that $\text{Ker}(T - cI)$ and $\text{Image}(T - cI)$ are T -invariant. □

Here's a good exercise:

Example. Consider the following linear operator on \mathbb{R}^2 :

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

Find all A -invariant subspaces of \mathbb{R}^2 . Ok so for this, what we will try to do is to find a proper subspace of \mathbb{R}^2 which will be A -invariant. Thus suppose $W \subseteq \mathbb{R}^2$ is a proper non-trivial subspace, thus $\dim W = 1$. Now if $AW \subseteq W$, therefore there is only one choice that $AW = W$ as the only proper subspace of W is $\{0\}$. Since W is of dimension 1, therefore A on W gives a linear operator on W which is surjective, hence injective, hence bijective and hence invertible. So since we can write $W = \langle w \rangle \subseteq V$, therefore $Aw = cw$ for some $c \in \mathbb{R}$, that is, A has an eigenvalue in \mathbb{R} . But notice that it is not true because $\Phi(x) = x^2 + 1$. Hence W cannot be invariant for A . □

Let $T \in L(V)$ and $W \subseteq V$ be T -invariant. Then we obtain $T_W \in L(W)$, by restricting T to W . We will now find relationship between the matrix representation of T and T_W for a T -invariant subspace $W \subseteq V$.

Proposition 2.2.4. Let $T \in L(V)$ where V is a finite-dimensional F -vector space. Suppose $\mathcal{B} = \{b_1, \dots, b_n\}$ is an ordered basis of V and let $W \subseteq V$ be a T -invariant subspace whose basis is given by $\mathcal{B}' = \{b_1, \dots, b_r\}$ where $r \leq n$. Then,

1. the matrices $[T]_{\mathcal{B}}$ and $[T_W]_{\mathcal{B}'}$ are related as follows:

$$[T]_{\mathcal{B}} = \begin{bmatrix} [T_W]_{\mathcal{B}'} & C \\ 0 & D \end{bmatrix}$$

where C is of shape $r \times (n - r)$ and D is of shape $n - r \times n - r$,

2. the characteristic polynomial of T_W $\Phi_W(x)$ divides the characteristic polynomial of T $\Phi(x)$,
3. the minimal polynomial of T_W $p_W(x)$ divides the minimal polynomial of T $p(x)$.

Proof. 1. Denote $A = [T]_{\mathcal{B}}$ and $B = [T_W]_{\mathcal{B}'}$. Then, for any b_j , $j \leq r$, we have $Tb_j = \sum_{i=1}^n A_{ij}b_i$. Since $Tb_j \in W$, therefore $\sum_{i=1}^n A_{ij}b_i \in W$. This means that $A_{ij} = 0$ for all $n \geq i > r$, thus $Tb_j = \sum_{i=1}^r A_{ij}b_i$ for all $1 \leq j \leq r$.

2. To see this, define $A = [T]_{\mathcal{B}}$ and $B = [T_W]_{\mathcal{B}'}$. Then,

$$\begin{aligned} \det(A - xI) &= \det \begin{bmatrix} B - xI & C \\ 0 & D - xI \end{bmatrix} \\ &= \det(B - xI) \det(D - xI) \end{aligned}$$

which tells us that $\Phi_W(x) | \Phi(x)$.

3. The minimal polynomial $p_W(x)$ is the generator of $\text{Ann}(T_W)$, that is, if $f(x) \in F[x]$ is such that $f(T_W) = 0$, then $p_W(x) | f(x)$. Now, take any $f(x) \in \text{Ann}(T) = \langle p(x) \rangle$. Then $f(T) = 0$. That means that for all $v \in V$, $f(T)(v) = 0$. But this means that for all $w \in W$, $f(T)(w) = 0$, therefore $f(T_W) : W \rightarrow W$ is zero. Hence $f(x) \in \text{Ann}(T_W)$. Thus $\text{Ann}(T) \subseteq \text{Ann}(T_W)$. Hence $p_W(x) | p(x)$. \square

Lecture # 3, 16/08/2022

2.3 Characterization of diagonal operators in terms of minimal polynomial

We now develop tools which will allow us to characterize all diagonalizable and triangulable operators on a f.d. F -vector space. The main ingredient in all of this is the notion of T -conductors.

Let $T \in L(V)$ be a linear operator on a finite dimensional F -vector space. Let $W \subseteq V$ be a T -invariant subspace of V . Let $\alpha \notin W$ be an element of V . Then, the **T -conductor of α into W** is the following ideal of $F[x]$:

$$S_T(\alpha; W) := \{f(x) \in F[x] \mid f(T)(\alpha) \in W\}.$$

Indeed $S_T(\alpha; W)$ is an ideal of $F[x]$ as the following lemma shows:

Lemma 2.3.1. Let $T \in L(V)$ where V is a finite dimensional F -vector space and $W \subseteq V$ is T -invariant. Then,

1. for all $f \in F[x]$, W is $f(T)$ -invariant,
2. for all $\alpha \in V$, the T -conductor of α into W is an ideal of $F[x]$.

Proof. 1. Take any $f(x) \in F[x]$ and $w \in W$. Then $f(T)w = \sum_{i=0}^n c_i T^i(w)$. Since $T(W) \subseteq W$, thus $T^i w \in W$ for all i . Hence $f(T)w \in W$, showing that $f(T)(W) \subseteq W$ and that if W is T -invariant, then W is also $f(T)$ -invariant for all $f \in F[x]$.

2. To show $S_T(\alpha; W)$ is an ideal of $F[x]$, take $f \in S_T(\alpha; W)$ and $g \in F[x]$. Then $f(T)g(T)\alpha = g(T)f(T)\alpha = g(T)(f(T)\alpha)$, where $f(T)\alpha \in W$ as f is in the T -conductor. Now, by 1. above, W is $g(T)$ -invariant as well, so $g(T)(f(T)\alpha) \in W$. Hence $g(T)f(T) = f(T)g(T) \in S_T(\alpha; W)$. \square

Since $F[x]$ is a PID and $S_T(\alpha; W)$ is an ideal of $F[x]$, therefore $S_T(\alpha; W) = \langle g(x) \rangle$, for some $g(x) \in F[x]$. Even though this generator won't be used much, we will still be calling it the **T -conductor α into W** .

Since for all $f \in \text{Ann}(T)$, $f(T)\alpha = 0$ as $f(T) = 0$, therefore $f \in S_T(\alpha; W)$, so $\text{Ann}(T) \subseteq S_T(\alpha; W)$ for all $\alpha \in V$ and for all T -invariant W .

Let us now introduce a lemma which will be our key to characterization of diagonal and triangular operators. In-fact, we will generalize it after some time and that will too depend on this lemma.

Lemma 2.3.2. Let $T \in L(V)$ be a linear operator on a finite dimensional F -vector space V . If $W \subsetneq V$ is T -invariant and the minimal polynomial $p(x)$ of T can be written as

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

where $c_i \in F$, then $\exists \alpha \notin W$ such that $(T - c_i I)\alpha \in W$ for some $i = 1, \dots, k$.

Proof. We will construct one such $\alpha \notin W$ satisfying the above condition. Notice that the condition is demanding us to show that there is a polynomial of form $x - c_i$ in the conductor $S_T(\alpha; W)$. So this observation motivates our proof. First, take any $\beta \notin W$ and consider the conductor $S_T(\beta; W) = \langle g(x) \rangle$. Since $\text{Ann}(T) = \langle p(x) \rangle$ is contained in $S_T(\beta; W)$, therefore $g(x) | p(x)$ which is to say that

$$g(x) = (x - c_1)^{e_1} \dots (x - c_k)^{e_k}$$

where $0 \leq e_i \leq r_i$ for all $i = 1, \dots, k$. Can all e_i be zero? If they are then $g(x) = 1$ and hence $S_T(\beta; W) = F[x]$ and hence $x \in S_T(\beta; W)$ and hence $\beta \in W$, a contradiction to our starting assumption. So there $i_0 = 1, \dots, k$ such that $e_{i_0} \neq 0$. Consider then the polynomial $h(x)$ as in

$$g(x) = (x - c_{i_0})h(x).$$

We are close to the form needed in the question. Now, we need only show that $h(T)\beta \notin W$ because then, taking $\alpha := h(T)\beta$, we will get that $g(T)\beta = (T - c_{i_0}I)\alpha$ where $g(T)\beta \in W$ as $g \in S_T(\beta; W)$. So we reduce to showing that $h(T)\beta \notin W$. Well it is clear that $h(T)\beta \notin W$ because if to the contrary we find that $h(T)\beta \in W$, then $h(x) \in S_T(\beta; W) = \langle g(x) \rangle$ and hence we obtain a contradiction to the minimality of $g(x)$. So we are done. \square

We now come to the main result, a characterization of diagonal operators.

Theorem 2.3.3. Let $T \in L(V)$ be a linear operator on a finite dimensional F -vector space. Then, the following are equivalent:

1. The minimal polynomial $p(x)$ of T is of form

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

where $c_1, \dots, c_k \in F$ are distinct.

2. The operator T is diagonalizable.

Proof. The implication $2 \implies 1$ is exactly the content of Lemma 2.1.4, so we need only show $1 \implies 2$. We need to show that

$$T = \text{Ker}(T - c_1 I) + \cdots + \text{Ker}(T - c_k I).$$

That seems a bit difficult because we need to show that every vector is an eigenvector just by knowing that minimal polynomial is split linearly and simply over F . So we try to prove the contrapositive, which is: If T is not diagonalizable, then $p(x)$ has a repeated root. Now how is this easier to prove? Well, we have the Lemma 2.3.2 on our side, and it tells us exactly that for a given T -invariant subspace W , there is a $\alpha \notin W$ such that $x - c_i \in S_T(\alpha; W)$. Since we are given that T is not diagonalizable, therefore the direct sum of eigenspaces is not whole of T and one does have a feeling that sum of eigenspaces will naturally be T -invariant, and this feeling will open a door to a possible proof. So we are motivated to prove this contrapositive. So let us check whether our feeling is correct or not. Let $W := \bigoplus_{i=1}^k \text{Ker}(T - c_i I)$. Since T is not diagonalizable, therefore $W \subsetneq V$. Now, W is also T -invariant because for any $v_1 + \dots + v_k \in W$, $T(v_1 + \dots + v_k) = Tv_1 + \dots + Tv_k = a_1 v_1 + \dots + a_k v_k \in W$. Now, using Lemma 2.3.2, we get an $\alpha \notin W$ such that $\beta := (T - c_{i_0} I)\alpha \in W$ for some $1 \leq i_0 \leq k$. Now remember that we have to show that there is a repeated root of the minimal polynomial $p(x)$. The previous discussion motivates us to claim that if we write

$$p(x) = (x - c_{i_0})h(x),$$

then $h(x)$ should have c_{i_0} as a root also. If that's the case, then we're done. So let us see if that's actually true or not. Since we have that $(T - c_{i_0})\alpha \in W$, so we try to observe what we can get after putting T in above equation:

$$\begin{aligned} (T - c_{i_0} I)h(T) &= p(T) \\ &= 0. \end{aligned}$$

So $(T - c_{i_0} I)h(T)$ is the zero operator. Now, observe that $h(x) - h(c_{i_0}) = (x - c_{i_0})q(x)$ for some $q(x) \in F[x]$. Since $0 = p(T)\alpha = (T - c_{i_0} I)h(T)\alpha$, therefore $h(T)\alpha \in \text{Ker}(T - c_{i_0} I) \subset W$. Thus,

$$\begin{aligned} h(T)\alpha - h(c_{i_0})\alpha &= (T - c_{i_0} I)q(T)\alpha \\ &= q(T)(T - c_{i_0} I)\alpha \\ &= q(T)\beta \end{aligned}$$

and since $\beta \in W$ with W being T -invariant, therefore W is $q(T)$ -invariant by Lemma 2.3.1, 1. Thus, $q(T)\beta \in W$. So in total we have $h(T)\alpha \in W$ and $h(T)\alpha - h(c_{i_0})\alpha \in W$, thus giving us that $h(c_{i_0})\alpha \in W$. If $h(c_{i_0}) \neq 0$, then $\alpha \in W$. Use contrapositive of this easy statement and the fact that $\alpha \notin W$ to conclude that $h(c_{i_0}) = 0$. \square

Lecture # 4, 22/08/2022

We begin with the following remark:

Remark 2.3.4. So far we have seen two ways to check whether a linear operator $T : V \rightarrow V$ on a finite dimensional F -vector space is diagonalizable or not. Denote $\Phi(x) = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$ to be the characteristic polynomial of T . Then, T is diagonalizable if and only if:

1. (Lemma 2.0.2) : for each $i = 1, \dots, k$, $d_i = \dim \text{Ker}(T - c_i I)$,
2. (Theorem 2.3.3) : $(T - c_1 I) \dots (T - c_k I) = 0$.

We will see a similar characterization for triangulable operators now.

2.4 Characterization of triangulable operators in terms of minimal polynomial

The Lemma 2.3.2 is very important to us, as we will see soon in the proof of the even more important theorem below:

Theorem 2.4.1. Let $T \in L(V)$ be a linear operator on a finite dimensional F -vector space. Then, the following are equivalent:

1. The minimal polynomial $p(x)$ of T splits into linear factors over F as follows:

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

where $c_i \in F$.

2. There exists a basis \mathcal{B} of V such that $[T]_{\mathcal{B}}$ is upper triangular.

Proof. ($1 \Rightarrow 2$) We will successively use Lemma 2.3.2 to construct the said basis. First of all, it is beneficial to actually look at what we want. Suppose we already have a basis \mathcal{B} such that $[T]_{\mathcal{B}}$ is upper triangular. We would then have:

$$[T]_{\mathcal{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

So if we write basis \mathcal{B} as $\mathcal{B} = \{b_1, \dots, b_n\}$ and since we know that the j^{th} -column of $[T]_{\mathcal{B}}$ is given by coordinates of Tb_j in \mathcal{B} , so we get that

$$Tb_j = \sum_{i=1}^j a_{ij} b_i.$$

Thus, $Tb_j \in \text{Span}(b_1, \dots, b_j)$. This is a crucial information because using this, we will construct b_j 's inductively by an intriguing use of Lemma 2.3.1. So let us begin.

Consider the trivial T -invariant subspace $W = \{0\}$. Using Lemma 2.3.2 on W (which uses the hypothesis that $p(x)$ is split over F into linear factors), we get $\alpha_1 \neq 0$ and an eigenvalue a_{11} of T (so $a_{11} = c_i$ for some $1 \leq i \leq k$) such that $(T - a_{11}I)\alpha_1 \in W = \{0\}$. So $T\alpha_1 = a_{11}\alpha_1$. Next

consider the $W_1 := \text{Span}(\alpha_1)$. So $T\alpha_1 \in W_1$. It is easy to see that $T(W_1) \subseteq W_1$ as $W_1 = \langle \alpha_1 \rangle$ and we know that $T\alpha_1 \in W_1$. So W_1 is a T -invariant subspace. Again using Lemma 2.3.2, we get $\alpha_2 \notin W_1$ and eigenvalue a_{22} of T such that $T\alpha_2 - a_{22}\alpha_2 \in W_1$, that is $T\alpha_2 \in \text{Span}(\alpha_1, \alpha_2)$. Denote $W_2 := \text{Span}(\alpha_1, \alpha_2) = \langle \alpha_1, \alpha_2 \rangle$. Note that $W_2 \supset W_1$. We again see that W_2 is T -invariant because $T\alpha_1 \in W_1 \subset W_2$ and $T\alpha_2 \in W_2$. Continuing like this, we will get at end a basis \mathcal{B} of V via this process which will terminate when we would have in the next step obtained the whole space V . The basis vectors here are $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$, but they are not just any vectors, they follow the special property that:

$$T\alpha_j \in \text{Span}(\alpha_1, \dots, \alpha_j).$$

Hence, when represented T as a matrix using this basis, we will obtain an upper triangular matrix².

(2 \Rightarrow 1) Suppose \mathcal{B} is a basis of V in which $[T]_{\mathcal{B}}$ is an upper triangular matrix. We need to show that minimal polynomial $p(x)$ splits linearly over F . We will use Theorem 2.1.3 here. First we see that $\Phi(x) = \det([T]_{\mathcal{B}} - xI) = \prod_{i=1}^n (x - a_{ii})$. So $\Phi(x) = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$ where $c_i \in F$. By Theorem 2.1.3, we see that $p(x)$ will have to be split into linear factors over F . \square

With the above major theorem done, we can see some easy corollaries:

Corollary 2.4.2. Let F be an algebraically closed field. If $T \in L(V)$ where V is an F -vector space, then T is triangulable.

Proof. Consider $p(x)$ to be the minimal polynomial of T . Since $p(x) \in F[x]$ and F is algebraically closed, we have that $p(x)$ splits into linear factors over $F[x]$. By Theorem 2.4.1, we are done. \square

We can actually prove Cayley-Hamilton theorem over algebraically closed field using the above theorem:

Corollary 2.4.3. (*Cayley-Hamilton Theorem*) Let F be an algebraically closed field, V be an F -vector space and $T \in L(V)$ be a linear operator on V . Then, $\Phi(x) \in \text{Ann}(T)$.

Proof. Since by Corollary 2.4.2 we have that T is triangulable, so we reduce to showing that if T is a triangulable operator, then $\Phi(x) \in \text{Ann}(T)$. Since characteristic polynomial of T is same as characteristic polynomial of $[T]_{\mathcal{B}}$, so in particular, taking \mathcal{B} to be the basis such that $A := [T]_{\mathcal{B}}$ is upper triangular, we reduce to showing that $\Phi(x)$ for A which is upper triangular is in $\text{Ann}(A) = \text{Ann}(T)$. So write

$$\begin{aligned} \Phi(x) = \det(A - xI) &= \det \left(\begin{bmatrix} a_{11} - x & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} - x & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} - x & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} - x \end{bmatrix} \right) \\ &= \prod_{i=1}^n (a_{ii} - x). \end{aligned}$$

²The importance of Lemma 2.3.2, and hence of T -conductors, cannot be overstated here.

We need to show that $\Phi(A) = 0$. In particular, we need to see that

$$(a_{11}I - A) \dots (a_{nn}I - A) = 0.$$

But that follows quite easily. Hence we are done. \square

2.5 Simultaneous diagonalizability and triangulability

Suppose T_1 and T_2 are two linear operators on a finite dimensional F -vector space such that there exists a basis \mathcal{B} such that both $[T_1]_{\mathcal{B}}$ and $[T_2]_{\mathcal{B}}$ are diagonal. Then, we call T_1 and T_2 to be **simultaneously diagonalizable**. Notice that if T_1 and T_2 are simultaneously diagonalizable, then they commute, because $T_1T_2 = PAP^{-1}PBP^{-1} = PABP^{-1} = PBAP^{-1} = PBP^{-1}PAP^{-1} = T_2T_1$ where A and B are diagonal. We will see in this section that commutativity of operators is also a sufficient condition for simultaneous diagonalizability and also triangulability.

Our goal is to prove theorems akin to Theorems 2.3.3 and 2.4.1. For that, in order to follow same proof strategy, we need to generalize the Lemma 2.3.2 to our case. This is the first thing we do.

Lemma 2.5.1. Let V be a finite dimensional F -vector space and let $\mathcal{F} \subset L(V)$ be a family of commuting triangulable linear operators. If $W \subseteq V$ is an \mathcal{F} -invariant subspace of V , then $\exists \alpha \notin W$ such that

$$T\alpha \in \text{Span}(\alpha, W) \quad \forall T \in \mathcal{F}.$$

Proof. We will again use Lemma 2.3.2 here (the only place where we will need the triangulability hypothesis). But first note that we can reduce to \mathcal{F} being a finite set of commuting operators because $L(V)$ is of finite dimension. In particular, we can reduce to the maximal linearly independent set of vectors in \mathcal{F} , that is, the basis of $\text{Span}(\mathcal{F})$. If we can prove that such an α exists for that basis, then same alpha will work for any $T \in \mathcal{F}$. So having been reduced to \mathcal{F} being finite, say $\mathcal{F} = \{T_1, \dots, T_r\}$, we also reduce to finding $\alpha \notin W$ such that $T_j\alpha \in \text{Span}(\alpha, W)$ for all $j = 1, \dots, r$. So how do we construct such an α ? Well, if we have learned anything from Theorem 2.4.1 is that we should approach most of such constructions in an inductive manner as we will be crucially using the finite-dimensionality of V to say that such a construction will indeed end. So having said that, the construction of such an α will be as follows.

Begin with T_1 . We know that W is T_1 -invariant, so using Lemma 2.3.2³, we get $\alpha_1 \notin W$ such that $T_1\alpha_1 \in \text{Span}(\alpha_1, W)$. But of-course this doesn't tell us that, say $T_2\alpha_1 \in \text{Span}(\alpha_1, W)$. So how should we produce next iteration of α ? Well we need to include that T_2 also throw the *to-be-sought* α_2 inside $\text{Span}(\alpha_2, W)$. So clearly, we should first focus on *all* those elements $\alpha \in V$ which are thrown by T_1 inside $\text{Span}(\alpha, W)$, that is, denote:

$$V_1 := \{\alpha \in V \mid T_1\alpha \in \text{Span}(\alpha, W)\}.$$

Observe that $\alpha_1 \in V_1$ and $W \subsetneq V_1$. But now if we similarly define V_2 to be all those elements α thrown inside $\text{Span}(\alpha, W)$ by T_2 , then we may get that V_1 and V_2 may be incomparable, and that

³At this point it should be clear that how important this lemma actually is.

will not be ideal. So we instead of focusing on T_2 , we focus on $T_2|_{V_1}$. But there's a problem, how are we sure that $T_2(V_1) \subseteq V_1$? Well, we see that V_1 is \mathcal{F} -invariant(!) To see this, we will have to use that fact that \mathcal{F} is a commuting family. Take $\alpha \in V_1$ and $T_i \in \mathcal{F}$, we wish to show that $T_i\alpha \in V_1$. Since $T_1\alpha - c_1\alpha \in W$, therefore $T_iT_1\alpha - c_1T_i\alpha \in T_iW \subseteq W$ as W is \mathcal{F} -invariant. By commutativity of \mathcal{F} , we have $T_1T_i\alpha - c_1T_i\alpha \in W$, therefore $T_i\alpha \in V_1$, thus making \mathcal{F} a T_i -invariant subspace. Hence, $T_2|_{V_1}$ is a valid linear operator on V_1 , where $V_1 \supsetneq W$. Now, we apply Lemma 2.3.2 again on $T_2|_{V_1}$ where W is now treated as a $T_2|_{V_1}$ -invariant subspace of V_1 . Then we again get $\alpha_2 \notin W$ but $\alpha_2 \in V_1$ (!) such that $T_2|_{V_1}(\alpha_2) \in \text{Span}(\alpha_2, W)$. We then again define the following set:

$$V_2 := \{\alpha \in V_1 \mid T_2|_{V_1}(\alpha) \in \text{Span}(\alpha, W)\}$$

Observe that $\alpha_1, \alpha_2 \in V_2$ and $W \subsetneq V_2 \subsetneq V_1$. Continue like this and we will get a sequence of subspaces $W \subsetneq V_r \subsetneq V_{r-1} \subsetneq \cdots \subsetneq V_2 \subsetneq V_1$ and correspondingly $\alpha_r, \alpha_{r-1}, \dots, \alpha_2, \alpha_1$ where $\alpha_i \in V_i$ and $\alpha_i \notin W$ for all $1 \leq i \leq r$. Hence, we have that $\alpha_r \in V_r$ and $V_r \subsetneq V_i$ for each $i = 1, \dots, r$. This means that $\alpha_r \notin W$ is such an element of V such that $T_i\alpha_r \in \text{Span}(\alpha_r, W)$ for each $i = 1, \dots, r$. Hence, we are done. \square

Lecture # 5, 23/08/2022

So we now see the simultaneous triangulability theorem, which says that any commuting family of triangulable operators is simultaneously triangulable:

Theorem 2.5.2. Let V be a finite-dimensional F -vector space and let $\mathcal{F} \subseteq L(V)$ be a commuting family of triangulable linear operators on V . Then, there exists an ordered basis \mathcal{B} of V such that

$$[T]_{\mathcal{B}} \text{ is triangulable } \forall T \in \mathcal{F}.$$

Proof. Using Lemma 2.5.1, one needs to use the carbon-copy of the strategy of proof of Theorem 2.4.1. \square

Corollary 2.5.3. Let V be a finite dimensional F -vector space where F is algebraically closed. Suppose $\mathcal{F} \subseteq L(V)$ is a commuting family of linear operators on V , then there exists an invertible matrix P such that

$$PAP^{-1} \text{ is upper triangular } \forall A \in \mathcal{F}.$$

Proof. Since F is algebraically closed, so for any $T \in L(V)$, the minimal polynomial of T , $p(x)$ is split into linear factors over F . Hence each operator T is triangulable by Theorem 2.4.1. So \mathcal{F} is a commuting family of triangulable operators on V . Now use Theorem 2.5.2 to deduce that there exists a basis \mathcal{B} of V such that for each $A \in \mathcal{F}$, $[A]_{\mathcal{B}}$ is upper triangular. By using the base change matrix P , we get that $P^{-1}AP$ is upper triangular. \square

The next result is on simultaneous diagonalization. It tells you that a commuting family of diagonalizable operators is always simultaneously diagonalizable, that is, this is a companion to the Theorem 2.5.2.

Theorem 2.5.4. Let V be a finite dimensional F -vector space and suppose $\mathcal{F} \subseteq L(V)$ is a commuting family of diagonalizable operators. Then, there exists a basis \mathcal{B} of V such that

$$[T]_{\mathcal{B}} \text{ is diagonal } \forall T \in \mathcal{F}.$$

Proof. The proof of this theorem is also akin to the ones we did above, by constructing the required basis each time and critically making use of finite dimensionality. A similar setup will work here, but we will actually use induction here. In-fact we will use strong-induction here. So we begin first by taking any $T \in \mathcal{F}$. We know that T is diagonalizable, therefore by Theorem 2.0.4, we have that T has all eigenvalues inside field F and let us denote the distinct eigenvalues by $c_1, \dots, c_k \in F$. We next use induction on $\dim V$. If $\dim V = 1$, then representation of T in any basis will be a scalar and hence there is nothing to be proved here as matrices of form 1×1 are always diagonalizable. Next, suppose $\dim V = n > 1$. Moreover, suppose that for all subspaces $W \subseteq V$ of dimension $\leq n - 1$, we have that the above hypothesis is true; for each commuting family \mathcal{F}_W of diagonalizable operators on W , there exists a basis \mathcal{B}_W of W such that $[T_W]_{\mathcal{B}_W}$ is diagonal for all $T_W \in \mathcal{F}_W$.

Hopefully spelling out what we already have would have made it clear that we will use the diagonalizability of each $T \in \mathcal{F}$ and will reduce to their eigenspaces where we know the hypothesis is true. So let us spell it out more concretely. First, observe that $W_i := \text{Ker}(T - c_i I) \neq 0$ for each $i = 1, \dots, k$ as T is diagonal. Now, having obtained a proper subspace of V and knowing that we have a family of commuting operators, it is natural to ask whether it is invariant under that family. Well, we now observe that each W_i is indeed \mathcal{F} -invariant because for each $S \in \mathcal{F}$, we have that S commutes with T (given assumption) and hence S commutes with $T - c_i I$. Thus, by Lemma 2.2.1, $\text{Ker}(T - c_i I) = W_i$ is S -invariant. But we somehow need a basis on each W_i (which should make $T|_{W_i}$ diagonal) using the inductive step, so that we can join each of the basis together to get that T represented in that basis is diagonal. By this, it is clear that we should look at the following commuting family of operators:

$$\mathcal{F}|_{W_i} := \{S|_{W_i} : W_i \rightarrow W_i \mid S \in \mathcal{F}\}.$$

The fact that $S|_{W_i}$ is a linear operator on W_i follows from the fact that each W_i is \mathcal{F} -invariant. So indeed, $\mathcal{F}|_{W_i}$ is a commuting family of diagonalizable operators on W_i , where $\dim W_i < n$. So by (strong) inductive hypothesis, we get that there exists a basis \mathcal{B}_i of W_i such that for all $S|_{W_i} \in \mathcal{F}|_{W_i}$, we have that $[S|_{W_i}]_{\mathcal{B}_i}$ is diagonal.

Finally, we use the fact that T is diagonal, therefore

$$V = \bigoplus_{i=1}^k \text{Ker}(T - c_i I) = \bigoplus_{i=1}^k W_i$$

so that the basis \mathcal{B}_i of W_i can be joined together to give basis $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_k)$ of V . Now, we observe that

$$[S]_{\mathcal{B}} = \begin{bmatrix} [S|_{W_1}]_{\mathcal{B}_1} & 0 & \cdots & 0 \\ 0 & [S|_{W_2}]_{\mathcal{B}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [S|_{W_k}]_{\mathcal{B}_k} \end{bmatrix}$$

where each $[S|_{W_i}]_{\mathcal{B}_i}$ is diagonal, hence making $[S]_{\mathcal{B}}$ diagonal, for all $S \in \mathcal{F}$. It is important to note what we did here. We took one diagonalizable linear operator $T \in \mathcal{F}$ and constructed a basis out of it for V in which each $S \in \mathcal{F}$ is diagonal. So we have indeed constructed a basis such that each each member in \mathcal{F} is diagonal. Thus we are done. \square

2.6 Direct sum decomposition

The first reason that we wish to decompose our old little finite dimensional F -vector space is because, as we will see, there are special type of linear operators (projections) whose collection will directly correspond to any such direct sum decomposition of V (Theorem 2.6.5). Next, we will see that diagonalizability of an operator literally depends on the existence of this so-called projection operators (Theorem 2.7.2), which as we said above, is same as giving a direct sum decomposition. So this is the reason we are doing this section. Well the thing is, most students know what an internal direct sum is in vector spaces. So we will be a bit quick here and will go to the funky part soon.

Remember that two subspaces $W_1, W_2 \subseteq V$ are said to be independent if $w_1 + w_2 = 0$ implies $w_1 = w_2 = 0$ where $w_i \in W_i$.

Lemma 2.6.1. Let V be a f.d. F -vector space and suppose $W_1, \dots, W_k \subseteq V$ are independent. If $\alpha = \alpha_1 + \dots + \alpha_k$ where $\alpha_i \in W_i$, then the above expression is unique.

Proof. Nothing to see here. □

We have a simple characterization of independent subspaces:

Lemma 2.6.2. Let V be a f.d. F -vector space and $W_1, \dots, W_k \subseteq V$ be a set of subspaces. Then the following are equivalent:

1. Subspaces $W_1, \dots, W_k \subseteq V$ are independent.
2. For all $j = 1, \dots, k$, $W_j \cap (W_1 + \dots + W_{j-1}) = 0$.
3. If \mathcal{B}_i is an ordered basis of W_i , then $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_k)$ is an ordered basis of $W_1 + \dots + W_k$.

Proof. (1 \Rightarrow 2) : Take any $\alpha \in W_j \cap (W_1 + \dots + W_{j-1})$. Then $\alpha = \alpha_j = \alpha_1 + \dots + \alpha_{j-1}$. Then we get by independence that each α_i above should be zero.

(2 \Rightarrow 3) : Since $W_1 \cap W_2 = 0$, therefore \mathcal{B}_1 and \mathcal{B}_2 are disjoint, so $(\mathcal{B}_1, \mathcal{B}_2)$ forms a basis for $W_1 + W_2$. Keep using the given relation and at the end we will get that $(\mathcal{B}_1, \dots, \mathcal{B}_k)$ is a basis for $W_1 + \dots + W_k$.

(3 \Rightarrow 1) : Take $0 = \alpha_1 + \dots + \alpha_k$, where $\alpha_i \in W_i$. Therefore α_i is a linear combination of vectors from \mathcal{B}_i , so that sum of linearly independent basis vectors is 0 and hence each scalar is zero and we are done. □

If $\{W_i\}_{i=1}^k$ is a collection of subspaces satisfying any of the above criterion, then $W_1 + \dots + W_k$ is called **internal direct sum** and is denoted as $\bigoplus_{i=1}^k W_i$.

2.6.1 Projection operators

A **projection operator** E is literally just an idempotent operator, $E^2 = E$. There are some obvious properties of E .

Lemma 2.6.3. Let V be a f.d. F -vector space and let E be a projection:

1. If $\beta \in \text{Image}(E)$, then $E\beta = \beta$.
2. $V = \text{Image}(E) \oplus \text{Ker}(E)$.
3. $I - E$ is also a projection.

4. $I + E$ is invertible with inverse being $I - \frac{1}{2}E$.

Proof. $\alpha = \alpha - E\alpha + E\alpha$. Rest is left to the reader. \square

Ok, so we now notice that each decomposition of V into $R \oplus N$ comes from image and kernel of some unique projection.

Lemma 2.6.4. Let V be a f.d. F -vector space. If $V = R \oplus N$, then there exists a unique projection $E \in L(V)$ such that $R = \text{Image}(E)$ and $N = \text{Ker}(E)$.

Proof. Consider $E : V \rightarrow V$ such that $v = r + n \mapsto r$. Clearly, $\text{Image}(E) = R$ and $\text{Ker}(E) = N$. \square

We then have the first important theorem:

Theorem 2.6.5. Let V be a f.d. F -vector space. Then, $V = W_1 \oplus \cdots \oplus W_k$ for some subspaces $W_i \subseteq V$ if and only if there exists projections $E_1, \dots, E_k \in L(V)$ such that

1. $I = E_1 + \cdots + E_k$,
2. $E_i E_j = 0$ for $i \neq j$.

Proof. (L \Rightarrow R) : Define $E_i : V \rightarrow V$ which maps $\alpha_1 + \cdots + \alpha_k \mapsto \alpha_i$. The problem of well-definedness is handled by Lemma 2.6.1. Clearly, $E_i^2 = E_i$. Moreover, for $i \neq j$, $E_i E_j(\alpha) = E_i(\alpha_j) = 0$ as the i^{th} component of α_j is 0. So $E_i E_j = 0$. Finally, for any $\alpha \in V$, $E_i \alpha = \alpha_i$. Therefore $\alpha = \alpha_1 + \cdots + \alpha_k = E_1 \alpha + E_k \alpha = (E_1 + \cdots + E_k) \alpha$. Hence $E_1 + \cdots + E_k = I$.

(R \Rightarrow L) Suppose we have projections $E_1, \dots, E_k \in L(V)$ such that $E_1 + \cdots + E_k = I$ and $E_i E_j = 0$ for $i \neq j$. Now denote $W_i = \text{Image}(E_i) \subseteq V$. We claim that W_i 's are independent and moreover $W_1 + \cdots + W_k = V$. First let us show they are independent. So take any $\alpha_i \in W_i$ such that $\alpha_1 + \cdots + \alpha_k = 0$. Then there are $v_i \in V$ for each $i = 1, \dots, k$ such that $E_i v_i = \alpha_i$. Then we have $E_1 v_1 + \cdots + E_k v_k = 0$, and so applying E_2 for example gives us $0 + E_2 E_2 v_2 + 0 + \cdots + 0 = 0$ and thus $\alpha_2 = E_2 v_2 = 0$. Similarly, all α_i 's are zero. So W_i 's are indeed independent. Next we need to show that W_i 's span the whole space V . This is simple because for any $v \in V$, $v = E_1 v + \cdots + E_k v$. So indeed W_i 's span the whole space V . \square

Finally, let us end with the following independent lemma regarding projections, that they all are diagonalizable:

Lemma 2.6.6. Let $E \in L(V)$ be a projection where V is a f.d. F -vector space. Then E is diagonalizable.

Proof. Let E be a projection. Then $x^2 - x \in F[x]$ is an annihilator of E because $E^2 - E = 0$. Therefore the minimal polynomial $p(x) | x(x - 1)$. In any case, $p(x)$ will be a product of disjoint linear factors and therefore E is diagonalizable using Theorem 2.3.3. \square

2.7 Invariant direct sum decomposition

We have been focusing on arbitrary direct sum decomposition in the last section. Here we don't look at any direct sum decomposition of V , but those which are all invariant subspaces of some linear operator T . This is quite natural to seek because we know that a diagonalizable operator T induces a decomposition of V (into just the eigenspaces of T) in which each subspace is T -invariant.

So in-turn it is natural to ask whether if for a operator T , there is an direct sum decomposition of V where each subspace is T -invariant, does that make T diagonalizable? So that is why seeking *invariant* direct sum decomposition is important, and thus justifies the time that this section demands.

So suppose $T \in L(V)$ and we have a direct-sum decomposition of V as $V = W_1 \oplus \cdots \oplus W_k$. If each W_i is T -invariant, then $T_i := T|_{W_i} : W_i \rightarrow W_i$ is such that $\forall \alpha \in V$ with decomposition $\alpha = \alpha_1 + \cdots + \alpha_k$, we get $T\alpha = T\alpha_1 + \cdots + T\alpha_k = T_1\alpha_1 + \cdots + T_k\alpha_k$. Therefore we actually get a decomposition of T as $T = T_1 + \cdots + T_k$. Under these conditions we thus define $T = T_1 + \cdots + T_k$ as the **direct sum decomposition of T w.r.t. $V = \bigoplus_{i=1}^k W_i$** . Moreover, we then denote the decomposition into W_i 's as a **T -invariant decomposition of V** .

Ok, now observe that if we have a T -invariant direct sum decomposition of V as $V = W_1 \oplus \cdots \oplus W_k$ and moreover \mathcal{B}_i is a basis of W_i , then two things happens. First, $(\mathcal{B}_1, \dots, \mathcal{B}_k)$ is a basis of V . Second $A_i = [T_i]_{\mathcal{B}_i}$, where $T_i = T|_{W_i}$ combines to give

$$[T]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}.$$

That is, we get that representation of T can be achieved as a block diagonal form and we can thus reduce to analyzing the blocks themselves. In other words, having a T -invariant direct sum decomposition reduces the task of analyzing the operator T to it's components T_i .

Ok, so we saw in Theorem 2.6.5 that whether we may talk about a direct sum decomposition or equivalently projections with two more conditions, we will be doing the same thing. What we now like to know is how does a T -invariant decomposition of V turns out to be in terms of the corresponding projections. This is the goal of the following theorem:

Theorem 2.7.1. Let $f : X \rightarrow Y$ where V is f.d. F -vector space. Let $V = W_1 \oplus \cdots \oplus W_k$ and let E_1, \dots, E_k be the projections corresponding to each W_i by Theorem 2.6.5⁴. Then, each subspace W_i is T -invariant if and only if each E_i commutes with T .

Proof. (L \Rightarrow R) Suppose W_i 's form a T -invariant direct sum decomposition of V . We need to show that all of the associated projections commute with T . So first recall how E_i was defined; $E_i : V \rightarrow W_i$ maps as $\alpha = \alpha_1 + \cdots + \alpha_k \mapsto \alpha_i$. Therefore, for any $\alpha \in V$, we have $TE_i\alpha = T(\alpha_i) \in W_i$ as W_i is T -invariant and $W_i = \text{Image}(E_i)$. Next $E_iT(\alpha) = E_i(T\alpha_1 + \cdots + T\alpha_k)$. But since each W_i is T -invariant, therefore $T\alpha_i \in W_i$. Since $E_i\beta = 0$ for any $\beta \notin W_i$, therefore we get $E_iT(\alpha) = E_iT\alpha_i$. Since $E_i|_{W_i} = \text{id}_{W_i}$, therefore we further get that $E_iT\alpha_i = T\alpha_i$. Thus indeed $TE_i\alpha = E_iT\alpha$ and thus $E_iT = TE_i$. Note that we implicitly used above that $E_iE_j = 0$ for $i \neq j$ and also the fact that $E_i^2 = E_i$.

(R \Rightarrow L) Suppose the projections E_i corresponding to the decomposition all indeed commutes with T . Then we need to show that each $W_i = \text{Image}(E_i)$ is T -invariant. So take any $\alpha \in W_i =$

⁴which we hitherto call the **projections associated to decomposition $V = W_1 \oplus \cdots \oplus W_k$** .

$\text{Image}(E_i)$. Then there exists $\beta \in V$ such that $E_i\beta = \alpha$. Thus $T\alpha = TE_i\beta = E_iT\beta \in \text{Image}(E_i) = W_i$. Hence done. \square

Using the above results, we can finally give the main theorem of direct-sum decomposition, which characterizes diagonalizable operators in terms of existence of projections of the form as mentioned in Theorem 2.6.5, with an extra bell and whistle.

Theorem 2.7.2. Let V be a f.d. F -vector space. Then, we have the following two conclusions, which also contains each others' converse:

1. If $T \in L(V)$ is a diagonalizable operator with distinct eigenvalues $c_1, \dots, c_k \in F$, then there exists non-zero projections E_1, \dots, E_k such that
 - (a) $T = c_1E_1 + \dots + c_kE_k$,
 - (b) $I = E_1 + \dots + E_k$,
 - (c) $E_iE_j = 0$ for $i \neq j$.
 - (d) $\text{Image}(E_i) = \text{Ker}(T - c_iI)$ for all $i = 1, \dots, k$.
2. If there exists distinct $c_1, \dots, c_k \in F$ and projections E_1, \dots, E_k such that
 - (a) $T = c_1E_1 + \dots + c_kE_k$,
 - (b) $I = E_1 + \dots + E_k$,
 - (c) $E_iE_j = 0$ for $i \neq j$,
 then there exists a diagonalizable operator $T \in L(V)$ whose all distinct eigenvalues are exactly c_i 's and $\text{Image}(E_i)$'s are exactly the eigenspaces of T .

Proof. 1. Suppose T is diagonalizable. Then $V = \bigoplus_{i=1}^k \text{Ker}(T - c_iI)$. By Theorem 2.6.5, we have the corresponding projections $E_i : V \rightarrow V$ for each $i = 1, \dots, k$, which already satisfies (b) and (c). Now take any $v \in V$. Then $v = v_1 + \dots + v_k$ where $v_i \in \text{Ker}(T - c_iI)$. Therefore $Tv = Tv_1 + \dots + Tv_k = c_1v_1 + \dots + c_kv_k$.

2. Suppose we have k many projections as mentioned in the second condition. We need to show that the operator T is diagonalizable. First we should materialize how projections E_i are actually connected to operator T . This is done by the condition (a), because $c_i \in F$ are all distinct eigenvalues of T . So first thing that we do is to show somehow that each eigenspace of T is non-zero:

$$\text{Act. 1 : } \forall i = 1, \dots, k, \text{Ker}(T - c_iI) \neq \{0\}.$$

To see this, we need to show that T defined as $T = c_1E_1 + \dots + c_kE_k$ will have each subspace $\text{Ker}(T - c_iI)$ as non-zero. We need only show how $\text{Ker}(T - c_1I) \neq \{0\}$. Since for all $v \in V$, $Tv = c_1E_1v + \dots + c_kE_kv$ and since $TE_i = c_iE_i^2 = c_iE_i$. Moreover, $E_iT = c_iE_i^2 = c_iE_i$, therefore T and E_i commutes with each other. Notice that $E_i^2 = E_i$ follows from the given conditions on E_i . Now, for any $v \in V$, $E_1v \in V$ is such that $T(E_1v) = E_1(Tv) = E_1(c_1E_1v + \dots + c_kE_kv) = c_1E_1^2v = c_1E_1v$. Therefore $\text{Image}(E_1) \subseteq \text{Ker}(T - c_1I)$ and since E_1 is non-zero, therefore $\text{Ker}(T - c_1I) \neq \{0\}$. This concludes act 1.

$$\text{Act. 2 : } c_1, \dots, c_k \in F \text{ are ALL eigenvalues of } T.$$

So let $c \in F$ be any eigenvalue of T . We need to show that there exists $i = 1, \dots, k$ such that $c = c_i$. In particular, we also need to show that there exists $\alpha \in V$ such that $T\alpha = c\alpha$ and $\alpha \in \text{Ker}(T - c_iI)$. How do we do that? Since c is an eigenvalue of T therefore $\text{Ker}(T - cI) \neq \{0\}$. That is there exists $\alpha \in V$ which is non-zero such that $T\alpha = c\alpha$. Then we observe that $\alpha =$

$E_1\alpha + \dots + E_k\alpha$, so applying T to it we get $T\alpha = TE_1\alpha + \dots TE_k\alpha$. But $T\alpha = c_1E_1\alpha + \dots + c_kE_k\alpha$ as well. Therefore $0 = (T - c_1I)E_1\alpha + \dots + (T - c_kI)E_k\alpha$. Now by Theorem 2.6.5, we get that $W_i = \text{Image}(E_i)$ forms a direct sum decomposition of V . Therefore W_i 's are independent and therefore in the above expression, all $(T - c_iI)E_i\alpha = 0$. Since E_i and T commutes, therefore $(T - c_iI)E_i\alpha = E_i((T - c_iI)\alpha) = E_i(T\alpha - c_i\alpha) = E_i(c\alpha - c_i\alpha) = (c - c_i)E_i\alpha = 0$ for all $i = 1, \dots, k$. Now that's a very clear indication that we are on right path. Ok, so next we see that it cannot happen that all $E_i\alpha = 0$, because $I = E_1 + \dots + E_k$ and $\alpha \neq 0$. Thus, there exists $j = 1, \dots, k$ such that $E_j\alpha \neq 0$. But then $(c - c_j)E_j\alpha = 0$, so $c = c_j$. This concludes act 2.

Act. 3 : $V = \bigoplus_{i=1}^k \text{Ker}(T - c_iI)$.

Since in Act 1, we showed that $\text{Image}(E_i) \subseteq \text{Ker}(T - c_iI)$, and since Theorem 2.6.5 tells us that $\bigoplus_{i=1}^k \text{Image}(E_i) = V$, therefore $V = \bigoplus_{i=1}^k \text{Ker}(T - c_iI)$ where we know that eigenspaces are always independent. This concludes act 3.

We have by now shown that T is diagonalizable. We now need only show that E_i 's are exactly the eigenspaces of T . To do this, we recall that in Act 1 above, we already showed that $\text{Image}(E_i) \subseteq \text{Ker}(T - c_iI)$. So we need to show the converse now. It's enough to do this for $i = 1$. Ok, so take any $\alpha \in \text{Ker}(T - c_1I)$. We have $T\alpha = c_1\alpha$. We need to show that $\alpha \in \text{Image}(E_1)$. So again, we first write $T\alpha = c_1E_1\alpha + c_2E_2\alpha + \dots + c_kE_k\alpha = c_1\alpha$ and then using $c_1\alpha = c_1E_1\alpha + \dots + c_1E_k\alpha$, we again get (as a similar situation happened in Act 2), that $(c_2 - c_1)E_2\alpha + \dots + (c_k - c_1)E_k\alpha = 0$. Again, using Theorem 2.6.5, we get that each $(c_i - c_1)E_i\alpha = 0$. But since $c_i \neq c_1$ for all $i \neq 1$, therefore $E_i\alpha = 0$ for all $i \neq 1$. Since $\alpha = E_1\alpha + \dots + E_k\alpha$ therefore $\alpha = E_1\alpha$. We are now (quite tiredly), done. \square

We end this section with a good observation about the structure of the projections associated to a diagonalizable operator:

Lemma 2.7.3. Let $T \in L(V)$ be a diagonalizable operator where V is an F -vector space. Let $c_1, \dots, c_k \in F$ be the distinct eigenvalues of T and let E_1, \dots, E_k be the projections of V corresponding to the diagonalizable operator T obtained by Theorem 2.7.2. Then, for all $g(x) \in F[x]$, $g(T) = g(c_1)E_1 + \dots + g(c_k)E_k$.

Proof. Consider $g(x) = x^2$, for example. Then $T^2 = (c_1E_1 + \dots + c_kE_k)^2 = c_1^2E_1^2 + \dots + c_k^2E_k^2 +$ other terms of form E_iE_j for $i \neq j$, where these other terms will be zero by the fact that $E_iE_j = 0$ for $i \neq j$. So we will only be left with $c_1^2E_1 + \dots + c_k^2E_k$. A similar thing happens for T^r , where to write formally, you will have to use the multinomial theorem. Rest is easy. \square

Lecture # 7, 05/09/2022

2.8 Primary decomposition

The general philosophy of our approach to try to characterize all linear operators on a finite dimensional vector space has been to find a characterization of a class of operators in the terms of some algebraic objects, namely polynomials. In all of these discussions we always worked with just the right examples and situations where indeed the polynomial splits nicely over the base field into linear factors. But, if you arbitrarily construct a matrix, then more often than not you will find it to have a very complicated minimal and or characteristic polynomial; in particular, it may not be

able to split into linear factors over your base field. These cases have been nicely neglected so far in our goal of characterizing linear operators over a finite dimensional vector space. We now bring these cases to light and prove the most important theorem therein.

Theorem 2.8.1. (*Primary decomposition theorem*) Let V be a finite dimensional F -vector space and let $T \in L(V)$ be a linear operator. If the minimal polynomial $p(x) \in F[x]$ of T is given by

$$p(x) = p_1(x)^{r_1} \cdots p_k(x)^{r_k}$$

where $p_i(x) \in F[x]$ are all monic irreducible polynomials, then:

1. we have a direct sum decomposition

$$V = \text{Ker}(p_1(T)^{r_1}) \oplus \cdots \oplus \text{Ker}(p_k(T)^{r_k}),$$

2. the above decomposition is a T -invariant direct sum decomposition,
3. the minimal polynomial of the operator $T_i := T|_{\text{Ker}(p_i(T)^{r_i})}$ is exactly $p_i(x)^{r_i}$, for all $i = 1, \dots, k$.

Proof. See Theorem 12, [HK71], pp 220. □

Let us now look at some corollaries of the primary decomposition.

Corollary 2.8.2. Suppose $T \in L(V)$ is a linear operator satisfying the hypothesis of Theorem 2.8.1. If $U \in L(V)$ commutes with T , then $\text{Ker}(p_i(T)^{r_i})$ is U -invariant for all $i = 1, \dots, k$.

Proof. By Theorem 2.8.1, we have that $V = \bigoplus_{i=1}^k \text{Ker}(p_i(T)^{r_i})$ is a T -invariant direct sum decomposition. Now denote the corresponding projections given by Theorem 2.6.5 as E_1, \dots, E_k . By Theorem 2.7.1, we also have that E_i and T commutes. Since U and T commutes, therefore E_i and U commutes. But again using Theorem 2.7.1, we observe that since E_i and U commutes, therefore the decomposition $V = \bigoplus_{i=1}^k \text{Ker}(p_i(T)^{r_i})$ is U -invariant. □

Corollary 2.8.3. Let $T \in L(V)$ be a linear operator on a finite dimensional F -vector space V . If minimal polynomial of T is of the form

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

and E_1, \dots, E_k denotes the projections corresponding to the decomposition of V as given by Theorem 2.8.1, then, denoting $D = c_1 E_1 + \cdots + c_k E_k$, we have that $N = T - D$ is a nilpotent operator.

Proof. The corresponding projections, by Theorem 2.6.5, follows $I = E_1 + \cdots + E_k$. Therefore $T = TE_1 + \cdots + TE_k$. Now, $N = T - D = (T - c_1 I)E_1 + \cdots + (T - c_k I)E_k$. Now, for clarity, consider $(T - D)^2 = (T - c_1 I)^2 E_1 + \cdots + (T - c_k I)^2 E_k$ and the rest is zero because $E_i E_j = 0$ for $i \neq j$. Similarly, we can say $T^r = (T - c_1 I)^r E_1 + \cdots + (T - c_k I)^r E_k$. Now, if we take $r \geq r_i$ for all $i = 1, \dots, k$, then for any $\alpha \in V$, $T^r \alpha = \sum_{i=1}^r (T - c_i I)^r E_i(\alpha)$, where $\text{Image}(E_i) = \text{Ker}((T - c_i I)^{r_i})$. Since $r \geq r_i$ for each i , therefore each term of the above sum is 0. We are done. □

In some sense, the following is the actual primary decomposition theorem, as it tells you that every operator whose minimal polynomial is split over base field can be broken down into a diagonal and a nilpotent part, and that too in a unique way.

Theorem 2.8.4. Let V be a finite dimensional F -vector space and $T : V \rightarrow V$ be a linear operator whose minimal polynomial is a product of linear factors (is split over F). Then, there exists a unique diagonalizable operator D and a unique nilpotent operator N such that $T = D + N$ and $DN = ND$.

Proof. See Theorem 13, [HK71], pp 222. □

Lecture # ??, ??/??/????

3 Rational & Jordan forms

[Put motivating stuff here and link the main theorems, i.e. cyclic decomp. thm.]

Let V be a finite dimensional F -vector space and $T \in L(V)$. Suppose $\alpha \in V$ is given to us. In relation to α and T and V we can ask the following question: *What is the smallest T -invariant subspace of V which contains α ?* Well, of-course the answer is the intersection of all T -invariant subspaces containing α . But we can get more detailed than that:

Lemma 3.0.1. Let V be a f.d. F -vector space, $T \in L(V)$ and $\alpha \in V$. Denote the smallest T -invariant subspace of V which contains α to be $Z(\alpha; T)$. Then,

$$Z(\alpha; T) = \{g(T)\alpha \in V \mid g(x) \in F[x]\}.$$

Proof. Since $Z(\alpha; T)$ is T -invariant and contains α by definition, therefore for all $n \in \mathbb{N}$, $T^n\alpha \in Z(\alpha; T)$. This implies that $g(T)\alpha \in Z(\alpha; T)$ for all $g(x) \in F[x]$. The set of all $g(T)\alpha$ for all $g(x) \in F[x]$ naturally is a subspace of V which moreover contains α because $x \in F[x]$ and it is T -invariant because $T(g(T)\alpha) = f(T)\alpha$ where $f(x) = xg(x) \in F[x]$. \square

3.1 Cyclic subspaces and cyclic vectors

With the above Lemma 3.0.1, we define for an operator $T \in L(V)$ and $\alpha \in V$ the **T -cyclic subspace of V generated by α** to be the smallest T -invariant subspace containing α , that is, just $Z(\alpha; T)$. If it happens that $Z(\alpha; T) = V$, then one calls the vector $\alpha \in V$ to be a **cyclic vector for T** or that α is a **T -cyclic vector of V** .

In this subsection we will study some of the equivalent conditions for a linear operator to have a cyclic vector. In this regard, Theorem 3.1.1, Corollary 3.1.2 and Theorem 3.1.4 are important.

Note that $Z(\alpha; T)$ is the span of all $T^i\alpha$ for all positive integer i . One then realizes that $Z(\alpha; T)$ is a subspace of V and since V is finite dimensional, then so is $Z(\alpha; T)$. This tells us that there is a finite basis of $Z(\alpha; T)$. So does there exists a basis of $Z(\alpha; T)$ consisting of vectors $\alpha, T\alpha, T^2\alpha, \dots, T^n\alpha$? If yes, then where does this n comes from? We answer these and more, by, of-course, utilizing the algebra in the finitely generated reduced F -algebra $F[x]^5$. In particular, we go from the subspace $Z(\alpha; T)$ to an ideal, which is literally the T -conductor of α into 0, but we would like to pretend that we never studied that so we give the following definition. Let $\alpha \in V$ and $T \in L(V)$. We define the **T -annihilator of α** to be the ideal

$$M(\alpha; T) := \{g(x) \in F[x] \mid g(T)\alpha = 0\}.$$

This is literally just $S_T(\alpha; 0)$ by the way. Since $F[x]$ is a PID and $M(\alpha; T)$ is an ideal, therefore there exists a polynomial $p_\alpha(x) \in F[x]$ such that

$$M(\alpha; T) = \langle p_\alpha(x) \rangle. \tag{2}$$

⁵I'm sorry I had to do that.

We call this $p_\alpha(x)$ to be the T -**annihilator** of α as well by heavy abuse (of notation).

Now because of the trivial observation that $M(\alpha; T) \supseteq \text{Ann}(T)$, we get that $p_\alpha(x)|p(x)$ where $\langle p(x) \rangle = \text{Ann}(T)$.

It turns out that the questions we asked above are answered by the T -annihilator of α , $p_\alpha(x)$, via the following theorem. This theorem tells us the relation between the T -annihilator $p_\alpha(x)$ of $\alpha \in V$ and the cyclic subspace $Z(\alpha; T) \subseteq V$.

Theorem 3.1.1. Let V be a finite dimensional F -vector space and $T : V \rightarrow V$ be a linear operator. Let $\alpha \in V$ be a non-zero vector and consider $M(\alpha; T) = \langle p_\alpha(x) \rangle$. Then,

1. $\deg p_\alpha(x) = \dim Z(\alpha; T)$.
2. If $\deg p_\alpha(x) = k$, then $\alpha, T\alpha, T^2\alpha, \dots, T^{k-1}\alpha$ forms a basis of $Z(\alpha; T)$.
3. If $U := T|_{Z(\alpha; T)}$, then the minimal polynomial of U is $p_\alpha(x)$.

Proof. We prove statements 1 and 2 together. In particular, we will show that if $\deg p_\alpha(x) = k$, then $\alpha, T\alpha, \dots, T^{k-1}\alpha$ forms a basis of Z . For this, consider any polynomial $g(x) \in F[x]$. Then by division with $p_\alpha(x)$, we get $g(x) = p_\alpha(x)q(x) + r(x)$ where $\deg r(x) < k$. Since by Lemma 3.0.1, we get $g(T)\alpha \in Z(\alpha; T)$, so $g(T)\alpha = p_\alpha(T)q(T)\alpha + r(T)\alpha = q(T)p_\alpha(T)\alpha + r(T)\alpha = r(T)\alpha$ as $p_\alpha(x) \in M(\alpha; T)$. So for any $g(T)\alpha \in Z(\alpha; T)$, we get that there is a polynomial $r(x) \in F[x]$ of degree less than k such that $g(T)\alpha = r(T)\alpha \in Z(\alpha; T)$. Hence, the set $\{\alpha, T\alpha, T^2\alpha, \dots, T^{k-1}\alpha\}$ spans the whole $Z(\alpha; T)$. Next, we need to show it's linear independence. Suppose $c_i \in F$ such that $c_0\alpha + c_1T\alpha + c_2T^2\alpha + \dots + c_{k-1}T^{k-1}\alpha = 0$. We thus have that $g(x) = c_0 + c_1x + c_2x^2 + \dots + c_{k-1}x^{k-1} \in F[x]$ is a polynomial such that $g(T)\alpha = 0$. This means that $g(x) \in M(\alpha; T)$. Hence $p_\alpha(x)|g(x)$. But, $\deg p_\alpha(x) = k$ whereas $\deg g(x) = k-1$. So this implies $g(x) = 0$ which is equivalent to saying that $c_i = 0$ for all $i = 0, 1, \dots, k-1$. Hence $\{\alpha, T\alpha, T^2\alpha, \dots, T^{k-1}\alpha\}$ is a basis for $Z(\alpha; T) \subseteq V$.

To prove 3, consider any polynomial $g(x) \in \text{Ann}(U)$. Then $g(U)v = 0$ for all $v \in Z(\alpha; T)$. If $v \in Z(\alpha; T)$, then there is a polynomial $h(x) \in F[x]$ such that $v = h(T)\alpha$. Hence $0 = g(U)v = g(U)h(T)\alpha = g(T)h(T)\alpha$. Since $M(\alpha; T) = \langle p_\alpha(x) \rangle$, therefore $p_\alpha(x)|g(x)h(x)$. So set $v = \alpha \in Z(\alpha; T)$. Then, $h(x) = 1$. Thus, $p_\alpha(x)|g(x)$. Hence we are done. \square

A simple corollary of the above theorem is the following:

Corollary 3.1.2. Let V be a f.d. F -vector space and $T \in L(V)$. If $\alpha \in V$ is a T -cyclic vector for V , then the characteristic and minimal polynomial of T are same.

Proof. We have that $Z(\alpha; T) = V$. By Theorem 3.1.1, 3, we get that the minimal polynomial of T is $p_\alpha(x)$, the T -annihilator of α . So $\text{Ann}(T) = \langle p_\alpha(x) \rangle$. By Theorem 3.1.1, 1, we see that $\deg p_\alpha(x) = \dim V$. Since $\Phi(x) \in \text{Ann}(T)$, $\deg \Phi(x) = \dim V = \deg p_\alpha(x)$, $p_\alpha(x)|\Phi(x)$ and both $\Phi(x)$ and $p_\alpha(x)$ are monic, therefore $p_\alpha(x) = \Phi(x)$. \square

We will soon see that the above is also a sufficient condition for having T -cyclic vector for V .

In our goal of finding equivalent criterion for when a linear operator has a cyclic vector, we will now introduce another concept, which materializes in Theorem 3.1.4.

3.1.1 Companion matrix of a monic polynomial

We first begin with the following definition.

Definition 3.1.3. (Companion matrix) Let $p(x) \in F[x]$ be a monic polynomial given by

$$p(x) = c_0 + c_1x + \dots + c_{k-1}x^{k-1} + x^k.$$

Then, the companion matrix of $p(x)$ is defined to be the following $k \times k$ matrix

$$\text{Comp}(p(x)) := \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{k-1} \end{bmatrix}.$$

The main reason for introducing this matrix is because of the following theorem, which tells us that an operator has a cyclic vector if and only if it can be represented in some basis as the companion matrix of its minimal polynomial.

Theorem 3.1.4. Let V be a f.d. F -vector space and $T \in L(V)$ be a linear operator. Then, T has a cyclic vector if and only if there exists an ordered basis \mathcal{B} of V such that $[T]_{\mathcal{B}} = \text{Comp}(p(x))$ where $p(x) \in F[x]$ is the minimal polynomial of T .

Proof. (L \Rightarrow R) Suppose $\alpha \in V$ is a T -cyclic vector. Then by Theorem 3.1.1, 1 & 2, we have an ordered basis $\mathcal{B} := \{\alpha, T\alpha, T^2\alpha, \dots, T^{k-1}\alpha\}$ where $k = \deg p_{\alpha}(x)$ and $p_{\alpha}(x)$ is the T -annihilator of α which is also the minimal polynomial of T . Now, we claim that $[T]_{\mathcal{B}} = \text{Comp}(p_{\alpha}(x))$. For this, first write $\alpha_i = T^i\alpha$ for $i = 0, \dots, k-1$, thus $\alpha_0 = \alpha$. Then, $T\alpha_i = \alpha_{i+1}$ for each $i = 0, \dots, k-2$. We also have $T\alpha_{k-1} = T^k\alpha$. Now denote $p_{\alpha}(x) = c_0 + c_1x + \dots + c_{k-1}x^{k-1} \in F[x]$, so that $p_{\alpha}(T)\alpha = c_0\alpha + c_1T\alpha + \dots + c_{k-1}T^{k-1}\alpha + T^k\alpha = 0$, so that in particular we get $T^k\alpha = -c_0\alpha_0 - c_1\alpha_1 - \dots - c_{k-1}\alpha_{k-1}$. This proves that $[T]_{\mathcal{B}} = \text{Comp}(p_{\alpha}(x))$.

(R \Rightarrow L) Suppose that \mathcal{B} is a basis $\mathcal{B} = \{\alpha_0, \alpha_1, \dots, \alpha_{k-1}\}$ such that $[T]_{\mathcal{B}} = \text{Comp}(p(x))$. We need to show that T has a cyclic vector. We claim that α_0 is a T -cyclic vector. First we observe that $T\alpha_i = \alpha_{i+1}$ for $i = 0, \dots, k-2$. This follows from the structure of the companion matrix $\text{Comp}(p(x))$. We also have $T\alpha_{k-1} = T^k\alpha_0 = -c_0\alpha_0 - c_1\alpha_1 - \dots - c_{k-1}\alpha_{k-1} = -c_0\alpha_0 - c_1T\alpha_0 - \dots - c_{k-1}T^{k-1}\alpha_0$. So, we see that we can write the basis \mathcal{B} as $\mathcal{B} = \{\alpha_0, T\alpha_0, \dots, T^{k-1}\alpha_0\}$. Now since $Z(\alpha_0; T)$ is generated by all vectors of the form $\alpha_0, T\alpha_0, \dots, T^{k-1}\alpha_0, \dots$ and \mathcal{B} is a basis of V , so $V = Z(\alpha_0; T)$. \square

A natural question to ask about the companion matrices is that what are the characteristic and minimal polynomials of it and what is the connection between these two and the generating monic polynomial of the companion matrix. The following proposition answers that.

Proposition 3.1.5. Let F be a field. If $A \in M_n(F)$ is a companion matrix of some monic polynomial of degree n , say $p(x) \in F[x]$, then the characteristic and minimal polynomials of A are both $p(x)$.⁶

⁶Not sure if the instructor covered this, but this is how I proved this.

Proof. Let $\alpha = e_1 = (1, 0, 0, \dots, 0)$ in the standard basis \mathcal{B} of F^n . We claim that α forms a cyclic vector of A . To do this, we need only show that $\{\alpha, A\alpha, A^2\alpha, \dots\}$ generates F^n . This is simple to see as $A^i\alpha = e_{1+i}$ for all $i = 0, \dots, n-1$. Thus, $Z(\alpha; A) = F^n$ and hence A has a cyclic vector. Now, Theorem 3.1.1, 3, tells us that the minimal polynomial of A is $p_\alpha(x)$, the A -annihilator of $\alpha = e_1 \in V$. So we need only show that $p_\alpha(x)$ is equal to $p(x)$. Now notice that $Ae_j = e_{j+1}$ for $j = 1, \dots, n-1$ and $Ae_n = -\sum_{i=1}^n c_{i-1}e_i$. Now, it is a matter of a small calculation to see that $p(A)\alpha := p(A)e_1 = 0$ and it is at this point we use the specific structure of the A as the companion matrix of $p(x)$. Thus $p(x) \in M(\alpha; T)$ and thus $p_\alpha(x)|p(x)$. Conversely, if $g(x) \in M(\alpha; A)$, then $g(A)\alpha = 0$. Then, dividing $g(x)$ by $p(x)$ we obtain $g(x) = p(x)q(x) + r(x)$ where $\deg r(x) < \deg p(x) = n$. Now, $g(A)\alpha = p(A)q(A)\alpha + r(A)\alpha = 0$ which further implies $r(A)\alpha = 0$. Hence $r(x) \in M(\alpha; A)$. But by Theorem 3.1.1, 1, $\deg p_\alpha(x) = n$. But since $\deg r(x) < n$, therefore it implies that $\deg p_\alpha(x) < n$ as $p_\alpha(x)|r(x)$. This can only happen when $r(x) = 0$. Hence $p(x)|g(x)$. Thus $p(x)|p_\alpha(x)$. Hence we have $p_\alpha(x) = p(x)$. Since A has a cyclic vector, so by Corollary 3.1.2, $p(x)$ is equal to both the characteristic and minimal polynomials of A . \square

3.2 Cyclic decomposition theorem

In this section we will show the second decomposition theorem, after primary decomposition, in which we will decompose a f.d F -vector space into direct sum of cyclic subspaces and the corresponding cyclic vectors and their annihilators will satisfy the descending divisibility criterion. This will prove to be an important result as using it we would be able to prove results in Section 3.2.1 like the Corollary 3.2.5 which brings to light the importance of cyclic vectors and the Corollary 3.2.6 which generalizes the Cayley-Hamilton theorem. In this section, we also see the first major theorem which steps towards fulfilling the philosophy of the course, as mentioned in Remark 3.2.10. Unfortunately, due to the sheer size of the proof of this theorem, we would not do its proof here, which is not much enlightening.

We begin with the following definition.

Definition 3.2.1. (*T -admissible subspaces*) Let V be a f.d. F -vector space. A subspace $W \subseteq V$ is said to be T -admissible if

1. $W \subseteq V$ is T -invariant,
2. for any $\alpha \in V$ and $f(x) \in F[x]$, $f(T)\alpha \in W$ implies that $\exists \beta \in W$ such that $f(T)\alpha = f(T)\beta$.

Example. Let V be an F -vector space. The trivial subspace $W = \{0\}$ is a T -admissible subspace of V . The T -invariance of W is clear. Now suppose $v \in V$ and $f(x) \in F[x]$ is such that $f(T)v \in W$, i.e $f(T)v = 0$. Then, the only vector in W , which is 0, is such that $f(T)0 = f(T)v = 0$. Hence $W = \{0\}$ is indeed a T -admissible subspace of V .

Here's the main theorem.

Theorem 3.2.2. (*Cyclic decomposition theorem*) Let V be a f.d. F -vector space. If $W_0 \subsetneq V$ is a proper T -admissible subspace of V , then there exists non-zero vectors $\alpha_1, \dots, \alpha_r \in V$ such that

1. we have

$$V = W_0 \oplus Z(\alpha_1; T) \oplus Z(\alpha_2; T) \oplus \dots \oplus Z(\alpha_r; T),$$

2. if we denote $p_i(x) \in F[x]$ to be the T -annihilator of α_i for each $i = 1, \dots, r$, then

$$p_i(x) \mid p_{i-1}(x) \quad \forall i = 2, \dots, r$$

where we have that, in particular, $p_1(x)$, the T -annihilator of α_1 , is the minimal polynomial of T .

3. the integer r and the T -annihilators $p_1(x), \dots, p_r(x)$ are unique w.r.t properties 1., 2. and $\alpha_i \neq 0 \quad \forall i = 1, \dots, r$. We call these polynomials the **invariant factors of T** .

Proof. Theorem 3, p.p. 233 of [HK71]. □

Remark 3.2.3. It is important to note that one can keep $W_0 = \{0\}$ in order to get the cyclic decomposition of V with respect to T . In other words, every linear operator on a f.d. F -vector space has a cyclic decomposition.

3.2.1 Corollaries of cyclic decomposition

Let us in this section discuss some of the corollaries that one obtain from a clear understanding of Theorem 3.2.2. These are very interesting results and in particular throws light and concludes the constructions and questions raised in Section 3.1.

The first corollary completely characterizes the T -admissible subspaces of V .

Corollary 3.2.4. Let V be a f.d. F -vector space. Then, $W \subseteq V$ is a T -admissible subspace of V if and only if W is T -invariant and it has a complementary T -invariant subspace $W' \subseteq V$, i.e. $W \oplus W' = V$.

Proof. (R \Rightarrow L) Suppose that $W \subseteq V$ has a T -invariant complement as in $W \oplus W' = V$. Then for any $v \in V$, we have $v = w + w'$ where $w \in W$ and $w' \in W'$. Now take any $f(x) \in F[x]$, we then have $f(T)v = f(T)w + f(T)w'$. Since W and W' are T -invariant, therefore $f(T)w \in W$ and $f(T)w' \in W'$. Now, if $f(T)v \in W$, then $f(T)w + f(T)w' \in W$ and thus $f(T)w' \in W$. But since $W \cap W' = \{0\}$, hence $f(T)w' = 0$. Thus $f(T)v = f(T)w$ where $w \in W$.

(L \Rightarrow R) If $W = V$, then we are done. Otherwise, let $W \subsetneq V$. It is in this place that we use the cyclic decomposition theorem (Theorem 3.2.2). Since W is given to us to be a T -admissible subspace of V , then there is a decomposition $V = W \oplus \bigoplus_{i=1}^r Z(\alpha_i; T)$ for $\alpha_i \in V$. Since each $Z(\alpha_i; T)$ is T -invariant, hence so is their direct sum. So we define $W' := \bigoplus_{i=1}^r Z(\alpha_i; T)$ and call it a day. □

Next corollary is very important as it brings to conclusion the whole goal of Section 3.1 of finding an equivalent characterization of when an operator has a cyclic vector, and more.

Corollary 3.2.5. Let V be a f.d. F -vector space and let $T : V \rightarrow V$ be a linear operator. We have the following conclusions:

1. There exists an $\alpha \in V$ such that the T -annihilator of α is equal to the minimal polynomial of T .
2. T has a cyclic vector if and only if the characteristic and minimal polynomials of T are same.

Proof. 1. This is just Theorem 3.2.2, 2, in the special case when we choose $W_0 = \{0\}$, which is a valid proper T -admissible subspace of V .

2. (L \Rightarrow R) This is just Corollary 3.1.2.

(R \Rightarrow L) Suppose T is such that characteristic and minimal polynomials of T are equal. Let them be denoted by $\Phi(x)$ and $p(x)$ respectively. Now using Theorem 3.2.2 with $W_0 = \{0\}$ we can write $V = \bigoplus_{i=1}^r Z(\alpha_i; T)$. Then, we have

$$\begin{aligned} \dim V &= \deg \Phi(x) \\ &= \deg p(x) \\ &= \dim Z(\alpha_1; T). \end{aligned}$$

Thus $Z(\alpha_1; T) = V$ and hence T has a cyclic vector, namely α_1 appearing in the cyclic decomposition of T w.r.t $W_0 = \{0\}$. \square

We next study an important generalization of Cayley-Hamilton theorem.

Corollary 3.2.6. (*Generalized Cayley-Hamilton theorem*) Let V be a f.d. F -vector space and let $T : V \rightarrow V$ be a linear operator. Denote by $p(x)$ the minimal polynomial and by $\Phi(x)$ the characteristic polynomial of T . Then,

1. $p(x) \mid \Phi(x)$,
2. $p(x)$ and $\Phi(x)$ has same prime factors,
3. if $p(x) = f_1^{r_1}(x) \cdots f_k^{r_k}(x)$ is the prime factorization of $p(x)$, then $\Phi(x) = f_1^{d_1}(x) \cdots f_k^{d_k}(x)$ where

$$d_i = \frac{\dim \text{Ker}(f_i(T)^{r_i})}{\deg f_i(x)}.$$

Proof. To get 1., we first decompose $V = \bigoplus_{i=1}^r Z(\alpha_i; T)$ using Theorem 3.2.2, for non-zero $\alpha_i \in V$. Also denote by $p_i(x) \in F[x]$ the T -annihilator of $\alpha_i \in V$ where we have that $p_1(x)$ is the minimal polynomial of T , so $p(x) = p_1(x)$, and that $p_i(x) \mid p_{i-1}(x)$ for each $i = 2, \dots, r$. Now denote $U_i = T|_{Z(\alpha_i; T)}$ which is well-defined as any T -cyclic subspace of V is T -invariant. Now, by Theorem 3.1.1, 3, we have that the minimal polynomial of U_i is $p_i(x)$. Moreover, by Corollary 3.1.2 or more generally by Corollary 3.2.5, 2., we also get that the characteristic polynomial of U_i denoted by $\Phi_i(x)$ is equal to $p_i(x)$. Now we claim that if we denote the characteristic polynomial of T by $\Phi(x)$, then $\Phi(x) = \Phi_1(x) \cdots \Phi_r(x)$. To see this, take any basis \mathcal{B}_i of $Z(\alpha_i; T)$. Then, since each $Z(\alpha_i; T)$ are independent, therefore we get a basis of V as in $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_r)$. Now, in this basis, we get

$$[T]_{\mathcal{B}} = \begin{bmatrix} [U_1]_{\mathcal{B}_1} & 0 & \cdots & 0 \\ 0 & [U_2]_{\mathcal{B}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [U_r]_{\mathcal{B}_r} \end{bmatrix}.$$

Now, we see that the characteristic polynomial of $[T]_{\mathcal{B}}$ (and thus of T) is exactly $\Phi_1(x) \cdots \Phi_r(x)$. So this proves the claim. Next, since $p_1(x) = \Phi_1(x)$ thus divides $\Phi(x)$, and since $p_1(x)$ is exactly the minimal polynomial of T , hence 1. is done.

Now to conclude 2. is also a matter of few observations. First let $f(x) \in F[x]$ be any prime factor of $p(x)(= p_1(x))$. Then, it is a prime factor of $p_1(x) = \Phi_1(x)$ and hence a prime factor of $\Phi(x)$. Conversely, let $f(x)$ be any prime factor of $\Phi(x)$. Then it is a prime factor of some $\Phi_i(x) = p_i(x)$, thus $f(x)|p_i(x)$. But since $p_i(x)|p_{i-1}(x)$, hence, $f(x)|p_1(x)$ where we know that $p_1(x) = p(x)$. Hence 2. is done as well.

Finally, to conclude 3., we would require the help of an old friend, the primary decomposition theorem (Theorem 2.8.1). Since $p(x) = f_1^{r_1}(x) \dots f_k^{r_k}(x)$ where each $f_i(x)$ is irreducible, therefore by the primary decomposition theorem, we can decompose $V = \bigoplus_{i=1}^k \text{Ker}(f_i(T)^{r_i})$. Now, denote $T_i := T|_{\text{Ker}(f_i(T)^{r_i})}$, which is well-defined by Theorem 2.8.1, 2. By Theorem 2.8.1, 3, we also have that the minimal polynomial of T_i is $f_i(x)^{r_i}$. By statement 2. of this result which we just proved above, and applying it on $T_i : \text{Ker}(f_i(T)^{r_i}) \rightarrow \text{Ker}(f_i(T)^{r_i})$, we get that the characteristic polynomial $\Phi_i(x)$ has to be $f_i(x)^{d_i}$ where $d_i \geq r_i$. Moreover, we get that $\Phi(x) = \Phi_1(x) \dots \Phi_k(x)$. Now, since the degree of characteristic polynomial is equal to the dimension of the vector space, therefore

$$\begin{aligned} \deg \Phi_i(x) &= \dim \text{Ker}(f_i(T)^{r_i}) \\ d_i \cdot \deg f_i &= \dim \text{Ker}(f_i(T)^{r_i}) \\ d_i &= \frac{\dim \text{Ker}(f_i(T)^{r_i})}{\deg f_i}. \end{aligned}$$

This concludes the proof. \square

The proof Corollary 3.2.6, 1 also gives us the following statement, which is good to keep in mind.

Corollary 3.2.7. Let V be a f.d. F -vector space and let $T : V \rightarrow V$ be a linear operator. If in the cyclic decomposition of T , the invariant factors are $p_1(x), \dots, p_r(x)$, then the characteristic polynomial of T is

$$\Phi(x) = p_1(x) \dots p_r(x).$$

Proof. See proof of Corollary 3.2.6, 1. \square

3.2.2 Cyclic decomposition in terms of matrices

Let V be a f.d. F -vector space and let $T : V \rightarrow V$ be a linear operator. By cyclic decomposition (Theorem 3.2.2) on $W_0 = \{0\}$, we get that

$$V = \bigoplus_{i=1}^r Z(\alpha_i; T)$$

where let us denote the T -annihilator of α_i by $p_i(x) \in F[x]$. Now, consider the restricted operator $T_i := T|_{Z(\alpha_i; T)}$ on cyclic subspace $Z(\alpha_i; T)$. It is clear that $Z(\alpha_i; T)$ has a T_i -cyclic vector, namely α_i . Now, by Theorem 3.1.1, 2 and Theorem 3.1.4, we get several conclusions. Firstly, that $\mathcal{B}_i :=$

$\{\alpha_i, T\alpha_i, \dots, T^{k_i-1}\alpha_i\}$ is a basis of $Z(\alpha_i; T)$ where $k_i := \deg p_i(x) = \dim Z(\alpha_i; T)$ and $p_i(x)$ is the monic T -annihilator of α_i . Second, $p_i(x)$ is also the minimal polynomial of T_i . Third, we get that

$$A_i := [T_i]_{\mathcal{B}_i} = \text{Comp}(p_i(x)).$$

Now, denote $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_r)$ to be the basis of V formed by collating the bases of individual parts. Then, we get that

$$[T]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & 0 & \dots & 0 \\ 0 & A_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_r \end{bmatrix}$$

where keep in mind that $A_i = \text{Comp}(p_i(x))$ is a matrix of size $k_i \times k_i$, thus all A_i s are not necessarily of same size! Recall also that $p_i(x)|p_{i-1}(x)$ for all $i = 2, \dots, r$. We thus give the definition of rational form of a matrix.

Definition 3.2.8. (Rational form) Let F be a field and let A be an $n \times n$ matrix over the field F . If A can be written as

$$A = \begin{bmatrix} \text{Comp}(p_1(x)) & 0 & \dots & 0 \\ 0 & \text{Comp}(p_2(x)) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Comp}(p_r(x)) \end{bmatrix}$$

where $p_1(x), \dots, p_r(x)$ are non-scalar monic polynomials such that $p_i(x)|p_{i-1}(x)$ for all $i = 2, \dots, r$, then A is said to be in rational form.

Indeed, the above definition makes sense because for each $n \times n$ matrix A over field F can only be similar to a unique matrix in rational form, as the following theorem tells us.

Theorem 3.2.9. (Uniqueness of rational form) Let F be a field and let B be an $n \times n$ matrix over F . Then there exists a unique $n \times n$ matrix A in rational form such that B is similar to A . That is, in the monoid $M_n(F)$, each conjugacy class⁷ contains a unique matrix in rational form(!)

Proof. It is in this place that we realize the importance of the statement 3. of Theorem 3.2.2, and it follows straightforwardly from it, by keeping in mind the datum required for a matrix to be in rational form. \square

Remark 3.2.10. The Theorem 3.2.9 can be seen as the first major theorem in the direction of the philosophy of the course, which is the humble goal to "classify all linear operators on a f.d. vector space".

⁷Let G be a monoid and let $g, h \in G$, then the relation

$$g \sim h \iff \exists p \in G^\times \text{ s.t. } h = pgp^{-1}$$

is an equivalence relation on G , where $G^\times = \{g \in G \mid \exists h \in G, gh = hg = e\}$. We call each equivalence class a conjugacy class.

3.3 Jordan form

In rational forms, we found one answer to our goal of characterizing all linear operators on a f.d. F -vector space. In particular, one can gain knowledge of the rational form of an operator by various arguments about the dimension of the base vector space and the allowed sizes of the cyclic subspaces (which, along with the divisibility criterion of the respective T -annihilators, can often give you the few cases which the cyclic decomposition of an operator can have). However, we now study a form of an operator which is in some sense even simpler than rational form, albeit we would require more hypotheses on the operator⁸.

Construction 3.3.1. (*Jordan form of a nilpotent operator*) Let V be a n -dimensional F -vector space and $N : V \rightarrow V$ be a nilpotent operator. Let $p(x) = x^k$ and $\Phi(x) = x^n$ be the minimal and characteristic polynomials of N , respectively ($k \leq n$). The cyclic decomposition of N (Theorem 3.2.2) tells us

$$V = \bigoplus_{i=1}^r Z(\alpha_i; N)$$

where $\alpha_i \in V$ and denote $p_i(x) \in F[x]$ to be the T -annihilator of α_i . By divisibility criterion, we also get $p_1(x) = p(x) = x^k$ and $p_i(x) = x^{k_i}$ where $k_i \leq k_{i-1}$. That is,

$$k = k_1 \geq k_2 \geq \cdots \geq k_{r-1} \geq k_r.$$

With this, the rational form of N will be given by

$$\begin{aligned} \text{Rat}(N) &= \begin{bmatrix} \text{Comp}(p_1(x)) & 0 & \cdots & 0 \\ 0 & \text{Comp}(p_2(x)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \text{Comp}(p_r(x)) \end{bmatrix} \\ &= \begin{bmatrix} \text{Comp}(x^k) & 0 & \cdots & 0 \\ 0 & \text{Comp}(x^{k_2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \text{Comp}(x^{k_r}) \end{bmatrix} \end{aligned}$$

where

$$\text{Comp}(x^{k_i}) = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{k_i \times k_i}$$

for each $i = 1, \dots, r$. Hence, the conclusion of this is the following:

Rational form of a nilpotent operator is determined by natural r and r descending naturals $k_1 \geq \cdots \geq k_r$ which adds upto n .

⁸We also test out a new environment.

Question : How is r related to the nilpotent operator N ?

The answer is clear from the fact that $\text{Rat}(N)$ has exactly r zero columns and all other columns are independent, that is, the $\dim \text{Ker}(N) = r$.

We now, in some sense, conclude one main goal of the course by using everything we did in the past sections, in the following construction:

Construction 3.3.2. (*Jordan form of an arbitrary operator*) Let V be a f.d. F -vector space of $\dim V = n$ and let $T : V \rightarrow V$ be a linear operator. Further suppose that the characteristic polynomial of T is factored completely into linear factors in $F[x]$ as follows:

$$\Phi(x) = (x - c_1)^{d_1} \dots (x - c_k)^{d_k},$$

where $c_1, \dots, c_k \in F$ are all distinct eigenvalues of T . Hence by Corollary 3.2.6, we also get that the minimal polynomial of T is

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

where $r_i \leq d_i$ for each i .

In such a case, we first use the primary decomposition theorem (Theorem 2.8.1) in order to get the following decomposition of V :

$$V = \bigoplus_{i=1}^k \text{Ker}((T - c_i I)^{r_i}).$$

Denote by $W_i := \text{Ker}((T - c_i I)^{r_i})$ for each $i = 1, \dots, k$ and denote $T_i := T|_{W_i} : W_i \rightarrow W_i$. Now, consider the operator

$$N_i := T_i - c_i I : W_i \rightarrow W_i.$$

Note that the dimension of W_i is d_i . This follows by Corollary 3.2.6, 3.

By the primary decomposition, we also have that the minimal polynomial of $T_i = T|_{W_i}$ is equal to $(x - c_i)^{r_i}$. Since $N_i^{r_i} = 0$ on W_i , therefore N_i as defined above is a nilpotent operator on W_i and has minimal polynomial $x^{r_i} \in F[x]$. This is true for each $i = 1, \dots, k$. Now, by executing Construction 3.3.1 on $N_i : W_i \rightarrow W_i$, we get that there are following data attached to N_i :

1. the positive natural d_i , where $d_i = \dim W_i$,
2. the positive natural n_i , where $n_i := \dim \text{Ker}(N_i)$,
3. a sequence of naturals $r_i = k_1^{(i)} \geq k_2^{(i)} \geq \dots \geq k_{n_i}^{(i)}$ which add up to d_i ,
4. a basis \mathcal{B}_i of W_i , in which N_i is represented as

$$[N_i]_{\mathcal{B}_i} = \begin{bmatrix} \text{Comp}\left(x^{k_1^{(i)}}\right) & 0 & \dots & 0 \\ 0 & \text{Comp}\left(x^{k_2^{(i)}}\right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Comp}\left(x^{k_{n_i}^{(i)}}\right) \end{bmatrix}.$$

That is, we get for $N_i : W_i \rightarrow W_i$ the following tuple of information: $(d_i, n_i, \{k_j^{(i)}\}_{j=1}^{r_i}, \mathcal{B}_i)$. This is for each $i = 1, \dots, k$.

If $[N_i]_{\mathcal{B}_i}$ is as given above, then $[T_i]_{\mathcal{B}_i}$ is given by:

$$\begin{aligned}
 [T_i]_{\mathcal{B}_i} &= \begin{bmatrix} \text{Comp}(x^{k_1^{(i)}}) & 0 & \dots & 0 \\ 0 & \text{Comp}(x^{k_2^{(i)}}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Comp}(x^{k_{n_i}^{(i)}}) \end{bmatrix} + c_i I_{d_i} \\
 &= \begin{bmatrix} \text{Comp}(x^{k_1^{(i)}}) + c_i I_{k_1^{(i)}} & 0 & \dots & 0 \\ 0 & \text{Comp}(x^{k_2^{(i)}}) + c_i I_{k_2^{(i)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{Comp}(x^{k_{n_i}^{(i)}}) + c_i I_{k_{n_i}^{(i)}} \end{bmatrix}_{d_i \times d_i}.
 \end{aligned}$$

Now note that, for each $i = 1, \dots, k$, for each $j = 1, \dots, n_i$, we get that the block

$$\begin{aligned}
 J_j^{(i)} &:= \text{Comp}(x^{k_j^{(i)}}) + c_i I_{k_j^{(i)}} \\
 &= \begin{bmatrix} c_i & 0 & 0 & \dots & 0 & 0 \\ 1 & c_i & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c_i & 0 \\ 0 & 0 & 0 & \dots & 1 & c_i \end{bmatrix}_{k_j^{(i)} \times k_j^{(i)}}.
 \end{aligned}$$

So, we can write

$$[T_i]_{\mathcal{B}_i} = \begin{bmatrix} J_1^{(i)} & 0 & \dots & 0 \\ 0 & J_2^{(i)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{n_i}^{(i)} \end{bmatrix}_{d_i \times d_i}.$$

With this information, we can now construct a basis of V by gluing together all of the bases \mathcal{B}_i . Let the resultant basis be \mathcal{B} . Clearly, the matrix $[T]_{\mathcal{B}}$ will be given by:

$$[T]_{\mathcal{B}} = \begin{bmatrix} [T_1]_{\mathcal{B}_1} & 0 & \dots & 0 \\ 0 & [T_2]_{\mathcal{B}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & [T_k]_{\mathcal{B}_k} \end{bmatrix}_{n \times n}.$$

The matrix $[T]_{\mathcal{B}}$ is said to be the **Jordan form** of operator T .

Conversely, one defines a matrix to be in Jordan form as follows:

Definition 3.3.3. (Matrix in Jordan form) Let $A_{n \times n}$ be a matrix over field F . Then A is said to be in Jordan form if there exists

1. $k \in \mathbb{N}$,
2. $c_1, \dots, c_k \in F$,
3. $d_i \in \mathbb{N}$ for each $i = 1, \dots, k$
4. $n_i \in \mathbb{N}$ for each $i = 1, \dots, k$
5. sequence $k_1^{(i)} \geq k_2^{(i)} \geq \dots \geq k_{n_i}^{(i)}$ for each $i = 1, \dots, k$ such that

$$\sum_{j=1}^{n_i} k_j^{(i)} = d_i$$

such that

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}_{n \times n}$$

where

1. A_i is given by

$$A_i := \begin{bmatrix} J_1^{(i)} & 0 & \cdots & 0 \\ 0 & J_2^{(i)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{n_i}^{(i)} \end{bmatrix}_{d_i \times d_i}$$

2. $J_j^{(i)}$ is given by

$$J_j^{(i)} := \begin{bmatrix} c_i & 0 & \cdots & 0 & 0 \\ 1 & c_i & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & c_i \end{bmatrix}_{k_j^{(i)} \times k_j^{(i)}}.$$

Theorem 3.3.4. Let $T : V \Rightarrow V$ be a linear operator on a f.d. F -vector space. Then there exists a basis \mathcal{B} of V such that the matrix $[T]_{\mathcal{B}}$ is in Jordan form.

Proof. This is exactly the Construction 3.3.2. □

This concludes the Part 1 of the course, where we found that each conjugacy class of the monoid $L(V)$ has two unique matrices, one in rational form and one in Jordan form. Thus completing the initial goal we set out for us in the beginning of the course.

Part II

Inner product spaces

One of the main goals of this part is to answer the following question: *What condition should one require on an operator T on a \mathbb{C} -IPS in order to get that T is diagonalizable?* The answer is normality! This will be the main theorem of this part.

4 Inner products, orthonormal basis, functionals, adjoints and uniqueness

Since people already know what is an inner product, so we will breeze through some of the initial material here. Please note that at times our proofs in this part might become very sketchy, but that is to be expected by this long into the course. Let F be a field which is either \mathbb{R} or \mathbb{C} . Let V be an F vector space. An **inner product** structure on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow F$$

such that it satisfies the following three axioms:

1. (*Linearity in first argument*) $\forall \alpha, \beta, \gamma \in V$ and $c \in F$, we must have

$$\langle c\alpha + \beta, \gamma \rangle = c\langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle.$$

2. (*Conjugation symmetry*) $\forall \alpha, \beta \in V$, we must have

$$\langle \alpha, \beta \rangle = \overline{\langle \beta, \alpha \rangle}.$$

3. (*Positivity*) $\forall \alpha \neq 0$, we must have

$$\langle \alpha, \alpha \rangle > 0,$$

that is, if $\langle \alpha, \alpha \rangle = 0$, then $\alpha = 0$.

Denote by $\|\alpha\| := \langle \alpha, \alpha \rangle^{1/2}$. An important property to keep in mind while working with inner products is the polarisation identity.

Lemma 4.0.1. Let V be a \mathbb{C} -IPS, then for all $\alpha, \beta \in V$, we have

$$4\langle \alpha, \beta \rangle = \|\alpha + \beta\|^2 - \|\alpha - \beta\|^2 + i\|\alpha + i\beta\|^2 - i\|\alpha - i\beta\|^2.$$

The proof is straightforward. □

We also recall the usual Gram-Schmidt process to find an orthonormal basis of a f.d. \mathbb{C} -IPS.

Theorem 4.0.2. (*Gram-Schmidt process*) Let V be a f.d. \mathbb{C} -IPS. Then, there exists a basis \mathcal{B} of V consisting of orthogonal and unit norm vectors.

One can find its proof in any basic linear algebra textbook. It is good to keep in mind that using the above result, we can most of the times reduce down to calculations with an orthonormal collection of vectors, which is at times much, much simpler.

4.1 Linear functionals

A **linear functional** on an F -IPS V is a linear map $f : V \rightarrow F$.

Lemma 4.1.1. Let V be an F -IPS and let $\beta \in V$. Then,

$$\begin{aligned} f_\beta : V &\longrightarrow F \\ \alpha &\longmapsto \langle \alpha, \beta \rangle. \end{aligned}$$

is a linear functional. □

We see that on f.d. F -IPS, all linear functionals are of the form of Lemma 4.1.1. This is *baby* Riesz representation theorem.

Theorem 4.1.2. Let V be a f.d. F -IPS. If $f : V \rightarrow F$ is a linear functional, then there exists a unique $\beta \in V$ such that $f = f_\beta := \langle -, \beta \rangle$.

Proof. (Brief sketch) The main idea is to take an orthonormal basis of V by Theorem 4.0.2 and do back-calculation on what that $\beta \in V$ should be. □

4.2 Adjoint operators

The following theorem tells us that corresponding to each linear operator $T : V \rightarrow V$ on a f.d. F -IPS, there is a unique linear operator T^* which *switches* in the inner product.

Theorem 4.2.1. Let V be a f.d. F -IPS. If $T : V \rightarrow V$ is a linear operator, then there exists a unique linear operator $T^* : V \rightarrow V$ satisfying the following for all $\alpha, \beta \in V$:

$$\langle T\alpha, \beta \rangle = \langle \alpha, T^*\beta \rangle.$$

Proof. The idea is to construct a linear functional and then use the Theorem 4.1.2. In particular, we have the following functional for each $\beta \in V$, where f_β is as earlier

$$V \xrightarrow{T} V \xrightarrow{f_\beta} F.$$

Now, by Theorem 4.1.2, we get that there is a unique $\gamma_\beta \in V$ such that $f_\beta T = f_{\gamma_\beta}$. Now define $T^* : V \rightarrow V$ by $\beta \mapsto \gamma_\beta$. One can see it is linear and it is unique by uniqueness of γ_β . □

Remark 4.2.2. (Matrix of the adjoint) Once we know that each linear operator has a unique adjoint, it is natural to ask how for a basis \mathcal{B} , are the following two matrices related

$$[T]_{\mathcal{B}} \text{ \& } [T^*]_{\mathcal{B}}?$$

The answer is in the following result.

Proposition 4.2.3. Let V be a f.d. F -IPS and let $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ be any orthonormal basis of V .

1. If $A := [T]_{\mathcal{B}}$, then

$$A_{ij} = \langle T\alpha_j, \alpha_i \rangle.$$

2. The matrices $[T]_{\mathcal{B}}$ and $[T^*]_{\mathcal{B}}$ are related by:

$$[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*.$$

where $*$ of a matrix means the usual Hermitian conjugate.

Proof. Trivial and use 1. in 2. □

We see that an invariant subspace also behaves nicely with adjoints:

Lemma 4.2.4. Let V be a f.d. \mathbb{C} -IPS, $T : V \rightarrow V$ be a linear operator and $W \subseteq V$ be a linear subspace. Then

$$W \text{ is } T\text{-invariant} \iff W^\perp \text{ is } T^*\text{-invariant}.$$

Proof. This is almost a tautology, as if we take $\alpha \in W$ and $\beta \in W^\perp$, then $0 = \langle \alpha, T^*\beta \rangle = \langle T\alpha, \beta \rangle = 0$. □

4.2.1 Properties of adjoints on \mathbb{C} -IPS

Let us discuss the properties of adjoints of operators on a \mathbb{C} -IPS. The first one instructs us how to do algebra with them.

Proposition 4.2.5. Let V be a f.d. \mathbb{C} -IPS and let $T, U : V \rightarrow V$ be two operators and $c \in \mathbb{C}$. Then,

1. $(T + U)^* = T^* + U^*$,
2. $(cT)^* = \bar{c}T^*$,
3. $(TU)^* = U^*T^*$,
4. $T^{**} = T$.

Proof. In all of these, the main result that we will use is Theorem 4.2.1 which makes sure that each operator has a unique adjoint. With this, all of them are routine, but instructive. □

Remark 4.2.6. One of the ways one tries to understand the stuff about operators and its adjoints is by the following analogy:

Operators on a f.d. \mathbb{C} -IPS and Hermitian adjoints behaves as complex numbers and complex conjugates.

This is already testified in Proposition 4.2.5 and is further seen by the following lemma, which mimics how each complex number z can be written as $z = a + bi$ where $\bar{a} = a$ and $\bar{b} = b$ (i.e. $a, b \in \mathbb{R}$).

Lemma 4.2.7. Let $T : V \rightarrow V$ be a linear operator on a f.d. \mathbb{C} -IPS. Then, there exists operators $U_1, U_2 : V \rightarrow V$ such that

$$T = U_1 + iU_2$$

and $U_1 = U_1^*$ and $U_2 = U_2^*$.

Proof. Just consider the following candidates

$$U_1 = \frac{T + T^*}{2} \text{ \& } U_2 = \frac{T - T^*}{2i}$$

and these can be verified to satisfy the needed property. □

4.3 Normal & unitary operators - 1

Motivated by above lemma, we give the following two definitions, which will be the central theme of the conclusion of this part.

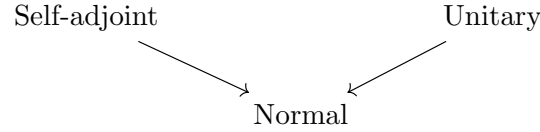
Definition 4.3.1. (Normal and unitary operators) Let V be an F -IPS and let $T : V \rightarrow V$ be a linear operator. Then, T is called normal if

$$T^*T = TT^*$$

and T is called unitary if

$$T^*T = TT^* = I.$$

Remark 4.3.2. Apart from normal and unitary, we can also define self-adjoint operators, which, as the name suggests, are required to be such that it and its adjoint are both actually equal. We will study all these three operators in great detail in the next few pages. For now, it is clear that



We'll slowly add more members to the above diagram.

We now build some of the tools that we may use later. It is usually the case while working with operators on a f.d. \mathbb{C} -IPS that we need to show that two operators are equal. In such a situation, a nice way of showing this is to somehow show that their difference, say T , is zero. One can do this especially by using the following lemma.

Lemma 4.3.3. Let V be a f.d. \mathbb{C} -IPS. If $\forall \alpha \in V, \langle T\alpha, \alpha \rangle = 0$, then $T = 0$.

Proof. Take any $\alpha, \beta \in V$. Now we have $0 = \langle T(\alpha + i\beta), \alpha + i\beta \rangle = -i\langle T\alpha, \beta \rangle + i\langle T\beta, \alpha \rangle$. We may write $\langle T\alpha, \beta \rangle = \langle T\alpha, \alpha + \beta \rangle$ as $\langle T\alpha, \alpha \rangle = 0$. Similarly, write $\langle T\beta, \alpha \rangle = \langle T\beta, \alpha + \beta \rangle$. We thus get that

$$\langle T\alpha, \beta \rangle = \langle T\beta, \alpha \rangle.$$

Now, do the same on $\langle T(\alpha + \beta), \alpha + \beta \rangle$ in order to get from both of them that $\langle T\alpha, \beta \rangle = 0$ for all $\alpha, \beta \in V$. Thus put $\beta = T\alpha$ to get rid of this. \square

Another tool is the following lemma which gives a characterization of normal operators.

Lemma 4.3.4. Let V be a f.d. \mathbb{C} -IPS. Then, T is normal if and only if $\forall \alpha \in V, \|T\alpha\| = \|T^*\alpha\|$.

Proof. (R \Rightarrow L) T is normal iff $TT^* - T^*T = 0 \Leftarrow \forall \alpha \in V, \langle (TT^* - T^*T)\alpha, \alpha \rangle = 0 \iff \langle TT^*\alpha, \alpha \rangle = \langle T^*T\alpha, \alpha \rangle \iff \langle T^*\alpha, T^*\alpha \rangle = \langle T\alpha, T\alpha \rangle$, and the last statement is exactly the hypothesis.

(L \Rightarrow R) If T is normal, then $\langle T\alpha, T\alpha \rangle = \langle T^*T\alpha, \alpha \rangle = \langle TT^*\alpha, \alpha \rangle = \langle T^*\alpha, T^*\alpha \rangle$. \square

A useful lemma about normal operators is that an "obvious" method of calculation is indeed true:

Lemma 4.3.5. Let $T : V \rightarrow V$ be a normal operator on a f.d. \mathbb{C} -IPS. Then $\lambda \in \mathbb{C}$ is an eigenvalue of T with eigenvector $v \in V$, if and only if $\bar{\lambda}$ is an eigenvalue of T^* with eigenvector $v \in V$. That is,

$$Tv = \lambda v \iff T^*v = \bar{\lambda}v.$$

Proof. This follows from Lemma 4.3.4 applied on $T - \lambda I$ which is also normal if and only if T is. It then tells us that $\text{Ker}(T - \lambda I) = \text{Ker}(T^* - \bar{\lambda}I)$. \square

One of the important fact about normal operators is that the eigenspaces are orthogonal.

Proposition 4.3.6. Let $T : V \rightarrow V$ be a linear operator on a f.d. \mathbb{C} -IPS. If $\lambda, \eta \in \mathbb{C}$ are two distinct eigenvalues of T and T is normal, then

$$\text{Ker}(T - \lambda I) \perp \text{Ker}(T - \eta I).$$

Proof. Take $\alpha \in \text{Ker}(T - \lambda I)$ and $\beta \in \text{Ker}(T - \eta I)$, then we wish to show that $\langle \alpha, \beta \rangle = 0$. We do this as follows:

$$\lambda \langle \alpha, \beta \rangle = \langle \lambda \alpha, \beta \rangle = \langle T\alpha, \beta \rangle = \langle \alpha, T^*\beta \rangle = \langle \alpha, \bar{\eta}\beta \rangle = \bar{\eta} \langle \alpha, \beta \rangle.$$

Since $\lambda \neq \eta$, therefore $\langle \alpha, \beta \rangle = 0$. \square

5 Best approximation and self-adjoint operators

We will understand here how to approximate a vector, while staying inside a subspace, which is closest to it. In the process we will construct a linear map for each subspace of the vector space, which has some spectacular properties (Theorem 5.1.5). This operator will just keep on giving and will motivate us to study a class of operators on an IPS called self-adjoint operators. These operators are very nice as they are normal (see Remark 4.3.2), and as we noted in the beginning, they will be shown to satisfy some very good properties over a \mathbb{C} -IPS.

5.1 Best approximation

In an F -IPS, it is possible to find for an element $\alpha \in V$ and a subspace $W \subseteq V$, an element of W which is *closest* to α . For that, we recall that in an IPS, two vectors $v, w \in V$ are said to be **perpendicular**, if $\langle \alpha, \beta \rangle = 0$. For a subspace $W \subseteq V$, one denotes $W^\perp = \{v \in V \mid \langle v, w \rangle = 0 \forall w \in W\}$. One then defines the following:

Definition 5.1.1. (*W -approximation of α*) Let V be an F -IPS, $\alpha \in V$ and $W \subseteq V$ be a subspace. The W -approximation of α is defined to be a vector $\beta \in W$ such that for all $\gamma \in W$,

$$\|\beta - \alpha\| \leq \|\gamma - \alpha\|.$$

The following theorem tells us when will W -approximation of α exist, if it does, then whether it is unique and how to find it given an orthonormal basis.

Theorem 5.1.2. Let V be a f.d. F -IPS and let $\alpha \in V$ and $W \subseteq V$. Then,

1. A W -approximation of α exists and is equal to $\beta \in W$ if and only if $\alpha - \beta \perp W$.

2. If a W -approximation of α exists, then it is unique.
3. If $W \subseteq V$ is finite dimensional and $\mathcal{B} = \{\beta_1, \dots, \beta_n\}$ is an orthonormal basis of W , then the W -approximation of α is given by

$$\beta = \sum_{i=1}^n \langle \alpha, \beta_i \rangle \beta_i.$$

Proof. 2. Let us first prove the uniqueness using 1. Let $\beta_1, \beta_2 \in V$ be two elements such that both of them are W -approximations of α . We will show that $\beta_1 - \beta_2 = 0$, by showing that $\langle \beta_1 - \beta_2, \gamma \rangle = 0$ for all $\gamma \in V$ (then take $\gamma = \beta_1 - \beta_2$). This is simple by 1.

1. (R \Rightarrow L) Take any $\gamma \in W$ and then open $\alpha - \gamma = (\alpha - \beta) + (\beta - \gamma)$ where note that $\beta - \gamma \in W$ and $\alpha - \beta \perp W$ and the open as the following:

$$\begin{aligned} \|\alpha - \gamma\|^2 &= \|(\alpha - \beta) + (\beta - \gamma)\|^2 \\ &= \|\alpha - \beta\|^2 + \|\beta - \gamma\|^2 + \underbrace{\langle \alpha - \beta, \beta - \gamma \rangle}_0 + \underbrace{\langle \beta - \gamma, \alpha - \beta \rangle}_0 \\ &= \|\alpha - \beta\|^2 + \|\beta - \gamma\|^2. \end{aligned}$$

Hence $\|\alpha - \gamma\| \geq \|\alpha - \beta\|$.

(L \Rightarrow R) $\forall \gamma \in W$ we have $\|\alpha - \beta\| \leq \|\alpha - \gamma\|$. This gives us the following : $\|\gamma - \alpha\|^2 - \|\beta - \alpha\|^2 \geq 0$. Now open the square as above to get $2\Re\langle \gamma - \alpha, \beta - \alpha \rangle + \|\alpha - \gamma\|^2 \geq 0$ and do a serious back-calculation on $\gamma - \alpha$ to get the desired data.

3. Use the statement 1 to show that β is a *right* candidate, then use statement 2 to show that β is the candidate. \square

Let us now call the W -approximation of α by a name which is closer to home.

Definition 5.1.3. (Orthogonal projection of α onto W) Let V be a f.d. F -IPS and let $\alpha \in V$ and $W \subseteq V$ be a subspace. The W -approximation of α is also called the orthogonal projection of α onto W .

Construction 5.1.4. Let V be an F -IPS with $\alpha \in V$ and $W \subseteq V$ a subspace. We can construct the following mapping:

$$\begin{aligned} E_W : V &\longrightarrow V \\ \alpha &\longmapsto \text{orthogonal projection of } \alpha \text{ onto } W. \end{aligned}$$

This mapping is very special (as testified by the following result) and is called the **orthogonal projection onto W** .

Theorem 5.1.5. Let $W \subseteq V$ be a finite dimensional subspace of an F -IPS V . Let $E_W : V \rightarrow V$ be the orthogonal projection onto W . Then,

1. E_W is an idempotent linear operator,
2. $W^\perp = \text{Ker}(E_W)$,
3. $V = W \oplus W^\perp$.

Proof. 1. Idempotency of E_W is straightforward. Use the explicit form of the orthogonal projection of α onto W as stated in Theorem 5.1.2, 3 to get that E_W is linear.

2.& 3. This is straightforward once one notices that for any $\alpha \in V$, we can write it as $\alpha = \alpha - E_W\alpha + E_W\alpha$ where $\alpha - E_W\alpha \in \text{Ker}(E_W)$ and $E_W\alpha \in \text{Image}(E_W)$. But since $E_W\alpha = 0 \iff \alpha \in W^\perp$ by Theorem 5.1.2, 3, therefore we have 2. The 3. is simple. \square

Let us now see that E_W is a self-adjoint operator.

Lemma 5.1.6. Let V be a f.d. \mathbb{C} -IPS and $W \subseteq V$ be a subspace. Let $E_W : V \rightarrow V$ denote the orthogonal projection of V onto W . Then, E_W is a self-adjoint operator.

Proof. We wish to show that for any $\alpha, \beta \in V$ that $\langle E_W\alpha, \beta \rangle = \langle \alpha, E_W\beta \rangle$. By uniqueness of adjoints, we will be done. This is simple, because of the following observation : $\langle E_W\alpha, \beta \rangle = \langle E_W\alpha, E_W\beta + (\beta - E_W\beta) \rangle$, where by Theorem 5.1.5, we see that $\beta - (E_W\beta) \in W^\perp$ and $E_W\alpha, E_W\beta \in W$. Now after this we get

$$\begin{aligned} \langle E_W\alpha, \beta \rangle &= \langle E_W\alpha, E_W\beta \rangle \\ &= \langle E_W\alpha, E_W\beta \rangle + \underbrace{\langle \alpha - E_W\alpha, E_W\beta \rangle}_0. \end{aligned}$$

This last step is very important and must be kept in mind all the time while dealing with inner products; you can write $0 = \langle \alpha, \alpha^\perp \rangle$ (!) The proof then follows by opening the RHS of the above equation. \square

We now see the spectrum (the set of eigenvalues) of self-adjoint and unitary operators.

Lemma 5.1.7. Let V be a f.d. \mathbb{C} -IPS. Let $T : V \rightarrow V$ be a linear operator.

1. If T is self-adjoint, then $\lambda(T) \subseteq \mathbb{R}$.
2. If T is unitary, then $\lambda(T) \subset S^1 \subseteq \mathbb{C}$.

Proof. 1. Begin with $\lambda\langle \alpha, \alpha \rangle$ for $\alpha \in \text{Ker}(T - \lambda I)$. Everything is straightforward from there.

2. As usual, since we wish to show that $\lambda\bar{\lambda} = 1$, so begin with $\lambda\bar{\lambda}\langle \alpha, \alpha \rangle$ where α is an eigenvector for λ , and after some work it is equal to $\langle \alpha, \alpha \rangle$. \square

We now prove the main theorem of this part.

6 Spectral theorem for normal operators

We earlier saw that over a f.d. \mathbb{C} -vector space, every operator T is triangulable, that is, there is a basis \mathcal{B} over which $[T]_{\mathcal{B}}$ is triangular (Corollary 2.4.2). We now show the same thing, but more stronger when V is an \mathbb{C} -IPS, that there is an *orthonormal* basis \mathcal{B} such that $[T]_{\mathcal{B}}$ is triangular. When T becomes normal, then this theorem will also become special.

We have the first main theorem towards understanding when operators on a f.d. \mathbb{C} -IPS becomes diagonal.

Theorem 6.0.1. Let $T : V \rightarrow V$ be a linear operator on a f.d. \mathbb{C} -IPS. Then there exists an orthonormal basis \mathcal{B} such that $[T]_{\mathcal{B}}$ is upper triangular.

Proof. Theorem 21, pp 316 of [HK71]. \square

Another theorem that we observe is the following, which when combined with the above will get us the main result of this part. In-fact this gives us more than what we wanted, as not only it tells us that each normal operator is diagonalizable, but that diagonalizability only comes from normal operators.

Theorem 6.0.2. Let V be a f.d. \mathbb{C} -IPS and $T : V \rightarrow V$ be a linear operator. Let \mathcal{B} be an orthonormal basis such that $[T]_{\mathcal{B}}$ is upper triangular. Then,

$$T \text{ is normal} \iff [T]_{\mathcal{B}} \text{ is diagonal.}$$

Proof. Follows from Theorem 20, pp 315 of [HK71] and by Theorem 6.0.1 above. \square

We can state above in the matrix form as follows.

Corollary 6.0.3. If A is an $n \times n$ normal matrix over \mathbb{C} , then there exists an unitary $n \times n$ matrix U

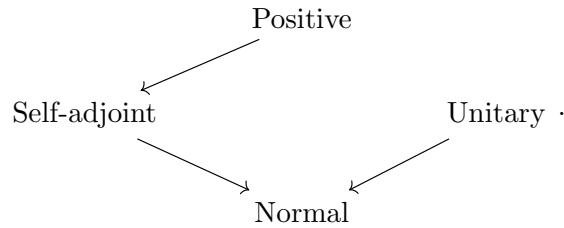
$$U^*AU = \Lambda$$

where Λ is the matrix with diagonal entries equal to the eigenvalues of A and columns of U are obtained by the eigenvectors corresponding to the eigenvalues in Λ .

Before moving on to the next theorem, let us give an important definition.

Definition 6.0.4. (Positive operator) Let V be a f.d. \mathbb{C} -IPS and let $T : V \rightarrow V$ be a linear operator. If $\forall \alpha \in V$, we have that $\langle T\alpha, \alpha \rangle \geq 0$, then T is said to be positive.

Remark 6.0.5. The following addendum to Remark 4.3.2 can be seen quite easily:



That is, a positive operator is self-adjoint. This follows by investigating the following implication that is obtained from the hypothesis: since $\langle T\alpha, \alpha \rangle \in \mathbb{R}$, therefore

$$\begin{aligned} \mathbb{R} \ni \langle T\alpha, \alpha \rangle &= \overline{\langle \alpha, T\alpha \rangle} \\ &= \langle \alpha, T\alpha \rangle \\ &= \langle T^*\alpha, \alpha \rangle \\ &= \overline{\langle \alpha, T^*\alpha \rangle} \\ &= \langle \alpha, T^*\alpha \rangle. \end{aligned}$$

We then conclude by the uniqueness of the adjoints. One can further see that the eigenvalues of a positive operator is always positive. This can be observed by letting α be an eigenvector of T .

Let us now state the more general spectral theorem for normal operators.

Theorem 6.0.6. (*Spectral theorem*) Let $T : V \rightarrow V$ be a normal linear operator on a f.d. \mathbb{C} -IPS. By Theorem 6.0.2, T is diagonalizable, so let $c_1, \dots, c_k \in \mathbb{C}$ be its distinct eigenvalues. Denote $W_i := \text{Ker}(T - c_i I) \subseteq V$ and let $E_i : V \rightarrow V$ be the orthogonal projection of V onto W_i for all $i = 1, \dots, k$. Then,

1. $W_i \perp W_j$ for all $i \neq j = 1, \dots, k$,
2. $T = c_1 E_1 + \dots + c_k E_k$, which is called the spectral resolution of the normal operator T ,
3. $I = E_1 + \dots + E_k$,
4. $V = W_1 \oplus \dots \oplus W_k$, that is T is diagonalizable,
5. the projections E_1, \dots, E_k are uniquely determined by T .

Proof. Statement 1. is just Proposition 4.3.6. Rest follows from Theorem 2.7.2. \square

A useful fact one can deduce from above is that all the orthogonal projections E_i in the above theorem are polynomials of T .

Corollary 6.0.7. Let $T : V \rightarrow V$ be a normal linear operator on a f.d. \mathbb{C} -IPS. Denote $W_i := \text{Ker}(T - c_i I) \subseteq V$ and let $E_i : V \rightarrow V$ be the orthogonal projection of V onto W_i for all $i = 1, \dots, k$. Then,

$$E_j = e_j(T)$$

where $e_j(x) = \prod_{i \neq j} \frac{x - c_i}{c_j - c_i} \in \mathbb{C}[x]$ is the Lagrange's interpolation polynomial.

Proof. By Theorem 6.0.6, we know that T can be written as $c_1 E_1 + \dots + c_k E_k$. This result then follows. \square

Let us revisit the Remark 6.0.5. We saw that each one of the unitary, self-adjoint and positive operators are normal. However, what conditions do we need to find on a normal operator T in order to conclude that it will be unitary, self-adjoint or positive? The following result answers that:

Proposition 6.0.8. Let V be a f.d. \mathbb{C} -IPS and let $T : V \rightarrow V$ be a linear operator. If T is normal, then we have the following equivalences:

1. T is self-adjoint if and only if $\lambda(T) \subseteq \mathbb{R}$,
2. T is unitary if and only if $\lambda(T) \subseteq S^1 \subseteq \mathbb{C}$,
3. T is positive if and only if $\lambda(T) \subseteq [0, \infty)$.

Proof. In each of the equivalences we need only prove the $R \Rightarrow L$ part as the converse has been done in Lemma 5.1.7 and in Remark 6.0.5. Let us state the main idea in each of these. In all of these, the main ingredient is the spectral resolution of the normal operator T (Theorem 6.0.6). In particular, for 1, we will use the fact that since T has an orthonormal basis of eigenvectors by Theorem 6.0.6, we get that any $\alpha \in V$ can be written as $\alpha = \sum_i d_i \alpha_i$. Now we want to show that $\langle T\alpha, \alpha \rangle = \langle \alpha, T\alpha \rangle$. It follows that for them to be equal, the eigenvalues c_i must be self-conjugate, that is real. One uses spectral resolution for the 2. and 3. \square

6.1 Unitary operators - 2

We will now study unitary operators in a different light. It will turn out that an unitary operator is exactly an inner product structure preserving linear isomorphism, hence giving us a geometric interpretation of unitary operators and algebraically, this simply tells us that if T is unitary, then $\langle T\alpha, T\beta \rangle = \langle \alpha, \beta \rangle$, which clearly would be very helpful. Let us first define the relevant category for our purposes. Denote by \mathbf{IPS}_F be the category of inner product spaces over F and inner product preserving linear transformations between them. A linear transform between IPS V and W , $T : V \rightarrow W$, is to preserve inner product if for all $\alpha, \beta \in V$, we have

$$\langle T\alpha, T\beta \rangle = \langle \alpha, \beta \rangle.$$

Lemma 6.1.1. If $T : V \rightarrow W$ is an inner product preserving linear transformation and T is surjective, then T is an isomorphism of inner product spaces.

Proof. We need only show that $\text{Ker}(T) = 0$. If $\alpha \in V$ is such that $T\alpha = 0$ then $\langle T\alpha, T\alpha \rangle = 0$, but $\langle T\alpha, T\alpha \rangle = \langle \alpha, \alpha \rangle$. \square

We now come to the main result. We'll later see that IPS isomorphisms are exactly unitary operators.

Theorem 6.1.2. Let $T : V \rightarrow W$ be a linear transform between two IPS V, W over F , both of which are of same dimension. Then, the following are equivalent.

1. T preserves inner products,
2. T is an F -IPS isomorphism (isomorphism in \mathbf{IPS}_F),
3. T takes every orthonormal basis of V to an orthonormal basis of W ,
4. \exists an orthonormal basis of V which is mapped by T to an orthonormal basis of W .

Proof. (1. \Rightarrow 2.) Since V and W have same dimension and $\|T\alpha\| = \|\alpha\|$ for all $\alpha \in V$, therefore $T\alpha = 0 \iff \|T\alpha\| = \|\alpha\| = 0 \iff \alpha = 0$. Hence T is an injective linear transform between two equal and finite dimensional vector spaces. Hence T is linear isomorphism and hence T is an IPS isomorphism.

(2. \Rightarrow 3.) and (3. \Rightarrow 4.) Trivial.

(4. \Rightarrow 1.) To check that T preserves inner products, reduce to the calculations on the orthonormal basis that T preserves. \square

Corollary 6.1.3. Any two finite dimensional IPS over F of same dimension are isomorphic.

Proof. Since any linear transform can be made by defining a set map between the respective basis, therefore we can conclude by Theorem 6.1.2, 4. \square

If T is an IPS map, then it preserves norms. But what about the converse? That is, if T is a linear transform which preserves norms, then is it an inner product preserving linear transform (i.e. an IPS map) as well? The answer is yes. Hence a linear isometry between IPS is exactly an IPS map.

Proposition 6.1.4. Let $T : V \rightarrow W$ be an F -linear transformation. Then, T is an F -IPS map if and only if $\|T\alpha\| = \|\alpha\|$ for all $\alpha \in V$.

Proof. One side is trivial, other side requires polarization identity (Lemma 4.0.1). \square

We finally see the main result of this section, which characterizes F -IPS isomorphisms with unitary operators.

Theorem 6.1.5. Let V be an F -IPS and $T : V \rightarrow V$ be a linear operator. Then, T is an F -IPS isomorphism if and only if T is an unitary operator.

Proof. (L \Rightarrow R) Since T is linear isomorphism which preserves inner products, therefore, for any $\alpha, \beta \in V$, we have

$$\begin{aligned}\langle T\alpha, \beta \rangle &= \langle T\alpha, TT^{-1}\beta \rangle \\ &= \langle \alpha, T^{-1}\beta \rangle.\end{aligned}$$

Thus by uniqueness of adjoints (Theorem 4.2.1), $T^* = T^{-1}$, that is, $TT^* = T^*T = I$.
(R \Rightarrow L) Super trivial. \square

A small characterization of unitary operators in terms of matrices are in order.

Corollary 6.1.6. An operator T on a \mathbb{C} -IPS V is unitary if and only if in every orthonormal basis \mathcal{B} of V , the matrix $[T]_{\mathcal{B}}$ is unitary.

Proof. This is an artifact of the fact that $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$ by Proposition 4.2.3, 2. \square

7 Polar decomposition

In this final topic of the course, we will talk about an another property of operators on a f.d. \mathbb{C} -IPS, that makes them behave like complex numbers (see Remark 4.2.6). We know two things about an arbitrary complex number $z \in \mathbb{C}$:

1. If $z \in [0, \infty)$, the z has a *unique square root* $w \in \mathbb{C}$, that is, $w^2 = z$,
2. z has a *polar decomposition*, that is we can write

$$z = re^{i\theta}$$

where $r \in [0, \infty)$ and $e^{i\theta} \in S^1 \subseteq \mathbb{C}$. Moreover, r is always unique for z whereas $e^{i\theta}$ will not be.

The way the second statement is written incites one to wonder whether an arbitrary operator T on a f.d. \mathbb{C} -IPS can be written as

$$T = NU$$

where N is a positive operator and U is a unitary operator. The answer is yes, it is true, and this decomposition is the title of the concluding section of this course.

Let us first find the analogue of statement 1 above.

7.1 Square root of positive operators

The following theorem asserts the existence and uniqueness of square root of positive operators.

Theorem 7.1.1. Let V be a f.d. \mathbb{C} -IPS and let $T : V \rightarrow V$ be a linear operator on V . If T is a positive operator, then there exists a unique positive operator $N : V \rightarrow V$ such that

$$N^2 = T.$$

Proof. Since T is positive, therefore it is self-adjoint and hence it is normal. Now by the spectral theorem (Theorem 6.0.6), we have the spectral resolution of T as $T = c_1 E_1 + \cdots + c_k E_k$. Since T is positive, therefore $c_i \geq 0$ for all i . Now consider the operator $N = \sum_i \sqrt{c_i} E_i$. Taking square of N , we see that $N^2 = (\sum_i \sqrt{c_i} E_i)^2 = \sum_{i,j} \sqrt{c_i c_j} E_i E_j$. Since $E_i E_j = \delta_{ij}$ by Theorem 6.0.6, 1, Theorem 5.1.5 and by Theorem 5.1.2, 1, therefore $N^2 = \sum_i \sqrt{c_i c_i} E_i^2 = \sum_i |c_i| E_i$ and since T is positive, therefore $|c_i| = c_i$. This shows existence. The uniqueness follows from the uniqueness of E_i for T as given by Theorem 6.0.6, 5. \square

This shows the existence of square roots of positive operators.

7.2 Polar decomposition of complex operators

Let us finally prove the analogue of statement 2 in the opening discussion.

Theorem 7.2.1. (*Polar decomposition*) Let $T : V \rightarrow V$ be a linear operator on a f.d. \mathbb{C} -IPS. Then there are

1. a unitary operator $U : V \rightarrow V$,
2. a positive operator $N : V \rightarrow V$,

such that $T = UN$ ⁹. Moreover, N is unique for T . If T is invertible, then U is also unique.

Proof. Theorem 14, pp 342 of [HK71]. \square

Remark 7.2.2. We have some concluding remarks about this theorem.

1. The polar decomposition of T is not unique unless T is invertible.
2. Even if T is invertible so as to make N and U unique, they may not commute.
3. But, for an operator T , we have that $UN = NU$ if and only if T is normal.

Proof. (L \Rightarrow R) Since N is a positive operator, so it is self-adjoint. We thus have $TT^* = (UN)(UN)^* = UNN^*U^* = UN^*N^*U^* = UN^*U^*N^* = UU^*NN = NN = N^*N = N^*U^*UN = T^*T$.

(R \Rightarrow L) If T is normal, then $T^*T = (UN)^*(UN) = NU^*UN = NN$ and $TT^* = (UN)(UN)^* = UNN^*U^* = UNNU^*$. Hence $UN^2U^* = N^2$. Now, here's a remarkable line of thought. Since $T^*T = TT^*$ is a positive operator, therefore by Theorem 7.1.1, there exists a unique square root. Since $N^2 = UN^2U^* = (UNU^*)^2$ and UNU^* is positive, therefore by uniqueness of square roots, $N = UNU^* \Rightarrow NU = UN$. Let us prove why UNU^* is positive. It is clear that UNU^* is self-adjoint, thus it is normal. Now, by Corollary 6.0.3, we see that \exists unitary V such that $UNU^* = V\Lambda V^*$ where Λ is the diagonal matrix with entries being the eigenvalues of UNU^* . After rearrangement, we see $N = U^*V\Lambda V^*U$. So the operator N , which is positive (and hence normal and hence diagonalizable), is similar to a diagonal matrix Λ . Since N is also similar to the diagonal matrix consisting of eigenvalues of N , which are all positive, therefore Λ is similar to another diagonal matrix of all positive entries. It follows that Λ is equal to that diagonal matrix and hence all eigenvalues of UNU^* are positive. We then conclude by Proposition 6.0.8, 3. \square

⁹One another way of looking at this is that every linear operator in $L(V)$ is a product of two normal operators, so if $L(V)_n$ denotes the set of all normal operators, then $L(V)_n \cdot L(V)_n = L(V)$.

4. It follows that each operator T on V can also be written as

$$T = NU$$

where N is positive and U unitary. Indeed, by applying the Theorem 7.2.1 on T^* , we get that we can write

$$T^* = U_1 N_1.$$

But this implies that $T = N_1^* U_1^* = N_1 U_1^*$ because N_1 is positive (so self-adjoint). But since U_1^* is also unitary, therefore we get the said claim.

References

[HK71] K. Hoffman, R. Kunze, S.S. Roy, *Linear Algebra*, 1971.