Chapter 7

Linear Operators

在這一章中, 我們探討特別的一種但很常用的 linear transformation, 稱為 linear operator. 它是定義域與對應域相同的 linear transformation. 這一章我們介紹其基本性質，下一章再更進一步探討對角化問題。

7.1. Change of Basis

在這一節中, 我們介紹 change of basis 的概念, 了解到一個 linear operator 換了 ordered basis 後其表現矩陣的關係. 這個概念能幫助我們以後處理矩陣對角化的問題.

我們知道一個 linear transformation, 當我們用不同的 ordered bases 所得的 matrix representation 會不同. 假設 β,β' 為 V 的兩組 ordered bases, 而 γ,γ' 為 W 的兩組 ordered basis. 對於 linear transformation $T: V \to W$, 其對應於這兩對 ordered bases 的 matrix representations $[T]_\beta^\gamma$ 和 $[T]_{\beta'}^{\gamma'}$ 之間會有甚麼關係呢? 首先我們考慮 identity map $\mathrm{id}: V \to V$. 注意雖然是 identity map, 但其 matrix representation 未必會是 identity matrix. 事實上, 當我們定義域和對應域都選同一組 ordered basis $\beta = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$, 則由於 $\mathrm{id}(\mathbf{v}_i) = \mathbf{v}_i$, 故其 matrix representation 是 identity matrix. 但若定義域是使用 β 這一組 ordered basis, 而對應域選的是 $\beta' = \{\mathbf{v}'_1,\ldots,\mathbf{v}'_n\}$ 這一組 ordered basis, identity map 對應於 β,β' 的 matrix representation $[\mathrm{id}]_\beta^{\beta'}$ 其 i-th column 雖然仍和 $\mathrm{id}(\mathbf{v}_i) = \mathbf{v}_i$ 有關, 不過卻是要將 $\mathbf{v}_i$ 寫成以 $\{\mathbf{v}'_1,\ldots,\mathbf{v}'_n\}$ 為 ordered basis 的坐標表示法 $[\mathbf{v}_i]_{\beta'}$. 所以當 β 和 β' 相異時, $[\mathrm{id}]_\beta^{\beta'}$ 不是 identity matrix. 現對任意 $\mathbf{v} \in V$, 因 $\mathbf{v}$ 對於 β 的坐標表示法為 $[\mathbf{v}]_\beta$, 依 matrix representation 的性質 (Proposition 6.3.14) 可得

$$[\mathrm{id}]_\beta^{\beta'}[\mathbf{v}]_\beta = [\mathrm{id}(\mathbf{v})]_{\beta'} = [\mathbf{v}]_{\beta'}.$$

也就是說, 矩陣 $[\mathrm{id}]_\beta^{\beta'}$ 可以將 V 中元素對於 β 的坐標表示轉換成對於 β' 的坐標表示, 也因此我們稱 $[\mathrm{id}]_\beta^{\beta'}$ 為 *change-of-basis matrix.*

要注意 $\mathrm{id}:V\to V$ 是 isomorphism, 所以由 Theorem 6.3.19 (3), 我們得 $[\mathrm{id}]_{\beta}^{\beta'}$ 為 invertible 且

$$([\mathrm{id}]_{\beta}^{\beta'})^{-1}=[\mathrm{id}^{-1}]_{\beta'}^{\beta}=[\mathrm{id}]_{\beta'}^{\beta} \tag{7.1}$$

也就是說將 β 的坐標表示轉換成對於 β' 的坐標表示的 change-of-basis matrix 的 inverse 就是 β' 的坐標表示轉換成對於 β 的坐標表示的 change-of-basis matrix.

我們回到原先的問題, 假設 $T:V\to W$ 為 linear transformation 且 β,β' 為 V 的兩組 ordered bases, 而 γ,γ' 為 W 的兩組 ordered basis. 我們要探討 $[T]_{\beta}^{\gamma}$ 和 $[T]_{\beta'}^{\gamma'}$ 之間的關係. 由於 $\mathrm{id}_V:V\to V$, $T:V\to W$ 和 $\mathrm{id}_W:W\to W$ 之合成 $\mathrm{id}_W\circ T\circ \mathrm{id}_V:V\to W$ 仍為 $T:V\to W$, 所以由 Theorem 6.3.19 (2) 得

$$[\mathrm{id}_W]_{\beta}^{\gamma'}[T]_{\beta}^{\gamma}[\mathrm{id}_V]_{\beta'}^{\beta}=[T]_{\beta'}^{\gamma'}.$$

這就是所謂的 change-of basis formula, 我們將之完整敘述如下.

Theorem 7.1.1 (Change-of-basis Formula)**.** 假設 $T:V\to W$ 為 *linear transformation* 且 β,β' 為 V 的兩組 *ordered bases,* 而 γ,γ' 為 W 的兩組 *ordered basis,* 則存在 *invertible matrix* P,Q 使得 $[T]_{\beta'}^{\gamma'}=Q([T]_{\beta}^{\gamma})P$, 其中 P 為將 β' 的坐標表示轉換成 β 的坐標表示的 *change-of-basis matrix* $[\mathrm{id}_V]_{\beta'}^{\beta}$, 而 Q 為將 γ 的坐標表示轉換成 γ' 的坐標表示的 *change-of-basis matrix* $[\mathrm{id}_W]_{\gamma}^{\gamma'}$.

Example 7.1.2. 在 Example 6.3.13 中我們考慮 linear transformation $T:P_2(\mathbb{R})\to P_3(\mathbb{R})$, 其中 $T(p(x))=(x+1)p(x-1)$, $\forall p(x)\in P_2(\mathbb{R})$. 另外我們考慮 $P_2(\mathbb{R})$ 的兩組 ordered bases $\varepsilon=(x^2,x,1)$, $\beta=(p_1(x),p_2(x),p_3(x))$ 其中

$$p_1(x)=\frac{1}{2}(x^2-x),\quad p_2(x)=-x^2+1,\quad p_3(x)=\frac{1}{2}(x^2+x)$$

以及 $P_3(\mathbb{R})$ 的兩組 ordered bases $\varepsilon'=(x^3,x^2,x,1)$, $\beta'=(q_1(x),q_2(x),q_3(x),q_4(x))$ 其中

$$q_1(x)=\frac{-x^3+3x^2-2x}{6},q_2(x)=\frac{x^3-2x^2-x+2}{2},q_3(x)=\frac{-x^3+x^2+2x}{2},q_4(x)=\frac{x^3-x}{6}.$$

在 Example 6.3.13 中我們得到

$$[T]_{\varepsilon}^{\varepsilon'}=\begin{bmatrix}1&0&0\\-1&1&0\\-1&0&1\\1&-1&1\end{bmatrix},\quad [T]_{\beta}^{\beta'}=\begin{bmatrix}0&0&0\\1&0&0\\0&2&0\\0&0&3\end{bmatrix}.$$

因 $[p_1(x)]_{\varepsilon}=\begin{bmatrix}1/2\\-1/2\\0\end{bmatrix}$, $[p_2(x)]_{\varepsilon}=\begin{bmatrix}-1\\0\\1\end{bmatrix}$, $[p_3(x)]_{\varepsilon}=\begin{bmatrix}1/2\\1/2\\0\end{bmatrix}$ 依定義 β 到 ε 的 change-of-basis matrix 為 $[\mathrm{id}_{P_2(\mathbb{R})}]_{\beta}^{\varepsilon}=\begin{bmatrix}1/2&-1&1/2\\-1/2&0&1/2\\0&1&0\end{bmatrix}$. 另外若 $x^3=c_1q_1(x)+c_2q_2(x)+c_3q_3(x)+c_4q_4(x)$, 則因 $q_1(-1)=1,q_2(-1)=q_3(-1)=q_4(-1)=0$, 將 $x=-1$ 代入前式得 $c_1=-1$, 同理我們可得 $c_2=0,c_3=1,c_4=8$, 亦即 $[x^3]_{\beta'}=\begin{bmatrix}-1\\0\\1\\8\end{bmatrix}$. 用同樣方法求 $x^2,x,1$ 對於 β' 的坐標表示法,

我們得 ε' 到 β' 的 change-of-basis matrix 為 $[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix}$. 我們也可以先寫下 β' 到 ε' 的 change-of-basis matrix $[\mathrm{id}_{P_3(\mathbb{R})}]_{\beta'}^{\varepsilon'} = \begin{bmatrix} -1/6 & 1/2 & -1/2 & 1/6 \\ 1/2 & -1 & 1/2 & 0 \\ -1/3 & -1/2 & 1 & -1/6 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ 再取 inverse 得 $[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'}$. 最後我們驗算

$$[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'}[T]_{\varepsilon}^{\varepsilon'}[\mathrm{id}_{P_2(\mathbb{R})}]_{\beta}^{\varepsilon} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{-1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} = [T]_{\beta}^{\beta'}.$$

前面提過, 我們經常談論的一種 linear transformation 是其定義域及對應域為相同的 vector space. 這樣的 linear transformation 我們特別稱之為 *linear operator*. 關於 linear operator 我們通常對於定義域及對應域會選同樣的一組 ordered basis. 此時利用 Theorem 7.1.1, 我們得以下之結果.

Corollary 7.1.3. *假設 $T: V \to V$ 為 linear transformation 且 β, β' 為 V 的兩組 ordered bases. 則存在 invertible matrix P 使得 $[T]_{\beta'}^{\beta'} = P^{-1}([T]_{\beta}^{\beta})P$, 其中 P 為將 β' 的坐標表示轉換成 β 的坐標表示的 change-of-basis matrix $[\mathrm{id}_V]_{\beta'}^{\beta}$.*

Proof. 考慮 Theorem 7.1.1 其中 $W = V$, $\gamma = \beta$ 以及 $\gamma' = \beta'$ 的情形. 此時 $Q = [\mathrm{id}_V]_{\beta}^{\beta'}$ 由式子 (7.1), 知 $Q = ([\mathrm{id}_V]_{\beta'}^{\beta})^{-1} = P^{-1}$, 得證本定理. □

給定一個 $n \times n$ matrix A 我們知道它可以代表某一個 dimension 為 n 的 vector space V 上的 linear operator $T: V \to V$, 對於 V 的某一組 ordered basis 的 matrix representation. 若 P 為 $n \times n$ invertible matrix, 則我們稱 $B = P^{-1}AP$ 和 A 為 *similar*. 意味著我們也可將 B 視為 $T: V \to V$ 的一個 matrix representation 只是選取 V 不同的 ordered basis 而已.

有時一個 linear operator, 若選取夠好的一組 ordered basis, 我們可以得到更好的 matrix representation 以至於更容易了解這個 linear transformation. 例如 Orthonormal basis 也可幫助我們處理 linear operator 的問題. 考慮 $T: V \to V$ 為 linear operator. 當我們給定 $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 為 V 的 ordered basis, 我們便可得到 T 對 β 的表現矩陣 $A = [T]_{\beta}$. 其中 A 的 j-th column, 就是 $T(\mathbf{v}_j)$ 用 β 寫下的坐標. 也就是說若 $T(\mathbf{v}_j) = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 則 A 的 j-th column 就是 $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. 特別的, 當 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 V 的一組 orthonormal basis, 我們很容易將 $T(\mathbf{v}_j)$ 寫成 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 的線性組合, 事實上 Proposition 4.3.6 告訴我們 $T(\mathbf{v}_j) = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 其中 $c_i = \langle T(\mathbf{v}_j), \mathbf{v}_i \rangle = \langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$. 也就是說 A 的 j-th column 其 i-th entry 為 $c_i = \langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$, 因此 $[T]_{\beta}$ 的 (i,j)-th entry 就是 $\langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$.

Proposition 7.1.4. *假設 V 為 inner product space 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 orthonormal basis. 若 $T:V\to V$ 為 linear operator 且考慮 V 的 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 則 T 用 β 所得的 matrix representation $[T]_\beta$ 其 (i,j)-th entry 為 $\langle\mathbf{v}_i,T(\mathbf{v}_j)\rangle$.*

Question 7.1. *在 Proposition 7.1.4 中若 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 是由 orthogonal basis 所形成, 則 $[T]_\beta$ 的 (i,j)-th entry 應為何?*

另外有的 linear operator 可以找到好的基底使其 matrix representation 為對角矩陣. 有關於這個課題, 等以後談到對角化時我們再進一步探討. 我們先看一個簡單的例子.

Example 7.1.5. 考慮 linear operator $T:\mathbb{R}^2\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x\\y\end{bmatrix})=\frac{1}{25}\begin{bmatrix}9x+12y\\12x+16y\end{bmatrix}$. 若利用標準基底 $\varepsilon=(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix})$ 我們得 $[T]_\varepsilon^\varepsilon=\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix}$. 然而若用 $\beta=(\begin{bmatrix}3\\4\end{bmatrix},\begin{bmatrix}-4\\3\end{bmatrix})$ 這組 ordered basis 可由 $T(\begin{bmatrix}3\\4\end{bmatrix})=\begin{bmatrix}3\\4\end{bmatrix}$, $T(\begin{bmatrix}-4\\3\end{bmatrix})=\begin{bmatrix}0\\0\end{bmatrix}$, 得 $[T]_\beta^\beta=\begin{bmatrix}1&0\\0&0\end{bmatrix}$. 我們很容易由

$$[T\circ T]_\beta^\beta=([T]_\beta^\beta)([T]_\beta^\beta)=\begin{bmatrix}1&0\\0&0\end{bmatrix}^2=\begin{bmatrix}1&0\\0&0\end{bmatrix}=[T]_\beta^\beta$$

推得 $T\circ T=T$. 事實上從 β 這組 ordered basis 我們很容易看出 T 就是將 $\mathbb{R}^2$ 上的向量對 $\begin{bmatrix}3\\4\end{bmatrix}$ 的投影. 另外令 $P=[\mathrm{id}]_\beta^\varepsilon=\begin{bmatrix}3&-4\\4&3\end{bmatrix}$, 我們得

$$\begin{bmatrix}1&0\\0&0\end{bmatrix}=[T]_\beta^\beta=([\mathrm{id}]_\varepsilon^\beta)([T]_\varepsilon^\varepsilon)([\mathrm{id}]_\beta^\varepsilon)=P^{-1}(\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix})P,$$

所以 $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ 和 $\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix}$ 為 similar.

Corollary 7.1.3 告訴我們，一個 linear operator 選取不同的 ordered basis, 其表現矩陣會有 similar 的關係. 反過來, 當給定 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 我們可以考慮 $L_A:\mathbb{F}^n\to\mathbb{F}^n$, 其定義為 $L_A(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$ 這一個 linear operator. 此時 L_A 對 $\mathbb{F}^n$ 的 standard ordered basis $\varepsilon=(\mathbf{e}_1,\ldots,\mathbf{e}_n)$ 的表現矩陣 $[L_A]_\varepsilon^\varepsilon$ 就是 A. 現若 $B\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 B 和 A similar, 亦即存在 invertible matrix U 滿足 $B=U^{-1}AU$. 現考慮 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 其中 $\mathbf{v}_i$ 是 U 的 i-th column, 則依定義 $U=[\mathrm{id}_{\mathbb{F}^n}]_\beta^\varepsilon$, 也因此由 Corollary 7.1.3 知 B 是 L_A 用 β 所得的表現矩陣, 即 $B=[L_A]_\beta^\beta$. 從這裡我們知道, 以後要探討兩個相似矩陣的問題, 我們都可以將之視為是同一個 linear operator 利用不同的 ordered basis 所得的表現矩陣.

7.2. Characteristic Polynomial

給定 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 以及 $\mathbf{v}\in\mathbb{F}^n$, 在數學上常會探討 $A^k\mathbf{v}$ 其中 k 為任意正整數的問題. 例如 Fibonacci sequence $F_1,F_2,\ldots.$ 是一組滿足 $F_{k+1}=F_k+F_{k-1}$ 的遞迴數列. 若我們令 $A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$ 且對任意 $k\geq 2$ 令 $\mathbf{v}_k=\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}$, 則

$$A\mathbf{v}_k=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}=\begin{bmatrix}F_k+F_{k-1}\\F_k\end{bmatrix}=\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}=\mathbf{v}_{k+1}.$$

因此我們有 $\mathbf{v}_3 = A\mathbf{v}_2$, $\mathbf{v}_4 = A\mathbf{v}_3 = A(A\mathbf{v}_2) = A^2\mathbf{v}_2, \ldots$, 這樣一直下去可得 $\mathbf{v}_{k+1} = A^{k-1}\mathbf{v}_2$. 也就是說對於任意 $k \geq 2$, 我們只要能算出 $A^{k-1}\mathbf{v}_2$, 就能求出 F_{k+1} 為何.

一搬來說當 k 越大時, 計算 $A^k\mathbf{v}$ 就越困難. 不過在一種特殊其況, 即當存在 $\lambda \in \mathbb{R}$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$ 時, 我們有 $A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v}) = \lambda^2\mathbf{v}$. 同理我們會有 $A^3\mathbf{v} = \lambda^3\mathbf{v}, \ldots, A^k\mathbf{v} = \lambda^k\mathbf{v}$. 也就是說在這種情況之下, 就很容易計算出 $A^k\mathbf{v}$. 因此我們對於怎樣的 $\mathbf{v}$ 會存在 $\lambda \in \mathbb{R}$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$ 特別有興趣. 所以有以下的定義.

Definition 7.2.1. 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. 若對於非零向量 $\mathbf{v} \in \mathbb{F}^n$, 存在 $\lambda \in \mathbb{F}$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 A 的一個 *eigenvector*, 而此 λ 稱為 A 的一個 *eigenvalue*.

注意, 依定義 A 的 eigenvector 一定是非零向量. 又若 $\mathbf{v}$ 是 A 的一個 eigenvector, 且 $\lambda, \lambda' \in \mathbb{R}$ 滿足 $A\mathbf{v} = \lambda\mathbf{v} = \lambda'\mathbf{v}$, 則由 $(\lambda - \lambda')\mathbf{v} = \mathbf{0}$ 以及 $\mathbf{v} \neq \mathbf{0}$, 可得 $\lambda = \lambda'$. 因此對於 A 的一個 eigenvector $\mathbf{v}$ 一定有也僅有一個實數 λ 會滿足 $A\mathbf{v} = \lambda\mathbf{v}$. 此時我們稱 eigenvector $\mathbf{v}$ 所對應的 eigenvalue 為 λ.

Question 7.2. 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, $\mathbf{v} \in \mathbb{F}^n$ 為非零向量滿足 $A\mathbf{v} = \mathbf{0}$. 是否 $\mathbf{v}$ 為 *eigenvector?* 其所對應的 *eigenvalue* 為何?

Example 7.2.2. 考慮

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

我們有

$$A\mathbf{v}_1 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2\mathbf{v}_1.$$

所以 $\mathbf{v}_1$ 是 A 的一個 eigenvector, 而 -2 是其對應的 eigenvalue. 同樣的

$$A\mathbf{v}_2 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} = 5 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 5\mathbf{v}_2.$$

所以 $\mathbf{v}_2$ 是 A 的一個 eigenvector, 而 5 是其對應的 eigenvalue. 然而

$$A\mathbf{v}_3 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \end{bmatrix} \notin \mathrm{Span}(\begin{bmatrix} 4 \\ 3 \end{bmatrix}).$$

所以 $\mathbf{v}_3$ 不是 A 的一個 eigenvector.

首先我們看一些 eigenvector 和 eigenvalue 的性質.

Proposition 7.2.3. 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, $\mathbf{v} \in \mathbb{F}^n$ 是 A 的 *eigenvector* 且其 *eigenvalue* 為 λ.

(1) 若 $c \in \mathbb{F}$ 且 $c \neq 0$, 則 $c\mathbf{v}$ 亦為 A 的一個 *eigenvalue* 為 λ 的 *eigenvector.*

(2) 若 $\mathbf{v}' \in \mathbb{F}^n$ 亦為 A 的一個 *eigenvalue* 為 λ 的 *eigenvector* 且 $\mathbf{v} + \mathbf{v}' \neq 0$, 則 $\mathbf{v} + \mathbf{v}'$ 亦為 A 的一個 *eigenvalue* 為 λ 的 *eigenvector.*

Proof. 依假設我們知道 $\mathbf{v} \neq \mathbf{0}$ 且 $A\mathbf{v} = \lambda\mathbf{v}$.

(1) 令 $\mathbf{w} = c\mathbf{v}$, 由於 $c \neq 0$ 且 $\mathbf{v} \neq \mathbf{0}$, 我們知 $\mathbf{w} \neq \mathbf{0}$. 現考慮

$$A\mathbf{w} = A(c\mathbf{v}) = c(A\mathbf{v}) = c(\lambda\mathbf{v}) = \lambda(c\mathbf{v}) = \lambda\mathbf{w}.$$

得證 $\mathbf{w} = c\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector.

(2) 令 $\mathbf{u} = \mathbf{v} + \mathbf{v}'$. 依假設 $A\mathbf{v}' = \lambda\mathbf{v}'$ 且 $\mathbf{u} \neq \mathbf{0}$. 現考慮

$$A\mathbf{u} = A(\mathbf{v} + \mathbf{v}') = A\mathbf{v} + A\mathbf{v}' = \lambda\mathbf{v} + \lambda\mathbf{v}' = \lambda(\mathbf{v} + \mathbf{v}') = \lambda\mathbf{u}.$$

得證 $\mathbf{u} = \mathbf{v} + \mathbf{v}'$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector.

□

Question 7.3. 假設 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{F}^n$ 皆為 A 的一個以 *eigenvalue* 為 λ 的 *eigenvector.* 證明若 $\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$ 且 $\mathbf{w} \neq \mathbf{0}$, 則 $\mathbf{w}$ 也是 A 的一個以 *eigenvalue* 為 λ 的 *eigenvector.*

要注意, Question 7.3 並不是說任意兩個 eigenvector 的線性組合仍為 eigenvector. 必須是它們所對應的 eigenvalue 是一樣的才會對. 例如在 Example 7.2.2 中雖然 $\mathbf{v}_1, \mathbf{v}_2$ 都是 A 的 eigenvector, 但 $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2$ 就不是 A 的 eigenvector.

在 Example 7.2.2 中 $A \in \mathsf{M}_{2\times 2}(\mathbb{R})$, 而 $\mathbf{v}_1, \mathbf{v}_2$ 不平行, 所以 $\mathbf{v}_1, \mathbf{v}_2$ 形成 $\mathbb{R}^2$ 的一組 basis. 另一方面 $\mathbf{v}_1, \mathbf{v}_2$ 都是 A 的 eigenvector. 這樣的矩陣 A 是很特別的, 我們對有這樣特點的 matrix 給了以下的定義.

Definition 7.2.4. 假設 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. 若存在 $\mathbb{F}^n$ 的一組 basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 A 的 eigenvectors, 則稱 A 為 *diagonalizable* (可對角化).

為甚麼稱為 diagonalizable 呢? 這是因為若 $\mathbf{v}_1, \ldots, \mathbf{v} \in \mathbb{F}^n$ 是 $\mathbb{F}^n$ 的一組 basis 且皆為 A 的 eigenvectors, 又假設它們所對應的 eigenvalues 分別為 $\lambda_1, \ldots, \lambda_n$. 亦即 $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \ldots, A\mathbf{v}_n = \lambda_n\mathbf{v}_n$. 此時由矩陣乘法的定義我們有

$$A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \\ | & | & & | \end{bmatrix}.$$

另一方面若考慮 (i,i)-th entry 為 λ_i 的 $n \times n$ diagonal matrix D(即對角線第 i 個位置為 λ_i 而對角線外其餘位置皆為 0), 則我們有

$$\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} D = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_3\mathbf{v}_n \\ | & | & & | \end{bmatrix}.$$

因此若令 $C = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}$, 則我們有 $AC = CD$. 現又因 C 的 column 之間為 linearly independent 且有 n 個 column, 我們得 C 的 rank 為 n, 因此由 C 為 $n \times n$ matrix 得知 C 為 invertible (參見 Theorem 2.5.2). 因此我們可將 $AC = CD$ 改寫成 $D = C^{-1}AC$. 反之,

若存在一個 $n\times n$ invertible matrix 使得 $C^{-1}AC$ 為 diagonal matrix D, 則因 C 為 $n\times n$ invertible matrix, 所以 C 的 n 個 column vectors 形成 $\mathbb{F}^n$ 的一組 basis. 又因為 $AC = CD$, 由上面矩陣乘法的性質知 C 的 i-th column 就會是 A 以 D 的 (i,i)-th entry 為 eigenvalue 的 eigenvector. 所以 C 的 column vectors 就是 $\mathbb{F}^n$ 的一組 basis 且為 A 的 eigenvectors, 也就是說 A 為 diagonalizable. 前面曾經提過, 形如 $U^{-1}AU$ (其中 U 為 $n\times n$ invertible matrix) 這樣的 matrix 就稱為和 A 為 similar 的 matrix. 因此由這裡的討論, 我們知道 A 為 diagonalizable 就等同於 A 和一個 diagonal matrix 是 similar. 這也就是 diagonalizable 這個名稱的原因.

Example 7.2.5. 考慮 Example 7.2.2 中

$$A = \begin{bmatrix}1 & 3\\4 & 2\end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix}1\\-1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}3\\4\end{bmatrix}.$$

由於 $\mathbf{v}_1, \mathbf{v}_2$ 為 A 的 eigenvectors 且可形成 $\mathbb{R}^2$ 的一組 basis, 我們知 A 為 diagonalizable. 事實上若令 $C = \begin{bmatrix}1 & 3\\-1 & 4\end{bmatrix}$, 則由

$$AC = \begin{bmatrix}1 & 3\\4 & 2\end{bmatrix}\begin{bmatrix}1 & 3\\-1 & 4\end{bmatrix} = \begin{bmatrix}-2 & 15\\2 & 20\end{bmatrix} = \begin{bmatrix}1 & 3\\-1 & 4\end{bmatrix}\begin{bmatrix}-2 & 0\\0 & 5\end{bmatrix} = C\begin{bmatrix}-2 & 0\\0 & 5\end{bmatrix}.$$

故得 $C^{-1}AC = \begin{bmatrix}-2 & 0\\0 & 5\end{bmatrix}$ 為 diagonal matrix.

要如何找到一個 $n\times n$ matrix 的 eigenvector 及其對應的 eigenvalue 呢? 其實一般的找法是先找到 eigenvalue, 然後再找出與其對應的 eigenvector. 首先觀察若 $\lambda \in \mathbb{F}$ 是 A 的 eigenvalue, 表示存在一個非零向量 $\mathbf{v} \in \mathbb{F}^n$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$. 由於 $I_n\mathbf{v} = \mathbf{v}$, 所以看成矩陣的運算 $\lambda\mathbf{v} = (\lambda I_n)\mathbf{v}$. 因此 $A\mathbf{v} = \lambda\mathbf{v}$ 就等同於 $(A-\lambda I_n)\mathbf{v} = \mathbf{0}$. 換言之, λ 是 A 的 eigenvalue 等同於由 $n\times n$ matrix $A-\lambda I_n$ 所對應的 linear system $(A-\lambda I_n)\mathbf{x} = \mathbf{0}$ 有 nontrivial solution $\mathbf{x} = \mathbf{v}$. 由 Theorem 2.5.9, 這又等同於 $A-\lambda I_n$ 不是 invertible, 再由 Theorem 5.2.6(1) 知這也等同於 $\det(A-\lambda I_n) = 0$. 總言之, 要找到 A 的 eigenvalue λ 就是要找到 λ 滿足 $\det(A-\lambda I_n) = 0$.

要怎樣找到 λ 滿足 $\det(A-\lambda I_n) = 0$ 呢? 假設 $A = [a_{ij}]$, 若我們將 t 視為變數, 考慮 $\det(A-tI_n)$. 由於

$$A - tI_n = \begin{bmatrix} a_{11}-t & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22}-t & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}-t\end{bmatrix}$$

利用數學歸納法, 我們可以證明 $\det(A-tI_n)$ 會是一個以 t 為變數的 n 次實係數多項式. 而若 $t = \lambda$ 為此多項式的一實數根, 則 λ 就會滿足 $\det(A-\lambda I_n) = 0$, 也就是說 λ 就會是 A 的一個 eigenvalue. 反之, 若 λ 就會是 A 的一個 eigenvalue, 就表示 $t = \lambda$ 會是多項式 $\det(A-tI_n)$ 的一個根. 由此可知多項式 $\det(A-tI_n)$ 可以讓我們完全掌握 A 的 eigenvalue, 我們因而給它一個特別的定義.

Definition 7.2.6. 假設 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$, 考慮以 t 為變數的多項式 $p_A(t) = \det(A-tI_n)$. 我們稱 $p_A(t)$ 為 A 的 *characteristic polynomial* (特徵多項式)..

從上面的討論我們知道 $\lambda \in \mathbb{F}$ 為 characteristic polynomial $p_A(t)$ 的一個根若且唯若 λ 為 A 的 eigenvalue. 這裡要注意要談論 eigenvalue 是必須強調在哪一個 field 的 eigenvalue. 例如當 $A \in \mathsf{M}_{n\times n}(\mathbb{R})$, 其 characteristic polynomial $p_A(t)$ 是一個實係數多項式, 不過 $p_A(t)$ 有可能有非實數的虛根. 此時這個虛根不會是 A 在 $\mathbb{R}^n$ 中的 eigenvector 所對應的 eigenvalue. 事實上如果 $\lambda \in \mathbb{C}\setminus\mathbb{R}$ 是 $p_A(t)$ 的一個虛根, 此時假設存在 $\mathbf{v} \in \mathbb{R}^n$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$. 由於 $A\mathbf{v} \in \mathbb{R}^n$, 但 $\lambda\mathbf{v} \notin \mathbb{R}^n$, 所以 $A\mathbf{v} = \lambda\mathbf{v}$ 不可能成立. 不過依前面的探討我們知道一定會有 $\mathbf{w} \in \mathbb{C}^n$ 滿足 $A\mathbf{w} = \lambda\mathbf{w}$. 在這個課程裡, 當我們探討矩陣 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 的 eigenvalue 時, 若沒有特別說明, 都僅討論在 $\mathbb{F}$ 的 eigenvalue. 例如當我們討論實矩陣時, 我們考慮 eigenvalue 僅考慮 characteristic polynomial 的實根.

Example 7.2.7. 考慮 $B = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$, 此時 A 的 characteristic polynomial 為

$$p_B(t) = \det(B - tI_3) = \det \begin{bmatrix} -1-t & 4 & 2 \\ 1 & 3-t & 1 \\ -1 & 2 & 2-t \end{bmatrix}.$$

對第一個 row 降階求行列式得

$$p_A(t) = (-1-t)\det \begin{bmatrix} 3-t & 1 \\ 2 & 2-t \end{bmatrix} - 4\det \begin{bmatrix} -1 & 1 \\ -1 & 2-t \end{bmatrix} + 2\det \begin{bmatrix} -1 & 3-t \\ -1 & 2 \end{bmatrix}.$$

化簡可得 $p_A(t) = -(t-1)^2(t-2)$. 也因此 $t = 1$ 和 $t = 2$ 為 A 的 characteristic polynomial 的二實根, 也因此得 A 有兩個 eigenvalues $1, 2$.

接下來我們說明當 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 時, 其 characteristic polynomial $\det(A - tI_n)$ 確實是 t 的多項式. 首先觀察當我們在利用降階求 determinant 時, 其實是一些乘積之和. 利用數學歸納法可得這些乘積是由每一個 column 中的某個元素相乘而得而且它們都不會在同一個 row. 例如當我們計算 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的 characteristic polynomial $\det \begin{bmatrix} a-t & b \\ c & d-t \end{bmatrix}$ 時不難發現會貢獻 t 的最高次項乘積的是 $(a-t)(d-t)$ 而另一個乘積 bc 就僅影響到常數項, 因此其最高次項 t^2 與次高次項 t 的係數就完全由 $(a-t)(d-t)$ 的 t^2 與 t 的係數即 $at^2 - (a+d)t$ 所決定. 現考慮 3×3 matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ 的 characteristic polynomial. 利用對第一個 row 降階的方式我們有

$$\det \begin{bmatrix} a-t & b & c \\ d & e-t & f \\ g & h & i-t \end{bmatrix} = (a-t)\det \begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix} - b\det \begin{bmatrix} d & f \\ g & i-t \end{bmatrix} + c\det \begin{bmatrix} d & e-t \\ g & h \end{bmatrix}.$$

從前面 2×2 的情形我們看出 $\det \begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix}$ 的 t^2 與次高次項 t 的係數就完全由 $(e-t)(i-t)$ 的 t^2 與 t 的係數所決定, 因此 $(a-t)(e-t)(i-t)$ 貢獻出 t^3 和 t^2 的係數. 而 $\det \begin{bmatrix} d & f \\ g & i-t \end{bmatrix}$ 和 $\det \begin{bmatrix} d & e-t \\ g & h \end{bmatrix}$ 最多僅有 t 的一次出現, 因此得 $\det(A - tI_3)$ 的 t^3 和 t^2 的係數完全由 $(a-t)(e-t)(i-t)$ 所決定. 也就是說 A 的 chacteristic polynomial $p_A(t)$ 為 3 次多項式且其最高次的兩項為 $(-1)^3t^3 + (-1)^2(a+e+i)t^2$. 這裡 a, e, i 為 A 的 diagonal entries, 它們之和 $a+e+i$ 我們稱為 A 的 trace, 用 $\mathrm{tr}(A)$ 來表示. 利用數學歸納法, 我們可得當 $A=[a_{ij}]$ 為 $n\times n$ matrix 時, A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 為 t 的 n 次實係數多項式, 且其最高次的兩項是由 $(a_{11}-t)(a_{22}-t)\cdots(a_{nn}-t)$ 所貢獻因此為 $(-1)^n t^n+(-1)^{n-1}(a_{11}+\cdots+a_{nn})t^{n-1}$. 由於 A 的 diagonal entries 之和 $a_{11}+\cdots+a_{nn}$ 我們定為 $\mathrm{tr}(A)$, 因此有以下之結論.

Proposition 7.2.8. 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$. 則 A 的 *characteristic polynomial* 為 t 的 n 次實係數多項式. 其 t^n 項係數為 $(-1)^n$, t^{n-1} 項係數為 $(-1)^{n-1}\mathrm{tr}(A)$ 而常數項係數為 $\det(A)$.

Proof. 令 $p_A(t)=\det(A-tI_n)$, 由前面的討論我們僅剩討論 $p_A(t)$ 的常數項. 由於 $p_A(t)$ 是多項式所以它的常數項是 $p_A(0)=\det(A-0I_n)=\det(A)$. □

Question 7.4. 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$. 試問 A 最多會有幾個相異的 *eigenvalues?*

Example 7.2.9. 考慮 Example 7.2.2 中 $A=\begin{bmatrix}1&3\\4&2\end{bmatrix}$ 的 characteristic polynomial $p_A(t)$. 由於 $\mathrm{tr}(A)=1+2=3$ 以及 $\det(A)=2-12=-10$, 利用 Proposition 7.2.8 可得

$$p_A(t)=(-1)^2t^2+(-1)3t+(-10)=t^2-3t-10.$$

事實上利用 characteristic polynomial 的定義直接計算可得

$$p_A(t)=\det\begin{bmatrix}1-t&3\\4&2-t\end{bmatrix}=(1-t)(2-t)-12=t^2-3t-10.$$

分解後可得 $-2,5$ 為 A 的 eigenvalues.

接下來我們介紹一個和 eigenvalue 有關的定義. 若 $\lambda\in\mathbb{F}$ 是 A 的 eigenvalue. 由於 $t=\lambda$ 會是 A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 的一個根. 由因式定理知 $t-\lambda$ 會整除 $p_A(t)$. 若 $(t-\lambda)^m$ 可整除 $p_A(t)$, 但 $(t-\lambda)^{m+1}$ 不能整除 $p_A(t)$, 則我們稱 eigenvalue λ 的 *algebraic multiplicity* (代數重根數) 為 m. 當然了當 $t=\lambda$ 是 $p_A(t)$ 的一個單根, 我們就說 λ 的 algebraic multiplicity 為 1. 例如 Example 7.2.7 中 B 有兩個 eigenvalue 1 和 2, 其中 eigenvalue 1 的 algebraic multiplicity 為 2, 而 eigenvalue 2 的 algebraic multiplicity 為 1. 而 Example 7.2.9 中 A 的兩個 eigenvalue $-2,5$ 其 algebraic multiplicity 皆為 1. 有關 algebraic multiplicity 的性質, 以後我們還會進一步討論.

Question 7.5. *Identity matrix* I_n 的 *eigenvalue* 有哪些? 其 *algebraic multiplicity* 為何?

最後我們介紹一些和 characteristic polynomial 有關的性質. 一般來說兩個 $n\times n$ matrices 的 characteristic polynomial 可能不相同. 不過在一種特殊情況之下, 它們的 characteristic polynomial 會一樣. 前面提過當 A,B 為 $n\times n$ matrices, 若存在 $n\times n$ 的 invertible matrix U, 使得 $B=U^{-1}AU$, 則我們稱 A,B 為 *similar* (關於這個定義的原因我們以後會再詳述). 此時我們可得 A 和 B 的 characteristic polynomial 是相同的.

Proposition 7.2.10. 假設 A,B 為 $n\times n$ *matrices* 且存在 $n\times n$ 的 *invertible matrix* U 滿足 $B=U^{-1}AU$. 則 A 和 B 有相同的 *characteristic polynomial.*

Proof. 依假設 B 的 characteristic polynomial 為 $\det(B-tI_n)=\det(U^{-1}AU-tI_n)$. 然而依矩陣乘法性質

$$U^{-1}(A-tI_n)U=U^{-1}AU-U^{-1}(tI_n)U=U^{-1}AU-tU^{-1}I_nU=U^{-1}AU-tI_n.$$

因此再由 determinant 的性質 (Theorem 5.2.6) 得

$$\det(B-tI_n)=\det(U^{-1}(A-tI_n)U)=\det(U^{-1})\det(A-tI_n)\det(U)=\det(A-tI_n).$$

得證 A 和 B 有相同的 characteristic polynomial. □

另一個會有相同的 characteristic polynomial 的情況就是 A 和 A^{t} 有相同的 characteristic polynomial.

Proposition 7.2.11. 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 則 A 和 A^{t} 有相同的 *characteristic polynomial*

Proof. 利用 trnaspose 的性質 $(A-tI_n)^{\mathrm{t}}=A^{\mathrm{t}}-tI_n^{\mathrm{t}}=A^{\mathrm{t}}-tI_n$ (Proposition 2.2.4), 故利用 Theorem 5.2.6 (3), 我們有

$$P_{A^{\mathrm{t}}}(t)=\det(A^{\mathrm{t}}-tI_n)=\det((A-tI_n)^{\mathrm{t}})=\det(A-tI_n)=P_A(t).$$

□

Question 7.6. 試說明 A 和 A^{t} 有相同的 *eigenvalues* 且對每個 *eigenvalue* 其在 A 和 A^{t} 的 *algebraic multiplicity* 也相同.

7.3. Eigenspace 和 Eigenvector

我們了解了如何找到一個 $n\times n$ matrix 的 eigenvalue 之後, 接下來便是要找出這些 eigenvalue 所對應的 eigenvectors.

假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\lambda\in\mathbb{F}$ 為 A 的一個 eigenvalue. 由於 $\det(A-\lambda I_n)=0$, 我們知聯立方程組 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 存在非零的 nontrivial solution. 現假設 $\mathbf{v}\in\mathbb{F}^n$ 為非零向量且 $\mathbf{x}=\mathbf{v}$ 為 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 的一組解. 此即表示 $\mathbf{v}$ 滿足 $(A-\lambda I_n)\mathbf{v}=\mathbf{0}$, 亦即 $A\mathbf{v}=\lambda\mathbf{v}$. 故此時 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector. 反之, 若 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector, 則 $\mathbf{x}=\mathbf{v}$ 必為 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 的一組 nontrivial solution. 因此我們只要掌握 $n\times n$ matrix $A-\lambda I_n$ 的 nullspace (即 $\{\mathbf{v}\in\mathbb{F}^n\mid(A-\lambda I_n)\mathbf{v}=\mathbf{0}\}$) 中的非零向量就會是 A 相對於 λ 的 eigenvector. 由於 nullspace 是 vector space, 因此我們有以下的定義.

Definition 7.3.1. 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\lambda\in\mathbb{F}$ 為 A 的一個 eigenvalue. 則 $A-\lambda I_n$ 的 nullspace 稱為 A 對於 eigenvalue λ 的 *eigenspace*. 我們用 $E_A(\lambda)$ 來表示.

要注意對於 λ 的 eigenspace 並不是由以 λ 為 eigenvalue 的 eigenvectors 所組成. 這是因為零向量 $\mathbf{0}$ 不是 eigenvector, 但 vector space 必須包含 $\mathbf{0}$. 所以對於 λ 的 eigenspace 應該是由所有以 λ 為 eigenvalue 的 eigenvectors 和 $\mathbf{0}$ 所組成. 那為什麼要讓它形成 vector space 呢? 因為 vector space 有其方便性, 例如有了 vector space 我們就可以利用 dimension 來知道它的大小. 因此我們定義 $E_A(\lambda)$ 的 dimension 為 eigenvalue λ 的 *geometric multiplicity*

(幾何重根數). 要注意 eigenvalue λ 的 algebraic multiplicity 無法讓我們知道 λ 所對應的 eigenvectors 的多寡, 而是 λ 的 geometric multiplicity 可以提供這一個訊息. 有關於 Eigenspace 以及

Example 7.3.2. 考慮 $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$. 由前面 Example 7.2.7, Example 7.2.9 我們已計算出 A 和 B 的 characteristic polynomial 分別為 $p_A(t) = (x+2)(x-5)$, $p_B(t) = -(t-1)^2(t-2)$. 接下來我們分別計算 A 和 B 的 eigenspace.

首先考慮 A 對於 eigenvalue -2 的 eigenspace, 亦即找出 $A - (-2I_2) = \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix}6$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. 可得 $E_A(1) = \mathrm{Span}(\begin{bmatrix} 1 \\ -1 \end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 -2 的 eigenvector 就是那些和 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 平行的 nonzero vector. 由於 $\dim(E_A(-2)) = 1$, 我們也得到 A 對於 eigenvalue -2 的 geometric multiplicity 為 1. 至於 A 對於 eigenvalue 5 的 eigenspace, 亦即找出 $A - 5I_2 = \begin{bmatrix} -4 & 3 \\ 4 & -3 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} -4 & 3 \\ 0 & 0 \end{bmatrix}$. 因此得 $E_A(5) = \mathrm{Span}(\begin{bmatrix} 3 \\ 4 \end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 5 的 eigenvector 就是那些和 $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ 平行的 nonzero vector, 我們也得到 A 對於 eigenvalue 5 的 geometric multiplicity 為 1. 在 Example 7.2.2 中我們舉出 A 的 eigenvector 的例子其實是這樣得到的.

接著考慮 B 對於 eigenvalue 1 的 eigenspace, 亦即找出 $B - I_3 = \begin{bmatrix} -2 & 4 & 2 \\ -1 & 2 & 1 \\ -1 & 2 & 1 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. 可得 $E_B(1) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix})$. 也就是說 B 對於 eigenvalue 為 1 的 eigenvector 就是那些由 $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 的 linear combination 所得的 nonzero vector. 例如 $\mathbf{v} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 就滿足

$$B\mathbf{v} = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \mathbf{v}.$$

由於 $\dim(E_B(1)) = 2$, 我們也得到 B 對於 eigenvalue 1 的 geometric multiplicity 為 2. 至於 B 對於 eigenvalue 2 的 eigenspace, 亦即找出 $B - 2I_3 = \begin{bmatrix} -3 & 4 & 2 \\ -1 & 1 & 1 \\ -1 & 2 & 0 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$. 因此得 $E_B(2) =$

Span($\begin{bmatrix}2\\1\\1\end{bmatrix}$). 也就是說 B 對於 eigenvalue 為 2 的 eigenvector 就是那些和 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 B 對於 eigenvalue 2 的 geometric multiplicity 為 1.

在 Proposition 7.2.3 中我們知道兩個有相同 eigenvalue 的 eigenvectors 其線性組合只要不是 $\mathbf{0}$, 就會是有同樣 eigenvalue 的 eigenvector. 所以一般在探討一個 $n\times n$ 矩陣的 eigenvector 時, 我們只要寫下其 eigenspace 的一組基底即可.

以前我們提過有關 matrix 的問題都可以轉換成 linear transformation 的問題, 反之亦然. 回顧一下當 V 是 over $\mathbb{F}$ 的 vector space, 則一個 linear transformation $T:V\to V$, 稱為一個 linear operator. 特別地當 $\dim_{\mathbb{F}}(V)=n$, 且 β 是 V 的一組 ordered basis, 則 T 利用這組 ordered basis 所得的 matrix representation $[T]_\beta$ 會是一個 $n\times n$ matrix. 注意這裡因為定義域和對應域都是 V 所以兩邊是選同樣的 ordered basis, 因此我們將原本 matrix representation 的表示法 $[T]_\beta^\beta$ 省略寫成 $[T]_\beta$. 方陣 $[T]_\beta$ 的 eigenvalue 和 eigenvector 會和 T 有甚麼關係呢? 假設 $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$, 而 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 是 $[T]_\beta$ 的一個 eigenvector 且其 eigenvalue 為 λ, 此時我們有

$$[T]_\beta\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\lambda\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\begin{bmatrix}\lambda c_1\\\vdots\\\lambda c_n\end{bmatrix}.$$

若令 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 回顧一下在 Proposition 6.3.14 中告訴我們這表示 $T(\mathbf{v})$ 用 β 這組 ordered basis 的坐標表示應該是 $\begin{bmatrix}\lambda c_1\\\vdots\\\lambda c_n\end{bmatrix}$. 也就是說

$$T(\mathbf{v})=\lambda c_1\mathbf{v}_1+\cdots+\lambda c_n\mathbf{v}_n=\lambda(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=\lambda\mathbf{v}.$$

反之, 若 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in V$ 滿足 $T(\mathbf{v})=\lambda\mathbf{v}$, 則由 Proposition 6.3.14 知 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 是 $[T]_\beta$ 的一個 eigenvector 且其 eigenvalue 為 λ. 也因此我們有以下的定義.

Definition 7.3.3. 假設 V 是一個 vector space over $\mathbb{F}$ 且 $T:V\to V$ 是一個 linear operator. 若對 $\mathbf{v}\in V$, 存在 $\lambda\in\mathbb{F}$ 滿足 $T(\mathbf{v})=\lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 T 的一個 *eigenvector*, 且 λ 為其 *eigenvalue*.

Example 7.3.4. 考慮 Linear operator $T:P_2(\mathbb{R})\to P_2(\mathbb{R})$ 定義為

$$T(f(x))=f(x)+(x+1)f'(x),\quad\forall f(x)\in P_2(\mathbb{R}).$$

此時令 $g(x)=x^2+2x+1$, 則

$$T(g(x))=(x^2+2x+1)+(x+1)(2x+2)=3(x^2+2x+1)=3g(x).$$

故 x^2+2x+1 是 T 的一個 eigenvector 且其 eigenvalue 為 3.

在 Proposition 7.2.3, 我們提到關於方陣的 eigenvalue 和 eigenvector 的性質. 事實上這性質對 linear operator 也是對的, 我們有以下的性質. 由於證明方法和矩陣的情形一致, 我們就不再證明了.

Proposition 7.3.5. 假設 V 是一個 *vector space over* $\mathbb{F}$ 且 $T: V \to V$ 是一個 *linear operator.* 又假設 $\mathbf{v}_1, \mathbf{v}_2$ 為 T 的 *eigenvectors* 且其 *eigenvalue* 皆為 $\lambda \in \mathbb{F}$. 若 $c_1, c_2 \in \mathbb{F}$ 且 $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \neq \mathbf{0}$, 則 $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ 也會是 T 的一個以 λ 為 *eigenvalue* 的 *eigenvector.*

回顧一下在 $\mathsf{M}_{n\times n}(\mathbb{F})$ 的情形, 我們有所謂 diagonalizable matrix, 也就是說這樣的矩陣可以在 $\mathbb{F}^n$ 找到一組由 eigenvectors 所組成的 basis. 同樣的對於 linear operator, 我們也有以下的定義.

Definition 7.3.6. 假設 V 是一個 vector space over $\mathbb{F}$ 且 $T: V \to V$ 是一個 linear operator. 若 V 中存在一組 basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 T 的 eigenvectors, 則稱 T 為 *diagonalizable* (可對角化).

為何這樣子的 linear operator 會稱為 diagonalizable 呢? 其原因比矩陣的情況更容易讓人理解. 事實上如果 $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 是 V 的一組 ordered basis 且 $\mathbf{v}_i$ 皆為 T 的 eigenvector. 假設 λ_i 就是 $\mathbf{v}_i$ 所對應的 eigenvalue, 亦即 $T(\mathbf{v}_i) = \lambda_i\mathbf{v}_i$, $\forall i \in \{1, \ldots, n\}$. 此時考慮 T 利用 β 所得的 matrix representation $[T]_\beta$. 回顧一下 $[T]_\beta$ 的 1-st column 是 $T(\mathbf{v}_1) = \lambda_1$ 用 β 寫下的坐標表示, 即 $\begin{bmatrix}\lambda_1\\0\\\vdots\\0\end{bmatrix} = \lambda_1\mathbf{e}_1$. 而對一般的 $i \in \{1, \ldots, n\}$, $[T]_\beta$ 的 i-th column 就是 $T(\mathbf{v}_i) = \lambda_i\mathbf{v}_1$ 用 β 寫下的坐標表示, 即 $\lambda_i\mathbf{e}_i$. 也因此 $[T]_\beta$ 就是 $\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \lambda_n\end{bmatrix}$ 這樣的 diagonal matrix.

要怎樣找一個 linear operator $T: V \to V$ 的 eigenvalue 和 eigenvector 呢? 從前面一開始的說明可以知道, 任取 V 的一組 ordered basis β, 只要考慮其 matrix representation $[T]_\beta$ 的 eigenvalue 和 eigenvector 就可以還原成 T 的 eigenvalue 和 eigenvector 了, 我們看以下的例子.

Example 7.3.7. 考慮 Example 7.3.4 中的 linear operator $T: P_2(\mathbb{R}) \to P_2(\mathbb{R})$, 以及 $P_2(\mathbb{R})$ 的 standard basis $\varepsilon = (1, x, x^2)$. 由於依定義 $T(1) = 1$, $T(x) = 2x+1$, $T(x^2) = 3x^2+2x$, 我們得 $[T]_\varepsilon = \begin{bmatrix}1 & 1 & 0\\0 & 2 & 2\\0 & 0 & 3\end{bmatrix}$. 因為 $[T]_\varepsilon$ 是上三角矩陣, 很容易求得其 characteristic polynomial 為 $(1-t)(2-t)(3-t)$. 得知 $[T]_\varepsilon$ 的 eigenvalue 為 $1, 2, 3$ (事實上這也是 T 的 eigenvalue). 接下來我們利用解 $[T]_\varepsilon$ 的 eigenspace 得 $[T]_\varepsilon$ 的 eigenvectors. 對於 eigenvalue 1 所得的 eigenspace 就是 $\begin{bmatrix}0 & 1 & 0\\0 & 1 & 2\\0 & 0 & 2\end{bmatrix}$ 的 null space, 即 $\mathrm{Span}(\begin{bmatrix}1\\0\\0\end{bmatrix})$. 然而 $\begin{bmatrix}1\\0\\0\end{bmatrix}$ 是 1 在 $P_2(\mathbb{R})$ 利用 ε 所

得的坐標表示. 故知 Span(1) 中的非 0 元素是 T 的 eigenvector 且其 eigenvalue 為 1. 事實上我們有 $T(1)=1$, 對於 $[T]_\varepsilon$ 的 eigenvalue 2 所得的 eigenspace 就是 $\begin{bmatrix}-1&1&0\\0&0&2\\0&0&1\end{bmatrix}$ 的 null space, 即 $\mathrm{Span}(\begin{bmatrix}1\\1\\0\end{bmatrix})$. 然而 $\begin{bmatrix}1\\1\\0\end{bmatrix}$ 是 $x+1$ 在 $P_2(\mathbb{R})$ 利用 ε 所得的坐標表示. 故知 $\mathrm{Span}(x+1)$ 中的非 0 元素是 T 的 eigenvector 且其 eigenvalue 為 2. 事實上我們有 $T(x+1)=2(x+1)$, 對於 $[T]_\varepsilon$ 的 eigenvalue 3 所得的 eigenspace 就是 $\begin{bmatrix}-2&1&0\\0&-1&2\\0&0&0\end{bmatrix}$ 的 null space, 即 $\mathrm{Span}(\begin{bmatrix}1\\2\\1\end{bmatrix})$. 然而 $\begin{bmatrix}1\\2\\1\end{bmatrix}$ 是 x^2+2x+1 在 $P_2(\mathbb{R})$ 利用 ε 所得的坐標表示. 故知 $\mathrm{Span}(x^2+2x+1)$ 中的非 0 元素是 T 的 eigenvector 且其 eigenvalue 為 3. 事實上在 Example 7.3.4 中我們算過 $T(x^2+2x+1)=3(x^2+2x+1)$,

因為 $\{1,x+1,x^2+2x+1\}$ 是 T 的 eigenvectors 且是 $P_2(\mathbb{R})$ 的一組 basis, 所以我們知 T 是 diagonalizable. 事實上若考慮 ordered basis $\beta=(1,x+1,x^2+2x+1)$, 則 $[T]_\beta=\begin{bmatrix}1&0&0\\0&2&0\\0&0&3\end{bmatrix}$.

最後我們要強調, 在求 linear operator $T:V\to V$ 的 eigenvalue 和 eigenvector 時, 不必擔心選取 V 的 ordered basis 為何. 這是因為 eigenvalue 和 eigenvector 的定義和 T 有關, 而和 V 的 ordered basis 無關. 所以即使選取 V 的 ordered basis 不同會造成不同的矩陣表示, 所得的 eigenvalue 和 eigenvector 都可得到同樣 T 的 eigenvalue 和 eigenvector. 事實上我們知道, 當 V 選取不同的 ordered basis β,β', 雖然 $[T]_\beta$ 和 $[T]_{\beta'}$ 會不同, 但它們會是 similar, 所以它們會有同樣的 characteristic polynomial (Proposition 7.2.10), 因此有同樣的 eigenvalues. 最後要提醒的是, 在選取 V 的 ordered basis 使用表現矩陣來求 eigenvalue 和 eigenvector 時, 定義域和對應域都要使用同樣的 ordered basis, 否則這樣的表現矩陣所求得的 eigenvalue 和 eigenvector 和 T 的 eigenvalue 和 eigenvector 的定義是不吻合的.

基於上面的探討, 我們可以定義一個 linear operator $T:V\to V$ 的 characteristic polynomial $P_T(t)$. 其定義的方法就是任取一個 V 的 ordered basis β, 若 $A=[T]_\beta^\beta$, 則定義 $P_T(t)=P_A(t)$. 注意這樣定義出來的 characteristic polynomial 和 β 的選取無關. 主要的原因是若取 V 的另一組 ordered basis, 其表現矩陣會和 A 是 similar. 所以利用 Proposition 7.2.10 知, similar matrix 的 characteristic polynomial 是一樣的, 所以 $P_T(t)$ 不會因選取的 ordered basis 不同而有所不同. 注意, 前面我們提過, T 的 eigenvalue 就是其表現矩陣 A 的 characteristic polynomial 的根, 所以依此定義我們也可以說 T 的 eigenvalue 就是 T 的 characteristic polynomial 的根. 這裡唯一要注意的是 T 的 eigenvector 並不是 A 的 eigenvector. 事實上 T 的 eigenvector 並需用 ordered basis β 寫成 $\mathbb{F}^n$ 上的坐標表示法後, 才會是 A 的 eigenvector.

7.4. Cayley-Hamilton Theorm

在這節中我們將介紹 Cayley-Hamilton Theorem. 首先我們先介紹 linear operator 的 invariant subspace, 再利用 invariant subspace 的概念證明 linear operator 的 Cayley-Hamilton Theorem, 再因此推得矩陣的 Cayley-Hamilton Theorem.

在探討函數的理論時, 通常當定義域很大時, 我們可以透過所謂的 restriction 將函數限制在較小的範圍來了解該函數. 給定一個函數 $f: X \to Y$, 以及 X 中的子集合 S, 所謂 f 的 restriction on S, 用 $f|_S$ 表示, 就是將 f 的定義域縮小到 S, 其他對於 f 的映射方式都沒有改變. 也就是說 $f|_S$ 是一個定義域為 S 的函數 $f|_S : S \to Y$, 且對於任意 $s \in S$, $f|_S(s) = f(s)$, 不過若 $x \in X$ 但 $x \notin S$, 則 $f|_S(x)$ 是無定義的. 現若 $T : V \to V$ 是 linear operator, W 為 V 的 subspace, 則 $T|_W$ 依然會是一個 linear transformation (只是定義域在 W 上). 不過 $T|_W$ 未必會是一個 linear operator, 因為 T 未必會將 W 中的元素映射到 W. 如此一來, 我們就不能將過去探討 linear operator 的理論運用在 $T|_W$ 上了. 為了達到 $T|_W$ 仍為 linear operator 的目的, 我們必須選有特殊性質的 W (即 T 會將 W 的元素映射到 W), 這樣就能套用 linear operator 的理論了. 因此我們有以下的定義.

Definition 7.4.1. 假設 V 是一個 vector space over $\mathbb{F}$ 且 $T : V \to V$ 是 linear operator. 若 W 是 V 的 subspace 且滿足 $T(W) \subseteq W$ (即 $T(\mathbf{w}) \in W, \forall \mathbf{w} \in W$), 則稱 W 為一個 T-invariant subspace.

要注意當 $T : V \to V$ 是 linear operator,, Definition 7.4.1, 告訴我們 W 是 T-invariant, 表示 $T(W) \subseteq W$, 並不是說 $T(W) = W$, 也不是說 $T(\mathbf{w}) = \mathbf{w}$, $\forall \mathbf{w} \in W$. 請大家不要誤解. 也就是說要檢查 V 的 subspace W 是否為 T-invariant subspace, 我們僅要檢查是否所有 W 的元素 $\mathbf{w}$ 經由 T 的映射 (即 $T(\mathbf{w})$) 依然在 W 中. 當然了, 因 T 為 linear operator, 對任意 $\mathbf{v} \in V$, 皆有 $T(\mathbf{v}) \in V$, 故 V 本身是 T-invariant. 還有因為 T 是 linear tansformation, 我們知道 $T(\mathbf{0}) = \mathbf{0}$, 所以 zero space $\{\mathbf{0}\}$ 也是 T-invariant. 另外若 $\lambda \in \mathbb{F}$ 是 T 的 eigenvector, 則 λ 所對應的 eigenspace $E_T(\lambda) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$ 也會是 T-invariant subspace. 這是因為若 $\mathbf{v} \in E_T(\lambda)$, 則 $T(\mathbf{v}) = \lambda\mathbf{v}$. 由於 $E_T(\lambda)$ 是 V 的 subspace 且 $\mathbf{v} \in E_T(\lambda)$, 自然有 $\lambda\mathbf{v} \in E_T(\lambda)$, 亦即 $T(\mathbf{v}) \in E_T(\lambda)$. 故 $E_T(\lambda)$ 亦為 T-invariant. 另外我們過去熟悉的 T 的 range $\mathsf{R}(T)$, 也是 T-invariant, 這是因為對任意 $\mathbf{v} \in V$, 自然有 $T(\mathbf{v}) \in T(V) = \mathsf{R}(T)$. 當然當 $\mathbf{v} \in \mathsf{R}(T)$, 由於 $T : V \to V$ 是 linear operator, 故 $\mathsf{R}(T) \subseteq V$, 因此我們依然有 $T(\mathbf{v}) \in \mathsf{R}(T)$. T 的 null space $\mathsf{N}(T)$ 也是 T-invariant. 這是因為對任意 $\mathbf{v} \in \mathsf{N}(T)$, 由於 $T(\mathbf{v}) = \mathbf{0}$ 且 $\mathbf{0} \in \mathsf{N}(T)$ (別忘了 $T(\mathbf{0}) = \mathbf{0}$), 故 $T(\mathbf{v}) \in \mathsf{N}(T)$.

除了前面舉的幾個例子外, 還有哪些 T-invariant subspace 呢? 前面提過考慮 T-invariant subspace 就是想將 T 的定義域縮小. 所以給定 $\mathbf{v} \in V$, 我們很想知道甚麼是包含 $\mathbf{v}$ 最小的 T-invariant subspace. 假設 W 是包含 $\mathbf{v}$ 的 T-invariant subspace. 當然了, 我們有 $\mathbf{v} \in W$. 不過由 W 是 T-invariant, 我們自然要有 $T(\mathbf{v}) \in W$ (因 $\mathbf{v} \in W$). 再由 $T(\mathbf{v}) \in W$ 以及 W 是 T-invariant, 我們有 $T(T(\mathbf{v})) = T^2(\mathbf{v}) \in W$. 如此一直下去我們知 $T^m(\mathbf{v}) \in W, \forall m \in \mathbb{N}$. 由此我們知, 若 W 是包含 $\mathbf{v}$ 的 T-invariant subspace, 則 W 必須包含 $\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}$

這個集合 (即 $\{T^i(\mathbf{v}) \mid i \in \mathbb{N}\}$) 中所有的元素. 不過 W 是 subspace, 所以也必須包含所有這些元素所 span 的 subspace, 所以我們有以下的定義.

Definition 7.4.2. 假設 $T: V \to V$ 是 linear operator. 對任意 $\mathbf{v} \in V$, 令

$$C(T,\mathbf{v}) = \mathrm{Span}\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}.$$

我們稱 $C(T,\mathbf{v})$ 為 the T-cyclic space generated by $\mathbf{v}$.

要注意若令 $S = \{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}$, 雖然 S 中可能有無窮多個元素, 不過依 span 的定義, $C(T,\mathbf{v}) = \mathrm{Span}(S)$ 中的元素是 S 中有限多個元素的線性組合. 因此若 $\mathbf{w} \in C(T,\mathbf{v}) = \mathrm{Span}(S)$, 表示存在 $c_0, c_1, \ldots, c_m \in \mathbb{F}$ 使得 $\mathbf{w} = c_0\mathbf{v} + c_1 T(\mathbf{v}) + \cdots + c_m T^m(\mathbf{v})$ (其中可能有些 $c_i = 0$). 故得 $T(\mathbf{w}) = c_0 T(\mathbf{v}) + c_1 T^2(\mathbf{v}) + \cdots + c_m T^{m+1}(\mathbf{v}) \in \mathrm{Span}(S) = C(T,\mathbf{v})$. 也因此得證 $c(T,\mathbf{v})$ 是 T-invariant subspace. 前面提過包含 $\mathbf{v}$ 的 T-invariant subspace 必包含 S, 故我們得到下面的定理.

Proposition 7.4.3. 假設 $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 則 $C(T,\mathbf{v})$ 是包含 $\mathbf{v}$ 最小的 *T-invariant subspace.*

Question 7.7. 假設 V 是 *vector space over* $\mathbb{F}$, $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 證明對任意 $\mathbf{w} \in C(T,\mathbf{v})$, 皆存在係數在 $\mathbb{F}$ 的多項式 $f(x)$ 使得 $\mathbf{w} = f(T)(\mathbf{v})$. 依此得 $C(T,\mathbf{v}) = \{f(T)(\mathbf{v}) \mid f(x) \in \mathbb{F}[x]\}$.

當 V 是 finite dimensional vector space over $\mathbb{F}$, $C(T,\mathbf{v})$ 為其 subspace, 故 $C(T,\mathbf{v})$ 也是 finite dimensional. 如何知道 $C(T,\mathbf{v})$ 的維度呢? 當然了, 若 $\mathbf{v} = \mathbf{0}$, 則 $T^i(\mathbf{v}) = T^i(\mathbf{0}) = \mathbf{0}$, $\forall i \in \mathbb{N}$, 故此時 $C(T,\mathbf{v}) = \{\mathbf{0}\}$, 即 $\dim(C(T,\mathbf{v})) = 0$. 因此我們僅考慮 $\mathbf{v} \neq \mathbf{0}$ 的情況. 首先我們考慮 $\mathbf{v}, T(\mathbf{v})$ 是否為 linearly independent. 若 $\mathbf{v}, T(\mathbf{v})$ 不是 independent, 由於 $\mathbf{v} \neq \mathbf{0}$, 故知存在 $c \in \mathbb{F}$ 使得 $T(\mathbf{v}) = c\mathbf{v}$. 由此知 $T^i(\mathbf{v}) = c^i\mathbf{v} \in \mathrm{Span}(\mathbf{v})$, $\forall i \in \mathbb{N}$. 因此得 $C(T,\mathbf{v}) = \mathrm{Span}(\mathbf{v})$, 即 $\dim(C(T,\mathbf{v})) = 1$. 而若 $\mathbf{v}, T(\mathbf{v})$ 為 independent, 則我們考慮 $\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v})$ 是否為 independent. 若它們不是 independent, 則由 $\mathbf{v}, T(\mathbf{v})$ 為 independent 知 $T^2(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$ (Lemma 3.5.4). 因此存在 $c, d \in \mathbb{F}$ 使得 $T^2(\mathbf{v}) = c\mathbf{v} + dT(\mathbf{v})$. 此時

$$T^3(\mathbf{v}) = T(T^2(\mathbf{v})) = cT(\mathbf{v}) + dT^2(\mathbf{v}) = cT(\mathbf{v}) + d(c\mathbf{v} + dT(\mathbf{v})) = dc\mathbf{v} + (c + d^2)T(\mathbf{v}).$$

因此得 $T^3(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$. 再利用數學歸納法, 我們可以證明 $T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$, $\forall i \in \mathbb{N}$, 因此知此時 $C(T,\mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$, 即 $\dim(C(T,\mathbf{v})) = 2$. 我們可以一直這樣探討下去得到以下的定理.

Proposition 7.4.4. 假設 V 是 *vector space over* $\mathbb{F}$, $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 則 $\dim(C(T,\mathbf{v})) = m$ 若且唯若 m 是最大的 i 使得 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{i-1}(\mathbf{v})$ 為 *linear independent:* 也就是說 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 *linear independent* 但 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 *linearly independent.*

事實上若 $\dim(C(T,\mathbf{v})) = m$, 則 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 $C(T,\mathbf{v})$ 的一組 *basis.*

Proof. 首先我們用數學歸納法證明若 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 則

$$T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})), \forall i \in \mathbb{N}.$$

由於已知 $i \leq m$ 成立, 我們直接歸納假設 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 成立. 現考慮 $T^{k+1}(\mathbf{v})$. 由於 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 存在 $c_0, c_1, \ldots, c_{m-1} \in \mathbb{F}$ 使得 $T^k(\mathbf{v}) = c_0\mathbf{v} + c_1 T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$. 故

$$T^{k+1}(\mathbf{v}) = T(T^k(\mathbf{v})) = c_0 T(\mathbf{v}) + c_1 T^2(\mathbf{v}) + \cdots + c_{m-2}T^{m-1}(\mathbf{v}) + c_{m-1}T^m(\mathbf{v}).$$

由於 $T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 和 $T^m(\mathbf{v})$ 皆屬於 $\mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 得證

$$T^{k+1}(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})).$$

也因此證明了 $T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})), \forall i \in \mathbb{N}$. 因而得知

$$C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})).$$

現假設 $\dim(C(T, \mathbf{v})) = m$, 我們先說明 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent. 若它們不是 independent, 表示存在 $k \leq m-1$ 使得 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v}))$, 依前面討論, 此表示 $C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v}))$, 亦即 $C(T, \mathbf{v})$ 是由 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v})$ 這 k 個向量所展成. 然而 $k \leq m-1$, 此與 $\dim(C(T, \mathbf{v})) = m$ 的假設相矛盾, 故知 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent. 既然 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 independent 又可展成 $C(T, \mathbf{v})$, 故知它們形成 $C(T, \mathbf{v})$ 的一組 basis. 然而 $T^m(\mathbf{v}) \in C(T, \mathbf{v})$, 故知 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 linearly independent.

反之, 假設 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent, 但 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 linearly independent. 由 Lemma 3.5.4, 我們知 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$. 因此再由前面討論得 $C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$. 亦即 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 不只是 independent 且可展成 $C(T, \mathbf{v})$. 故 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 形成 $C(T, \mathbf{v})$ 的一組 basis, 得證 $\dim(C(T, \mathbf{v})) = m$. □

Question 7.8. 證明 $\dim(C(T, \mathbf{v})) = 1$ 若且唯若 $\mathbf{v}$ 是 T 的 *eigenvector.*

既然 $C(T, \mathbf{v})$ 是 T-invariant, 我們知 $T|_{C(T,\mathbf{v})} : C(T, \mathbf{v}) \to C(T, \mathbf{v})$ 是 linear operator. 那麼 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 會是甚麼呢? 要求 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial, 我們要先找到 $T|_{C(T,\mathbf{v})}$ 的定義域 $C(T, \mathbf{v})$ 的一組 ordered basis, 在利用這組 ordered basis 得到 $T|_{C(T,\mathbf{v})}$ 的表現矩陣, 再求該矩陣的 characteristic polynomial. 根據 Proposition 7.4.4, 若 $\dim(C(T, \mathbf{v})) = m$, 我們很自然的選 $(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 這一組 $C(T, \mathbf{v})$ 的 ordered basis.

接著我們來看 $T|_{C(T,\mathbf{v})}$ 用 $\beta = (\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 這一組 ordered basis 其表現矩陣 $[T|_{C(T,\mathbf{v})}]_\beta$ 為何? 首先 $[T|_{C(T,\mathbf{v})}]_\beta$ 的 1-st column 是 β 的第一個向量 (即 $\mathbf{v}$) 經由 T 映射後所得的向量 (即 $T(\mathbf{v})$) 用 ordered basis β 所得的坐標表示. 由於 $T(\mathbf{v})$ 恰好是 β 的第二個向量, 故其坐標表示為 $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$. 同理得 $[T|_{C(T,\mathbf{v})}]_\beta$ 的 i-th column 為 $\mathbf{e}_{i+1}$, 其中

$1 \leq i \leq m-1$. 至於 $[T|_{C(T,\mathbf{v})}]_\beta$ 的最後一個 column, 應該是 β 的最後一個向量 (即 $T^{m-1}(\mathbf{v})$) 經由 T 映射後所得的向量 (即 $T(T^{m-1}(\mathbf{v})) = T^m(\mathbf{v})$) 用 ordered basis β 所得的坐標表示. 然而 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 若假設 $T^m(\mathbf{v}) = c_0\mathbf{v} + c_1 T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$, 其中 $c_0, c_1, \ldots, c_{m-1} \in \mathbb{F}$, 則 $[T|_{C(T,\mathbf{v})}]_\beta$ 的最後一個 column 就是 $\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{bmatrix}$. 因此得

$$[T|_{C(T,\mathbf{v})}]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1} \end{bmatrix}.$$

如何求這樣的矩陣的 characteristic polynomial 呢? 我們首先考慮 2×2 矩陣的情形. 若 $A_2 = \begin{bmatrix} 0 & c_0 \\ 1 & c_1 \end{bmatrix}$, 直接計算可得 A_2 的 characteristic polynomial 為

$$\det(A_2 - tI_2) = \det\begin{bmatrix} -t & c_0 \\ 1 & c_1 - t \end{bmatrix} = t^2 - c_1 t - c_0.$$

而若 $A_3 = \begin{bmatrix} 0 & 0 & c_0 \\ 1 & 0 & c_1 \\ 0 & 1 & c_2 \end{bmatrix}$, 考慮矩陣 $A_3 - tI_3 = \begin{bmatrix} -t & 0 & c_0 \\ 1 & -t & c_1 \\ 0 & 1 & c_2 - t \end{bmatrix}$ 的 determinant, 由於我們可以用數學歸納法處理, 所以不直接計算而是採用降階的方式處理. 對 $A_3 - tI_3$ 的 1-st row 降階求 determinant 得

$$\det(A_3 - tI_3) = (-t)\det\begin{bmatrix} -t & c_1 \\ 1 & c_2 - t \end{bmatrix} + c_0 \det\begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}.$$

其中第一個矩陣是前面 2×2 的情況可得其 determinant 為 $t^2 - c_2 t - c_1$, 而第二個矩陣式上三角矩陣故其 determinant 為 1. 因此可得

$$\det(A_3 - tI_3) = (-t)(t^2 - c_2 t - c_1) + c_1 = -(t^3 - c_2 t^2 - c_1 t - c_0).$$

利用數學歸納法我們可以得到以下之結果.

Proposition 7.4.5. 假設 V 是 *vector space over* $\mathbb{F}$, $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 若 $\dim(C(T,\mathbf{v})) = m$ 且 $T^m(\mathbf{v}) = c_0\mathbf{v} + c_1 T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$, 則 $T|_{C(T,\mathbf{v})}$ 的 *characteristic polynomial* 為

$$(-1)^m(t^m - c_{m-1}t^{m-1} - \cdots - c_1 t - c_0).$$

Proof. 我們繼續剛才的討論, 利用數學歸納法求 $m\times m$ 矩陣

$$[T|_{C(T,\mathbf{v})}]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1} \end{bmatrix}.$$

的 characteristic polynomial. 對 1-st row 展開得

$$\det\begin{bmatrix} -t & 0 & \cdots & 0 & c_0 \\ 1 & -t & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -t & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1}-t \end{bmatrix} =$$

$$(-t)\det\begin{bmatrix} -t & 0 & \cdots & 0 & c_1 \\ 1 & -t & \cdots & 0 & c_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -t & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1}-t \end{bmatrix} + (-1)^{m+1}c_0\det\begin{bmatrix} 1 & -t & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -t \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

注意等式右邊的矩陣都是降階後的 $(m-1)\times(m-1)$ matrix. 其中第一個是歸納假設成立的 $(m-1)\times(m-1)$ matrix, 所以其 determinant 為 $(-1)^{m-1}(t^{m-1}-c_{m-1}t^{m-2}-\cdots-c_2t-c_1)$. 而第二個矩陣式對角線為 1 的 upper triangular matrix 故其 determinant 為 1. 因此得證其 characteristic polynomial 為 $(-1)^m(t^m-c_{m-1}t^{m-1}-\cdots-c_2t^2-c_1t-c_0)$. □

利用 Proposition 7.4.5, 我們馬上可得以下結論.

Corollary 7.4.6. 假設 $T: V\to V$ 為 *linear operator* 且 $\mathbf{v}\in V$ 為非零向量. 令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}: C(T,\mathbf{v})\to C(T,\mathbf{v})$ 的 *characteristic polynomial.* 則 $g(T)(\mathbf{v})=\mathbf{0}$.

Proof. 假設 $\dim(C(T,\mathbf{v}))=m$ 且 $T^m(\mathbf{v})=c_0\mathbf{v}+c_1T(\mathbf{v})+\cdots+c_{m-1}T^{m-1}(\mathbf{v})$. Proposition 7.4.5 告訴我們 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 為 $g(x)=(-1)^m(x^m-c_{m-1}x^{m-1}-\cdots-c_1x-c_0)$. 因此

$$\begin{aligned} g(T)(\mathbf{v}) &= (-1)^m(T^m-c_{m-1}T^{m-1}-\cdots-c_1T-c_0\mathrm{id}_{C(T,\mathbf{v})})(\mathbf{v}) \\ &= (-1)^m(T^m(\mathbf{v})-c_{m-1}T^{m-1}(\mathbf{v})-\cdots-c_1T(\mathbf{v})-c_0\mathbf{v})=\mathbf{0}. \end{aligned}$$

□

Question 7.9. 假設 $T: V\to V$ 為 *linear operator* 且 $\mathbf{v}\in V$ 為非零向量. 令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}$ 的 *characteristic polynomial.* 證明 $g(T)|_{C(T,\mathbf{v})}=\mathbf{0}$.

Linear operator 限制在較小的 T-invariant subspace 基本上和原來的 operator 是相同的, 因此它們的 characteristic polynomial 之間應該有關係. 接下來, 我們便是要探討它們之間的關係. 假設 $T: V\to V$ 是 linear operator, 且 W 為 T-invariant subspace. 要討論 T 和 $T|_W$ 的 characteristic polynomial, 我們需找 W 和 V 的 ordered basis, 然後得到相對應的表現矩陣, 再得到它們的 characteristic polynomial. 因為 W 是 V 的 subspace, 我們又期待它們之間的 characteristic polynomial 相關. 自然的, 我們可以先找 W 的一組 ordered basis $\beta=(\mathbf{w}_1,\ldots,\mathbf{w}_k)$ 再將 β 擴大成 V 的一組 ordered basis $\gamma=(\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1},\ldots,\mathbf{w}_n)$. 現假設 $A=[T|_W]_\beta$, 注意 A 的 i-th column 就是 $T(\mathbf{w}_i)$ 用 β 所得的坐標表示 (即 $\mathbb{F}^k$ 中的向量). 現考慮 T 用 ordered basis γ 所得的矩陣表示 $[T]_\gamma$. 要注意當 $1\leq i\leq k$ 時, $[T]_\gamma$ 的 i-th column 和 A 的 i-th column 一樣是 $T(\mathbf{w}_i)$, 不同的是它應該是 $T(\mathbf{v}_i)$ 用 γ 的坐標表示 (即 $\mathbb{F}^n$

中的向量). 由於 $T(\mathbf{w}_i)\in W$, 其用 γ 寫下的線性組合, 僅需要用到前面 k 個向量, 即 β 的線性組合, 所以 $T(\mathbf{w}_i)$ 用 γ 的坐標表示基本上和用 β 坐標表示相同, 只是後面 $k+1,\ldots,m$ 這 $m-k$ 個 entry 須補上 0. 至於 $[T]_\gamma$ 在 k_1-th column 之後的 columns 我們就不知道會是怎樣, 不過這並不會影響我們要探討的問題. 總而言之, 若 $[T|_W]\beta = A$, 我們可以將 $[T]_\gamma$ 寫成 $\left[\begin{array}{c|c} A & M_1 \\ \hline \mathbf{0} & M_2 \end{array}\right]$ 這樣的形式. 注意這裡 A 是 $k\times k$ matrix, $\mathbf{0}$ 是 $(n-k)\times k$ matrix, 而 M_1,M_2 分別為 $k\times(n-k)$ 和 $(n-k)\times(n-k)$ matrix. 因此 $[T]_\gamma$ 的 characteristic polynomial 應為

$$\det\left[\begin{array}{c|c} A-tI_k & M_1 \\ \hline \mathbf{0} & M_2-tI_{n-k} \end{array}\right].$$

回顧前面提過在計算形如 $\left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & C \end{array}\right]$ 這樣的矩陣的 determinant 時, 當 A, C 分別為 $k\times k$ 和 $(n-k)\times(n-k)$ matrix 時我們可以先將前 k 個 row 用 elementary row operations 將 A 化成 echelon form, 再將後面 $n-k$ 個 row 用 elementary row operations 將 C 化為 echelon form. 因為這樣最後是一個 upper triangular matrix, 其 determinant 就是對角線相乘, 因此知 $\det\left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & C \end{array}\right] = (\det A)(\det C)$. 因此前面算 T 的 characteristic polynomial 便會是 $\det(A-tI_k)\det(M_2-tI_{n-k})$. 因為 $\det(A-tI_k)$ 就是 $T|_W$ 的 characteristic polynomial, 因此我們有以下的結果.

Proposition 7.4.7. 假設 V 為 *vector space over* $\mathbb{F}$, $T:V\to V$ 為 *linear operator* 且 W 為 T*-invariant subspace,* 則 $T|_W$ 的 *characteristic polynomial* 會是 T 的 *characteristic polynomial* 的因式. 也就是說, 若 $f(x)$ 是 T 的 *characteristic polynomial* 且 $g(t)$ 是 $T|_W$ 的 *characteristic polynomial,* 則存在係數在 $\mathbb{F}$ 的多項式 $h(x)$ 滿足 $f(x)=h(x)g(x)$.

特別地, 當 $\mathbf{v}\in V$ 我們可以考慮 Proposition 7.4.7 中 $W=C(T,\mathbf{v})$ 的情形, 也就是說若 $g(x)$ 是 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 而 $p_T(x)$ 是 T 的 characteristic polynomial, 則存在多項式 $h(x)$ 滿足 $p_T(x)=h(x)g(x)$. 利用這個多項式的等式, 我們會想要將 T 代入兩邊的多項式, 不過我們必須先釐清 $h(T)g(T)$ 是甚麼意思. 當然了, 依定義 $h(T)$, $g(T)$ 都是定義域為 V 的 linear operator. 我們要知道的是, 如果兩個多項式 $h(x)$, $g(x)$ 相乘會是 $f(x)$, 那麼 $f(T),h(T),g(T)$ 這三個 linear operator 會有甚麼關係. 首先在定義將 linear operator T 代入一個多項式時, 我們用到了將 T 代入 x^i 會得到 T^i. 這裡 x^i 是多項式的乘法, 而 T^i 是 linear operator 的合成. 簡單來說就是多項式相乘代入 linear operator 後會得到的是 linear operator 的合成. 我們剛剛是觀察單項式, 不過多項式只是單項式相加, 而多項式的加法和乘法有所謂的分配律 (即 $f_1(x)(f_2(x)+f_3(x)) = f_1(x)f_2(x)+f_1(x)f_3(x)$). linear operators 的合成也有所謂的分配律 (即若 T_1,T_2,T_3 皆為定義域為 V 的 linear operators. 則 $T_1\circ(T_2+T_3)=T_1\circ T_2+T_1\circ T_3$.) 也就是說多項式的乘法和 linear operator 的合成是符合同樣的運算規則的. 例如 $h(x)=x^2-2x$, $g(x)=x+1$ 且 $f(x)=h(x)g(x)=x^3-x^2-2x$. 我們檢查 $h(T)$ 和 $g(T)$ 的合成. 依定義

$$h(T)\circ g(T) = (T^2-2T)\circ(T+\mathrm{id}_V) = T^2\circ(T+\mathrm{id}_V)-2T\circ(T+\mathrm{id}_V).$$

上式整理後可得 $T^3+T^2-2T^2-2T = T^3-T^2-2T = f(T)$.

由前面的討論我們知若 $g(x)$ 是 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 而 $p_T(x)$ 是 T 的 characteristic polynomial, 則存在多項式 $h(x)$ 滿足 $p_T(x)=h(x)g(x)$. 將 T 代入 $p_T(x)=h(x)g(x)$ 這個等式可得 $p_T(T)=h(T)\circ g(T)$. 有了這個 linear operator 的關係式, 我們就可以證明 Cayley-Hamilton Theorem.

Theorem 7.4.8 (Cayley-Hamilton Theorem)**.** 假設 V 為 *vector space over* $\mathbb{F}$ 且 $T:V\to V$ 為 *linear operator.* 考慮 T 的 *characteristic polynomial* $p_T(x)$, 我們有 $p_T(T):V\to V$ 為 *zero operator,* 亦即對任意 $\mathbf{v}\in V$, $p_T(T)(\mathbf{v})=\mathbf{0}$.

Proof. 由於當 $\mathbf{v}=\mathbf{0}$, 時自然有 $p_T(T)(\mathbf{v})=\mathbf{0}$ (因 $p_T(T)$ 是 linear transformation), 故僅考慮 $\mathbf{v}\in V$ 且 $\mathbf{v}\neq O$. 此時考慮 $\mathbf{v}$ 所產生的 T-cyclic space $C(T,\mathbf{v})$, 並令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial. 由 Proposition 7.4.7 知存在 $h(x)$ 係數在 $\mathbb{F}$ 的 polynomial 使得 $p_T(x)=h(x)g(x)$. 因此知 $p_T(T)=h(T)\circ g(T)$. 此時 $p_T(T)(\mathbf{v})=(h(T)\circ g(T))(\mathbf{v})=h(T)(g(T)(\mathbf{v}))$. 然而由 Corollary 7.4.6 知 $g(T)(\mathbf{v})=\mathbf{0}$, 故得證 $p_T(T)(\mathbf{v})=h(T)(g(T)(\mathbf{v}))=h(T)(\mathbf{0})=\mathbf{0}$. □

假設 V 是 over $\mathbb{F}$ 的 vector space 且 $\dim(V)=n$, 我們知道定義在 V 上的 linear operators 的性質和 $M_{n\times n}(\mathbb{F})$ 上的矩陣的性質有相對照的關係. 所以 Cayley-Hamilton Theorem 對矩陣應該也是對的. 亦即若 $p_A(x)$ 是 A 的 characteristic polynomial, 則 $p_A(A)$ 會是零矩陣 $\mathbf{0}$. 或許大家會覺得這很好證明, 因為依定義 $p_A(x)=\det(A-xI_n)$, 所以 $p_A(A)=\det(A-AI_n)=\det(\mathbf{0})=0$. 這樣的證明是不對的. 依定義 $p_A(A)$ 是將 A 代入 $p_A(x)$ 這個多項式所得的 $n\times n$ 矩陣, 因此定理是說 $p_A(A)$ 是 $n\times n$ 的零矩陣. 而這個證明裡所得的 $\det(A-A)=0$ 是 $\mathbb{F}$ 的零元素. 兩個談的是不同的事 (僅在 $n=1$ 時相同). 這個證明的問題是當我們考慮 $\det(A-xI_n)$ 時, 是將 x 看待成未知的 $\mathbb{F}$ 中的元素, 因此 $\det(A-xI_n)$ 可寫成 $\det\begin{bmatrix} a_{11}-x & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}-x\end{bmatrix}$, 再將之展開得 x 的多項式. 因此在這一步驟當 x 代任何 $\mathbb{F}$ 中的元素進入上述的矩陣是行得通的, 但是 x 不能帶入矩陣 A. 因為 $\det\begin{bmatrix} a_{11}-A & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}-A\end{bmatrix}$ 是沒意義的. 也就是 $A-xI_n$ 我們視為是 A 這個矩陣減去對角線皆為 x 的對角矩陣 xI_n 所得的矩陣. 巧的是當 x 代 A 時 xI_n 是 $AI_n=A$ 有定義, 所以會讓人誤以為這樣可以證明 Cayley-Hamilton Theorem.

要證明矩陣形式的 Cayley-Hamilton Theorem, 可以利用 adjoint matrix 和 determinant 的關係處理, 不過由於我們已證明 linear operator 的情形, 所以這裡我們利用 linear operator 和 matrix 的關係處理, 順便讓大家熟悉 linear operator 和 matrix 之間的轉換關係. 首先給定 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$, 考慮和 T 有關的 linear transformation $T:\mathbb{F}^n\to\mathbb{F}^n$, 其定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 在此定義之下, 使用 $\mathbb{F}^n$ 的 standard ordered basis ε 我們會有 T 的表現矩陣 $[T]_\varepsilon=A$. 因此依定義 T 的 characteristic polynomial $p_T(x)$ 就是 A 的 characteristic polynomial $p_A(x)$. 利用 linear operator 的 Cayley-Hamilton Theorem (Theorem 7.4.8), 我們知 $p_T(T)=p_A(T)$ 是 zero operator, 所以 $p_A(T)$ 用 ε 所得的表現矩陣 $[p_A(T)]_\varepsilon$ 就是零矩陣. 然而 $[p_A(T)]_\varepsilon$ 和 $p_A(A)$ 有什麼關係呢? 回顧一下, 當 $T:V\to W$, $T':W\to U$ 為 linear transformation 時考慮 α, β, γ 分別為 V, W, U 的 ordered basis, 則我們有 $[T' \circ T]_\alpha^\gamma = [T']_\beta^\gamma [T]_\alpha^\beta$. 現在 $T : \mathbb{F}^n \to \mathbb{F}^n$ 為 linear operator, 故考慮 $\alpha = \beta = \gamma = \varepsilon$ 以及 $T' = T$, 的情形我們會有

$$[T^2]_\varepsilon = [T \circ T]_\varepsilon = [T]_\varepsilon [T]_\varepsilon = A^2.$$

同理我們會有 $[T^k]_\varepsilon = A^k$. 另外我們也知兩個 linear transformations 的線性組合, 用相同的 ordered basis 其表現矩陣, 就是它們個別的表現矩陣做相同的線性組合. 因此若 $p_A(x) = (-1)^n(x^n + \cdots + c_1 x + c_0)$, 則 $p_A(T) = (-1)^n(T^n + \cdots + c_1 T + c_0 \mathrm{id}_{\mathbb{F}^n})$. 因此 $p_A(T)$ 利用 standard ordered basis ε 所得的表現矩陣 $[p_A(T)]_\varepsilon$ 就會是

$$(-1)^n([T]_\varepsilon^n + \cdots + c_1 [T]_\varepsilon + c_0 I_n) = (-1)^n(A^n + \cdots + c_1 A + c_0 I_n) = p_A(A).$$

得證 $p_A(A) = \mathbf{0}$, 因此有以下的定理.

Theorem 7.4.9 (Cayley-Hamilton Theorem)**.** 假設 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 且 $p_A(x)$ 為 A 的 *characteristic polynomial,* 則 $p_A(A)$ 為零矩陣.

當 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 為 diagonalizable 時, 我們可以將 A 對角化求得 A 的高次方 A^k, 而不必真的將多個 A 相乘 (參見 Example 8.7.6). 當 A 不是 diagonalizable, 要計算 A^k 我們可以利用 Theorem 7.4.9 以及長除法將其次數降下來, 這樣也可很快算出 A^k, 我們看下面的 Example.

Example 7.4.10. 考慮 $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$. 由於 A 的 characteristic polynomial $p_A(x) = x^2 - 2x + 5$ 在 $\mathbb{R}$ 中無法分解, 故知 A 在 $\mathbb{R}$ 不是 diagonalizable. 不過由 Cayley-Hamilton Theorem, 我們知 $A^2 - 2A + 5I_2 = \mathbf{0}$. 我們可以利用此式計算 A 的高次方. 例如計算 A^5. 首先利用長除法, 將 x^5 除以 $p_A(x) = x^2 - 2x + 5$, 可得 $x^5 = (x^3 + 2x^2 - x - 12)(x^2 - 2x + 5) - 19x + 60$. 兩邊代入 A 得 $A^5 = (A^3 + 2A^2 - A - 12I_2)(A^2 - 2A + 5I_2) - 19A + 60I_2$. 由於 $A^2 - 2A + 5I_2 = \mathbf{0}$, 故得 $A^5 = -19A + 60I_2 = \begin{bmatrix} 41 & -38 \\ 38 & 41 \end{bmatrix}$.