JORDAN CANONICAL FORM

CHIU HSIANG-TZU

1. Triangular Form

Definition 1.1. *let $T : V \to V$ be a linear operator. A subspace $W \subseteq V$ is said to be invariant (or stable) under T if $T(W) \subseteq W$*

Remark. $\{0\}$, V, $Ker(T)$, $Im(T)$, and E_λ are T-invarient.

Definition 1.2. *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. We say that V is triangularizable if and only if there exists an ordered basis β such that $[T]_\beta^\beta$ is upper triangular.*

Example. Consider $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^4$. Let β be the standard ordered basis of V and let $\beta = \{e_1, e_2, e_3, e_4\}$, where e_i is $(0, ..., 0, 1, 0, ..., 0)$ with the nonzero component at position i. We compute the matrix representation of $[T]_\beta^\beta$ as follows.

$$[T]_\beta^\beta = \begin{bmatrix} 1 & 1-i & 2 & 0 \\ 0 & 1 & i & 0 \\ 0 & 0 & 1-i & 3+i \\ 0 & 0 & 0 & 1-i \end{bmatrix}$$

Clearly, the matrix is upper triangular. Notice that $T(e_1) = e_1$, that is, e_1 is an eigenvector of T, $T(e_2) = (1-i)e_1 + e_2 \in span(\{e_1, e_2\})$, $T(e_3) \in span(\{e_1, e_2, e_3\})$, and $T(e_4) \in span(\{e_1, e_2, e_3, e_4\})$. Let W_i be the subspace of $\mathbb{C}^4$ spanned by the first i vectors in the standard ordered basis, that is, $T(e_i) \in W_i$ for all i. Clearly, $T(W_i) \subseteq W_i$, where $T(W) = \{T(w) \mid w \in W\} = Im(T|_W)$ for all subsapce of V.

Proposition 1.1. *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and $\beta = \{x_1, \cdots, x_n\}$ be a basis for V. Then $[T]_\beta^\beta$ is upper triangle if and only if the subspace $w_i = span(\{x_1, \cdots, x_i\})$ is T-invariant.*

Proof. It is trivial. $\square$

Note that the subspace W_i in Proposition 1.1 are related as follows:

$$\{0\} \subseteq W_1 \subseteq \cdots \subseteq W_{n-1} \subseteq W_n = V.$$

We say that W_i forms an increasing sequence of subspaces. On the other hand, in Proposition 1.1, to show that a given linear operator is triangularizable. We must to construct the increasing sequence of T-invariant subspaces

$$\{0\} \subseteq W_1 \subseteq \cdots \subseteq W_{n-1} \subseteq W_n = V.$$

One also have to introduce the restriction of T to an T-invariant subspace $W \subseteq V$. Clearly, $T|_W = T_W : W \to W$ is a new linear mapping, where W is a T-invariant subspace of V.

Proposition 1.2. *Let $T : V \to V$ be a linear operator and W be a T-invariant subspace of a finite dimensional vector space V. Then the characteristic polynomial of $T|_W$ divides the characteristic polynomial of T.*

Proof. One can see Theorem 5.21 in Friedbreg. □

Corollary 1.1. *Every eigenvalue of $T|_W$ is also an eigenvalue of T,that is, the eigenvalue of $T|_W$ is some subset of the eigenvalue of T on V.*

Let $T : V \to V$ be a linear operator, where V is a finite dimensional vector space. Let $\lambda_1, \cdots, \lambda_n$ be distinct eigenvalues and m_i be the multiplicity of λ_i, as a root of the characteristic polynomial of T. Then T is diagonalizable if and only if $m_1 + \cdots + m_n = \dim(V)$ and $\dim E_{\lambda_i} = m_i$ for all i. The proof combines with Theorem 5.9 to 5.11 in Friedberg. Moreover, it means that

(1) V can be rewritten as the direct sum of eigenspaces
(2) m_i is algebraic multiplicity
(3) $\dim(E_{\lambda_i})$ is geometric multiplicity.

Theorem 1.1 (**Schur's Lemma**). *Let V be a finite dimensional vector space over $\mathbb{F}$ and $T : V \to V$ be an linear operator. Then T is triangularizable if and only if the characteristic polynomial of T has* $\dim(V)$ *roots* (*counted with multiplicities*) *in* $\mathbb{F}$

Remark. If $\mathbb{F} = \mathbb{C}$ (algebraic closure), then, by the fundamental theorem of algebra, every matrix $A \in M_{n\times n}(\mathbb{C})$ can be triangularizable. However, if $\mathbb{F} = \mathbb{R}$ ($x^2 + 1$ does not split on $\mathbb{R}$), then we can consider the rotation matix $R(\phi)$, where $0 < \phi < \pi$. Since the rotation matrix in $\mathbb{R}^2$, says

$$R = R(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \in M_{2\times 2}(\mathbb{R}),$$

where $0 < \phi < \pi$. There does not exist any vector $v \neq 0$ such that $Rv = \lambda v$ for some λ. Since

$$\begin{aligned} \det(R - \lambda I_2) &= | \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} | \\ &= | \begin{bmatrix} \cos\phi - \lambda & -\sin\phi \\ \sin\phi & \cos\phi - \lambda \end{bmatrix} | \\ &= \cos^2\phi - 2\lambda\cos\phi + \lambda^2 + \sin^2\phi \\ &= \lambda^2 - 2\cos\phi\lambda + 1, \end{aligned}$$

so we use the discriminant of a quadratic polynomial to get that $D = 4\cos^2\phi - 4 < 0$, that is, there are no solution in $\mathbb{R}$. Hence the characteristic polynomial of R does not split over $\mathbb{R}$.

Lemma. *Let $T : V \to V$ be as in theorem 1.1 and assume that the characteristic polynomial of T has $n = \dim(V)$ roots in $\mathbb{F}$. If $W \subsetneq V$ is an invariant subspace under T, then there exists a non-zero vector x in V such that $x \notin W$ and $W + span(\{x\})$ is also T-invariant.*

Proof. let $\alpha = \{x_1, \cdots, x_k\}$ be a basis for W and extend α by adjointing $\alpha' = \{x_{k+1}, \cdots, x_n\}$ to form a basis $\beta = \alpha \cup \alpha'$ for V. Let $W' = span(\alpha')$. Clearly, $V = W \oplus W'$ because we know that the fact that

> Let β_1 and β_2 be disjoint bases for subspaces W_1 and W_2, respectively, of a vector space V. If $\beta_1 \cup \beta_2$ is a basis for V, then $V = W_1 \oplus W_2$.

We define a linear operator $P : V \to V$ by

$$P(a_1x_1 + \cdots + a_nx_n) = a_1x_1 + \cdots + a_kx_k.$$

Clearly, $W' = Ker(P)$, $W = Im(P)$, and $P^2 = P$. Hence P is the projection on W with kernal W'. Moreover $I - P$ is also the projection on W' with kernel W. Since

$$\begin{aligned} &(I - P)(a_1x_1 + \cdots + a_nx_n) \\ &= I(a_1x_1 + \cdots + a_nx_n) - P(a_1x_1 + \cdots + a_nx_n) \\ &= a_1x_1 + \cdots + a_nx_n - a_1x_1 + \cdots + a_kx_k \\ &= a_{k+1}x_{k+1} + \cdots + a_nx_n, \end{aligned}$$

$W = Ker(I - P)$, $W' = Im(I - P)$, and $(I - P)^2 = (I - P)(I - P) = I - P^2 - P + P = I - P^2 = I - P$. If the basis of V is orthonormal (the Grame-Schmidt process), then it is clear that $W' = W^\perp$ by theorem 6.7 in Friedbreg. Since V is a finite dimensional vector space, so $(W^\perp)^\perp = W$ for all subspace W of V. It implies that P

is an orthogonal projection.Let $S = (I - P) \circ T \equiv (I - P)T$. Since $Im(I - P) = W'$, so $Im(S) \subseteq Im(I - P) = W'$, that is, W' is S-invariant subspace ($S(W') \subseteq W'$). Now, we claim that the set of eigenvalues of $S|_W$ is a subset of the root of eigenvalues of T. Since W is T-invariant (by assumption), so we compute the matrix representation of $[T]_\beta^\beta$ is the form

$$[T]_\beta^\beta = \begin{bmatrix} A & B \\ O & C \end{bmatrix}.$$

Clearly, $A = [T|_W]_\alpha^\alpha$ is $k \times k$ block matrix and $C = [S|_{W'}]_{\alpha'}^{\alpha'}$ is a $(n-k) \times (n-k)$ block matrix. Hence

$$\det(T - \lambda I) = \det(T|_W - \lambda I) \cdot \det(S|_{W'} - \lambda I).$$

From Corollary 1.1, we are done. Since all the eigenvalues of T lie in the field $\mathbb{F}$ ($\because$ the characteristic polynomial n roots), so by previous discussion, the same is true of all the the eigenvalues of $S|_{W'}$. Then there exists a nonzero vector x in W' ($x \notin W$) such that $Sx = \lambda x$, for some $\lambda \in \mathbb{F}$. It implies that

$$\begin{aligned} (I - P)(Tx) = \lambda x &\implies Tx - PTx = \lambda x \\ &\implies Tx = \lambda x + PTx \in \mathit{span}(\{x\}) + W. \end{aligned}$$

Finally, we show that $W + \mathit{span}(\{x\})$ is T-invariant. For all $y \in W + \mathit{span}(\{x\})$, there exists $z \in W$ and $\lambda \in \mathbb{F}$ such that $y = z + \lambda x$. Then

$$\begin{aligned} T(y) = T(z + \lambda x) &= T(z) + \lambda T(x) \\ &= T(z) + \lambda(x) + \lambda PT(x) \in \mathit{span}(\{x\}) + W. \end{aligned}$$

□

Now we go back to prove Theorem 1.1 (**Schur's Lemma**).

Proof. Suppose that T is triangularizable. Then there exists an ordered basis β for V such that $[T]_\beta$ is an upper triangular matrix. Hence the eigenvalues of T are the diagonal entries of this matrix. They are elements of $\mathbb{F}$. It means that the characteristic polynomial splits over $\mathbb{F}$. Conversely, suppose that the condition holds. Let λ be eigenvalues of T, x_i be an eigenvector of T correspond with λ and $W_1 = \mathit{span}(\{x_1\})$. Clearly, W_1 is T-invariant. From lemma, there exists a non-zero vector $x \notin W_1$ such that $W_1 + \mathit{span}(\{x_1\})$ is also T-invariant. Continuing the process,we get an increasing sequence of T-invariant subspace of V, that is, $W_1 \subseteq \cdots \subseteq W_k$ with $W_i = \mathit{span}(\{x_1, \cdots, x_i\})$ for all $i \geq 1$. Again, by lemma, there exists a nonzero vector W_k such that $W_k \ni W_{k+1} = W_k + \mathit{span}(\{x_{k+1}\})$ is also T-invariant. We continue the process until we have produced a basis for V. Hence, by Proposition 1.1, T is triangularizable. $\square$

Corollary 1.2. *If $T : V \to V$ is triangularizable with eigenvalues λ_i with respective multiplicities m_i, then there eixsts an order basis β for V such that $[T]_\beta$ is upper triangular matrix, and the diagonal entries of $[T]_\beta$ are m_1 λ_1's followed by m_2 λ_2's and so on.*

Under the assumption that all the eigenvalues of $T : V \to V$ lie in the foeld $\mathbb{F}$ over which V is defined. We recall that if T is a linear operator (or matrix) and $p(t) = a_n t^n + a_{n-1}t^{n-1}T^{n-1} + \cdots + a_0$ is a polynomial, then we can define a new linear mapping

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

Theorem 1.2 (**Cayley-Hamilton theorem**). *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p(t) = \det(T - tI)$ be its characteristic polynomial. Assume that $p(t)$ has $\dim(V)$ roots in the field $\mathbb{F}$ over which V is defined. Then $p(T) = 0$, which is a zero transformation on V.*

Remark. It follows form Exercise 6-4-16 in Friedberg.

Proof. It suffices to show that $p(T)(x) = 0$ for all the vectors in some basis of V. By Theorem 1.1 (**Schur's Lemma**), there exists an ordered basis $\beta = \{x_1, \cdots, x_n\}$ for V such that $W_i = span(\{x_1, \cdots, x_i\})$ is T-invariant for all $1 \leq i \leq n$. Since all the eigenvalues of T lie in $\mathbb{F}$, so

$$p(t) = \pm 1 \cdot (t - \lambda) \cdots (t - \lambda_n)$$

for some $\lambda_i \in \mathbb{F}$ (not necessary distinct). If the factors here are ordered in the same fashion as the diagonal entries of $[T]_\beta^\beta$, then

$$T(x_i) = \lambda_i x_i + y_{i-1},$$

where $y_i \in W_{i-1}$ and $i \geq 2$, in particular, $T(x_1) = \lambda_1 x_1$. Now, we use the induction on i. For $i = 1$, we get that

$$\begin{aligned} p(T)(x_1) &= \pm(T - \lambda_1 I) \cdots (T - \lambda_n I)(x_1) \\ &= \pm(T - \lambda_2 I) \cdots (I - \lambda_n I)(T - \lambda_1)(x_1) = 0. \end{aligned}$$

The last equality follows that the powers of T commutes with each other and with I. Suppose that $p(T)(x_i) = 0$ for all $i \leq k$. We compute $p(T)(x_{k+1})$. It is clear that only the factors $(T - \lambda_1 I), \cdots, (T - \lambda_k I)$ are needed to send x_i to 0 for $i \leq k$. As before, we can rearrange the factors in $p(T)$ to obtain

$$\begin{aligned} p(T)(x_{k+1}) &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(x_{k+1}) \\ &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(T-\lambda_{k+1} I)(x_{k+1}) \\ &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(y_k) = 0. \end{aligned}$$

The last equality follows that $T(x_{k+1}) = \lambda_{k+1}x_{k+1} + y_k$ and $y_k \in W_k$. By induction, the other factors in $p(T)$ send all the vectors in this subspace to 0. We are done. □

Remark. There are another proof of Theorem 1.2 (**Cayley-Hamilton theorem**). One can see Theorem 5.23 in Friedbreg.

Remark. Suppose that $A \in M_{n\times n}(\mathbb{F})$. If A is invertible, then $\det(A) \neq 0$. We consider that the characteristic polynomial of A

$$p(A) = \det(A - tI) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0.$$

Clearly, $a_0 = \det(A)$. By Theorem 1.2 (**Cayley-Hamilton theorem**), we get that

$$p(A) = (-1)^n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0.$$

Moreover, we also get that

$$A^{-1} = \frac{-1}{\det(A)}((-1)^n A^{n-1} + \cdots + a_1 I).$$

One can see Excercise 5-1-20, 5-1-21 and 5-4-18 in Friedbreg.