IV.D. Applications to linear algebra

Let $T \in \text{End}_{\mathbb{F}}(V)\backslash\{0\}$ be a nontrivial linear transformation of a finite-dimensional vector space V over a field $\mathbb{F}$. Take $\{x_i\}_{i=1}^n \subset V$ to be a basis and $B := {}_{\underline{x}}[T]$ to be the corresponding matrix, with entries $b_{ij} \in \mathbb{F}$. We have that $V = \oplus_{i=1}^n \mathbb{F}x_i = \sum_{i=1}^n \mathbb{F}[\lambda]x_i$, where V has the structure of an $\mathbb{F}[\lambda]$-module by $P(\lambda)v := P(T)v$ (for any polynomial $P(\lambda) \in \mathbb{F}[\lambda]$). Since $\mathbb{F}[\lambda]$ is *not* f.g. as an $\mathbb{F}$-module, *V must be a (f.g.) torsion $\mathbb{F}[\lambda]$-module.*

We have a short-exact sequence

$$K := \ker(\eta) \hookrightarrow \mathbb{F}[\lambda]^n \overset{\eta}{\twoheadrightarrow} V$$

$$\mathbf{e}_i \mapsto x_i$$

of $\mathbb{F}[\lambda]$-modules, in which K must be free with generators $\{f_i\}_{i=1}^n$. To obtain the (d_j) which will be annihilators of the $\mathbb{F}[\lambda]z_j$ in the structure theorem decomposition, we must find (then put in normal form) a matrix whose columns express the $\{f_j\}$ in terms of the $\{\mathbf{e}_i\}$. To wit:

IV.D.1. LEMMA. *$A := \lambda\mathbb{1}_n - B$ is a relations matrix for V.*

PROOF. We need to specify the $\{f_j\}$. Put

$$f_j := \lambda\mathbf{e}_j - \sum_{i=1}^n b_{ij}\mathbf{e}_i.$$

Clearly $\eta(f_j) = \lambda\eta(\mathbf{e}_j) - \sum_i b_{ij}\eta(\mathbf{e}_i) = T(x_j) - \sum_i b_{ij}x_i = 0$, by definition of B. So $f_j \in K$ ($\forall j$).

To see that they generate K, suppose $0 = \eta(\sum_j P_j(\lambda)\mathbf{e}_j)$ for some polynomials P_j. By repeatedly applying $\lambda^k\mathbf{e}_j = \lambda^{k-1}\lambda\mathbf{e}_j = \lambda^{k-1}f_j + \sum_i b_{ij}\lambda^{k-1}\mathbf{e}_j$, we may rewrite this as $0 = \eta(\sum_j Q_j(\lambda)f_j + \sum_i \beta_i\mathbf{e}_i)$ with $\beta_i \in \mathbb{F}$. That is, $0 = \sum_j Q_j(T)\cancelto{0}{\eta(f_j)} + \sum_i \beta_i x_i \implies \beta_i = 0$ ($\forall i$). Hence $\sum_j P_j(\lambda)\mathbf{e}_j = \sum_j Q_j(\lambda)f_j \in \mathbb{F}[\lambda]\langle f_1, \ldots, f_n\rangle$. $\square$

IV.D.2. REMARK. In fact, we can prove that $\{f_j\}$ *is a base for K over* $\mathbb{F}[\lambda]$: given $\sum_j h_j(\lambda)f_j = 0$, we have

$$(\sum_i h_i(\lambda)\lambda\mathbf{e}_i =)\ \sum_j h_j(\lambda)\lambda\mathbf{e}_j = \sum_{i,j} h_j(\lambda)b_{ij}\mathbf{e}_i \quad (\text{in } \mathbb{F}[\lambda]^n)$$

$\implies h_i(\lambda)\lambda = \sum_j h_j(\lambda)b_{ij}$ (in $\mathbb{F}[\lambda]$) for each i. But this is impossible: consider i such that h_i is of maximal degree: then deg(LHS) > deg(RHS).

Apply the normal form algorithm to obtain bases $\{e_i'\}$ and $\{f_j'\}$ (for $\mathbb{F}[\lambda]^n$ resp. K) related by

$$\text{(IV.D.3)} \qquad \begin{aligned} Q(\lambda \mathbb{1}_n - B)P &= \operatorname{diag}(d_1, \ldots, d_{k_0}, d_{k_0+1}, \ldots, d_n) \\ &= \operatorname{diag}(1, \ldots, 1, \delta_1, \ldots, \delta_s) \end{aligned}$$

in the notation of the structure theorem and its proof (with $k = n$ and $\ell = s$ since V is torsion). That is, $f_i' = d_i e_i'$ is our new base for $K = \ker(\eta)$.

Now put $\eta(e_i') =: x_i'$. This is not a basis for V as a vector space ($\mathbb{F}$-module), since $x_1', \ldots, x_{k_0}' = 0$. However, the remaining nonzero elements $x_{k_0+1}' =: z_1, \ldots, x_n' =: z_s$ must generate V as an $\mathbb{F}[\lambda]$-module; and indeed by the structure theorem we have

$$\text{(IV.D.4)} \quad V = \mathbb{F}[\lambda]z_1 \oplus \cdots \oplus \mathbb{F}[\lambda]z_s \cong \mathbb{F}[\lambda]/(\delta_1) \oplus \cdots \oplus \mathbb{F}[\lambda]/(\delta_s),$$

with $\delta_1 \mid \cdots \mid \delta_s$ nonzero nonunits, i.e. polynomials of positive degree.

The canonical forms. The direct sum decomposition in (IV.D.4) also expresses V as an internal direct sum of s subspaces, whose dimensions obviously must add to n. We start by examining the matrix of the restriction of T to one such subspace.

Pick an $i \in \{1, \ldots, s\}$ and write

$$\delta_i =: F(\lambda) = \lambda^m + F_{m-1}\lambda^{m-1} + \cdots + F_0.$$

Since δ_i has degree m,

$$\text{(IV.D.5)} \qquad z_i,\ Tz_i,\ \ldots,\ T^{m-1}z_i$$

are linearly independent over $\mathbb{F}$ and span $\mathbb{F}[\lambda]z_i$. Moreover,

$$\begin{aligned} 0 = F(\lambda)z_i = F(T)z_i \quad \implies \quad T(T^{m-1}z_i) &= T^m z_i = (T^m - F(T))z_i \\ &= -F_0 z_i - F_1 T z_i - \cdots - F_{m-1}T^{m-1}z_i. \end{aligned}$$

We conclude that in the basis (IV.D.5) of $\mathbb{F}[\lambda]z_i$, the restriction $T|_{\mathbb{F}[\lambda]z_i}$ has matrix

$$C_F := \begin{pmatrix} 0 & & & & & -F_0 \\ 1 & 0 & & & & -F_1 \\ & 1 & 0 & & & -F_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & 0 & -F_{m-2} \\ & & & & 1 & -F_{m-1} \end{pmatrix}, \tag{IV.D.6}$$

which is called the **companion matrix** of the monic polynomial F.

Doing the same thing for each of the subspaces in (IV.D.3) we find first that

$$\underline{\tilde{z}} := \{z_1, Tz_1, \ldots, T^{\deg(\delta_1)-1}; \ \ldots; \ z_s, Tz_s, \ldots, T^{\deg(\delta_s)-1}z_s\}$$

is a basis of V; in particular, $\sum_{i=1}^{s} \deg(\delta_s) = n$. Writing T in this basis produces a block diagonal matrix with $\deg(\delta_i) \times \deg(\delta_i)$ blocks

$$_{\underline{\tilde{z}}}[T] = \mathrm{diag}(C_{\delta_1}, \ldots, C_{\delta_s}) \tag{IV.D.7}$$

which is called the **rational canonical form** of the original matrix B. The point, of course, is that this new matrix is *similar* to B: taking $S := {}_{\underline{\tilde{z}}}[\mathrm{id}_V]_{\underline{x}}$ the change-of-basis matrix, we have $SBS^{-1} =$ (IV.D.7).

Next, assume that the $\{\delta_i\}$ can be completely factored into *linear* factors[20] over $\mathbb{F}$. In this case we get more useful bases for each subspace $\mathbb{F}[\lambda]z_i$ by decomposing it into primary cyclic submodules.

For example, if $\delta_i = (\lambda - \alpha_1)^{e_1}(\lambda - \alpha_2)^{e_2}$, take $y = (\lambda - \alpha_2)^{e_2}z_i$ and $w = (\lambda - \alpha_1)^{e_1}z_i$, and observe that by (IV.C.26),

$$\mathbb{F}[\lambda]z_i = \mathbb{F}[\lambda]y \oplus \mathbb{F}[\lambda]w \cong \frac{\mathbb{F}[\lambda]}{((\lambda - \alpha_1)^{e_1})} \oplus \frac{\mathbb{F}[\lambda]}{((\lambda - \alpha_2)^{e_2})}. \tag{IV.D.8}$$

Clearly $\{y, (T - \alpha_1)y, (T - \alpha_1)^2 y, \ldots, (T - \alpha_1)^{e_1 - 1}y\}$ is an $\mathbb{F}$-basis for $\mathbb{F}[\lambda]y$, and similarly for $\mathbb{F}[\lambda]w$. Writing the restriction of T to $\mathbb{F}[\lambda]y$

[20] This is always true if $\mathbb{F}$ is *algebraically closed*, which is to say that every polynomial over $\mathbb{F}$ has a root in $\mathbb{F}$. For instance, $\mathbb{C}$ is algebraically closed.

with respect to this basis gives the $e_1 \times e_1$ matrix

$$J_{e_1}(\alpha_1) := \begin{pmatrix} \alpha_1 & & & & \\ 1 & \alpha_1 & & & \\ & 1 & \ddots & & \\ & & \ddots & \alpha_1 & \\ & & & 1 & \alpha_1 \end{pmatrix} \tag{IV.D.9}$$

since

$$\begin{cases} Ty = (T-\alpha_1)y + \alpha_1 y \\ T((T-\alpha_1)y) = (T-\alpha_1)^{e_1} y + \alpha_1(T-\alpha_1)y \\ \text{etc.} \end{cases}$$

Repeating this process for each δ_i yields, as before, a basis for V. Writing T with respect to this basis produces a block diagonal matrix

$$\operatorname{diag}(J_{e_1}(\alpha_1), J_{e_2}(\alpha_2), \ldots) \tag{IV.D.10}$$

called the **Jordan canonical form**, which is again *similar* to B.

IV.D.11. DEFINITION. The Jordan form reveals the **generalized eigenspaces** E_α of V with respect to T. We set

$$E_\alpha(T) := \mathcal{A}_{(\lambda-\alpha)}(V) = \{v \in V \mid (\lambda-\alpha)^k v = 0 \text{ for some } k \in \mathbb{N}\}.$$

Clearly this is the span of the basis elements corresponding to the blocks $J_{e_i}(\alpha_i)$ in (IV.D.10) with $\alpha_i = \alpha$, so that

$$\dim(E_\alpha(T)) = \sum_{i\colon \alpha_i = \alpha} e_i.$$

To summarize, Jordan canonical form corresponds to the primary cyclic decomposition of V as an $\mathbb{F}[\lambda]$-module, and the rational canonical form to the (less refined) decomposition in the structure theorem. Let's try a basic

IV.D.12. EXAMPLE. We recall from Example IV.C.17, that for

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

we have $\mathrm{nf}(\lambda\mathbb{1}_3 - B) = \mathrm{diag}(1,\ \lambda,\ \lambda^2 - 3\lambda)$. Hence

$$V = \mathbb{F}[\lambda]z_1 \oplus \mathbb{F}[\lambda]z_2 \cong \mathbb{F}[\lambda]/(\lambda) \oplus \mathbb{F}[\lambda]/(\lambda^2 - 3\lambda),$$

and with respect to the basis $\underset{\sim}{z} = \{z_1, z_2, Tz_2\}$ we get the rational canonical form

$$\left(\begin{array}{c|cc} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 1 & 3 \end{array}\right)$$

since $T(z_1) = 0$, $T(z_2) = Tz_2$, and $T(Tz_2) = T^2z_2 = 3Tz_2$ (from $T^2 - 3T = 0$ on $\mathbb{F}[\lambda]z_2$).

For the Jordan form, we factor $\delta_2(\lambda) = \lambda^2 - 3\lambda = \lambda(\lambda - 3)$ to further decompose V into primary cyclic modules:

$$\begin{aligned} V &= \mathbb{F}[\lambda]z_1 \oplus \mathbb{F}[\lambda](T - 3\,\mathrm{id}_V)z_2 \oplus \mathbb{F}[\lambda]Tz_2 \\ &\cong \mathbb{F}[\lambda]/(\lambda) \oplus \mathbb{F}[\lambda]/(\lambda) \oplus \mathbb{F}[\lambda]/(\lambda - 3). \end{aligned}$$

Of course, T kills the first two generators and $T(Tz_2) = 3(Tz_2)$ so the Jordan form is

$$\left(\begin{array}{c|c:c} 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hdashline 0 & 0 & 3 \end{array}\right)$$

and $\dim(E_0(B)) = 2$, $\dim(E_3(B)) = 1$. After all, if a matrix can be diagonalized, the Jordan form is diagonal. This happens precisely when the δ_i (taken individually) have no repeated linear factors.

One thing you may wonder is how to *find* the basis (or change-of-basis matrix) which puts B in rational or Jordan canonical form. We have (writing $\theta\colon K \hookrightarrow \mathbb{F}[\lambda]^n$ for the inclusion)

$$\begin{aligned} {}_{e'}[\theta]_{f'} &= \mathrm{nf}(\lambda\mathbb{1}_3 - B) = Q(\lambda\mathbb{1}_3 - B)P \\ &= {}_{e'}[\mathrm{id}_{\mathbb{F}[\lambda]^n}]_{\mathbf{e}} \cdot {}_{\mathbf{e}}[\theta]_f \cdot {}_f[\mathrm{id}_K]_{f'}\,, \end{aligned}$$

so that $Q = {}_{e'}[\mathrm{id}]_{\mathbf{e}} \implies$ columns of $Q^{-1} = {}_{\mathbf{e}}[\mathrm{id}]_{e'}$ yield the e'-basis (written in the $\mathbf{e}$-basis). One builds the basis $\underset{\sim}{z}$ for the rational (or Jordan) form from the $x_i' := \eta(e_i')$ for $i = k_0 + 1, \ldots, n$. Noting that

$x_j := \eta(\mathbf{e}_j)$, if (say) the last column of Q^{-1} is

$$_{\mathbf{e}}[e_n'] = \begin{pmatrix} p_1(\lambda) \\ \vdots \\ p_n(\lambda) \end{pmatrix} = \sum \lambda^k \begin{pmatrix} a_1^{(k)} \\ \vdots \\ a_n^{(k)} \end{pmatrix} ,$$

then applying η yields[21]

$$\underline{x}[x_n'] = \sum_k \underline{x}[T^k] \begin{pmatrix} a_1^{(k)} \\ \vdots \\ a_n^{(k)} \end{pmatrix} = \sum_k B^k \begin{pmatrix} a_1^{(k)} \\ \vdots \\ a_n^{(k)} \end{pmatrix} .$$

But this is a bit ugly and there are often better ways to proceed:

IV.D.13. EXAMPLE. The matrix

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

has characteristic polynomial

$$p_B(\lambda) = \det(\lambda \mathbb{1}_4 - B) = (\lambda - 1)^3 (\lambda - 2).$$

This guides the selection of our basis: this is straightforward for eigenvalue 2, as

$$E_2(B) = \ker(B - 2\mathbb{1}_4) = \left\langle \left(\begin{smallmatrix} -2 \\ 1 \\ 1 \\ 0 \end{smallmatrix} \right) \right\rangle =: \langle v_1 \rangle.$$

For eigenvalue 1, first find bases for kernels of powers of $(B - \mathbb{1})$:

$$\ker(B - \mathbb{1}) = \left\langle \left(\begin{smallmatrix} 0 \\ -1 \\ 0 \\ 1 \end{smallmatrix} \right) \right\rangle$$

$$\subset \ \ker((B - \mathbb{1})^2) = \left\langle \left(\begin{smallmatrix} 0 \\ -1 \\ 0 \\ 1 \end{smallmatrix} \right), \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \\ 0 \end{smallmatrix} \right) \right\rangle$$

$$\subset \ E_1(B) = \ker((B - \mathbb{1})^3) = \left\langle \left(\begin{smallmatrix} 0 \\ -1 \\ 0 \\ 1 \end{smallmatrix} \right), \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \\ 0 \end{smallmatrix} \right), \left(\begin{smallmatrix} 0 \\ 0 \\ 0 \\ 1 \end{smallmatrix} \right) \right\rangle .$$

[21] This is essentially what [**Jacobson**] does in the Example on his pp. 198-199, though as usual his convention is the transpose of that used in these notes.

(So far, the bases for kernels are easily computed by taking rref of $B - 2\mathbb{1}$, $B - \mathbb{1}$, $(B - \mathbb{1})^2$, and $(B - \mathbb{1})^3$.) It is the last basis vector for $E_1(B)$ that generates it as a $\mathbb{Q}[\lambda]$-module, and we choose its cyclic images as our remaining basis vectors for V:

$$v_2 := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \mapsto v_3 := (B - \mathbb{1})v_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \mapsto v_4 := (B - \mathbb{1})^2 v_2 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Taking S to be the matrix with columns given by the $\{v_i\}$, we get

$$B = S \left(\begin{array}{c|ccc} 2 & & & \\ \hline & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \end{array} \right) S^{-1}.$$

The minimal polynomial. We previously used this term for an element of an algebraic extension of a field. But it makes sense for any finitely generated torsion module M over a PID R, by the structure theorem. In the notation of IV.C.18, since the free part is zero ($t = 0$ and $\ell = s$), each direct summand is annihilated by some δ_i. Since all of these divide δ_s, we have $\delta_s M = \{0\}$. Conversely, if $rM = \{0\}$ for some $r \in R$, then $r \in (\delta_1) \cap \cdots \cap (\delta_s) = (\delta_s)$. So $(\delta_s) \subset R$ is the set of all elements annihilating M.

So in the special case under study here (cf. (IV.D.4)), $(d_n) \subset \mathbb{F}[\lambda]$ is the annihilator of V. An immediate consequence is the

IV.D.14. THEOREM. *$d_n(T)$ $(= \delta_s(T))$ is the zero transformation, and if $F \in \mathbb{F}[\lambda]$ satisfies $F(T) = 0$, then $d_n \mid F$. The same holds with "B" [resp. "matrix"] replacing "T" [resp. "transformation"].*

PROOF. For the second part, just note that $F(\lambda)V = \{0\} \iff F(T)v = 0\ (\forall v \in V) \iff F(T)x_i = 0\ (\forall i) \iff F(B) = 0$. $\square$

IV.D.15. DEFINITION. $d_n(\lambda)$ is the **minimal polynomial** of T (or B). We will henceforth write this m_T (or m_B).

IV.D.16. PROPOSITION. *(a)* $m_B(\lambda) = \dfrac{\det(\lambda\mathbb{1} - B)}{\left\{ \begin{array}{c} \text{monic gcd of } (n-1) \times (n-1) \\ \text{minors of } \lambda\mathbb{1} - B \end{array} \right\}}$.

(b) m_B and $p_B := \det(\lambda\mathbb{1} - B)$ are invariant under similarity.

PROOF. (a) This is just $d_n = \Delta_n / \Delta_{n-1}$.
(b) If $B' = SBS^{-1}$, $S \in \mathrm{GL}_n(\mathbb{F})$, then B and B' are matrices of the same T (with respect to different bases of V). The invariant factors d_i in $\mathbb{F}[\lambda]$ are defined for the $\mathbb{F}[\lambda]$-module V, which itself depends only on T. □

Notice that the coefficients of powers of λ in $p_B(\lambda)$ are therefore *polynomials in the entries of B that are invariant under similarity transformation* (conjugation by an invertible S). These include the trace and determinant.

Finally we have the

IV.D.17. COROLLARY (Cayley-Hamilton). $p_B(B) = 0$.

PROOF. We have $p_T(\lambda) := \det(\lambda \mathrm{id}_V - T) = d_{s+1}(\lambda) \cdots d_n(\lambda)$, hence $p_T(T) = d_{s+1}(T) \cdots d_n(T) = 0$ (since $d_n(T) = 0$). □

This looks much simpler than the proofs in linear algebra courses, because we have already proved a more difficult result using module theory. In any case, writing $p_B(B) = \det(B\mathbb{1} - B) = \det(0) = 0$ is still wrong, because you have to first expand the determinant and *then* substitute in the matrix, not the other way around!