Further Linear Algebra

Kexing Ying

May 15, 2020

Contents

1 Introduction

As we have learnt from last year, linear algebra is a important subject regarding matrices, vector spaces, linear maps, and this year we will also take a look at some geometrical interpretations of these concepts.

1.1 Matrices

Definition 1.1 (Similar matrices)**.** Let $A, B \in \mathbb{F}^{n \times n}$ for some field $\mathbb{F}$. We say A is similar to B if and only if there exists some $P \in \mathbb{F}^{n \times n}$ such that,

$$B = P^{-1} A P.$$

We recall that *similar* is an equivalence relation and similar matrices shares many useful properties such as

- same determinant
- same characteristic polynomial
- same Eigenvalues
- same rank

and many more. As similar matrices share so many properties, one major aim in linear algebra is to find a “nice” matrix B given any arbitrary square matrix A such that A and B are similar. We first saw a version of this question last year through the *diagonalisation* of matrices. However, as we have seen, not all matrices are diagonalisable, therefore, in this course, we will take a look at some *weaker* versions that are more general.

A version of our aim is the triangular theorem which states that; given $A \in \mathbb{C}^{n\times n}$ (note that this theorem is not true for arbitrary field), there exists (and not uniquely) some upper triangular matrix $B \in \mathbb{C}^{n\times n}$ such that A is similar to B.

Another version of this aim is the *Jordan Canonical Form* theorem. It turns out if $A \in \mathbb{C}^{n\times n}$, then A is similar to a *unique* matrix in the Jordan canonical form. This theorem is powerful due to the canonical nature of this theorem. One immediate result of this theorem is that we can check whether two matrices are similar to each other by checking where or not they have the same *JCF* (which is computationally easy to do).

However, we see that neither of the above version are theorems over arbitrary fields. The *Rational Canonical Form* attempts to solve this.

Definition 1.2 (Companion matrix)**.** Given an arbitrary field $\mathbb{F}$, $p \in \mathbb{F}[X]$ such that p is monic (i.e. the coefficient of the highest term of p is 1) and $\deg p = k$, the companion matrix of p is the the $k \times k$ matrix

$$C(p) = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}$$

where a_i is the coefficient of p of the term X^i in $\mathbb{F}$.

The companion matrix is a nice matrix and it we can in fact show that the characteristic polynomial of the companion matrix of some p is p.

Theorem 1. *Let $A \in \mathbb{F}^{n\times n}$ with characteristic polynomial p. Then, there exists a polynomial factorisation such that $p = \prod_{i=1}^{k} p_i$ and*

$$A \sim \begin{bmatrix} C(p_1) & 0 & \cdots & 0 \\ 0 & C(p_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C(p_k) \end{bmatrix}.$$

Furthermore, it turns out this factorisation is unique under certain assumptions which we will take a look at in the course.

1.2 Geometry

Recall the dot product on $\mathbb{R}^n$ where given $u, v \in \mathbb{R}^n$, $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$. Furthermore, recall we also took a look at *orthogonal* and *symmetric* matrices last year. All of these, of course, has geometric interpretations and we will in this part of the course generalise and axiomatise these to the theory of *inner product spaces* of V over $\mathbb{R}$. We will also extend this theory to arbitrary fields $\mathbb{F}$ - the *Theory of Bilinear Forms*.

2 More on Vector Spaces

From this point forward, we write $\sum \mu S$ as a shorthand for $\sum_{s \in S} \mu_s s$ for some suitable set S and indexed value μ. We will also write $T \in \mathrm{End}(V)$ for T is an endomorphism of V, that is, a linear map $T : V \to V$.

2.1 Algebraic & Geometric Multiplicities of Eigenvalues

We recall some basic definitions and properties of Eigenvectors.

Definition 2.1. Let V be some vector space, $T \in \mathrm{End}(V)$ a linear map and λ an Eigenvalue of T. Then the λ-Eigenspace of T is the subspace of V,

$$E_\lambda := \{v \in V \mid (\lambda I_V - T)v = \mathbf{0}\}.$$

We see that this is a subspace as it is the kernel of the linear map $\lambda I_V - T$.

Theorem 2. *Let V be some vector space, $T \in \mathrm{End}(V)$ a linear map. Suppose that $\{v_1, \cdots, v_k\}$ are Eigenvectors corresponding to distinct Eigenvalues $\lambda_1, \cdots, \lambda_k$, then it is linearly independent.*

Proof. We will prove by contrapositive. Suppose that $\{v_1, \cdots, v_k\}$ are Eigenvectors that are linearly independent. Then by definition, there exists a minimal set of $\{\mu_i \mid i \in I\}$, such that $\sum_{i \in I} \mu_i v_i = 0$ (we see that $\mu_i \neq 0$ for all i as otherwise it is not minimal). Now, let $j \in I$, then by rewriting, we have $v_j = \sum_{i \neq j} \mu'_i v_i$. Thus,

$$\lambda_j \sum_{i \neq j} \mu'_i v_i = \lambda_j v_j = T(v_j) = T\left(\sum_{i \neq j} \mu'_i v_i\right) = \sum_{i \neq j} \mu'_i T(v_i) = \sum_{i \neq j} \mu'_i \lambda_i v_i.$$

So, by rearranging, $0 = \sum_{i \neq j} (\lambda_i - \lambda_j) \mu'_i v_i$. Now, if for all $i \neq j$, $\lambda_i \neq \lambda_j$, we have found a smaller subset of $\{v_1, \cdots, v_k\}$ that is linearly dependent, contradicting our assumption, so there must be some i such that $\lambda_i = \lambda_j$. $\square$

Corollary 2.1. *Let V be a n-dimensional vector space. Then if the characteristic polynomial of the linear map $T \in \mathrm{End}(V)$ has n distinct roots, then T is diagonalisable.*

We define *algebraic* and *geometric* multiplicity for Eigenvalues.

Definition 2.2 (Algebraic and Geometric Multiplicity)**.** Let $T \in \mathrm{End}(V)$ be a linear map with characteristic polynomial χ_T, such that $\chi_T(\lambda) = 0$ (i.e. λ is an Eigenvalue of T).

The algebraic multiplicity of λ is the number $a(\lambda)$ such that

$$\chi_T(x) = (x - \lambda)^{a(\lambda)} q(x),$$

for some polynomial $q(x)$ where $q(\lambda) \neq 0$.

The geometric multiplicity of λ is

$$g(\lambda) = \dim E_\lambda.$$

Proposition 2.1. *Let $T \in \mathrm{End}(V)$ be a linear map with an Eigenvalue λ, then $g(\lambda) \leq a(\lambda)$.*

Proof. Let $r = g(\lambda) = \dim E_\lambda$, then there exits linearly independent vectors $v_1, \cdots, v_r$ which forms a basis of E_λ. Suppose we extend this to a basis of V,

$$B = \{v_1, \cdots, v_r, w_1, \cdots, w_s\},$$

then by working out $T(b)$ for all $b \in B$, we find $T(v_i) = \lambda v_i$, and $T(w_i) = \sum \mu_i B$ so,

$$[T]_B = \left[\begin{array}{cccc|cccc} \lambda & 0 & \cdots & 0 & \mu_1(v_1) & \mu_2(v_1) & \cdots & \mu_s(v_3) \\ 0 & \lambda & \cdots & 0 & \mu_1(v_2) & \mu_2(v_2) & \cdots & \mu_s(v_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & \mu_1(v_r) & \mu_2(v_r) & \cdots & \mu_s(v_r) \\ \hline & & & & \mu_1(w_1) & \mu_2(w_1) & \cdots & \mu_s(w_3) \\ & \multicolumn{2}{c}{\mathbf{0}} & & \mu_1(w_2) & \mu_2(w_2) & \cdots & \mu_s(w_2) \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \mu_1(w_s) & \mu_2(w_s) & \cdots & \mu_s(w_s) \end{array}\right].$$

We will refer to the four quadrants as $[\lambda], A, \mathbf{0}$ and C respectively.

Thus, by considering the characteristic polynomial of this, we have

$$\chi_{[T]_B} = \det(xI - [T]_B) = (x - \lambda)^r \det(xI - C),$$

implying the algebraic multiplicity of λ is at least r. □

Theorem 3. *Let $\dim V = n$ and $T \in \mathrm{End}(V)$ be a linear map with distinct Eigenvalues $\lambda_1, \cdots, \lambda_r$. Suppose that the characteristic polynomial of T is*

$$\chi_T = \prod_i (x - \lambda_i)^{a(\lambda_i)},$$

(so, $\sum_i a(\lambda_i) = n$). Then the following are equivalent,

1. *T is diagonalisable;*
2. *$\sum_i g(\lambda_i) = n$;*
3. *for all i, $g(\lambda_i) = a(\lambda_i)$.*

Proof. $2 \iff 3$ is trivial so let us consider the other cases.

$1 \implies 2$. Suppose T is diagonalisable, then there exists some B, a basis of V consisting of Eigenvectors of T. Then, we can partition B into $F_{\lambda_i} := \{v \in B \mid T(v) = \lambda_i v\}$ for all Eigenvalues of T. By noting that the subspace induced by F_{λ_i} is a subspace of the λ_i Eigenspace E_{λ_i}, we have,

$$\sum_i g(\lambda_i) = \sum_i \dim E_{\lambda_i} \geq \sum_i \dim F_{\lambda_i} = n.$$

Now, as $\sum_i g(\lambda_i) \leq \sum_i a(\lambda_i) = n$ by the previous proposition, it follows $\sum_i g(\lambda_i) = n$.

$2 \implies 1$. Suppose $\sum_i g(\lambda_i) = n$. Let B_i be the basis of E_{λ_i} for all λ_i an Eigenvalue and let $B = \bigcup B_i$. We can straight away see that $|B| = n$ so it suffices to show that B is linearly independent. Suppose otherwise, then there exists an index set $I \subseteq \{1, \cdots, r\}$,

$$\sum_{i \in I} \sum \mu_i B_i = 0$$

where $\sum \mu_i B_i \neq 0$ for all i. Now as $\sum \mu_i B_i \in E_{\lambda_i}$, this is a sum of Eigenvectors with distinct Eigenvalues. However, by theorem 2, these Eigenvectors are there fore linearly independent, so they must be zero. # □

2.2 Direct Sums

Recall that we can add subspaces of a vector space together forming another subspace, that is, given $U_1, U_2 \leq V$, $U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\} \leq V$. Direct sums is a particular case of this and is closely linked to *block-diagonal* matrices.

Definition 2.3 (Direct Sum of Vector Space)**.** Let V be a vector space with subspaces $U_1, \cdots, U_k$. We write

$$V = \bigoplus_{i=1}^{k} U_i$$

for the direct sum of subspaces if every $v \in V$, there exists unique $u_i \in U_i$ for all i such that $v = \sum u_i$.

Proposition 2.2. *Let V be a vector space with subspaces V_1, V_2, then $V = V_1 \oplus V_2$ if and only if $V_1 \cap V_2 = \{0_v\}$ and $\dim V_1 + \dim V_2 = \dim V$.*

Proof. Follow your nose. □

Proposition 2.3. *Let V be a vector space with subspaces $V_1, \cdots, V_k$, then $V = \bigoplus_{i=1}^{k} V_i$ if and only if $\sum_{i=1}^{k} \dim V_i = \dim V$ and if B_i is a basis of V_i then $\bigcup_{i=1}^{k} B_i$ is a basis of V.*

Proof. ($\implies$). Suppose $V = \bigoplus_{i=1}^{k} V_i$, then for all i, j, $V_i \cap V_j = \{0_V\}$, thus $B_i \cap B_j = \varnothing$ for all $i \neq j$, implying $\sum_{i=1}^{k} \dim V_i = \dim V$ and $|\bigcup B_i| = \dim V$ so it suffices to show that $\bigcup B_i$ is linearly independent. However, this is trivial as if $\bigcup B_i$ is linearly dependent, then there are two distinct ways of writing 0 as a sum of vectors in V_i. #

($\impliedby$). Suppose $\sum_{i=1}^{k} \dim V_i = \dim V$ and $\bigcup_{i=1}^{k} B_i$ is a basis of V, then it follows,

$$V = V_1 + V_2 + \cdots V_k.$$

Suppose for contradiction there is two representations of $v \in V$ where $v = \sum v_i = \sum v_i'$. Then, $v = \sum_i \sum \mu_i B_i = \sum_i \sum \mu_i' B_i$, and thus, $0 = \sum_i \sum \mu_i B_i - \sum_i \sum \mu_i' B_i = \sum_i \sum (\mu_i - \mu_i') B_i$. By rewriting, $0 = \sum (\mu_i - \mu_i') \bigcup B_i$, implying $\bigcup B_i$ is linearly dependent. # □

Definition 2.4 (Invariant Subspace)**.** Let V be a vector space with subspace W and let $T \in \mathrm{End}(V)$ be a linear map. We say W is T-invariant if and only if

$$T(W) \subseteq W.$$

We write $T_W : W \to W$ as the restriction of T to W.

A example of an invariant subspace is the Eigenspace of a linear map since $T(E_\lambda) = \{T(v) \mid v \in E_\lambda\} = \{\lambda v\} \subseteq E_\lambda$.

Theorem 4. *Let $T \in \mathrm{End}(V)$ be a linear map and suppose $V = \bigoplus V_i$ where for all i, V_i is T-invariant. Let B_i be a basis of V_i, and $A_i = [T_{V_i}]_{B_i}$, then*

$$[T]_{\bigcup B_i} = \mathrm{diag}(A_1, A_2, \cdots, A_k).$$

Proof. Follows directly from the T-invariant property of V_i. $\square$

From the proposition above, we see the close link between direct sums and block diagonal matrices. To further highlight the fact, from this point forward, we write $\bigoplus_{i=1}^k A_i = \mathrm{diag}(A_!, A_2, \cdots, A_k)$ where A_i are block matrices.

Corollary 4.1. *Let $A = \bigoplus_{i=1}^r A_i$ and let $\pi \in S_r$. Then $A \sim A' := \bigoplus_{i=1}^r A_{\pi(i)}$.*

Proof. Let the vector space V in the above theorem be the span of the columns of A and V_i the span of columns of A_i with the missing entries filled with zero. Then it is not hard to see $V = \bigoplus V_i$. Now, by letting $T \in \mathrm{End}(V) : v \mapsto Av$, we see that for all i, V_i is T-invariant and $T_{V_i} = v \mapsto A_i v$, so, by taking the basis B to be the standard basis, we have $A = [T]_B$. Now, by permuting the standard basis by pi, resulting in the basis B', we have $A' = [T]_{B'}$, and so, by letting P be the change of basis matrix from $B \to B'$, we have shown $A \sim A'$. $\square$

2.3 Quotient Spaces

Just like other algebraic graphs we have can construct a quotient structure on vector spaces.

Let V be a vector space and $W \leq V$, then let $\sim_W : V \to V \to \mathrm{Prop}$ be the binary relation such that

$$v_1 \sim_W v_2 \iff v_1 + W = v_2 + W,$$

where $v + W = \{v + w \mid w \in W\}$ for all $v \in V$.

By manually checking, we find this is an equivalence relation and the set $V/\sim_W$ equipped with the natural addition and scalar multiplication form a vector space. We will write V/W for this quotient space.

Definition 2.5. Given a quotient space V/W, there exists a linear map

$$q_W : V \to V/W : v \mapsto v + W.$$

Proposition 2.4. *Let V be a finite dimensional vector space with the subspace W, then $\dim V/W = \dim V - \dim W$.*

Proof. Let B_W be a basis of W and B_V the extension basis of V from B_W. Then we easily see that $V/W \subseteq \mathrm{sp}(q_W(B_V \setminus B_W))$ as for all $v \in V$, $v = \sum \mu B_V$, so $v + W = q_W(v) = q_W(\sum \mu B_V) = \sum \mu q_W(B_W) + \sum \mu q_W(B_V \setminus B_W) = \sum 0_{V/W} + \sum \mu q_W(B_V \setminus B_W) \in \mathrm{sp}(q_W(B_V \setminus B_W))$.

Now suppose $q_W(B_V \setminus B_W)$ is not linearly independent in V/W, then, there exists μ, $0 = \sum \mu q_W(B_V \setminus B_W) = q_W(\sum \mu(B_V \setminus B_W))$, so $\sum \mu(B_V \setminus B_W) \in \ker q_W = W$. If $\sum \mu(B_V \setminus B_W) = 0_V$, then $B_V \setminus B_W$ is not linearly dependent, a contradiction so, $\sum \mu(B_V \setminus B_W) \neq 0_V$. Now, as $\sum \mu(B_V \setminus B_W) \in W$, there is some λ, $\sum \mu(B_V \setminus B_W) = \sum \lambda B_W$, so $\sum \mu(B_V \setminus B_W) - \sum \lambda B_W = 0$ implying B_V is not linearly independent. # □

With the above proposition, we have found a method to find a basis of a quotient space V/W by extending the basis of W.

Let us now consider quotient spaces' relation with linear maps.

Definition 2.6 (Quotient Map)**.** Let V be a vector space and W a subspace of V. Suppose $T \in \mathrm{End}(V)$ is a linear map and W is T-invariant. Then there is an induced quotient map

$$\bar{T} : V/W \to V/W : q_W(v) \mapsto q_W(T(v)).$$

To see that this is well defined, let $u, v \in V$, $q_W(u) = q_W(v)$, then $u - v \in W$ implying $T(u - v) \in W$ as W is T-invariant. Thus, $0_{V/W} = q_W(T(u - v)) = q_W(T(u) - T(v)) = q_W(T(u)) - q_W(T(v))$ implying $\bar{T}(u) = \bar{T}(v)$.

Theorem 5. *Let V be a vector space W a subspace that is T-invariant for some $T \in \mathrm{End}(V)$ a linear map. Let B_W be a basis of W, B the extended basis of V from B_W, and $\bar{B}$ the basis of V/W as constructed by proposition 2.4. Then*

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right],$$

where A is some matrix.

Proof. Consider where $T(v)$ lands whenever $u \in B_w \subseteq W$, and where $\bar{T}(v)$ lands for the rest of the basis vectors. □

Corollary 5.1. *Let $T \in \mathrm{End}(V)$ be a linear map and $W \leq V$ is T-invariant, then $\chi_T = \chi_{T_W} \chi_{\bar{T}}$ where χ_f denotes the characteristic polynomial of the linear map f.*

2.4 Triangularisation Theorem

We have now arrived at the first major theorem of this course, that under certain conditions we can always triangularise matrices. We will in general work with upper triangular matrices when referring to triangular matrices.

Proposition 2.5. *Let $A = [a_{i,j}], B = [b_{i,j}] \in M_n(\mathbb{F})$ be triangular, then*

- $\chi_A(x) = \prod_{i=1}^{n} (x - a_{i,i})$;
- $\det A = \prod_{i=1}^{n} a_{i,i}$;
- *AB is also triangular with diagonal $a_{i,i} b_{i,i}$.*

The Triangularisation theorem states:

Theorem 6. *Let V be a finite dimensional vector space over some field $\mathbb{F}$, and let $T \in \mathrm{End}(V)$ be a linear map. Suppose the the characteristic polynomial of T, χ_T factorises into a product of linear factors, i.e. there exists $\lambda_i \in \mathbb{F}$,*

$$\chi_T(x) = \prod (x - \lambda_i),$$

then, there exists a basis B of V such that $[T]_B$ is upper triangular.

Straight away, we see a version of this in terms of matrices instead of linear maps in which the matrix is *similar* to a triangular matrix. We also note that, for some fields, such as the complex numbers $\mathbb{C}$, we can always triangularise any matrix (by *FTA*). This might not be the case for other fields such as the real numbers.

Proof. We induct on the dimension of V. The theorem is trivial when $\dim V = 1$, so let us consider the case when $\dim V = k + 1$ under the inductive hypothesis.

As χ_T factorises, T has an Eigenvalue λ and some Eigenvector $v \in V$. Let $W = \mathrm{sp}(v)$ be a T-invariant subspace of V. Then, by proposition 2.4, V/W has dimension k and we have the induced quotient map $\bar{T} : V/W \to V/W$. Now, by corollary 5.1, $\prod(x - \lambda_i) = \chi_T(x) = \chi_{\bar{T}}(x)\chi_{T_W}(x) = \chi_{T_W}(x)(x - \lambda)$. So, $\chi_{T_W}(x)$ is a polynomial of degree k which factorises. Then by our inductive hypothesis, there exists a basis $\bar{B}$ such that $[\bar{T}]_{\bar{B}}$ is triangular. Then by theorem 5, we have found a basis B, $[T]_B$ is triangular. $\square$

Corollary 6.1. *Let $A \in M_n(\mathbb{C})$ with Eigenvalues λ_i. Then $\sum g(\lambda_i)\lambda_i = \mathrm{tr}(A)$.*

Proof. By the triangularisation theorem, $A = PQP^{-1}$ where Q is triangular. As A and Q have the same Eigenvectors, it suffices to show that $\mathrm{tr}(A) = \mathrm{tr}(Q)$. But this follows as $\mathrm{tr}(A) = \mathrm{tr}(PQP^{-1}) = \mathrm{tr}(P^{-1}PQ) = \mathrm{tr}(Q)$. $\square$

3 Polynomials

3.1 Cayley-Hamilton Theorem

Recall that given a polynomial $p(x) = \sum_{i=0}^n a_i x^i$, we write $p(T)$ as the linear map $\sum_{i=0}^n a_i T^i$ for the linear map $T \in \mathrm{End}(V)$ and similarly for matrices. Then, the Cayley-Hamilton theorem states the famous result that, given a linear map $T \in \mathrm{End}(V)$, if χ_T is the characteristic polynomial of T, then $\chi_T(T) = 0$. We will prove this theorem within this chapter.

Straight away, we see that the result is trivial if the matrix in question is diagonal (or thus, similar to a diagonal matrix) since if $A = \mathrm{diag}(\lambda_i)$, $p(A) = \mathrm{diag}(p(\lambda_i))$ for any polynomial p. In fact, by similar argument, we find the theorem is also true for triangular matrices, and thus, by the triangularisation theorem, the Cayley-Hamilton theorem is true for vector spaces over the complex numbers (see problem sheet 3). However, this is less trivial for general matrices over arbitrary fields which we shall provide a proof here.

Lemma 3.1. *Let $T \in \mathrm{End}(V)$ be a linear map such that there does not exists a proper non-trivial T-invariant subspace of V. Suppose $\dim V = n$, then the set $B := \{v, T(v), T^2(v), \cdots, T^{n-1}(v)\}$ forms a basis of V for any non-zero $v \in V$.*

Proof. As $|B| = n$, it suffices to show that it is linearly independent. Suppose otherwise, then there exists some μ, such that $\sum \mu B = 0$. Then, we can choose i such that i is the largest number in which $T^i(v) = \sum_{i \neq j} \mu_j T^j(v)$. But then, for all $u \in \text{sp}(B), u = \sum \lambda T^i(v)$, so $T(u) = T(\sum \lambda T^i(v)) = \sum \lambda T^{i+1}(v)$. Now, as $T^{n+1}(v) = T^{n+1-k}(T^k(v)) = T^{n+1-k}(\sum_{i \neq j} \mu_j T^j(v)) = \sum_{i \neq j} \mu_j T^{n+1-k+j}(v) \in \text{sp}(B)$ as $n + 1 - k + j \leq n$ for all j as k is the largest. Thus, sp(B) is a proper and non-trivial T-invariant subspace. # ☐

Proof. (Cayley-Hamilton Theorem). Let $T \in \text{End}(V)$ and χ_T be the characteristic polynomial of T. We will induct on the dimension of V, n.

The $n = 1$ case is trivial, so let us suppose the inductive hypothesis for dimensions $\leq k$ and we will prove this theorem for $n = k$.

Suppose first that there exists a proper and non-trivial T-invariant subspace W of V and suppose it has basis B_W. We can then extend this basis to a basis B of V such that

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right].$$

Now, we recall that $\chi_T = \chi_{T_W} \chi_{\bar{T}}$, so,

$$\begin{aligned}
\chi_T([T]_B) &= \chi_{T_W}([T]_B)\chi_{\bar{T}}([T]_B) \\
&= \left[\begin{array}{c|c} \chi_{T_W}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \chi_{T_W}([\bar{T}]_{\bar{B}}) \end{array}\right] \left[\begin{array}{c|c} \chi_{\bar{T}}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \chi_{\bar{T}}([\bar{T}]_{\bar{B}}) \end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{0} & A \\ \hline \mathbf{0} & \chi_{T_W}([\bar{T}]_{\bar{B}}) \end{array}\right] \left[\begin{array}{c|c} \chi_{\bar{T}}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \\
&= \mathbf{0},
\end{aligned}$$

where we write A for arbitrary block matrix and the second to last equality is due to the inductive hypothesis.

Suppose now that there does not exists a non-trivial proper T-invariant subspace of V. Then by lemma 3.1, the set $B := \{v, T(v), \cdots, T^n(v)\}$ forms a basis of V. Now, we see that $[T]_B$ is a companion matrix resulting in $\chi_{[T]_B}(x) = \sum a_i x^i$, where a_i are chosen such that $T^{n+1} = \sum -a_i T^i(v)$. But $T^{n+1} = \sum -a_i T^i(v) \iff \sum_{i=0}^{n+1} -a_i T^i(v) = 0$ where we let $a_{n+1} = 1$, so we have $\chi_{[T]_B}([T]_B(v)) = \sum_{i=0}^{n+1} -a_i T^i(v) = 0$.

Thus, by the law of excluded middle, we have Cayley-Hamilton. ☐

3.2 Some Theories on Polynomials

Let $\mathbb{F}$ be a field, then we denote the ring formed by the polynomials over $\mathbb{F}$ as $\mathbb{F}[X]$. And we will develop the theories of greatest common divisor, least common multiple, and polynomial prime factorisation for this ring.

Theorem 7 (Euclidean Algorithm)**.** *Let $f, g \in \mathbb{F}[X]$ such that $\deg g \geq 1$. Then there exists $q, r \in \mathbb{F}[X]$, $f = qg + r$ where $r = 0$ or $\deg r < \deg g$.*

Proof. We induct on the degree of f, n. For $n = 0$, we can choose $q = 0$ and $r = f$ and we are done. Let's now assume $n = k + 1$ alongside the inductive hypothesis.

Let's write $f(x) = a_{k+1}x^{k+1} + \cdots$ and $g(x) = a_m x^m + \cdots$. Then, we can write $f_1 = f - a_{k+1}b_m^{-1}x^{n-m}g$, where $\deg f_1 \leq \deg f$. So by the inductive hypothesis, there exists some $q, r \in \mathbb{F}[X]$ such that $f_1 = qg + r$ and $\deg r \leq \deg g$. Then, $f = f_1 + a_{k+1}b_m^{-1}x^{n-m}g = (q + a_{k+1}b_m^{-1}x^{n-m})g + r$. $\square$

Definition 3.1 (Greatest Common Divisor)**.** Let $f, g \in \mathbb{F}[X] \setminus \{0\}$. Then we say $d \in \mathbb{F}[X]$ is the greatest common divisor of f and g, $\gcd(f, g)$ if and only if $d \mid f, d \mid g$ and for all $e \mid f$ and $e \mid g$, $e \mid d$.

Straight away, we see that, unlike the integers, the greatest common divisor is of two polynomials is not unique as we can simply multiply the gcd by any scalar and receive another gcd. However, if we quotient out by this relation ($\sim: \mathbb{F}[X] \to \mathbb{F}[X] \to \mathrm{Prop} : f, g \mapsto \exists \lambda \in \mathbb{F} \setminus \{0\}, f = \lambda g$), the gcd turns out to be unique (see problem sheet 3) and exists.

Theorem 8. *If $f, g \in \mathbb{F}[X] \setminus \{0\}$, then the $\gcd(f, g)$ exists.*

Proof. Same argument as the integers by repeatedly applying the Euclidean algorithm. $\square$

Definition 3.2 (Coprime)**.** We call two polynomials $f, g \in \mathbb{F}[X]$ to be coprime if and only if $\gcd(f, g) = 1$.

Theorem 9 (Bezout's)**.** *If $f, g, d \in \mathbb{F}[X]$ such that $d = \gcd(f, g)$, then there exists $r, s \in \mathbb{F}[X]$ such that $d = rf + sg$.*

Now that we have established some basic properties about polynomials, we would like to consider what it might mean to be a prime polynomial.

Definition 3.3 (Irreducible)**.** A polynomial $f \in \mathbb{F}[X]$ is irreducible over $\mathbb{F}$ if and only if $\deg f \geq 1$ and there does not exists $g, h \in \mathbb{F}[X]$, $\deg g, \deg h < \deg f$ such that $f = gh$.

Remark. *We see that this definition of irreducibility is consistent with the one we have defined in ring theory since, $\langle f \rangle$ is not a maximal ideal if and only if there exists $g \in \mathbb{F}[X]$, $\langle f \rangle \subset \langle g \rangle \subset \mathbb{F}[X]$ and hence, $f \in \langle g \rangle$ implying there exists $h \in \mathbb{F}[X], f = gh$.*

Given $p \in \mathbb{Q}[X]$, it is usually difficult to decide whether or not it is irreducible. However, there is some tools that can help us determine the irreducibility of some rational polynomials.

Theorem 10. *Let $p \in \mathbb{Q}[X]$ be a monic polynomial with integer coefficients. Then,*

- *if $\alpha \in \mathbb{Q}$ is a root of p, then $\alpha \in \mathbb{Z}$;*
- *if p is irreducible over $\mathbb{Q}$, then it has a monic factorisation q, r, where q, r also have integer coefficients.*

Proof. The first part follows easily while the other is Gauss' lemma. $\square$

Theorem 11. *Let $p \in \mathbb{F}[X]$ be irreducible, and $a, b \in \mathbb{F}[X]$, if $p \mid ab$, then $p \mid a$ or $p \mid b$.*

Proof. Suppose $p \nmid a$, then $\gcd(p, q) = 1$ and by Bezout's, there exists $r, s \in \mathbb{F}[X]$ such that $rp + sa = 1$ so $b = rpb + sab$. Now, as $p \mid ab$, there exists $q \in \mathbb{F}[X]$ such that $ab = pq$ and so $b = (rb + sq)p$ and thus $p \mid b$. $\square$

Theorem 12 (Unique factorisation Theorem for Polynomials)**.** *Let $f \in \mathbb{F}[X]$ with $\deg f \geq 1$, then there exists a unique sequence of polynomials $(p_i)_{i=1}^r \subset \mathbb{F}[X]$, such that $f = \prod p_i$.*

Proof. Let us first prove existence. We induct on the degree of f. For $\deg f = 1$, the result is trivial so let use consider the theorem with $\deg f = k + 1$ with the inductive hypothesis. Now, if f is irreducible, the result follows so suppose otherwise. Then f can be factorised into two polynomials with degree less than that of f. But, by the inductive hypothesis, these two polynomial can be factorised, so, by multiplying their factors, can f be factorised.

Let us now prove uniqueness. Suppose now $f = \prod_{i=1}^r p_i = \prod_{j=1}^s q_j$. Then, by considering that for all i $p_i \mid q_j$ for some j, the result follows. $\square$

Lastly, we conclude on defining the least common multiple for polynomials.

Definition 3.4 (Least Common Mulitple)**.** Let $f, g \in \mathbb{F}[X]$, then the least common multiple of f and g, $\mathrm{lcm}(f, g)$ is the polynomial h such that $f \mid h, g \mid h$ and for all $k \in \mathbb{F}[X]$, if $f \mid k$ and $g \mid k$, then $h \mid k$.

Similarly to the greatest common multiple, we find the least common multiple exists and is unique (up to scalar multiplication).

3.3 Minimal Polynomial

We will in this section take a look at the minimal polynomial and some of its applications.

Definition 3.5 (Minimal Polynomial)**.** The minimal polynomial of $T \in \mathrm{End}(V)$ is the non-zero monic polynomial $m_T(x)$ of least degree such that $m_T(T) = x \mapsto 0$.

Straight away, we see $m_T \mid \chi_T$ as, by the Euclidean algorithm, there exists $q, r \in \mathbb{F}[X]$ such that $\chi_T = qm_T + r$ where $\deg r < \deg m_T$ if $r \neq 0$. By evaluating with T on both sides, we have $0 = r(T)$, so r annihilates T. But this contradicts the minimality of m_T, so $r = 0$. This can be generalised to any annihilators of T.

Lemma 3.2. *Let $T \in \mathrm{End}(V)$ and let $p \in \mathbb{F}[X]$ such that $p(T) = 0$. Then, a minimal polynomial of T divides p.*

Proof. By the Euclidean algorithm, there exists, $q, r \in \mathbb{F}[X]$, where $r = 0$ or $\deg r < \deg m_T$, such that $p = qm_T + r$. Thus, $0 = p(T) = q(T)m_T(T) + r(T) = r(T)$. # (Minimality of m_T.) $\square$

Furthermore, we can deduce that the minimal polynomial is unique.

Theorem 13. *Let $T \in \mathrm{End}(V)$ and let m_T and m_T' be minimal polynomials of T, then $m_T = m_T'$.*

Proof. By the previous lemma, we have $m_T \mid m'_T$ so there exists $p \in \mathbb{F}[X]$, such that $m_T = pm'_T$. Now, as both m_T and m'_T are the minimal polynomials, they must have the same degree, thus $\deg p = 0$ and hence, is a constant. But, since both m_T and m'_T are monic, $p = 1$, so $m_T = m'_T$. $\square$

As one can imagine, the minimal polynomial for matrices is defined similarly and alike many other properties, is shared for similar matrices.

Theorem 14. *Let $A, B \in M_n\mathbb{F}$ and $A \sim B$. Then the minimal polynomial of A, m_A is the same as the minimal polynomial of B, m_B.*

Proof. Suppose $A = P^{-1}BP$. Then, $0 = m_A(A) = m_A(P^{-1}BP) = P^{-1}m_A(B)P$, so $m_A(B) = 0$ as P is invertible. By symmetry, $m_B(A) = 0$ and thus, the result follows by divisibility. $\square$

As the minimal polynomial is pretty powerful, it will be helpful to be able to compute the minimal polynomial. We will now develop some tools to help us find this minimal polynomial.

Theorem 15. *Let $T \in \mathrm{End}(V)$, then λ is an Eigenvalue of T if and only if $m_T(\lambda) = 0$.*

Proof. ($\implies$) If λ is an Eigenvalue of T then there exists non-zero $v \in V$, $T(v) = \lambda v$, then $0 = m_T(T)(v) = \sum a_i T^i(v) = \sum a_i \lambda^i v = m_T(\lambda)v$. Now, as v is non-zero, $m_T(\lambda) = 0$.

($\impliedby$) Backwards direction follows straight away as if $m_T(\lambda) = 0$, as $m_T \mid \chi_T$ there exists some polynomial p such that $\chi_T = m_T p$ and thus, $\chi_T(\lambda) = m_T(\lambda)p(\lambda) = 0$, and hence, is an Eigenvalue. $\square$

With that, in order to find the minimal polynomial, it will be insightful to find the characteristic polynomial as the minimal polynomial shares roots and divides it (and then you can check all the cases).

Proposition 3.1. *The minimal polynomial of the companion matrix of some polynomial p is p.*

Before we can prove proposition 3.1, let us prove another useful lemma.

Lemma 3.3. *Let $A \in M_n(\mathbb{F})$ and suppose there exists some $v \in \mathbb{F}^n$, such that the set $S := \{A^i v \mid i \leq k\}$ is linearly independent for some $k < n$. Then all polynomials that annihilates A has degree at least $k + 1$.*

Proof. Suppose there exists $p \in \mathbb{F}[X]$ with degree $l \leq k$ such that $p(A) = 0$. Then $0 = p(A)v = \sum A^i v$. # $\square$

Proof. (Proposition 3.1). By the previous lemma, it suffices to find some v such that $\{C(p)^i v \mid i \leq n - 1\}$ is linearly independent. Straight away, we see $v = e_1$ suffices so we are done. $\square$

While, we proved that the minimal polynomial share linear factors with the characteristic polynomial, we hope the same is true for all irreducible factors. This turns out to be true.

Theorem 16. *Let $T \in \operatorname{End}(V)$, then, for all $p \in \mathbb{F}[X]$ such that p is a irreducible factor of χ_T, $p \mid m_T$.*

Proof. We will prove this theorem by cases on whether there exists a proper non-trivial T-invariant subspace of V, so first, let us suppose such a subspace W exists.

Let us induct on the dimension of V, n. The theorem is trivial for $n = 1$, so let us consider the case for $n = k + 1$ with the inductive hypothesis holding for all $n = m \leq k$. Again, as W is T-invariant, we have

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right].$$

Now, suppose we write $m_T(x) = \sum a_i x^i$, then,

$$\begin{aligned} 0 = m_T([T]_B) &= \sum a_i [T]_B^i \\ &= \left[\begin{array}{c|c} \sum a_i [T_W]_{B_W}^i & A' \\ \hline \mathbf{0} & \sum a_i [\bar{T}]_{\bar{B}}^i \end{array}\right] \\ &= \left[\begin{array}{c|c} m_T([T_W]_{B_W}) & A' \\ \hline \mathbf{0} & m_T([\bar{T}]_{\bar{B}}) \end{array}\right], \end{aligned}$$

so m_T annihilates both $[T_W]_{B_W}$ and $[\bar{T}]_{\bar{B}}$. Now, by considering $p \mid \chi_T = \chi_{[T_W]_{B_W}} \chi_{[\bar{T}]_{\bar{B}}}$, where p is irreducible, $p \mid \chi_{[T_W]_{B_W}}$ or $p \mid \chi_{[\bar{T}]_{\bar{B}}}$. Either way, by the inductive hypothesis, $p \mid m_{[T_W]_{B_W}}$ or $p \mid m_{[\bar{T}]_{\bar{B}}}$, both of which divides m_T as m_T annihilates their respecctive matrices.

Let us now consider the case in which there does not exists a proper non-trivial T-invariant subspace of V. But, then, similar to the proof of the Cayley-Hamilton theorem, we can construct some basis B such that $[T]_B$ is the companion matrix of $\chi_{[T]_B}$. Now as the companion matrix has the same minimal polynomial as the characteristic polynomial, we are done! □

Before we end this section, I would like to prove a powerful result regarding diagonalisability.

Theorem 17. *Let $T \in \operatorname{End}(V)$, and suppose there exists a non-zero polynomial p of degree greater than 0 such that p annihilates T and p can be factored into distinct roots. Then T is diagonalisable.*

Proof. Suppose we write $p(x) = \prod (x - \lambda_i)$, then, it suffices to show that $\dim \ker(\prod (T - \lambda_i)) \leq \sum \dim \ker(T - \lambda_i)$ as the kernel of $\prod (T - \lambda_i)$ is V while the latter is the sum of the dimensions of the Eigenspaces. This can be proved by showing $\dim \ker T_1 T_2 \leq \dim \ker T_1 + \dim \ker T_2$ and applying induction. □

Proposition 3.2. *Let $T_1, T_2 \in \operatorname{End}(V)$, then $\dim \ker T_1 T_2 \leq \dim \ker T_1 + \dim \ker T_2$.*

Proof. It is easy to see that $\ker T_2 \leq \ker T_1 T_2$ so it suffices to show $\dim \ker T_1 T_2 / \ker T_2 \leq \dim \ker T_1$ since for $W \leq V$, $\dim V/W = \dim V - \dim W$.

Consider the map $\phi : \ker T_1T_2/\ker T_2 \to \dim\ker T_1 : v + \ker T_2 \mapsto T_2(v)$. ϕ has well-defined range as for all $v \in \ker T_1T_2$, $T_2(v) \in \ker T_1$. We will now show ϕ is well-defined overall and injective all at once. Let $v_1, v_2 \in \ker T_1T_2$, then,

$$v_1 + \ker T_2 = v_2 + \ker T_2 \iff v_1 - v_2 \in \ker T_2 \iff T_2(v_1 - v_2) = 0 \iff T_2(v_1) = T_2(v_2).$$

Furthermore, it is easy to see that ϕ is a linear map since

$$\begin{aligned}\phi((v_1 + \ker T_2) + (v_2 + \ker T_2)) &= \phi((v_1 + v_2) + \ker T_2)\\ &= T_2(v_1 + v_2) = T_2(v_1) + T_2(v_2)\\ &= \phi(v_1 + \ker T_2) + \phi(v_2 + \ker T_2).\end{aligned}$$

so ϕ is a injective linear map. Now, as the image of an linearly independent set under a injective linear map is also linearly independent, we can construct an linearly independent set in $\ker T_1$ with cardinality $\dim\ker T_1T_2/\ker T_2$ by taking the image of its basis over ϕ. Thus, $\dim\ker T_1T_2/\ker T_2 \le \dim\ker T_2$ and the result follows from proposition 2.4. □

4 Canonical Forms of Vector Spaces

As mentioned in the introduction, the highlight of this course are the canonical form theorems on general vector spaces (that is we would like to show any matrix over some field is similar to a particularly "nice" block diagonal matrix). This will allow us to deduce properties about general matrices and correspondingly, linear maps.

In order to achieve this, we will first develop some theories on how to decompose a general vector space, that is, given vector space V, we would like to find V_i such that $V = \bigoplus V_i$.

4.1 Primary Decomposition

Theorem 18 (Primary Decomposition Theorem)**.** *Let V be a finite dimensional vector space of the field $\mathbb{F}$, and let $T \in \mathrm{End}(V)$ with minimal polynomial m_T. Suppose m_T has the irreducible factors p_i such that,*

$$m_T = \prod_{i=1}^{k} p_i^{n_i},$$

where $p_i \neq p_j$ for all $i \neq j$. Then,

- $V = \bigoplus_{i=1}^{k} \ker(p_i(T)^{n_i})$*;*
- $\ker(p_i(T)^{n_i})$ *is T-invariant;*
- *each restriction of T on $\ker(p_i(T)^{n_i})$ has minimal polynomial $p_i^{n_i}$.*

Straight away, we see that this decomposition is unique by the unique factorisation theorem, so, we can call it a canonical decomposition. Furthermore, if all factors of m_T are linear, the factorisation becomes,

$$m_T = \prod_{i=1}^{k} (x - \lambda_i)^{n_i},$$

for distinct λ_i. In this cases, the individual decomposition becomes $\ker(T - \lambda I)^{n_i}$ which are the *generalised Eigenspace* of T. Lastly, we notice a direct corollary of this theorem is theorem 17, so we have alternatively an (arguably) easier proof of this.

Before we can prove the primary decomposition theorem, let us prove the following lemma.

Lemma 4.1. *Let $T \in \mathrm{End}(V)$ and suppose $p_1, p_2 \in \mathbb{F}[X]$ are coprime such that $p_1(T)p_2(T) = 0$, then $V = \ker p_1(T) \oplus \ker p_2(T)$. Furthermore, if $m_T = p_1 p_2$, then the restriction of T to $\ker p_i(T)$ has minimal polynomial p_i for $i = 1, 2$.*

Proof. By Bezout's, there exists $s, t \in \mathbb{F}[X]$ such that $1 = sp_1 + tp_2$. So, by evaluating at T, we have $I = s(T)p_1(T) + t(T)p_2(T)$. Then for all $v \in V$, $v = Iv = s(T)p_1(T)(v) + t(T)p_2(T)(v) = v_1 + v_2$. Now, as $p_1 p_2 = 0$, we have $p_2(v_1) = p_2(s(T)p_1(T)(v)) = p_2 s(T) p_2 p_1(T)(v) = 0$, we have $v_1 \in \ker p_2(T)$ and similarly, $v_2 \in \ker p_1(T)$, and so, $V = \ker p_1(T) + \ker p_2(T)$. Also, suppose $v \in V_1 \cap V_2$, then $v = Iv = s(T)p_1(T)(v) + t(T)p_2(T)(v) = 0$, so $V = \ker p_1(T) \oplus \ker p_2(T)$.

Now, suppose $m_T = p_1 p_2$ and let us denote the minimal polynomial of T restricted on $\ker p_i(T)$ as m_i for $i = 1, 2$. By definition $p_i(T_i) := p_i(T \mid_{\ker p_i(T)}) = 0$, so $m_i \mid p_i$. So, as p_1 and p_2 are coprime, so are m_1 and m_2, thus, $m_T = \mathrm{lcm}(m_1, m_2) = m_1 m_2$ and the result follows. $\square$

We can see straight away how this lemma can help us prove the theorem.

Proof. (Primary Decomposition Theorem). Let us write $m_T = \prod_{i=1}^{k} p_i^{n_i}$ and we will induct on k. The theorem is trivial for for the base case so we begin by letting $k = n + 1$. Then, by considering the previous lemma on p_{n+1} and $\prod_{i=1}^{n} p_i$, the theorem follows. $\square$

4.2 Jordan Canonical Form

Recall that the triangularisation theorem is limited in many ways. Not only does it not apply for matrices who's characteristic polynomial cannot be factored, for those matrices the theorem does apply, the triangularisation is not unique. We attempt to improve upon this with the Jordan canonical form theorem.

Definition 4.1 (Jordan Blocks)**.** Let $\mathbb{F}$ be a field and let $\lambda \in \mathbb{F}$. Then $J_n(\lambda) \in M_n(\mathbb{F})$ is a Jordan block if and only if

$$J_n(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}.$$

Jordan blocks are a nice form of matrices and we can derive several properties straight away.

Proposition 4.1. *Let $J = J_n(\lambda)$ be some Jordan block, then*

- $\chi_J(x) = (x - \lambda)^n = m_J$;

- *λ is the only Eigenvalue and $a(\lambda) = n, g(\lambda) = 1$;*
- *$(J - \lambda I)e_{i+1} = e_i$ for $i = 1, \cdots, n-1$ and $(J - \lambda I)e_1 = 0$.*

Proof. Obvious. $\square$

Before we state the Jordan canonical form theorem, let us recall some properties of block diagonal matrices. Let A_i have characteristic polynomial χ_i and let $A = \bigoplus A_i$, then

- $\chi_A = \prod \chi_i$;
- $m_A = \mathrm{lcm}(m_i)$;
- for all λ and Eigenvalue of A, $\dim E_\lambda(A) = \sum \dim E_\lambda(A_i)$;
- for all $q \in \mathbb{F}[X]$, $q(A) = \bigoplus q(A_i)$.

With that, we can state the Jordan canonical form theorem.

Theorem 19 (Jordan canonical form theorem)**.** *Let $A \in M_n(\mathbb{F})$ and the characteristic polynomial is a product of linear factors over $\mathbb{F}$. Then, there exists some J, $A \sim J = \bigoplus J_{n_i}(\lambda_i)$ and this decomposition is unique up to the ordering of the Jordan blocks.*

We call this unique decomposition the Jordan canonical form of A.

Remark. *The condition on which the characteristic polynomial must be factorisable is necessary as otherwise it would not even be triangulisable. Also, we note that the λ_i is not necessarily distinct.*

Of course, there is an equivalent theorem for linear maps. In fact, we shall prove the linear map version now.

Theorem 20. *Let V be a n-dimensional vector space over $\mathbb{F}$ and $T \in \mathrm{End}(V)$ such that χ_T is a product of linear factors over $\mathbb{F}$. Then, there exists a basis B of V such that $[T]_B = J = \bigoplus J_{n_i}(\lambda_i)$, and furthermore, J is uniquely determined by T (up to ordering).*

Proof. We shall provide a proof for uniqueness first.

Suppose that T has only one Eigenvalue λ and

$$[T]_B \sim J = \bigoplus J_{n_i}(\lambda)$$

for some basis of V, B. Let us write $J = \bigoplus_{i=1}^r J_i(\lambda)^{a_i}$ where a_i is the number of λ-blocks with size i. Then, let us define

$$m_i := \dim \mathrm{Im}(J - \lambda I)^i.$$

Now, by considering $J_r(0)^i$, for $i = 1, \cdots, r$ we have

$$\begin{aligned}
m_{r-1} &= a_r, \\
m_{r-2} &= a_{r-1} + 2a_r, \\
m_{r-3} &= a_{r-2} + 2a_{r-1} + 3a_r,
\end{aligned}$$

and so on. In general, we see

$$m_i = \sum_{j=1}^{r-i} j a_{i+j}$$

for $i = 0, \cdots, r-1$. Thus, m_i uniquely determines a_i; and as m_i is computable and is uniquely determined by $(J - \lambda I)^i$, we have a_i is unique.

Now, let us relax the number of Eigenvalues condition. Suppose T has at least two Eigenvalues with one of them being λ, then we can write

$$[T]_B \sim J = J_\lambda \oplus L,$$

where L is the direct sum of all the other Jordan blocks. Now, as $L - \lambda I$ is invertible (easy to see as the determinant is non-zero), it has full rank l. Thus, by defining

$$r_i = \dim \mathrm{Im}(J - \lambda I)^i,$$

we have $r_i = \dim \mathrm{Im}(J_\lambda - \lambda I)^i + l$, allowing us to compute $m_i = \dim \mathrm{Im}(J_\lambda - \lambda I)^i = r_i - l$. $\square$

In most cases, it might not be necessary to go through the steps as described above to find the Jordan canonical form of a matrix.

Proposition 4.2. *Let $A \in M_n(\mathbb{F}) \sim J$ where J is in Jordan Canonical Form. Suppose we write*

$$J = (J_{n_1}(\lambda) \oplus \cdots \oplus J_{n_\alpha}(\lambda)) \oplus (J_{m_1}(\mu) \oplus \cdots \oplus J_{m_\beta}(\mu)) \oplus \cdots,$$

where $\lambda, \mu, \cdots$ are Eigenvalues of A, then,

- *the sum of the size of the λ-blocks equals the algebraic multiplicity of λ, i.e. $n_1 + \cdots + n_\alpha = a(\lambda)$;*
- *the number of λ-blocks equals the geometric multiplicity of λ, i.e. $\alpha = g(\lambda)$;*
- $\max\{n_1, \cdots, n_\alpha\} = r$ *where $(x - \lambda)^r$ is the highest power of $x - \lambda$ dividing m_A.*

Proof. The proofs are rather straight forward.

- Obvious by considering the characteristic polynomial of $J_k(\lambda)$.
- True by considering each λ-block has geometric multiplicity one, and that J and A share the same λ-Eigenspace.
- Since for all n, $J_n(\lambda)$ has minimal polynomial $(x - \lambda)^n$, we have the power dividing m_J is $\mathrm{lcm}((x - \lambda)^{n_i}, \cdots, (x - \lambda)^{n_\alpha})$ which equals $(x - \lambda)^{\max\{n_1, \cdots, n_\alpha\}}$.

$\square$

Now, let us finish the proof of the JCF theorem by proving that a linear map is similar to its JCF (we can say this since we had proved uniqueness).

Proof. First we reduce the proof to the case where $T \in \mathrm{End}(V)$ has only one Eigenvalue. Suppose

$$m_T(x) = \prod_{i=1}^{n} (x - \lambda_i)^{n_i},$$

where λ_i are distinct. By the primary decomposition theorem

$$V = \bigoplus_{i=1}^{n} \ker(T - \lambda_i I).$$

Then, if B_i is a basis of $\ker(T - \lambda_i I)$, then $\bigcup_i B_i$ is a basis of V. Furthermore, as $\ker(T - \lambda_i I)$ are T-invariant, we can write

$$[T]_B = \bigoplus_{i=1}^{n} [T_{V_i}]_{B_i},$$

and for all i, $[T_{V_i}]_{B_i}$ has minimal polynomial $(x - \lambda_i)^{n_i}$, that is $[T_{V_i}]_{B_i}$ has only one Eigenvalue. Hence, it suffices to show the theorem whenever T has only one Eigenvalue λ.

If T only has one Eigenvalue λ, then $\chi_T(x) = (x - \lambda)^n$ where $n = \dim V$. Suppose we define $S := T - \lambda I$, then straight away, we see $S^n = 0$ by Cayley-Hamilton so S is nilpotent, and hence, it suffices to show that S is similar to a JCF with 0-Jordan blocks.

By considering the 0-Jordan blocks are cyclic matrices, that is, we see $[S]_B = J_1(0) \oplus \cdots$ where $B := \{v_1, \cdots, v_n\}$ if and only if $S(v_i) = S(v_{i+1}$ for all $i = 0, \cdots n - 1$ and $S(v_{n_1}) = 0$, we can write the basis B as the union

$$\bigcup_{i=1}^{k} \{v_1, S(v_1), \cdots, S^{n_i - 1}(v)\}.$$

So, it suffices to find such $\{v_1, \cdots, v_k\}$ (we call basis generated using S by this set a Jordan basis of V).

To prove this, we shall induct on $n = \dim V$. The base case is trivial so assume the existence of $\{v_1, \cdots, v_k\}$ for all $\dim V < n$. Consider $\mathrm{Im}(S) = S(V)$ is a proper subspace of V (as S is nilpotent). So, by the inductive hypothesis, there exists a Jordan basis of $S(V)$ generated by $U = \{u_1, \cdots, u_r\}$. Now as, for all $u_i \in U$, $u_i \in S(V)$, there exists $v_i \in V$, such that $S(v_i) = u_i$. Furthermore, as $S^{m_i}(u_i) = 0$, we see that $S^{m_i - 1}(u_i) \in \ker S$, and thus, we can extend this to a basis of $\ker S$ by adding $w_1, \cdots, w_s$.

Finally, by letting $B = \{v_1, \cdots, v_r, w_1 \cdots, w_s\}$, we see that B generates a Jordan basis if and only if $S^i(B)$ linearly independent (since $\dim \ker S + \dim \mathrm{Im} S = n$). Suppose then there exists μ such that $\sum \mu S^i(B) = 0$. Then, by applying S on both sides, we receive $\sum \mu' S^i(U) = 0$ implying the non-vanishing terms has zero coefficient. Furthermore, we see the vanishing vectors is a basis of $\ker S$, so they also have zero coefficient, and hence $\mu = 0$ and $S^i(B)$ is linearly independent. Thus, we have found a Jordan basis of S, so we are done! □

Now that we have proved that the JCF theorem, we would like to compute the Jordan basis of V such that $[T]_B$ is in JCF for some appropriate $T \in \mathrm{End}(V)$.

Let $S \in \mathrm{End}(V)$ be nilpotent,

1. Compute the chain of subspaces

$$V \geq S(V) \geq S^2(V) \geq \cdots \geq S^r(V) \geq 0$$

where $S^{r+1}(V) = 0$.

2. Find a basis of $S^r(V)$ and for each i, add v_i to U where $S(v_i) = u_i$. Furthermore, if needed, extend U such that it has span greater than $\ker S^{r-1}$. This results in U to be a Jordan basis of $S^{r-1}(V)$.
3. Repeat the above step until we get a Jordan basis of V.

4.3 Cyclic Subspaces

So far we have seen a satisfactory canonical form for matrices who's characteristic polynomial factors into linear roots. However, this is often not the case. So, in the next few sections, we shall develop theories for the *rational canonical form* theorem.

Definition 4.2. Let V be a finite dimensional vector space over some field $\mathbb{F}$, $T \in \text{End}(V)$, $0 \neq v \in V$ and define the subspace

$$\begin{aligned} Z(v,T) :=& \{p(T)(v)) \mid p \in \mathbb{F}[X]\} \\ =& \text{sp}(v, T(v), T^2(v), \cdots), \end{aligned}$$

and we call it the T-cyclic subspace of V generated by v.

Straight away, we see that $Z(v,T)$ is T-invariant and thus, we can restrict T by $Z(v,T)$ (denoted by T_v). Also, if v is an Eigenvector of T, then $Z(v,T) = \text{sp}(v)$.

Definition 4.3 (T-annihilator)**.** The T-annihilator of v is the smallest degree, monic polynomial $m_v(x)$ such that $m_v(T)(v) = 0$.

Consider the sequence $v, T(v), T^2(v), \cdots$. As V is finite dimensional there exists some k, where k is the smallest natural number such that $T^k(v) \in \text{sp}(v, \cdots, T^{k-1}(v))$. Then, there exists μ, such that $T^k(v) = -\sum \mu T^i(v)$, and so, $(\sum^n \mu T^i)v = 0$. Hence, by the choice of k, $m_v(x) = x^k \sum \mu x^i$.

Proposition 4.3. *Given k as defined above, $B = \{v, \cdots, T^{k-1}v\}$ is a basis of $Z(v,T)$. Furthermore, $[T_v]_B$ is the companion matrix of m_v and the minimal polynomial of T_v is m_v.*

Proof. Obvious except perhaps for the last part. However, we have proved that the minimal polynomial of the companion matrix is the polynomial it is associated with, that is m_v. $\square$