## VI.C. Rational canonical form

Let's agree to call a transformation $T: \mathrm{F}^{n} \rightarrow \mathrm{~F}^{n}$ semisimple if there is a basis $\mathcal{B}=\left\{\vec{v}_{1}, \ldots, \vec{v}_{n}\right\}$ such that

$$
T \vec{v}_{1}=\lambda_{1} \vec{v}_{1}, T \vec{v}_{2}=\lambda_{2} \vec{v}_{2}, \ldots, T \vec{v}_{n}=\lambda_{n} \text { for some scalars } \lambda_{i} \in \mathrm{~F} .
$$

$T$ is completely transparent when looked at relative to this basis. In terms of matrices, if $A=[T]_{\hat{e}}$ and $D=[T]_{\mathcal{B}}=\operatorname{diag}\left\{\lambda_{1}, \ldots, \lambda_{n}\right\}$, 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 $\mathrm{F}=\mathbb{C}$,

- Answer 1: when the geometric multiplicities (of the eigenvalues of A) equal the algebraic multiplicities.
- Answer 2: when $m_{A}\left(\operatorname{not} f_{A}\right)$ 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 $n f(\lambda I I-A)$ ain't easy either.
- Answer 3: 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

$$
\begin{equation*}
V=V_{1} \oplus \cdots \oplus V_{r} \tag{VI.C.1}
\end{equation*}
$$

means that
(i) $V=V_{1}+\cdots+V_{r}$
(ii) $\sum_{i} \operatorname{dim} V_{i}=\operatorname{dim} V$, 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 $\mathcal{B}=\left\{\mathcal{B}_{1}, \ldots, \mathcal{B}_{r}\right\}$ for $V$ such that $\mathcal{B}_{i}$ are (respectively) bases for $V_{i}$. A linear transformation $T: V \rightarrow V$ is said to respect the direct sum decomposition if $T\left(V_{i}\right) \subseteq V_{i}$ (for each $i$ ); that is, if each $V_{i}$ is closed under the action of $T$. In that case,
(VI.C.2) $\quad[T]_{\mathcal{B}}=\left(\begin{array}{ccc}{\left[\left.T\right|_{V_{1}}\right]_{\mathcal{B}_{1}}} & & 0 \\ & \ddots & \\ 0 & & {\left[\left.T\right|_{V_{r}}\right]_{\mathcal{B}_{r}}}\end{array}\right)$
— 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 $\vec{v} \in V$ :

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

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

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

but

$$
\vec{v}, T \vec{v}, \ldots, T^{m-1} \vec{v}, T^{m} \vec{v} \quad \text { are dependent }
$$

and so

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

$\overline{10}$ smaller matrices: evidently the dimensions of the $i^{\text {th }}$ block $\operatorname{are} \operatorname{dim} V_{i} \times \operatorname{dim} V_{i}$.
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 $\alpha_{m}$, 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

$$
W=\operatorname{span}\left\{\vec{v}, T \vec{v}, \ldots, T^{m-1} \vec{v}\right\}
$$

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

$$
T^{m} \vec{v}=-a_{m-1} T^{m-1} \vec{v}-\ldots-a_{1} T \vec{v}-a_{0} \vec{v}
$$

In fact $W$ contains $T^{m+1} \vec{v}, T^{m+2} \vec{v}, \ldots$ 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 $\left.T\right|_{W}$ with respect to the basis $\mathcal{B}=$ $\left\{\vec{v}, T \vec{v}, \ldots, T^{d-1} \vec{v}\right\}=:\left\{\vec{v}_{1}, \ldots, \vec{v}_{m}\right\}$. Clearly

$$
T \vec{v}_{1}=\vec{v}_{2}, T \vec{v}_{2}=\vec{v}_{3}, \ldots, T \vec{v}_{m-1}=\vec{v}_{m}
$$

while

$$
\begin{gathered}
T \vec{v}_{m}=T\left(T^{m-1} \vec{v}\right)=T^{m} \vec{v}=-a_{m-1} T^{m-1} \vec{v}-a_{m-2} T^{m-2} \vec{v}-\ldots-a_{0} \vec{v} \\
=-a_{m-1} \vec{v}_{m}-a_{m-2} \vec{v}_{m-1}-\ldots-a_{0} \vec{v}_{1}
\end{gathered}
$$

Therefore
(VI.C.4)

$$
\left[\left.T\right|_{W}\right]_{\mathcal{B}}=\left(\begin{array}{cccc}
0 & \cdots & 0 & -a_{0} \\
1 & & 0 & -a_{1} \\
& \ddots & & \vdots \\
0 & & 1 & -a_{m-1}
\end{array}\right)=: M(p)
$$

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

$$
\begin{equation*}
p(\lambda)=\lambda^{m}+a_{m-1} \lambda^{m-1}+\ldots+a_{0} . \tag{VI.C.5}
\end{equation*}
$$

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

Proof. We proceed by computing $n f(\lambda I-M(p))$ : starting with

$$
\lambda \mathbb{I}-M(p)=\left(\begin{array}{cccccc}
\lambda & & & & 0 & a_{0} \\
-1 & \lambda & & & & a_{1} \\
& \ddots & \ddots & & & \vdots \\
& & \ddots & \lambda & & a_{m-3} \\
& & & -1 & \lambda & a_{m-2} \\
0 & & & & -1 & \lambda+a_{m-1}
\end{array}\right),
$$

add $\lambda$ times the bottom row to the $2^{\text {nd }}$ row from bottom, to obtain

$$
\left(\begin{array}{cccccc}
\lambda & & & & 0 & a_{0} \\
-1 & \lambda & & & & a_{1} \\
& \ddots & \ddots & & & \vdots \\
& & \ddots & \lambda & & a_{m-3} \\
& & & -1 & 0 & \lambda^{2}+a_{m-1} \lambda+a_{m-2} \\
0 & & & & -1 & \lambda+a_{m-1}
\end{array}\right)
$$

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

Noting that the top-right entry is $p(\lambda)$, add multiples of the $1^{\text {st }}$ thru $(m-1)^{\text {st }}$ 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

$$
\left(\begin{array}{cccc}
1 & & & 0 \\
& \ddots & & \\
& & 1 & \\
0 & & & p(\lambda)
\end{array}\right)=n f(\lambda \mathbb{I}-M(p))
$$

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 $\left.T\right|_{W}$, $p\left(\left.T\right|_{W}\right)=0$; in other words, $p(T)$ annihilates $W$, which is also clear from the fact that it annihilated $W^{\prime}$ s cyclic vector $\vec{v}$ above.

Rational Canonical Form. For simplicity let $V=\mathrm{F}^{n}, A \in M_{n}(\mathrm{~F})$ be any matrix, and $T: V \rightarrow V$ be the corresponding transformation ( $A=[T]_{\hat{e}}$ as usual). Recall that the diagonal entries of $n f(\lambda \mathbb{I}-A)$ give the invariant factors $\Delta_{k}(\lambda)$ of $\lambda \mathbb{I}-A$, and let $\left\{\Delta_{r}, \ldots, \Delta_{n}\right\}$ be the nontrivial $(\neq 1)$ ones among them. We now show that to each (nontrivial) $\Delta_{k}$ 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 $(\lambda \mathbb{I}-A) \in M_{n}(\mathrm{~F}[\lambda])$ into a matrix in normal form:
(VI.C.8)

$$
P(\lambda \mathbb{I}-A) Q=n f(\lambda \mathbb{I}-A)=\operatorname{diag}\left\{1, \ldots, 1, \Delta_{r}(\lambda), \ldots, \Delta_{n}(\lambda)\right\}
$$

where $P, Q \in M_{n}(\mathrm{~F}[\lambda])$ are invertible, and $\sum_{k} \operatorname{deg} \Delta_{k}(\lambda)=n$.

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

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

$$
\operatorname{diag}\left\{\lambda \mathbb{I}_{m_{r}}-M\left(\Delta_{r}\right), \ldots, \lambda \mathbb{I}_{m_{n}}-M\left(\Delta_{n}\right)\right\}
$$

consisting of blocks of dimensions $m_{r} \times m_{r}, \ldots, m_{n} \times m_{n}$ down the diagonal. So we have

$$
\begin{equation*}
{ }^{\prime} P(\lambda \mathbb{I}-A)^{\prime} Q=\lambda \mathbb{I}_{n}-\operatorname{diag}\left\{M\left(\Delta_{r}\right), \ldots, M\left(\Delta_{n}\right)\right\} \tag{VI.C.9}
\end{equation*}
$$

where ${ }^{\prime} P,{ }^{\prime} Q \in M_{n}(\mathrm{~F}[\lambda])$ are still invertible.
Now suppose we could replace the left-hand side of (VI.C.9) by $S^{-1}\left(\lambda \mathbb{I}_{n}-A\right) S$ for $S \in M_{n}(\mathrm{~F})$ with scalar entries. Here is what that would tell us. Let $\mathcal{B}=\left\{\vec{v}_{1}, \ldots, \vec{v}_{n}\right\}$ be the basis given by the columns of $S$. Then since $S^{-1}\left(\lambda \mathbb{I}_{n}-A\right) S=\lambda \mathbb{I}_{n}-S^{-1} A S$,

$$
\begin{gathered}
\lambda \mathbb{I}_{n}-S^{-1} A S=\lambda \mathbb{I}_{n}-\operatorname{diag}\left\{M\left(\Delta_{r}\right), \ldots, M\left(\Delta_{n}\right)\right\} \\
\Longrightarrow S^{-1} A S=\operatorname{diag}\left\{M\left(\Delta_{r}\right), \ldots, M\left(\Delta_{n}\right)\right\}
\end{gathered}
$$

or

$$
[T]_{\mathcal{B}}=\left(\begin{array}{ccc}
\boxed{M\left(\Delta_{r}\right)} & &  \tag{VI.C.10}\\
& \ddots & \\
& & \boxed{M\left(\Delta_{n}\right)}
\end{array}\right)
$$

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

$$
\begin{gathered}
\mathcal{B}=\left\{\vec{v}_{1}, \ldots, \vec{v}_{m_{r}} ; \vec{v}_{m_{r}+1}, \ldots, \vec{v}_{m_{r}+m_{r+1}} ; \ldots ; \vec{v}_{n-m_{n}+1}, \ldots, \vec{v}_{n}\right\} \\
=:\left\{\mathcal{B}_{r} ; \mathcal{B}_{r+1} ; \ldots ; \mathcal{B}_{n}\right\},
\end{gathered}
$$

then define $W_{k}=\operatorname{span}\left\{\mathcal{B}_{k}\right\}$, 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 - $\mathcal{B}_{r}=\left\{\vec{v}_{1}, \ldots, \vec{v}_{m_{r}}\right\}=\left\{\vec{v}_{1}, T \vec{v}_{1}, \ldots, T^{m_{r}-1} \vec{v}_{1}\right\}$ (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

$$
A=\left(\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
-2 & -2 & 0 & 1 \\
-2 & 0 & -1 & 2
\end{array}\right)
$$

and consider the basis $\mathcal{B}$ given by

$$
\vec{v}_{1}=\left(\begin{array}{c}
1 \\
0 \\
-2 \\
0
\end{array}\right), \vec{v}_{2}=\left(\begin{array}{l}
0 \\
1 \\
0 \\
0
\end{array}\right), \vec{v}_{3}=\left(\begin{array}{c}
0 \\
1 \\
-2 \\
0
\end{array}\right), \vec{v}_{4}=\left(\begin{array}{c}
0 \\
1 \\
-2 \\
2
\end{array}\right) .
$$

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

$$
S^{-1} A S=[T]_{\mathcal{B}}=\left(\begin{array}{cccc}
\boxed{1} & 0 & 0 & 0 \\
0 & \begin{array}{|ccc|}
0 & 0 & 1 \\
0 & 1 & 0
\end{array} & -3 \\
0 & 0 & 1 & 3 \\
\hline
\end{array}\right)
$$

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, $M\left(\Delta_{3}\right)$ and $M\left(\Delta_{4}\right)$.

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 $(\mathrm{F}[\lambda])^{n}$. (This is really not a vector space, but a module over the ring $F[\lambda]$.) There is a map

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

given by writing out a "vector" in components and formally replacing $\lambda$ 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 $(\lambda I I-$ A) in normal form were invertible, so we may consider ${ }^{12} P^{-1}=$ $P_{1}^{-1} \ldots P_{M}^{-1}$. Let $\mathcal{C}$, $\Gamma$ be the bases $\left\{\vec{c}_{1}, \ldots, \vec{c}_{n}\right\},\left\{\vec{\gamma}_{1}, \ldots, \vec{\gamma}_{n}\right\}$ for $(F[\lambda])^{n}$ consisting of the "column vectors" of $P^{-1}$ and $Q$, respectively. We write $S_{\mathcal{C}}^{-1}={ }_{\mathcal{C}}[\mathbb{I}]_{\hat{e}}=P, S_{\Gamma}={ }_{\hat{e}}[\mathbb{I}]_{\Gamma}=Q$, and also $[\lambda I I-T]_{\hat{e}}$ for $\lambda \mathbb{I}_{n}-A$; now (VI.C.8) becomes
(VI.C.12)

$$
\left(\begin{array}{ccccc}
1 & & & & \\
& \ddots & & & \\
\\
& & 1 & & \\
& & & \Delta_{r}(\lambda) & \\
& & & & \ddots \\
\\
& & & & \\
\Delta_{n}(\lambda)
\end{array}\right)=S_{\mathcal{C}}^{-1}\left([\lambda \mathbb{I}-T]_{\hat{e}}\right) S_{\Gamma}={ }_{\mathcal{C}}[\lambda \mathbb{I}-T]_{\Gamma}
$$

We are going to use (VI.C.12) to show that the vectors $\eta\left(\vec{c}_{r}\right), \ldots, \eta\left(\vec{c}_{n}\right)$ generate $V\left(=\mathrm{F}^{n}\right)$ under the action of $T$.

To this end let

- $W_{k}:=$ space generated by $\eta\left(\vec{c}_{k}\right)$ under the action of $T$, and
- $m_{k}:=\operatorname{deg} \Delta_{k}(\lambda)$ (which we do not yet know $=\operatorname{dim} W_{k}$ ).

[^0]Noting that $\vec{c}_{k}$ (resp. $\vec{\gamma}_{k}$ ) are the columns of $S_{\mathcal{C}}$ (resp. $S_{\Gamma}$ ), denote the entries of $S_{\mathcal{C}}, S_{\Gamma}$ by $C_{i j}, \Gamma_{i j} \in \mathrm{~F}[\lambda]$, so that

$$
\vec{\gamma}_{j}=\Gamma_{1 j} \hat{e}_{1}+\ldots+\Gamma_{n j} \hat{e}_{n}, \quad \vec{c}_{j}=C_{1 j} \hat{e}_{1}+\ldots+C_{n j} \hat{e}_{n} .
$$

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 $j=1, \ldots, r-1, k=r, \ldots, n$ :
(VI.C.13)

$$
\begin{aligned}
& (\lambda \mathbb{I}-T)\left(\Gamma_{1 j}(\lambda) \hat{e}_{1}+\ldots+\Gamma_{n j}(\lambda) \hat{e}_{n}\right)=C_{1 j}(\lambda) \hat{e}_{1}+\ldots+C_{n j}(\lambda) \hat{e}_{n}, \\
& (\lambda \mathbb{I}-T)\left(\Gamma_{1 k}(\lambda) \hat{e}_{1}+\ldots+\Gamma_{n k}(\lambda) \hat{e}_{n}\right)= \\
& \Delta_{k}(\lambda) C_{1 k}(\lambda) \hat{e}_{1}+\ldots+\Delta_{k}(\lambda) C_{n k}(\lambda) \hat{e}_{n} .
\end{aligned}
$$

One can employ a formal argument like that in our proof of CayleyHamilton, to show it is valid to substitute in $T$ for $\lambda$ to get (since $T I I-T=0)$
(VI.C.14) $0=C_{1 j}(T) \hat{e}_{1}+\cdots+C_{n j}(T) \hat{e}_{n}=\eta\left(\vec{c}_{j}\right), \quad j=1, \ldots, r-1$,
and
(VI.C.15) $0=\Delta_{k}(T)\left(C_{1 k}(T) \hat{e}_{1}+\cdots+C_{n k}(T) \hat{e}_{n}\right)$

$$
=\Delta_{k}(T) \eta\left(\vec{c}_{k}\right), \quad k=r, \ldots, n .
$$

Basically, this amounts to taking $\eta$ of both sides of both equations (VI.C.13). ${ }^{13}$
${ }^{13}$ 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

$$
\eta\left(\vec{c}_{j}\right)=0, \quad 1 \leq j \leq r-1
$$

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

$$
\begin{gathered}
(\lambda \mathbb{I}-T)\left(\sum_{i} \Gamma_{i j}(\lambda) \hat{e}_{i}\right)=\sum_{i} C_{i j}(\lambda) \hat{e}_{i} \quad \Longrightarrow \\
\sum_{i}\left(\lambda \Gamma_{i j}(\lambda)-C_{i j}(\lambda)\right) \hat{e}_{i}=\sum_{\ell} \Gamma_{\ell j}(\lambda) T \hat{e}_{\ell}=\sum_{i, \ell} \Gamma_{\ell j}(\lambda) A_{i \ell} \hat{e}_{i} \quad \Longrightarrow
\end{gathered}
$$ for each $i, j(1 \leq j \leq r-1)$ the polynomial equation

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

So we may discard the first $r-1$ columns of $S_{\mathcal{C}}=P^{-1}$, since their images under $\eta$ are zero by (VI.C.14). For the remaining $\vec{c}_{k}$, $\Delta_{k}(T)$ annihilates $\eta\left(\vec{c}_{k}\right)$ by (VI.C.15) so that

$$
\left\{\eta\left(\vec{c}_{k}\right), T \eta\left(\vec{c}_{k}\right), \ldots, T^{m_{k}} \eta\left(\vec{c}_{k}\right)\right\} \text { are dependent. }
$$

We do not yet know that

$$
\left\{\eta\left(\vec{c}_{k}\right), \operatorname{T\eta }\left(\vec{c}_{k}\right), \ldots, T^{m_{k}-1} \eta\left(\vec{c}_{k}\right)\right\} \text { are independent, }
$$

but we do know that $W_{k}=$ their span and so

$$
\operatorname{dim} W_{k} \leq m_{k}=\operatorname{deg} \Delta_{k}(\lambda)
$$

Since $\sum_{k} \operatorname{deg} \Delta_{k}=n$,

$$
\sum_{k} \operatorname{dim} W_{k} \leq n
$$

Moreover, $\Delta_{k}(T)$ annihilates $W_{k}$ (since it annihilates $\eta\left(\vec{c}_{k}\right)$ ).
Now let $R_{i j}$ be the (polynomial) entries of $\left.P=S_{\mathcal{C}}^{-1}=\mathcal{C}^{[I I}\right]_{\hat{e}}$, so that for all $i=1, \ldots, n$

$$
\hat{e}_{i}=R_{1 i}(\lambda) \vec{c}_{1}+\ldots+R_{n i}(\lambda) \vec{c}_{n} .
$$

Applying $\eta$ to both sides, we have

$$
\begin{aligned}
\hat{e}_{i} & =R_{1 i}(T) \eta\left(\vec{c}_{1}\right)+\ldots+R_{n i}(T) \eta\left(\vec{c}_{n}\right) \\
& =R_{r i}(T) \eta\left(\vec{c}_{r}\right)+\ldots+R_{n i}(T) \eta\left(\vec{c}_{n}\right)
\end{aligned}
$$

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

$$
W_{r}+\ldots+W_{n}=V
$$

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

$$
T \Gamma_{i j}(T)-C_{i j}(T)=\sum_{\ell} A_{i \ell} \Gamma_{\ell j}(T) \quad \Longrightarrow \quad C_{i j}(T)=\sum_{\ell}\left(T \delta_{i \ell}-A_{i \ell}\right) \Gamma_{\ell j}(T) .
$$

Therefore $\eta\left(\vec{c}_{j}\right):=$

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

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

and so

$$
\sum_{k} \operatorname{dim} W_{k} \geq \operatorname{dim} V=n
$$

Combining our inequalities,

$$
\begin{equation*}
\sum_{k} \operatorname{dim} W_{k}=n \tag{VI.C.16}
\end{equation*}
$$

and so the sum is direct:

$$
V=W_{r} \oplus \cdots \oplus W_{n}
$$

Also (VI.C.16) forces $\operatorname{dim} W_{k}=\operatorname{deg} \Delta_{k}(\lambda)$, so that

$$
\mathcal{B}_{k}=\left\{\eta\left(\vec{c}_{k}\right), T \eta\left(\vec{c}_{k}\right), \ldots, T^{m_{k}-1} \eta\left(\vec{c}_{k}\right)\right\}
$$

is a basis for $W_{k}$ (i.e., it was an independent set after all). Thus

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

as promised, and

$$
\mathcal{B}=\left\{\mathcal{B}_{r}, \ldots, \mathcal{B}_{n}\right\}
$$

is the basis for $V=\mathrm{F}^{n}$.
Let's sum up, in terms of a theorem about transformations, and an algorithm for the corresponding matrices:
VI.C.17. Theorem. For $T: V \rightarrow V(\operatorname{dim} V=n)$ there is a decomposition

$$
V=W_{r} \oplus \cdots \oplus W_{n}
$$

into T-cyclic subspaces, such that the minimal and characteristic polynomials of $\left.T\right|_{W_{k}}$ are both just the $k^{\text {th }}$ invariant factor $\Delta_{k}$ of $\lambda I-T$.

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

$$
P(\lambda I I-A) Q=\left(\begin{array}{cccccc}
1 & & & & & 0 \\
& \ddots & & & & \\
& & 1 & & & \\
& & & \Delta_{r}(\lambda) & & \\
& & & & \ddots & \\
0 & & & & & \Delta_{n}(\lambda)
\end{array}\right) .
$$

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

$$
\vec{c}_{k}=C_{1 k}(\lambda) \hat{e}_{1}+\ldots+C_{n k}(\lambda) \hat{e}_{n}
$$

and compute

$$
\eta\left(\vec{c}_{k}\right)=C_{1 k}(T) \hat{e}_{1}+\ldots+C_{n k}(T) \hat{e}_{n}:
$$

after applying the $T^{\prime}$ '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 $\eta\left(\vec{c}_{k}\right)=\vec{c}_{k}$.)

Let $m_{k}=\operatorname{deg} \Delta_{k}$, and write (for $k=r, \ldots, n$ )

$$
\mathcal{B}_{k}=\left\{\eta\left(\vec{c}_{k}\right), \operatorname{T\eta }\left(\vec{c}_{k}\right), \ldots, T^{m_{k}-1} \eta\left(\vec{c}_{k}\right)\right\} ;
$$

put these together to form a basis

$$
\mathcal{B}=\left\{\mathcal{B}_{r}, \ldots, \mathcal{B}_{n}\right\} \text { for } \mathrm{F}^{n} .
$$

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

$$
\mathcal{R}=\operatorname{diag}\left\{M\left(\Delta_{r}\right), \ldots, M\left(\Delta_{n}\right)\right\}
$$

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: $(A=)[T]_{\hat{e}}={ }_{\hat{e}}[\mathbb{I}]_{\mathcal{B}} \cdot[T]_{\mathcal{B}} \cdot \mathcal{B}[\mathbb{I}]_{\hat{e}}$, where $T$ is "transparent" relative to $\mathcal{B}$.
VI.C.19. EXAMPLE. We now put our favorite example

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

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

$$
n f(\lambda I I-A)=\left(\begin{array}{lll}
1 & & \\
& \lambda & \\
& & \lambda^{2}-3 \lambda
\end{array}\right)
$$

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 $(1-\lambda) \times\{$ 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 $(\lambda-1) \times\{$ 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 matrices ${ }^{14}$

$$
\begin{aligned}
& C=R_{1}^{-1} R_{2}^{-1} R_{3}^{-1} R_{4}^{-1} R_{5}^{-1} R_{6}^{-1} \\
& =\left(\begin{array}{lll} 
& 1 & \\
1 & & \\
& & 1
\end{array}\right)\left(\begin{array}{ccc}
-1 & & \\
& 1 & \\
& & 1
\end{array}\right)\left(\begin{array}{ccc}
1 & & \\
\lambda-1 & 1 & \\
& & \\
& & 1
\end{array}\right)\left(\begin{array}{ccc}
1 & & \\
& 1 & \\
-1 & & 1
\end{array}\right)\left(\begin{array}{ccc}
1 & & \\
& 1 & \\
& -1 & 1
\end{array}\right)\left(\begin{array}{lll}
1 & & \\
& -1 & \\
& & 1
\end{array}\right) \\
& =\left(\begin{array}{ccc}
\lambda-1 & -1 & 0 \\
-1 & 0 & 0 \\
-1 & 1 & 1
\end{array}\right) \quad \Longrightarrow \quad \vec{c}_{1}=\left(\begin{array}{c}
\lambda-1 \\
-1 \\
-1
\end{array}\right), \vec{c}_{2}=\left(\begin{array}{c}
-1 \\
0 \\
1
\end{array}\right), \vec{c}_{3}=\left(\begin{array}{l}
0 \\
0 \\
1
\end{array}\right) .
\end{aligned}
$$

Now according to the entries of $n f(\lambda I I-A)$, the invariant factors of $A$ are $\Delta_{1}=1, \Delta_{2}=\lambda, \Delta_{3}=\lambda^{2}-3 \lambda$. This means that we keep $\vec{c}_{2}=\eta\left(\vec{c}_{2}\right)$ and $\vec{c}_{3}=\eta\left(\vec{c}_{3}\right)$, 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 $\Delta_{1}=1 \Longrightarrow \eta\left(\vec{c}_{1}\right)=0$.

[^1]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 $\lambda$ :

$$
\begin{aligned}
\eta\left(\vec{c}_{1}\right) & =(T-1) \hat{e}_{1}+(-1) \hat{e}_{2}+(-1) \hat{e}_{3} \\
& =T \hat{e}_{1}-\left(\hat{e}_{1}+\hat{e}_{2}+\hat{e}_{3}\right) .
\end{aligned}
$$

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

$$
=\left(\begin{array}{lll}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{array}\right)\left(\begin{array}{l}
1 \\
0 \\
0
\end{array}\right)-\left\{\left(\begin{array}{l}
1 \\
0 \\
0
\end{array}\right)+\left(\begin{array}{l}
0 \\
1 \\
0
\end{array}\right)+\left(\begin{array}{l}
0 \\
0 \\
1
\end{array}\right)\right\}=\left(\begin{array}{l}
1 \\
1 \\
1
\end{array}\right)-\left(\begin{array}{l}
1 \\
1 \\
1
\end{array}\right)=\overrightarrow{0}
$$

Written in this way it looks kind of like magic.
Now for the remaining cases $k=2,3, \operatorname{dim}\left(W_{k}\right)=\#\left\{\right.$ vectors in $\left.\mathcal{B}_{k}\right\}$ must be the degree $m_{k}$ of $\Delta_{k} \cdot{ }^{15}$ Since $m_{2}=1$ and $m_{3}=2$,

$$
\mathcal{B}_{2}=\left\{\eta\left(\vec{c}_{2}\right)\right\}=\left\{\left(\begin{array}{r}
-1 \\
0 \\
1
\end{array}\right)\right\}, \quad \mathcal{B}_{3}=\left\{\eta\left(\vec{c}_{3}\right), T \eta\left(\vec{c}_{3}\right)\right\}=\left\{\left(\begin{array}{l}
0 \\
0 \\
1
\end{array}\right),\left(\begin{array}{l}
1 \\
1 \\
1
\end{array}\right)\right\} .
$$

These three vectors together give our basis and

$$
S=\left(\begin{array}{ccc}
-1 & 0 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1
\end{array}\right) \quad \xrightarrow{\operatorname{rref}} \quad S^{-1}=\left(\begin{array}{ccc}
-1 & 1 & 0 \\
1 & -2 & 1 \\
0 & 1 & 0
\end{array}\right)
$$

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

$$
\left(-a_{0}\right) \quad \text { and } \quad\left(\begin{array}{cc}
0 & -a_{0} \\
1 & -a_{1}
\end{array}\right)
$$

Therefore in our case

$$
M\left(\Delta_{2}\right)=(0) \quad \text { and } \quad M\left(\Delta_{3}\right)=\left(\begin{array}{ll}
0 & 0 \\
1 & 3
\end{array}\right)
$$

${ }^{15}$ That is, $\mathcal{B}_{k}$ is $\left\{\eta\left(\vec{c}_{k}\right), T \eta\left(\vec{c}_{k}\right), \ldots, T^{m_{k}-1} \eta\left(\vec{c}_{k}\right)\right\}$.

The decomposition is therefore (in the form $A=S \mathcal{R} S^{-1}$ )

$$
\left(\begin{array}{lll}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{array}\right)=\left(\begin{array}{ccc}
-1 & 0 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1
\end{array}\right)\left(\begin{array}{ccc}
\boxed{0} & 0 & 0 \\
0 & \begin{array}{|cc|}
0 & 0 \\
0 & 1
\end{array} 3
\end{array}\right)\left(\begin{array}{ccc}
-1 & 1 & 0 \\
1 & -2 & 1 \\
0 & 1 & 0
\end{array}\right) .
$$

## Exercises

(1) Compute the rational canonical form for

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

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

$$
A=\left(\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
-2 & -2 & 0 & 1 \\
-2 & 0 & -1 & -2
\end{array}\right)
$$

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

$$
A=\left(\begin{array}{llll}
1 & 0 & 0 & 0 \\
c & 1 & 0 & 0 \\
0 & c & 1 & 0 \\
0 & 0 & c & 1
\end{array}\right)
$$

How does it depend on $c$ ?
(4) Let $T: V \rightarrow V$ (with $V \cong \mathbb{R}^{n}$ ) be semisimple.
(a) If $T$ has a cyclic vector $\vec{v} \in V$ (i.e. $\vec{v}, T \vec{v}, \ldots, T^{n-1} \vec{v}$ span $V$ ), show that $T$ has $n$ distinct eigenvalues.
(b) If $T$ has $n$ distinct eigenvalues, then taking $\left\{\vec{v}_{1}, \ldots, \vec{v}_{n}\right\}$ 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 $n=2 k$, show that $A$ is similar (over $\mathbb{R}$ ) to

$$
J_{n}:=\left(\begin{array}{cc}
0 & -\mathbb{I}_{k} \\
\mathbb{I}_{k} & 0
\end{array}\right) .
$$

(6) Let $T: P^{3} \rightarrow P^{3}$ be $\frac{d}{d x}$. Write its matrix $A$ in the "standard" basis $\left\{x^{3}, x^{2}, x, 1\right\}=\mathcal{B}$, 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 $5 \times 5$ matrix $A$, without giving you the matrix: $f_{A}(\lambda)=(\lambda-3)^{3}(\lambda-2)^{2}$.
(a) What are the possible rational canonical forms $\mathcal{R}$ ? (Start by listing the possibilities for $n f(\lambda I I-A)$.) There should be six.
(b) This says that there are six "distinct" (= nonsimilar) transformations of $\mathbb{R}^{5}$ 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!]


[^0]:    ${ }^{11}$ Here $g_{i}(T) \hat{e}_{i}$ is just a polynomial in the transformation $T$ [or matrix $A$ ] acting on the $i^{\text {th }}$ standard basis vector $\in \mathrm{F}^{n}$.
    ${ }^{12}$ Finding $P^{-1}$ means taking the product of the elementary matrices corresponding to the reverse of the operations you did - only the row operations!

[^1]:    ${ }^{14}$ As an alternative (to avoid multiplying matrices), one could perform these as row operations on the identity matrix: $R_{1}^{-1}\left(\cdots\left(R_{6}^{-1} \mathbb{I}\right)\right)$.

