INTRODUCTION TO JORDAN CANONICAL FORM

CHIU HSIANG-TZU

1. Triangular Form

Definition 1.1. *let $T : V \to V$ be a linear operator. A subspace $W \subseteq V$ is said to be invariant (or stable) under T if $T(W) \subseteq W$*

Remark. $\{0\}$, V, $Ker(T)$, $Im(T)$, and E_λ are T-invarient.

Definition 1.2. *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. We say that V is triangularizable if and only if there exists an ordered basis β such that $[T]_\beta^\beta$ is upper triangular.*

Example. Consider $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^4$. Let β be the standard ordered basis of V and let $\beta = \{e_1, e_2, e_3, e_4\}$, where e_i is $(0, ..., 0, 1, 0, ..., 0)$ with the nonzero component at position i. We compute the matrix representation of $[T]_\beta^\beta$ as follows.

$$[T]_\beta^\beta = \begin{bmatrix} 1 & 1-i & 2 & 0 \\ 0 & 1 & i & 0 \\ 0 & 0 & 1-i & 3+i \\ 0 & 0 & 0 & 1-i \end{bmatrix}$$

Clearly, the matrix is upper triangular. Notice that $T(e_1) = e_1$, that is, e_1 is an eigenvector of T, $T(e_2) = (1-i)e_1 + e_2 \in span(\{e_1, e_2\})$, $T(e_3) \in span(\{e_1, e_2, e_3\})$, and $T(e_4) \in span(\{e_1, e_2, e_3, e_4\})$. Let W_i be the subspace of $\mathbb{C}^4$ spanned by the first i vectors in the standard ordered basis, that is, $T(e_i) \in W_i$ for all i. Clearly, $T(W_i) \subseteq W_i$, where $T(W) = \{T(w) \mid w \in W\} = Im(T|_W)$ for all subsapce of V.

Proposition 1.1. *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and $\beta = \{x_1, \cdots, x_n\}$ be a basis for V. Then $[T]_\beta^\beta$ is upper triangle if and only if the subspace $W_i = span(\{x_1, \cdots, x_i\})$ is T-invariant.*

Proof. It is trivial. □

Note that the subspace W_i in Proposition 1.1 are related as follows:

$$\{0\} \subseteq W_1 \subseteq \cdots \subseteq W_{n-1} \subseteq W_n = V.$$

We say that W_i forms an increasing sequence of subspaces. On the other hand, in Proposition 1.1, to show that a given linear operator is triangularizable. We must to construct the increasing sequence of T-invariant subspaces

$$\{0\} \subseteq W_1 \subseteq \cdots \subseteq W_{n-1} \subseteq W_n = V.$$

One also have to introduce the restriction of T to an T-invariant subspace $W \subseteq V$. Clearly, $T|_W = T_W : W \to W$ is a new linear mapping, where W is a T-invariant subspace of V.

Proposition 1.2. *Let $T : V \to V$ be a linear operator and W be a T-invariant subspace of a finite dimensional vector space V. Then the characteristic polynomial of $T|_W$ divides the characteristic polynomial of T.*

Proof. One can see Theorem 5.21 in Friedbreg. □

Corollary 1.1. *Every eigenvalue of $T|_W$ is also an eigenvalue of T,that is, the eigenvalue of $T|_W$ is some subset of the eigenvalue of T on V.*

Let $T : V \to V$ be a linear operator, where V is a finite dimensional vector space. Let $\lambda_1, \cdots, \lambda_n$ be distinct eigenvalues and m_i be the multiplicity of λ_i, as a root of the characteristic polynomial of T. Then T is diagonalizable if and only if $m_1 + \cdots + m_n = \dim(V)$ and $\dim E_{\lambda_i} = m_i$ for all i. The proof combines with Theorem 5.9 to 5.11 in Friedberg. Moreover, it means that

(1) V can be rewritten as the direct sum of eigenspaces
(2) m_i is algebraic multiplicity
(3) $\dim(E_{\lambda_i})$ is geometric multiplicity.

Theorem 1.1 (**Schur's Lemma**). *Let V be a finite dimensional vector space over $\mathbb{F}$ and $T : V \to V$ be an linear operator. Then T is triangularizable if and only if the characteristic polynomial of T has* $\dim(V)$ *roots* (*counted with multiplicities*) *in* $\mathbb{F}$

Remark. If $\mathbb{F} = \mathbb{C}$ (algebraic closure), then, by the fundamental theorem of algebra, every matrix $A \in M_{n\times n}(\mathbb{C})$ can be triangularizable. However, if $\mathbb{F} = \mathbb{R}$ ($x^2 + 1$ does not split on $\mathbb{R}$), then we can consider the rotation matix $R(\phi)$, where $0 < \phi < \pi$. Since the rotation matrix in $\mathbb{R}^2$, says

$$R = R(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \in M_{2\times 2}(\mathbb{R}),$$

where $0 < \phi < \pi$. There does not exist any vector $v \neq 0$ such that $Rv = \lambda v$ for some λ. Since

$$\begin{aligned}
\det(R - \lambda I_2) &= | \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} | \\
&= | \begin{bmatrix} \cos\phi - \lambda & -\sin\phi \\ \sin\phi & \cos\phi - \lambda \end{bmatrix} | \\
&= \cos^2\phi - 2\lambda\cos\phi + \lambda^2 + \sin^2\phi \\
&= \lambda^2 - 2\cos\phi\lambda + 1,
\end{aligned}$$

so we use the discriminant of a quadratic polynomial to get that $D = 4\cos^2\phi - 4 < 0$, that is, there are no solution in $\mathbb{R}$. Hence the characteristic polynomial of R does not split over $\mathbb{R}$.

Lemma. *Let $T : V \to V$ be as in theorem 1.1 and assume that the characteristic polynomial of T has $n = \dim(V)$ roots in $\mathbb{F}$. If $W \subsetneqq V$ is an invariant subspace under T, then there exists a non-zero vector x in V such that $x \notin W$ and $W + span(\{x\})$ is also T-invariant.*

Proof. let $\alpha = \{x_1, \cdots, x_k\}$ be a basis for W and extend α by adjointing $\alpha' = \{x_{k+1}, \cdots, x_n\}$ to form a basis $\beta = \alpha \cup \alpha'$ for V. Let $W' = span(\alpha')$. Clearly, $V = W \oplus W'$ because we know that the fact that

> Let β_1 and β_2 be disjoint bases for subspaces W_1 and W_2, respectively, of a vector space V. If $\beta_1 \cup \beta_2$ is a basis for V, then $V = W_1 \oplus W_2$.

We define a linear operator $P : V \to V$ by

$$P(a_1x_1 + \cdots + a_nx_n) = a_1x_1 + \cdots + a_kx_k.$$

Clearly, $W' = Ker(P)$, $W = Im(P)$, and $P^2 = P$. Hence P is the projection on W with kernal W'. Moreover $I - P$ is also the projection on W' with kernel W. Since

$$\begin{aligned}
&(I - P)(a_1x_1 + \cdots + a_nx_n) \\
&= I(a_1x_1 + \cdots + a_nx_n) - P(a_1x_1 + \cdots + a_nx_n) \\
&= a_1x_1 + \cdots + a_nx_n - a_1x_1 + \cdots + a_kx_k \\
&= a_{k+1}x_{k+1} + \cdots + a_nx_n,
\end{aligned}$$

$W = Ker(I - P)$, $W' = Im(I - P)$, and $(I - P)^2 = (I - P)(I - P) = I - P^2 - P + P = I - P^2 = I - P$. If the basis of V is orthonormal (the Grame-Schmidt process), then it is clear that $W' = W^\perp$ by Theorem 6.7 in Friedbreg. Since V is a finite dimensional vector space, so $(W^\perp)^\perp = W$ for all subspace W of V. It implies that

P is an orthogonal projection. Let $S = (I - P) \circ T \equiv (I - P)T$. Since $Im(I - P) = W'$, so $Im(S) \subseteq Im(I - P) = W'$, that is, W' is S-invariant subspace ($S(W') \subseteq W'$). Now, we claim that the set of eigenvalues of $S|_W$ is a subset of the root of eigenvalues of T. Since W is T-invariant (by assumption), so we compute the matrix representation of $[T]_\beta^\beta$ is the form

$$[T]_\beta^\beta = \begin{bmatrix} A & B \\ O & C \end{bmatrix}.$$

Clearly, $A = [T|_W]_\alpha^\alpha$ is $k \times k$ block matrix and $C = [S|_{W'}]_{\alpha'}^{\alpha'}$ is a $(n-k) \times (n-k)$ block matrix. Hence

$$\det(T - \lambda I) = \det(T|_W - \lambda I) \cdot \det(S|_{W'} - \lambda I).$$

From Corollary 1.1, we are done. Since all the eigenvalues of T lie in the field $\mathbb{F}$ ($\because$ the characteristic polynomial n roots), so by previous discussion, the same is true of all the the eigenvalues of $S|_{W'}$. Then there exists a nonzero vector x in W' ($x \notin W$) such that $Sx = \lambda x$, for some $\lambda \in \mathbb{F}$. It implies that

$$\begin{aligned} (I - P)(Tx) = \lambda x &\implies Tx - PTx = \lambda x \\ &\implies Tx = \lambda x + PTx \in span(\{x\}) + W. \end{aligned}$$

Finally, we show that $W + span(\{x\})$ is T-invariant. For all $y \in W + span(\{x\})$, there exists $z \in W$ and $\lambda \in \mathbb{F}$ such that $y = z + \lambda x$. Then

$$\begin{aligned} T(y) = T(z + \lambda x) &= T(z) + \lambda T(x) \\ &= T(z) + \lambda(x) + \lambda PT(x) \in span(\{x\}) + W. \end{aligned}$$

□

Now we go back to prove Theorem 1.1 (**Schur's Lemma**).

Proof. Suppose that T is triangularizable. Then there exists an ordered basis β for V such that $[T]_\beta$ is an upper triangular matrix. Hence the eigenvalues of T are the diagonal entries of this matrix. They are elements of $\mathbb{F}$. It means that the characteristic polynomial splits over $\mathbb{F}$. Conversely, suppose that the condition holds. Let λ be eigenvalues of T, x_i be an eigenvector of T correspond with λ and $W_1 = span(\{x_1\})$. Clearly, W_1 is T-invariant. From lemma, there exists a non-zero vector $x \notin W_1$ such that $W_1 + span(\{x_1\})$ is also T-invariant. Continuing the process,we get an increasing sequence of T-invariant subspace of V, that is, $W_1 \subseteq \cdots \subseteq W_k$ with $W_i = span(\{x_1, \cdots, x_i\})$ for all $i \geq 1$. Again, by lemma, there exists a nonzero vector W_k such that $W_k \ni W_{k+1} = W_k + span(\{x_{k+1}\})$ is also T-invariant. We continue the process until we have produced a basis for V. Hence, by Proposition 1.1, T is triangularizable. $\square$

Corollary 1.2. *If $T : V \to V$ is triangularizable with eigenvalues λ_i with respective multiplicities m_i, then there eixsts an order basis β for V such that $[T]_\beta$ is upper triangular matrix, and the diagonal entries of $[T]_\beta$ are m_1 λ_1's followed by m_2 λ_2's and so on.*

Under the assumption that all the eigenvalues of $T : V \to V$ lie in the foeld $\mathbb{F}$ over which V is defined. We recall that if T is a linear operator (or matrix) and $p(t) = a_n t^n + a_{n-1} t^{n-1} T^{n-1} + \cdots + a_0$ is a polynomial, then we can define a new linear mapping

$$p(T) = a_n T^n + a_{n-1} T^{n-1} + \cdots + a_0 I.$$

Theorem 1.2 (**Cayley-Hamilton theorem**)**.** *Let $T : V \to V$ be a linear operator on a finite dimensional vector space V and let $p(t) = \det(T - tI)$ be its characteristic polynomial. Assume that $p(t)$ has $\dim(V)$ roots in the field $\mathbb{F}$ over which V is defined. Then $p(T) = 0$, which is a zero transformation on V.*

Remark. It follows form Exercise 6-4-16 in Friedberg.

Proof. It suffices to show that $p(T)(x) = 0$ for all the vectors in some basis of V. By Theorem 1.1 (**Schur's Lemma**), there exists an ordered basis $\beta = \{x_1, \cdots, x_n\}$ for V such that $W_i = span(\{x_1, \cdots, x_i\})$ is T-invariant for all $1 \le i \le n$. Since all the eigenvalues of T lie in $\mathbb{F}$, so

$$p(t) = \pm 1 \cdot (t - \lambda) \cdots (t - \lambda_n)$$

for some $\lambda_i \in \mathbb{F}$ (not necessary distinct). If the factors here are ordered in the same fashion as the diagonal entries of $[T]_\beta^\beta$, then

$$T(x_i) = \lambda_i x_i + y_{i-1},$$

where $y_i \in W_{i-1}$ and $i \ge 2$, in particular, $T(x_1) = \lambda_1 x_1$. Now, we use the induction on i. For $i = 1$, we get that

$$\begin{aligned} p(T)(x_1) &= \pm(T - \lambda_1 I) \cdots (T - \lambda_n I)(x_1) \\ &= \pm(T - \lambda_2 I) \cdots (I - \lambda_n I)(T - \lambda_1)(x_1) = 0. \end{aligned}$$

The last equality follows that the powers of T commutes with each other and with I. Suppose that $p(T)(x_i) = 0$ for all $i \le k$. We compute $p(T)(x_{k+1})$. It is clear that only the factors $(T - \lambda_1 I), \cdots, (T - \lambda_k I)$ are needed to send x_i to 0 for $i \le k$. As before, we can rearrange the factors in $p(T)$ to obtain

$$\begin{aligned} p(T)(x_{k+1}) &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(x_{k+1}) \\ &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(T-\lambda_{k+1}I)(x_{k+1}) \\ &= \pm(T-\lambda_1 I)\cdots(T-\lambda_n I)(y_k) = 0. \end{aligned}$$

The last equality follows that $T(x_{k+1}) = \lambda_{k+1}x_{k+1} + y_k$ and $y_k \in W_k$. By induction, the other factors in $p(T)$ send all the vectors in this subspace to 0. We are done. $\square$

Remark. There are another proof of Theorem 1.2 (**Cayley-Hamilton theorem**). One can see Theorem 5.23 in Friedbreg.

Remark. Suppose that $A \in M_{n\times n}(\mathbb{F})$. If A is invertible, then $\det(A) \neq 0$. We consider that the characteristic polynomial of A

$$p(A) = \det(A - tI) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0.$$

Clearly, $a_0 = \det(A)$. By Theorem 1.2 (**Cayley-Hamilton theorem**), we get that

$$p(A) = (-1)^n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0.$$

Moreover, we also get that

$$A^{-1} = \frac{-1}{\det(A)}((-1)^n A^{n-1} + \cdots + a_1 I).$$

One can see Excercise 5-1-20, 5-1-21 and 5-4-18 in Friedbreg.

2. A Canonical Form for Nilpotent mappings

We look at the linear mapping $N : V \to V$, which have only one distinct eigenvalue $\lambda = 0$ with multiplicity $n = dim(V)$. If N is such mapping, then by Theorem 1.2 (**Cayley-Hamilton theorem**), $N^n = 0$. We call that N is nilpotent. Conversely, if N is nilpotent with $N^k = 0$ for some $k \geq 1$, then every eigenvalue of N is equal to 0. Hence, a mapping has one eigenvalue $\lambda = 0$ with multiplicity n if and only if N is nilpotent.

If $N^k = 0$ for some $k \geq 1$. Let λ be an eigenvalue of N with respect to eigenvector v, that is, $Nv = \lambda v$. Then

$$N^k v = \cdots = \lambda^k v.$$

Since eigenvector cannot be zero, so $\lambda = 0$. Clearly, the algebraic multiplicity is n.

Example. If S and T are nilpotent linear transformations which commute, then ST and $S+T$ are nilpotent linear transformations.

Proof. Let S and T be nilpotent linear transformations which commute, that is, there exist $m, n > 0$ such that $S^m = 0$ and $T^n = 0$ and $ST = TS$. Now, we have two cases as follows: $(ST)^m = S^m T^m = 0 \cdot T^m = 0$ if $m \geq n$ and $(ST)^n = S^n T^n = S^n \cdot 0 = 0$ if $m < n$. Hence ST is a nilpotent. Moreover

$$(S+T)^{m+n} = \sum_{k=0}^{m+n} C_k^{m+n} S^k T^{m+n-k}$$

If $k \geq m$, then $S^k = S^m \cdot S^{k-m} = 0$. If $k \leq m$, then $m+n-k \geq n$, so $T^{m+n-k} = T^n \cdot T^{m-k} = 0$. Hence $(S+T)^{m+n} = 0$ in each case. we get that $S+T$ is also nilpotent. $\square$

Let $\dim(V) = n$ and $N : V \to V$ be a nilpotent mapping. By above discussion, we know that $N^n = 0$, that is, $N^n(x) = 0$ for all x in V. Now, for each x in V, either $x = 0$ or there is a unique (the smallest) integer $1 \leq k \leq n$ such that $N^k(x) = 0$ but $N^{k-1}(x) \neq 0$. The existence of integer follows form the well ordering principle. It follows then, that if $x \neq 0$. The set

$$\{N^{k-1}(x), N^{k-2}(x), \cdots, N(x), x\}$$

consists of distinct nonzero vectors. In fact, the set is also linearly independent.

If not, that is, there exist numbers r, s with $1 \leq r, s \leq k-1$ such that $N^r(x) = N^s(x)$. Without loss of generality, we can consider $r = s-1$. Then

$$N^{k-s}(N^s(x)) = 0 = N^{k-s}(N^{s-1}(x)) = N^{k-1}(x).$$

It is a contradiction to the smallest integer k. Hence the set $\{N^{k-1}(x), N^{k-2}(x), \cdots, N(x), x\}$ consists of distinct nonzero vectors.

Definition 2.1. *Let $N, x \neq 0$ and k be as before. Then*

(a) *The set $\{N^{k-1}(x), N^{k-2}(x), \cdots, N(x), x\}$ is called the cycle generated by x is called the initial vector of the cycle.*
(b) *The subspace span$(\{N^{k-1}(x), N^{k-2}(x), \cdots, N(x), x\})$ is called the cyclic subspace generated by x and denoted $C(x)$.*
(c) *The integer k is called the length of the cycle.*

Proposition 2.1. *With all notationas before:*

(a) $N^{k-1}(x)$ *is an eigenvector of* N *with eigenvalue* $\lambda = 0$.
(b) $C(x)$ *is an invariant subspace of* V *under* N.
(c) *The cycle generated by* $x \neq 0$ *is a linearly independent set. Hence* $\dim(C(x)) = k$*, the length of the cycle.*

Proof.

(a) $N(N^{k-1}(x)) = N^k(x) = 0 \cdot N^{k-1}(x)$.
(b) Given $v \in C(x)$. Then there exist constants $a_0, a_1, \cdots, a_{k-1}$ in $\mathbb{F}$ such that $v = a_0 x + a_1 N(x) + \cdots + a_{k-1}N^{k-1}(x)$. Hence we get that
$$\begin{aligned} N(v) &= N(a_0 x + \cdots + a_{k-1}N^{k-1}(x)) \\ &= a_0 N(x) + \cdots + a_{k-2}N^{k-1}(x) + a_{k-1}N^k(x) \\ &= a_0 N(x) + \cdots + a_{k-2}N^{k-1}(x) \in C(x). \end{aligned}$$
(c) We will prove this by the induction on the length of the cycle. If $k = 1$, then $C(x) = \{x\}$. Since $x \neq 0$, so this set is linearly independent. Now, assume that the result has been proved for cycles of length m. Consider a cycle of length $m+1$, say $\{N^m(x), \cdots, x\}$. Then there exists $a_0, a_1, \cdots, a_m$ in $\mathbb{F}$ such that
$$a_m N^m(x) + \cdots a_1 N(x) + a_0 x = 0.$$
Hence we get that
$$\begin{aligned} 0 &= N(a_m N^m(x) + \cdots a_1 N(x) + a_0 x) \\ &= a_{m+1}N^{m+1}(x) + \cdots a_1 N^2(x) + a_0 N(x) \\ &= a_{m-1}N^m(x) + \cdots a_1 N^2(x) + a_0 N(x) \end{aligned}$$
By the induction hypothesis, $a_{m-1} = \cdots = a_0 = 0$. Since $N^m(x) \neq 0$, so $a_m = 0$. Hence we are done.

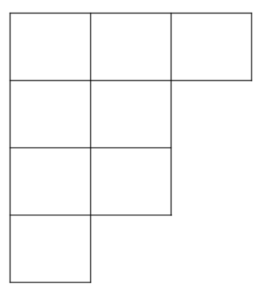
$\square$

Let N be a nilpotent mapping and $C(x)$ be the cyclic subspace generated by some $x \in V$ and let α be the cycle generated by x viewed as a basis for $C(x)$ by Proposition 2.1. Since
$$N(N^i(x)) = N^{i+1}(x)$$
for all $i \geq 1$, so we see that
$$[N|_C(x)]_\alpha^\alpha = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Thus, the matrix of N restricted to a cyclic subspace is not only upper-triangular, but an upper-triangular matrix of a very special form.

Proposition 2.2. *Let* $\alpha_i = \{N^{k_i-1}(x_i), N^{k-2}(x), \cdots, N(x_i), x_i\}$ $(1 \leq i \leq r)$ *be cycles of length* k_i*, respectively. If the set of eigenvectors* $\{N^{k_1-1}(x_1), \cdots, N^{k_r-1}(x_r)\}$ *is linear independent, then* $\alpha_1 \cup \cdots \cup \alpha_r$ *is also linearly independent.*

Proof. To make this proof somewhat more visual, we introduce a convenient pictorial representation of the vectors in the cycles, called a "cycle tableau" Let us assume that the cycle arranged so that $k_1 \leq k_2 \leq \cdots \leq k_r$. Then, the cycle tableau consists of r rows of boxes, where the boxes in the i-th row respectively the vectors in the i-th cycle, that is, there are k_i boxes in the i-th row of the tableau. We will always arrange the tableau so that the left-most boxes in each row are in the same column. In other word, cycle tableaux are left-justified. For example, if we have four cycles of lengths $3, 2, 2, 1$, then the correswponding cycle tableau is



Note that the boxes in the left-hand column represent the eigenvectors. Now, applying the mapping N to a vector in any one of the cycle (respectively a linear combination of thoses vectors) corresponding to shifting one box to the left on the corresponding row (respectively rows). Since $N^{k_i-1}(x_i)$ map to 0 under N ($\because$ it is an eigenvector), so the vectors in the left-hand column get pushed over the edge and disappear. Suppose we have a linear combination of the vectors in $\alpha \cup \cdots \cup \alpha_r$ that sums to zero. There is some power of N, say, N^l that will shift the right-most box or boxes in the tableau into the left-most column. (In general $l = k_1 - 1$.) Applying N^l to the linear combination gives a linear combination of the eigenvectors $N^{k_i-1}(x_i)$ that sums to zero. Since those eigenvectors is linearly independent from our assumption, so those coefficients must be zero. In the same way, we apply N^{l-j} to thee original linear combination for each j $(1 \geq j \geq l)$ in turn will shift each column in the tableau into the left-most columns, so all the coefficients must be zero. Hence $\alpha_1 \cup \cdots \cup \alpha_r$ is linearly independent. $\square$

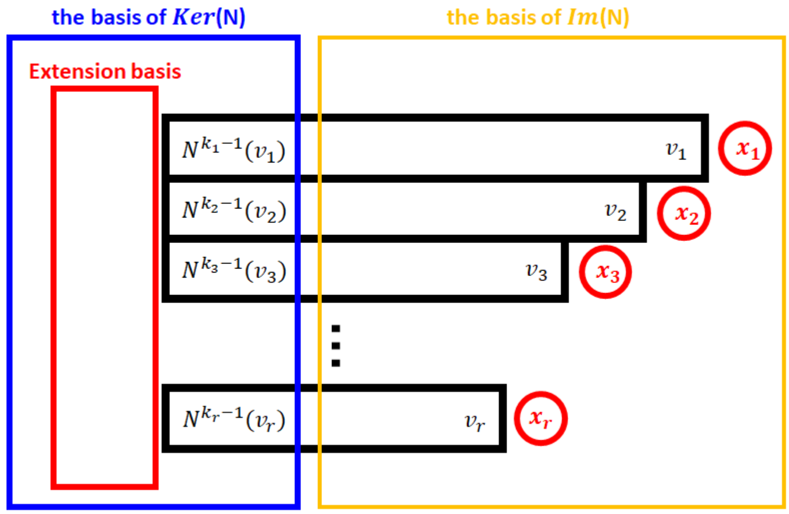
Definition 2.2. *Let $N : V \to V$ be a nilpotent mapping on a finite dimensional vector space V. We call a basis β for V a canonical basis with respect to N if β is the union of a collection of non-overlapping cycles for N.*

Theorem 2.1 (**Canonical form for nilpotent mappings**). *Let $N : V \to V$ be a nilpotent mapping on a finite dimensional vector space. Then there exists a canonical basis β of V with respect to N.*

Proof. We use the induction on $\dim(V)$. If $\dim(V) = 1$, then $N \equiv 0$ and any basis for V is canonical. Now assume that the theorem has been proved for all spaces of dimension less than k. Consider $\dim(V) = k$ and a nilpotent mapping $N : V \to V$. Since N is nilpotent, so $\dim(Im(N)) < k$. ($\because$ nilpotent mapping N is not invertiable $\therefore$ it cannot be full rank.) Furthermore, $Im(N)$ is an invariant subspace and $N|_{Im(N)}$ is also a nilpotent mapping. By induction hypothesis, there exists a canonical basis γ for $N|_{Im(N)}$. Write $\gamma = \gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_r$, wher γ_i are non-overlapping cycles for $N|_{Im(N)}$. One have to extend γ to find a basis for V.

(1) If v_i is the inital vector if the cycle γ_i, then since v_i belongs in $Im(N)$, $v_i = N(x_i)$ for some x_i in V for all $1 \le i \le n$. Let $\sigma = \{x_1, \cdots, x_r\}$. Note that $\sigma \cup \gamma$ is a collection of r non-overlapping cycles for N with the i-th cycle given by $\gamma_i \cup \{x_i\}$. (If not, that is, $x_i = x_j$, $1 \le i, j \le r$, then $N(x_i) = v_i = v_j = N(x_j)$. It is a contradiction to non-overlapping cycles. Hence $x_i \ne x_j$ for all $1 \le i, j \le r$.)
(2) The final vector of the cycle γ_i form a linearly independent subset of $Ker(N)$. However, there may be further eigenvectors in $Ker(N)$ that are not in $Im(N)$. Let α be a linearly independent subset of $Ker(N)$ such that α together with the final vectors of the cycles γ_i forms a basis $Ker(N)$.

Finally, we claim that $\beta = \alpha \cup \gamma \cup \sigma$ is the desired canonical basis. Since each of the vectors in α is a cycle of length 1, so it follows that $\beta = \alpha \cup (\gamma \cup \sigma)$ is the union of a collection of non-overlapping cycle of N by Proposition 2.2. Now. β contains $\dim(Im(N)) = \dim(V) - \dim(Ker(N))$ vectors in γ, together with the r vectors in σ and the $\dim(Ker(N)) - r$ vectors in α. Hence β contains $\dim(V)$ vectors in all. We are done.



□

Lemma. *Consider the cycle tableau with respect to a canonical basis for a nilpotent mapping $N : V \to V$. As before, let r be the number of rows and let k_i be the number of boxes in the i-th row ($k_1 \geq k_2 \geq \cdots \geq k_r$). For each j ($1 \leq j \leq k_1$), the number of boxes in the j-th column of the tableau is* $\dim(Ker(N^j)) - \dim(Ker(N^{j-1}))$.

Proof. Suppose that $N^l = 0$ but $N^{l-1} \neq 0$. Then we have

$$\{0\} \subset Ker(N) \subset Ker(N^2) \subset \cdots \subset Ker(N^{l-1}) \subset Ker(N^l) = V.$$

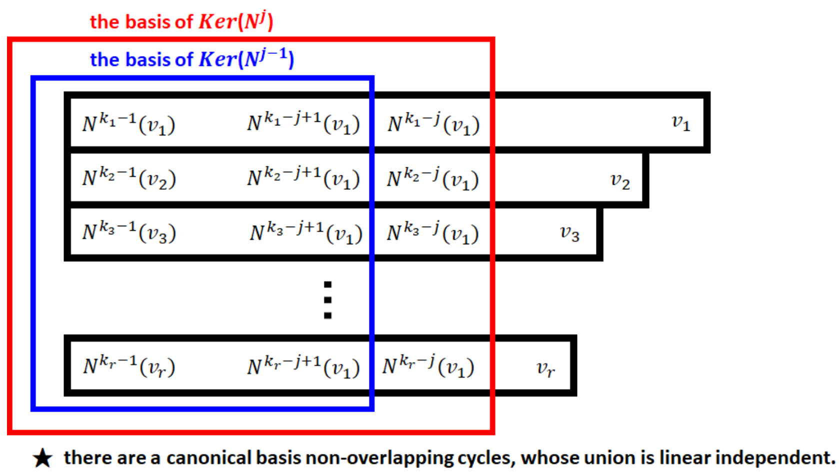
It suffices to prove that $Ker(N^k) \subset Ker(N^{k+1})$ for all $1 \leq k \leq l-1$. Given $x \in Ker(N^k)$. Then

$$\begin{aligned} N^k(x) = 0 &\implies N(N^k(x)) = N(0) = 0 \\ &\implies N^{k+1}(x) = 0 \\ &\implies x \in Ker(N^{k+1}). \end{aligned}$$

Since $N^l = 0$ but $N^{l-1} \neq 0$, so there exists $x \in Ker(N^l)$ but $x \notin Ker(N^{l-1})$. Then $N^{l-i}(x) \in Ker(N^i)$ and $N^{l-i}(x) \notin Ker(N^{i-1})$ for all $1 \leq i \leq l-1$. Hence each inclusion os strict. Moreover, one can also prove that

$$\{0\} = Im(N^l) \subset Im(N^{l-1}) \subset \cdots \subset Im(N^2) \subset Im(N) \subset V.$$

Thus, for each j, $\dim(Ker(N^j)) - \dim(Ker(N^{j-1}))$ measures the new vectors in $Ker(N^j)$ that are not in $Ker(N^{j-1})$. Now, it is easy to see that if we represent the vectors in a canonical basis for V by the corresponding cycle tableau, then the vectors represented by the boxes in the first j columns (from the left) of the tableau are all in $Ker(N^j)$.



As Proposition 2.2, we see that N^j will push all of these vectors off the left-hand edge of the tableau, that is, N^j sends to these vectors to 0. By the same reasoning, the vectors represented by the boxes in the j-th column of the tableau represent a set of linearly independent vectors in $Ker(N^j) \setminus Ker(N^{j-1})$ whcih together with the basis of $Ker(N^j)$ form a linearly independent set. The number of boxes in the j-th column is less than $\dim(Ker(N^j)) - \dim(Ker(N^{j-1}))$. On the other hand, we find that

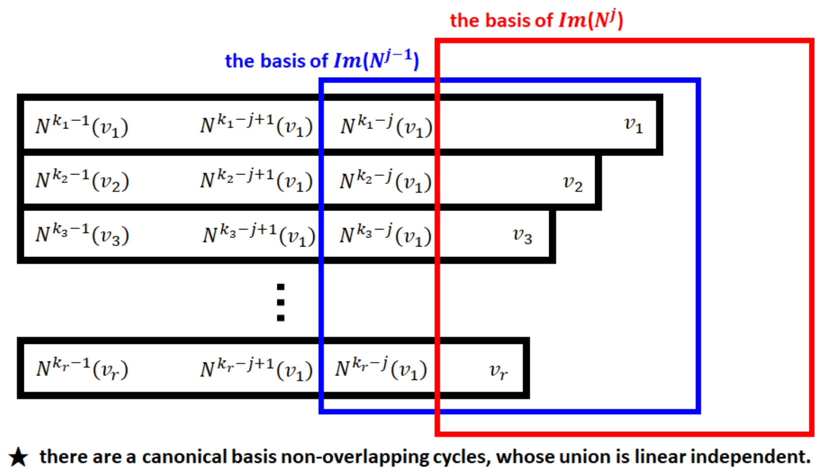
$$\begin{aligned}
&\sum_{j=1}^{l}(\dim(Ker(N^j)) - \dim(Ker(N^{j-1}))) \\
&= \dim(Ker(N)) - 0 + \dim(Ker(N^2)) - \dim(Ker(N))+ \\
&\qquad \cdots + \dim(Ker(N^l)) - \dim(Ker(N^{l-1})) \\
&= \dim(Ker(N^l)) = \dim(V).
\end{aligned}$$

Since our previous discussion and the fact that the total number boxex in the tableau is euqal to $\dim(V)$, so

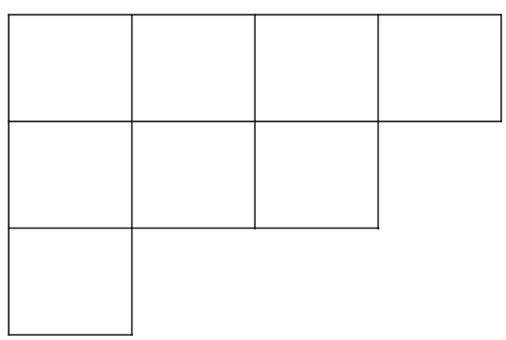
$$\begin{aligned}
\dim(V) &= \sum_{j=1}^{l}(\text{The number of boxes in the } j\text{-th column}) \\
&\leq \sum_{j=1}^{l}(\dim(Ker(N^j)) - \dim(Ker(N^{j-1}))) = \dim(V).
\end{aligned}$$

Hence the result follows. $\square$

Remark. Let $N : V \to V$ be a nilpotent mapping. Since V is a finite dimensional vector space, so we use the dimension theorem to get that $\dim(Ker(N^j)) - \dim(Ker(N^{j-1})) = \dim(Im(N^{j-1})) - \dim(Im(N^j))$.



Example. If the cycle tableau with respect to a canonical basis for N as follows:



We get that $\dim(Ker(N)) = 3$, $\dim(Ker(N^2)) = 5$, $\dim(Ker(N^3)) = 7$ and $\dim(Ker(N^4)) = 8$. Conversely, if we had computed these dimensions for a nilpotent mapping N on a vector space of dimension 8, then the cycle talbeau would be uniquely determined by this information. Indeed, suppose that $\dim(Ker(N)) = 3$, $\dim(Ker(N^2)) = 5$, $\dim(Ker(N^3)) = 7$ and $\dim(Ker(N^4)) = 8$ holds. When the columnsare assembled to form the tableau, by Lemma, we can get the same cycle tableau as before.

Corollary 2.1. *The canonical form of a nilpotent mapping is unique (provided the cyles in the canonical basis are arranged so the lengths satisfy $k_1 \geq k_2 \geq \cdots \geq k_r$).*

Proof. The number of boxes in each column of the cycle tableau (and hance entire tableau and the canonical form) is determined by the integers $\dim(Ker(N^j))$ for all $j \geq 1$. $\square$

Corollary 2.2. *N is a nilpotent if and only if there exists an invertible matrix $Q \in M_{n\times n}(\mathbb{F})$ such that $Q^{-1}NQ$ is in canonical form.*

Proof. It is trivial. $\square$

3. Jordan Canonical Form

Proposition 3.1. *Let $T : V \to V$ be a linear mapping whose characteristic polynomial ha $\dim(V)$ roots (λ_i with respective multiplicities m_i for all $1 \leq i \leq k$) in the field $\mathbb{F}$ over which V is defined.*

(a) *There exist subspaces $V_i' \subseteq V$ $(1 \leq i \leq k)$ such that*
  (1) *Each V_i' in invariant under T,*
  (2) *$T|_{V_i'}$ has exactly one distinct eigenvalue λ_i, and*
  (3) *$V = V_1' \oplus \cdots \oplus V_k'$.*

(b) *There exists a baswis β for V such that $[T]_\beta^\beta$ has a direct sum decomposition into upper triangular blocks of the form*

$$\begin{bmatrix} \lambda & * & * & \cdots & * & * \\ 0 & \lambda & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & * \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}.$$

(*The entries above the diagonal are arbitrary and all entries in the matrix other than those in the diagonal blocks are zero.*)

Remark. From above Proposition 3.1, (1) implies that V_i' is called a generalized eigenspace and (2) implies that $T|_{V_i'}$ is a nilpotent mapping.

Before begining the proof, we will indicate with a picture exactly what the relation is between this proposition and our previous results on triangularization. Since the characteristic polynomial has $\dim(V)$ roots, so there exists a basis α for V such that $[T]_\alpha^\alpha$ is upper-triangular of the form

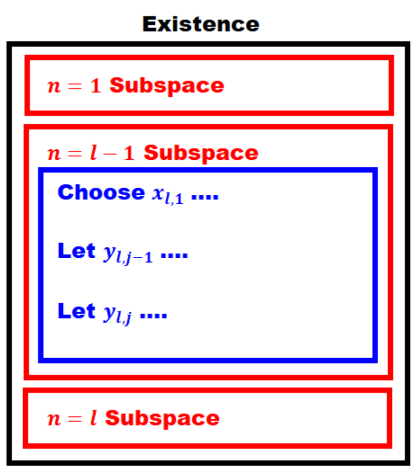
$$\begin{bmatrix} \lambda_1 & * & * & \cdots & * & * \\ 0 & \lambda_1 & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_k & * \\ 0 & 0 & 0 & \cdots & 0 & \lambda_k \end{bmatrix}. \tag{1}$$

(The $*$ above the diagonal represents the fact that those entries of the matrix are arbitrary.) For each $1 \leq k \leq k$ we draw in an $m_i \times m_i$

block on the main diagonal of the matrix in the rows and columns corresponding ti the entries λ_i on the diagonal. Then, the content of Proposition 3.1 is that by choosing another basis, we may obtain another upper-triangular matrix, but in which now all the entries outside these blocks on the diagonal are zero.

$$\begin{bmatrix} \lambda_1 & * & * & & & & & \\ 0 & \ddots & * & & & & & \\ 0 & 0 & \lambda_1 & & & & & \\ & & & \lambda_2 & * & * & & \\ & & & 0 & \ddots & * & & \\ & & & 0 & 0 & \lambda_2 & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \end{bmatrix}. \tag{2}$$

In order to replaqce the unwanted nonzero entries in (1) by zeros, we use an inductive procedure that adds a correction term to each member of the basis α to produce a corresponding vector in the new basis β. There is a flowchart proof of Theorem 3.1. We use the induction twice in the red and blue block .



Proof.

(a) Since the characteristic polynomial splits, so, by Theorem 1.1 (**Schur's Lemma**), T is triangularizable. Then there exists a basis $\alpha = \{x_{1,1}, \cdots, x_{1,m_1}, \cdots, x_{k,1}, \cdots, x_{k,m_k}\}$ for V such that $[T]_\alpha^\alpha$ is upper-triangular with diagonal entries $\lambda_1, \cdots, \lambda_1, \cdots, \lambda_k, \cdots, \lambda_k$ in that order.Let $V_i = span(\{x_{i,1}, \cdots, x_{i,m_i}\})$ for each $1 \leq i \leq k$. By construction the subspaces $V_1, V_1 + V_2, \cdots V_1 +$

$V_2 + \cdots + V_k$ are all T-invariant by Proposition 1.1. Now, we have to prove the existence of our desired subspaces V_i' for all $1 \le i \le k$.

Claim.
(1) Each V_i' is T-invariant for all $1 \le i \le k$,
(2) $T|_{V_i'}$ has exactly one distinct eigenvalue λ_i, and
(3) For each $1 \le i \le k$, $V_1 + V_2 + \cdots + V_i = V_1' + V_2' + \cdots + V_i'$.

Clearly, the condition (2) implies that $\dim(V_i') \le m_i$, so that if condition (3) holds as well. then V will be the direct sum of the subspaces V_i'. ($\because$ $V = V_1 \oplus \cdots \oplus V_k$ is and only if $\dim(V) = \dim(V_1) + \cdots \dim(V_k)$. One can see Excercise $5-2-20$ in Friedberg.) We will construct these subspace and the desired basis $\beta = \{y_{1,1}, \cdots, y_{1,m_1}, \cdots, y_{k,1}, \cdots, y_{k,m_k}\}$ by an inductive process. First, let $V_1' = V_1$ and let $y_{1,j} = x_{1,j}$ for all $1 \le j \le m_1$. Then it is trivial. Now, suppose we have constructed subspaces $V_1', \cdots, v_{l-1}'$, satisfying (1) and (2) and such that

$$V_1 + V_2 + \cdots + V_{l-1} = V_1' + V_2' + \cdots + V_{l-1}'.$$

(i) Consider the vector $x_{l,1}$. Since $x_{l,1}$ belongs to α, so we know that

$$T(x_{l,1}) = \lambda_l x_{l,1} + z,$$

where $z \in V_1 + V_2 + \cdots + V_{l-1}$. Hence, $(T - \lambda_l I)(x_{l,1}) = z$. However, by our construction, the only eigenvalues of $T|_{V_1 + \cdots V_{l-1}}$ are $\lambda_1, \cdots, \lambda_{l-1}$ so since λ_l is distinct from these, it implies that

$$S \equiv (T - \lambda_l I)|_{V_1 + \cdots V_{l-1}}$$

is invertible.

> Suppose not, that is, $Ker(S) = \{v\}$, where $v \ne 0$. Then $S(v) = 0$, that is, $Tv = \lambda_l v$ on $V_1 + V_2 + \cdots + V_{l-1}$. It means that $T|_{V_1 + \cdots V_{l-1}}$ has an eignevalue λ_l. Is is a contradiction. Hence $Ker(S) = \{0\}$, that is, S is one to one. Since V is a finite dimensional vector space, so S is invertible on $V_1 + \cdots V_{l-1}$.

Since S is onto on $V_1 + \cdots V_{l-1}$, so there exists a vector w in $V_1 + \cdots V_{l-1}$ such that

$$(T - \lambda_l I)(w) = z.$$

Then

$$(T-\lambda_l I)(w) = z = (T-\lambda_l I)(x_{l,1}),$$

that is, $(T-\lambda_l I)(w - x_{l,1}) = 0$ Let $y_{l,1} = x_{l,1} - w$, where w is the vector of the correction term.

(ii) Now, inductively again, assume we have constructed vectors $y_{l,1}, \cdots, y_{l,j-1}$ such that $T(y_{l,n}) = \lambda_l y_{l,n} + u_n$, where $u_n \in span(\{y_{l,1}, \cdots, y_{l,n-1}\})$ for all $1 \leq n \leq j-1$ and

$$\begin{aligned}
&span(\{y_{l,1}, \cdots, y_{l,j-1}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j-1}\}) + V_1 + \cdots + V_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j-1}\}) + V'_1 + \cdots + V'_{l-1}.
\end{aligned}$$

We know that

$$T(x_{l,j}) = \lambda_j x_{l,j} + u + z,$$

where $u \in span(\{y_{l,1}, \cdots, y_{l,j-1}\})$ and $z \in V'_1 + \cdots + V'_{l-1}$ by the construction of the triangular basis α and induction. As before discussion,

$$(T-\lambda_l I)(x_{l,j}) = u + z.$$

Since the only eigenvalues of $T|_{V'_1+\cdots+V'_{l-1}}$ are $\lambda_1, \cdots, \lambda_{l-1}$, so, by the same reason, there exists a vector $w \in V'_1 + \cdots + V'_{l-1}$ such that $(T-\lambda_l I)(w) = z$. Then

$$(T-\lambda_l I)(x_{l,j} - w) = u,$$

where $u \in span(\{y_{l,1}, \cdots, y_{l,j-1}\})$. Let $y_{l,j} = x_{l,j} - w$. We get that

$$\begin{aligned}
&span(\{y_{l,1}, \cdots, y_{l,j}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j}\}) + V_1 + \cdots + V_{l-1}.
\end{aligned}$$

$$\begin{aligned}
&span(\{y_{l,1}, \cdots, y_{l,j}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{y_{l,1}, \cdots, x_{l,j} - w\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{y_{l,1}, \cdots y_{l,j-1}\}) + span(\{x_{l,j}\}) \\
&\qquad + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j-1}\} \cup \{x_{l,j}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j}\}) + V'_1 + \cdots + V'_{l-1} \\
&= span(\{x_{l,1}, \cdots, x_{l,j}\}) + V_1 + \cdots + V_{l-1}.
\end{aligned}$$

Hence the induction can continue and we find the subspace

$$V_l' = span(\{y_{l,1}, \cdots y_{l,m_l}\})$$

after m_l steps in all. By construction, V_l' is invariant under T since it is $T - \lambda_l I$ invariant. Thus, we find the desired basis for V.

(b) This statement follows immediately from part (a).

□

Remark. Clearly, V_i' is the generalized eigenspaces of T for all $1 \leq i \leq k$. We know that $T|_{V_i'}$ has only one eigenvalue λ_i such that $[T - \lambda_i I|_{V_i'}]_{\beta_i}^{\beta_i}$ is an upper-triangular matrix with diagonal entries 0, where β_i consists of eigenvectors with λ_i. One can see the matrix representation of T and $T - \lambda_i I$ on V_i', as follows:

$$[T|_{V_i'}]_{\beta_i}^{\beta_i} = \begin{bmatrix} \lambda_i & * & * & \cdots & * & * \\ 0 & \lambda_i & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & * \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}$$

and

$$[T - \lambda_i I|_{V_i'}]_{\beta_i}^{\beta_i} = \begin{bmatrix} 0 & * & * & \cdots & * & * \\ 0 & 0 & * & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & * \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Hence $T - \lambda_i I|_{V_i'}$ has the only eigenvalue 0 with multiplicities m_i if and only if $T - \lambda_i I|_{V_i'}$ is a nilpotent mapping. Since $m_i = \dim(V)$, it is easy to see that $(T - \lambda_i I|_{V_i'})^{m_i} = 0$ (the zero mapping on V_i') and that, furthermore, $V_i' = Ker((T - \lambda_i I)^{m_i})$ (as a mapping on all of V). This is true because of the observation we used several times in the proof of Proposition 3.1. Since the λ_i are distinct on the sum of the V_i' for $i \neq j$, $T - \lambda_i I$ is invertible. The reason follows form Theorem 7.1 in Friedbreg or the direct sum of V. ($\because V_i \cap \sum_{j \neq i} V_j' = \{0\} \therefore T - \lambda_i I|_{\sum_{j \neq i} V_j'}$ is invertible.)

Definition 3.1. *Let $T : V \to V$ be a linear transformation on a finite dimensional vector space V. Let λ be an eignevlaue of T with multiplicity m.*

- *The λ- generalized eigenspace, denoted K_λ, is the kernal of the mapping $(T - \lambda I)^m$ on V.*

- *The nonzero elements of K_λ are called generalized eigenvectors of T.*

In other words, a generalized eigenvector of T is any vector $x \neq 0$ such that $(T-\lambda I)^m(x) = 0$. Note that this will certainlly be the case if $(T-\lambda I)^k(x) = 0$ for some $k \leq m$. In particular, the eigenvectors with eigenvalue λ are also generalized eigenvectors, and we have $E_\lambda \subseteq K_\lambda$ for each eigenvalue.

Proposition 3.2.

(a) *For each eigenvalue λ of T, K_λ is a invariant subspace of V.*
(b) *If λ_i $(1 \leq i \leq k)$ are the distinct eigenvalues of T, then $V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \cdots \oplus K_{\lambda_k}$.*
(c) *if λ is an eigenvector of multiplicity m, then* $\dim(K_\lambda) = m$.

Proof. It is trivial because Proposition 6.3 and Definition 3.1. □

From above discussion, we know that $N_i = T - \lambda_i I|_{K_{\lambda_i}}$ is a nilpotent mapping. Since N_i is a nilpotent mapping, so there exists a canonical basis γ_i for K_{λ_i} such that

$$[N_i]_{\gamma_i}^{\gamma_i} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

As a result,

$$\begin{aligned} [T|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i} &= [N_i + \lambda_i I|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i} \\ &= [N_i|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i} + [\lambda_i I|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i} \\ &= \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}, \end{aligned}$$

denoted by $(*)$.

Definition 3.2.

(a) *A matrix of the form $(*)$ is called a Jordan block matrix.*
(b) *A matrix $A \in M_{n\times n}(\mathbb{F})$ is said to be in Jordan canonical form if A is a direct sum of Jordan block matrix.*

Theorem 3.1 (**Jordan Canonical Form**). *Let $T : V \to V$ be a linear mapping on a finite dimensional vector space V whose characteristic polynomail has $\dim(V)$ roots in the field $\mathbb{F}$ over which V is defined.*

(a) *There exists a basis γ (called a canonical basis) of V such that $[T]_\gamma^\gamma$ has a direct sum decomposition into Jordan block matrices.*
(b) *In this decomposition the number of Jordan blocks and their sizes are uniquely determined by T. (This order in which the blocks appear in the matrix may be different for different canonical bases, however.)*

Proof.

(a) In each generalized eigenspace K_{λ_i}, as we have noted, by Theorem 3.1, there is a basis γ_i such that $[T|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i}$ has a direct sum decomposition into Jordan blocks with diagonal entries λ_i. In addition, we know that V is direct sum of the subspaces K_{λ_i}. Let $\gamma = \bigcup_{i=1}^k \gamma_i$. Then γ is a basis for V because of Theorem 5.10 in Friedberg. We see that $[T]_\gamma^\gamma$ has desired form.
(b) This statement follows from Corollary 2.1.

$\square$

Theorem 3.2. *Let $A \in M_{n\times n}(\mathbb{C})$. We say that A has a unique Jordan Chevalley decomposition if there exist a unique diagonalizable matrix D and nilpotent matrix N such that $A = D + N$ and $ND = DN$.*

Proof. One can see Linear Algebra (Kenneth Hoffman) p.262 to 269.

$\square$