VI. CANONICAL FORMS

VI.C. Rational canonical form

Let's agree to call a transformation $T: F^n \to F^n$ semisimple if there is a basis $\mathcal{B} = {\vec{v}_1, ..., \vec{v}_n}$ such that

 $T\vec{v}_1 = \lambda_1\vec{v}_1$, $T\vec{v}_2 = \lambda_2\vec{v}_2$, ..., $T\vec{v}_n = \lambda_n$ for some scalars $\lambda_i \in F$.

T is completely transparent when looked at relative to this basis. In terms of matrices, if $A = [T]_{\hat{e}}$ and $D = [T]_{\mathcal{B}} = \text{diag}\{\lambda_1, \dots, \lambda_n\}$, then

$$A = S D S^{-1},$$

and A is similar to a diagonal matrix.

What if *A* is *not* diagonalizable (\Leftrightarrow *T* is not semisimple)? Is there some basis in terms of which the action of *T* is still "transparent"? An equivalent question: is there a "canonical" [= standard] *sort* of matrix to which *any A* is similar? In the next three sections, we present two distinct solutions to this problem: the *rational* and *Jordan canonical forms* of *A*. That is, any *A* is similar to (essentially) unique matrices \mathcal{R} and \mathcal{J} obeying certain rules. One warning: in case *A* is diagonalizable, i.e. $A \sim D$, the rational form \mathcal{R} does *not* reduce to that (or any) diagonal matrix. Later we will find that, by contrast, the Jordan form \mathcal{J} *does* equal *D* in this case.

Another basic question: *when* is A diagonalizable? Momentarily assuming $F = \mathbb{C}$,

• <u>Answer 1</u>: when the geometric multiplicities (of the eigenvalues of *A*) equal the algebraic multiplicities.

• <u>Answer 2</u>: when m_A (not f_A) has no repeated roots (see Exercise VI.B.4). Although this avoids the computation involved in Answer 1 (with which you are by now familiar), finding $nf(\lambda I - A)$ ain't easy either.

• <u>Answer 3</u>: the Spectral Theorem, to be discussed later in this text.

Now we go over a couple of basic building blocks we'll need for our discussion of rational canonical form.

Direct-Sum Decomposition. The notation

(VI.C.1)
$$V = V_1 \oplus \cdots \oplus V_r$$

means that

- (i) $V = V_1 + \cdots + V_r$
- (ii) $\sum_{i} \dim V_{i} = \dim V$, or equivalently
- (ii') there is only one way of writing a given $\vec{v} \in V$ as $\vec{v}_1 + \ldots + \vec{v}_r$ (with $\vec{v}_i \in V_i$).

We say that *V* is the *direct sum* of the V_i . If this holds then there exists a basis $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_r\}$ for *V* such that \mathcal{B}_i are (respectively) bases for V_i . A linear transformation $T : V \to V$ is said to *respect* the direct sum decomposition if $T(V_i) \subseteq V_i$ (for each *i*); that is, if each V_i is *closed* under the action of *T*. In that case,

(VI.C.2)
$$[T]_{\mathcal{B}} = \begin{pmatrix} [T|_{V_1}]_{\mathcal{B}_1} & 0 \\ & \ddots & \\ 0 & [T|_{V_r}]_{\mathcal{B}_r} \end{pmatrix}$$

— that is, the matrix of *T* (with respect to \mathcal{B}) consists of *blocks*¹⁰ along the diagonal, representing its behavior on each subspace *V*_{*i*}.

Cyclic Subspaces and Vectors. Let $T: V \rightarrow V$ be a transformation, and look at the successive images (under *T*) of $\vec{v} \in V$:

$$\vec{v}$$
, $T\vec{v}$, $T^2\vec{v}$, ...

Now *V* is finite-dimensional, so the collections $\{\vec{v}, T\vec{v}, ..., T^i\vec{v}\}$ have to stop being independent at some point: i.e., there is some *m* such that

 $\vec{v}, T\vec{v}, \ldots, T^{m-1}\vec{v}$ are independent,

but

 \vec{v} , $T\vec{v}$, ..., $T^{m-1}\vec{v}$, $T^m\vec{v}$ are dependent

and so

$$0 = \alpha_m T^m \vec{v} + \alpha_{m-1} T^{m-1} \vec{v} + \ldots + \alpha_1 T \vec{v} + \alpha_0 \vec{v},$$

¹⁰smaller matrices: evidently the dimensions of the *i*th block are dim $V_i \times \dim V_i$.

where not all $\alpha_i = 0$. In fact $\alpha_m \neq 0$, because otherwise we would contradict independence of $\vec{v}, \ldots, T^{m-1}\vec{v}$. Dividing by α_m , we see that

$$0 = p(T)\vec{v}$$

where *p* is the monic polynomial

$$p(\lambda) = \lambda^m + \frac{\alpha_{m-1}}{\alpha_m} \lambda^{m-1} + \ldots + \frac{\alpha_0}{\alpha_m} =: \lambda^m + a_{m-1} \lambda^{m-1} + \ldots + a_0.$$

Let

$$W = span\{\vec{v}, T\vec{v}, \dots, T^{m-1}\vec{v}\};$$

it is called a (*T*-)*cyclic subspace* of *V* with *cyclic vector* \vec{v} , and we say that \vec{v} "generates" *W* (under the action of *T*).

VI.C.3. REMARK. Clearly
$$T^m \vec{v} \in W$$
 since $p(T)\vec{v} = 0 \implies$
 $T^m \vec{v} = -a_{m-1}T^{m-1}\vec{v} - \ldots - a_1T\vec{v} - a_0\vec{v}.$

In fact *W* contains $T^{m+1}\vec{v}$, $T^{m+2}\vec{v}$, ... simply by applying *T* to both sides of this equation repeatedly. So if we "cut off" at the degree (= m) of the first polynomial relation amongst the $T^i\vec{v}$ and consider the span (= W) of the first *m* of these, we have already the space spanned by *all* the $T^i\vec{v}$.

Let's look at the matrix of $T|_W$ with respect to the basis $\mathcal{B} = \{\vec{v}, T\vec{v}, \dots, T^{d-1}\vec{v}\} =: \{\vec{v}_1, \dots, \vec{v}_m\}$. Clearly

$$Tec v_1 = ec v_2$$
 , $Tec v_2 = ec v_3$, \ldots , $Tec v_{m-1} = ec v_m$,

while

$$T\vec{v}_m = T(T^{m-1}\vec{v}) = T^m\vec{v} = -a_{m-1}T^{m-1}\vec{v} - a_{m-2}T^{m-2}\vec{v} - \dots - a_0\vec{v}$$
$$= -a_{m-1}\vec{v}_m - a_{m-2}\vec{v}_{m-1} - \dots - a_0\vec{v}_1.$$

Therefore

(VI.C.4)
$$[T |_W]_{\mathcal{B}} = \begin{pmatrix} 0 & \cdots & 0 & -a_0 \\ 1 & 0 & -a_1 \\ & \ddots & & \vdots \\ 0 & 1 & -a_{m-1} \end{pmatrix} =: M(p);$$

this is called the *companion matrix* of the (monic) polynomial

(VI.C.5)
$$p(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} + \ldots + a_0.$$

VI.C.6. PROPOSITION. $p(\lambda)$ is both the minimal and characteristic polynomial of its companion matrix.

PROOF. We proceed by computing $nf(\lambda \mathbb{I} - M(p))$: starting with

$$\lambda \mathbb{I} - M(p) = \begin{pmatrix} \lambda & 0 & a_0 \\ -1 & \lambda & & a_1 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \lambda & & a_{m-3} \\ & & & -1 & \lambda & a_{m-2} \\ 0 & & & -1 & \lambda + a_{m-1} \end{pmatrix},$$

add λ times the bottom row to the 2nd row from bottom, to obtain

$$\left(egin{array}{cccccccc} \lambda & & 0 & & a_0 \ -1 & \lambda & & & a_1 \ & \ddots & \ddots & & & \vdots \ & & \ddots & \lambda & & a_{m-3} \ & & & -1 & 0 & \lambda^2 + a_{m-1}\lambda + a_{m-2} \ 0 & & & -1 & \lambda + a_{m-1} \end{array}
ight);$$

repeat all the way up (add λ times $(i + 1)^{st}$ row to the i^{th}) to get

$$\begin{pmatrix} 0 & 0 & \lambda^m + a_{m-1}\lambda^{m-1} + \ldots + a_0 \\ -1 & 0 & \lambda^{m-1} + a_{m-1}\lambda^{m-1} + \ldots + a_1 \\ & \ddots & \ddots & & \vdots \\ & & \ddots & 0 & \lambda^3 + a_{m-1}\lambda^2 + a_{m-2}\lambda + a_{m-3} \\ & & -1 & 0 & \lambda^2 + a_{m-1}\lambda + a_{m-2} \\ 0 & & -1 & \lambda + a_{m-1} \end{pmatrix}.$$

Noting that the top-right entry is $p(\lambda)$, add multiples of the 1st thru $(m-1)^{st}$ columns to the right-hand column to kill all the *other* entries

(in that column), obtaining

$$\left(\begin{array}{cccc} 0 & \cdots & 0 & p(\lambda) \\ -1 & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & -1 & 0 \end{array}\right)$$

Finally, multiplying the first (m - 1) columns by (-1) and performing row swaps, we arrive at

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & & \\ & & 1 \\ 0 & & & p(\lambda) \end{pmatrix} = nf(\lambda \mathbb{I} - M(p)).$$

Thus $m_{M(p)}$ [= lower-right-hand entry]= $p(\lambda)$ and $f_{M(p)}(\lambda)$ [= product of diagonal entries]= $p(\lambda)$, as claimed.

VI.C.7. REMARK. Since $p(\lambda)$ is the minimal polynomial of $T|_W$, $p(T|_W) = 0$; in other words, p(T) annihilates W, which is also clear from the fact that it annihilated W's cyclic vector \vec{v} above.

Rational Canonical Form. For simplicity let $V = F^n$, $A \in M_n(F)$ be any matrix, and $T: V \to V$ be the corresponding transformation $(A = [T]_{\hat{e}}$ as usual). Recall that the diagonal entries of $nf(\lambda I - A)$ give the invariant factors $\Delta_k(\lambda)$ of $\lambda I - A$, and let $\{\Delta_r, \ldots, \Delta_n\}$ be the nontrivial ($\neq 1$) ones among them. We now show that to each (nontrivial) Δ_k there is associated a *T*-cyclic subspace $W_k \subseteq V$ having $\Delta_k(\lambda)$ as both characteristic and minimal polynomial, and moreover that $V = W_r \oplus \cdots \oplus W_n$. The rational canonical form is then just the reflection of this decomposition of *V* (into cyclic subspaces) in matrix form.

In §VI.B we showed how to use row and column operations to transform $(\lambda \mathbb{I} - A) \in M_n(\mathsf{F}[\lambda])$ into a matrix in normal form: (VI.C.8)

 $P(\lambda \mathbb{I} - A)Q = nf(\lambda \mathbb{I} - A) = \text{diag}\{1, \dots, 1, \Delta_r(\lambda), \dots, \Delta_n(\lambda)\}$

where $P, Q \in M_n(\mathsf{F}[\lambda])$ are invertible, and $\sum_k \deg \Delta_k(\lambda) = n$.

Set $m_k := \text{deg } \Delta_k(\lambda)$. By swapping rows and columns one may change the right-hand side of (VI.C.8) to

diag{
$$\underbrace{1,\ldots,1,\Delta_r(\lambda)}_{m_r}$$
; ...; $\underbrace{1,\ldots,1,\Delta_n(\lambda)}_{m_n}$ }.

Reversing the row/column operations we did on the companion matrices, we may change this to a matrix

diag{
$$\lambda \mathbb{I}_{m_r} - M(\Delta_r), \ldots, \lambda \mathbb{I}_{m_n} - M(\Delta_n)$$
}

consisting of *blocks* of dimensions $m_r \times m_r, \ldots, m_n \times m_n$ down the diagonal. So we have

(VI.C.9)
$$P(\lambda \mathbb{I} - A)'Q = \lambda \mathbb{I}_n - \text{diag}\{M(\Delta_r), \dots, M(\Delta_n)\}$$

where '*P*, ' $Q \in M_n(\mathsf{F}[\lambda])$ are still invertible.

Now *suppose* we could replace the left-hand side of (VI.C.9) by $S^{-1}(\lambda \mathbb{I}_n - A)S$ for $S \in M_n(\mathsf{F})$ with *scalar* entries. Here is what that *would* tell us. Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be the basis given by the columns of *S*. Then since $S^{-1}(\lambda \mathbb{I}_n - A)S = \lambda \mathbb{I}_n - S^{-1}AS$,

$$\lambda \mathbb{I}_n - S^{-1}AS = \lambda \mathbb{I}_n - \operatorname{diag}\{M(\Delta_r), \dots, M(\Delta_n)\}$$
$$\implies S^{-1}AS = \operatorname{diag}\{M(\Delta_r), \dots, M(\Delta_n)\},$$

or

(VI.C.10)
$$[T]_{\mathcal{B}} = \begin{pmatrix} M(\Delta_r) \\ & \ddots \\ & M(\Delta_n) \end{pmatrix}.$$

Combined with what we have seen about cyclic subspaces and direct sums, this tells us that *V* decomposes into a direct sum of *T*-cyclic subspaces as described.

Namely, if the columns of our proposed *S* yield

$$\mathcal{B} = \{ \vec{v}_1, \dots, \vec{v}_{m_r} ; \vec{v}_{m_r+1}, \dots, \vec{v}_{m_r+m_{r+1}} ; \dots ; \vec{v}_{n-m_n+1}, \dots, \vec{v}_n \}$$

=: { $\mathcal{B}_r; \mathcal{B}_{r+1}; \dots; \mathcal{B}_n$ },

then define $W_k = span\{\mathcal{B}_k\}$, and $V = W_r \oplus \cdots \oplus W_n$. From the supposed form of $[T]_{\mathcal{B}}$, it is clear that:

- $\Delta_k(T)$ annihilates the subspace W_k (cf. Remark VI.C.7);
- $\vec{v}_1, \vec{v}_{m_r+1}, \ldots, \vec{v}_{n-m_r+1}$ are cyclic vectors for $W_r, W_{r+1}, \ldots, W_n$; and
- $\mathcal{B}_r = \{ \vec{v}_1, \dots, \vec{v}_{m_r} \} = \{ \vec{v}_1, T\vec{v}_1, \dots, T^{m_r 1}\vec{v}_1 \}$ (etc. for the other \mathcal{B}_k).

All of this should be regarded as provisional and tentative, since we shall soon present the basis differently. After all, we still have to prove it exists!

VI.C.11. EXAMPLE. Before things get more technical, we need some light at the end of the tunnel. Let

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

and consider the basis \mathcal{B} given by

$$\vec{v}_1 = \begin{pmatrix} 1\\0\\-2\\0 \end{pmatrix}, \ \vec{v}_2 = \begin{pmatrix} 0\\1\\0\\0 \end{pmatrix}, \ \vec{v}_3 = \begin{pmatrix} 0\\1\\-2\\0 \end{pmatrix}, \ \vec{v}_4 = \begin{pmatrix} 0\\1\\-2\\2 \end{pmatrix}$$

You can easily check that $A\vec{v}_1 = \vec{v}_1$, $A\vec{v}_2 = \vec{v}_3$, $A\vec{v}_3 = \vec{v}_4$, and $A\vec{v}_4 = 3\vec{v}_4 - 3\vec{v}_3 + \vec{v}_2$. Taking *S* to be the matrix with these vectors as its columns, we therefore have

$$S^{-1}AS = [T]_{\mathcal{B}} = \begin{pmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

If you did Exercise VI.B.3(b), you know that the two nontrivial invariant factors of *A* are $\Delta_3(A) = \lambda - 1$ and $\Delta_4(A) = (\lambda - 1)^3 = \lambda^3 - 3\lambda^2 + 3\lambda - 1$. The boxed submatrices are their companion matrices, $M(\Delta_3)$ and $M(\Delta_4)$.

Finding the Basis. We now embark on the quest to find the basis \mathcal{B} in general so that $[T]_{\mathcal{B}}$ takes the form (VI.C.10). We'll need to work with a space of "vectors" with polynomial coefficients, namely $(\mathsf{F}[\lambda])^n$. (This is really not a vector space, but a module over the ring $\mathsf{F}[\lambda]$.) There is a map

$$(\mathsf{F}[\lambda])^n \xrightarrow{\eta} \mathsf{F}^n$$

given by writing out a "vector" in components and formally replacing λ by T:¹¹

$$\vec{c} = g_1(\lambda)\hat{e}_1 + \ldots + g_n(\lambda)\hat{e}_n \longmapsto \eta(\vec{c}) := g_1(T)\hat{e}_1 + \ldots + g_n(T)\hat{e}_n.$$

Recall that the row operations $P = P_M \cdots P_1$ used to put $(\lambda \mathbb{I} - A)$ in normal form were invertible, so we may consider¹² $P^{-1} = P_1^{-1} \cdots P_M^{-1}$. Let C, Γ be the bases $\{\vec{c}_1, \ldots, \vec{c}_n\}$, $\{\vec{\gamma}_1, \ldots, \vec{\gamma}_n\}$ for $(\mathsf{F}[\lambda])^n$ consisting of the "column vectors" of P^{-1} and Q, respectively. We write $S_C^{-1} = {}_C[\mathbb{I}]_{\hat{e}} = P$, $S_{\Gamma} = {}_{\hat{e}}[\mathbb{I}]_{\Gamma} = Q$, and also $[\lambda \mathbb{I} - T]_{\hat{e}}$ for $\lambda \mathbb{I}_n - A$; now (VI.C.8) becomes (VI.C.12)

$$\begin{pmatrix} 1 & & & \\ & \ddots & & & \\ & & 1 & & \\ & & \Delta_r(\lambda) & & \\ & & & \ddots & \\ & & & & \Delta_n(\lambda) \end{pmatrix} = S_{\mathcal{C}}^{-1}([\lambda \mathbb{I} - T]_{\hat{\ell}})S_{\Gamma} = {}_{\mathcal{C}}[\lambda \mathbb{I} - T]_{\Gamma}$$

We are going to use (VI.C.12) to show that the vectors $\eta(\vec{c}_r), \ldots, \eta(\vec{c}_n)$ generate $V(=\mathsf{F}^n)$ under the action of *T*.

To this end let

- W_k := space generated by $\eta(\vec{c}_k)$ under the action of *T*, and
- $m_k := \deg \Delta_k(\lambda)$ (which we do not yet know = dim W_k).

¹¹Here $g_i(T)\hat{e}_i$ is just a polynomial in the transformation T [or matrix A] acting on the *i*th standard basis vector $\in \mathsf{F}^n$.

¹²Finding P^{-1} means taking the product of the elementary matrices corresponding to the reverse of the operations you did — *only* the row operations!

Noting that \vec{c}_k (resp. $\vec{\gamma}_k$) are the columns of S_C (resp. S_Γ), denote the entries of S_C , S_Γ by C_{ij} , $\Gamma_{ij} \in \mathsf{F}[\lambda]$, so that

$$\vec{\gamma}_j = \Gamma_{1j}\hat{e}_1 + \ldots + \Gamma_{nj}\hat{e}_n, \quad \vec{c}_j = C_{1j}\hat{e}_1 + \ldots + C_{nj}\hat{e}_n.$$

Now observe that according to (VI.C.12), $\lambda \mathbb{I} - T$ takes

$$\vec{\gamma}_1 \mapsto \vec{c}_1, \ldots, \vec{\gamma}_{r-1} \mapsto \vec{c}_{r-1}; \ \vec{\gamma}_r \mapsto \Delta_r(\lambda)\vec{c}_r, \ldots, \vec{\gamma}_n \mapsto \Delta_n(\lambda)\vec{c}_n.$$

We expand this as follows for all j = 1, ..., r - 1, k = r, ..., n:

(VI.C.13)

$$(\lambda \mathbb{I} - T) \left(\Gamma_{1j}(\lambda) \hat{e}_1 + \ldots + \Gamma_{nj}(\lambda) \hat{e}_n \right) = C_{1j}(\lambda) \hat{e}_1 + \ldots + C_{nj}(\lambda) \hat{e}_n,$$

$$(\lambda \mathbb{I} - T) \left(\Gamma_{1k}(\lambda) \hat{e}_1 + \ldots + \Gamma_{nk}(\lambda) \hat{e}_n \right) = \Delta_k(\lambda) C_{1k}(\lambda) \hat{e}_1 + \ldots + \Delta_k(\lambda) C_{nk}(\lambda) \hat{e}_n.$$

One can employ a formal argument like that in our proof of Cayley-Hamilton, to show it is valid to substitute in *T* for λ to get (since $T\mathbb{I} - T = 0$)

(VI.C.14)
$$0 = C_{1j}(T)\hat{e}_1 + \dots + C_{nj}(T)\hat{e}_n = \eta(\vec{c}_j), \quad j = 1, \dots, r-1,$$

and

(VI.C.15)
$$0 = \Delta_k(T) (C_{1k}(T)\hat{e}_1 + \dots + C_{nk}(T)\hat{e}_n)$$

= $\Delta_k(T)\eta(\vec{c}_k), \quad k = r, \dots, n.$

Basically, this amounts to taking η of both sides of both equations (VI.C.13).¹³

$$\eta(\vec{c}_j) = 0, \quad 1 \le j \le r-1$$

goes as follows: Since $T\hat{e}_{\ell} = \sum_{i} A_{i\ell}\hat{e}_{i}$,

$$(\lambda \mathbb{I} - T) \left(\sum_{i} \Gamma_{ij}(\lambda) \hat{e}_{i} \right) = \sum_{i} C_{ij}(\lambda) \hat{e}_{i} \implies$$

 $\sum_{i} (\lambda \Gamma_{ij}(\lambda) - C_{ij}(\lambda)) \hat{e}_{i} = \sum_{\ell} \Gamma_{\ell j}(\lambda) T \hat{e}_{\ell} = \sum_{i,\ell} \Gamma_{\ell j}(\lambda) A_{i\ell} \hat{e}_{i} \implies$ for each *i*, *j* (1 ≤ *j* ≤ *r* − 1) the *polynomial* equation

$$\lambda \Gamma_{ij}(\lambda) - C_{ij}(\lambda) = \sum_{\ell} A_{i\ell} \Gamma_{\ell j}(\lambda)$$

 $^{^{13}}$ Unfortunately, justifying this is a bit ugly — unless you use theory of modules which makes such things automatic. (Think of it as motivation to learn abstract algebra!) The formal argument (without module theory) for

So we may discard the first r - 1 columns of $S_C = P^{-1}$, since their images under η are zero by (VI.C.14). For the remaining \vec{c}_k , $\Delta_k(T)$ annihilates $\eta(\vec{c}_k)$ by (VI.C.15) so that

$$\{\eta(\vec{c}_k), T\eta(\vec{c}_k), \ldots, T^{m_k}\eta(\vec{c}_k)\}$$
 are dependent.

We do *not* yet know that

$$\left\{\eta(ec{c}_k), T\eta(ec{c}_k), \ldots, T^{m_k-1}\eta(ec{c}_k)
ight\}$$
 are independent,

but we *do* know that W_k = their span and so

$$\dim W_k \le m_k = \deg \Delta_k(\lambda).$$

Since $\sum_k \deg \Delta_k = n$,

$$\sum_k \dim W_k \leq n.$$

Moreover, $\Delta_k(T)$ annihilates W_k (since it annihilates $\eta(\vec{c}_k)$).

Now let R_{ij} be the (polynomial) entries of $P = S_{\mathcal{C}}^{-1} = {}_{\mathcal{C}}[\mathbb{I}]_{\hat{e}}$, so that for all i = 1, ..., n

$$\hat{e}_i = R_{1i}(\lambda)\vec{c}_1 + \ldots + R_{ni}(\lambda)\vec{c}_n.$$

Applying η to both sides, we have

$$\hat{e}_i = R_{1i}(T)\eta(\vec{c}_1) + \ldots + R_{ni}(T)\eta(\vec{c}_n)$$
$$= R_{ri}(T)\eta(\vec{c}_r) + \ldots + R_{ni}(T)\eta(\vec{c}_n)$$

since the first (r - 1) { $\eta(\vec{c}_j)$ } are zero. Therefore (sums of) powers of *T* acting on the { $\eta(\vec{c}_r), \ldots, \eta(\vec{c}_n)$ } give all the { $\hat{e}_1, \ldots, \hat{e}_n$ }, and thereby span all of $V = F^n$! In other words,

$$W_r + \ldots + W_n = V,$$

$$\sum_{i} C_{ij}(T)\hat{e}_{i} = \sum_{i,\ell} \left(T\delta_{i\ell} - A_{i\ell} \right) \Gamma_{\ell j}(T)\hat{e}_{i} = \sum_{\ell} \Gamma_{\ell j}(T) \left[\sum_{i} \left(T\delta_{i\ell} - A_{i\ell} \right) \hat{e}_{i} \right] = 0$$

because $\sum_{i} (T\delta_{i\ell} - A_{i\ell})\hat{e}_{i} = T\hat{e}_{\ell} - \sum_{i} A_{i\ell}\hat{e}_{i} = 0.$

holds. That is, the two sides are the same polynomial. It now makes perfect sense to substitute in *T* for λ to obtain

 $T\Gamma_{ij}(T) - C_{ij}(T) = \sum_{\ell} A_{i\ell} \Gamma_{\ell j}(T) \implies C_{ij}(T) = \sum_{\ell} (T\delta_{i\ell} - A_{i\ell}) \Gamma_{\ell j}(T).$ Therefore $\eta(\vec{c}_j) :=$

and so

$$\sum_k \dim W_k \ge \dim V = n.$$

Combining our inequalities,

(VI.C.16)
$$\sum_{k} \dim W_{k} = n$$

and so the sum is direct:

$$V=W_r\oplus\cdots\oplus W_n.$$

Also (VI.C.16) forces dim $W_k = \deg \Delta_k(\lambda)$, so that

$$\mathcal{B}_k = \left\{ \eta(\vec{c}_k), T\eta(\vec{c}_k), \dots, T^{m_k - 1}\eta(\vec{c}_k) \right\}$$

is a basis for W_k (i.e., it was an independent set after all). Thus

$$\left[T\left|_{W_{k}}\right]_{\mathcal{B}_{k}}=M(\Delta_{k})$$

as promised, and

$$\mathcal{B} = \{\mathcal{B}_r, \ldots, \mathcal{B}_n\}$$

is the basis for $V = F^n$.

Let's sum up, in terms of a theorem about transformations, and an algorithm for the corresponding matrices:

VI.C.17. THEOREM. For $T : V \rightarrow V$ (dim V = n) there is a decomposition

$$V=W_r\oplus\cdots\oplus W_n$$

into *T*-cyclic subspaces, such that the minimal and characteristic polynomials of $T|_{W_k}$ are both just the k^{th} invariant factor Δ_k of $\lambda \mathbb{I} - T$.

Algorithm. Given $A \in M_n(\mathsf{F})$, apply the normal form procedure to $\lambda \mathbb{I} - A$ to obtain

$$P(\lambda \mathbb{I} - A)Q = \begin{pmatrix} 1 & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & \Delta_r(\lambda) & & \\ & & & & \ddots & \\ 0 & & & & \Delta_n(\lambda) \end{pmatrix}$$

Find the inverse P^{-1} of the row operations, call its entries C_{ij} (some will be polynomials in λ) and columns \vec{c}_k . Throw out $\vec{c}_1, \ldots, \vec{c}_{r-1}$; write each of the remaining people component-by-component

$$\vec{c}_k = C_{1k}(\lambda)\hat{e}_1 + \ldots + C_{nk}(\lambda)\hat{e}_n,$$

and compute

$$\eta(\vec{c}_k) = C_{1k}(T)\hat{e}_1 + \ldots + C_{nk}(T)\hat{e}_n :$$

after applying the *T*'s to the \hat{e}_i 's rewrite the result as a column vector. (This sounds hard, but in actual practice, is not: e.g., if \vec{c}_k happens only to have scalar entries, then $\eta(\vec{c}_k) = \vec{c}_k$.)

Let $m_k = \deg \Delta_k$, and write (for k = r, ..., n)

$$\mathcal{B}_k = \left\{\eta(\vec{c}_k), T\eta(\vec{c}_k), \ldots, T^{m_k-1}\eta(\vec{c}_k)\right\};$$

put these together to form a basis

$$\mathcal{B} = \{\mathcal{B}_r, \ldots, \mathcal{B}_n\}$$
 for F^n .

Set

$$S := S_{\mathcal{B}} = \left\{ \begin{array}{c} \text{matrix whose columns} \\ \text{are the vectors of } \mathcal{B} \end{array} \right\},$$

and write

 $A = S \cdot \mathcal{R} \cdot S^{-1},$

where

$$\mathcal{R} = \operatorname{diag}\{M(\Delta_r), \ldots, M(\Delta_n)\}$$

consists of companion-matrix blocks along the diagonal (cf. (VI.C.4) and (VI.C.10)).

VI.C.18. REMARK. You should of course think of this as a change of basis: $(A =)[T]_{\hat{e}} = {}_{\hat{e}}[\mathbb{I}]_{\mathcal{B}} \cdot {}_{\mathcal{B}}[\mathbb{I}]_{\hat{e}}$, where *T* is "transparent" relative to \mathcal{B} .

VI.C.19. EXAMPLE. We now put our favorite example

$$A = \left(\begin{array}{rrr} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right)$$

into rational canonical form. At the end of §VI.B, we put (for this *A*) $\lambda \mathbb{I} - A$ into normal form,

$$nf(\lambda \mathbb{I} - A) = \begin{pmatrix} 1 & & \\ & \lambda & \\ & & \lambda^2 - 3\lambda \end{pmatrix},$$

in 4 (somewhat condensed) steps. We list the row operations involved in each of these steps, numbering them in the order in which they occur:

Step I : R_1 = swapping rows 1 and 2

Step II : (Involves three row operations.) R_2 = multiplying row 1 by (-1), R_3 = adding $(1 - \lambda) \times \{\text{row 1}\}$ to row 2, R_4 = adding row 1 to row 3

Step III : R_5 = adding row 2 to row 3

Step IV : R_6 = multiplying row 2 by (-1).

What are the inverses of these row operations? This is routine by now! For example, R_5^{-1} = subtracting row 2 from row 3, while R_3^{-1} = adding ($\lambda - 1$) × {row 1} to row 2. Things like R_1 and R_6 are the same in reverse. To find $C = P^{-1}$ one simply writes out the corresponding product of matrices¹⁴

$$C = R_1^{-1} R_2^{-1} R_3^{-1} R_4^{-1} R_5^{-1} R_6^{-1}$$

$$= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda - 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda - 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} \lambda - 1 \\ -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \vec{c}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Now according to the entries of $nf(\lambda \mathbb{I} - A)$, the invariant factors of A are $\Delta_1 = 1$, $\Delta_2 = \lambda$, $\Delta_3 = \lambda^2 - 3\lambda$. This means that we keep $\vec{c}_2 = \eta(\vec{c}_2)$ and $\vec{c}_3 = \eta(\vec{c}_3)$, which will generate \mathbb{R}^3 under the action of A (recall that $\eta(\vec{c}) = \vec{c}$ if \vec{c} has only scalar entries); and we throw out \vec{c}_1 because $\Delta_1 = 1 \implies \eta(\vec{c}_1) = 0$.

¹⁴As an alternative (to avoid multiplying matrices), one could perform these as row operations on the identity matrix: $R_1^{-1}(\cdots(R_6^{-1}\mathbb{I}))$.

Let's check this, in order to see a computation of $\eta(\vec{c})$ where \vec{c} has actual polynomial entries! First break \vec{c}_1 into components with respect to the standard basis:

$$\vec{c}_1 = (\lambda - 1)\hat{e}_1 + (-1)\hat{e}_2 + (-1)\hat{e}_3.$$

Next substitute *T* for λ :

$$\eta(\vec{c}_1) = (T-1)\hat{e}_1 + (-1)\hat{e}_2 + (-1)\hat{e}_3$$
$$= T\hat{e}_1 - (\hat{e}_1 + \hat{e}_2 + \hat{e}_3).$$

Finally, rewrite everything in matrix-vector notation (with respect to \hat{e}) to compute the result:

$$= \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \vec{0}.$$

Written in this way it looks kind of like magic.

Now for the remaining cases k = 2, 3, dim $(W_k) = #\{$ vectors in $\mathcal{B}_k \}$ must be the degree m_k of Δ_k .¹⁵ Since $m_2 = 1$ and $m_3 = 2$,

$$\mathcal{B}_2 = \{\eta(\vec{c}_2)\} = \left\{ \begin{pmatrix} -1\\0\\1 \end{pmatrix} \right\}, \quad \mathcal{B}_3 = \{\eta(\vec{c}_3), T\eta(\vec{c}_3)\} = \left\{ \begin{pmatrix} 0\\0\\1 \end{pmatrix}, \begin{pmatrix} 1\\1\\1 \end{pmatrix} \right\}$$

These three vectors together give our basis and

$$S = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{rref} S^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Finally for the companion matrices. For arbitrary monic polynomials $\lambda + a_0$ and $\lambda^2 + a_1\lambda + a_0$ of degrees 1 and 2, these are

$$(-a_0)$$
 and $\begin{pmatrix} 0 & -a_0 \\ 1 & -a_1 \end{pmatrix}$.

Therefore in our case

$$M(\Delta_2) = (0)$$
 and $M(\Delta_3) = \begin{pmatrix} 0 & 0 \\ 1 & 3 \end{pmatrix}$.

¹⁵That is, \mathcal{B}_k is $\{\eta(\vec{c}_k), T\eta(\vec{c}_k), \dots, T^{m_k-1}\eta(\vec{c}_k)\}.$

The decomposition is therefore (in the form $A = SRS^{-1}$)

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \boxed{0} & 0 & 0 \\ 0 & \boxed{0} & 0 \\ 0 & \boxed{1} & 3 \end{bmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercises

(1) Compute the rational canonical form for

$$A = \left(\begin{array}{rrr} -1 & -1 & -1 \\ -2 & 0 & -1 \\ 6 & 3 & 4 \end{array} \right).$$

That is, present it as $S\mathcal{R}S^{-1}$, where \mathcal{R} consists of companion matrix blocks along the diagonal and *S* is a change-of-basis matrix. Interpret the result in terms of a direct-sum decomposition of \mathbb{R}^3 under the action of the transformation *T* with matrix $[T]_{\hat{e}} = A$.

(2) Same problem with

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

and \mathbb{R}^4 instead of \mathbb{R}^3 . [Note: from Exercise VI.B.2(a), you already know the normal form for $\lambda \mathbb{I}_4 - A$.]

(3) Once more, same problem with

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

How does it depend on *c*?

(4) Let T: V → V (with V ≅ ℝⁿ) be semisimple.
(a) If T has a cyclic vector v ∈ V (i.e. v, Tv,..., Tⁿ⁻¹v span V), show that T has n distinct eigenvalues.

EXERCISES

(b) If *T* has *n* distinct eigenvalues, then taking $\{\vec{v}_1, \ldots, \vec{v}_n\}$ to be an eigenbasis, show that $\vec{v} = \sum_{i=1}^n \vec{v}_i$ is a cyclic vector for *T*.

- (5) Let $A \in M_n(\mathbb{R})$ be such that $A^2 = -\mathbb{I}_n$.
 - (a) Prove that *n* is even.

(b) Writing n = 2k, show that A is similar (over \mathbb{R}) to

$$J_n := \left(egin{array}{cc} 0 & -\mathbb{I}_k \ \mathbb{I}_k & 0 \end{array}
ight).$$

- (6) Let $T : P^3 \to P^3$ be $\frac{d}{dx}$. Write its matrix A in the "standard" basis $\{x^3, x^2, x, 1\} = \mathcal{B}$, and put it in rational canonical form (i.e. $A = S\mathcal{R}S^{-1}$), without computation! [Hint: what is T^4 ?]
- (7) Suppose I tell you the characteristic polynomial of a 5 × 5 matrix *A*, without giving you the matrix: f_A(λ) = (λ − 3)³(λ − 2)².
 (a) What are the possible rational canonical forms *R*? (Start by listing the possibilities for *nf*(λI − A).) There should be six.
 (b) This says that there are six "distinct" (= nonsimilar) transformations of ℝ⁵ with eigenvalue list 3, 3, 3, 2, 2 (repeated according to multiplicity). Which ones are diagonalizable? [Hint: look at "Answer 2" at the beginning of the section!]