Chapter VII. Nondiagonalizable Operators.

VII.1. Basic Definitions and Examples.

Nilpotent operators present the first serious obstruction to attempts to diagonalize a given linear operator.

1.1. Definition. *A linear operator $T : V \to V$ is* **nilpotent** *if $T^k = 0$ for some $k \in \mathbb{N}$; it is* **unipotent** *if $T = I + N$ with N nilpotent.*

Obviously T is unipotent $\Leftrightarrow T - I$ is nilpotent.

Nilpotent operators cannot be diagonalized unless T is the zero operator (or $T = I$, if unipotent). Any analysis of normal forms must examine these operators in detail. Nilpotent and unipotent matrices $A \in \mathrm{M}(n,\mathbb{F})$ are defined the same way. As examples, all strictly upper triangular matrices (with zeros on the diagonal) as well as those that are strictly lower triangular, are nilpotent in view of the following observations.

1.2. Exercise. If A has upper triangular form with zeros on and below the diagonal, prove that

$$A^2 = \begin{pmatrix} 0 & 0 & & * \\ & \cdot & \cdot & \\ & & \cdot & 0 \\ 0 & & & 0 \end{pmatrix} \qquad A^3 = \begin{pmatrix} 0 & 0 & 0 & & * \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & 0 \\ & & & \cdot & 0 \\ 0 & & & & 0 \end{pmatrix},$$

etc, so that $A^n = 0$. $\square$

Matrices of the same form, but with 1's on the diagonal all correspond to unipotent operators.

We will see that if $N : V \to V$ is nilpotent there is a basis $\mathfrak{X}$ such that

$$[N]_{\mathfrak{X}} = \begin{pmatrix} 0 & & * \\ & \cdot & \\ & & \cdot \\ 0 & & 0 \end{pmatrix},$$

but this is not true for all bases. Furthermore, a lot more can be said about the terms $(*)$ for suitably chosen bases.

1.3. Exercise. In $\mathrm{M}(n,\mathbb{F})$, show that the sets of upper triangular matrices:

(a) The **strictly upper triangular group** $\mathcal{N} = \left\{ \begin{pmatrix} 1 & & * \\ & \cdot & \\ & & \cdot \\ 0 & & 1 \end{pmatrix} \right\}$ with entries in $\mathbb{F}$.

(b) The **full upper triangular group** in $\mathrm{M}(n,\mathbb{F})$, $\mathcal{P} = \left\{ \begin{pmatrix} a_{1,1} & & * \\ & \cdot & \\ & & \cdot \\ 0 & & a_{n,n} \end{pmatrix} \right\}$ with entries in $\mathbb{F}$ such that $\prod_{i=1}^n a_{i,i} \neq 0$.

are both subgroups in $\mathrm{GL}(n,\mathbb{F})$, with $\det(A) = \prod_{i=1}^n a_{i,i} \neq 0$ for elements of either group. Verify that $\mathcal{N}$ and $\mathcal{P}$ are closed under taking products and inverses. $\square$

1.4. Exercise. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ in $\mathrm{M}(2,\mathbb{F})$. This is a nilpotent matrix and in any ground field the only root of its characteristic polynomial

$$p_A(\lambda) = \det(A - \lambda I) = \lambda^2$$

is $\lambda = 0$. There is a nontrivial eigenvector $e_1 = (1,0)$, corresponding to eigenvalue $\lambda = 0$, because $\ker(A) = \mathbb{F} \cdot e_1$ is nontrivial (as it must be for any nilpotent operator). But you can easily verify that scalar multiples of e_1 are the only eigenvectors, so there is no basis of eigenvectors. A cannot be diagonalized by any similarity transformation, Regardless of the ground field $\mathbb{F}$. $\square$

"Stable Range" and "Stable Kernel" of a Linear Map. If $T : V \to V$ is a linear operator on a finite dimensional vector space (arbitrary ground field), let $K_i = K(T^i) = \ker(T^i)$ and $R_i = R(T^i) = \mathrm{range}(T^i)$ for $i = 0, 1, 2, \cdots$. Obviously these spaces are nested

$$\begin{aligned} &(0) \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_i \subseteq K_{i+1} \subseteq \cdots \\ &V \supseteq R_1 \supseteq R_2 \supseteq \cdots \supseteq R_i \supseteq R_{i+1} \supseteq \cdots\ , \end{aligned}$$

and if $\dim(V) < \infty$ they must each stabilize at some point, say with $K_r = K_{r+1} = \cdots$ and $R_s = R_{s+1} = \cdots$ for some integers r and s. In fact if r is the first (smallest) index such that $K_r = K_{r+1} = \cdots$ the sequence of ranges must also stabilize at the same point because $|V| = |K_i| + |R_i|$ at each step. With this in mind, we define (for finite dimensional V)

$$\begin{aligned} R_\infty &= \bigcap_{i=1}^{\infty} R_i \;=\; R_r \;=\; R_{r+1} \;=\; \cdots \qquad (\textbf{Stable range} \text{ of } T) \\ K_\infty &= \bigcup_{i=1}^{\infty} K_i \;=\; K_r \;=\; K_{r+1} \;=\; \cdots \qquad (\textbf{Stable kernel} \text{ of } T) \end{aligned}$$

1.5. Proposition. *$V = R_\infty \oplus K_\infty$ and the spaces R_∞, K_∞ are T-invariant. Furthermore $R_{i+1} \neq R_i$ and $K_{i+1} \neq K_i$ for $i < r$.*

Note: This splitting is sometimes referred to as the "Fitting decomposition" (after a guy named Fitting).

Proof: To see there is a non-trivial jump $R_{i+1} \subsetneqq R_i$ at every step until $i = r$ if suffices to show that $R_{i+1} = R_i$ at some step implies $R_i = R_j$ for all $j \geq i$ (a similar result for kernels then follows automatically). It suffices to show that $R_i = R_{i+1} \Rightarrow R_{i+1} = R_{i+2}$. Obviously, $R_{i+2} \subseteq R_{i+1}$ for all i; to prove the reverse inclusion $R_{i+1} \subseteq R_{i+2}$, let $v \in R_{i+1}$. Then there is some $w_1 \in V$ such that $v = T^{i+1}(w_1) = T\big(T^i(w_1)\big)$. By hypothesis $R_{i+1} = T^{i+1}(V) = R_i = T^i(V)$ so there is some $w_2 \in V$ such that $T^i(w_1) = T^{i+1}(w_2)$. Thus

$$v = T^{i+1}(w_2) = T(T^i(w_1)) = T(T^{i+1}(w_2)) = T^{i+2}(w_2) \in R_{i+2}$$

So, $R_{i+1} \subseteq R_{i+2}$, $R_i = R_{i+1} = R_{i+2}$, and by induction $R_i = R_{i+1} = \cdots = R_\infty$.

For T-invariance of $R_\infty = R_r$ and $K_\infty = K_r$, T maps $R_i \to R_{i+1} \subseteq R_i$ for all i; taking $i = r$, we get $T(R_\infty) = R_\infty$. As for the kernels, if $v \in K_{i+1}$ then $0 = T^{i+1}(v) = T^i(T(v))$. As a consequence, $T(v) \in K_i$ and $T(K_{i+1}) \subseteq K_i$ for all i. For $i \geq r$, we have $K_i = K_{i+1} = K_\infty$, so $T(K_\infty) = K_\infty$ as claimed.

To see $V = K_\infty \oplus R_\infty$ we show (i) $R_\infty + K_\infty = V$ and (ii) $R_\infty \cap K_\infty = \{0\}$. For (ii), if $v \in R_\infty = R_r$ there is some $w \in V$ such that $T^r(w) = v$; but if $v \in K_\infty = K_r$, then $T^r(v) = 0$ and hence $T^r(v) = 0$. Consequently $T^{2r}(w) = T^r(v) = 0$. We now observe that $T : R_i \to R_{i+1}$ is a bijection for $i \geq r$ so $\ker(T|_{R_r}) = \ker(T|_{R_\infty}) = \{0\}$. In fact, if $i \geq r$ then $R_i = R_{i+1}$ and $T : R_i \to R_{i+1}$ is a surjective linear map, and if $T : R_i \to R_{i+1} = R_i$ is surjective it is automatically a bijection. Now in the preceding discussion $v = T^r(w) \in R_r$ and $T^r : R_r \to R_{2r} = R_r$ is a bijection, so

$$0 = T^{2r}(w) = T^r(T^r(w)) = T^r(v)$$

Then $v = 0$, hence $R_\infty \cap K_\infty = \{0\}$

For $(ii) \Rightarrow (i)$, we know

$$\begin{aligned} |R_\infty + K_\infty| &= |R_r + K_r| = |R_r| + |K_r| - |K_r \cap R_r| \\ &= |K_\infty| + |R_\infty| = |K_r| + |R_r| = |V| \end{aligned}$$

(by the Dimension Theorem). We conclude that $R_\infty + K_\infty = V$, proving (i). $\square$

1.6. Lemma. *$T|_{K_\infty}$ is a nilpotent operator on K_∞ and $T|_{R_\infty}$ is a bijective linear map of $R_\infty \to R_\infty$. Hence, every linear operator T on a finite dimensional space V, over any field, has a direct sum decomposition.*

$$T = (T|R_\infty) \oplus (T|K_\infty)$$

such that $T|_{K_\infty}$ is nilpotent and $T|_{R_\infty}$ bijective on R_∞.

Proof: $T^r(K_\infty) = T^r(\ker(T^r)) = \{0\}$ so $(T|_{K_\infty})^r = 0$ and $T|_{K_\infty}$ is nilpotent of degree $\leq r$, the index at which the ranges stabilize at R_∞.

2. Some Observations about Nilpotent Operators.

2.1. Lemma. *If $N : V \to V$ is nilpotent, the unipotent operator $I + N$ is invertible.*

Proof: If $N^k = 0$ the geometric series $I + N + N^2 + \ldots + N^{k-1} + \ldots = \sum_{k=0}^{\infty} N^k$ is finite and a simple calculation shows that

$$(I - N)(I + N + \cdots + N^{k-1}) = I - N^k = I \ .$$

Hence

$$(I - N)^{-1} = I + N + \cdots + N^{k-1} \quad \square \tag{1}$$

if $N^k = 0$. $\square$

2.2. Lemma. *If $T : V \to V$ is nilpotent then $p_T(\lambda) = \det(T - \lambda I)$ is equal to $(-1)^n \lambda^n$ ($n = \dim(V)$), and $\lambda = 0$ is the only eigenvalue (over any field $\mathbb{F}$). [It is an eigenvalue since $\ker(T) \neq \{0\}$ and the full subspace of $\lambda = 0$ eigenvectors is precisely $E_{\lambda=0}(T) = \ker(T)$].*

Proof: Take a basis $\mathfrak{X} = \{e_1, \cdots, e_n\}$ that runs first through $K(T) = K_1 = \ker(T)$, then augments to a basis in $K_2 = \ker(T^2)$, etc. With respect to this basis $[T]_{\mathfrak{X}\mathfrak{X}}$ is an upper triangular matrix with zero matrices blocks on the diagonal (see Exercise 2.4 below). Obviously, $T - \lambda I$ has diagonal values $-\lambda$, so $\det(T - \lambda I) = (-1)^n \lambda^n$ as claimed. $\square$

Similarly a unipotent operator T has $\lambda = 1$ as its only eigenvalue (over any field) and its characteristic polynomial is $p_T(x) = 1$ (constant polynomial $\equiv 1$). The sole eigenspace $E_{\lambda=1}(T)$ is the set of *fixed points* $\mathrm{Fix}(T) = \{v : T(v) = v\}$.

2.3. Exercise. Prove that

(a) A nilpotent operator T is diagonalizable (for some basis) if and only if $T = 0$.

(b) T is unipotent if and only if T is the identity operator $I = \mathrm{id}_V$ $\square$

2.4. Exercise. If $T : V \to V$ is a nilpotent linear operator on a finite dimensional space let $\mathfrak{X} = \{e_1, \ldots, e_n\}$ is a basis that passes through successive kernels $K_i = \ker(T^i)$, $1 \le i \le d = \deg(T)$. Prove that $[T]_{\mathfrak{X}}$ is upper triangular with $m_i \times m_i$ zero-blocks on the diagonal, $m_i = \dim(K_i/K_{i-1})$.
Hints: The problem is to devise efficient notation to handle this question. Partition the indices $1, 2, \ldots, n$ into consecutive intervals $J_1, \ldots, J_d$ ($d = \deg(T)$) such that $\{e_j : j \in J_1\}$ is a basis for K_1, $\{e_i : i \in J_1 \cup J_2\}$ is a basis for K_2, etc. Matrix coefficients T_{ij} are determined by the system of vector equations

$$T(e_i) = \sum_{j=1}^{n} T_{ji} e_j \qquad (1 \le i \le n = \dim(V))$$

What do the inclusions $T(K_i) \subseteq K_{i-1}$ tell you about the coefficients T_{ij}? $\square$

Let $T : V \to V$ be nilpotent. The powers T^k eventually "kill" every vector $v \neq 0$, so there is an $m \in \mathbb{N}$ such that $\{v, T(v), \cdots, T^{m-1}(v)\}$ are nonzero and $T^m(v) = 0$. The **nilpotence degree** $\deg(T)$ is the smallest exponent $d = 0, 1, 2, \cdots$ such that $T^d = 0$.

2.5. Proposition. *Let $T : V \to V$ be nilpotent and $v_0 \neq 0$. If $v_0, T(v_0), \cdots, T^{m-1}(v_0)$ are all nonzero and $T^m(v_0) = 0$ define $W(v_0) = \mathbb{F}{-}\mathrm{span}\{v_0, T(v_0), \cdots, T^{m-1}(v_0)\}$. This subspace is T-invariant and the vectors $\{v_0, T(v_0), \cdots, T^{m-1}(v_0)\}$ are independent, hence a basis for this "cyclic subspace" determined by v_0 and the action of T.*

Proof: The $\{T^k(v_0) : 0 \le k \le n-1\}$ span $W(v_0)$ by definition. They are independent because if $0 = c_0 + c_1T(v_0) + \cdots + c_{m-1}T^{m-1}(v_0)$ for some choice of $c_k \in \mathbb{F}$, then

$$\begin{aligned} 0 = T^{m-1}(0) &= T^{m-1}\big(c_0v_0 + c_1T(v_0) + \cdots + c_{m-1}T^{m-1}(v_0)\big) \\ &= c_0T^{m-1}(v_0) + c_1 \cdot 0 + \cdots + c_{m-1} \cdot 0 \ , \end{aligned}$$

which implies $c_0 = 0$ since $T^{m-1}(v_0) \neq 0$ by minimality of the exponent m. Next, apply T^{m-2} to the original sum, which has now the form $c_1T(v_0) + \cdots + c_{m-1}T^{m-1}(v_0)$; we get

$$T^{m-2}(0) = T^{m-2}\big(c_1T(v_0) + \cdots + c_{m-1}T^{m-1}(v_0)\big) = c_1T^{m-1}(v_0) + 0 + \cdots + 0$$

and then $c_1 = 0$. We can apply the same process repeatedly to get $c_0 = c_1 = c_2 = \cdots = c_{m-1} = 0$. $\square$

Obviously $W(v_0)$ is T-invariant and $T_0 = T|_{W(v_0)}$ is nilpotent (with degree $m = \deg(T_0) \le \deg(T)$) because for each basis vector $T^k(v_0)$ we have $T_0^m(T^k(v_0)) = T^k(T^m(v_0)) = 0$; but in fact $\deg(T_0) = m$ because $T_0^{m-1}(v_0) \neq \{0\}$. Now consider the ordered basis

$$\mathfrak{X} = \{e_1 = T^{m-1}(v_0), e_2 = T^{m-2}(v_0), \cdots, e_m = v_0\} \text{ in } W(v_0) \ .$$

Since $T(e_{k+1}) = e_k$ for each $k \ge 1$ and $T(e_1) = 0$, the matrix $[T]_{\mathfrak{X},\mathfrak{X}}$ has the form

$$[T]_{\mathfrak{X}} = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & & \cdot & \\ 0 & 0 & 0 & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & 1 \\ 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

The action on these ordered basis vectors is :

$$0 \xleftarrow{T} e_1 \xleftarrow{T} e_2 \xleftarrow{T} \cdots \xleftarrow{T} e_{m-1} \xleftarrow{T} e_m = v_0$$

The "top vector" $e_m = v_0$ is referred to as a **cyclic vector** for the invariant subspace $W(v_0)$. Any matrix having the form

$$\begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & & \cdot & \\ 0 & 0 & 0 & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & 1 \\ 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

is called an **elementary nilpotent matrix**.

Cyclic Vectors and Cyclic Subspaces for General Linear Operators. To put this in its proper context we leave the world of nilpotent operators for a moment.

2.6. Definition. *If* $\dim(V) < \infty$*,* $T : V \rightarrow V$ *is a linear operator, and* $W \subseteq V$ *a nonzero* T*-invariant subspace, we say* W *is a* **cyclic subspace** *if it contains a "cyclic vector"* $v_0 \in W$ *such that* $W = \mathbb{F}\text{-span}\{v_0, T(v_0), T^2(v_0), \cdots\}$.

Only finitely many iterates $T^i(v_0)$ under the action of T can be linearly independent, so there will be a first (smallest) exponent $k = k(v_0)$ such that $\{v_0, T(v_0), \cdots, T^{k-1}(v_0)\}$ are linearly independent and $T^k(v_0)$ is a linear combination of the previous vectors.

2.7. Proposition. *Let* $T : V \rightarrow V$ *be an arbitrary linear operator on a finite dimensional vector space. If* $v_0 \in V$ *is non-zero there is a unique exponent* $k = k(v_0) \geq 1$ *such that* $\{v_0, T(v_0), \cdots, T^{k-1}(v_0)\}$ *are linearly independent and* $T^k(v_0)$ *is a linear combination of these vectors. Obviously,*

$$W = \mathbb{F}\text{−span}\{T^j(v_0) : j = 0, 1, 2, \cdots\} = \mathbb{F}\text{−span}\{v_0, T(v_0), \cdots, T^{k-1}(v_0)\}$$

and $\dim(W) = k$*. Furthermore,* $T(W) \subseteq W$ *and* W *is a cyclic subspace in* V*.*

Proof: By definition of $k = k(v_0)$, $T^k(v_0)$ is a linear combination $T^k(v_0) = \sum_{j=0}^{k-1} c_j T^j(v_0)$. Arguing recursively,

$$\begin{aligned} T^{k+1}(v_0) &= T(T^k(v_0)) = \sum_{j=0}^{k-1} c_j T^{j+1}(v_0) \\ &= (c_{k-1}T^k(v_0)) + \big(\text{linear combinations of } v_0, T(v_0), \cdots, T^{k-1}(v_0)\,\big) \end{aligned}$$

Since we already know $T^k(v_0)$ lies in $\mathbb{F}\text{-span}\{v_0, T(v_0), \cdots, T^{k-1}(v_0)\}$, so does $T^{k+1}(v_0)$. Continuing this process, we find that all iterates $T^i(v_0)$ $(i \geq k)$ lie in W. By definition $v_0, T(v_0), \cdots, T^{k-1}(v_0)$ are linearly independent and span W, so $\dim(W) = k$. $\square$

When T is nilpotent there is a simpler alternative description of the cyclic subspace W generated by the action of T on $v_0 \neq 0$. Since $T^d = 0$ on all of V when $d = \deg(T)$, there is a smallest exponent l such that $\{v_0, T(v_0), \cdots, T^{\ell-1}(v_0)\}$ are nonzero and $T^\ell(v_0) = T^{\ell+i}(v_0) = 0$ for all $i \geq 0$. These vectors are independent and the next vector $T^\ell(v_0) = 0$ lies in $\mathbb{F}\text{-span}\{v_0, T(v_0), \cdots, T^{\ell-1}(v_0)\}$, so ℓ is precisely the exponent of the previous lemma and $C = \mathbb{F}\text{-span}\{v_0, T(v_0), \cdots, T^{\ell-1}(v_0)\}$ is the cyclic subspace generated by v_0.

3. Structure of Nilpotent Operators.

Resuming the discussion of nilpotent operators, we first observe that if $T : V \rightarrow V$ is nilpotent and nonzero the chain of kernels $K_i = \ker(T^i)$,

$$\{0\} = K_0 \subsetneqq K_1 = \ker(T) \subsetneqq K_2 \subsetneqq \cdots \subsetneqq K_d = V \qquad (d = \deg(T))$$

terminates at V in finitely many steps. The difference sets partition $V \sim (0)$ into disjoint "layers"

$$V \sim (0) = \big(K_d \sim K_{d-1}\big) \cup \cdots \cup \big(K_i \sim K_{i-1}\big) \cup \cdots \cup \big(K_1 \sim K_0\big)$$

where $K_0 = (0)$. The layers $K_i \sim K_{i-1}$ correspond to the quotient spaces K_i/K_{i-1}, and by examining the action of T on these quotients we will be able to determine the structure of the operator T.

3.1. Exercise. If v_0 is in the "top layer" $V \sim K_{d-1}$, prove that $\mathbb{F}$-span$\{T^j(v_0) : j \geq 0\}$ has dimension d and *every* such v_0 is a cyclic vector under the iterated action of T on W. $\square$

Since $\dim(K_{d-1}) < \dim(K_d) = \dim(V)$, K_{d-1} is a very thin subset of V and has "measure zero" in V when $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. If you could pick a vector $v_0 \in V$ "at random," you would have $v_0 \in V \sim K_{d-1}$ "with probability 1," and every such choice of v_0 would generate a cyclic subspace of dimension d. "Unsuccessful" choices, which occur with "probability zero," yield cyclic subspaces $W(v_0)$ of dimension $< d$.

We now state the main structure theorem for nilpotent operators .

3.2. Theorem (Cyclic Subspace Decomposition). *Given a nilpotent linear operator $T : V \rightarrow V$ on a finite dimensional vector space V, there is a decomposition $V = V_1 \oplus \cdots \oplus V_r$ into cyclic T-invariant subspaces. Obviously the restrictions $T_i = T|_{V_i}$ are nilpotent, with degrees*

$$m_i = \dim(V_i) = (\textit{smallest exponent } m \textit{ such that } T^m \textit{ kills the cyclic generator } v_i \in V_i)$$

These degrees are unique when listed in descending order $m_1 \geq m_2 \geq \cdots \geq m_r > 0$ (repeats allowed), and $\sum_{i=1}^{r} m_i = \dim(V)$.

While it is nice to know such structure exists, it is equally important to develop a constructive procedure for finding suitable cyclic subspaces $V_1, \cdots, V_r$. This is complicated by the fact that the cyclic subspaces are not necessarily unique, unlike the eigenspaces $E_\lambda(T)$ associated with a diagonalizable operator. Any algorithm for constructing suitable V_i will necessarily involve some arbitrary choices.

The rest of this section provides a proof of Theorem 3.2 that yields on an explicit construction of the desired subspaces. There are some very elegant proofs of Theorem 3.2, but they are existential rather than constructive and so are less informative.

3.3. Corollary. *If $T : V \rightarrow V$ is nilpotent, there is a decomposition into cyclic spaces $V = V_1 \oplus \ldots \oplus V_r$, so there is a basis $\mathfrak{X}$ such that $[T]_{\mathfrak{X}}$ consists of elementary nilpotent diagonal blocks.*

$$[T]_{\mathfrak{X}} = \begin{pmatrix} \boxed{B_1} & 0 & 0 & \cdot & & 0 \\ 0 & \boxed{B_2} & 0 & & \cdot & \\ 0 & 0 & 0 & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \\ 0 & \cdot & \cdot & \cdot & 0 & \boxed{B_r} \end{pmatrix}$$

with

$$B_i = \begin{pmatrix} 0 & 1 & 0 & & & 0 \\ 0 & 0 & 1 & & & \cdot \\ 0 & 0 & 0 & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & 1 \\ 0 & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}$$

We start with the special case in which T has the largest possible degree of nilpotence.

3.4. Lemma. *If T is nilpotent and $\deg(T) = \dim(V)$, there is a cyclic vector in V and a basis such that $[T]_{\mathfrak{X}}$ has the form B_i of an elementary nilpotent matrix.*

Proof: If $\deg(T) = d$ is equal to $\dim(V)$, the spaces $K_i = \ker(T^i)$ increase with $|K_{i+1}| \geq 1 + |K_i|$ at each step in the chain $\{0\} \subsetneqq K_1 \subseteq \cdots \subseteq K_{d-1} \subseteq K_d = V$. There are $d = \dim(V)$ steps so we must have $|K_{i+1}| = 1 + |K_i|$. Take any vector $v_0 \in V \sim K_{d-1}$. Then $T^d(v_0) = 0$ but by definition of K_{d-1}, $v_0, T(v_0), \cdots, T^{d-1}(v_0)$ are all nonzero, so v_0 is a cyclic vector for the iterated action of T. $\square$

If $T : V \to V$ is nilpotent of degree d, the idea behind proof of Theorem 3.1 is to look at the kernels $K_i = \ker(T^i)$.

$$V = K_d \supsetneqq K_{d-1} \supsetneqq \cdots \supsetneqq K_2 \supsetneqq K_1 = \ker(T) \supsetneqq \{0\}$$

As the kernels get smaller, more of V is "uncovered" (the difference set $V \sim K_s$ and the quotient V/K_s get bigger) and the action in V/K_s reveals more details about the full action of T on V.

It will be important to note that $T(K_i) \subseteq K_{i-1}$ (since $0 = T^i(x) = T^{i-1}(T(x))$ and $T(x) \in K_{i-1}$). Furthermore, $x \notin K_i$ implies that $0 \neq T^i(x) = T^{i-1}(T(x))$ so that $T(x) \notin K_{i-1}$. Thus

$$T \text{ maps } K_{i+1} \sim K_i \text{ into } K_i \sim K_{i-1} \text{ for all } i. \tag{2}$$

But it is not generally true that $T(K_j) = K_{j-1}$.

3.5. Definition. *Let $T : V \to V$ be an arbitrary linear map and W a T-invariant subspace. We say that vectors $e_1, \cdots, e_m$ in V are:*

1. **Independent (mod W)** *if their images $\overline{e}_1, \cdots, \overline{e}_m$ in V/W are linearly independent. Since $\sum_i c_i \overline{e}_i = 0$ in V/W if and only if $\sum_i c_i e_i \in W$ in V, that means:*

$$\sum_{i=1}^{m} c_i e_i \in W \;\Rightarrow\; c_1 = \cdots = c_m = 0 \qquad (c_i \in \mathbb{F})$$

2. **Span V (mod W)** *if $\mathbb{F}$-span$\{\overline{e}_i\} = V/W$, which means: given $v \in V$, there are $c_i \in \mathbb{F}$ such that $\left(v - \sum_i c_i e_i\right) \in W$, or $\overline{v} = \sum_{i=0} c_i \overline{e}_i$ in V/W.*

3. *A* **basis for V (mod W)** *if the images $\{\overline{e}_i\}$ are a basis in V/W, which happens if and only if* 1. *and* 2. *hold.*

3.6. Exercise. Let $W \subseteq \mathbb{R}^5$ be the solution set of system

$$\begin{cases} x_1 + x_3 &= 0 \\ x_1 - x_4 &= 0 \end{cases}$$

and let $\{e_i\}$ be the standard basis in $V = \mathbb{R}^5$.

1. Find vectors v_1, v_2 that are a basis for V (mod W).

2. Is $\mathfrak{X} = \{e_1, e_2, e_3, v_1, v_2\}$ a basis for V where v_1, v_2 are the vectors in (1.)?

3. Find a basis $\{f_1, f_2, f_3\}$ for the subspace W. $\square$

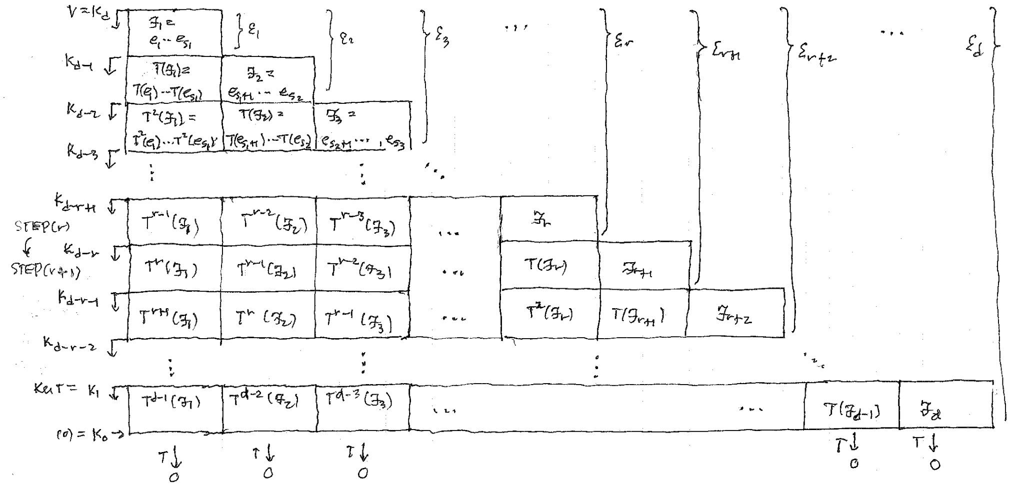

Figure 7.1. Steps in the construction of a basis that decomposes vector space V into cyclic subspaces under the action of a nilpotent linear operator $T : V \to V$. The subspaces K_i are the kernels of the powers T^i for $1 \le i \le d = \deg(T)$, with $K_d = V$ and $K_0 = (0)$.

3.7. Exercise. Let $T : V \to V$ be an arbitrary linear map and W a T-invariant subspace. Independence of vectors $f_1, \cdots, f_r$ mod a T-invariant subspace $W \subseteq V$ implies the independence (mod W') for any smaller T-invariant subspace $W' \subseteq W \subseteq V$. □

Proof of Theorem 3.2. Below we will construct two related sets of vectors $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \cdots$ and $\mathcal{E}_1 = \mathcal{F}_1 \subseteq \mathcal{E}_2 \subseteq \mathcal{E}_3 \subseteq \cdots \subseteq \mathcal{E}_r$ such that $\mathcal{E}_r$ is a basis for V aligned with the kernels $K_d = V \supseteq K_{d-1} \supseteq \cdots \supseteq K_1 = \ker(T) \supseteq \{0\}$. When the construction terminates, the vectors in $\mathcal{E}_r$ will be a basis for all of V that provides the desired decomposition into cyclic subspaces.

(Initial) Step1: Let $\mathcal{F}_1 = \mathcal{E}_1 = \{e_i : i \in \text{ index set } I_1\}$ be any set of vectors in $V \sim K_{d-1}$ that are a basis for V (mod K_{d-1}), so their images $\{\overline{e}_i\}$ are a basis in V/K_{d-1}. Obviously the index set I_1 has cardinality $|I_1| = |V/K_{d-1}| = |V| - |K_{d-1}|$, the dimension of the quotient space.

You might feel more comfortable indicating the index sets $I_1, I_2, \cdots$ being constructed here as consecutive blocks of integers, say $I_1 = \{1, 2, \cdots, s_1\}$, $I_2 = \{s_1 + 1, \cdots, s_2\}$ etc, but this notation becomes really cumbersome after the first two steps. And in fact there is no need to explicitly name the indices in each block. From here on you should refer to the chart shown in Figure 7.1, which lists all the players that will emerge in our discussion.

Step 2: The T-images $T(\mathcal{F}_1)$ lie in the layer $T(V \sim K_{d-1}) \subseteq K_{d-1} \sim K_{d-2}$, as noted in (2). In this step we shall verify two assertions.

> Claim (i): The vectors in $T(\mathcal{F}_1) = \{T(e_i) : i \in I_1\} \subseteq K_{d-1} \sim K_{d-2}$ are independent (mod K_{d-2}).

If these vectors are not already representatives of a basis for K_{d-1}/K_{d-2} we can adjoin additional vectors $\mathcal{F}_2 = \{e_i : i \in I_2\} \subseteq K_{d-1} \sim K_{d-2}$ chosen so that $T(\mathcal{F}_1) \cup \mathcal{F}_2$ corresponds to a basis for K_{d-1}/K_{d-2}; otherwise we take $\mathcal{F}_2 = \emptyset$.

> Claim (ii): The vectors $\mathcal{E}_2 = \mathcal{F}_2 \cup \big[\mathcal{E}_1 \cup T(\mathcal{F}_1)\big] = \mathcal{E}_1 \cup \big[T(\mathcal{F}_1) \cup \mathcal{F}_2\big]$ are a basis for all of V (mod K_{d-2}).

Remarks: In Linear Algebra I we saw that if $W \subseteq V$ and $\{e_1, \cdots, e_r\}$ is a basis for W, we can adjoin successive "outside vectors" $e_{r+1}, \cdots, e_s$ to get a basis for V. (These can even be found by deleting some of the vectors in a pre-ordained basis in V.) Then the images $\{\overline{e}_{r+1}, \cdots, \overline{e}_s\}$ are a basis for the quotient space V/W. That is how we proved the dimension formula $|V| = |W| + |V/W|$ for finite dimensional V.] $\square$

Proof: Claim (i). If $\sum_{i\in I_1} a_i T(e_i) = T(\sum_{i\in I_1} a_i e_i) \equiv 0 \pmod{K_{d-2}}$ then $\sum_{i\in I_1} a_i e_i$ is in K_{d-2} and also lies in the larger space $K_{d-1} \supseteq K_{d-2}$. But by definition vectors in $\mathcal{F}_1 = \{e_i : i \in I_1\}$ are independent (mod K_{d-1}), so we must have $a_i = 0$ for $i \in I_1$, proving independence (mod K_{d-1}) of the vectors in $T(\mathcal{F}_1)$.

Proof: Claim (ii). Suppose there exist coefficients $a_i^{(1)}, a_i^{(2)}, b_i \in \mathbb{F}$ such that

$$\sum_{i\in I_1} a_i^{(1)} e_i + \sum_{i\in I_2} a_i^{(2)} e_i + \sum_{i\in I_1} b_i T(e_i) \equiv 0 \ \ (mod\ K_{d-2}), \tag{3}$$

This sum lies in K_{d-2}, hence also in the larger subspace K_{d-1}, and the last two terms are already in K_{d-1} because $\mathcal{F}_2 \cup T(\mathcal{F}_1) \subseteq K_{d-1} \sim K_{d-2}$. Thus

$$\sum_{i\in I_1} a_i^{(1)} e_i \equiv 0 \ \ (mod\ K_{d-1})\ ,$$

and since the e_i, $i \in I_1$, are independent (mod K_{d-1}) we must have $a_i^{(1)} = 0$ for all $i \in I_1$. Now the sum (3) reduces to its last two terms, which all lie in K_{d-1}. But by construction, $\mathcal{F}_2 \cup T(\mathcal{F}_1)$ is a basis for K_{d-1} (mod K_{d-2}), which implies $a_i^{(2)} = 0$ for $i \in I_2$ and $b_i = 0$ for $i \in I_1$. Thus $\mathcal{E}_2 = \mathcal{F}_1 \cup [T(\mathcal{F}_1) \cup \mathcal{F}_2]$ is an independent set of vectors (mod K_{d-2}).

It remains to show $\mathcal{E}_2$ spans V (mod K_{d-2}). If $v \in V$ is not contained in K_{d-1} there is some $v_1 \in \mathbb{F}\text{-span}\{\mathcal{F}_1\}$ such that $v - v_1 \equiv 0 \pmod{K_{d-1}}$, so $v - v_1 \in K_{d-1}$. If this difference is lies outside of K_{d-2} we can find some $v_2 \in T(\mathcal{F}_1) \cup \mathcal{F}_2$ such that $v = (v_1 + v_2) \in K_{d-2}$. Thus $v = v_1 + v_2 \pmod{K_{d-2}}$, and since $v_1 + v_2 \in \mathbb{F}\text{-span}\{\mathcal{F}_1 \cup T(\mathcal{F}_1) \cup \mathcal{F}_2\}$, statement (ii) is proved. $\square$

That completes Step 2. Further inductive steps fill in successive rows in Figure 7.1. They involve no new ideas, but things can get out of hand unless the notation is carefully managed. Below we include a complete discussion of the general inductive step in this process, which could be skipped on first reading. It is followed by a final paragraph proving uniqueness of the multiplicities m_i (which you should read).

The General Inductive Step in Proving Theorem 3.2. This should probably be read with the chart from Figure 7.1 in hand to keep track of the players.

Continuing the recursive construction of basis vectors: at step r we have defined sets of vectors $\mathcal{F}_i \subseteq K_{d-i+1} \sim K_{d-i}$ for $1 \le i \le r$ with the properties $\mathcal{E}_1 = \mathcal{F}_1$ and

$$\mathcal{E}_r = \mathcal{E}_{r-1} \cup \left[T^{r-1}(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_{r-1}) \cup \mathcal{F}_r\right]$$

is a basis for V/K_{d-r}. At the next step we take the new vectors

$$T^{r-1}(\mathcal{F}_1) \cup T^{r-2}(\mathcal{F}_2) \cup \cdots \cup \mathcal{F}_r \subseteq K_{d-r+1} \sim K_{d-r}$$

created in the previous step and form their T-images

$$T^r(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_r) \subseteq K_{d-r} \sim K_{d-r-1}$$

To complete the inductive step we show:

1. These vectors are independent (mod K_{d-r-1})

2. We then adjoin additional vectors $\mathcal{F}_{r+1} \subseteq K_{d-r} \sim K_{d-r-1}$ as needed to produce a basis for K_{d-r}/K_{d-r-1}, taking $\mathcal{F}_{r+1} = \emptyset$ if the vectors $T^r(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_r)$ are already representatives for a basis in K_{d-r}/K_{d-r-1}. The vectors

$$\mathcal{E}_{r+1} = \mathcal{E}_r \cup \big[T^r(\mathcal{F}_1) \cup \ldots \cup T(\mathcal{F}_r) \cup \mathcal{F}_{r+1}\big]$$

will then be a basis for V (mod K_{d-r-1}).

Proof details:

1. If the vectors $T^r(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_r)$ are not representatives for an independent set of vectors in K_{d-r}/K_{d-r-1} there are coefficients $\{c_i^{(1)} : i \in I_1\}, \cdots, \{c_i^{(r)} : i \in I_r\}$ such that

$$\sum_{i \in I_1} c_i^{(1)} T^r(e_i) + \ldots + \sum_{i \in I_r} c_i^{(r)} T(e_i) \equiv 0 \quad (mod\ K_{d-r-1})$$

So, this sum is in K_{d-r-1} and in K_{d-r}. But $T^{r-1}\{e_i : i \in I_1\} \cup \cdots \cup \{e_i : i \in I_r\}$ are independent vectors (mod K_{d-r}) by hypothesis, and are a basis for K_{d-r+1}/K_{d-r}. We may rewrite the last congruence as

$$T\Big[\sum_{i \in I_1} c_i^{(1)} T^{r-1}(e_i) + \ldots + \sum_{i \in I_r} c_i^{(r)} e_i\Big] \equiv 0 \quad (\text{mod } K_{d-r-1})$$

So, $T[\cdots] \in K_{d-r-1}$, hence $[\cdots] \in K_{d-r}$ too. By independence of the e_i (mod K_{d-r}), we must have $c_i^{(j)} = 0$ in $\mathbb{F}$ for all i, j. Thus the vectors $T^r(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_r)$ are independent (mod K_{d-r-1}) as claimed.

2. To verify independence of the updated set of vectors

$$\mathcal{E}_{r+1} = \mathcal{E}_r \cup \big[T^r(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_r) \cup \mathcal{F}_{r+1}\big]$$

in V/K_{d-r-1}, suppose some linear combination $S = S' + S''$ is zero (mod K_{d-r-1}) where S' is a sum over vectors in $\mathcal{E}_r$ and S'' a sum over vectors in $T^r(\mathcal{F}_1) \cup \cdots \cup \mathcal{F}_{r+1}$. Then $S \equiv 0$ (mod K_{d-r-1}) implies $S \equiv 0$ (mod K_{d-r}), and then by independence of vectors in $\mathcal{E}_r$ (mod K_{d-r}), all coefficients in S' are zero. The remaining term S'' in the reduced sum lies in $K_{d-r} \sim K_{d-r-1}$, and by independence of $T^r(\mathcal{F}_1) \cup \cdots \cup \mathcal{F}_{r+1}$ in K_{d-r}/K_{d-r-1} all coefficients in S'' are also zero. Thus $\mathcal{E}_{r+1} \subseteq V$ corresponds to an independent set in K_{d-r}/K_{d-r-1}.

Dimension counting reveals that

$$\begin{aligned}
|V/K_{d-1}| &= |\mathcal{F}_1| \\
|K_{d-1}/K_{d-2}| &= |T(\mathcal{F}_1)| + |\mathcal{F}_2| \ = |\mathcal{F}_1| + |\mathcal{F}_2| \\
&\ \vdots \\
|K_{d-r}/K_{d-r-1}| &= |\mathcal{F}_1| + \ldots + |\mathcal{F}_{r+1}|
\end{aligned} \tag{4}$$

Thus $|V/K_{d-r-1}| = |V/K_{d-1}| + \cdots + |K_{d-r}/K_{d-r-1}|$ is precisely the number $|\mathcal{E}_{r+1}|$ of basis vectors appearing in the first $r+1$ rows from the top of the chart in Figure 7.1). But this is also equal to $\dim(V/V_{d-r-1})$, so $\mathcal{E}_{r+1}$ is a basis for V/V_{d-r-1} and Step(r+1) of the induction is complete.

The Cyclic Subspace Decomposition. A direct sum decomposition of V into cyclic subspaces can now be read out of Figure 7.1, in which basis vectors have been constructed row-by-row. Consider what happens when we partition into columns. For each $e_i \in \mathcal{F}_1$, $(i \in I_1)$, we have $e_i, T(e_i), T^2(e_i), \cdots, T^{d-1}(e_i) \neq 0$ and $T^d(e_i) = 0$, so these vectors span a cyclic subspace $E(e_i)$ such that $T|_{E(e_i)}$ has nilpotent degree d with e_i as its cyclic vector. Since the vectors that span $E(e_i)$ are part of a basis $\mathcal{E}_d$ for all of V, we obtain a direct sum of cyclic T-invariant subspaces $\bigoplus_{i\in I_1} E(e_i) \subseteq V$ ($|I_1| = |\mathcal{F}_1|$ subspaces).

Vectors $e_i \in \mathcal{F}_2$ $(i \in I_2)$ generate cyclic subspaces $E(e_i)$ such that $\dim\big(E(e_i)\big) = \deg(T|_{E(e_i)}) = d-1$; these become part of

$$\bigoplus_{i\in I_1} E(e_i) \oplus \bigoplus_{i_2\in I_2} E(e_2) \ ,$$

etc. At the last step, the vectors $e_i \in \mathcal{F}_d$ $(i \in I_d)$ determine T-invariant one-dimensional cyclic spaces such that $T(\mathbb{F}e_i) = (0)$, with nilpotence degree $= 1$ – i.e. the spaces $E(e_i) = \mathbb{F}e_i$ all lie within $\ker(T)$. The end result is a cyclic subspace decomposition

$$\Big(\bigoplus_{i_1\in I_1} E(e_{i_1})\Big) \oplus \Big(\bigoplus_{i_2\in I_2} E(e_{i_2})\Big) \oplus \ldots \oplus \Big(\bigoplus_{i_d\in I_d} E(e_{i_d})\Big) \tag{5}$$

of the entire space V, since all basis vectors in $\mathcal{E}_r$ are accounted for. (Various summands in (5) may of course be trivial.)

Uniqueness: A direct sum decomposition $V = \bigoplus_{j=1}^{s} E_j$ into T-invariant cyclic subspaces can be refined by gathering together those E_i of the same dimension, writing

$$V = \bigoplus_{k=1}^{d} \mathcal{H}_k \quad \text{where} \quad \mathcal{H}_k = \bigoplus \{E_i : \dim(E_i) = \deg(T|E_i) = k\}$$

for $1 \leq k \leq d = \deg(T)$.

3.8. Proposition. *In any direct sum decomposition $V = \bigoplus_{j=1}^{s} E_j$ into cyclic T-invariant subspaces, the number of spaces of dimension $\dim(E_i) = k$, $1 \leq k \leq d = \deg(T)$ can be computed in terms of the dimensions of the quotients K_i/K_{i-1}. These numbers are the same for all cyclic decompositions.*

Proof: Let us regard Figure 7.1 as a $d \times d$ array of "cells" with C_{ij} the cell in Row(i) (from the top) and Col(j) (from the left) in the array; the "size" $|C_{ij}|$ of a cell is the number of basis vectors it contains. Note that

(*i*) $|C_{ij}| = 0$ if the cell lies above the diagonal, with $j > i$, because those cells are empty (others may be empty too).

(*ii*) $|C_{ij}| = |\mathcal{F}_j|$ for all cells on and below the diagonal in Col(j) of the array. In particular $|C_{j1}| = |\mathcal{F}_1|$ for all nonempty cells in Col(1), $|C_{j2}| = |\mathcal{F}_2|$ for those in Col(2), etc.

By our construction, it is evident that vectors in the nonempty cells in Row(r) of Figure 7.1 correspond to a basis for the quotient space K_{d-r}/K_{d-r-1}. Counting the total number of basis vectors in Row(r) we find that

$$\dim(K_{d-r}/K_{d-r-1}) = |C_{r1}| + \ldots + |C_{r+1,r+1}| = |\mathcal{F}_1| + \ldots + |\mathcal{F}_{r+1}| \ ,$$

We may now recursively compute the values of $|C_{rj}|$ and $|\mathcal{F}_j|$ from the dimensions of the quotent spaces K_i/K_{i-1}. But as noted above, each $e_i \in \mathcal{F}_k$ lies in the diagonal cell C_{kk} and generates a distinct cyclic space in the decomposition. $\square$

That completes the proof of Theorem 3.2.

Remarks. To summarize,

1. We define $K_i = \ker(T^i)$ for $1 \le d = i$ nilpotence degree of T.
2. The following relations hold.

$$\begin{array}{rcl}
\mathcal{E}_1 &=& \mathcal{F}_1 \subseteq V \sim K_{d-1} \text{ determines a basis for } V/K_{d-1}, \\
\mathcal{E}_2 &=& \mathcal{E}_1 \cup \big[\, T(\mathcal{F}_1) \cup \mathcal{F}_2 \,\big] \subseteq V \sim K_{d-2} \text{ determines a basis for } V/K_{d-2}, \\
&\vdots& \\
\mathcal{E}_{r+1} &=& \mathcal{E}_r \cup \big[\, T^r(\mathcal{F}_1) \cup T^{r-1}(\mathcal{F}_2) \cup \cdots \cup \mathcal{F}_{r+1} \,\big] \subseteq V \sim K_{d-r} \text{ determines a basis for } V/K_{d-r-1} \\
&\vdots& \\
\mathcal{E}_d &=& \mathcal{E}_{d-1} \cup \big[\, T^{d-1}(\mathcal{F}_1) \cup \cdots \cup T(\mathcal{F}_{d-1}) \cup \mathcal{F}_d \,\big] \text{ is a basis for all of } V. \ \square
\end{array}$$

In working examples it usually helps to start by determining a basis $\mathcal{B}^{(0)} = \mathcal{B}^{(1)} \cup \ldots \cup \mathcal{B}^{(d)}$ for V aligned with the kernels so that $\mathcal{B}^{(1)}$ is a basis for K_1, $\mathcal{B}^{(2)}$ determines a basis for K_2/K_1, etc. This yields a convenient basis in V to start the construction.

3.9. Example. Let $V = \mathbb{F}^5$ and $T : V \to V$ the operator $T = L_A$,

$$T(x_1, \cdots, x_5) = (0, x_3 + x_4, 0, x_3, x_1 + x_4)$$

whose matrix with respect to the standard basis $\mathfrak{X} = \{e_1, \cdots, e_5\}$ in $\mathbb{F}^5$ is

$$A = [T]_{\mathfrak{X}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Show that T is nilpotent, then determine $\deg(T)$ and the kernels

$$V = K_d \supseteq K_{d-1} \supseteq \cdots \supseteq K_1 \supseteq \{0\}$$

Find a basis $\mathfrak{Y}$ such that $[T]_{\mathfrak{Y}}$ has block diagonal form, with each block B_i an elementary nilpotent matrix. This is the *Jordan canonical form* for a nilpotent linear operator.

Discussion: First find bases for the kernels $K_i = \ker(T^i)$. We have

$$\begin{array}{rcl}
K_1 = \ker(T) &=& \{\mathbf{x} : x_3 + x_4 = 0,\ x_3 = 0,\ x_1 + x_4 = 0\} \\
&=& \{\mathbf{x} : x_4 = x_3 = 0, x_1 + x_4 = 0\} \ = \ \{x : x_1 = x_3 = x_4 = 0\} \\
&=& \{(0, x_2, 0, 0, x_5) : x_2, x_5 \in \mathbb{F}\} = \mathbb{F}\text{-span}\{e_2, e_5\}
\end{array}$$

Iteration of T yields

$$\begin{array}{rcl}
T(\mathbf{x}) &=& (0, x_3 + x_4, 0, x_3, x_1 + x_4) \\
T^2(\mathbf{x}) &=& T(T(\mathbf{x})) \ = \ (0, x_3, 0, 0, x_3) \\
T^3(\mathbf{x}) &=& (0, \cdots, 0)
\end{array}$$

for $\mathbf{x} \in \mathbb{F}^5$. Clearly T is nilpotent with $\deg(T) = 3$, and

$$\begin{array}{ll}
|K_1| = 2 : & K_1 = \mathbb{F}\text{-span}\{e_2, e_5\} = \{\mathbf{x} : x_1 = x_3 = x_4 = 0\} \\
|K_2| = 4 : & K_2 = \ker(T^2) = \{\mathbf{x} : x_3 = 0\} = \mathbb{F}\text{-span}\{e_1, e_2, e_4, e_5\} \\
|K_3| = 5 : & K_3 = \mathbb{F}^5
\end{array}$$

In this example, $\mathfrak{X} = \{e_2, e_5; e_1, e_4; e_2\} = \mathcal{B}^{(1)} \cup \mathcal{B}^{(2)} \cup \mathcal{B}^{(3)}$ is an ordered basis for V aligned with the K_i running through $(0) \subseteq K_1 \subseteq K_2 \subseteq K_3 = V$. From this we can determine the families $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ of Theorem 3.2.

STEP 1: Since $|K_3/K_2| = 1$ any nonzero vector in the layer $K_3 \sim K_2 = \{\mathbf{x} : x_3 \neq 0\}$ yields a basis vector for K_3/K_2. We shall take $\mathcal{F}_1 = \{e_3\}$ chosen from the standard basis $\mathfrak{X}$, and then $\mathcal{E}_1 = \{e_3\}$ too. (Any $\mathbf{x}$ with $x_3 \neq 0$ would also work.)

STEP 2: The image set $T(\mathcal{F}_1) = T(e_3) = e_2 + e_4$ lies in the next layer

$$\begin{aligned} K_2 \sim K_1 &= \{\mathbf{x} : x_3 = 0\} \sim \mathbb{F}\text{-span}\{e_2, e_5\} \\ &= \{\mathbf{x} : x_3 = 0 \text{ and } x_1, x_4 \text{ are not both } = 0\} \end{aligned}$$

Since $|T(\mathcal{F}_1)| = 1$ and $\dim(K_2/K_1) = |K_2| - |K_1| = 4 - 2 = 2$, we must adjoin one suitably chosen new vector $\mathbf{x}$ from layer $K_2 \sim K_1$ to $T(\mathcal{F}_1)$ to get the desired basis for K_2/K_1. Then $\mathcal{F}_2 = \{\mathbf{x}\}$ and

$$\mathcal{E}_2 = (\mathcal{F}_1 \cup T(\mathcal{F}_1)) \cup \mathcal{F}_2 = \{e_3, e_2 + e_4, \mathbf{x}\}$$

$\mathcal{E}_2$ is a basis for V/K_2 as in first inductive step of Theorem 3.2.

A suitable vector $\mathbf{x} = (x_1, \ldots, x_5)$ in $K_2 \sim K_1$, $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ must have $x_3 = 0$ (so $\mathbf{x} \in K_2$) and x_1, x_3, x_4 not all zero (so $x \notin K_1$). This holds if and only if $(x_3 = 0)$ and (x_1, x_4 are not both 0). But we must also insure that our choice of $\mathbf{x}$ makes $\{e_3, e_2 + e_4, \mathbf{x}\}$ independent (mod K_1). The following lemma is helpful.

3.10. Lemma. *Let $V = \mathbb{F}^n$, W a subspace, $\mathfrak{X} = \{v_1, \cdots, v_r\}$ vectors in V, and let $M = \mathbb{F}$-span$\{v_1, \cdots, v_r\}$ (so $r = |V/W|$). Let $\mathfrak{Y} = \{w_1, \cdots, w_{n-r}\}$ be a basis for W. Then the following assertion are equivalent.*

1. *$\mathfrak{X}$ determines a basis for V/W.*
2. *$\mathfrak{Y} \cup \mathfrak{X} = \{v_1, \cdots, v_r, w_1, \cdots, w_{n-r}\}$ is a basis for V.*
3. *$V = W \oplus M$ (direct sum of subspaces).*

Proof: In Linear Algebra I we showed that the images $\overline{v}_1, \cdots, \overline{v}_r$ are a basis for V/W if and only if $\{v_1, \cdots .v_r\} \cup \mathfrak{Y}$ are a basis for V. It is obvious that $(ii) \Leftrightarrow (iii)$. $\square$

3.11. Corollary. *In the setting of the lemma the "outside vectors" $v_1, \cdots, v_r \in V \sim W$ are a basis for V ((mod W), so the images $\{\overline{v}_1, \cdots, \overline{v}_r\}$ are a basis for V/W, if and only if the $n \times n$ matrix A whose rows are $R_1 = v_1, \cdots, R_r = v_r, R_{r+1} = w_1, \cdots, R_n = w_{n-r}$ has rank equal to n.*

Armed of this observation (and the known basis $\{e_2, e_5\}$ for K_1), we seek a vector $\mathbf{x} = (x_1, \ldots, x_5)$ with x_1, x_4 not all equal to 0, such that

$$A = \begin{pmatrix} e_3 \\ e_2 + e_4 \\ (x_1, x_2, 0, x_4, x_5) \\ e_2 \\ e_5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ x_1 & x_2 & 0 & x_4 & x_5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

has Rowrank(A) = 5. Symbolic row operations put this into the form

$$\begin{pmatrix} x_1 & x_2 & 0 & x_4 & x_5 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

which has $rank = 5$ if and only if $x_1 \neq 0$.

Thus we may take e_1 as the additional vector we seek, and then

$$\mathcal{F}_1 = \{e_3\} \quad T(\mathcal{F}_1) = \{e_2 + e_4\} \quad \mathcal{F}_3 = \{e_1\} \ ,$$

and $\mathcal{E}_2 = \big[\mathcal{F}_1 \cup T(\mathcal{F}_1)\big] \cup \mathcal{F}_2$. That completes Step 2. (Actually any $\mathbf{x}$ with $x_1 \neq 0$, $x_3 = 0$ would work.)

Step 3: In the next layer $K_1 \sim K_0$ we have the vectors

$$T^2(\mathcal{F}_1) = \{T^2(e_3) = T(e_2 + e_4) = e_2 + e_5\} \quad \text{and} \quad T(\mathcal{F}_2) = \{T(e_1)\} = \{e_5\}$$

Since, $|K_1/K_0| = |K_1| = 2$ there is no need to adjoin additional vectors from this layer, so $\mathcal{F}_3 = \emptyset$. The desired basis in V is

$$\mathcal{E}_3 = \mathcal{F}_1 \cup \big[\,T(\mathcal{F}_1) \cup \mathcal{F}_2\,\big] \cup \big[\,T^2(\mathcal{F}_1) \cup T(\mathcal{F}_2)\,\big] = \{e_3; e_2 + e_4, e_1; e_2 + e_5, e_5\}$$

The iterated action of T sends

$$e_3 \to T(e_3) = e_2 + e_4 \to T^2(e_3) = e_2 + e_5 \quad \text{and} \quad e_1 \to T(e_1) = e_5$$

The cyclic subspaces are

$$\begin{aligned} E_1 &= \mathbb{F}\text{ -span}\{e_3, T(e_3), T^2(e_3)\} \ = \{e_3, e_2 + e_4, e_2 + e_5\} \\ E_2 &= \mathbb{F}\text{-span}\{e_1, T(e_1) = e_5\} \end{aligned}$$

and $V = E_1 \oplus E_2$. With respect to this basis $[T]_{\mathfrak{X}}$ has the block diagonal form

$$[T_{\mathfrak{X}}] = \left(\begin{array}{ccc|cc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) ,$$

each diagonal block being an elementary nilpotent matrix. The number and size of such blocks are uniquely determined but the bases are not unique, nor are the cyclic subspaces in the splitting $V = E_1 \oplus E_2$. $\square$

3.12. Exercise. Let W be the 3-dimensional subspace in $V = \mathbb{F}^5$ determined by the equations

$$\left\{ \begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 3x_1 + 5x_3 &= 0 \end{aligned} \right.$$

which is equivalent to the matrix equation $A\mathbf{x} = 0$ with

$$A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 3 & 0 & 5 & -1 \end{pmatrix}$$

(a) Find vectors $\{v_1, v_2, v_3\}$ that are a basis for W.

(b) Find 2 vectors $\{v_4, v_5\}$ that form a basis for V (mod W).

(c) Find two of the standard basis vectors $\{e_1, e_2, e_3, e_4, e_5\}$ in $\mathbb{F}^5$ that are a basis for V (mod W).

3.13. Exercise. Do either of the vectors in

$$f_1 = 2e_1 - 3e_2 + e_3 + e_4 \qquad f_2 = -e_1 + 2e_2 + 5e_3 - 2e_4$$

in $\mathbb{F}^5$ lie in the subspace W determined by the system of the previous exercises? Do these vectors form a basis for $\mathbb{F}^5$ (mod W)? $\Box$

3.14. Exercise. Which of the following matrices A are nilpotent?

$$\text{(a)} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{(b)} \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{(c)} \begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \\ -1 & -2 & 1 \end{pmatrix} \quad \text{(d)} \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & 4 \end{pmatrix}$$

If A is nilpotent, find a basis for $\mathbb{F}^3$ that puts A into block diagonal form with elementary nilpotent blocks. What is the resulting block diagonal form if the blocks are listed in order of decreasing size? $\Box$

3.15. Exercise. If N_1, N_2 are nilpotent is $N_1 N_2$ nilpotent? What if N_1 and N_2 commute? $\Box$

3.16. Corollary. If N_1, N_2 are nilpotent operators $N_k : V \to V$ and their commutator $[N_1, N_2] = N_1 N_2 - N_2 N_1$ is $= 0$.

(a) Prove that linear combination $c_1 N_1 + c_2 N_2$ are also nilpotent.

(b) If $N_1, \cdots, N_r$ are nilpotent and commute pairwise, so $[N_i, N_j] = 0$ for $i \neq j$, prove that all operators in $\mathbb{F}$-span$\{N_1, \cdots, N_r\}$ are nilpotent. $\Box$

3.17. Exercise. Let $V = \mathcal{P}_n(\mathbb{F})$ be the space of polynomials $f = \sum_{i=0}^n c_i x^i \in \mathbb{F}[x]$ of degree $\leq n$.

(a) Show that the differentiation operator

$$D : V \to V, Df = df/dx = c_1 + 2c_2\, x + \cdots + n \cdot c_n x^{n-1}$$

is nilpotent with $\deg(D) = n + 1$ (Note: $\dim(V) = n + 1$).

(b) Prove that any constant coefficient differential operator $L : V \to V$ of the form $a_1 D + a_2 D^2 + \cdots + a_n D^n$ (*no constant term* $a_0 I$) is nilpotent on V.

(c) Does this remain true if a nonzero constant term $c_0 I$ is allowed? $\Box$

3.18. Exercise. In the space of polynomials $\mathcal{P}_n(\mathbb{R})$ consider the subspaces

$$\begin{aligned} V_1 &= \{f : f(x) = f(-x), \text{ the even polynomials}\} \\ V_2 &= \{f : f(-x) = -f(x), \text{ the odd polynomials}\} \end{aligned}$$

Prove that these subspaces are invariant under differentiation, and that $\mathcal{P}_n$ is their direct sum $V_1 \oplus V_2$. $\Box$

3.19. Exercise. Show $\mathrm{Tr}(A) = 0$, for any nilpotent linear operator $A : V \to V$ of a finite dimensional space. Is the converse true? $\Box$

VII.4 A Review of the Diagonalization Problem.

We will give a general structure theorem for linear operators T over a field $\mathbb{F}$ large enough that the characteristic polynomials $p_T = \det(T - xI)$ splits into linear factors $f(x) = c \cdot \prod_{i=1}^s (x - a_i)^{m_i}$ in $\mathbb{F}[x]$. This is always true if $\mathbb{F} = \mathbb{C}$, but p_T need not split over other fields; and even if $p_T(x)$ does split, that alone is not enough to guarantee T is diagonalizable. In this section we briefly review diagonalizability of linear operators over a general field $\mathbb{F}$, which means that there is a basis of eigenvectors in V (or equivalently that the eigenspaces $E_\lambda(T)$ span V so $V = \sum_\lambda E_\lambda(T)$). If you already have a good understanding of these matters you may want to skip to Section VII.5 where we discuss the *generalized eigenspaces* that lead to the Jordan Decomposition. However, you should at least read the next theorem and its proof since the techniques used are the basis for the more complicated proof that generalized eigenspaces are independent, part of a direct sum decomposition of V.

Diagonalization.

4.1. Definition. *Let $T : V \to V$ be a linear operator on vector space over $\mathbb{F}$. If $\lambda \in \mathbb{F}$, the* **λ-eigenspace** *is $E_\lambda = \{v \in V : (T - \lambda I)v = 0\}$. Then λ is an* **eigenvalue** *if $E_\lambda(T) \neq \{0\}$ and $\dim_{\mathbb{F}}\big(E_\lambda(T)\big)$ is its* **geometric multiplicity**. *We often refer to $\mathrm{sp}_{\mathbb{F}}(T) = \{\lambda \in \mathbb{F} : E_\lambda \neq \{0\}\}$ as the* **spectrum** *of T over $\mathbb{F}$.*

4.2. Exercise. Show that every eigenspace E_λ is a vector subspace in V that is T-invariant. If $\mathfrak{X} = \{e_1, \cdots, e_r, \cdots, e_n\}$ is a basis for V that first passes through E_λ, show that the matrix of T takes the form

$$[T]_{\mathfrak{X}} = \left(\begin{array}{ccc|cc} \lambda & 0 & 0 & * & * \\ \cdot & \cdot & \cdot & * & * \\ 0 & 0 & \lambda & * & * \\ \hline 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \end{array}\right)$$

The geometric multiplicity of λ is $\dim(E_\lambda(T))$. We have already seen that when $\mathbb{F} = \mathbb{R}$ the operator $T = (90^\circ$ rotation acting on $\mathbb{R}^2)$ has no eigenvalues in $\mathbb{R}$, so $\mathrm{sp}_{\mathbb{R}}(T) = \emptyset$.

An operator T is diagonalizable if there is a basis $\mathfrak{X} = \{e_1, \cdots, e_n\}$ consisting of eigenvectors e_i, so $T(e_i) = \lambda_i e_i$ with respect to this basis. Then $[T]_{\mathfrak{X}}$ has the diagonal form

$$[T]_{\mathfrak{X}} = \begin{pmatrix} \lambda_1 & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_n \end{pmatrix}$$

in which there may be repeats among the λ_i. Conversely, any basis such that $[T]_{\mathfrak{X}}$ takes this form consist entirely of eigenvectors for T. A more sophisticated choice of basis vectors puts $[T]_{\mathfrak{X}}$ into *block diagonal form*. First a simple observation:

4.3. Exercise. If $T : V \to V$ is a linear operator on a finite dimensional space, show that the following statements are equivalent.

(a) There is a basis in V consisting of eigenvectors.

(b) The eigenspaces for T span V, so that

$$V = \sum_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)\ .$$

Note: There actually is something to be proved here: (b) requires more care selecting basis vectors than (a). $\square$

So, if T is diagonalizable and $\{\lambda_1, \ldots, \lambda_r\}$ are its *distinct* eigenvalues in $\mathbb{F}$, may choose a basis of eigenvectors e_i that first runs through E_{λ_1}, then through E_{λ_2}, etc. It is obvious that this choice yields a "block disagonal" matrix

$$[T]_{\mathfrak{X}} = \begin{pmatrix} \lambda_1 \boxed{I_{m_1 \times m_1}} & & & 0 \\ 0 & \lambda_2 \boxed{I_{2 \times m_2}} & & \\ & & \ddots & \\ 0 & & & \lambda_r \boxed{I_{m_r \times m_r}} \end{pmatrix}$$

in which $m_i = \dim\big(E_{\lambda_i}(T)\big)$. $\square$

These observations do not quite yield the definitive characterization of diagonalizability.

(6) Diagonalizability Criterion. *A linear operator T on a finite dimensional space is diagonalizable over $\mathbb{F} \Leftrightarrow V$ is the* direct sum *of its distinct eigenspaces: $V = \bigoplus_{i=1}^r E_{\lambda_i}(T)$.*

The implication ($\Leftarrow$) is trivial, but in the reverse direction we have so far only shown that (diagonalizable) $\Rightarrow V$ is *spanned* by its eigenspaces, so $V = \sum_{i=1}^r E_{\lambda_i}(T)$ and every v has at least one decomposition $v = \sum_i v_i : wq$ with $v_i \in E_{\lambda_i}(T)$. In a direct sum $\bigoplus_i E_{\lambda_i}(T)$ the decomposition is *unique*, and in particular $0 = \sum_i v_i$ with $v_i \in E_{\lambda_i}(T) \Rightarrow$ each term $v_i = 0$.

4.4. Exercise. Finish the proof of the Diagonalizability Criterion (6). If $V = \sum_i E_{\lambda_i}(T)$ prove that every $v \in V$ has a unique decomposition $v = \sum_i v_i$ such that $v_i \in E_{\lambda_i}(T)$. $\square$

4.5. Proposition. If $\{\lambda_1, \cdots, \lambda_r\}$ are the distinct eigenvalues in $\mathbb{F}$ for a linear operator $T : V \to V$ on a finite dimensional vector space, and if the eigenspaces E_{λ_i} span V, then V is a direct sum $E_{\lambda_1} \bigoplus \cdots \bigoplus E_{\lambda_r}$. Furthermore,

1. $\dim(V) = \sum_{i=1}^r \dim(E_{\lambda_i}) = \sum_{i=1}^r$ (geometric multiplicity of λ_i)
2. T is diagonalizable over $\mathbb{F}$.

Proof: Since $V = \sum_i E_{\lambda_i}$ every vector in V has a decomposition $v = \sum_{i=1}^r v_i$ with $V_i \in E_{\lambda_i}(T)$, so we need only prove uniqueness of this decomposition, which in turn reduces to proving that the v_i are "independent" in the sense that

$$0 = v_1 + \cdot + v_r \text{ with } v_i \in E_{\lambda_i} \;\Rightarrow\; v_1 = \cdots = v_r = 0$$

Note that for $\mu, \lambda \in \mathbb{F}$, the linear operators $(T - \lambda I)$, $(T - \mu I)$ commute with each other, since I commutes with everybody. Now suppose $\sum_{i=1}^r v_i = 0$ with $T(v_i) = \lambda_i v_i$.

Fix an index i and apply the operator $S = \prod_{j \neq i}(T - \lambda_j I)$ to the sum. We get

$$0 = S(0) = S\big(\sum_{k=1}^r v_k\big) = \sum_{k=1}^r S(v_k) \tag{7}$$

But if $k \neq i$, we can write

$$S(v_k) = \prod_{\ell \neq i}(T - \lambda_\ell I)v_k = \big[\prod_{\ell \neq k,i}(T - \lambda_k I)\big] \cdot (T - \lambda_k)v_k = 0$$

Hence the sum (7) reduces to

$$0 = \sum_k S(v_k) = S(v_i) + 0 + \cdots + 0 = \prod_{\ell \neq i}(T - \lambda_\ell I)v_i$$

Observe that we may write $(T - \lambda_\ell) = i) = (T - \lambda_i) + (\lambda_i - \lambda_\ell)I$, for all ℓ, so this becomes

$$0 = \big[\prod_{\ell \neq i}(T - \lambda_i) + (\lambda_i - \lambda_\ell)I\big]\, v_i = 0 + \big[\prod_{\ell \neq i}(\lambda_i - \lambda_\ell)\big]\, v_i \tag{8}$$

(because $(T - \lambda_i)v_i = 0$). The constant $c = \prod_{\ell \neq i}(\lambda_i - \lambda_\ell)$ must be nonzero because $\lambda_\ell \neq \lambda_i$. Therefore (7) $\Rightarrow v_i = 0$. This works for every $1 \leq i \leq r$ so the v_i are independent, as required. $\square$

4.6. Exercise. Let V be finite dimensional, $\{\lambda_1, \cdots, \lambda_r\}$ the distinct eigenvalues in $\mathbb{F}$ for an $\mathbb{F}$-linear operator $T : V \to V$. Let $E = \sum_{i=1}^r E_{\lambda_i}$ be the span of the eigenspaces. $(E \subseteq V)$. Show that

(a) E is T-invariant .

(b) $T|_E$ is diagonalizable. $\square$

4.7. Exercise. Let $T : V \to V$ be an linear operator on a finite dimensional vector space over $\mathbb{F}$, with $n = \dim_{\mathbb{F}}(V)$. If T has n distinct eigenvalues, prove that

(a) $V = \bigoplus_{i=1}^{n} E_{\lambda_i}$,

(b) The eigenspace are all one-dimensional, and

(c) T is diagonalizable. $\square$

4.8. Exercise. If a basis $\mathfrak{X}$ for V passes through the successive eigenspaces $E_{\lambda_1}(T), \cdots, E_{\lambda_r}(T)$, and we then adjoin vectors outside of the subspace $E = \sum_{\lambda_i \in \mathrm{sp}(T)} E_{\lambda_i}(T)$ to get a basis for V, explain why the matrix of T has the form

$$[T]_{\mathfrak{X}} = \left(\begin{array}{ccc|cc} \boxed{\lambda_1 I_{m_1 \times m_1}} & 0 & 0 & * & * \\ \cdot & \cdot & \cdot & * & * \\ 0 & 0 & \boxed{\lambda_r I_{m_r \times m_r}} & * & * \\ \hline 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \end{array}\right)$$

where $m_i = \dim(E_{\lambda_i}(T))$. $\square$

4.9. Definition. *If $T : V \to V$ is linear operator on a finite dimensional vector space, every root $\alpha \in \mathbb{F}$ of the characteristic polynomial $p_T(x) = \det(T - x\,I)$ is an eigenvalue for T, so $p_T(x)$ is divisible (without remainder) by $(x - \alpha)$.Repeated division by $(x - \alpha)$ may be possible, and yields a factorization $p_T(x) = (x - \alpha)^{m_\alpha} Q(x)$ where $Q \in \mathbb{F}[x]$ does not have α as a root, and thus is not divisible by $(x - \alpha)$. The exponent m_α is the* **algebraic multiplicity** *of the eigenvalue α.*

Now suppose $\mathbb{F}$ is an algebraically closed field (every nonconstant polynomial $f \in \mathbb{F}[x]$ has a root $\alpha \in \mathbb{F}$), for example $\mathbb{F} = \mathbb{C}$. It follows that every f over such a field splits completely into linear factors $f = c \cdot \prod_{i=1}(x - \alpha_i)$ where $\alpha_1, \cdots, \alpha_n$ are the roots of $f(x)$ in $\mathbb{F}$ (repeats allowed). If $T : V \to V$ is a linear operator on a finite dimensional vector space over such a field, and $\lambda_1, \ldots, \lambda_r$ are its *distinct* eigenvalues in $\mathbb{F}$, the characteristic polynomial splits completely

$$p_T(x) = \det(T - xI) = c \cdot \prod_{j=1}^{r} (x - \lambda_j)^{m_j}$$

where $m_j =$ the algebraic multiplicity of λ_j and $\sum_j m_j = \dim(V)$.

4.10. Corollary. *Let $T : V \to V$ be a linear operator on a finite dimensional space V over $\mathbb{F} = \mathbb{C}$. If the characteristic polynomial*

$$p_T(x) = \det(T - x\,I) = c \cdot \prod_{j=1}^{r} (x - \lambda_j)^{m_j}$$

has DISTINCT *roots (so $m_j = 1$ for all j), then $r = n = \dim_{\mathbb{C}}(V)$ and T is diagonalizable.*

Algebraic vs Geometric Multiplicity.

4.11. Proposition. *If $\lambda \in \mathbb{F}$ is an eigenvalue for linear operator $T : V \to V$, its algebraic multiplicity as a root of $p_T(x) = \det(T - x\,I)$ is $\geq$ (geometric multiplicity of λ) $=$*

$\dim E_\lambda$.

Proof: Fix an eigenvalue λ. Then $E = E_\lambda(T)$ is T-invariant and $T|_E = \lambda \cdot \mathrm{id}_E$. So, if we take a basis $\{e_1, \cdots, e_m\}$ in E_λ and then add vectors $e_{m+1}, \cdots, e_n$ to get a basis $\mathfrak{X}$ for V, we have

$$[T]_{\mathfrak{X}} = \left(\begin{array}{c|c} \lambda I_{m\times m} & * \\ \hline 0 & * \end{array}\right) \qquad \big(m = \dim(E_\lambda(T))\big)$$

Here m is the geometric multplicity of λ and the characteristic polynomial is

$$p_T(x) = \det \left(\begin{array}{ccc|ccc} (\lambda - x) & & 0 & & & \\ & \cdot & & & * & \\ 0 & & (\lambda - x) & & & \\ \hline & & & (a_{m+1,m+1} - x) & & * \\ & 0 & & & \cdot & \\ & & & * & & (a_{n,n} - x) \end{array}\right)$$

This determinant can be written as

$$\det(T - xI) = \sum_{\pi \in S_n} sgn(\pi) \cdot (T - xI)_{1,\pi(1)} \cdot \ldots \cdot (T - xI)_{n,\pi(n)} \tag{9}$$

Each term in this sum involves a product of matrix entries, one selected from each row. If the spots occupied by the selected entries in (9) are marked with a "$\square$," the marked spots provide a "template" for making the selection, and there is one template for each permutation $\pi \in S_n$: in Row(i), mark the entry in Col(j) with $j = \pi(i)$.

The only $n \times n$ templates that can contribute to the determinant of our block-upper triangular matrix $(T - xI)$ are those in which the first m diagonal spots have been marked (otherwise the corresponding product of terms will include a zero selected from the lower left block). The remaining marked spots must then be selected from the lower right block $(*)$ – i.e. from Row(i) and Col(j) with $m + 1 \leq i, j \leq n$, as indicated in the following diagram.

$$\left(\begin{array}{ccc|cccc} \square & & 0 & & & & \\ & \cdot & & & * & & \\ 0 & & \square & & & & \\ \hline & & & \cdot & \cdot & \square & \cdot \\ & 0 & & \square & \cdot & \cdot & \cdot \\ & & & \cdot & \square & \cdot & \cdot \end{array}\right)$$

Thus $p_T(x) = \det(T - xI)$ has the general form $(x - \lambda)^m \cdot G(x)$, in which the factor $G(x)$ might involve additional copies of λ. We conclude that

$$\big(\text{algebraic multiplicity of } \lambda\big) \geq m = \big(\text{ geometric multiplicity of } \lambda\big)\ ,$$

as claimed. $\square$

4.12. Example. Let $T = L_A : \mathbb{R}^3 \to \mathbb{R}^3$, with $A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$ If $\mathfrak{X} = \{e_1, e_2, e_3\}$ is the standard Euclidean basis then $[T]_{\mathfrak{X}} = A$ and the characteristic polynomial is

$$\begin{aligned}
p_T(x) &= \det(A - x\,I) = \det\begin{pmatrix} 4-x & 0 & 1 \\ 2 & 3-x & 2 \\ 1 & 0 & 4-x \end{pmatrix} \\
&= [(4-x)(3-x)(4-x) + 0 + 0\,] - [(3-x) + 0 + 0\,] \\
&= (12 - 7x + x^2)(4-x) - 3 + x \\
&= 48 - 28x + 4x^2 - 12x + 7x^2 - x^3 - 3 + x \\
&= -x^3 + 11x^2 - 39x + 45
\end{aligned}$$

To determine $\mathrm{sp}(T)$ we need to find roots of a cubic; however we can in this case guess a root λ and then long divide by $(x-\lambda)$. After a little trial and error it turns out that $\lambda = 3$ is a root, with $p_T(3) = -27 + 99 - 117 + 45 = 0$ and $p_T(x) = -(x-3)(x^2 - 8x + 15) = -(x-3)^2(x-5)$.

Eigenvalues in $\mathbb{F} = \mathbb{R}$ (or $\mathbb{F} = \mathbb{Q}$) are $\lambda_1 = 3, \lambda_2 = 5$ with algebraic multiplicities $m_1 = 2, m_2 = 1$. For the geometric multiplicities we must compute the eigenspaces $E_{\lambda_k}(T)$.

Case 1: $\lambda = 3$. We solve the system $(A - 3I)\mathbf{x} = 0$ by row reduction.

$$[A - 3I\,] = \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} \boxed{1} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Columns in the row reduced system that *do not* meet a "step corner" $\boxed{\ast}$ correspond to free variables in the solution; thus x_2, x_3 can take any value in $\mathbb{F}$ while $x_1 = -x_3$. Thus

$$\begin{aligned}
E_{\lambda=3} &= \ker(A - 3I) \;=\; \{(-v_3, v_2, v_3) : v_2, v_3 \in \mathbb{F}, v_1 = -v_3\} \\
&= \mathbb{F}\cdot(-1,0,1) \oplus \mathbb{F}\cdot(0,1,0) \;=\; \mathbb{F}(-e_1 + e_3) \oplus \mathbb{F}e_2
\end{aligned}$$

These vectors are a basis and $2 = \dim(E_{\lambda=3}) =$ (geometric multiplicity) $=$ (algebraic multplicity).

Case 2: $\lambda = 5$. Solving $(A - 5I)\mathbf{v} = 0$ by row reduction yields

$$[A{-}5I\,] = \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 2 & -2 & 2 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 0 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} \boxed{1} & 0 & -1 & 0 \\ 0 & \boxed{1} & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Now there is only one free variable x_3, with $x_2 = 2x_3$ and $x_1 = x_3$. Thus

$$E_{\lambda=5} = \{(x_3, 2x_3, x_3) : x_3 \in \mathbb{F}\} \;=\; \mathbb{F}\cdot(1,2,1)$$

and

$$1 = \dim(E_{\lambda=5}) = (\text{geometric multiplicity of } \lambda = 5) = (\text{algebraic multiplicity}).$$

Diagonalization: A basis $\mathfrak{Y}$ consisting of eigenvectors is given by

$$f_1 = -e_1 + e_3 \qquad f_2 = e_2 \qquad f_3 = e_1 + 2e_2 + e_3\ .$$

and for this basis we have

$$[T]_{\mathfrak{Y}} = \left(\begin{array}{cc|c} 3 & 0 & 0 \\ 0 & 3 & 0 \\ \hline 0 & 0 & 5 \end{array}\right)$$

while $[T]_{\mathfrak{X}} = A$ with respect to the standard Euclidean basis $\{e_i\}$.

It is sometimes important to know the similarity transform $SAS^{-1} = [T]_{\mathfrak{Y}}$ that effects the transition between bases. The matrix S can be found by writing

$$\begin{aligned} [T]_{\mathfrak{Y}\mathfrak{Y}} &= [\mathrm{id} \circ T \circ \mathrm{id}\,]_{\mathfrak{Y}\mathfrak{Y}} \;=\; [\mathrm{id}\,]_{\mathfrak{Y}\mathfrak{X}} \cdot [T]_{\mathfrak{X}\mathfrak{X}} \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \\ &= [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot A \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \end{aligned}$$

Then $S = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}}$, with $S_{\mathfrak{X}\mathfrak{Y}} = S^{-1}$ because

$$[\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \cdot [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{X}} = I_{3\times 3} \qquad (3 \times 3 \text{ identity matrix})$$

(All this is discussed in Chapter II.4 of the *Linear Algebra I Notes*.)

The easiest matrix to determine is usually $S^{-1} = [\mathrm{id}\,]_{\mathfrak{X}\mathfrak{Y}}$ which can be written down immediately if we know how to write basis vectors in $\mathfrak{Y}$ in terms of those in the standard basis $\mathfrak{X}$ in $\mathbb{F}^3$). In the present example we have

$$\left\{ \begin{array}{lcl} id(f_1) &=& -e_1 + 0 \cdot e_2 + e_3 \\ id(f_2) &=& 0 + e_2 + 0 \\ id(f_3) &=& e_1 + 2e_2 + e_3 \end{array} \right. \qquad \Rightarrow \; S^{-1} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} \;=\; \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$$

It is useful to note that the matrix $[\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}}$ is just the transpose of the coefficient array in the system of vector identities that express the f_i in terms of the e_j.

We can now find the desired inverse $S = (S^{-1})^{-1}$ by row operations (or by Cramer's rule) to get

$$S = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ -1 & 1 & -1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

and then

$$SAS^{-1} \;=\; \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

as expected. That concludes our analysis of this example. $\square$

4.13. Exercise. Fill in the details needed to compute S.

4.14. Example. Let $A = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$ with $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then

$$A^2 = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad A^3 = 0\ ,$$

so with respect to the standard basis in $\mathbb{F}^2$ the matrix of the map $T = L_A : \mathbb{F}^2 \rightarrow \mathbb{F}^2$ is $[T]_{\mathfrak{X}} = A$ and the characteristic polynomial is:

$$\det(A - x\,I) = \det \begin{pmatrix} 2-x & 4 \\ -1 & -2-x \end{pmatrix} \;=\; -4 + x^2 + 4 = x^2$$

Thus, $\lambda = 0$ is a root (over $\mathbb{R}$ or $\mathbb{C}$) with (algbraic multiplicity) $= 2$, but the geometric multiplicity is $\dim\big(E_{\lambda=0}\big) = 1$. When we solve the system $(A - \lambda I)\mathbf{x} = A\mathbf{x} = 0$ taking $\lambda = 0$ to determine $E_{\lambda=0} = \ker(L_A)$, row reduction yields

$$\left(\begin{array}{cc|c} 2-\lambda & 4 & 0 \\ -1 & -2-\lambda & 0 \end{array} \right) = \left(\begin{array}{cc|c} 2 & 4 & 0 \\ -1 & -2 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} \boxed{1} & 2 & 0 \\ 0 & 0 & 0 \end{array} \right)$$

For this system x_2 is a free variable and $x_1 = -2x_2$, so

$$E_{\lambda=0} = \mathbb{F} \cdot (-2, 1) = \mathbb{F} \cdot (2e_1 - e_2) \ ,$$

and $\dim\big(E_{\lambda=0}\big) = 1 <$ (algebraic multiplicity) $= 2$. There are no other eigenvalues so the best we can do in trying to reduce T is to find a basis such that $[T]_{\mathfrak{Y}}$ has form

$$\left(\begin{array}{c|c} 0 & * \\ \hline 0 & * \end{array}\right)$$

by taking $\mathfrak{Y} = \{f_1, f_2\}$ where $f_1 = (2, 1) = 2e_1 + e_2$ and f_2 is any other vector independent of f_1.

However, T is a nilpotent operator (verify this), so we can do better with a slightly different basis $\mathfrak{Z}$ that puts A into the Jordan canonical form for nilpotent operators (as in Theorem 3.2). In the present example this is

$$[T]_{\mathfrak{Z}} = \begin{pmatrix} \boxed{0} & 1 \\ 0 & \boxed{0} \end{pmatrix} \qquad \text{(an elementary nilpotent matrix)}$$

with two 1×1 blocks of zeros on the diagonal. In fact, in the notation of Theorem 3.2 we have kernels $(0) \subseteq K_1 = \ker(T) \subseteq K_2 = \ker(T^2) = V$ with

$$K_1 = E_{\lambda=0} = \mathbb{F} \cdot f_1 \qquad \text{and} \qquad K_2 = \mathbb{F}^2$$

So, if f_2 is any vector transverse to $\ker(T)$, we have $T(f_2) \in \ker(T) = \mathbb{F} \cdot f_1$. But $T(f_2) \neq 0$ since $f_2 \notin K_1$, and by scaling f_2 appropriately we can make $T(f_2) = f_1$. Then $\mathfrak{Z} = \{f_1, f_2\}$ is a basis that puts $[T]_{\mathfrak{Z}}$ into the form shown above. $\square$

4.15. Exercise. *Repeat the analysis of the previous exercise for the matrix* $A = \begin{pmatrix} 4 & 4 \\ -1 & 0 \end{pmatrix}$. $\square$

That concludes our review of Diagonalization.

VII-5 Generalized Eigenspace Decomposition I.

The Fitting Decomposition (Proposition 1.5) is a first step in trying to decompose a linear operator $T : V \to V$ over an arbitrary field.

5.1. Proposition (Fitting Decomposition). *Given linear $T : V \to V$ on a finite dimensional vector space over any field, then $V = N \oplus S$ for T-invariant subspaces N, S such that $T|_S : S \to S$ is a bijection (invertible linear operator on S), and $T|_N : N \to N$ is nilpotent.*

The relevant subspaces are the "stable kernel" and "stable range" of T,

$$\begin{aligned} K_\infty &= \bigcup_{i=1}^{\infty} K_i, \quad \big(K_i = \ker(T^i) \quad \text{with } \{0\} \subsetneqq K_1 \subsetneqq \cdots \subsetneqq K_r = K_{r+1} = \cdots = K_\infty\big) \\ R_\infty &= \bigcap_{i=1}^{\infty} R_i, \quad \big(R_i = \text{range}(T^i) \quad \text{with } \{0\} \supsetneqq R_1 \supsetneqq \cdots \supsetneqq R_r = R_{r+1} = \cdots = R_\infty\big) \end{aligned}$$

(see Section VII-1). Obviously, $T = (T|_{R_\infty}) \oplus (T|_{K_\infty})$ which splits T into canonically defined nilpotent and invertible parts.

5.2. Exercise. Prove that the Fitting decomposition is unique: If $V = N \oplus S$, both T-invariant, such that $T|_N$ is nilpotent and $T|_S : S \to S$ invertible show that $N = K_\infty(T)$ and $S = R_\infty(T)$. $\square$

Given a linear operator $T : V \to V$ we may apply these remarks to the operators $(T - \lambda I)$ associated with eigenvalues λ in $\mathrm{sp}_{\mathbb{F}}(T)$. The eigenspace $E_\lambda(T) = \ker(T - \lambda I)$ is the first in an ascending chain of subspaces, shown below.

$$\{0\} \subsetneqq \ker(T-\lambda) = E_\lambda(T) \subsetneqq \ker(T-\lambda)^2 \subsetneqq \cdots \subsetneqq \ker(T-\lambda)^r = \cdots = K_\infty(T-\lambda)$$

5.3. Definition. *If $\lambda \in \mathbb{F}$ the "stable kernel" of $(T - \lambda I)$*

$$K_\infty(\lambda) = \bigcup_{m=1}^{\infty} \ker(T - \lambda I)^m$$

is called the **generalized λ-eigenspace**, *which we shall hereafter denote by $M_\lambda(T)$. Thus,*

$$\begin{aligned} M_\lambda(T) &= \{v \in V : (T-\lambda I)^k v = 0 \text{ for some } k \in \mathbb{N}\} \\ &\supseteq E_\lambda(T) = \{v : (T-\lambda I)v = 0\} \end{aligned} \tag{10}$$

We refer to any $\lambda \in \mathbb{F}$ such that $M_\lambda(T) \neq (0)$ as a **generalized eigenvalue** *for T. But note that $M_\lambda(T) \neq \{0\} \Leftrightarrow E_\lambda(T) \neq \{0\} \Leftrightarrow \det(T - \lambda I) = 0$, so these are just the usual eigenvalues of T in $\mathbb{F}$.*

Generalized eigenspaces have the following properties.

5.4. Lemma. *The spaces $M_\lambda(T)$ are T-invariant.*

Proof: T commutes with all the operators $(T - \lambda)^m$, which commute with each other. Thus, $v \in M_\lambda(T) \Rightarrow (T - \lambda I)^k v = 0$ for some $k \in \mathbb{N} \Rightarrow$

$$(T - \lambda I)^k T(v) = T(T - \lambda I)^k v = T(0) = 0$$

Hence $T(v) \in M_\lambda(T)$. $\square$

We now show that $T|_{M_\lambda}$ has a nice upper triangular form with respect to a suitably chosen basis in M_λ.

5.5. Proposition. *Every generalized eigenspace $M_\lambda(T)$, $\lambda \in \mathrm{sp}(T)$, has a basis $\mathfrak{X}$ such that the matrix of $T|_{M_\lambda(T)}$ has upper triangular form*

$$[T|_{M_\lambda}]_{\mathfrak{X}} = \begin{pmatrix} \lambda & & & * \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \lambda \end{pmatrix}$$

Proof: We already know that any nilpotent operator N on a finite dimensional vector space can be put into strictly upper triangular form by a suitable choice of basis.

$$[N]_{\mathfrak{X}} = \begin{pmatrix} 0 & & & * \\ & \cdot & & \\ & & \cdot & \\ 0 & & & 0 \end{pmatrix}$$

Now write

$$T|_{M_\lambda} = (T - \lambda I)|_{M_\lambda} + \lambda I|_{M_\lambda}$$

in which $V = M_\lambda$, $N = (T - \lambda I)|_{M_\lambda}$ and $I|_{M_\lambda}$ is the identity operator on M_λ. Since $[I|_{M_\lambda}]_{\mathfrak{X}} = I_{m\times m}$ for any basis, a basis that puts $(T-\lambda I)|_{M_\lambda}$ into strict upper triangular form automatically yields

$$[T|_{M_\lambda}]_{\mathfrak{X}} = \big[(T - \lambda I)|_{M_\lambda}\big]_{\mathfrak{X}} + \lambda I = \begin{pmatrix} \lambda & & * \\ & \cdot & \\ 0 & & \lambda \end{pmatrix} \quad \square$$

The most precise result of this sort is obtained using the cyclic subspace decomposition for nilpotent operators (Theorem 3.2) to guide our choice of basis. As a preliminary step we might pick a basis $\mathfrak{X}$ aligned with the kernels

$$(0) \subsetneq K_1 = \ker(T) \subsetneq K_2 = \ker(T^2) \subsetneq \ldots \subsetneq K_d = V$$

where $d = \deg(T)$. As we indicated earlier in Exercise 2.4, $[T]_{\mathfrak{X}}$ is then upper triangular with zero blocks $Z_i, 1 \leq i \leq d = \deg(T)$, on the diagonal. Applying this to a generalized eigenspace $M_\lambda(T)$, the matrix of the nilpotent operator $T - \lambda I$ becomes upper triangular with zero blocks on the diagonal. Writing $T = (T - \lambda I) + \lambda I$ as above we see that the matrix of T with respect to any basis $\mathfrak{X}$ running through successive kernels $K_i = \ker(T - \lambda I)^i$ must have the form

$$\begin{aligned}\big[T|_{M_\lambda}\big]_{\mathfrak{X}} &= \lambda \cdot I_{n\times n} + [\,T - \lambda I\,]_{\mathfrak{X}} \\ &= \begin{pmatrix} \lambda \cdot \boxed{I_{m_1\times m_1}} & & & * \\ & \lambda \cdot \boxed{I_{m_2\times m_2}} & & \\ & & \ddots & \\ 0 & & & \lambda \cdot \boxed{I_{m_r\times m_r}} \end{pmatrix} \end{aligned} \qquad (11)$$

with $m_i = \dim(K_i/K_{i-1}) = \dim(K_i) - \dim(K_{i-1})$ and $n = \dim(V) = \sum_i m_i$. The shape of the "block upper triangular form" (11) is completely determined by the dimensions of the kernels $K_i = K_i(T - \lambda I)$.

Note that (11) can be viewed as saying $T|_{M_\lambda} = \lambda I_\lambda + N_\lambda$ where $I_\lambda = \mathrm{id}_{M_\lambda}$, λI_λ is a scalar operator on M_λ, and $N_\lambda = (T - \lambda I)|_{M_\lambda}$ is a nilpotent operator whose matrix with respect to the basis $\mathfrak{X}$ is similar to the matrix in (11), but with $m_i \times m_i$ zero-blocks on the diagonal. The restriction $T|_{M_\lambda}$ has an "additive decomposition" $T|_{M_\lambda} =$ (*diagonal*) + (*nilpotent*) into commuting scalar and nilpotent parts,

$$T|_{M_\lambda} = \lambda \cdot I + N_\lambda = \begin{pmatrix} \lambda & & * \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda \end{pmatrix} + \begin{pmatrix} 0 & & * \\ & \cdot & \\ & & \cdot \\ 0 & & 0 \end{pmatrix}$$

Furthermore, the nilpotent part N_λ turns out to be a polynomial function of $\big(T|_{M_\lambda}\big)$, so both components of this decomposition also commute with $T|M_\lambda$. There is also a "multiplicative decomposition" $T|_{M_\lambda} =$ (*diagonal*) $\cdot$ (*unipotent*) $= (\lambda I) \cdot U_\lambda$ where U_λ is the unipotent operator $(I + N_\lambda)$; for the corresponding matrices we have

$$\begin{pmatrix} \lambda & & * \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda \end{pmatrix} \cdot \begin{pmatrix} 1 & & * \\ & \cdot & \\ & & \cdot \\ 0 & & 1 \end{pmatrix}$$

Note: The off-diagonal entries $(*)$ in N_λ and U_λ need not be the same in these two decompositions.

As we show below, this description of the way T acts on M_λ can be refined to provide much more information about the off-diagonal terms $(*)$, but we will also see that for many purposes the less explicit block upper triangular form (11) will suffice, and is easy to compute since we only need to determine the kernels K_i.

Now consider what happens if we take a basis $\mathfrak{Y}$ in M_λ corresponding to a cyclic subspace decomposition of the nilpotent operator

$$N_\lambda = (T - \lambda I)|_{M_\lambda} = (T|_{M_\lambda}) - \lambda I_\lambda \qquad (I_\lambda = I|_{M_\lambda})$$

Then $[\lambda I_\lambda]_{\mathfrak{Y}}$ is λ times the identity matrix (as it is for any basis in M_λ) while $[N_\lambda]_{\mathfrak{Y}}$ consists of diagonal blocks, each an elementary nilpotent matrix N_i.

$$[N_\lambda]_{\mathfrak{Y}} = [(T-\lambda I)|_{M_\lambda}]_{\mathfrak{Y}} = \begin{pmatrix} \boxed{N_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \boxed{N_r} \end{pmatrix}$$

and

$$N_i = \begin{pmatrix} 0 & 1 & & & 0 \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ 0 & & & & 0 \end{pmatrix}$$

of size $d_i \times d_i$, with N_i a 1×1 zero matrix when $d_i = 1$. This yields the "**Jordan block decomposition**" of $T|_{M_\lambda}$

$$(12) \qquad [T|_{M_\lambda}]_{\mathfrak{Y}} = \lambda [I|_{M_\lambda}]_{\mathfrak{Y}} + [N_\lambda]_{\mathfrak{Y}} = \begin{pmatrix} \boxed{T_1} & & & & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \boxed{T_m} & \\ 0 & & & & \lambda \cdot \boxed{I_{r\times r}} \end{pmatrix}$$

with $T_i = \lambda \cdot I_{d_i \times d_i} + (\textit{elementary nilpotent})$ when $d_i > 1$,

$$T_i = \begin{pmatrix} \lambda & 1 & & & 0 \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ 0 & & & & \lambda \end{pmatrix}$$

The last block in (12) is exceptional. The other T_i correspond to the restrictions $T|_{C_i(\lambda)}$ to cyclic subspaces of dimension $d_i > 1$ in a cyclic subspace decomposition

$$M_\lambda(T) = C_1(\lambda) \oplus \ldots \oplus C_m(\lambda)$$

of the generalized eigenspace. However, some of the cyclic subspaces might be one-dimensional, and any such $C_i(\lambda)$ is contained in the ordinary eigenspace $E_\lambda(T)$. If there are r such degenerate cyclic subspaces we may lump them together into a single subspace

$$E = \bigoplus \{C_i(\lambda) : \dim(C_i(\lambda)) = 1\} \subseteq E_\lambda(T) \subseteq M_\lambda(T)$$

such that $\dim(E) = s$ and $T|_E = \lambda \cdot I_E$. It should also be evident that

$$s + d_1 + \ldots + d_m = \dim\big(M_\lambda(T)\big)$$

This is the Jordan Canonical Form for the restriction $\big(T|_{M_\lambda}\big)$ of T to a single generalized eigenspace. If $M_\lambda(T) \neq 0$ the description (12) is valid for any ground field $\mathbb{F}$, since it is really a result about the nilpotent operator $(T-\lambda)_{M_\lambda}$. Keep in mind that the T-invariant subspaces in a cyclic subspace decomposition of M_λ (or of any nilpotent operator) are not unique, but the number of cyclic subspaces in any decomposition and their dimensions are uniques, and we get the same matrix form (12) for a suitably chosen basis.

VII-6. Generalized Eigenspace Decomposition of T.

So far we have only determined the structure of T restricted to a single generalized eigenspace $M_\lambda(T)$. Several obstacles must be surmounted to arrive at a similar structure for T on all of V.

- If the generalized eigenspaces $M_{\lambda_i}(T)$ fail to span V, knowing the behavior of T only on their span
$$M = \sum_{\lambda \in \mathrm{sp}(T)} M_\lambda(T)$$
leaves the gobal behavior of T beyond reach.

- It is equally important to prove (as we did in Proposition 4.5 for ordinary eigenspaces) that the span of the generalized eigenspaces is in fact a DIRECT sum,
$$M = \bigoplus_{\lambda \in \mathrm{sp}(T)} M_\lambda(T)$$
That means the actions of T on different M_λ are *independent* and can be examined separately, yielding a decomposition $T|_M = \bigoplus_{\lambda\in\mathrm{sp}(T)} \left(T|_{M_\lambda}\right)$.

Both issues will be resolved in our favor for operators $T : V \rightarrow V$ provided that the characteristic polynomial $p_T(x)$ splits into linear factors in $\mathbb{F}[x]$. This is always true if $\mathbb{F} = \mathbb{C}$; we are not so lucky for linear operators over $\mathbb{F} = \mathbb{R}$ or over a finite field such as $\mathbb{F} = \mathbb{Z}_p$. When this program succeeds the result is the Jordan Canonical Decomposition.

6.1. Theorem (Jordan Canonical Form). *If $T : V \rightarrow V$ is a linear operator on a finite dimensional space whose characteristic polynomial $p_T(x) = \det(T - x\,I)$ splits over $\mathbb{F}$, then V is a direct sum of its generalized eigenspaces*
$$V = \bigoplus_{\lambda \in \mathrm{sp}(T)} M_\lambda(T)\ ,$$
and since the $M_\lambda(T)$ are T-invariant we obtain a decomposition of T itself
$$T = \bigoplus_{\lambda \in \mathrm{sp}(T)} T|_{M_\lambda(T)} \tag{13}$$
into operators, each of which can be put into Jordan upper triangular form (12) *by choosing bases compatible with a decomposition of M_λ into T-invariant cyclic subspaces.*

Proof that the generalized eigenspaces are independent components in a direct sum follows the same lines as a similar result for ordinary eigenspaces (Proposition VII-4.5), but with more technical complications. Proof that they span V will require some new ideas based on the Fitting decomposition.

6.2. Proposition (Independence of the M_λ). *The span $M = \sum_{\lambda\in\mathrm{sp}(T)} M_\lambda(T)$ of the generalized eigenspaces* (*which may be a proper subspace in V*) *is always a direct sum, $M = \bigoplus_{\lambda\in\mathrm{sp}(T)} M_\lambda(T)$.*

Proof: Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues in $\mathbb{F}$. By definition of "direct sum" we must show the components M_λ are independent, so that
$$0 = v_1 + \cdots + v_r,\ \text{with } v_i \in M_{\lambda_i} \ \Rightarrow\ \text{each term } v_i \text{ is zero.} \tag{14}$$

Fix an index k. For for each $1 \le i \le r$, let $m_j = \deg(T - \lambda_j I)|_{M_{\lambda_j}}$. If $v_k = 0$ we're done; and if $v_k \neq 0$ let $m \le m_k$ be the smallest exponent such that $(T - \lambda_k I)^m v_k = 0$ and $(T - \lambda_k I)^{m-1} v_k \neq 0$. Then $w = (T - \lambda_k I)^{m-1} v_k$ is a nonzero eigenvector in E_{λ_k}.

Define

$$A = \prod_{i \neq k} (T - \lambda_i I)^{m_i} \cdot (T - \lambda_k)^{m-1}$$

which is keyed to the particular index λ_k as above. We then have

$$\begin{aligned}
0 &= A(0) = 0 + Av_k \qquad (\text{since } Av_i = 0 \text{ for } i \neq k) \\
&= \prod_{i \neq k} (T - \lambda_i)^{m_i} \big((T - \lambda_k)^{m-1} v_k \big) \;=\; \prod_{i \neq k} (T - \lambda_i)^{m_i} w \\
&= \prod_{i \neq k} \big[(T - \lambda_k) + (\lambda_k - \lambda_i) \big]^{m_i} w \qquad (\text{a familiar algebraic trick}) \\
&= \prod_{i \neq k} \sum_{s=0}^{m_i} \binom{m_i}{s} (T - \lambda_k)^{m_i - s} (\lambda_k - \lambda_i)^s w \qquad (\text{binomial expansion of } [\cdots]^{m_i})
\end{aligned}$$

All terms in the binomial sum are zero except when $s = m_i$, so we get

$$0 = \big[\prod_{i \neq k} (\lambda_k - \lambda_i)^{m_i} \big] \cdot w$$

The factor $[\cdots]$ is nonzero because the λ_i are the distinct eigenvalues of T in $\mathbb{F}$, so w must be zero. This is a contradiction because $w \neq 0$ by definition. We conclude that *every* term v_k in (14) is zero, so the span M is a direct sum of the M_λ. $\square$

Further Properties of Characteristic Polynomials. Before takling up the proof of Theorem 6.1 we digress to develop a few more facts about characteristic polynomials, in order to work out the relationship between $\mathrm{sp}(T)$ and $\mathrm{sp}\big(T|_{R_\infty}\big)$ where $R_\infty = R_\infty(T - \lambda_1 I)$.

6.3. Lemma. *If* $A \in \mathrm{M}(n, \mathbb{F})$ *has form* $A = \left(\begin{array}{c|c} B & * \\ \hline 0 & C \end{array} \right)$ *where* B *is* $r \times r$ *and* C *is* $(n-r) \times (n-r)$, *then* $\det(A) = \det(B) \cdot \det(C)$.

6.4. Corollary. *If* $A \in \mathrm{M}(n, \mathbb{F})$ *is upper triangular with values* $c_1, \ldots, c_n$ *on the diagonal, then* $\det(A) = \prod_{i=1}^{n} c_i$.

Proof (Lemma 6.3): Consider an $n \times n$ template corresponding to some $\sigma \in S_n$. If any of the marked spots in columns $C_1, \cdots, C_r$ occur in a row R_i with $r+1 \le i \le n$, then $a_{ij} = a_{i,\sigma(i)} = 0$ and so is the corresponding term in $\sum_{\sigma \in S_n}(\cdots)$. Thus all columns $C_j, 1 \le j \le r$, must be marked in rows $R_1, \ldots, R_r$ if the template is to yield a nonzero term in $\det(A)$. It follows immediately that all columns C_j with $r+1 \le j \le n$ must be marked in rows R_i with $r+1 \le i \le n$ if σ is to contribute to

$$\det(A) = \sum_{\sigma \in S_n} sgn(\sigma) \cdot \prod_{i=1}^{n} a_{i,\sigma(i)}$$

Therefore only permutations σ that leave invariant the blocks of indices $[1, r]$, $[r+1, n]$ can contribute. These σ are composites of permutations $\mu = \sigma|_{[1,r]} \in S_r$ and $\tau = \sigma|_{[r+1,n]} \in S_{n-r}$, with

$$\sigma(k) \;=\; \mu \times \tau(l) = \begin{cases} \mu(k) & \text{if} \quad 1 \le k \le r \\ \tau(k-r) & \text{if} \quad r+1 \le k \le n \end{cases}$$

Furthermore, we have $sgn(\sigma) = sgn(\mu \times \tau) = sgn(\mu) \cdot sgn(\tau)$ by definition of sgn, because μ, τ operate on disjoint subsets of indices in $[1, n]$.

In the matrix A we have

$$\begin{aligned} B_{k,\ell} &= A_{k,\ell} \quad \text{for } 1 \leq k, \ell \leq r \\ C_{k,\ell} &= A_{k+r,\ell+r} \quad \text{for } r+1 \leq k, \ell \leq n \end{aligned}$$

so we get

$$\begin{aligned} \det(A) &= \sum_{(\mu,\tau) \in S_r \times S_{n-r}} sgn(\mu \times \tau) \cdot \Big(\prod_{i=1}^{r} B_{i,\mu(i)} \Big) \cdot \Big(\prod_{j=1}^{n-r} C_{j,\tau(j)} \Big) \\ &= \Big(\sum_{\mu \in S_r} sgn(\mu) \cdot \prod_{i=1}^{r} B_{i,\mu(i)} \Big) \cdot \Big(\sum_{\tau \in S_{n-r}} sgn(\tau) \cdot \prod_{j=1}^{n-r} C_{j,\tau(j)} \Big) \\ &= \det(B) \cdot \det(C) \quad \square \end{aligned}$$

6.5. Corollary. *If $T : V \to V$ is a linear operator on a finite dimensional vector space and $M \subseteq V$ is a T-invariant subspace, the characteristic polynomial $p_{T|_M}(x)$ divides $p_T(x) = \det(T - x\,I)$ in $\mathbb{F}[x]$.*

Proof: If $M \subseteq V$ is T-invariant and we take a basis $\mathfrak{X} = \{e_i\}$ that first spans M and then picks up additional vectors to get a basis for V, the matrix $[T]_{\mathfrak{X}}$ has block upper triangular form $\left(\begin{array}{c|c} A & * \\ \hline 0 & B \end{array}\right)$, and then

$$[\,T - xI\,]_{\mathfrak{X}} = \left(\begin{array}{c|c} A - xI & * \\ \hline 0 & B - xI \end{array}\right)$$

But it is trivial to check that $A - xI = [(T - xI)|_M]_{\mathfrak{X}'}$ where $\mathfrak{X}' = \{e_1, \cdots, e_r\}$ are the initial vectors that span M. Thus $\det(A - xI)] = \det\big((T - xI)|_M\big) = p_{T|M}(x)$ divides $p_T(x) = \det(A - xI) \cdot \det(B - xI)$. $\square$

6.6. Exercise. Let (V, M, T) be as in the previous corollary. Then T induces a linear map $\tilde{T}$ from $V/M \to V/M$ such that $\tilde{T}(v + M) = T(v) + M$, for $v \in V$. Prove that the characteristic polynomial $p_{\tilde{T}}(x) = \det_{V/M}(\tilde{T} - xI)$, also divides $p_T(x) = \det(A - xI) \cdot \det(B - xI)$. $\square$

6.7. Lemma. *If f and P are nonconstant polynomials in $\mathbb{F}[x]$ and P divides f, so $f(x) = P(x)Q(x)$ for some other $Q \in \mathbb{F}[x]$, then $P(x)$ must split over $\mathbb{F}$ if $f(x)$ does.*

Proof: If Q is constant there is nothing to prove. Nonconstant polynomials $f \neq 0$ in $\mathbb{F}[x]$ have unique factorization into irreducible polynomials $f = \prod_{i=1}^{r} F_i$, where F_i cannot be written as a product of nonconstant polynomials of lower degree. Each polynomial f, P, Q has such a factorization $P = \prod_{k=1}^{m} P_k$, $Q = \prod_{j=1}^{m} Q_j$ so $f = PQ = \prod_k P_k \cdot \prod_j Q_j$. Since f splits over $\mathbb{F}$ it can also be written as a product of linear factors $f(x) = \prod_{i=1}^{n}(x - \alpha_i)$ where $\{\alpha_i\}$ are the roots of $f(x)$ in $\mathbb{F}$, counted according to their multiplicities. Linear factors $(x - \alpha)$ are obviously irreducible and the two irreducible decomposition of $f(x)$ must agree. Thus P (and Q) are products of linear factors and $P(x)$ splits over $\mathbb{F}$. $\square$

This lemma has a useful Corollary.

6.8. Corollary. *If the characteristic polynomial $p_T(x)$ of a linear operator $T : V \to V$ splits over $\mathbb{F}$, so does $p_{T|_W}$ for any T-invariant subspace $W \subseteq V$.*

Over $\mathbb{F} = \mathbb{C}$, all non-constant polynomials split.

Proof of Theorem 6.1. T has eigenvalues because p_T splits and its distinct eigenvalues $\{\lambda_1, \cdots, \lambda_r\}$ are the distinct roots of p_T in $\mathbb{F}$. Recall that $E_\lambda \neq \{0\} \Leftrightarrow M_\lambda \neq \{0\}$.

Pick an eigenvalue λ_1 and consider the Fitting decomposition $V = K_\infty \oplus R_\infty$ with respect to the operator $(T - \lambda_1 I)$, so K_∞ is the generalized eigenspace $M_{\lambda_1}(T)$ while R_∞ is the stable range of $(T - \lambda_1 I)$. Both spaces are invariant under $T - \lambda_1 I$, and also under T since $\lambda_1 I$ commutes with T. It will be important to note that

$$\lambda_1 \textit{ cannot be an eigenvalue of } T|_{R_\infty},$$

for if $v \in R_\infty$ is nonzero then $(T - \lambda_1 I)v = 0 \Rightarrow v \in K_\infty \cap R_\infty = \{0\}$. Hence $\mathrm{sp}(T|_{R_\infty}) \subseteq \{\lambda_2, \cdots, \lambda_r\}$.

We now argue by induction on $n = \dim(V)$. There is little to prove if $n = 1$. [There is an eigenvalue, so $E_\lambda = V$ and $T = \lambda I$ on V.] So, assume $n > 1$ and the theorem has been proved for all spaces V' of dimension $< n$ and all operators $T' : V' \to V'$ such that $\det(T' - \lambda I)$ splits over $\mathbb{F}$. The natural move is to apply this inductive hypothesis to $T' = T|_{R_\infty(T - \lambda_1 I)}$ since $\dim(R_\infty) = \dim(V) - \dim(M_{\lambda_1}) < \dim(V) = n$. But to do so we must show $p_{T'}$ splits over $\mathbb{F}$. [If $\mathbb{F} = \mathbb{C}$, every polynomial in $\mathbb{C}[x]$ splits, and this issue does not arise.]

By Corollary 6.8 the characteristic polynomial of $T' = T|_{R_\infty}$ splits over $\mathbb{F}$, and by induction on dimension $R_\infty(T')$ is a direct sum of generalized eigenspaces for the restricted operator T'.

$$R_\infty(T') = \bigoplus_{\mu \in \mathrm{sp}(T')} M_\mu(T') \ ,$$

where $\mathrm{sp}(T')$ = the distinct roots of $p|_{T'}$ in $\mathbb{F}$. To compare the roots of p_T and $p_{T'}$, we invoke the earlier observation that $p_{T'}$ divides p_T. Thus the roots $\mathrm{sp}(T') = \mathrm{sp}(T|_{R_\infty})$ are a subset of the roots $\mathrm{sp}(T)$ of $p_T(x)$, and in particular every eigenvalue μ for T' is an eigenvalue for T. Let's label the distinct eigenvalues of T so that

$$\mathrm{sp}(T') = \{\lambda_s, \lambda_{s+1}, \cdots, \lambda_r\} \subseteq \mathrm{sp}(T) = \{\lambda_1, \cdots, \lambda_r\}$$

(with $s > 1$ because $\lambda_1 \notin \mathrm{sp}(T|_{R_\infty})$, as we observed earlier).

Furthermore, for each $\mu \in \mathrm{sp}(T')$ the generalized eigenspace $M_\mu(T')$ is a subspace of $R_\infty \subseteq V$, and must be contained in $M_\mu(T)$ because $(T' - \mu I)^k v = 0 \Rightarrow (T - \mu I)^k v = 0$ for all $v \in R_\infty$. Thus,

$$R_\infty = \bigoplus_{\mu \in \mathrm{sp}(T')} M_\mu(T') \subseteq \sum_{\mu \in \mathrm{sp}(T')} M_\mu(T) \subseteq \sum_{\lambda \in \mathrm{sp}(T)} M_\lambda(T)$$

($R_\infty = \bigoplus_{\mu \in \mathrm{sp}(T')} M_\mu(T')$ by the induction hypothesis). Therefore the generalized eigenspaces $M_\lambda, \ \lambda \in \mathrm{sp}(T)$, must span V because

$$\begin{aligned} V \ &= \ K_\infty \oplus R_\infty \ = \ M_{\lambda_1}(T) \oplus R_\infty \ \subseteq \ M_{\lambda_1}(T) \oplus \big(\bigoplus_{\mu \in \mathrm{sp}(T')} M_\mu(T') \big) \\ &\subseteq \ M_{\lambda_1}(T) + \big(\sum_{\mu \in \mathrm{sp}(T')} M_\mu(T) \big) \qquad \big(\text{because } M_\mu(T') \subseteq M_\mu(T)\big) \qquad (15) \\ &\subseteq \ M_{\lambda_1}(T) + \big(\sum_{\lambda \in \mathrm{sp}(T)} M_\lambda(T) \big) \ \subseteq \ V \qquad \big(\text{because } \mathrm{sp}(T') \subseteq \mathrm{sp}(T)\big) \end{aligned}$$

Conclusion: the $M_\lambda(T), \ \lambda \in \mathrm{sp}(T)$, span V so by Proposition 6.2 V is a direct sum of its generalized eigenspaces. That finishes the proof of Theorem 6.1. $\square$

It is worth noting that

$$\mathrm{sp}(T') = \{\lambda_2, \ldots, \lambda_r\} \quad \text{and} \quad M_{\lambda_i}(T') = M_{\lambda_i}(T) \text{ for } 2 \le i \le r\ .$$

Since $M_\mu(T') \subseteq M_\mu(T)$ for all $\mu \in \mathrm{sp}(T')$, and $M_{\lambda_1}(T) \cap V' = (0)$, λ_1 cannot appear in $\mathrm{sp}(T')$; on the other hand every $\mu \in \mathrm{sp}(T')$ must lie in $\mathrm{sp}(T)$.

Consequences. Some things can be proved using just the block upper-triangular form for T rather than the more detailed Jordan Canonical form.

6.9. Corollary. *If the characteristic polynomial of $T : V \to V$ splits over $\mathbb{F}$, and in particular if $\mathbb{F} = \mathbb{C}$, there is a basis $\mathfrak{X}$ such that $[T]_{\mathfrak{X}}$ has block upper triangular form*

$$[T]_{\mathfrak{X}} = \begin{pmatrix} \boxed{T_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \boxed{T_r} \end{pmatrix} \tag{16}$$

with blocks on the diagonal

$$T_i = \begin{pmatrix} \lambda_i & & & * \\ & \cdot\ \cdot & & \\ & & \cdot\ \cdot & \\ & & & \cdot \\ 0 & & & \lambda_i \end{pmatrix}$$

of size $m_i \times m_i$ such that

1. *$\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of T.*
2. *The block sizes are the algebraic multiplicities m_i of the λ_i in the splitting of the characteristic polynomial $p_T(t)$ (see the next corollary for details).*
3. *$p_T(x) = (-1)^n \cdot \prod_{j=1}^{r} (x - \lambda_i)^{m_j}$ with $n = m_1 + \cdots + m_r$.*

The blocks T_i may or may not have off-diagonal terms. $\square$

6.10. Corollary. *If the characteristic polynomial of an $n \times n$ matrix A splits over $\mathbb{F}$, there is a similarity transform $A \mapsto SAS^{-1}$, $S \in \mathrm{GL}(n, \mathbb{F})$, such that SAS^{-1} has the block upper-triangular form shown above.*

6.11. Corollary. *If the characteristic polynomial of $T : V \to V$ splits over $\mathbb{F}$, and in particular if $\mathbb{F} = \mathbb{C}$, then for every $\lambda \in \mathrm{sp}(T)$ we have*

$$(\textit{algebraic multiplicity of } \lambda) = \dim(M_\lambda) = m_i$$

where m_i is the block size in (16)).

Proof: Taking a basis such that $[T]_{\mathfrak{X}}$ has the form (16), $[\,T - xI\,]_{\mathfrak{X}}$ will have the same form, but with diagonal entries λ_i replaced by $(\lambda_i - x)$. Then

$$\det[T - xI]_{\mathfrak{X}} = \prod_{j=1}^{r} (\lambda_j - x)^{\dim(M_{\lambda_j})} = p_T(x)$$

since the block T_j correspond to $T|_{M_{\lambda_j}}$. Obviously, the exponent on $(\lambda_j - x)$ is also the multiplicity of λ_j in the splitting of the characteristic polynomial p_T. $\square$

6.12. Corollary. *If the characteristic polynomial of an $n \times n$ matrix A, with distinct eigenvalues $\mathrm{sp}_{\mathbb{F}}(A) = \{\lambda_1, \ldots, \lambda_r\}$, splits over $\mathbb{F}$ then*

1. $\det(A) = \prod_{i=1}^{r} \lambda_i^{m_i}$, *the product of eigenvalues counted according to their algebraic multiplicities* m_i.

2. $\mathrm{Tr}(A) = \sum_{i=1}^{r} m_i \lambda_i$, *the sum of eigenvalues counted according to their algebraic multiplicities* m_i.

3. *More generally, if* $\mathbb{F} = \mathbb{C}$ *there are explicit formulas for all coefficients of the characteristic polynomial when we write it in the form*

$$p_A(x) = \det(A - x\,I) = \sum_{i=0}^{n} (-1)^i c_i(A)\, x^i$$

If eigenvalues are listed according to their multiplicities $m_i = m(\lambda_i)$, *say as* $\mu_1, \ldots, \mu_n$ *with* $n = \dim(V)$,

$$\mu_1 = \ldots = \mu_{m_1} = \lambda_1 \qquad \mu_{m_1+1} = \ldots = \mu_{m_1+m_2} = \lambda_2 \qquad \textit{etc}$$

then $c_n = 1$ *and*

$$\begin{aligned} c_{n-1} &= \sum_{j=1}^{n} \mu_j = \mathrm{Tr}(A), \\ &\vdots \\ c_{n-k} &= \sum_{j_1 < \ldots < j_k} \mu_{j_1} \cdot \ldots \cdot \mu_{j_k}, \\ &\vdots \\ c_0 &= \mu_1 \cdot \ldots \cdot \mu_n = \det(A) \end{aligned}$$

These formulas fail to be true if $\mathbb{F} = \mathbb{R}$ *and* $p_T(x)$ *has non-real roots in* $\mathbb{C}$.

6.13. Corollary. *If the characteristic polynomial of an* $n \times n$ *matrix* A *splits over* $\mathbb{F}$, *then* $T : V \to V$ *is diagonalizable if and only if*

$$(\textit{algebraic multiplicity}) = (\textit{geometric multiplicity}) \qquad \textit{for each } \lambda \in \mathrm{sp}(T).$$

Both multiplicities are then equal to $\dim(E_\lambda(T))$.

Proof: $E_\lambda \subseteq M_\lambda$ for every eigenvalue, and by the previous corollary we know that

$$(\textit{geometric multiplicity}) = \dim(E_\lambda) \leq \dim(M_\lambda) = (\textit{algebraic multiplicity})\ .$$

Furthemore, $M_\lambda = \ker(T - \lambda I)^N$ for large $N \in \mathbb{N}$. Writing $V = M_{\lambda_1} \oplus \ldots \oplus M_{\lambda_r}$, the implication ($\Leftarrow$) follows because

$$\begin{aligned} (\textit{algebraic multiplicity}) &= (\textit{geometric multiplicity}) \\ &\Rightarrow \dim(E_{\lambda_i}) = \dim(M_{\lambda_i}) \ \text{ for all } i \\ &\Rightarrow M_{\lambda_i} = E_{\lambda_i} \text{ since } M_{\lambda_i} \supseteq E_{\lambda_i} \\ &\Rightarrow V = \bigoplus_{i=1}^{r} E_{\lambda_i} \text{ and } T \text{ is diagonalizable.} \end{aligned}$$

For ($\Rightarrow$): if T is diagonalizable we have $V = \bigoplus_{i=1}^{r} E_{\lambda_i}$, but $E_{\lambda_i} \subseteq M_{\lambda_i}$ for each i. Comparing this with the Jordan decomposition $V = \bigoplus_{i=1}^{r} M_{\lambda_i}$ we see that $M_{\lambda_i} = E_{\lambda_i}$.
□

6.14. Corollary. *If A is an $n \times n$ matrix whose characteristic polynomial splits over $\mathbb{F}$, let $\mathfrak{X}$ be a basis that puts $[T]_{\mathfrak{X}}$ into Jordan form, so that*

$$(17) \qquad [T]_{\mathfrak{X}} = \begin{pmatrix} \boxed{T_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \boxed{T_r} \end{pmatrix} \qquad \textit{with} \qquad T_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \cdot & \cdot & \\ & & \cdot & \cdot \\ & & & \cdot \;\; 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

Then with respect to the same basis the powers $I, T, T^2, \cdots$ take form:

$$(18) \qquad [T^k]_{\mathfrak{X}} = \begin{pmatrix} \boxed{T_1^k} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \boxed{T_r^k} \end{pmatrix} \qquad \textit{with} \qquad T_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \cdot & \cdot & \\ & & \cdot & \cdot \\ & & & \cdot \;\; 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

Proof: $[T_i^k]_{\mathfrak{X}} = \big([T_i]_{\mathfrak{X}}\big)^k$ for $k = 0, 1, 2, \cdots$. $\square$

In (18) there may be blocks of various sizes with the same diagonal values λ_i.

These particular powers $[T^k] = [T]^k$ are actually easy to compute. Each block T_i has the form $T_i = \lambda_i I + N_i$ with N_i an elementary nilpotent matrix, so we have

$$T_i^k = \sum_{j=0}^{k} \tbinom{k}{i} \lambda^{k-j} N_i^j \qquad \text{(binomial expansion)} ,$$

with $N_i^j = 0$ when $j \geq \deg(N_i)$.

6.15. Exercise. If N is an $r \times r$ elementary nilpotent matrix

$$N = \begin{pmatrix} 0 & 1 & & 0 \\ & \cdot & \cdot & \\ & & \cdot & \cdot \\ & & & \cdot \;\; 1 \\ 0 & & & 0 \end{pmatrix} \qquad \text{show that} \qquad N^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ & \cdot & \cdot & \\ & & \cdot & 1 \\ & & \cdot & 0 \\ 0 & & & 0 \end{pmatrix} ,$$

Each new multiple of N moves the diagonal file of $1's$ one step to the upper right, with $N^r = 0$ at the last step. $\square$

6.16. Exercise. If the characteristic polynomial of $T : V \to V$ splits over $\mathbb{F}$, there is a basis that puts $[T]_{\mathfrak{X}}$ in Jordan form, with diagonal blocks

$$A = \lambda I + N = \begin{pmatrix} \lambda & 1 & & 0 \\ & \cdot & \cdot & \\ & & \cdot & \cdot \\ & & & \cdot \;\; 1 \\ 0 & & & \lambda \end{pmatrix}$$

Compute the exponential matrix

$$\mathrm{Exp}(tA) = e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$$

for $t \in \mathbb{F}$.
Hint: If A and B commute we have $e^{A+B} = e^A \cdot e^B$; apply the previous exercise.
Note: Since N is nilpotent the exponential series is a finite sum. $\square$

The Spectral Mapping Theorem. Suppose A is an $n \times n$ matrix whose characteristic polynomial splits over $\mathbb{F}$, with

$$p_A(x) = \det(A - x\,I) = c \cdot \prod_{i=1}^{r} (\lambda_i - x\,I)^{m_i} \qquad \text{if } \mathrm{sp}(T) = \{\lambda_1, \ldots, \lambda_r\}\ .$$

Examination of the diagonal entries in the Jordan Canonical form (16), or even the upper triangular form (17), immediately reveals that

$$\mathrm{sp}(T^k) = \text{ the } \textit{distinct} \text{ entries in the list of powers } \lambda_1^k, \ldots, \lambda_r^k$$

Be aware that there might be repeated entries among the λ_i^k, even though the λ_i are distinct. (Devise an example in which $\mathrm{sp}(T^k)$ reduces to a single point even though $\mathrm{sp}(T)$ contains several distinct points.)

Therefore the characteristic polynomial of T^k is the product of the diagonal entries $(\lambda_i^k - x)$ in the Jordan form of $(T^k - x\,I)$,

$$p_{T^k}(x) \;=\; \det(T^k - x\,I) \;=\; \prod_{i=1}^{r} (\lambda_i^k - x)^{m_i}\ , \qquad (\text{where } m_i = \dim(M_{\lambda_i}(T))\ .$$

More can be said under the same hypotheses. If $f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m$ is any polynomial in $\mathbb{F}[t]$ we can form an operator $f(T)$ that maps $V \to V$ to obtain a natural corresponding $\Phi : \mathbb{F}[t] \to \mathrm{Hom}_{\mathbb{F}}(V)$ such that

$$\Phi(1) = I, \quad \Phi(t) = T, \quad \Phi(t^k) = T^k$$

and

$$\begin{aligned} \Phi(f_1 + f_2) &= \Phi(f_1) + \Phi(f_2) && \text{(sum of linear operators)} \\ \Phi(f_1 \cdot f_2) &= \Phi(f_1) \circ \Phi(f_2) && \text{(composition of linear operator)} \\ \Phi(cf) &= c \cdot \Phi(f) && \text{for all } c \in \mathbb{F} \end{aligned}$$

I.e. Φ is a homomorphism between associative algebras. With this in mind we can prove:

6.17. Theorem. (Spectral Mapping Theorem). *If the characteristic polynomial of a linear operator $T : V \to V$ splits over $\mathbb{F}$, and in particular if $\mathbb{F} = \mathbb{C}$, every polynomial $f(t) = \sum_{i=0}^{m} a_i t^i$ in $\mathbb{F}[t]$ determines an operator $\Phi(f)$ in $\mathrm{Hom}_{\mathbb{F}}(V, V)$,*

$$\Phi(f) = \sum_{i=0}^{m} a_i T^i$$

The correspondence $\Phi : \mathbb{F}[t| \to \mathrm{Hom}_{\mathbb{F}}(V)$ is a unital homomorphism of associative algebras and has the following SPECTRAL MAPPING PROPERTY

$$\mathrm{sp}(f(T)) = f\big(\mathrm{sp}(T)\big) = \{f(z) : z \in \mathrm{sp}(T)\} \tag{19}$$

In particular, this applies if T is diagonalizable over $\mathbb{F}$.

Proof: It suffices to choose a basis $\mathfrak{X}$ such that $[T]_{\mathfrak{X}}$ has block upper-triangular form

$$[T]_{\mathfrak{X}} \;=\; \begin{pmatrix} T_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & T_r \end{pmatrix} \qquad \text{with} \qquad T_i \;=\; \begin{pmatrix} \lambda_i & & & * \\ & \cdot\ \cdot & & \\ & & \cdot\ \cdot & \\ & & & \cdot \\ 0 & & & \lambda_i \end{pmatrix}$$

of size $m_i \times m_i$ since $[T^k]_{\mathfrak{X}} = [T]^k_{\mathfrak{X}}$ (matrix product) for $k = 0, 1, 2, \cdots$. Hence

$$[f(T)]_{\mathfrak{X}} = f\big([T]_{\mathfrak{X}}\big) = a_0 I + a_1 [T]_{\mathfrak{X}} + \cdots + a_m [T]^m_{\mathfrak{X}}$$

because $f(A) = a_0 I + a_1 A + \cdots + a_m A^m$ for any matrix A.

As in the previous corollary, it follows that $[f(T)]_{\mathfrak{X}}$ is made up of blocks on the diagonal, each of which is upper-triangular with diagonal values $f(\lambda_i)$; then the characteristic polynomial of $f(T)$ is

$$p_{f(T)}(x) = \det(f(T) - x\,I) = \prod_{i=1}^{r} (f(\lambda_i) - x\,I)^{m_i}, \qquad m_i = \dim(M_{\lambda_i})\ .$$

This is zero if and only if $x = f(\lambda_i)$ for some i, so $\mathrm{sp}(f(T)) = \{f(\lambda_i) : 1 \le i \le r\} = f(\mathrm{sp}(T))$. Obviously the characteristic polynomial of $f(T)$ splits over $\mathbb{F}$ if $p_T(t)$ splits.

Here $\lambda_i \in \mathrm{sp}(T) \Rightarrow f(\lambda_i) \in \mathrm{sp}(f(T))$, but the multiplicity of $f(\lambda_i)$ as an eigenvalue of $f(T)$ might be greater than the multiplicity of λ_i as an eigenvalues of T because we might have $f(\lambda_i) = f(\lambda_j)$, and then $\mu = f(\lambda_i)$ will have multiplicity at least $m_i + m_j$ in $\mathrm{sp}(f(T))$.

Another consequence is the Cayley-Hamilton theorem, which can be proved in other ways without developing the Jordan Canonical form. However this normal form suggests the underlying reason why the result is true, and makes its validity almost obvious. On the other hand, alternative proofs can be made to work for arbitrary $\mathbb{F}$ and T, without any assumptions about the characteristic polynomial $p_T(x)$. Since the result is true in this generality, we give both proofs.

6.18. Theorem. (Cayley-Hamilton). *For any linear operator $T : V \to V$ on a finite dimensional vector space, over any $\mathbb{F}$, we have*

$$p_T(T) = \big[\, p_T(t)\big|_{t=T} \big] \,=\, 0 \quad (\textit{the zero operator in } \mathrm{Hom}_{\mathbb{F}}(V,V))\ ,$$

Thus, applying the characteristic polynomial $p_T(x) = \det(T - x\,I)$ to the linear operator T itself yields the zero operator.

Proof: If $p_T(x)$ splits over $\mathbb{F}$ we have $p_T(x) = \prod_{i=1}^{r} (\lambda_i - x)^{m_i}$, where $m_i = \dim(M_{\lambda_i})$ and $\{\lambda_1, \cdots, \lambda_r\}$ are the (distinct) eigenvalues in $\mathrm{sp}_{\mathbb{F}}(T)$. We want to show that

$$0 = \prod_{i=1}^{r} (T - \lambda_i I)^{m_i} = \big[\, p_T(x)\big|_{x=T} \big]$$

But $V = \bigoplus_{i=1}^{r} M_{\lambda_i}$ and $(T - \lambda_i I)^{m_i}(M_{\lambda_i}) = (0)$ [Given a Jordan basis in M_{λ_i}, $A = [(T - \lambda_i I)|_{M_{\lambda_i}}]_{\mathfrak{X}}$ consists of elementary nilpotent blocks N_j on the diagonal; the size $d_j \times d_j$ of such a block cannot exceed $m_i = \dim(M_{\lambda_i})$, so $N_j^{d_j} = N_j^{m_i} = 0$ for each j.] Hence $\prod_{j=1}^{r} (T - \lambda_i I)^{m_i}\big(M_{\lambda_i}\big) = (0)$, so the operator $p_T(T)$ is zero on each M_{λ_i} and on all of V. $\Box$

If p_T does not split over $\mathbb{F}$, a different argument shows that $p_T(T)v = 0$ for all $v \in V$.

Alternative Proof (6.18): The result is obvious if $v = 0$. If $v \ne 0$ there is a largest $m \ge 1$ such that $v, T(v), T^2(v), \cdots, T^{m-1}(v)$ are linearly independent. Then

$$W = \mathbb{F}\text{-span}\{T^k(v) : k \in \mathbb{N}\} = \mathbb{F}\text{-span}\{T^k(v) : 0 \le k \le m-1\}$$

and $\{v, T(v), \cdots, T^{m-1}(v)\}$ is a basis for the cyclic subspace W. This space is clearly T-invariant, and as we saw before, $p_{(T|_W)}$ divides p_T, so that $p_T(x) = p_{(T|_W)}(x) \cdot Q(x)$ for some $Q \in \mathbb{F}[x]$. We now compute $p_{(T|_W)}(x)$. For the basis $\mathfrak{X} = \{v, T(v), \cdots, T^{m-1}(v)\}$

we note that $T^m(v) = T(T^{m-1}(v))$ is a unique linear combination of the previous vectors $T^k(v)$, say

$$T^m(v) + a_{m-1}T^{m-1}(v) + \cdots + a_1 T(v) + a_0 v = 0\ .$$

Hence,

$$\big[T|_W\big]_{\mathfrak{X}} = \begin{pmatrix} 0 & & & 0 & -a_0 \\ 1 & \cdot & & & -a_1 \\ & \cdot & \cdot & & \cdot \\ & & \cdot & \cdot & \cdot \\ 0 & & & 1 & -a_{m-1} \end{pmatrix} \quad \text{and} \quad \big[(T-x\,I)|_W\big]_{\mathfrak{X}} = \begin{pmatrix} -x & & & 0 & -a_0 \\ 1 & \cdot & & & -a_1 \\ & \cdot & \cdot & & \cdot \\ & & \cdot & \cdot & \cdot \\ 0 & & & 1 & -x-a_{m-1} \end{pmatrix}$$

6.19. Exercise. Show that

$$p_{(T|_W)}(x) = \det(T|_W - x\,I)_{\mathfrak{X}} = (-1)^m\big(t^m + a_{m-1}t^{m-1} + \cdot + a_1 t + a_0\big)\ . \quad \square$$

It follows that $p_{(T|_W)}(T)$ satisfies the equation

$$p_{(T|_W)}(T)v = (-1)^m\big(T^n(v) + a_{m-1}T^{m-1}(v) + \cdot + a_1 T(v) + a_0 v\big) = 0$$

by definition of the coefficients $\{a_j\}$. But then

$$p_T(T)v = Q(T)\cdot[\,p_{(T|_W)}(T)v] = Q(T)(0) = 0\ .$$

(Recall that $W = \mathbb{F}\text{-span}\{T^k(v)\}$ as in Propositions 2.5 and 2.7.) Since this is true for all $v \in V$, $p_T(T)$ is the zero operator in $\mathrm{Hom}_{\mathbb{F}}(V,V)$. $\square$

Remarks: If $T : V \rightarrow V$ is a linear operator on a finite dimensional vector space the polynomial $Q(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$ in $\mathbb{F}[x]$ of minimal degree such that $Q(T) = 0$ is called the *minimal polynomial* for T. The polynomial function of T defined above by substituting $x = T$

$$T^m + a_{m-1}T^{m-1} + \cdots + a_1 T + a_0 I = 0$$

is precisely the minimal polynomial for T. The Jordan form (12) can be used to determine the minimal polynomial, but the block upper-triangular form (11) is too crude for this purpose. (The problem is that the nilpotence degree $\deg(N)$ of a nilpotent matrix will be greater than the degree of the minimal polynomial unless there is a cyclic vector in V.) $\square$

6.20. Example. Let $T : V \rightarrow V$ be a linear map on $V = \mathbb{R}^4$ whose matrix with respect to the standard basis $\mathfrak{X} = \{e_1, \cdots, e_4\}$ has the form

$$A = [T]_{\mathfrak{X}} = \begin{pmatrix} 7 & 1 & 2 & 2 \\ 1 & 4 & -1 & -1 \\ -2 & 1 & 5 & -1 \\ 1 & 1 & 2 & 8 \end{pmatrix} \quad \text{so that} \quad A - 6I = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & 1 & 2 & 2 \end{pmatrix}$$

After some computational work which we omit, we find that

$$p_T(t) = \det(A - xI) = (x-6)^4 = x^4 - 4(6x^3) + 6(6^2x^2) - 4(6^3x) + 6^4\ ,$$

so $\mathrm{sp}_{\mathbb{R}}(T) = \{6\}$ with algebraic multiplicity $m = 4$. Thus $V = M_{\lambda=6}(T)$ and $(T - 6I)$ is nilpotent. We find $K_1 = \ker(T-6I) = E_{\lambda=6}(T)$ by row reduction of $[T-6I]_{\mathfrak{X}} = [A-6I\,]$,

$$[A - 6I] \rightarrow \begin{pmatrix} 1 & 1 & 2 & 2 \\ 0 & -3 & -3 & -3 \\ 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 1 & 2 & 2 \\ 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

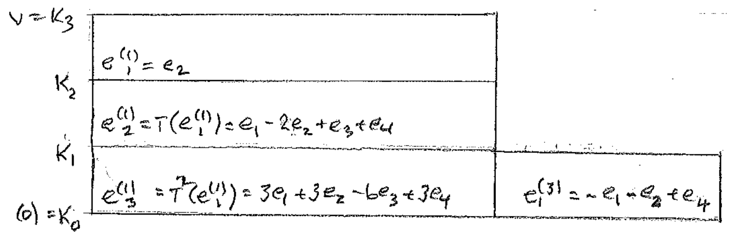

Figure 7.2. The version of Figure 7.1 worked out in Example 6.20. Although there are three columns, Column 2 is empty and has been deleted in the present diagram. All the basis vectors $e_j^{(i)}$ are shown

Thus,

$$\begin{aligned} K_1 &= \{v = (-s-t, -s-t, s, t) : s, t \in \mathbb{R}\} \\ &= \mathbb{R}\text{-span}\{f_1^{(1)} = -e_1 - e_2 + e_3 \,,\ f_2^{(1)} = -e_1 - e_2 + e_4\} \\ &= \mathbb{R}\text{-span}\{(-1,-1,1,0)\,,\ (-1,-1,0,1)\} \end{aligned}$$

and $\dim(K_1) = 2$. Next row reduce $\ker(A - 6I\,)^2$ to get

$$[A - 6I\,]^2 \rightarrow \begin{pmatrix} 0 & 3 & 3 & 3 \\ 0 & 3 & 3 & 3 \\ 0 & -6 & -6 & -6 \\ 0 & 3 & 3 & 3 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The first column (which meets no "step corner") corresponds to free variable x_1; the other free variables are x_3, x_4. Thus $K_2 = \ker(A - 6I)^2$ is

$$\begin{aligned} K_2 &= \{v = (a, -b-c, b, c) : a, b, c \in \mathbb{R}\} \\ &= \mathbb{R}\text{-span}\{f_1^{(2)} = e_1 \,,\ f_2^{(2)} = e_3 - e_2 \,,\ f_3^{(2)} = e_4 - e_2\} \ = \ \mathbb{R}\text{-span}\{e_1, e_3 - e_2, e_4 - e_2\} \end{aligned}$$

and $\dim(K_2) = 3$. Finally $(A - 6I)^3 = 0$, so $\deg(T - 6I\,) = 3$ and $K_3 = V$.

We now apply the procedure for finding cyclic subspaces outlined in Figure 7.1.

Step 1: Find a basis for V mod K_2. Since $\dim(V/K_2) = 1$ this is achieved by taking any $v \in V \sim K_2$. One such choice is $e_1^{(1)} = e_2 = (0,1,0,0)$, which obviously is not in K_2. Then compute its images under powers of $(T - 6I)$,

$$\begin{aligned} e_2^{(1)} &= (T - 6I)e_1^{(1)} \ = \ (1,-2,1,1) \ = \ e_1 - 2e_2 + e_3 + e_4 \in K_2 \sim K_1 \\ e_3^{(1)} &= (T - 6I)^2 e_1^{(1)} \ = \ (3,3,-6,3) \ = \ 3(e_1 + e_2 - 2e_3 + e_4) \in K_1 \sim \{0\} \end{aligned}$$

Step 2: There is no need to augment the vector

$$e_2^{(1)} = (T - 6I)e_1^{(1)} = (1,-2,1,1) \in K_2$$

to get a basis for K_2/K_1, because $\dim(K_2/K_1) = 1$.

Step 3: In $K_1 \sim \{0\}$ we must augment $e_3^{(1)} = (T - 6I)^2 e_1^{(1)}$ to get a basis for $K_1/K_0 \cong K_1$. We need a new vector $e_1^{(3)} \in K_1 \sim \{0\}$ such that $e_1^{(3)}$ and $e_3^{(1)}$ are independent mod $K_0 = (0)$ – i.e. vectors that are actually independent in V. We could try $e_1^{(3)} =$

$(-1,-1,0,1) = -e_1 - e_2 + e_4$ which is in $K_1 \sim K_0$. Independence holds if and only if the matrix M whose rows are $e_3^{(1)}$ and $e_1^{(3)}$ has $rank = 2$. But row operations yield

$$\begin{pmatrix} 1 & 1 & -2 & 1 \\ -1 & -1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 1 & -2 & 1 \\ 0 & 0 & -2 & 2 \end{pmatrix}$$

which has row $rank = 2$, as desired.

Thus $\{T^2(e_1^{(1)}), T(e_1^{(1)}), e_1^{(1)} \,;\, e_1^{(3)}\}$ is a basis for all of V such that

$$\begin{aligned} C_1 &= \mathbb{R}\text{-span}\{T^2(e_1^{(1)}), T(e_1^{(1)}), e_1^{(1)}\}, \\ C_2 &= \mathbb{R}\text{-span}\{e_1^{(3)}\} \end{aligned}$$

are independent cyclic subspaces, generated by the vectors $e_1^{(1)}$ and $e_1^{(3)}$. The ordered basis $\mathfrak{Y} = \{T^2(e_1^{(1)}), T(e_1^{(1)}), e_1^{(1)} \,;\, e_1^{(3)}\}$ puts $[T]_{\mathfrak{Y}}$ in Jordan canonical form

$$[T]_{\mathfrak{Y}} = \left(\begin{array}{ccc|c} 6 & 1 & 0 & 0 \\ 0 & 6 & 1 & 0 \\ 0 & 0 & 6 & 0 \\ \hline 0 & 0 & 0 & 6 \end{array}\right)$$

Basis vectors $T^2(e_1^{(1)})$ and $e_1^{(3)}$ are eigenvectors for the action of T and $E_{\lambda=6}(T) = \mathbb{F}\text{-span}\{e_1^{(3)}, T^2(e_1^{(3)})\}$ is 2-dimensional. $\square$

6.21. Exercise. Find the Jordan canonical form for the linear operators $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose matrices with respect to the standard bases $\mathfrak{X} = \{e_1, \cdots, e_n\}$ are

$$\text{(a)} \quad A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix} \qquad \text{(b)} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The Minimal Polynomial for $T : V \rightarrow V$. The space of linear operators $\mathrm{Hom}_{\mathbb{F}}(V, V) \cong \mathrm{M}(n, \mathbb{F})$ is finite dimensional, so there is a smallest exponent $m \in \mathbb{N}$ such that the powers $I, T, T^2, \ldots, T^{m-1}$ are linearly independent. Thus there are coefficients c_j such that $T^m + \sum_{j=0}^{m-1} c_j T^j = 0$ (the zero operator) The monic polynomial

$$x^m + \sum_{j=0}^{m-1} c_j x^j \text{ in } \mathbb{F}[x]$$

is the (unique) **minimal polynomial** $m_T(x)$ for this operator. Obviously $d = \deg(m_T)$ cannot exceed $n^2 = \dim(\mathrm{M}(n, \mathbb{F}))$, but it could be a lot smaller. The minimal polynomial for a matrix $A \in \mathrm{M}(n, \mathbb{F})$ is defined the same way, and it is easy to see that the minimal polynomial $m_T(x)$ for a linear operator is the same as the minimal polynomial of the associated matrix $A = [T]_{\mathfrak{X}}$, and this is so for *every* basis $\mathfrak{X}$ in V. Conversely the minimal polynomial $m_A(x)$ of a matrix coincides with that of the linear operator $L_A : \mathbb{F}^n \rightarrow \mathbb{F}^n$ such that $L_A(v) = A \cdot v$ (matrix product of $(n \times N) \times (n)$ column vector).

Computing $m_T(x)$ could be a chore, but it is easy if we know the Jordan form for T, and this approach also reveals interesting connections between the minimal polynomial $m_T(x)$ and the characteristic polynomial $p_T(x) = \det(T - x\,I)$. We have already seen that the characteristic polynomial is a "similarity invariant" for matrices (or for linear operators), so that

A similarity transformation, $A \mapsto SAS^{-1}$ yields a new matrix with the same characteristic polynomial, so $p_{SAS^{-1}}(x) = p_A(x)$ in $\mathbb{F}[x]$ for all invertible matrices S.

(See Section II.4 of the *Linear Algebra I Class Notes* for details regarding similarity transformations, and Section V.1 for invariance of the characteristic polynomial.) The minimal polynomial is also a similarity invariant, a fact that can easily be proved directly from the definitions.

6.22. Exercise. Explain why the minimal polynomial is the same for:

1. A matrix A and the linear operator $L_A : \mathbb{F}^n \to \mathbb{F}^n$.
2. A linear operator $T : V \to V$ on a finite dimensional vector space and its matrix $A = [T]_{\mathfrak{X}}$ with respect to any basis in V $\Box$

6.23. Exercise. Prove that the minimal polynomial $m_A(x)$ of a matrix $A \in \mathrm{M}(n, \mathbb{F})$ is a similarity invariant: $m_{SAS^{-1}}(x) = m_A(x)$ for any invertible $n \times n$ matirx S. $\Box$

We will use Example 6.22 to illustrate how the minimal polynomial can be found if the Jordan form for a matrix is known, but first let's compute and compare $m_A(x)$ and $p_A(x)$ for a diagonal matrix. If

$$A = \begin{pmatrix} \boxed{\lambda_1 I_{d_1}} & & 0 \\ & \ddots & \\ 0 & & \boxed{\lambda_r I_{d_r}} \end{pmatrix} \qquad \text{where } I_k = k \times k \text{ identity matrix.}$$

Obviously the characteristic polynomial, which depends only on the diagonal values of A, is $p_A(x) = \prod_{j=1}^{r} (\lambda_j - x)^{d_j}$; in contrast, we will show that the minimal polynomial is the product of the distinct factors,

$$m_A(x) = \prod_{j=1}^{r} (\lambda_j - x) \ ,$$

each taken with multiplicity one – i.e. for diagonal matrices, $m_A(x)$ is just $p_A(x)$, ignoring multiplicities.

MORE TEXT TO BE ADDED RE: MIN POLYN

VII-7 Applications of Jordan Form to Differential Equations.

Computing the exponential $e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$ of a matrix turns out to be important in many applications, one of which will be illustrated below. This is not an easy calculation if you try to do it by summing the series. In fact, computing a product of $n \times n$ matrices seems to require n^3 multiplication operations on matrix entries and computing a high power such as A^{200} directly could be a formidable task. But it can be done by hand for diagonal matrices, and for elementary nilpotent matrices (see Exercises 6.15 - 6.16), and hence for any matrix that is already in Jordan Canonical form. For a linear operator $T : V \to V$ the Jordan form is obtained by choosing a suitable basis in V; for a matrix A this amounts to finding an invertible matrix S such that the similarity transform $A \mapsto SAS^{-1}$ puts A into Jordan form. Similarity transforms are invertible operations that interact nicely with the matrix-exponential operation, with

$$Se^A S^{-1} = e^{SAS^{-1}} \qquad \text{for every } A \in \mathrm{M}(n, \mathbb{C}) \ . \tag{20}$$

Thus if we can determine the Jordan form B for A, we can compute e^A in four simple steps,

$$\begin{aligned} A \quad &\to \quad \text{Jordan form } B = SAS^{-1} \\ &\to \quad e^B = e^{SAS^{-1}} \quad \text{(a calculation that can be done by hand)} \\ &\to \quad e^A = S^{-1}e^BS = e^{S^{-1}BS} \qquad \text{(because } B = SAS^{-1}) \end{aligned}$$

Done. Note that e^D will have block upper-triangular form if D has Jordan form.

This idea was illustrated for *diagonalizable* matrices and operators in the *Linear Algebra I Notes*, but in the following example you will see that it is easily adapted to deal with matrices whose Jordan form can be determined. Rather that go through such a calculation just to compute e^A, we will go the extra parsec and show how the ability to compute matrix exponentials is the key to solving systems of constant coefficient differential equations.

Solving Linear Systems of Ordinary Differential Equations.

If $A \in \mathrm{M}(n, \mathbb{F})$ for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$ is a differentiable function with values in $\mathbb{F}^n$, the vector identity

$$\frac{d\mathbf{x}}{dt} = A \cdot \mathbf{x}(t) \qquad \mathbf{x}(0) = \mathbf{c} = (c_1, \ldots, c_n) \tag{21}$$

is a system of constant coefficient differential equations with initial conditions $x_k(0) = c_k$ for $k = 1, \ldots, n$. There is a unique solution for all $-\infty < t < \infty$, given by $\mathbf{x}(t) = e^{tA} \cdot \mathbf{c}$. This follows because the vector-valued map $\mathbf{x}(t)$ is infinitely differentiable, with

$$\frac{d}{dt}\Big(e^{tA}\Big) = A \cdot e^{tA} \quad \text{for all } t \in \mathbb{R} \ ,$$

from which we get

$$\frac{d\mathbf{x}}{dt} = \frac{d}{dt}\Big(e^{tA}\Big) \cdot \mathbf{c} = Ae^{tA} \cdot \mathbf{c} = A \cdot \mathbf{x}(t) \ .$$

Solving the differential equation (21) therefore reduces to computing e^{tA}, but how do you do that? As noted above, if A has a characteristic polynomial that splits over $\mathbb{F}$ (or if $\mathbb{F} = \mathbb{C}$), we can find a basis that puts A in Jordan canonical form. That means we can, with some effort, find a nonsingular $S \in \mathrm{M}(n, \mathbb{F})$ such that $B = SAS^{-1}$ consists of diagonal blocks

$$B = \begin{pmatrix} \boxed{B_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & \boxed{B_r} \end{pmatrix} ,$$

each having the form

$$B_k = \begin{pmatrix} \lambda_k & 1 & & & 0 \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ 0 & & & & \lambda_k \end{pmatrix} = \lambda_k I + N_k$$

where $\lambda_k \in \mathrm{sp}_{\mathbb{F}}(A)$, and N_k is either an elementary (cyclic) nilpotent matrix, or a 1×1 block consisting of the scalar λ_k, in which the elementary nilpotent part is degenerate. (Recall the discussion surrounding equation (12)). Obviously,

$$Se^{tA}S^{-1} = S\Big(\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k\Big) S^{-1} = \sum_{k=0}^{\infty} \frac{t^k}{k!} SA^kS^{-1} = e^{t\,SAS^{-1}} ,$$

but $SA^kS^{-1} = \big(SAS^{-1}\big)^k$ for $k = 0, 1, 2 \ldots$, so

$$Se^{tA}S^{-1} = \sum_{k=0}^{\infty} \frac{t^k(SAS^{-1})^k}{k!} = \begin{pmatrix} e^{tB_1} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & e^{tB_r} \end{pmatrix} = e^{tB}$$

Here $e^{tB} = e^{t(\lambda I+N)} = e^{t\lambda I} \cdot e^{tN}$ because $e^{A+B} = e^A \cdot e^B$ when A and B commute, and then we have

$$e^{tB} = e^{\lambda t}I \cdot \Big[\sum_{j=0}^{d-1} \frac{t^j}{j!} N^j\Big] = e^{\lambda t}I \cdot \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{d-1}}{(d-1)!} \\ & \cdot & \cdot & & \vdots \\ & & \cdot & \cdot & \frac{t^2}{2!} \\ & & & 1 & t \\ 0 & & & & 1 \end{pmatrix}$$

of size $d \times d$. These matrices can be computed explicitly and so can the scalars $e^{t\lambda}$. Then we can recover e^{tA} by reversing the similarity transform to get

$$e^{tA} = S^{-1}e^{tB}S$$

which requires computing two products of explicit matrices. The solution of the original differential equation

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x}(t) \quad \text{with inital condition} \quad \mathbf{x}(0) = \mathbf{c}$$

is then $\mathbf{x}(t) = e^{tA} \cdot \mathbf{c}$, as above.

7.1. Exercise. Use the Jordan Canonical form to find a list of solutions $A \in \mathrm{M}(2, \mathbb{C})$ to the matrix identity

$$A^2 + I = 0 \qquad \text{(the “square roots of } -I) \text{ ,}$$

such that every solution is similar to one in your list.
Note: $SA^2S^{-1} = (SAS^{-1})^2$. $\square$

Chapter VIII. Complexification.

VIII.1. Analysis of Linear Operators over $\mathbb{R}$.

How can one analyze a linear operator $T : V \to V$ when the characteristic polynomial $p_T(x)$ does not split over $\mathbb{F}$? One approach is via the "*Rational Canonical form*," which makes no attempt to replace the ground field $\mathbb{F}$ with a larger field of scalars $\mathbb{K}$ over which p_T might split; we will not pursue this topic in these *Notes*. A different approach, which we will illustrate for $\mathbb{F} = \mathbb{R}$, is to enlarge $\mathbb{F}$ by constructing a field of scalars $\mathbb{K} \supseteq \mathbb{F}$; then we may in an obvious way regard $\mathbb{F}[x]$ as a subalgebra within $\mathbb{K}[x]$, and since $\mathbb{K} \supseteq \mathbb{F}$ there is a better chance that $f(x)$ will split into linear factors in $\mathbb{K}[x]$

$$f(x) = c \cdot \prod_{j=1}^{d} (x - \mu_j)^{m_j} \qquad \text{with } c \text{ and } \mu_j \text{ in } \mathbb{K}. \tag{22}$$

It is in fact always possible to embed $\mathbb{F}$ in a field $\mathbb{K}$ that is **algebraically closed**, which means that every polynomial $f(x) \in \mathbb{K}[x]$ splits into linear factors belonging to $\mathbb{K}[x]$, as in (22).

The Fundamental Theorem of Algebra asserts that the complex number field is algebraically closed; but the real number system $\mathbb{R}$ is not – for instance $x^2 + x + 2 \in \mathbb{R}[x]$ cannot split into linear factors in $\mathbb{R}[x]$ because it has no roots in $\mathbb{R}$. However, it does split when regarded as an element of $\mathbb{C}[x]$,

$$x^2 + x + 1 = (x - z_+) \cdot (x - z_-) - \big(x - \tfrac{1}{2}(-1 + i\sqrt{3})\big) \cdot \big(x + \tfrac{1}{2}(-1 - i\sqrt{3})\big)$$

where $i = \sqrt{-1}$. In this simple example one can find the complex roots $z_\pm = \frac{1}{2}(-1 \pm i\sqrt{3})$ using the quadratic formula.

Any real matrix $A \in \mathrm{M}(n, \mathbb{R})$ can be regarded as a matrix in $\mathrm{M}(n, \mathbb{C})$ whose entries happen to be real. Thus the operator

$$L_A(\mathbf{x}) = A \cdot \mathbf{x} \qquad \text{(matrix multiplication)}$$

acting on $n \times 1$ column vectors can be viewed as a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$, but also as a "complexified" operator $T_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$ on the complexified space $\mathbb{C}^n = \mathbb{R}^n + i\mathbb{R}^n$. Writing vectors $\mathbf{z} = (z_1, \cdots, z_n)$ with complex entries $z_j = x_j + iy_j$ $(x_j, y_j \in \mathbb{R})$, we may regard $\mathbf{z}$ as a combination $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ with complex coefficients of the real vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$. The complexified operator $T_{\mathbb{C}} \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^n, \mathbb{C}^n)$ can then be expressed in terms of the original $\mathbb{R}$-linear map $T : \mathbb{R}^n \to \mathbb{R}^n$:

$$T_{\mathbb{C}}(x + iy) = T(x) + iT(y),\ T_{\mathbb{C}} \in \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^n, \mathbb{C}^n) \qquad \text{for } x, y \in V. \tag{23}$$

The result is a $\mathbb{C}$-linear operator $T_{\mathbb{C}}$ whose characteristic polynomial $p_{T_{\mathbb{C}}}(t) \in \mathbb{C}[t]$ turns out to be the same as $p_T(t)$ when we view $p_T \in \mathbb{R}[t]$ as a polynomial in $\mathbb{C}[t]$ that happens to have real coefficients. Since $p_{T_{\mathbb{C}}}(t)$ always splits over $\mathbb{C}$, all the preceding theory applies to $T_{\mathbb{C}}$. Our task is then to translate that theory back to the original real linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$.

1.1. Exercise. If T is a linear operator from $\mathbb{R}^n \to \mathbb{R}^n$ and $T_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$ is its complexification as defined in (23), verify that the characteristic polynomials $p_T(t) \in \mathbb{R}[t]$ and $p_{T_{\mathbb{C}}}(t) \in \mathbb{C}[t]$ are "the same" – i.e. that

$$p_{T_{\mathbb{C}}}(t) = p_T(t)$$

when we identify $\mathbb{R}[t] \subseteq \mathbb{C}[t]$.
Hint: Polynomials in $\mathbb{F}[t]$ are equal $\Leftrightarrow$ they have the same coefficients in $\mathbb{F}$. Here, $p_{T_\mathbb{C}}$ has coefficients in $\mathbb{C}$ while p_T has coefficients in $\mathbb{R}$, but we are identifying $\mathbb{R} = \mathbb{R} + i0 \subseteq \mathbb{C}$. $\Box$

Relations between $\mathbb{R}[t]$ and $\mathbb{C}[t]$. When we view a real coefficient polynomial $f(t) = \sum_{j=0}^{m} a_j t^j \in \mathbb{R}[t]$ as an element of $\mathbb{C}[t]$ it can have complex roots as well as real roots, but

> *When we view $f(t) \in \mathbb{R}[x]$ as an element of $\mathbb{C}[x]$, any non-real roots must occur in conjugate pairs $z_\pm = (u \pm iv)$ with u and v real. Such eigenvalues can have nontrivial multiplicities, resulting in factors $(t - z_+)^m \cdot (t - z_-)^m$ in the irreducible factorization of $f(t)$ in $\mathbb{C}[t]$.*

In fact, if $z = x + iy$ with x, y real and if $f(z) = 0$, the complex conjugate $\overline{z} = x - iy$ is also a root of f because

$$f(\overline{z}) = \sum_{j=0}^{m} a_j \overline{z}^j = \sum_{j=0}^{m} a_j \overline{z^j} = \overline{\sum_j a_j z^j} = \overline{f(z)} = 0$$

(Recall that $\overline{z + w} = \overline{z} + \overline{w}$, $\overline{zw} = \overline{z} \cdot \overline{w}$, and $(\overline{z})^- = z$ for $z, w \in \mathbb{C}$.)

Thus, $\#$(*non-real roots*) is even, if any exist, while the number of real roots is unrestricted, and might be zero. Thus the splitting of f in $\mathbb{C}[t]$ can written as

$$\begin{aligned} f(t) &= c \cdot \prod_{j=1}^{r} (t - \mu_j)^{m_j} (t - \overline{\mu_j})^{m_j} \cdot \prod_{k=r+1}^{s} (t - r_j)^{m_k} \\ &= c \cdot \prod_{j=1}^{r} \left[(t - \mu_j)(t - \overline{\mu_j}) \right]^{m_j} \cdot \prod_{k=r+1}^{s} (t - r_j)^{m_k} \end{aligned} \tag{24}$$

where the μ_j are complex and nonreal ($\overline{\mu} \neq \mu$), and the r_j are the distinct real roots of f. Obviously $n = \deg(f) = \sum_{j=1}^{r} 2m_j + \sum_{j=r+1}^{s} m_j$. Since f has real coefficients, all complex numbers must disappear when the previous equation is multiplied out. In particular, for each nonreal conjugate pair $\mu, \overline{\mu}$ we have

$$Q_\mu(t) = (t - \mu)(t - \overline{\mu}) = t^2 - 2\mathrm{Re}(\mu) + |\mu|^2 \ , \tag{25}$$

a quadratic with real coefficients. Hence,

$$f(t) = c \cdot \prod_{r+1}^{s} (t - r_j)^{m_j} \cdot \prod_{j=1}^{r} (Q_{\mu_j}(t))^{m_j}$$

is a factorization of $f(t)$ into linear and *irreducible* quadratic factors in $\mathbb{R}[t]$, and every $f \in \mathbb{R}[t]$ can be decomposed this way. This is the (unique) decomposition of $f(t)$ into irreducible factors in $\mathbb{R}[t]$: by definition, the $Q_\mu(t)$ have no real roots and cannot be a product of polynomials of lower degree in $\mathbb{R}[t]$, while the linear factors are irreducible as they stand.

1.2. Definition. *A nonconstant polynomial $f \in \mathbb{F}[t]$ is* **irreducible** *if it cannot be factored as $f(t) = g(t)h(t)$ with g, h nonconstant and of lower degree than f. A polynomial is* **monic** *if its leading coefficient is 1. It is well known that every monic $f \in \mathbb{F}[t]$ factors uniquely as $\prod_{j=1}^{r} h_j(t)^{m_j}$ where each h_j monic and irreducible in $\mathbb{F}[t]$. The exponent $m_j \geq 1$ is its multiplicity, and this factorization is unique.*

The simplest irreducibles (over any $\mathbb{F}$) are the linear polynomials $at + b$ (with $a \neq 0$ since "irreducibility" only applies to nonconstant polynomials). This follows from the *degree formula*

$$\text{Degree Formula:}\quad \deg(gh) = \deg(g) + \deg(h) \quad \text{for all nonzero } g, h \in \mathbb{F}[t].$$

If we could factor $at+b = g(t)h(t)$, either $g(t)$ or $h(t)$ would have degree 0, and the other would have degree $1 = \deg(at+b)$. Thus there are no nontrivial factorization of $at+b$. When $\mathbb{F} = \mathbb{C}$, *all* irreducible polynomials have *degree* $= 1$, but if $\mathbb{F} = \mathbb{R}$ they can have *degree* $= 1$ or 2.

1.3. Lemma. *The irreducible monic polynomials in* $\mathbb{R}[t]$ *have the form*

1. $(t - r)$ *with* $r \in \mathbb{R}$, *or*
2. $t^2 + bt + c$ *with* $b^2 - 4c < 0$. *These are precisely the polynomials* $(t-\mu)(t-\overline{\mu})$ *with* μ *a non-real element in* $\mathbb{C}$.

Proof: Linear polynomials $at + b\,(a \neq 0)$ in $\mathbb{R}[t]$ are obviously irreducible. If f has the form of (2.), the quadratic formula applied on be applied to $f(t) = t^2 + bt + c$ in $\mathbb{C}[x]$ to find its roots

$$\mu, \overline{\mu} = \frac{-b \pm \sqrt{b^2 - 4c}}{2} = \frac{-b \pm i\sqrt{4c - b^2}}{2}$$

There are three possible outcomes:

1. $f(x)$ has a single real root with multiplicity $m = 2$ when $b^2 - 4c = 0$ and then we have $f(t) = (t - \frac{1}{2}b)^2$;
2. There are two distinct real roots $r_\pm = \frac{1}{2}\big(-b \pm \sqrt{b^2 - 4c}\big)$ when $b^2 - 4c > 0$, and then $f(t) = (t - r_+)(t - r_-)$;
3. When the discriminant $b^2 - 4c$ is negative there are two distinct conjugate nonreal roots in $\mathbb{C}$,

   $$\mu = \frac{-b + \sqrt{b^2 - 4c}}{2} \quad \text{and} \quad \overline{\mu} = \frac{-b - i\sqrt{4c - b^2}}{2} \qquad (i = \sqrt{-1}),$$

   in which case $f(t) = (t - \mu)(t - \overline{\mu})$ has the form (25) in $\mathbb{R}[t]$.

The quadratic $f(t)$ is irreducible in $\mathbb{R}[t]$ when $f(t)$ has two nonreal roots; otherwise it would have a factorization $(t - r_1)(t - r_2)$ in $\mathbb{R}[t]$ and also in $\mathbb{C}[t]$. That would disagree with $(x - \mu)(x - \overline{\mu})$, contrary to unique factorization in $\mathbb{C}[t]$, and cannot occur. $\Box$

Complexification of Arbitrary Linear Operators over $\mathbb{R}$. We now discuss complexification of *arbitrary* vector spaces over $\mathbb{R}$ and complexifications $T_\mathbb{C}$ of the $\mathbb{R}$-linear operators $T : V \to V$ that act on them.

1.4. Definition. (Complexification). *Given an arbitrary vector space* V *over* $\mathbb{R}$ *its* **complexification** $V_\mathbb{C}$ *is the set of symbols* $\{\mathbf{z} = x + iy : x, y \in V\}$ *equipped with operations*

$$\begin{aligned}\mathbf{z} + \mathbf{w} = (x + iy) + (u + iv) &= (x + u) + i(y + v)\\ (a + ib)\cdot\mathbf{z} = (a + ib)(x + iy) &= (ax - by) + i(bx + ay),\ \textit{for } a + ib \in \mathbb{C}\end{aligned}$$

Two symbols $\mathbf{z} = (x + iy)$ *and* $\mathbf{z}' = (x' + iy')$ *designate the same element of* $V_\mathbb{C} \Leftrightarrow x' = x$ *and* $y' = y$.

1. *The* **real points** *in* $V_\mathbb{C}$ *are those of the form* $V + i0$. *This set is a vector space over* $\mathbb{R}$ (*but* NOT *over* $\mathbb{C}$), *because*

   $$\begin{aligned}(c + i0)\cdot(x + i0) &= (cx) + i0 \qquad \textit{for } c \in \mathbb{R}, x \in V\\ (x + i0) + (x' + i0) &= (x + x') + i0 \qquad \textit{for } x, x' \in V.\end{aligned}$$

   Clearly the operations (+) *and* (*scale by real scalar* c+i0) *match the usual operations in* V *when restricted to* $V + i0$.

2. *If $T : V \to V$ is an $\mathbb{R}$-linear operator its* **complexification** $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$ *is defined to be the map*

$$T_{\mathbb{C}}(x+iy) = T(x) + iT(y),\ \ x, y \in V\ , \tag{26}$$

which turns out to be a $\mathbb{C}$-linear operator on $V_{\mathbb{C}}$.

We indicate this by writing "$T_{\mathbb{C}} = T + iT$."

1.5. Exercise. If $M_1, \ldots, M_r$ are vector spaces over $\mathbb{R}$ prove that the complexification of $V = M_1 \oplus \ldots \oplus M_r$ is $V_{\mathbb{C}} = (M_1)_{\mathbb{C}} \oplus \ldots \oplus (M_r)_{\mathbb{C}}$. $\square$

1.6. Exercise. Prove that $V_{\mathbb{C}}$ is actually a vector space over $\mathbb{C}$. (Check each of the vector space Axioms.)
Hint: In particular, you must check that $(z_1 z_2) \cdot w = z_1 \cdot (z_2 \cdot w)$, for $z_1, z_2 \in \mathbb{C}$ and $w \in V_{\mathbb{C}}$, and that $(c + i0) scdot (x + i0) = (c \cdot x) + i0$ for $c \in \mathbb{R}$, so $V + i0$ is a subspace over $\mathbb{R}$ isomorphic to V. $\square$

1.7. Example. We verify that

1. $T_{\mathbb{C}}$ is in fact a $\mathbb{C}$-linear operator on the complex vector space $V_{\mathbb{C}}$.
2. When we identify V with the subset $V + i0$ in $V_{\mathbb{C}}$ via the map $j : v \mapsto v + i0$, the restriction $T_{\mathbb{C}}|_{V+i0}$ gets identified with the original operator T in the sense that

$$T_{\mathbb{C}}(v + i0) = T(v) + i \cdot 0 \quad \text{for all } v \in V\ .$$

Thus the folowing diagram is commutative, with $T_{\mathbb{C}} \circ j = j \circ T$

$$\begin{array}{ccc} V & \xrightarrow{j} & V + i0 \subseteq V_{\mathbb{C}} \\ T \downarrow & & \downarrow T_{\mathbb{C}} \\ V & \xrightarrow{j} & V + i0 \subseteq V_{\mathbb{C}} \end{array} \qquad \text{with } T_{\mathbb{C}} \circ j = j \circ T\ .$$

Discussion: Commutativity of the diagram is immediate from the definitions of $V_{\mathbb{C}}$ and $T_{\mathbb{C}}$. The messy part of proving (1.) is showing that $T_{\mathbb{C}}(z \cdot \mathbf{w}) = z \cdot T_{\mathbb{C}}(\mathbf{w})$ for $z \in \mathbb{C}, \mathbf{w} \in V_{\mathbb{C}}$, so we will only do that. If $z = a + ib \in \mathbb{C}$ and $\mathbf{w} = u + iv$ in $V_{\mathbb{C}}$ we get

$$\begin{aligned} T_{\mathbb{C}}\big((a+ib)(u+iv)\big), &= T_{\mathbb{C}}\big((au - bv) + i(bu + av)\big) \\ &= T(au - bv) + iT(bu + av) \ = \ aT(u) - bT(v) + ibT(u) + iaT(v) \\ &= (a + ib) \cdot (T(u) + iT(v)) \ = \ (a + ib) \cdot T_{\mathbb{C}}(u + iv) \quad \square \end{aligned}$$

1.8. Example. If $\{e_j\}$ is an $\mathbb{R}$-basis in V, then $\{\tilde{e_j} = e_j + i0\}$ is a $\mathbb{C}$-basis in $V_{\mathbb{C}}$. In particular, $\dim_{\mathbb{R}}(V) = \dim_{\mathbb{C}}(V_{\mathbb{C}})$.

Discussion: If $w = v + iw$ $(v, w \in V)$ there are real coefficients $\{c_j\}, \{d_j\}$ such that

$$w = \Big(\sum_j c_j e_j\Big) + i\Big(\sum_j d_j e_j\Big) = \sum_j (c_j + id_j)(e_j + i0)\ ,$$

so the $\{\tilde{e_j}\}$ span $V_{\mathbb{C}}$. As for independence, if we have

$$0 + i0 = \sum z_j \tilde{e_j} = \sum (c_j + id_j) \cdot (e_j + i0) = \Big(\sum_j c_j e_j\Big) + i\Big(\sum_j d_j e_j\Big)$$

in $V_{\mathbb{C}}$ for coefficients $z_j = c_j + id_j$ in $\mathbb{C}$, then $\sum_j c_j e_j = 0 = \sum_j d_j e_j$, which implies $c_j = 0$ and $d_j = 0$ because $\{e_j\}$ is a basis in V. Thus $z_j = 0$ for all j. $\square$

1.9. Example. If $V = \mathbb{R}^n$ then $V_{\mathbb{C}} = \mathbb{R}^n + i\mathbb{R}^n$ is, in a obvious sense, the same as $\mathbb{C}^n$.

If $A \in \mathrm{M}(n,\mathbb{R})$, we get a $\mathbb{R}$-linear operator $T = L_A$ that maps $v \to A \cdot v$ (matrix product of $n \times n$ times $n \times 1$), whose matrix with respect to the standard basis $\mathfrak{X} = \{e_j\}$ in $\mathbb{R}^n$ is $[T]_{\mathfrak{X}} = A$. If $\{\tilde{e_j} = e_j + i0\} = \mathfrak{Y}$ is the corresponding basis in $V_{\mathbb{C}}$, it is easy to check that we again have $[T_{\mathbb{C}}]_{\mathfrak{Y}} = A$ – i.e. $T_{\mathbb{C}}$ is obtained by letting the matrix A with real entries act on complex column vectors by matrix multiply (regarding A as a complex matrix that happens to have real entries). $\square$

1.10. Definition (Conjugation). *The* **conjugation operator** $J : V_{\mathbb{C}} \to V_{\mathbb{C}}$ *maps* $x + iy \to x - iy$. *It is an* $\mathbb{R}$*-linear operator on* $V_{\mathbb{C}}$*, with*

$$J(c \cdot w) = c \cdot J(w) \qquad \textit{if } c = c + i0 \in \mathbb{R} \ ,$$

but is **conjugate linear** *over* $\mathbb{C}$*, with*

$$\begin{aligned} J(z \cdot w) &= \overline{z} \cdot J(w) \quad \textit{for } z \in \mathbb{C}, w \in V_{\mathbb{C}} \\ J(w_1 + w_2) &= J(w_1) + J(w_2) \end{aligned}$$

Further properties of conjugation are easily verified from this definition:

(27)
1. $J^2 = J \circ J = \mathrm{id}$, so $J^{-1} = J$.
2. $w \in V_C$ is a real point if and only if $J(w) = w$.
3. $\dfrac{w + J(w)}{2} = x + i0$ and $\dfrac{w - J(w)}{2i} = y + i0$, if $w = x + iy$ in $V_{\mathbb{C}}$.

The operator J can be used to identity the $\mathbb{C}$-linear maps $S : V_{\mathbb{C}} \to V_{\mathbb{C}}$, of **real type**, those such that $S = T_{\mathbb{C}} = T + iT$ for some $\mathbb{R}$-linear $T : V \to V$.

1.11. Exercise. Whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, a matrix in $\mathrm{M}(n,\mathbb{F})$ determines a linear operator $L_A : \mathbb{F}^n \to \mathbb{F}^n$. Verify the following relationships between operators on $V = \mathbb{R}^n$ and $V_{\mathbb{C}} = \mathbb{C}^n = \mathbb{R}^n + i\mathbb{R}^n$.

1. If $\mathbb{F} = \mathbb{R}$, $\big(L_A\big)_{\mathbb{C}} = L_A + iL_A : \mathbb{C}^n \to \mathbb{C}^n$ is the same as the operator $L_A : \mathbb{C}^n \to \mathbb{C}^n$ we get by regarding A as a complex matrix all of whose entries are real.
2. Consider $A \in \mathrm{M}(n,\mathbb{C})$ and regard $\mathbb{C}^n$ as the complexification of $\mathbb{R}^n$. Verify that $L_A : \mathbb{C}^n \to \mathbb{C}^n$ is of *real type* $\Leftrightarrow$ all entries in A are real, so $A \in \mathrm{M}(n,\mathbb{R})$.
3. If S and T are $\mathbb{R}$-linear operators on a real vector space V, is the map

   $$(S + iT) : (x + iy) \to S(x) + iT(y)$$

   on $V_{\mathbb{C}}$ a $\mathbb{C}$-linear operator? If so, when is it of real-type? $\square$

1.12. Exercise. If $T : V \to V$ is an $\mathbb{R}$-linear operator on a real vector space, prove that

1. $\big(T_{\mathbb{C}}\big)^k = \big(T^k\big)_{\mathbb{C}}$ for all $k \in \mathbb{N}$
2. $e^{(T_{\mathbb{C}})} = \big(e^T\big)_{\mathbb{C}}$ on $V_{\mathbb{C}}$ $\square$

1.13. Lemma. *If* $T : V \to V$ *is* $\mathbb{R}$*-linear and* $T_{\mathbb{C}}$ *is its complexification, then* $T_{\mathbb{C}}$ *commutes with* J*,*

$$JT_{\mathbb{C}} = T_{\mathbb{C}}J \quad (\textit{or equivalently } JT_{\mathbb{C}}J = T_{\mathbb{C}}).$$

Conversely, if $S : V_{\mathbb{C}} \to V_{\mathbb{C}}$ *is any* $\mathbb{C}$*-linear operator the following statements are equivalent*

1. $S = T_{\mathbb{C}} = T + iT$ *for some real-linear* $T : V \to V$.

2. $SJ = JS$.

3. S *leaves invariant the* $\mathbb{R}$*-subspace of real vectors* $V + i0$ *in* $V_{\mathbb{C}}$.

Proof: For (1.) $\Rightarrow$ (2.) is trivial:

$$\begin{aligned} JT_{\mathbb{C}}J(x+iy) &= JT_{\mathbb{C}}(x-iy) \;=\; J(T(x)+iT(-y)) \;=\; J(T(x)-iT(y)) \\ &= T(x)+iT(y) \;=\; T_{\mathbb{C}}(x+iy) \end{aligned}$$

For (2.) $\Rightarrow$ (3.), suppose $SJ = JS$. We have $w \in V + i0$ if and only if $w = \frac{1}{2}\big(w + J(w)\big)$, and for these w we have

$$S(w) = \tfrac{1}{2}\big(S(w) + S(J(w)\big) = \tfrac{1}{2}\big(J(S(w)) + S(w)\big)\ .$$

By the properties (27), $S(w)$ is a vector in $V + i0$.

For (3.) $\Rightarrow$ (1.), if S leaves $V + i0$ invariant then $S(x + i0) = T(x) + i0$ for some uniquely determined vector $T(x) \in V$. We claim that $T : x \mapsto T(x)$ is an $\mathbb{R}$-linear map. In fact, if $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in V$, we have

$$S\big((c_1x_1 + c_2x_2) + i0\big) = T(c_1x_1 + c_2x_2) + i0\ ,$$

while S (being $\mathbb{C}$-linear) must also satisfy the identities

$$\begin{aligned} S((c_1x_1+c_2x_2)+i0) &= S((c_1x_1+i0)+(c_2x_2+i0)) \;=\; S(c_1x_1+i0)+S(c_2x_2+i0) \\ &= S((c_1+i0)\cdot(x_1+i0)+(c_2+i0)\cdot(x_2+i0)) \\ &= (c_1+i0)\cdot(Tx+i0)+(c_2+i0)\cdot(T(x_2)+i0) \\ &= \big(c_1T(x_1)+c_2T(x_2)\big)+i0 \end{aligned}$$

Thus T is $\mathbb{R}$-linear on V. Furthermore, $S = T_{\mathbb{C}}$ because

$$\begin{aligned} T_{\mathbb{C}}(x+iy) &= T(x)+iT(y) = (T(x)+i0)+i(T(y)+i0) \\ &= S(x+i0)+iS(y+i0) \qquad \text{(by } \mathbb{C}\text{-linearity of } S \text{ and definition of } T) \\ &= S\big((x+i0)+i(y+i0)\big) \;=\; S(x+iy) \end{aligned}$$

Thus $S : V_{\mathbb{C}} \to \mathbb{C}$ is of real type if and only if $JS = SJ$, and then $S = \big(S|_{V+i0)}\big)_{\mathbb{C}}$. $\square$

An Application. The complexified operator $T_{\mathbb{C}}$ acts on a complex vector space $V_{\mathbb{C}}$ and therefore can be put into Jordan form (or perhaps diagonalized) by methods worked out previously. We now examine the special case when $T_{\mathbb{C}}$ is diagonalizable, before taking up the general problem: if $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$ is diagonalizable, what can be said about the structure of the original $\mathbb{R}$-linear operator $T : V \to V$?

We start with an observation that holds for any $\mathbb{R}$-linear operator $T : V \to V$, whether or not $T_{\mathbb{C}}$ is diagonalizable.

1.14. Lemma. *If* $p_T(t) = \sum_{j=0}^{m} a_j t^j$ $(a_j \in \mathbb{R})$*, then* $p_{T_{\mathbb{C}}} = p_T$ *in the sense that* $p_{T_{\mathbb{C}}}(t) = \sum_{j=0}^{m}(a_j + i0)t^j$ *in* $\mathbb{C}[t] \supseteq \mathbb{R}[t]$.

Proof: Earlier we proved that if $\mathfrak{X} = \{e_j\}$ is an $\mathbb{R}$-basis in V then $\mathfrak{Y} = \{\tilde{e_j} = e_j + i0\}$ is a $\mathbb{C}$-basis in $V_{\mathbb{C}}$, and that $[T]_{\mathfrak{X}} = [T_{\mathbb{C}}]_{\mathfrak{Y}}$ because

$$\begin{aligned} T_{\mathbb{C}}(\tilde{e_j}) &= T_{\mathbb{C}}(e_j+i0) \;=\; T(e_j)+iT(0) \;=\; \Big(\sum_k t_{kj}\cdot e_k\Big)+i0 \\ &= \sum_k (t_{kj}+i0)(e_k+i0) \;=\; \sum t_{kj}\tilde{e_j} \end{aligned}$$

Thus $[T_{\mathbb{C}}]_{ij} = t_{ij} = [T]_{ij}$. Subtracting tI and taking the determinant, the outcome is the same whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $\square$

Hereafter we write p_T for $p_{T_{\mathbb{C}}}$ leaving the context to determine whether p_T is to be regarded as an element of $\mathbb{R}[x]$ or $\mathbb{C}[x]$. As noted earlier, p_T always splits in $\mathbb{C}[x]$, but might not split as an element of $\mathbb{R}[x]$. Furthermore, the nonreal roots in the factorization (24) come in conjugate pairs, so we may list the eigenvalues of $T_{\mathbb{C}}$ as follows, selecting one representative μ from each conjugate pair $\mu, \overline{\mu}$

$$\mu_1, \overline{\mu_1}, \cdots, \mu_r, \overline{\mu_r}; \lambda_{r+1}, \cdots, \lambda_s, \qquad \text{with } \lambda_i \text{ real and } \mu_j \neq \overline{\mu_j} \quad . \tag{28}$$

and repeating eigenvalues/pairs according to their multiplicity in $p_{T_{\mathbb{C}}}(t)$.

Now assume $T_{\mathbb{C}}$ is diagonalizable, with (complex) eigenspaces $E_{\lambda_i}, E_{\mu_j}, E_{\overline{\mu_j}}$ in $V_{\mathbb{C}}$ that yield a direct sum decomposition of $V_{\mathbb{C}}$. Now observe that if $\mu \neq \overline{\mu}$ then $w \in V_{\mathbb{C}}$ is an eigenvector for μ if and only if $J(w)$ is an eigenvector for $\overline{\mu}$ because

$$T_{\mathbb{C}}(J(w)) = J\big(T_{\mathbb{C}}(w)\big) = J(\mu w) = \overline{\mu} J(w) \tag{29}$$

Hence, $J(E_\mu(T_{\mathbb{C}})) = E_{\overline{\mu}}(T_{\mathbb{C}})$ and J is an $\mathbb{R}$-linear bijection between $E_\mu(T_{\mathbb{C}})$ and $E_{\overline{\mu}}(T_{\mathbb{C}})$. Observe that $J(E_\mu \oplus E_{\overline{\mu}}) = E_\mu \oplus E_{\overline{\mu}}$ even though neither summand need be J-invariant (although we do have $J(E_\lambda) = E_\lambda$ when λ is a real eigenvalue for $T_{\mathbb{C}}$). The complex subspaces $E_\mu \oplus E_{\overline{\mu}}$ are of a special "real type " in $V_{\mathbb{C}}$ owing to their conjugation-invariance.

1.15. Definition. *If W is a $\mathbb{C}$-subspace of the complexification $V_{\mathbb{C}} = V + iV$, its* **real points** *are those in $W_{\mathbb{R}} = W \cap (V + i0)$. This is a vector space over $\mathbb{R}$ that determines a complex subspace $(W_{\mathbb{R}})_{\mathbb{C}} \subseteq W$ by taking $\mathbb{C}$-linear combinations.*

$$(W_{\mathbb{R}})_{\mathbb{C}} = W_{\mathbb{R}} + iW_{\mathbb{R}} \subseteq W$$

In general, $W_{\mathbb{R}} + iW_{\mathbb{R}}$ can be a lot smaller than the original complex subspace W. We say that a complex subspace $W \subseteq V_{\mathbb{C}}$ is of **real-type** *if*

$$W = W_{\mathbb{R}} + iW_{\mathbb{R}}$$

where $W_{\mathbb{R}} = W \cap (V + i0)$ is the set of real points in W.

Thus a complex subspace of real type W is the complexification of its subspace of real points $W_{\mathbb{R}}$.

Subspaces of real type are easily identified by their conjugation-invariance.

1.16. Lemma. *A complex subspace W in a complexification $V_{\mathbb{C}} = V + iV$ is of real type if and only if $J(W) = W$.*

Proof: $W = W_{\mathbb{R}} + iW_{\mathbb{R}}$ so $J(W) = W_{\mathbb{R}} - iW_{\mathbb{R}} = W_{\mathbb{R}} + iW_{\mathbb{R}}$ since $W_{\mathbb{R}} = -W_{\mathbb{R}}$, proving ($\Rightarrow$). Conversely, for ($\Leftarrow$) : if $J(W) = W$ and we write $w \in W$ as $w = x + iy$ $(x, y \in V)$, both

$$\tfrac{1}{2}(w + J(w)) = x + i0, \quad \text{and} \quad \frac{1}{2i}(w - J(w)) = y - i0$$

are in $V + i0$, and both are in $W_{\mathbb{R}} = W \cap (V + i0)$. Since $w = (x + i0) + i(y + i0) = x + iy$, we conclude that $w \in W_{\mathbb{R}} + iW_{\mathbb{R}}$, so W is of real type. $\square$

The spaces $W = E_\mu \oplus E_{\overline{\mu}}$ (and $W = E_\lambda$ for real λ) are obviously of real type since $J(E_\mu) = E_{\overline{\mu}}$. We now compare what is happening in a complex subspace W with what goes on in the real subspace $W_{\mathbb{R}}$. Note that $T_{\mathbb{C}}(W_{\mathbb{R}}) \subseteq W_{\mathbb{R}}$ because

$$\begin{aligned} T_{\mathbb{C}}(W_{\mathbb{R}}) &= T_{\mathbb{C}}(W \cap (V + i0)) \;=\; T_{\mathbb{C}}(W) \cap T_{\mathbb{C}}(V + i0) \\ &\subseteq W \cap (T(V) + i0) \;\subseteq\; W \cap (V + i0) \;=\; W_{\mathbb{R}} \end{aligned}$$

Case 1: $\lambda \in \mathrm{sp}(T_{\mathbb{C}})$ is real. Proposition 8.17 below shows that $E_\lambda(T_{\mathbb{C}}) = \big(E_\lambda(T)\big)_{\mathbb{C}}$ if $\lambda \in \mathbb{R}$. We have previously seen that an arbitrary $\mathbb{R}$-basis $\{f_j\}$ for $E_\lambda(T)$ corresponds to a $\mathbb{C}$-basis $\tilde{f}_j = f_j + i0$ in $\big(E_\lambda(T)\big)_{\mathbb{C}} = E_\lambda(T_{\mathbb{C}})$. But

$$T_{\mathbb{C}}(\tilde{f}_j) = \lambda \cdot \tilde{f}_j = \lambda \cdot (f_j + i0) = \lambda f_j + i0 \ ,$$

since $\tilde{f}_j \in E_\lambda(T_{\mathbb{C}})$, while

$$T_{\mathbb{C}}(\tilde{f}_j) = (T + iT)(f_j + i0) = T(f_j) + i0 \ .$$

Hence $T(f_j) = \lambda f_j$ and $T = T_{\mathbb{C}}|_{(E_\lambda(T)+i0)}$ is diagonalizable over $\mathbb{R}$.

1.17. Proposition. *If $T : V \to V$ is a linear operator on a real vector space and λ is a real eigenvalue for $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$, then $\lambda \in \mathrm{sp}_{\mathbb{R}}(T)$ and*

$$E_\lambda(T_{\mathbb{C}}) = E_\lambda(T) + iE_\lambda(T) = (E_\lambda(T))_{\mathbb{C}}$$

In particular, $\dim_{\mathbb{C}}(E_\lambda(T)) = \dim_{\mathbb{R}}(E_\lambda(T))$ for real eigenvalues of T.

Proof: $\lambda + i0 \in \mathrm{sp}_{\mathbb{C}}(T_{\mathbb{C}}) \cap (\mathbb{R} + i0)$ if and only if there is a vector $u + iv \neq 0$ in $V_{\mathbb{C}}$ such that $T_{\mathbb{C}}(u+iv) = (\lambda+i0)(u+iv) = \lambda u + i\lambda v$. But because $T_{\mathbb{C}}(u+iv) = T(u)+iT(v)$ this happens if and only if $T(v) = \lambda v$ and $T(u) = \lambda u$, and since at least one of the vectors $u, v \in V$ is nonzero we get $\lambda + i0 \in \mathrm{sp}_{\mathbb{C}}(T_{\mathbb{C}}) \cap (\mathbb{R} + i0) \subseteq \mathrm{sp}_{\mathbb{R}}(T)$.

Conversely, suppose $x + iy \in E_\lambda(T_{\mathbb{C}})$ for real λ. Then

$$T_{\mathbb{C}}(x + iy) = (\lambda + i0)(x + iy) \ = \ \lambda x + i\lambda y \ ,$$

but we also have

$$T_{\mathbb{C}}(x + iy) = T_{\mathbb{C}}(x + i0) + iT_{\mathbb{C}}(y + i0) \ = \ T(x) + iT(y)$$

because $T_{\mathbb{C}}$ is $\mathbb{C}$-linear. This holds if and only if $T(x) = \lambda x$ and $T(y) = \lambda y$, so that

$$x + iy \in E_\lambda(T) + iE_\lambda(T) = (E_\lambda(T))_{\mathbb{C}} \ . \quad \square$$

1.18. Corollary. *We have $\mathrm{sp}_{\mathbb{C}}(T_{\mathbb{C}}) \cap (\mathbb{R} + i0) = \mathrm{sp}_{\mathbb{R}}(T)$ for any $\mathbb{R}$-linear operator $T : V \to V$ on a finite dimensional real vector space.*

1.19. Exercise. Let $V_{\mathbb{C}} = V + iV$ be the complexification of a real vector space V and let $S : V_{\mathbb{C}} \to V_{\mathbb{C}}$ be a $\mathbb{C}$-linear operator of real type, $S = T_{\mathbb{C}} = T + iT$ for some $T : V \to V$. Let $W = W_{\mathbb{R}} + iW_{\mathbb{R}} \subseteq V_{\mathbb{C}}$ be a complex subspace of real type that is S-invariant. Verify that

$$\text{(a)} \ \ S(W_{\mathbb{R}} + i0) \subseteq (W_{\mathbb{R}} + i0) \qquad \text{and} \qquad \text{(b)} \ \ S|_{(W_{\mathbb{R}}+i0)} = (T|_{W_{\mathbb{R}}}) + i0 \ . \quad \square$$

This will be the key to determining structure of an $\mathbb{R}$-linear operator $T : V \to V$ from that of its complexification $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$.

Consider now the situation not covered by Case 1 above.

Case 2: Nonreal conjugate pairs $\mu, \overline{\mu}$. The space $E_{\mu,\overline{\mu}} = E_\mu(T_{\mathbb{C}}) \oplus E_{\overline{\mu}}(T_{\mathbb{C}})$ is of real type and $T_{\mathbb{C}}$-invariant; $T_{\mathbb{C}}$ is an operator of real type on $V_{\mathbb{C}}$ by definition. Let us list the pairs of non-real eigenvalues $\mu, \overline{\mu}$ according to their multiplicities as in (28), and let

$$f_j^{(\mu)} = x_j^\mu + iy_j^\mu \qquad \text{with } (x_j, y_j \in V)$$

be a $\mathbb{C}$-basis for $E_\mu(T_{\mathbb{C}})$, so that

$$E_\mu(T_{\mathbb{C}}) = \bigoplus_{j=1}^{d} \mathbb{C} f_j^{(\mu)} \qquad \text{and} \qquad T_{\mathbb{C}}|_{E_\mu} = \mu \cdot I_{E_\mu} \ .$$

Since $J(E_\mu) = E_{\overline{\mu}}$, we get a matching $\mathbb{C}$-basis in $E_{\overline{\mu}} = \bigoplus_{j=1}^{d} \mathbb{C}J(f_j^{(\mu)})$ using (29).

$$J(f_j^{(\mu)}) = x_j^\mu - ix_j^\mu$$

Then for $1 \le i \le d = \dim_{\mathbb{C}}(E_\mu)$ define the 2-dimensional complex subspaces in $E_{\mu,\overline{\mu}}$

$$V_j^\mu = \mathbb{C}f_j^{(\mu)} \oplus \mathbb{C}J(f_j^{(\mu)}),\ j = 1, 2, \cdots, d = \dim_{\mathbb{C}}(E_\mu(T_{\mathbb{C}}))$$

These are $T_{\mathbb{C}}$-invariant and are of real type since they are J-invariant by definition. Clearly $E_\mu \oplus E_{\overline{\mu}} = \bigoplus_{j=1}^{d} V_j^\mu$. We claim that for each V_j^μ, we can find a $\mathbb{C}$-basis consisting of two *real* vectors in $(V_j^\mu)_{\mathbb{R}} = \big(V_j^\mu\big) \cap (V + i0)$ (something that cannot be done for the spaces $\mathbb{C}f_j^{(\mu)}$ or E_μ alone).

We prove $W = W_{\mathbb{R}} + iW_{\mathbb{R}}$. If $f_j^{(\mu)} = x_j + iy_j$, $J(f_j^{(\mu)}) = x_j - iy_j$, then $x_j = x_j + i0$ and $y_j = y_j + i0$ are in $(V_j^\mu)_{\mathbb{R}}$ but their $\mathbb{C}$-span includes $f_j^{(\mu)}$ and $J\big(f_j^{(\mu)}\big)$, and is obviously all of $V_j^{(\mu)}$; these real vectors are a $\mathbb{C}$-basis for $V_j^{(\mu)}$. They are also an $\mathbb{R}$-basis for the 2-dimensional space $(V_j^\mu)_{\mathbb{R}} = (\mathbb{R}x_j + \mathbb{R}y_j) + i0$ of real points in $V_j^{(\mu)}$.

Note that $x_j + i0 \in (V_j^\mu)_{\mathbb{R}}$ can be written as

$$\begin{aligned} x_j + i0 &= \tfrac{1}{2}\big(f_j^{(\mu)} + J(f_j^{(\mu)})\big), \text{ and similarly} \\ y_j + i0 &= \frac{1}{2i}\big(f_j^{(\mu)} - J(f_j^{(\mu)})\big) . \end{aligned}$$

As previously noted, $T_{\mathbb{C}}$ (resp. T) leaves V_j^μ (resp. $(V_j^\mu)_{\mathbb{R}}$) invariant.

We now determine the matrix of $T_j = T|_{(V_j^\mu)_{\mathbb{R}}}$ with respect to the ordered $\mathbb{R}$-basis $\mathfrak{X}_j = \{x_j^{(\mu)}, y_j^{(\mu)}\}$. If $\mu = a + ib$ with a, b real and $b \ne 0$, then $\overline{\mu} = a - ib$; suppressing the superscript "μ" for clarity, we then have

$$\begin{aligned} T_{\mathbb{C}}(x_j + iy_j) &= \mu(x_j + iy_j) \;=\; (a + ib)(x_j + iy_j) \;=\; (ax_j - by_j) + i(ay_j + bx_j) \\ T_{\mathbb{C}}(x_j - iy_j) &= \bar{\mu}(x_j - iy_j) \;=\; (a + ib)(x_j - iy_j) \;=\; (ax_j - by_j) - i(ay_j + bx_j) \end{aligned}$$

Write μ in polar form

$$\mu, \overline{\mu} = a \pm ib = re^{\pm i\theta} = r\cos(\theta) \pm ir\sin(\theta) .$$

$T_{\mathbb{C}}$ and J commute because $T_{\mathbb{C}}$ is of real type $T_{\mathbb{C}}$, and since $J(zw) = \overline{z}J(w)$ for $z \in \mathbb{C}$ we get

$$\begin{aligned} T(x_j) + i0 &= T_{\mathbb{C}}(x_j + i0) \;=\; T_{\mathbb{C}}\Big(\frac{f_j + J(f_j)}{2}\Big) \;=\; \frac{T_{\mathbb{C}}(f_j) + J(T_{\mathbb{C}}(f_j))}{2} \\ &= \tfrac{1}{2}\big[\,(a + ib)f_j + J\big((a + ib)f_j\big)\big] \\ &= \tfrac{1}{2}\big[\,(a + ib)(x_j + iy_j) + (a - ib)\cdot(x_j - iy_j)\,\big] \\ &= (ax_j - by_j) \\ &= x_j \cdot r\cos(\theta) - y_j \cdot r\sin(\theta) \end{aligned}$$

Similarly, we obtain

$$T(y_j) + i0 = T_{\mathbb{C}}\Big(\frac{w_j - J(w_j)}{2i}\Big) = (ay_j + bx_j) = x_j \cdot r\sin(\theta) + y_j \cdot r\cos(\theta)$$

We previously proved that $[T_{\mathbb{C}}]_{\{\tilde{e_i}\}} = [T]_{\{e_i\}}$ for any $\mathbb{R}$-basis in $\big(V_j^{(\mu)}\big)_{\mathbb{R}}$, so the matrix of $T_j : (V_j^\mu)_{\mathbb{R}} \to (V_j^\mu)_{\mathbb{R}}$ with respect to the $\mathbb{R}$-basis $\mathfrak{X}_j = \{x_j, y_j\}$ in $(V_j^{(\mu)})_{\mathbb{R}}$ is

$$\big[T|_{(V_j^\mu)_{\mathbb{R}}}\big]_{\mathfrak{X}_j} = r \cdot \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Reversing order of basis vectors yields basis $\mathfrak{X}'i_j = \{y_j^\mu, x_j^\mu\}$ such that the matrix of $T|_{(V_j^\mu)_\mathbb{R}}$ is a scalar multiple $r \cdot R(\theta)$ of the rotation matrix $R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ that corresponds to a counterclockwise rotation about the origin by θ radians, with $\theta \neq \pi, 0 \pmod{2\pi}$ because $\mu = a + ib$ is nonreal ($b \neq 0$).

Of course, with respect to the complex basis $\mathfrak{Y} = \{J(f_j^{(\mu)}), f_j^{(\mu)}\} = \{x_j - iy_j, x_j + iy_j\}$ in V_j^μ the operator $T_\mathbb{C}|_{V_j^\mu}$ is represented by the complex diagonal matrix

$$[T_\mathbb{C}|_{V_j^\mu}]_\mathfrak{Y} = \begin{pmatrix} r\,e^{-i\theta} & 0 \\ 0 & r\,e^{i\theta} \end{pmatrix} = \begin{pmatrix} \overline{\mu} & 0 \\ 0 & \mu \end{pmatrix}$$

To summarize: we have

$$\begin{aligned} V_\mathbb{C} &= \Big[\bigoplus_{\lambda \text{ real}} \big(\bigoplus_{j=1}^{d(\lambda)} \mathbb{C} \cdot f_j^{(\lambda)} \big) \Big] \oplus \Big[\bigoplus_{\mu \neq \overline{\mu} \text{ nonreal}} \big(\bigoplus_{j=1}^{d(\mu)} V_j^{(\mu)} \big) \Big] \\ &= \Big[\bigoplus_{\lambda \text{ real}} \big(E_\lambda(T) \big)_\mathbb{C} \Big] \oplus \Big[\bigoplus_{\mu \neq \overline{\mu} \text{ nonreal}} \big(\bigoplus_{j=1}^{d(\mu)} \big((V_j^{(\mu)})_\mathbb{R} \big)_\mathbb{C} \Big] \end{aligned}$$

where

$$V_j^{(\mu)} = \mathbb{C} f_j^{(\mu)} \oplus \mathbb{C} \cdot J(f_j^{(\mu)}) = \mathbb{C}(x_j^{(\mu)} + i0) \oplus \mathbb{C}(y_j^{(\mu)} + i0) \ ,$$

and all the spaces $\mathbb{C} f_j^{(\lambda)}$, $V_j^{(\mu)}$ are of real type. Restricting attention to the real points in $V_\mathbb{C}$ we arrive at a direct sum decomposition of the original real vector space V into T-invariant $\mathbb{R}$-subspaces

$$V = \Big[\bigoplus_{\lambda \text{ real}} E_\lambda(T) \Big] \oplus \Big[\bigoplus_{\mu \neq \overline{\mu} \text{ nonreal}} \big(\bigoplus_{j=1}^{d(\mu)} (V_j^{(\mu)})_\mathbb{R} \big) \Big] \tag{30}$$

We have arrived at the following decomposition of the $\mathbb{R}$-linear operator $T : V \to V$, when $T_\mathbb{C}$ is diagonalizable. Note: For each complex pair $(\mu, \overline{\mu})$ in $\mathrm{sp}(T_\mathbb{C})$, $E_\mu \oplus E_{\bar{\mu}}$ is of real type and we claim that $(E_\mu \oplus E_{\overline{\mu}}) \cap (V + i0) = \bigoplus_{j=1}^{d(\mu)} (V_j^\mu)_\mathbb{R}$. The sum on the right is direct so $\dim_\mathbb{R}(\bigoplus_{\mu, \overline{\mu}} \cdots) = 2d(\mu)$. Since we also have

$$\dim_\mathbb{R} \big(E_\mu \oplus E_{\overline{\mu}} \big)_\mathbb{R} = \dim_\mathbb{C} \big(E_\mu \oplus E_{\overline{\mu}} \big) = 2d(\mu) \ ,$$

the spaces coincide.

1.20. Theorem ($T_\mathbb{C}$ Diagonalizable). *If $T : V \to V$ is $\mathbb{R}$-linear and $T_\mathbb{C}$ is diagonalizable with eigenvalues labeled $\mu_1, \bar{\mu_1}, \cdots, \lambda_{r+1}, \cdots, \lambda_s$ as in* (28)*, there is an $\mathbb{R}$-basis $\mathfrak{X}$ such that $[T]_\mathfrak{X}$ has the block diagonal form*

$$\begin{pmatrix} \boxed{R_1} & & & & & & 0 \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \boxed{R_r} & & & \\ & & & & \boxed{D_1} & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ 0 & & & & & & \boxed{D_s} \end{pmatrix}$$

where

$$R_k = \begin{pmatrix} r_k R(\theta_k) & & 0 \\ & \cdot & \\ & & \cdot & \\ 0 & & r_k R(\theta_k) \end{pmatrix}$$

for $1 \leq k \leq r$, *and*

$$D_k = \begin{pmatrix} \lambda_k & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \lambda_k \end{pmatrix}$$

for $r+1 \leq k \leq s$.

Here, $\mu_k = r_k e^{i\theta_k}$ are representatives for the non-real pairs $(\mu, \overline{\mu})$ in $\mathrm{sp}(T_{\mathbb{C}})$.

When $T_{\mathbb{C}}$ is not diagonalizable, we apply the Jordan Canonical form for $T_{\mathbb{C}}$.

1.21. Lemma. *If* $T : V \to V$ *is a linear operator on a vector space over* $\mathbb{R}$, *let* $\mu \in \mathbb{C}$ *and* $M_\mu = \{w \in V_{\mathbb{C}} : (T_{\mathbb{C}} - \mu I)^k w = 0$ *for some* $k \in \mathbb{N}\}$. *Then* $w \in M_\mu \Leftrightarrow J(w) \in M_{\overline{\mu}}$, *so that*

$$J(M_\mu) = M_{\overline{\mu}} \quad \text{and} \quad J(M_{\overline{\mu}}) = M_\mu$$

Proof: The map $\Phi : S \mapsto JSJ = JSJ^{-1}$ is an *automorphism* of the algebra of all $\mathbb{C}$-linear maps $\mathrm{Hom}_{\mathbb{C}}(V_{\mathbb{C}}, V_{\mathbb{C}})$: it preserves products, $\Phi(S_1S_2) = \Phi(S_1)\Phi(S_2)$ because $J^2 = I \Rightarrow JS_1S_2J = (JS_1J)(JS_2J)$, and obviously $\Phi(S_1 + S_2) = \Phi(S_1) + \Phi(S_2)$, $\Phi(cS) = c\Phi(S)$ for $c \in \mathbb{C}$. In particular, $\Phi(S^k) = \Phi(S)^k$ for $k \in \mathbb{N}$. Then

$$J\big((T_{\mathbb{C}} - \mu I)^k\big)J = (J(T_{\mathbb{C}} - \mu I)J)^k = \big(JT_{\mathbb{C}}J - J(\mu I)J\big)^k = (T_{\mathbb{C}} - \overline{\mu} I)^k$$

($JT_{\mathbb{C}} = T_{\mathbb{C}}J$ because $T_{\mathbb{C}}$ is of real-type by definition), and

$$J(\mu I)J(w) = J(\mu \cdot J(w)) = \bar{\mu}J^2(w) = \bar{\mu}w, \ for\ w \in V_{\mathbb{C}}$$

Finally, we have $(T_{\mathbb{C}} - \mu I)^k w = 0$ if and only if

$$\begin{aligned} J(T_{\mathbb{C}} - \mu I)^k w = 0 \quad &\Leftrightarrow \quad J(T_{\mathbb{C}} - \mu I)J(J(w)) = 0 \\ &\Leftrightarrow \quad (T_{\mathbb{C}} - \overline{\mu})^k J(w) = 0 \Leftrightarrow J(w) \in M_{\overline{\mu}} \ . \end{aligned}$$

Hence $J(M_\mu) = M_{\overline{\mu}}$, which implies $M_\mu = J(M_{\overline{\mu}})$ since $J^2 = I$. $\square$

1.22. Theorem ($T_{\mathbb{C}}$ not Diagonalizable). *Let* $T : V \to V$ *lie a linear operator on a finite dimensional vector space over* $\mathbb{R}$ *and let* $\mu_1, \overline{\mu_1}, \cdots, \mu_r, \overline{\mu_r}, \lambda_{r+1}, \cdots, \lambda_s$ *be the eigenvalues of* $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$, *listed as in* (28). *Then there is an* $\mathbb{R}$*-basis for* V *that puts* $[T]_{\mathfrak{X}}$ *in block diagonal form:*

$$[T]_{\mathfrak{X}} = \begin{pmatrix} \boxed{A_1} & & 0 \\ & \cdot & \\ & & \cdot \\ 0 & & \boxed{A_m} \end{pmatrix},$$

in which each block A_j *has one of two possible block upper-triangular forms:*

$$A = \begin{pmatrix} \lambda & 1 & & 0 \\ & \cdot & \ddots & \\ & & \cdot & 1 \\ 0 & & & \lambda \end{pmatrix} \qquad \text{for real eigenvalues } \lambda \text{ of } T_{\mathbb{C}}\ ,$$

or

$$A = \begin{pmatrix} \boxed{R_{\mu,\overline{\mu}}} & I_2 & & 0 \\ & \cdot & \ddots & \\ & & \cdot & I_2 \\ 0 & & & \boxed{R_{\mu,\overline{\mu}}} \end{pmatrix} \qquad \textit{for conjugate pairs } \mu = e^{i\theta}, \overline{\mu} = e^{-i\theta}$$

where I_2 *is the* 2×2 *identity matrix and*

$$R_{\mu,\overline{\mu}} = r \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} .$$

Proof: The proof will of course employ the generalized eigenspace decomposition $V_{\mathbb{C}} = \bigoplus_{z \in \mathrm{sp}(T_{\mathbb{C}})} M_z(T_{\mathbb{C}})$. As above, we write $M_{\mu,\overline{\mu}} = M_\mu \oplus M_{\overline{\mu}}$ for each distinct pair of nonreal eigenvalues. The summands $M_\lambda, M_{\mu,\overline{\mu}}$ are of real type and are $T_{\mathbb{C}}$ invariant, so for each μ we may write $M_{\mu,\overline{\mu}}$ as the complexification of its space of real points

$$W_\mu = M_{\mu,\overline{\mu}} \cap (V + i0) \ . \tag{31}$$

Here $M_{\mu,\overline{\mu}} = W_\mu + iW_\mu$ is $T_{\mathbb{C}}$-invariant. Since $T_{\mathbb{C}}$ is of real type (by definition), it leaves invariant the $\mathbb{R}$-subspace $V + i0$; therefore the space of real points W_μ is also $T_{\mathbb{C}}$ invariant, with $T_{\mathbb{C}}|_{(W_\mu + i0)} = T|_{W_\mu}$. It follows that $T_{\mathbb{C}}|_{M_{\mu,\overline{\mu}}}$ is the complexification

$$\big(T|_{W_\mu}\big) + i\big(T|_{W_\mu}\big)$$

of T_{W_μ}. If we can find an $\mathbb{R}$-basis in the space (31) of real points $W_\mu \subseteq M_{\mu,\overline{\mu}}$ for which $T|_{W_\mu}$ takes the form described in the theorem, then our proof is complete. For this we may obviously restrict attention to a single subspace $M_{\mu,\overline{\mu}}$ in $V_{\mathbb{C}}$.

Case 1: A Real Eigenvalue λ. If λ is real then

$$\begin{aligned} (T_{\mathbb{C}} - \lambda)^k (x + iy) &= (T_{\mathbb{C}} - \lambda)^{k-1}\big[(T - \lambda)x + i(T - \lambda)y\big] = \cdots \\ &= (T - \lambda)^k x + i(T - \lambda)^k y \quad \text{for } k \in \mathbb{N} \ . \end{aligned}$$

Thus, $x + iy \in M_\lambda(T_{\mathbb{C}})$ if and only if $x, y \in M_\lambda(T)$, and the subspace of real points in $M_\lambda(T_{\mathbb{C}}) = M_\lambda(T) + iM_\lambda(T)$ is precisely $M_\lambda(T) + i0$. There is a $\mathbb{C}$-basis $\mathfrak{X} = \{f_j\}$ of real vectors in $M_\lambda(T_{\mathbb{C}})$ that yields the same matrix for $T_{\mathbb{C}}$ and for its restriction to this subspace of real points, and we have

$$T_{\mathbb{C}}|_{W_\lambda} = T|_{W_\lambda} = T|_{M_\lambda(T)} \ . \qquad \text{and} \qquad (T_{\mathbb{C}})|_{M_\lambda(T)} = T|_{M_\lambda(T)} \ .$$

Case 2: A Conjugate Pair $\mu, \overline{\mu}$. The space $M_{\mu,\overline{\mu}} = M_\mu(T_{\mathbb{C}}) \oplus M_{\overline{\mu}}(T_{\mathbb{C}})$ is of real type because $J(M_\mu) = M_{\overline{\mu}}$. In fact, if $v \in M_\mu$ then $(T_{\mathbb{C}} - \mu)^k v = 0$ for some k. But $T_{\mathbb{C}} J = J T_{\mathbb{C}}$, so

$$(T_{\mathbb{C}} - \overline{\mu})^k J(v) = (T_{\mathbb{C}} - \overline{\mu})^{k-1} J(T_{\mathbb{C}} - \mu)(v) = \cdots = J(T_{\mathbb{C}} - \mu)^k(v) = 0 \ .$$

Thus $M_{\mu,\overline{\mu}}$ is the complexification of its subspace of real points $V_\mu = M_{\mu,\overline{\mu}}(T_{\mathbb{C}}) \cap (V + i0)$.

By the Cyclic Subspace Decomposiiton (Theorem 3.2) a generalized eigenspace M_μ for $T_{\mathbb{C}}$ is a direct sum of $T_{\mathbb{C}}$-invariant spaces C_j that are cyclic under the action of the nilpotent operator $(T_{\mathbb{C}} - \mu I)$. In each C_j take a basis $\mathfrak{X}_j = \{f_1^{(j)}, \cdots, f_{d_j}^{(j)}\}$ that puts $(T_{\mathbb{C}} - \mu I)|_{C_j}$ into elementary nilpotent form

$$\begin{pmatrix} 0 & 1 & \cdot & 0 \\ & \cdot & \cdot & \cdot \\ & & \cdot & 1 \\ 0 & & & 0 \end{pmatrix} \qquad \text{so that} \qquad [T_{\mathbb{C}}] = \begin{pmatrix} \mu & 1 & \cdot & 0 \\ & \cdot & \cdot & \cdot \\ & & \cdot & 1 \\ 0 & & & \mu \end{pmatrix}$$

For the basis $\mathfrak{X}_j$ we have

$$(T_{\mathbb{C}} - \mu I)f_1 = 0 \qquad \text{and} \qquad (T_{\mathbb{C}} - \mu I)f_j = f_{j-1} \text{ for } j > 1,$$

which implies that $T_{\mathbb{C}}(f_1) = \mu f_1$ and $T_{\mathbb{C}}(f_j) = \mu f_j + f_{j-1}$ for $j > 1$.

If $f_j = x_j + iy_j \in W_\mu + iW_\mu$, we have $\bar{f}_j = J(f_j) = x_j - iy_j$ and $T_{\mathbb{C}}(f_j) = \mu f_j + f_{j-1}$, hence

$$T_{\mathbb{C}}(\bar{f}_j) = T_{\mathbb{C}}(J(f_j)) = J(T_{\mathbb{C}}(f_j)) = J(\mu f_j + f_{j-1}) = \overline{\mu} J(f_j) + J(f_{j-1}) \ .$$

Since real and imaginary parts must agree we get $T_{\mathbb{C}}(\bar{f}_j) = \overline{\mu}\bar{f}_j + \bar{f}_{j-1}$, as claimed.

Writing $\mu = a + ib$ with $b \neq 0$ (or in polar form, $\mu = re^{i\theta}$ with $\theta \notin \pi\mathbb{Z}$), we get

$$\begin{aligned} T(x_j) + iT(y_j) &= T_{\mathbb{C}}(x_j + iy_j) \ = \ T_{\mathbb{C}}(f_j) \ = \ \mu f_j + f_{j-1} \\ &= (a + ib)(x_j + iy_j) + (x_{j-1} + iy_{j-1}) \\ &= (ax_j - by_j) + i(bx_j + ay_j) + (x_{j-1} + iy_{j-1}) \end{aligned}$$

Since $\mu = re^{i\theta}$ and $\overline{\mu} = re^{-i\theta}$ this means

$$\begin{aligned} T(x_j) &= ax_j - by_j + x_{j-1} \ = \ x_j \cdot r\cos(\theta) - y_j \cdot r\sin(\theta) + x_{j-1} \\ T(y_j) &= bx_j + ay_j + y_{j-1} \ = \ x_j \cdot r\sin(\theta) + y_j \cdot r\cos(\theta) + y_{j-1} \end{aligned}$$

with respect to the $\mathbb{R}$-basis

$$\{x_1^{(1)}, y_1^{(1)}, \cdots, x_{d_1}^{(1)}, y_{d_1}^{(1)}, x_1^{(2)}; y_1^{(2)}, \cdots\}$$

in $V_\mu = M_{\mu,\overline{\mu}} \cap (V + i0) = \bigoplus_{j=1}^{d} V_J^{(\mu)}$. Thus the matrix $[T]_{\mathfrak{X}}$ consists of diagonal blocks of size $2d_j \times 2d_j$ that have the form

$$\begin{pmatrix} R & I_2 & \cdot & 0 \\ & R & I_2 & \\ & & \ddots & I_2 \\ 0 & & & R \end{pmatrix} = \begin{pmatrix} rR(\theta) & I_2 & \cdot & 0 \\ & rR(\theta) & I_2 & \\ & & \ddots & I_2 \\ 0 & & & rR(\theta) \end{pmatrix}$$

in which I_2 is the 2×2 identity matrix and

$$R = r \cdot R(\theta) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} r\,\cos(\theta) & -r\,\sin(\theta) \\ r\,\sin(\theta) & r\,\cos(\theta) \end{pmatrix}$$

if $\mu = a + ib = re^{i\theta}$. $\square$