Chapter 8

Week8

8.1. Monday for MAT3040

Reviewing.

- If $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, then

$$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

  for some basis $\mathcal{A}$. In other words, T is **triangularizable** with the diagonal entries $\lambda_1,\dots,\lambda_n$.

R I hope you appreciate this result. Consider the example below: In linear algebra we have studied that the matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable, and the characteristic polynomial is given by

$$\mathcal{X}_A(x) = (x-1)^2.$$

However, the theorem above claims that $\boldsymbol{A}$ is *triangularizable*, with diagonal entries 1 and 1. The diagonalization of $\boldsymbol{A}$ only uses the eigenvector of $\boldsymbol{A}$, but the 1-eigenspace has only 1 dimension. Fortunately, the triangularization gives a rescue such that we can make use of the generalized eigenvector

$(0,1)^{\mathrm{T}}$ (but not an eigenvector) of $\boldsymbol{A}$ by considering the mapping below:

$$
\begin{aligned}
U &= \operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}\right\}\\
\bar{A}:\quad & V/U \to V/U
\end{aligned}
$$

Here $(0,1)^{\mathrm{T}} + U$ is an eigenvector of $\bar{A}$, with eigenvalue 1.

Theorem 8.1 The linear operator T is triangularizable with diagonal entries $(\lambda_1, \dots, \lambda_n)$ if and only if

$$\mathcal{X}_T = (x - \lambda_1)\cdots(x - \lambda_n)$$

Proof. It suffices to show only the sufficiency. Suppose that there exists basis $\mathcal{A}$ such that

$$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \times & \times \\ 0 & \lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Then we compute the characteristic polynomial directly:

$$
\begin{aligned}
\mathcal{X}_T(x) &= \det[(xI - T)_{\mathcal{A},\mathcal{A}}] \\
&= \det \begin{pmatrix} x-\lambda_1 & \times & \times & \times \\ 0 & x-\lambda_2 & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & x-\lambda_n \end{pmatrix} \\
&= (x-\lambda_1)\cdots(x-\lambda_n)
\end{aligned}
$$

∎

8.1.1. Cayley-Hamiton Theorem

Proposition 8.1 — A Useful Lemma. Suppose that $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, then $\mathcal{X}_T(T) = 0$.

Proof. Since $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, we imply T is triangularizable under some basis $\mathcal{A}$. Note that

- $T \mapsto (T)_{\mathcal{A},\mathcal{A}}$ is an isomorphism between $\text{Hom}(V,V)$ and $M_{n\times n}(\mathbb{F})$,
- $(\underbrace{T\circ T\circ\cdots\circ T}_{m \text{ times}})_{\mathcal{A},\mathcal{A}} = [(T)_{\mathcal{A},\mathcal{A}}]^m$, for any m,

It suffices to show $\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}})$ is the zero matrix (why?):

$$\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}}) = ((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I}).$$

Observe the matrix multiplication

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_i \boldsymbol{I})\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1-\lambda_i & \times & \times & \times \\ 0 & \lambda_2-\lambda_i & \cdots & \times \\ 0 & \cdots & \ddots & \times \\ 0 & 0 & \cdots & \lambda_n-\lambda_i \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \text{span}\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_{i-1}\}$$

Therefore, for any $\boldsymbol{v}\in V$,

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})\boldsymbol{v} \in \text{span}\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_{n-1}\}.$$

Applying the same trick, we conclude that

$$((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}, \quad \forall \boldsymbol{v}\in V,$$

i.e., $\mathcal{X}_T((T)_{\mathcal{A},\mathcal{A}}) = ((T)_{\mathcal{A},\mathcal{A}} - \lambda_1 \boldsymbol{I})\cdots((T)_{\mathcal{A},\mathcal{A}} - \lambda_n \boldsymbol{I})$ is a zero matrix. ■

Now we are ready to give a proof for the Cayley-Hamiton Theorem:

Proof. Suppose that $\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{F}[x]$. By considering algebrically closed field $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we imply

$$\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \quad (8.1a)$$

$$= (x-\lambda_1)\cdots(x-\lambda_n), \quad \lambda_i \in \overline{\mathbb{F}} \quad (8.1b)$$

By applying proposition (8.1), we imply $\mathcal{X}_T(T) = 0$, where the coefficients in the formula $\mathcal{X}_T(T) = 0$ w.r.t. T are in $\overline{\mathbb{F}}$.

Then we argue that these coefficients are essentially in $\mathbb{F}$. Expand the whole map of $\mathcal{X}_T(T)$:

$$\mathcal{X}_T(T) = (T-\lambda_1 I)\cdots(T-\lambda_n I) \quad (8.2a)$$

$$= T^n - (\lambda_1 + \cdots + \lambda_n)T^{n-1} + \cdots + (-1)^n \lambda_1 \cdots \lambda_n I \quad (8.2b)$$

$$= T^n + a_{n-1}T^{n-1} + \cdots + a_0 I \quad (8.2c)$$

where the derivation of (8.2c) is because that the polynomial coefficients for (8.1a) and (8.1b) are all identical.

Therefore, we conclude that $\mathcal{X}_T(T) = 0$, under the field $\mathbb{F}$. ■

Corollary 8.1 $m_T(x) \mid \mathcal{X}_T(x)$. More precisely, if

$$\mathcal{X}_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k},\ e_i > 0, \forall i$$

where p_i's are distinct, monic, and irreducible polynomials. Then

$$m_T(x) = [p_1(x)]^{f_1} \cdots [p_k(x)]^{f_k}, \text{ for some } 0 < f_i \leq e_i, \forall i$$

Proof. The statement $m_T(x) \mid \mathcal{X}_T(x)$ is from Cayley-Hamiton Theorem. Therefore, $0 \leq f_i \leq e_i, \forall i$. Suppose on the contrary that $f_i = 0$ for some i. w.l.o.g., $i = 1$.

It's clear that $\gcd(p_1, p_j) = 1$ for $\forall j \neq 1$, which implies

$$a(x)p_1(x) + b(x)p_j(x) = 1, \text{ for some } a(x), b(x) \in \mathbb{F}[x].$$

Considering the field extension $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we have $p_1(x) = (x - \mu_1) \cdots (x - \mu_\ell)$. For any root μ_m of p_1, $m = 1, \dots, \ell$, we have

$$a(\mu_m)p_1(\mu_m) + b(\mu_m)p_j(\mu_m) = 1 \implies b(\mu_m)p_j(\mu_m) = 1 \implies p_j(\mu_m) \neq 0,$$

i.e., μ_m is not a root of p_j, $\forall j \neq 1$.

Therefore, μ_m is a root of $\mathcal{X}_T(x)$, but not a root of $m_T(x)$. Then μ_m is an eigenvalue of T, e.g., $T\boldsymbol{v} = \mu_m \boldsymbol{v}$ for some $\boldsymbol{v} \neq \mathbf{0}$. Recall that $m_{T,\boldsymbol{v}} = x - \mu_m$, we imply $m_{T,\boldsymbol{v}} = x - \mu_m \mid m_T(x)$, which is a contradiction. ■

■ **Example 8.1** We can use Corollary (8.1), a stronger version of Cayley-Hamiltion Theorem to determine the minimal polynomials:

1. For matrix $\boldsymbol{A} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, we imply $\mathcal{X}_A(x) = (x^2 + x + 1)^1$. Since $x^2 + x + 1$ is irreducible in $\mathbb{R}$, we have $m_A(x) = x^2 + x + 1$.
2. For matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

we imply $\mathcal{X}_A(x) = (x-1)^2(x-2)^2$.

By Corollary (8.1), we imply both $(x-1)$ and $(x-2)$ should be roots of $m_T(x)$, i.e., $m_A(x)$ may have the four options:

$$(x-1)^2(x-2)^2, \text{ or}$$
$$(x-1)(x-2)^2, \text{ or}$$
$$(x-1)^2(x-2), \text{ or}$$
$$(x-1)(x-2).$$

■

8.1.2. Primary Decomposition Theorem

We know that not every linear operator is diagonalizable, but diagonalization has some nice properties:

Definition 8.1 [diagonalizable] The linear operator $T : V \to V$ is diagonalizable over $\mathbb{F}$ if and only if there exists a basis $\mathcal{A}$ of V such that

$$(T)_{\mathcal{A},\mathcal{A}} = \text{diag}(\lambda_1, \dots, \lambda_n),$$

where λ_i's are not necessarily distinct.

Proposition 8.2 If the linear operator $T : V \to V$ is diagonalizable, then

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k),$$

where μ_i's are **distinct**.

Proof. Suppose T is diagonalizable, then there exists a basis $\mathcal{A}$ of V such that

$$(T)_{\mathcal{A},\mathcal{A}} = \text{diag}(\mu_1, \dots, \mu_1, \mu_2, \dots, \mu_2, \dots, \mu_k, \dots, \mu_k)$$

It's clear that $((T)_{\mathcal{A},\mathcal{A}} - \mu_1 \boldsymbol{I}) \cdots ((T)_{\mathcal{A},\mathcal{A}} - \mu_k \boldsymbol{I}) = \mathbf{0}$, i.e., $m_T(x) \mid (x - \mu_1) \cdots (x - \mu_k)$.

Then we show the minimality of $(x - \mu_1) \cdots (x - \mu_k)$. In particular, if $(x - \mu_i)$ is omitted for any $1 \le i \le k$, then it's easy to show

$$(T_{\mathcal{A},\mathcal{A}} - \mu_1 \boldsymbol{I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_{i-1} \boldsymbol{I})(T_{\mathcal{A},\mathcal{A}} - \mu_{i+1} \boldsymbol{I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_k \boldsymbol{I}) \ne \mathbf{0},$$

since all μ_i's are distinct. Therefore, $m_T(x)$ will not divide $(x - \mu_1) \cdots (x - \mu_{i-1})(x - \mu_{i+1}) \cdots (x - \mu_k)$ for any i, i.e.,

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k)$$

∎

R The converse of proposition (8.2) is also true, which is a special case for the Primary Decomposition Theorem.

Theorem 8.2 — Primary Decomposition Theorem. Let $T: V \to V$ be a linear operator with

$$m_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k},$$

where p_i's are distinct, monic, and irreducible polynomials. Let $V_i = \ker([p_i(x)]^{e_i}) \leq V, i = 1, \dots, k$, then

1. Each V_i is T-invariant ($T(V_i) \leq V_i$)
2. $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$
3. Consider $T\,|_{V_i}: V_i \to V_i$, then

$$m_{T|V_i}(x) = [p_i(x)]^{e_i}$$