Cayley-Hamilton Theorem

Seanie Lee

11 May 2019

Theorem 1. *$\boldsymbol{W}$ is $\boldsymbol{T}$-cyclic subspace of $\boldsymbol{V}$ generated by a nonzero vector $\boldsymbol{v} \in \boldsymbol{V}$ and dim$\boldsymbol{W} = k$*

1. *$\{\boldsymbol{v}, \boldsymbol{T}\boldsymbol{v}, \dots, \boldsymbol{T}^{k-1}\boldsymbol{v}\}$ is a basis for $\boldsymbol{W}$*
2. *$a_0\boldsymbol{v} + a_1\boldsymbol{T}\boldsymbol{v} + \cdots + a_{k-1}\boldsymbol{T}^{k-1}\boldsymbol{v} + \boldsymbol{T}^k\boldsymbol{v} = 0 \Longrightarrow \phi_{\boldsymbol{T}|\boldsymbol{W}}(t) = t^k + a_{k-1}t^{k-1} + \cdots + a_1 t + a_0$*

Proof. 1. Let l be the largest integer such that $\beta \coloneqq \{\boldsymbol{v}, \boldsymbol{T}\boldsymbol{v}, \dots, \boldsymbol{T}^{l-1}\boldsymbol{v}\}$ is linearly independent. Let $\boldsymbol{Z} \coloneqq$ spanβ. Then β is a basis for $\boldsymbol{Z}$. Since β is linearly independent set and $\beta \cup \{\boldsymbol{T}^i\boldsymbol{v}\}$ is linearly dependent, $\boldsymbol{T}^i\boldsymbol{v} \in$ spanβ for $i = l, l+1, \dots$. So, $\boldsymbol{T}^l\boldsymbol{v} \in \boldsymbol{Z}$. Note that $\boldsymbol{Z}$ is $\boldsymbol{T}$-invariant because of the following reason.

$$\begin{aligned} \boldsymbol{w} &= b_0\boldsymbol{v} + b_1\boldsymbol{T}\boldsymbol{v} + \cdots + b_{l-1}\boldsymbol{T}^{l-1} \in \boldsymbol{Z} \\ \boldsymbol{T}\boldsymbol{w} &= b_0\boldsymbol{T}\boldsymbol{v} + b_1\boldsymbol{T}^2\boldsymbol{v} + \cdots + b_{l-2}\boldsymbol{T}^{l-1}\boldsymbol{v} + b_{l-1}\boldsymbol{T}^l\boldsymbol{v} \in \boldsymbol{Z} \end{aligned}$$

Moreover $\boldsymbol{Z}$ is $\boldsymbol{T}$-invariant subspace containing $\boldsymbol{v}$ and $\boldsymbol{W}$ is the smallest $\boldsymbol{T}$-invariant subspace containing $\boldsymbol{v}$, $\boldsymbol{W} \leq \boldsymbol{Z}$ which implies that $k \leq l$. But dim$\boldsymbol{W} = k$ and $l \leq k$, thus $k = l$
$\therefore \beta$ is linearly independent subset of $\boldsymbol{W}$.
Since β is linearly independent and $\beta \cup \{\boldsymbol{T}^i\boldsymbol{v}\}$ is linearly dependent $\Longrightarrow \boldsymbol{T}^i\boldsymbol{v} \in$ spanβ for $i \geq k$.
$\therefore \beta$ is a basis for $\boldsymbol{W}$

2. $[\boldsymbol{T}|_{\boldsymbol{W}}]_\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix}$ and $\phi_{[\boldsymbol{T}|_{\boldsymbol{W}}]_\beta}(t) = \det \begin{pmatrix} t & 0 & \cdots & 0 & a_0 \\ -1 & t & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t + a_{k-1} \end{pmatrix}$

$$\text{Put } b_i := \det\begin{pmatrix} t & 0 & \cdots & 0 & a_i \\ -1 & t & \cdots & 0 & a_{i+1} \\ 0 & -1 & \cdots & 0 & a_{i+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t+a_{k-1} \end{pmatrix}_{(k-i)\times(k-i)}$$

$$b_0 = \begin{pmatrix} t & 0 & \cdots & 0 & a_0 \\ -1 & t & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t+a_{k-1} \end{pmatrix}$$

$$b_0 = t\det\begin{pmatrix} t & 0 & \cdots & 0 & a_1 \\ -1 & t & \cdots & 0 & a_2 \\ 0 & -1 & \cdots & 0 & a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t+a_{k-1} \end{pmatrix} + (-1)^{k+1}a_0\det\begin{pmatrix} -1 & t & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}$$

$$\begin{aligned} b_0 &= tb_1 + (-1)^{k+1+k-1}a_0 \\ b_0 &= tb_1 + a_0 \\ b_1 &= tb_2 + a_1 \\ &\vdots \\ b_{k-2} &= tb_{k-1} + a_{k-2} \\ b_{k-1} &= t + a_{k-1} \\ b_0 &= t^k + a_{k-1}t^{k-1} + \cdots + a_1t + a_0 \end{aligned}$$

$$\therefore \phi_{[\boldsymbol{T}|_{\boldsymbol{W}}]}(t) = t^k + a_{k-1}t^{k-1} + \cdots + a_1t + a_0$$

Definition 2. *Let $f(t) = a_0 + a_1t + \cdots + a_kt^k$ polynomial. Then $f(\boldsymbol{T}) := a_0\boldsymbol{I} + a_1\boldsymbol{T} + \cdots + a_k\boldsymbol{T}^k$*

Theorem 3 (Cayley-Hamilton Theorem)**.** *Let $\boldsymbol{T}$ a linear operator on a finite dimensional vector space $\boldsymbol{V}$ over $\boldsymbol{F}$ and let $\phi_{\boldsymbol{T}}(t)$ be the characteristic polynomial of $\boldsymbol{T}$. Then $\phi_{\boldsymbol{T}}(\boldsymbol{T}) = \boldsymbol{0}$, the zero transformation. That is $\boldsymbol{T}$ satisfies its characteristic equation.*

Proof. We show that $\phi_{\boldsymbol{T}}(\boldsymbol{T})(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in \boldsymbol{V}$. Suppose $\mathbf{v} = \mathbf{0}$. It is trivial to show that $\phi_{\boldsymbol{T}}(\boldsymbol{T})(\mathbf{v}) = \mathbf{0}$ because $\phi_{\boldsymbol{T}}(\boldsymbol{T})$ is linear map. So suppose $\mathbf{v} \neq \mathbf{0}$. Let $\boldsymbol{W}$ be the $\boldsymbol{T}$-cyclic subspace generated by $\mathbf{v}$ and suppose that $\dim \boldsymbol{W} = k$. Since $S := \{\mathbf{v}, \boldsymbol{T}\mathbf{v}, \ldots, \boldsymbol{T}^{k-1}\mathbf{v}\}$ is a basis for $\boldsymbol{W}$, there exits scalars $a_0, a_1, \ldots, a_{k-1}$ such that

$$a_0\mathbf{v} + a_1\boldsymbol{T}\mathbf{v} + \cdots + a_{k-1}\boldsymbol{T}^{k-1}\mathbf{v} + \boldsymbol{T}^k\mathbf{v} = \mathbf{0}$$

Since $g(t) = t^k + a_{k-1}t^{k-1} + \cdots + a_1t + a_0$ is characteristic polynomial of $\boldsymbol{T}|_{\boldsymbol{W}}$. Combining the two equations yields

$$g(\boldsymbol{T})(\mathbf{v}) = (\boldsymbol{T}^k + a_{k-1}\boldsymbol{T}^{k-1} + \cdots + a_1\boldsymbol{T} + a_0\boldsymbol{I})(\mathbf{v}) = \mathbf{0}$$

Since $\boldsymbol{W}$ is $\boldsymbol{T}$-invariant subspace of $\boldsymbol{V}$, $g(t)$ divides $\phi_{\boldsymbol{T}}(t)$ such that $\phi_{\boldsymbol{T}}(t) = h(t)g(t)$ where $h(t)$ is polynomial. So

$$\phi_{\boldsymbol{T}}(\boldsymbol{T})(\mathbf{v}) = h(\boldsymbol{T})g(\boldsymbol{T})(\mathbf{v}) = h(\boldsymbol{T})(g(\boldsymbol{T})(\mathbf{v})) = h(\boldsymbol{T})(\mathbf{0}) = \mathbf{0}$$

$\square$