7.4. Wednesday for MAT3040

Reviewing.

- Given the polynomial $f(x) \in \mathbb{F}[x]$, we extend it into the linear operator $f(T): V \to V$.
- The minimal polynomial $m_T(x)$ is defined to be the polynomial with least degree such that
$$m_T(T) = \mathbf{0}_{V \to V},$$
i.e., $[m_T(T)]\boldsymbol{v} = 0_{\boldsymbol{V}}, \forall \boldsymbol{v} \in V$.
- The minimial polynomial of a vector $\boldsymbol{v}$ relative to T is defined to be the polynomial $m_{T,\boldsymbol{v}}(x)$ with the least degree such that
$$m_{T,\boldsymbol{v}}(T)(\boldsymbol{v}) = 0$$
- If $f(T) = \mathbf{0}_{V \to V}$, then we imply $m_T(x) \mid f(x)$. If $[g(T)](\boldsymbol{w}) = 0_V$, following the similar argument, we imply $m_{T,\boldsymbol{w}}(x) \mid g(x)$.
- In particular, $m_T(T)\boldsymbol{w} = \mathbf{0}$, which implies $m_{T,\boldsymbol{w}}(x) \mid m_T(x)$.

7.4.1. Cayley-Hamiton Theorem

Let's raise an motivative example first:

■ **Example 7.7** Consider the matrix and its induced mapping $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. It has the characteristic polynomial
$$\mathcal{X}_A = (x-1)(x-2).$$

- Note that $m_A(x)$ cannot be with degree one, since otherwise $m_A(x) = x - k$ with some k, and

$$m_A(\boldsymbol{A}) = \boldsymbol{A} - k\boldsymbol{I} = \begin{pmatrix} 1-k & 0 \\ 0 & 2-k \end{pmatrix} \neq \boldsymbol{0}, \quad \forall k,$$

which is a contradiction.

- However, one can verify that the $m_A(x)$ is with degree 2:

$$m_A(x) = (x-1)(x-2).$$

- The minimial polynomial with eigenvectors can be with degree 1:

$$\boldsymbol{w} = [0,1]^{\mathrm{T}} \implies (A-2I)\boldsymbol{w} = \boldsymbol{0} \implies m_{A,\boldsymbol{w}}(x) = x-2$$

R More generally, given an eigen-pair $(\lambda, \boldsymbol{v})$, the minimal polynomial of an $\boldsymbol{v}$ has the explicit form

$$m_{T,\boldsymbol{v}}(x) = (x-\lambda) \implies (x-\lambda) \mid m_T(x)$$

Now we want to relate the characterstic polynomial $m_T(x)$ with $\mathcal{X}_T(x)$. Suppose that

$$\mathcal{X}_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_k)^{e_k} \in \mathbb{F}[x]. \tag{7.1}$$

Then we imply

- λ_i is an eigenvalue of T;
- $(x-\lambda_i) \mid m_T(x)$;

which implies that $(x-\lambda_1)\cdots(x-\lambda_k) \mid m_T(x)$.

Furthermore, does $m_T(x)$ possess other factors? In other words, does $(x-\lambda_i)^{f_i} \mid m_T(x)$ when $f_i > e_i$? Answer: No.

Theorem 7.1 — **Cayley-Hamilton.** $m_T(x) \mid \mathcal{X}_T(x)$. In particular, $\mathcal{X}_T(T) = \mathbf{0}$.

The nice equality in (7.1) does not necessarily hold. Sometimes $\mathcal{X}_T(x)$ cannot be factorized into linear factors in $\mathbb{F}[x]$, e.g., $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ in $\mathbb{R}$.

However, for every $f(x) \in \mathbb{F}[x]$, we can extend $\mathbb{F}$ into the algebraically closed set $\overline{F} \supseteq \mathbb{F}$ such that

$$f(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}$$

where $\lambda_i \in \overline{\mathbb{F}}$.

For example, for $f(x) = x^2 + 1 \in \mathbb{R}[x]$, we can extend $\mathbb{R}$ into $\mathbb{C}$ to obtain

$$f(x) = (x + i)(x - i).$$

Therefore, the general proof outline for the Cayley-Hamilton Theorem is as follows:

- Consider the case where $m_T(x), \mathcal{X}_T(x)$ are both in $\overline{F}[x]$
- Show that $m_T(x) \mid \mathcal{X}_T(x)$ under $\overline{F}[x]$.

Before the proof, let's study the invariant subspaces, which leads to the decomposition of charactersitc polynomial:

Assumption. From now on, we assume that V is finite dimensional by default.

Definition 7.11 [Invariant Subspace] An **invariant subspace** of a linear operator $T : V \to V$ is a subspace $W \leq V$ that is preserved by T, i.e., $T(W) \subseteq W$. We also call W as T-invariant. ▪

▪ **Example 7.8**
1. V itself is T-invariant.
2. For the eigenvalue λ, the associated λ-eigenspace $U = \ker(T - \lambda I)$ is T-invariant.
3. More generally, $U = \ker(g(T))$ is T-invariant for any polynomial g:

If $\boldsymbol{v} \in \ker(g(T))$, i.e., $g(T)\boldsymbol{v} = \mathbf{0}$, it suffices to show $T(\boldsymbol{v}) \in \ker(g(T))$:

$$\begin{aligned} g(T)[T(\boldsymbol{v})] &= (a_m T^m + \cdots + a_0 I)[T(\boldsymbol{v})] \\ &= (a_m T \circ T^m + \cdots + a_1 T \circ T + a_0 T \circ I)(\boldsymbol{v}) \\ &= T[g(T)\boldsymbol{v}] = T(\mathbf{0}) = \mathbf{0} \end{aligned}$$

4. For $\boldsymbol{v} \in \ker(T - \lambda I)$, $U = \text{span}\{\boldsymbol{v}\}$ is T-invariant.

Proposition 7.11 Suppose that $T : V \to V$ is a linear transformation and $W \leq V$ is T-invariant, then we construct the subspace mapping and the recipe mapping

$$\begin{aligned} T\,|_W\colon \quad & W \to W \\ \text{with} \quad & \boldsymbol{w} \mapsto T(\boldsymbol{w}) \end{aligned} \tag{7.2a}$$

$$\begin{aligned} \tilde{T} : \quad & V/W \to V/W \\ \text{with} \quad & \boldsymbol{v} + W \mapsto T(\boldsymbol{v}) + W \end{aligned} \tag{7.2b}$$

$$T\,|_W\colon W \to W$$

which leads to the decomposition of the charactersitic polynomial:

$$\mathcal{X}_T(x) = \mathcal{X}_{T|_W}(x)\mathcal{X}_{\tilde{T}}(x).$$

Proof. Suppose $\mathcal{C} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a basis of W, and extend it into the basis of V, denoted as

$$\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$$

Therefore, $\overline{\mathcal{B}} = \{\boldsymbol{v}_{k+1} + W, \ldots, \boldsymbol{v}_n + W\}$ is a basis of V/W. By Homework 2, Question 5, the representation $(T)_{\mathcal{B},\mathcal{B}}$ can be written as the block matrix

$$(T)_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} (T|_W)_{\mathcal{C},\mathcal{C}} & \times \\ \mathbf{0} & (\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} \end{pmatrix}_{(k+(n-k))\times(k+(n-k))}$$

Therefore, the characteristic polynomial of T can be calculated as:

$$\begin{aligned}\mathcal{X}_T(x) &= \det((T)_{\mathcal{B},\mathcal{B}} - xI) \\ &= \det((T|_U)_{\mathcal{C},\mathcal{C}} - xI) \cdot \det((\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} - xI)\end{aligned}$$

■

Proposition 7.12 Suppose that

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

where λ_i's are not necessarily distinct. Then there exists a basis of V, say $\mathcal{A}$, such that

$$(T)_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

R This proposition is the generalization of the eigenvalue decomposition studied in Linear Algebra:

> **Definition 7.12** [Eigenvalue Decomposition] A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) is said to admit an **eigenvalue decomposition** if there exists a nonsingular $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ and a collection of scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that
>
> $$\boldsymbol{A} = \boldsymbol{V}\Lambda\boldsymbol{V}^{-1}$$
>
> where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Proof. The proof is by induction on n, i.e., suppose the results hold for size $= n - 1$, and we aim to show this result holds for size $= n$.

1. **Step 1**: Argue that there exists the associated eigenvector $\boldsymbol{v}$ of λ_1 under the linear operator T.

   Consider any basis $\mathcal{M}$, by MAT2040, there exists associated eigenvector of λ_1, say $\boldsymbol{y} \in \mathbb{C}^n$ such that

$$(T)_{\mathcal{M},\mathcal{M}} \cdot \boldsymbol{y} = \lambda_1 \boldsymbol{y}$$

Since the operator $(\cdot)_{\mathcal{M}} : V \to \mathbb{C}^n$ is an isomorphism, there exists $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$ such that $(\boldsymbol{v})_{\mathcal{M}} = \boldsymbol{y}$. It follows that

$$(T)_{\mathcal{M},\mathcal{M}}(\boldsymbol{v})_{\mathcal{M}} = \lambda_1(\boldsymbol{v})_{\mathcal{M}} \implies (T\boldsymbol{v})_{\mathcal{M}} = (\lambda_1\boldsymbol{v})_{\mathcal{M}} \implies T\boldsymbol{v} = \lambda_1\boldsymbol{v}$$

2. **Step 2**: Dimensionality reduction of $\mathcal{X}_T(x)$: Construct $W = \text{span}\{\boldsymbol{v}\}$, which is T-invariant. By the proof of proposition (7.12), we imply there is a basis of V,say $B := \{\boldsymbol{v},\boldsymbol{h}_2,\dots,\boldsymbol{h}_n\}$, such that

$$(T)_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} (T|_W)_{\{\boldsymbol{v}\}} & \times \\ \mathbf{0} & (\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \times \\ \mathbf{0} & (\tilde{T})_{\overline{\mathcal{B}},\overline{\mathcal{B}}} \end{pmatrix}$$

where $\tilde{T} : V/W \to V/W$ admits the characteristic polynomial

$$\mathcal{X}_{\tilde{T}}(x) = (x-\lambda_2)\cdots(x-\lambda_n)$$

3. **Step 3:** Applying the induction, there exists basis $\overline{\mathcal{C}}$ of V/W, i.e.,

$$\overline{\mathcal{C}} = \{\boldsymbol{w}_2 + W,\dots,\boldsymbol{w}_n + W\}$$

such that

$$(\tilde{T})_{\overline{\mathcal{C}},\overline{\mathcal{C}}} = \text{diag}(\lambda_2,\dots,\lambda_n)$$

4. **Step 4:** Therefore, we construct the set $\mathcal{A} := \{\boldsymbol{v},\boldsymbol{w}_2,\dots,\boldsymbol{w}_n\}$. We claim that
   - $\mathcal{A}$ is a basis of V
   - $$(T)_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & (\tilde{T})_{\overline{\mathcal{C}},\overline{\mathcal{C}}} \end{pmatrix} = \text{diag}(\lambda_1,\dots,\lambda_n)$$

■

Proposition 7.13 Suppose that $\mathcal{X}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, then $\mathcal{X}_T(T) = \mathbf{0}$.

R One special case is that $\boldsymbol{A} = \text{diag}(\lambda_1,\dots,\lambda_n)$. The results for proposition (7.13) gives

$$(A - \lambda_1 \boldsymbol{I}) \cdots (A - \lambda_n \boldsymbol{I}) \text{ is a zero matrix}$$