Characteristic Polynomial of a Linear Map

In this note, let K be a field, let V be a K-vector space with basis $\beta = \{\beta_1, \dots, \beta_n\}$.

1 Introduction

Let $T\colon V \to V$ be a linear map. Recall that the matrix representation of T with respect to the basis β is given by

$$[T]_\beta^\beta = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

where the entries a_{ji} are uniquely determined by

$$T(\beta_i) = \sum_{j=1}^{n} a_{ji}\gamma_j \tag{1}$$

for all $1 \leq i \leq m$.

Note that this matrix representation of T *depends on a choice of a basis* β. Anytime you have a construction which depends on a particular choice of something, you should observe how your construction changes by making a different choice. With this philosophy in mind, let $\beta' = \{\beta_1', \dots, \beta_n'\}$ be another choice of basis of V. The matrix representation of T with respect to the basis β' is related to the matrix representation of T with respect to the basis β by

$$[T]_{\beta'}^{\beta'} = [1_V]_\beta^{\beta'}[T]_\beta^\beta[1_V]_{\beta'}^\beta.$$

In other words, setting $U = [1_V]_\beta^{\beta'}$[1] (so U is invertible and $U^{-1} = [1_V]_{\beta'}^\beta$), setting $M = [T]_\beta^\beta$, and setting $M' = [T]_{\beta'}^{\beta'}$, we see that

$$M' = UMU^{-1}. \tag{2}$$

In other words, M is conjugate to M'. Matrices which are conjugate to each other satisfy similar properties. For example, applying determinants to both sides of (2), we obtain

$$\begin{aligned} \det(M') &= \det(UMU^{-1}) \\ &= \det(U)\det(M)\det(U^{-1}) \\ &= \det(U)\det(U^{-1})\det(M) \\ &= \det(U)\det(U)^{-1}\det(M) \\ &= \det(M). \end{aligned}$$

Thus the determinant is invariant with respect conjugacy classes of matrices. In particular, we are justified in defining the **determinant of** T to be

$$\det(T) := \det[T]_\beta^\beta.$$

Again the reason why this definition makes sense is because it does not depend on a choice of basis.

The determinant of T is sometimes called an **invariant of** T. This is because it's construction does not depend on a choice of basis. It turns out that there is a more general invariant of T which includes the determinant of T, caled the **characteristic polynomial of** T. The goal of this note is to study this invariant.

[1] We call U the **change of basis matrix from the basis** β **to the basis** β'.

2 Characteristic Polynomial

Definition 2.1. Let $T\colon V \to V$ be a linear map. The **characteristic polynomial of** T is defined to be the polynomial

$$\chi_T(X) := \det(X\mathrm{I}_n - [T]_\beta^\beta).$$

Remark. The definition of characteristic polynomial of T involved a choice of basis, namely β. Thus, we had better check that this definition is independent of our choice of basis. Let $\beta' = \{\beta_1', \dots, \beta_n'\}$ be another choice of basis of V and let $U = [1_V]_\beta^{\beta'}$ be the change of basis matrix from β to β'. Setting $M = [T]_\beta^\beta$ and $M' = [T]_{\beta'}^{\beta'}$, we see that

$$\begin{aligned}\det(X\mathrm{I}_n - M') &= \det(U(X\mathrm{I}_n - M')U^{-1}) \\ &= \det(X\mathrm{I}_n - UM'U^{-1}) \\ &= \det(X\mathrm{I}_n - M).\end{aligned}$$

Thus the definition of $\chi_T(X)$ is independent of the choice of basis.

2.1 Eigenvalues

Definition 2.2. Let $T\colon V \to V$ be a linear map and let $\lambda \in K$. We say λ is an **eigenvalue of** T if there exists a nonzero vector $v \in V$ such that $Tv = \lambda v$. In this case we call v an **eigenvector of** T **corresponding to the eigenvalue** λ. We denote by E_λ to be the set of all eigenvectors of T corresponding to λ forms a subspace of V. Observe that $E_\lambda = \ker(T - \lambda)$, hence E_λ is in fact a subspace of V. We call this subspace the **eigenspace of** T **corresponding to** λ.

Remark. When context is clear, we often refer to λ, v, and E_λ as "an eigenvalue", "an eigenvector", and "an eigenspace" respectively.

Proposition 2.1. *Let $T\colon V \to V$ be a linear map and let λ be an eigenvalue of T. Then λ is also an eigenvalue of $[T]_\beta^\beta$.*

Proof. Choose an eigenvector v corresponding to the eigenvalue λ. Then

$$\begin{aligned}[T]_\beta^\beta [v]_\beta &= [Tv]_\beta \\ &= [\lambda v]_\beta \\ &= \lambda [v]_\beta.\end{aligned}$$

□

Proposition 2.2. *Let $T\colon V \to V$ be a linear map and let $\lambda \in K$. Then $\chi_T(\lambda) = 0$ if and only if λ is an eigenvalue of T.*

Proof. Setting $M = [T]_\beta^\beta$, we have

$$\begin{aligned}\chi_T(\lambda) = 0 &\iff \det(\lambda\mathrm{I}_n - M) = 0 \\ &\iff \ker(\lambda\mathrm{I}_n - M) \neq 0 \\ &\iff \lambda\mathrm{I}_n - M \text{ is not injective} \\ &\iff \text{there exists } \mathbf{v} \in K^n\backslash\{0\} \text{ such that } (\lambda\mathrm{I}_n - M)\mathbf{v} = 0 \\ &\iff \text{there exists } \mathbf{v} \in K^n\backslash\{0\} \text{ such that } M\mathbf{v} = \lambda\mathbf{v}. \\ &\iff \lambda \text{ is an eigenvalue of } M.\end{aligned}$$

□

Example 2.1. Consider the matrices $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. A quick calculation shows

$$\chi_A(X) = (X-1)^2 = \chi_B(X).$$

Thus the only root of $\chi_A(X) = \chi_B(X)$ is when $X = 1$. Proposition (2.2) implies 1 is an eigenvalue for both A and B (in fact it is the only one). On the other hand, note that $\ker(\mathrm{I}_2 - A) = 2$ and $\ker(\mathrm{I}_2 - B) = 1$.

2.2 Eigenspace

Definition 2.3. Let $T\colon V \to V$ be a linear map and let $\lambda \in K$. The **eigenspace of** λ is defined to be

$$E_\lambda := \ker(\lambda - T).$$

the dimension of E_λ is called the **geometric multiplicity of** λ and is denoted $\gamma_T(\lambda)$.

Remark. We often write $\lambda - T$ instead of $\lambda 1_V - T$ and we often write $\gamma(\lambda)$ instead of $\gamma_T(\lambda)$.

Proposition 2.3. *Let $T\colon V \to V$ be a linear map and let Λ denote the set of eigenvalues of T. Then the characteristic polynomial of T factors as*

$$\chi_T(X) = \prod_{\lambda \in \Lambda} (X - \lambda)^{\mu_T(\lambda)},$$

in a splitting field of K, where $\mu_T(\lambda) \in \mathbb{N}$ satisfy

$$\sum_{\lambda \in \Lambda} \mu_T(\lambda) = n.$$

We call $\mu_T(\lambda)$ the ***algebraic multiplicity of*** *λ.*

Remark. We often write $\mu(\lambda)$ instead of $\mu_T(\lambda)$.

3 Generalized Eigenvectors

3.1 $K[X]$-module

Let $T\colon V \to V$ be a linear map and let $K[X]$ be the polynomial ring in the indeterminate X with coefficients in K. We give V the structure of a $K[X]$-module by defining

$$p(X) \cdot v = p(T)(v) \tag{3}$$

for all $p(X) \in K[X]$ and for all $v \in V$. Let us check that the action (3) does indeed give V the structure of a $K[X]$-module. Obviously V is an abelian group since it is a K-vector space. Also we have $1 \cdot v = v$ for all $v \in V$. Let $p(X), q(X) \in K[X]$ and let $v, w \in V$. Write $p(X) = \sum_{i=0}^{l} c_i X^i$ and $q(X) = \sum_{j=0}^{m} d_j X^j$. Then

$$\begin{aligned}
(p(X) + q(X)) \cdot v &= (p(T) + q(T))(v) \\
&= \left(\sum_{i=0}^{l} c_i T^i + \sum_{j=0}^{m} d_j T^j \right)(v) \\
&= \sum_{i=0}^{l} c_i T^i(v) + \sum_{j=0}^{m} d_j T^j(v) \\
&= p(T)(v) + q(T)(v) \\
&= p(X) \cdot v + q(X) \cdot v
\end{aligned}$$

and

$$\begin{aligned}
p(X) \cdot (v + w) &= p(T)(v + w) \\
&= \sum_{i=0}^{l} c_i T^i(v + w) \\
&= \sum_{i=0}^{l} c_i (T^i(v) + T^i(w)) \\
&= \sum_{i=0}^{l} c_i T^i(v) + \sum_{i=0}^{l} c_i T^i(w) \\
&= p(T)(v) + p(T)(w) \\
&= p(X) \cdot v + p(X) \cdot w
\end{aligned}$$

and

$$
\begin{aligned}
p(X)\cdot(q(X)\cdot v) &= p(X)\cdot(q(T)(v)) \\
&= p(X)\cdot\sum_{j=0}^{m} d_j T^j(v) \\
&= \sum_{j=0}^{m} d_j (p(X)\cdot T^j(v)) \\
&= \sum_{j=0}^{m} d_j p(T)(T^j(v)) \\
&= \sum_{j=0}^{m} d_j \left(\sum_{i=0}^{l} c_i T^i(T^j(v))\right) \\
&= \sum_{j=0}^{m} d_j \sum_{i=0}^{l} c_i T^{i+j}(v) \\
&= \sum_{k=0}^{l+m} \left(\sum_{i=0}^{k} c_i d_{k-i}\right) T^k(v) \\
&= (p(X)q(X))\cdot v.
\end{aligned}
$$

Thus all of the required properties for V to be a $K[X]$-module under the action (3) are satisfied.

Proposition 3.1. *Let $p(X) \in K[X]$. Define*

$$\ker p(X) := \{v \in V \mid p(X)\cdot v = 0\}.$$

Then $\ker p(X)$ *is a linear subspace of V. In particular, if $p(X) = X - \lambda$ where λ is an eigenvalue of T, then*

$$ker(p(X)) = E_\lambda$$

where E_λ is the eigenspace corresponding to λ.

Proof. First note that $\ker(p(X))$ is nonzero since $0 \in \ker(p(X))$. Let $v, w \in \ker(p(X))$ and let $a, b \in K$. Write $p(X) = \sum_{i=0}^{l} c_i X^i$. Then

$$
\begin{aligned}
p(X)\cdot(av+bw) &= p(T)(av+bw) \\
&= \sum_{i=0}^{l} c_i T^i(av+bw) \\
&= \sum_{i=0}^{l} c_i (aT^i(v) + bf^i(w)) \\
&= a\sum_{i=0}^{l} c_i T^i(v) + b\sum_{i=0}^{l} c_i T^i(w) \\
&= a(p(X)\cdot v) + b(p(X)\cdot w) \\
&= 0+0 \\
&= 0.
\end{aligned}
$$

Thus $av + bw \in \ker(p(X))$. Therefore $\ker(p(X))$ is a linear subspace of V. In the case where $p(X) = X - \lambda$ for some eigenvalue λ of T, then we have

$$
\begin{aligned}
v \in \ker(p(X)) &\iff v \in \ker(X-\lambda) \\
&\iff (X-\lambda)\cdot v = 0 \\
&\iff (T-\lambda)(v) = 0 \\
&\iff T(v) = \lambda v.
\end{aligned}
$$

Thus $v \in \ker(p(X))$ if and only if v is an eigenvector of T with eigenvalue λ. Therefore $\ker(p(X)) = E_\lambda$. $\square$

Proposition 3.2. *Let $p(X)$ and $q(X)$ be polynomials in $K[X]$ so that* $\gcd(p(X), q(X)) = 1$. *Then we have*

$$ker(p(X)q(X)) = ker(p(X)) + ker(q(X)), \tag{4}$$

where the sum (4) is direct.

Proof. Write $p(X) = \sum_{i=0}^{l} c_i X^i$ and $q(X) = \sum_{j=0}^{m} d_j X^j$. We first show that $\ker(p(X)) + \ker(q(X)) \subseteq \ker(p(X)q(X))$. Let $v \in \ker(p(X)) + \ker(q(X))$. Write $v = v_1 + v_2$ where $v_1 \in \ker(p(X))$ and $v_2 \in \ker(q(X))$. Then

$$\begin{aligned}
(p(X)q(X)) \cdot v &= p(X) \cdot (q(X) \cdot v) \\
&= p(X) \cdot (q(X) \cdot (v_1 + v_2)) \\
&= p(X) \cdot (q(X) \cdot v_1 + q(X) \cdot v_2) \\
&= p(X) \cdot (q(X) \cdot v_1) \\
&= (p(X)q(X)) \cdot v_1 \\
&= (q(X)p(X)) \cdot v_1 \\
&= q(X) \cdot (p(X) \cdot v_1) \\
&= q(X) \cdot 0 \\
&= 0.
\end{aligned}$$

This implies $v \in \ker(p(X)q(X))$. Thus $\ker(p(X)) + \ker(q(X)) \subseteq \ker(p(X)q(X))$.

Now we show $\ker(p(X)q(X)) \subseteq \ker(p(X)) + \ker(q(X))$. Choose $a(X), b(X) \in K[X]$ so that

$$a(X)p(X) + b(X)q(X) = 1. \tag{5}$$

Such a choice is possible since $\gcd(p(X), q(X)) = 1$. Let $v \in \ker(p(X)q(X))$. Using (5), write $v = v_1 + v_2$ where

$$v_1 = (b(X)q(X)) \cdot v \quad \text{and} \quad v_2 = (a(X)p(X)) \cdot v.$$

Then $v_2 \in \ker(q(X))$ since

$$\begin{aligned}
q(X) \cdot v_2 &= q(X) \cdot ((a(X)p(X)) \cdot v) \\
&= (q(X)a(X)p(X)) \cdot v \\
&= (a(X)p(X)q(X)) \cdot v \\
&= a(X) \cdot (p(X)q(X) \cdot v) \\
&= a(X) \cdot 0 \\
&= 0.
\end{aligned}$$

Similarly, $v_1 \in \ker(p(X))$ since

$$\begin{aligned}
p(X) \cdot v_1 &= p(X) \cdot ((b(X)q(X)) \cdot v) \\
&= (p(X)b(X)q(X)) \cdot v \\
&= (b(X)p(X)q(X)) \cdot v \\
&= b(X) \cdot (p(X)q(X) \cdot v) \\
&= b(X) \cdot 0 \\
&= 0.
\end{aligned}$$

Therefore $v \in \ker(p(X)) + \ker(q(X))$, and this implies $\ker(p(X)q(X)) \subseteq \ker(p(X)) + \ker(q(X))$.

To see that (4) is a direct sum, let $v \in \ker(p(X)) \cap \ker(q(X))$. Then

$$\begin{aligned}
v &= 1 \cdot v \\
&= (a(X)p(X) + b(X)q(X)) \cdot v \\
&= (a(X)p(X)) \cdot v + (b(X)q(X)) \cdot v \\
&= a(X) \cdot (p(X) \cdot v) + b(X) \cdot (q(X) \cdot v) \\
&= a(X) \cdot 0 + b(X) \cdot 0 \\
&= 0 + 0 \\
&= 0.
\end{aligned}$$

Thus $\ker(p(X)) \cap \ker(q(X)) = 0$ and so the sum (4) is direct. □

Proposition 3.3. *Let $c(X) \in K[X]$ be any nonzero polynomial such that $c(T) = 0$. Suppose*

$$c(X) = p_1(X)p_2(X)\cdots p_m(X)$$

where each $p_i(X) \in K[X]$ and $\gcd(p_i(X), p_j(X)) = 1$ for all pairs $1 \leq i < j \leq m$. Then

$$V = ker(p_1(X)) + ker(p_2(X)) + \cdots + ker(p_m(X)), \tag{6}$$

where the sum (6) is direct.

Proof. We first prove by induction on $m \geq 2$ that for polynomials $p_i(X) \in K[X]$ such that $\gcd(p_i(X), p_j(X)) = 1$ for all $1 \leq i < j \leq m$, we have

$$\ker(p_1(X)p_2(X)\cdots p_m(X)) = \ker(p_1(X)) \oplus \ker(p_2(X)) \oplus \cdots \oplus \ker(p_m(X)), \tag{7}$$

where we use $\oplus$ to denote that the sum is direct. The base case $m = 2$ was established in Proposition (3.2). Now assume (7) is true for some $m \geq 2$. Let $p_i(X) \in K[X]$ such that $\gcd(p_i(X), p_j(X)) = 1$ for all $1 \leq i < j \leq m+1$. Since $\gcd(p_1(X), p_i(X)) = 1$ for all $2 \leq i \leq m+1$, we have $\gcd(p_1(X), p_2(X)\cdots p_{m+1}(X)) = 1$. Therefore

$$\begin{aligned}\ker(p_1(X)p_2(X)\cdots p_{m+1}(X)) &= \ker(p_1(X)) \oplus \ker(p_2(X)\cdots p_{m+1}(X)) \\ &= \ker(p_1(X)) \oplus \ker(p_2(X)) \oplus \cdots \oplus \ker(p_{m+1}(X)),\end{aligned}$$

where we used the base case on the first line and where we used the induction hypothesis to get from the first line to the second line.

To finish the problem, we just need to show that $V = \ker(c(X))$. Let $v \in V$. Then

$$\begin{aligned}c(X) \cdot v &= c(f)(v) \\ &= 0(v) \\ &= 0\end{aligned}$$

implies $v \in \ker(c(X))$. Therefore $V \subseteq \ker(c(X))$, which implies $V = \ker(c(X))$. □

Lemma 3.1. *Let $W_1, \ldots, W_t$ be subspaces of a vector space V. For each $1 \leq i \leq t$, let*

$$\mathcal{B}_i := \{u_{ij} \mid 1 \leq j \leq m_i\}$$

be a basis for W_i where $m_i := \dim W_i$. Assume that

$$W := W_1 + \cdots + W_t$$

is a direct sum. Then $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_t$ is a basis for W.

Proof. It suffices to show that $\mathcal{B}$ is a linearly independent set since $\operatorname{span}(\mathcal{B}) = W$ is clear. Suppose

$$\sum_{i=1}^{t}\sum_{j=1}^{m_i} a_{ij}u_{ij} = 0. \tag{8}$$

for some $a_{ij} \in K$ where $1 \leq i \leq t$ and $1 \leq j \leq m_i$. Then for each $1 \leq i \leq t$, we must have $\sum_{j=1}^{m_i} a_{ij}u_{ij} = 0$. Indeed, if $\sum_{j=1}^{m_k} a_{kj}u_{kj} \neq 0$ for some $1 \leq k \leq t$, then we can rearrange (8) to get

$$\sum_{j=1}^{m_k} a_{kj}u_{kj} = -\sum_{\substack{1 \leq i \leq t \\ i \neq k}}\sum_{j=1}^{m_i} a_{ij}u_{ij},$$

and so

$$\begin{aligned}0 \neq \sum_{j=1}^{m_k} a_{kj}u_{kj} \\ &\in W_k \cap \sum_{\substack{1 \leq i \leq t \\ i \neq k}} W_i \\ &= \{0\},\end{aligned}$$

gives us our desired contradiction. Thus, for each $1 \leq i \leq t$, we have

$$\sum_{j=1}^{m_i} a_{ij}u_{ij} = 0.$$

But this implies $a_{ij} = 0$ for all $1 \leq j \leq m_i$ since $\mathcal{B}_i$ is a basis for all $1 \leq i \leq t$. Thus $a_{ij} = 0$ for all $1 \leq i \leq t$ and $1 \leq j \leq m_i$, and hence $\mathcal{B}$ is linearly independent. □

4 Jordan Canonical Form

Theorem 4.1. *Assume K is algebraically closed. Let $T\colon V \to V$ be a linear map and let Λ denote the set of all eigenvalues of T. Then*

$$V = \bigoplus_{\substack{1\leq j\leq \mu(\lambda)\\ \lambda\in\Lambda}} E_{\lambda,j}^{r(j)}$$

4.1 Constructing a basis for $\ker \varphi^m$

Construction: Assume K is algebraically closed. Let $T\colon V \to V$ be a linear map. Suppose the characteristic polynomial of T factors as

$$\chi_T(X) = (X-\lambda)^n.$$

Denote $\varphi := T - \lambda$. We want to construct a basis for $\ker\varphi^n = V$. Before doing so, we first make the following observation. For each $1 \leq i \leq n$, we have the short exact sequence

$$0 \to \ker\varphi^{i-1} \hookrightarrow \ker\varphi^i \to \ker\varphi^i \,/\, \ker\varphi^{i-1} \to 0. \tag{9}$$

It follows from (9) that

$$\begin{aligned}\sum_{i=1}^n \dim(\ker\varphi^i \,/\, \ker\varphi^{i-1}) &= \sum_{i=1}^n \dim(\ker\varphi^i) - \dim(\ker\varphi^{i-1})\\ &= \dim(\ker\varphi^n) - \dim(\ker\varphi^0)\\ &= n.\end{aligned} \tag{10}$$

Now we proceed to construct a basis for $\ker\varphi^n$ as follows: Let

$$m_1 := \max\{i \mid \dim(\ker\varphi^i/\ker\varphi^{i-1}) > 0\}.$$

Note that $1 \leq m_1 \leq n$. Indeed, we have $1 \leq m_1$ since the dimension of the eigenspace E_λ is nonzero and we have $m_1 \leq n$ since the characteristic polynomial kills V. If $m_1 = 1$, then

$$\begin{aligned}\dim E_\lambda &= \dim(\ker\varphi)\\ &= \sum_{i=1}^n \dim(\ker\varphi^i/\ker\varphi^{i-1})\\ &= n,\end{aligned}$$

by the dimension formula (10) above. In this case, T is diagonalizable, and we can find a basis of V consisting of eigenvectors. Thus assume $1 < m_1 \leq n$. Let $\{\overline{v}_1^{m_1}, \dots, \overline{v}_{k_1}^{m_1}\}$[2] be a basis of $\ker\varphi^m \,/\, \ker\varphi^{m-1}$. It follows from linear independence of $\{\overline{v}_1^{m_1}, \dots, \overline{v}_{k_1}^{m_1}\}$ that if

$$a_1\overline{v}_1^{m_1} + \cdots + a_{k_1}\overline{v}_{k_1}^{m_1} = 0 \tag{11}$$

for some $a_1, \dots, a_{k_1} \in K$, then we must have $a_1 = \cdots = a_{k_1} = 0$. In other words, if

$$a_1 v_1^{m_1} + \cdots + a_{k_1} v_{k_1}^{m_1} \in \ker(\varphi^{m_1-1})$$

for some $a_1, \dots, a_{k_1} \in K$, then we must have $a_1 = \cdots = a_{k_1} = 0$. In other words, if

$$a_1\varphi^{m_1-1}(v_1^{m_1}) + \cdots + a_{k_1}\varphi^{m_1-1}(v_{k_1}^{m_1}) = 0$$

for some $a_1, \dots, a_{k_1} \in K$, then we must have $a_1 = \cdots = a_{k_1} = 0$. Thus, $\{\varphi^{m_1-1}(v_1^{m_1}), \dots, \varphi^{m_1-1}(v_{k_1}^{m_1})\}$ is a linearly independent set in $\ker(\varphi)$. In fact, $\{\varphi^{m_1-i}(v_1^{m_1}), \dots, \varphi^{m_1-i}(v_{k_1}^{m_1})\}$ is a linearly independent set in $\ker(\varphi^i)$ for all

[2]When we write $\overline{v}_j^m$, it is understood that $v_j^m \in \ker\varphi^m$ is a representative of the coset $\overline{v}_j^m \in \ker\varphi^m \,/\, \ker\varphi^{m-1}$. Note that if $\{\overline{v}_1^m, \dots, \overline{v}_k^m\}$ is a linearly independent set $\ker\varphi^m \,/\, \ker\varphi^{m-1}$, then $\{v_1^m, \dots, v_k^m\}$ is a linearly independent set in $\ker\varphi^m$ since it is in the preimage of a linear map.

$0 \le i < m_1$ since $\{\varphi^{m_1-i}(v_1^{m_1}), \dots, \varphi^{m_1-i}(v_{k_1}^{m_1})\}$ is in the preimage of $\{\varphi^{m_1-1}(v_1^{m_1}), \dots, \varphi^{m_1-1}(v_{k_1}^{m_1})\}$ under the map $\varphi^{i-1}\colon \ker(\varphi^i) \to \ker(\varphi)$. Moreover, $\{\varphi^{m_1-i}(\overline{v}_1^{m_1}), \dots, \varphi^{m_1-i}(\overline{v}_{k_1}^{m_1})\}$ is a linearly independent set in $\ker(\varphi^i)/\ker(\varphi^{i-1})$ all $1 \le i < m_1$. Indeed, if

$$a_1\varphi^{m_1-i}(v_1^{m_1}) + \cdots + a_{k_1}\varphi^{m_1-i}(v_{k_1}^{m_1}) \in \ker(\varphi^{i-1})$$

for some $a_1, \dots, a_{k_1}$, then

$$\begin{aligned}
a_1\varphi^{m_1-1}(v_1^{m_1}) + \cdots + a_{k_1}\varphi^{m_1-1}(v_{k_1}^{m_1}) &= a_1\varphi^{i-1}(\varphi^{m_1-i}(v_1^{m_1})) + \cdots + a_{k_1}\varphi^{i-1}(\varphi^{m_1-i}(v_{k_1}^{m_1})) \\
&= \varphi^{i-1}(a_1\varphi^{m_1-i}(v_1^{m_1}) + \cdots + a_{k_1}\varphi^{m_1-i}(v_{k_1}^{m_1})) \\
&= 0
\end{aligned}$$

which implies $a_1 = \cdots = a_{k_1} = 0$. Since $\{\varphi^{m_1-i}(\overline{v}_1^{m_1}), \dots, \varphi^{m_1-i}(\overline{v}_{k_1}^{m_1})\}$ is a linearly independent set in $\ker(\varphi^i)/\ker(\varphi^{i-1})$ we have the following inequality

$$\dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) \ge \dim(\ker(\varphi^{m_1})/\ker(\varphi^{m_1-1})). \tag{12}$$

for all $1 \le i \le m_1$.

If the inequality (12) is an equality for all $1 \le i < m_1$, then we must have $m_1 = n$ and

$$\dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) = 1$$

by dimension formula (10) and the inequality (12). In this case, $\{v_1^n, \varphi(v_1^n), \dots, \varphi^n(v_1^n)\}$ gives us a basis for V and we are done. Otherwise, let

$$m_2 := \max\{i \mid \dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) > \dim(\ker(\varphi^{m_1})/\ker(\varphi^{m_1-1}))\}.$$

Note that $1 \le m_2 < m_1$. Extend $\{\varphi^{m_1-m_2}(\overline{v}_1^{m_1}), \dots, \varphi^{m_1-m_2}(\overline{v}_{k_1}^{m_1})\}$ to a basis of $\ker(\varphi^{m_2})/\ker(\varphi^{m_2-1})$, say

$$\{\varphi^{m_1-m_2}(\overline{v}_1^{m_1}), \dots, \varphi^{m_1-m_2}(\overline{v}_{k_1}^{m_1}), \overline{v}_1^{m_2}, \dots, \overline{v}_{k_2}^{m_2}\}. \tag{13}$$

If $m_2 = 1$, then (13) gives us our desired basis. Otherwise, by the same arguments as above, the set

$$\{\varphi^{m_1-m_2-i}(\overline{v}_1^{m_1}), \dots, \varphi^{m_1-m_2-i}(\overline{v}_{k_1}^{m_1}), \varphi^{m_2-i}(\overline{v}_1^{m_2}), \dots, \varphi^{m_2-i}(\overline{v}_{k_2}^{m_2})\}$$

is a linearly independent set in $\ker(\varphi^i)/\ker(\varphi^{i-1})$ for all $1 \le i < m_2$. Hence we have the following inequality

$$\dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) \ge \dim(\ker(\varphi^{m_2})/\ker(\varphi^{m_2-1}))$$

for all $1 \le i \le m_2$.

At some point this process must terminate, say at m_t for some $t > 1$. Thus we obtain a decreasing sequence

$$n > m_1 > m_2 > \cdots > m_t \ge 1,$$

$$m_2 := \max\{i \mid \dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) > \dim(\ker(\varphi^{m_1})/\ker(\varphi^{m_1-1}))\}.$$

Note that $1 \le m_2 < m_1$.

First note that for each $1 \le i \le n$, we have the short exact sequence

$$0 \to \ker(\varphi^{i-1}) \hookrightarrow \ker(\varphi^i) \to \ker(\varphi^i)/\ker(\varphi^{i-1}) \to 0 \tag{14}$$

It follows from (14) that

$$\begin{aligned}
\sum_{i=1}^n \dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) &= \sum_{i=1}^n \dim(\ker(\varphi^i)) - \dim(\ker(\varphi^{i-1})) \\
&= \dim(\ker(\varphi^n)) - \dim(\ker(\varphi^0)) \\
&= n.
\end{aligned}$$

For each $0 \le i < m$, we will lift a basis of $\ker(\varphi^{i+1})/\ker(\varphi^i)$ to a linearly independent set in $\ker(\varphi^{i+1})$. Then we will show that the union of all of these linearly independent subsets forms a basis of $\ker(\varphi^m)$.

The final basis will be

$$\bigcup_{s=1}^t \{\varphi^{m_s-i}(v_j^{m_s}) \mid 1 \le i \le k_s \text{ and } 1 \le j \le m_s\}$$

Example 4.1. Let $A\colon K^{10} \to K^{10}$ be given by the matrix

$$A := \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In this case, we have $m_1 = 4$, $m_2 = 2$, $m_3 = 1$, and $k_1 = 1$, $k_2 = 2$, $k_3 = 2$. Note that

$$m_1k_1 + m_2k_2 + m_3k_3 = \mu(1),$$

where $\mu(1) = 10$ is the algebraic multiplicity of the eigenvalue 1. We also note that

$$k_1 + k_2 + k_3 = \gamma(1),$$

where $\gamma(1) = 5$ is the geometric multiplicity of the eigenvalue 1, i.e. the dimension of the eigenspace E_1. The generalized eigenvectors are given by

$$\begin{aligned} v_1^4 &= e_4 \\ \varphi(v_1^4) &= e_3 \\ \varphi^2(v_1^4) &= e_2 \\ \varphi^3(v_1^4) &= e_1 \\ v_1^2 &= e_6 \\ \varphi(v_1^2) &= e_5 \\ v_2^2 &= e_8 \\ \varphi(v_2^2) &= e_7 \\ v_1^1 &= e_9 \\ v_2^1 &= e_{10} \end{aligned}$$

Using our notation as above, we can line up the generalized eigenvectors like so:

$$\begin{array}{cccccc} \ker(\varphi^4)/\ker(\varphi^3): & v_1^4 & & & & \\ & | & & & & \\ \ker(\varphi^3)/\ker(\varphi^2): & \varphi(v_1^4) & & & & \\ & | & & & & \\ \ker(\varphi^2)/\ker(\varphi): & \varphi^2(v_1^4) & v_1^2 & v_2^2 & & \\ & | & | & | & & \\ \ker(\varphi): & \varphi^3(v_1^4) & \varphi(v_1^2) & \varphi(v_2^2) & v_1^1 & v_2^1 \end{array}$$

Now assume that $m_1 = n$. Then it follows from the dimension formula (10) and the inequality (12) that

$$\dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) = 1$$

for all $1 \le i \le n$. In this case, $\{v_1^n, \varphi(v_1^n), \dots, \varphi^n(v_1^n)\}$ gives us a basis for V and we are done. So assume $1 < m_1 < n$. Let

$$m_2 := \max\{i \mid \dim(\ker(\varphi^i)/\ker(\varphi^{i-1})) > \dim(\ker(\varphi^{m_1})/\ker(\varphi^{m_1-1}))\}.$$

Note that $1 \le m_2 < m_1$.

5 Invariant Subspaces

Proposition 5.1. *Let $\Psi \colon V_1 \to V_2$ be an isomorphism from the vector space V_1 to the vector space V_2 and let $T \colon V_1 \to V_1$ be a linear map. Then the T-invariant subspaces of V_1 are in one-to-one correspondence with the $(\Psi \circ T \circ \Psi^{-1})$-invariant subspaces of V_2.*

Proof. Let $\mathrm{Inv}_T(V_1)$ denote the set of T-invariant subspaces of V_1 and let $\mathrm{Inv}_{\Psi\circ T\circ\Psi^{-1}}(V_2)$ denote the set of $(\Psi \circ T \circ \Psi^{-1})$-invariant subspaces of V_2. The isomorphism $\Psi \colon V_1 \to V_2$ induces a bijection $\Psi \colon \mathrm{Inv}_T(V_1) \to \mathrm{Inv}_{\Psi\circ T\circ\Psi^{-1}}(V_2)$ given by $W_1 \mapsto \Psi(W_1)$. Observe that this map lands in the target space. Indeed, if $W_1 \in \mathrm{Inv}_T(V_1)$, then

$$\begin{aligned}
(\Psi \circ T \circ \Psi^{-1})(\Psi(W_1)) &= (\Psi \circ T)(\Psi \circ \Psi^{-1})(W_1) \\
&= (\Psi \circ T)(W_1) \\
&= \Psi(T(W_1)) \\
&\subset \Psi(W_1).
\end{aligned}$$

The inverse map is given by $\Psi^{-1} \colon \mathrm{Inv}_{\Psi\circ T\circ\Psi^{-1}}(V_2) \to \mathrm{Inv}_T(V_1)$. □

Proposition 5.2. *Let $V = V_1 \oplus \cdots \oplus V_n$ be a direct sum of vectors spaces $V_1, \dots, V_n$. Let $T \colon V \to V$ be given by $T = \oplus_i T_i$ where $T_i \colon V_i \to V_i$ are linear maps for each $1 \le i \le n$. Then the T-invariant subspaces of V consist of subspaces of the form*

$$W = W_1 \oplus \cdots \oplus W_n \tag{15}$$

where W_i is a T_i-invariant subspace for each $1 \le i \le n$.

Proof. Let $W = W_1 \oplus \cdots \oplus W_n$ be a subspace of V such that W_i is T_i-invariant for all $1 \le i \le n$. Let $w \in W$ and write $w = w_1 + \cdots + w_n$ where $w_i \in W_i$ for all $1 \le i \le n$. Then

$$\begin{aligned}
T(w) &= T(w_1 + \cdots + w_n) \\
&= T(w_1) + \cdots + T(w_n) \\
&= T_1(w_1) + \cdots + T_n(w_n) \\
&\in W_1 \oplus \cdots \oplus W_n \\
&= W.
\end{aligned}$$

Thus W is T-invariant. Conversely, let $W = W_1 \oplus \cdots \oplus W_n$ be any T-invariant subspace of V. Then for any $1 \le i \le n$ and for any $w \in W_i$, we have

$$\begin{aligned}
T_i(w) &= T(w) \\
&\subseteq W.
\end{aligned}$$

Since $\mathrm{im}(T_i) \subseteq V_i$, this implies $T_i(w) \in W \cap V_i = W_i$. Thus W_i is T_i-invariant for all $1 \le i \le n$. □