VI.E. Jordan normal form

Set $V = \mathbb{C}^n$ and let $T\colon V \to V$ be any linear transformation, with distinct eigenvalues $\sigma_1, \ldots, \sigma_s$. In the last lecture we showed that V decomposes into stable eigenspaces for T:

$$V = W_1 \oplus \cdots \oplus W_s = \widetilde{\ker}(T - \sigma_1 \mathbb{I}) \oplus \cdots \oplus \widetilde{\ker}(T - \sigma_s \mathbb{I}).$$

Let $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_s\}$ be a basis for V subordinate to this direct sum and set $B_k = [T\,|_{W_k}]_{\mathcal{B}_k}$, so that

$$[T]_{\mathcal{B}} = \mathrm{diag}\{B_1, \ldots, B_s\}.$$

Each B_k has only σ_k as eigenvalue. In the event that $A = [T]_{\hat{e}}$ is diagonalizable, or equivalently $\widetilde{\ker}(T - \sigma_k\mathbb{I}) = \ker(T - \sigma_k\mathbb{I})$ (or their dimensions $\tilde{d}_k = d_k$) for all k, $\mathcal{B}$ is an eigenbasis and $[T]_{\mathcal{B}}$ is a diagonal matrix

$$\mathrm{diag}\{\underbrace{\sigma_1, \ldots, \sigma_1}_{\tilde{d}_1 = \dim W_1}\,;\, \ldots\,;\, \underbrace{\sigma_s, \ldots, \sigma_s}_{\tilde{d}_s = \dim W_s}\}.$$

Otherwise we must perform further surgery on the $\mathcal{B}_k$'s separately, in order to transform the blocks B_k (and so the entire matrix for T) into the "simplest possible" form.

What might this form be? Consider $T = \frac{d}{dx}$ acting on the space of functions $f(x)$ satisfying $f''' - 3\sigma f'' + 3\sigma^2 f' - \sigma^3 f = 0$ (or equivalently $(\frac{d}{dx} - \sigma)^3 f = 0$), with basis $\mathcal{B} = \{\frac{x^2}{2}e^{\sigma x}, xe^{\sigma x}, e^{\sigma x}\}$. The matrix

$$[T]_{\mathcal{B}} = \begin{pmatrix} \sigma & 0 & 0 \\ 1 & \sigma & 0 \\ 0 & 1 & \sigma \end{pmatrix}$$

is an example of a *little Jordan block*. This is clearly quite natural, and is the sort of thing that we would like to generalize.

The attentive reader will have noticed above that I have written $T - \sigma_k\mathbb{I}$ in place of $\sigma_k\mathbb{I} - T$. This is a strategic move: when dealing with characteristic polynomials it is far more convenient to write $\det(\lambda\mathbb{I} - A)$ to produce a *monic* polynomial. On the other hand, as you'll see now, it is better to work on the individual W_k's with the nilpotent transformation $T\,|_{W_k} - \sigma_k\mathbb{I} =: N_k$.

Decomposition of the Stable Eigenspaces (Take 1). Let's briefly omit subscripts and consider $T\colon W \to W$ with one eigenvalue σ, $\dim W = \tilde{d}$, $\mathcal{B}$ a basis for W and $[T]_{\mathcal{B}} = B$. The operator $N = T - \sigma\mathbb{I}$ (with matrix $[N]_{\mathcal{B}} = B - \sigma\mathbb{I}_{\tilde{d}}$) is nilpotent, and so $f_N(\lambda) = \lambda^{\tilde{d}}$ (since its only eigenvalue is 0). Since $f_N(\lambda)$ must be the product of the invariant factors (of $\lambda\mathbb{I} - N$), the normal form of $\lambda\mathbb{I} - N$ is quite limited:

$$nf(\lambda\mathbb{I} - N) = \mathrm{diag}\{1, \ldots, 1, \lambda^{q^{(r)}}, \ldots, \lambda^{q^{(\tilde{d})}}\}$$

where[20] $q^{(r)} + \ldots + q^{(\tilde{d})} = \tilde{d}$.

To express the corresponding rational canonical form for N we introduce the following notation:[21]

$$\mathcal{N}_\sigma^q := \begin{pmatrix} \sigma & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & \sigma \end{pmatrix} \qquad [q \times q \text{ matrix}].$$

Then

$$\text{(VI.E.1)} \qquad \mathcal{R}(N) = \mathrm{diag}\left\{\mathcal{N}_0^{q^{(r)}}, \ldots, \mathcal{N}_0^{q^{(\tilde{d})}}\right\};$$

for instance, if

$$nf(\lambda\mathbb{I} - N) = \mathrm{diag}\{1, 1, 1, \lambda^2, \lambda^3\},$$

then

$$\mathcal{R}(N) = \begin{pmatrix} \boxed{\begin{matrix} 0 & \\ 1 & 0 \end{matrix}} & \\ & \boxed{\begin{matrix} 0 & & \\ 1 & 0 & \\ & 1 & 0 \end{matrix}} \end{pmatrix}.$$

(Notice the apparent "gap" between the first and second 1's.)

[20]Note: when you see superscripts in parentheses in this section (viz., $(r), \ldots, (\tilde{d})$), they are not exponents. (This is so that we can also add subscripts later.)

[21]A possibly more standard notation here is $J_q(\sigma)$, where the "J" stands for Jordan. We reserve that letter here for the Jordan form of a given matrix A.

So for some change-of-basis matrix S we have

$$B - \sigma\mathbb{I}_{\tilde{d}} = [N]_{\mathcal{B}} = S\,\mathcal{R}(N)\,S^{-1}$$

or

$$B = S\,\mathcal{R}(N)\,S^{-1} + \sigma\mathbb{I}_{\tilde{d}} = S\,(\mathcal{R}(N) + \sigma\mathbb{I}_{\tilde{d}})\,S^{-1}.$$

Using (VI.E.1) this means that if, say, $S = P_{\mathcal{C}\to\mathcal{B}}$ then

$$[T]_{\mathcal{C}} = S^{-1}BS = \operatorname{diag}\left\{\mathcal{N}_0^{q^{(r)}}, \ldots, \mathcal{N}_0^{q^{(\tilde{d})}}\right\} + \sigma\mathbb{I}_{\tilde{d}}$$
$$= \operatorname{diag}\left\{\mathcal{N}_\sigma^{q^{(r)}}, \ldots, \mathcal{N}_\sigma^{q^{(\tilde{d})}}\right\};$$

for example,

$$\begin{pmatrix} \boxed{\begin{matrix} \sigma & \\ 1 & \sigma \end{matrix}} & \\ & \boxed{\begin{matrix} \sigma & & \\ 1 & \sigma & \\ & 1 & \sigma \end{matrix}} \end{pmatrix} = \operatorname{diag}\left\{\mathcal{N}_\sigma^2, \mathcal{N}_\sigma^3\right\}.$$

This is called a *big Jordan block*, and the "boxes" $\mathcal{N}_\sigma^q$ are little Jordan blocks. They have very simple characteristic polynomials, namely $(\lambda - \sigma)^q$ (after all, $T\colon W \to W$ has only eigenvalue σ). The little blocks correspond to a decomposition of W into N-cyclic subspaces

$$W = W^{(r)} \oplus \cdots \oplus W^{(\tilde{d})}$$

as in the rational canonical structure theorem (Thm. VI.C.17). $\mathcal{C}$ is a basis subordinate to this decomposition, consisting of a choice of "cyclic vector" for each $W^{(i)}$ and its successive images under N; more on this later.

Going back to a transformation $T\colon V \to V$ with multiple eigenvalues, what we are aiming at is a further decomposition of each of the $W_k = \widetilde{\ker}(T - \sigma_k\mathbb{I})$ into N_k-cyclic subspaces:

$$V = \left(W_1^{(r_1)} \oplus \cdots \oplus W_1^{(\tilde{d}_1)}\right) \oplus \cdots \oplus \left(W_s^{(r_s)} \oplus \cdots \oplus W_s^{(\tilde{d}_s)}\right).$$

With respect to some special basis $\mathcal{C}$ subordinate to this entire direct sum decomposition,[22] the matrix of T will be the *Jordan normal form*

$$\mathcal{J}(T) = \mathrm{diag}\{\mathcal{N}_{\sigma_1}^{q_1^{(r_1)}}, \ldots, \mathcal{N}_{\sigma_1}^{q_1^{(\tilde{d}_1)}} ; \ \ldots \ ; \mathcal{N}_{\sigma_s}^{q_s^{(r_s)}}, \ldots, \mathcal{N}_{\sigma_s}^{q_s^{(\tilde{d}_s)}}\},$$

where each semicolon separates two big Jordan blocks (and each comma two little Jordan blocks). Note that there is one big Jordan block for each eigenvalue. As it stands this looks formidable, and at least the basis $\mathcal{C}$ looks very hard to compute.

Decomposition of W_k (Take 2). In order to facilitate computation of the entire Jordan decomposition $A = S\mathcal{J}S^{-1}$, it is useful to have an approach to the "nilpotent building blocks" $N_k \colon W_k \to W_k$ that does *not* appeal to rational canonical form.

VI.E.2. DEFINITION. The **height** h of a nilpotent transformation $N \colon W \to W$ is defined by $N^h = 0$, $N^{h-1} \neq 0$. (The minimal polynomial of such a transformation is simply λ^h.)

One may also define the (N-)*height* $h(\vec{w})$ *of* $\vec{w} \in W$ by $N^{h(\vec{w})}\vec{w} = 0$, $N^{h(\vec{w})-1}\vec{w} \neq 0$. Clearly the height of N is just the supremum of the (N-)heights of vectors in W. (The height of a subspace of W, similarly, just means the supremum of heights of vectors in that subspace.)

Notation: We shall write $Z_N^{h_0}(\vec{w})$ for the N-cyclic subspace of W generated by $\vec{w}$ ($\vec{w}$ of height h_0); that is,

$$Z_N^{h_0}(\vec{w}) := span\{\vec{w}, N\vec{w}, \ldots, N^{h_0-1}\vec{w}\}.$$

VI.E.3. CLAIM. *Given $N \colon W \to W$ nilpotent of height h, with*

$$\{N^{h-1}\vec{w}_1, \ldots, N^{h-1}\vec{w}_t\}$$

a basis for $\mathrm{im}(N^{h-1})$. *Then we have a decomposition*

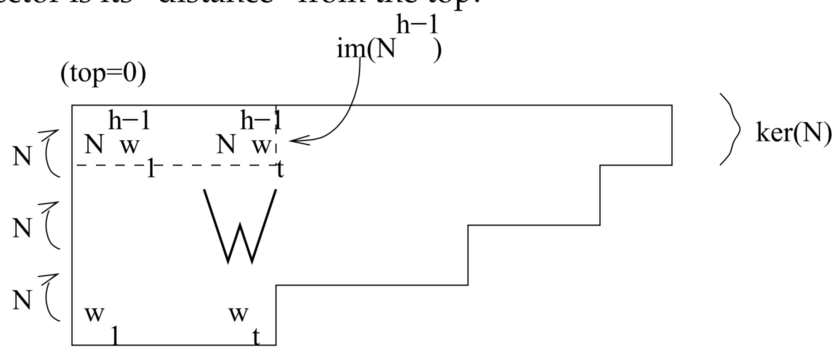
$$W = Z_N^h(\vec{w}_1) \oplus \cdots \oplus Z_N^h(\vec{w}_t) \oplus \underbrace{W^{(t+1)} \oplus \cdots \oplus W^{(m)}}_{\text{cyclic of height } \leq h-1} .$$

[22] Compatibility with the direct sum alone is of course not enough; "special" means the basis also has "cyclic" properties (to be described below)

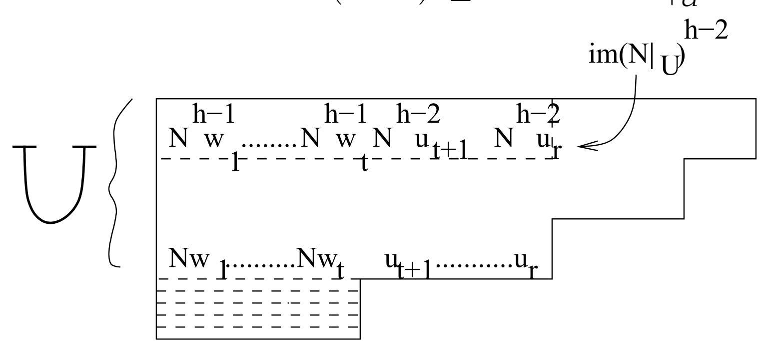
PROOF (BY INDUCTION ON h).

• base case ($h = 1$): $W = \ker N$. Take any basis for W; the spans of the individual basis vectors give the desired decomposition.

• inductive step: The following picture of W is useful, where N acts by sending you up one level, the top row is $\ker N$, and the height of a vector is its "distance" from the top.



In order to apply the inductive hypothesis (= the *Claim* for height $h-1$), we consider $U = \ker(N^{h-1}) \subseteq W$. Now $N|_U$ has height $h-1$:



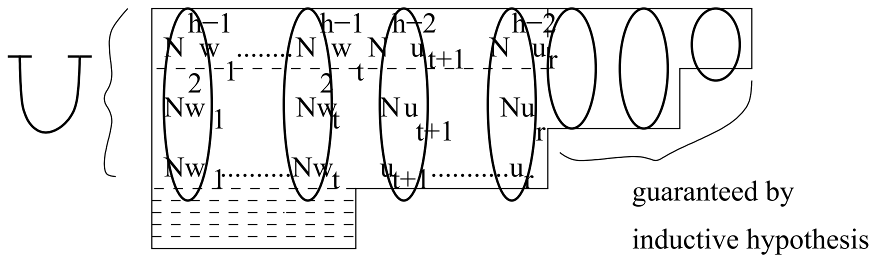
Set $\{\vec{u}_1, \ldots, \vec{u}_t\} = \{N\vec{w}_1, \ldots, N\vec{w}_t\} \subseteq U$, and complete

$$\{N^{h-2}\vec{u}_1, \ldots, N^{h-2}\vec{u}_t\} \; \Big(= \{N^{h-1}\vec{w}_1, \ldots, N^{h-1}\vec{w}_t\}\Big)$$

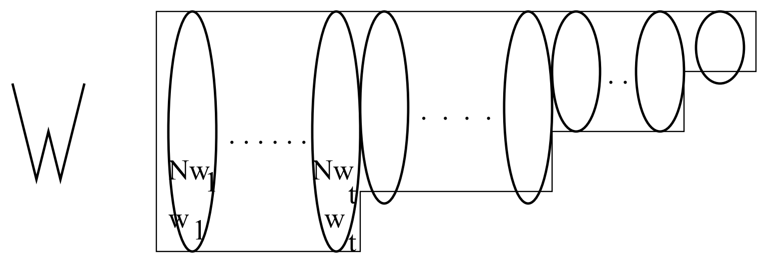
to a basis for $\mathrm{im}\{(N|_U)^{h-2}\}$, by adding some $\{N^{h-2}\vec{u}_{t+1}, \ldots, N^{h-2}\vec{u}_r\}$ as shown. By the inductive hypothesis,

$$U = Z_N^{h-1}(\vec{u}_1) \oplus \cdots \oplus Z_N^{h-1}(\vec{u}_t) \oplus Z_N^{h-1}(\vec{u}_{t+1}) \oplus \cdots \oplus Z_N^{h-1}(\vec{u}_r)$$
$$\oplus \underbrace{U^{(r+1)} \oplus \cdots \oplus U^{(m)}}_{N\text{-cyclic of height } \leq h-2} .$$

The picture is now



— U is the direct sum of the columns (= cyclic subspaces under the action of N). So why not tack $\vec{w}_1, \ldots, \vec{w}_t$ on to the bottom of the first t columns and obtain a decomposition



as promised?

Here's how to do this rigorously: since $N^{h-1}\vec{w}_1, \ldots, N^{h-1}\vec{w}_t$ are a *basis* for $\mathrm{im}(N^{h-1})$, no nontrivial linear combination of $\vec{w}_1, \ldots, \vec{w}_t$ can be in $\ker(N^{h-1}) = U$. Therefore

$$\text{(VI.E.4)} \qquad span\{\vec{w}_1, \ldots, \vec{w}_t\} \cap U = \{0\},$$

which implies that

$$\dim(span\{\vec{w}_1, \ldots, \vec{w}_t\}) + \dim U \;=\; \dim(span\{\vec{w}_1 \ldots \vec{w}_t\} + U).$$

Applying Rank+Nullity to N^{h-1}, we then find

$$\dim W \;=\; \dim(span\{\vec{w}_1, \ldots, \vec{w}_t\} + U)$$
$$\implies \quad W = span\{\vec{w}_1, \ldots, \vec{w}_t\} + U.$$

Together with (VI.E.4) this yields

$$W \;=\; span\{\vec{w}_1, \ldots, \vec{w}_t\} \oplus U,$$

which by our decomposition of U

$$= \left(span\{\vec{w}_1, \ldots, \vec{w}_t\} \oplus Z_N^{h-1}(\vec{u}_1) \oplus \cdots \oplus Z_N^{h-1}(\vec{u}_t) \right)$$
$$\oplus Z_N^{h-1}(\vec{u}_{t+1}) \oplus \ldots \oplus Z_N^{h-1}(\vec{u}_r) \oplus U^{(r+1)} \oplus \ldots \oplus U^{(m)}.$$

Since $\{\vec{u}_1, \ldots, \vec{u}_t\}$ was just $\{N\vec{w}_1, \ldots, N\vec{w}_t\}$, the direct sum in the parentheses is

$$Z_N^h(\vec{w}_1) \oplus \cdots \oplus Z_N^h(\vec{w}_t)$$

and we are done. □

The upshot of the *Claim* just proved is this (weaker)

VI.E.5. PROPOSITION. *$N: W \to W$ nilpotent $\Longrightarrow$*

$$W = W^{(1)} \oplus \cdots \oplus W^{(m)} = Z_N^{h_1}(\vec{w}_1) \oplus \cdots \oplus Z_N^{h_m}(\vec{w}_m);$$

that is, W decomposes into a direct sum of N-cyclic subspaces of various heights.[23]

So if we take

$$\mathcal{C} = \{\vec{w}_1, N\vec{w}_1, \ldots, N^{h_1-1}\vec{w}_1; \ \ldots \ ; \vec{w}_m, N\vec{w}_m, \ldots, N^{h_m-1}\vec{w}_m\},$$

then

$$[N]_{\mathcal{C}} = \text{diag}\left\{\mathcal{N}_0^{h_1}, \ldots, \mathcal{N}_0^{h_m}\right\}.$$

Given $T: V \to V$ with distinct eigenvalues $\{\sigma_1, \ldots, \sigma_s\}$ and associated stable eigenspaces $W_k = \widetilde{\ker}(T - \sigma_k \mathbb{I})$, we apply this Proposition to the nilpotent transformations $(T\,|_{W_k} - \sigma_k \mathbb{I}) = N_k: W_k \to W_k$ and find

$$\begin{aligned} V &= W_1 \oplus \cdots \oplus W_s \\ &= \left(W_1^{(1)} \oplus \cdots \oplus W_1^{(m_1)}\right) \oplus \cdots \oplus \left(W_s^{(1)} \oplus \cdots \oplus W_s^{(m_s)}\right) \end{aligned}$$

where the $W_k^{(i)}$ are N_k-cyclic. Let $\mathcal{C}_k^{(i)} = \{\vec{w}_k^{(i)}, T\vec{w}_k^{(i)}, \ldots, T^{h_k^{(i)}-1}\vec{w}_k^{(i)}\}$ be cyclic bases for the $W_k^{(i)}$, and combine them to produce $\mathcal{C}_k = \{\mathcal{C}_k^{(1)}, \ldots, \mathcal{C}_k^{(m_k)}\}$, and then $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$. Clearly $\mathcal{C}$ is a basis for

[23]This decomposition is *not* unique but the numbers m and $\{h_1, \ldots, h_m\}$ are. (See the end of the section for discussion of uniqueness.)

V and it is subordinate to the entire direct-sum decomposition just written. With respect to this basis,

$$[T]_{\mathcal{C}} = \text{diag}\left\{ \left[T\,|_{W_1}\right]_{\mathcal{C}_1}, \ldots, \left[T\,|_{W_s}\right]_{\mathcal{C}_s} \right\}$$

decomposes into big Jordan blocks, and each big block

$$\left[T\,|_{W_k}\right]_{\mathcal{C}_k} = \text{diag}\left\{ \left[T\,|_{W_k^{(1)}}\right]_{\mathcal{C}_k^{(1)}}, \ldots, \left[T\,|_{W_k^{(m_k)}}\right]_{\mathcal{C}_k^{(m_k)}} \right\}$$

decomposes into little blocks.

Now for $N\colon W \to W$ it is clear from our pictures that the "top" of each "N-cycle" contributes one dimension to the kernel of N.

VI.E.6. PROPOSITION. *$m = \{\#$ of $W^{(i)}$ in the decomposition of $W\}$* $= \dim(\ker N)$.

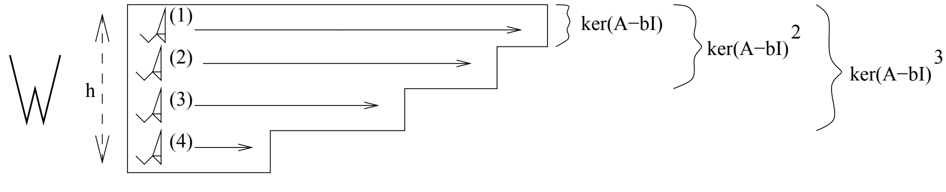
This has the following important consequence for $T\colon V \to V$.

VI.E.7. COROLLARY. *We have equality between*

- $m_k :=$ # of little Jordan blocks in the big block for σ_k *and*
- $d_k := \dim\{\ker(T - \sigma_k \mathbb{I})\} =$ geometric multiplicity of σ_k.

How to compute Jordan form (and an associated basis). We will work in terms of the matrix $A = [T]_{\hat{e}}$; our goal is to rewrite it as $P_{\mathcal{C}} \mathcal{J} P_{\mathcal{C}}^{-1}$, with $\mathcal{J} = [T]_{\mathcal{C}}$ consisting of diagonal Jordan blocks. We call $\mathcal{J}$ the *Jordan normal form of A*. Begin by computing and factoring $f_A(\lambda) = \det(\lambda\mathbb{I} - A)$, in order to find the eigenvalues $\{\sigma_1, \ldots, \sigma_s\}$. Pick some $\sigma = \sigma_k$ to concentrate on, set $W = W_k = \widetilde{\ker}(A - \sigma_k\mathbb{I})$, and use the following two steps to construct $\mathcal{C} = \mathcal{C}_k$. (Then repeat for the remaining eigenvalues.) Note that in the following $\mathcal{A}^{(i)}$ and $\mathcal{C}^{(i)}$ always denote finite collections of vectors (like a basis).

Step I. Again consider the picture of W:

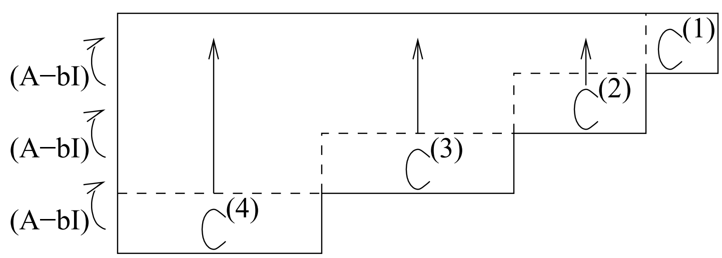


Find bases for the successive "levels" of W:

$$\mathcal{A}^{(1)} \text{ for } \ker(A-\sigma\mathbb{I}),\ \{\mathcal{A}^{(1)},\mathcal{A}^{(2)}\} \text{ for } \ker\{(A-\sigma\mathbb{I})^2\},$$
$$\{\mathcal{A}^{(1)},\mathcal{A}^{(2)},\mathcal{A}^{(3)}\} \text{ for } \ker\{(A-\sigma\mathbb{I})^3\}, \text{ and so on.}$$

This may require readjustment of bases along the way: the basis for $\ker(A-\sigma\mathbb{I})^2$ produced by the usual method may *not* include the elements of $\mathcal{A}^{(1)}$, etc. WARNING: the picture is slightly deceptive because — unlike *span*$(\mathcal{A}^{(1)})$, which is $\ker(A-\sigma I)$ — the spaces *span*$(\mathcal{A}^{(2)})$, *span*$(\mathcal{A}^{(3)})$, etc. are *not* uniquely determined (a different choice of $\mathcal{A}^{(2)}$, $\mathcal{A}^{(3)}$ etc. could change them). However, the number of vectors in $\mathcal{A}^{(i)}$ *is* unique: call this number a_i, and note that $a_1 \geq a_2 \geq a_3 \geq \cdots$.

Step II. Now we find the generators $\mathcal{C}^{(i)}$ for the cyclic subspaces of height i ($1 \leq i \leq h$). Here is a representative picture with $h = 4$:



First set $\mathcal{C}^{(h)} = \mathcal{A}^{(h)}$; this is a set of a_h vectors. Then take $\mathcal{C}^{(h-1)} \subseteq \mathcal{A}^{(h-1)}$ to be some subset such that

$$\{(A-\sigma\mathbb{I})\mathcal{C}^{(h)},\mathcal{C}^{(h-1)}\} \text{ is an independent set}$$

consisting of a_{h-1} vectors.

VI.E.8. REMARK. (1) $\mathcal{C}^{(h-1)}$ contains $(a_{h-1}-a_h)$ vectors. If $a_{h-1} = a_h$ then $\mathcal{C}^{(h-1)}$ will be empty.

(2) span$\{(A-\sigma\mathbb{I})\mathcal{C}^{(h)},\mathcal{C}^{(h-1)}\}$ will *not* in general be the same subspace as span$(\mathcal{A}^{(h-1)})$, although the dimensions are equal.

(3) You may wish to put subscripts on the $\mathcal{A}^{(h)}$'s, viz. $\mathcal{A}_k^{(h)}$ (and likewise $\mathcal{C}_k^{(h)}$ below), to avoid confusing the bases for different W_k's.

Continuing Step II, let $\mathcal{C}^{(h-2)} \subseteq \mathcal{A}^{(h-2)}$ be any subset (containing $a_{h-2} - a_{h-1}$ vectors) so that

$$\left\{ (A-\sigma\mathbb{I})^2\mathcal{C}^{(h)},\, (A-\sigma\mathbb{I})\mathcal{C}^{(h-1)},\, \mathcal{C}^{(h-2)} \right\} \text{ is an independent set}$$

consisting of a_{h-2} vectors, and so on. It may help to actually write the vectors inside diagrams similar to the ones we have drawn.

The desired basis (for W_k) is then $\mathcal{C}_k =$

$$\left\{ \mathcal{C}^{(1)}\,;\, \mathcal{C}^{(2)}, (A-\sigma\mathbb{I})\mathcal{C}^{(2)}\,;\, \mathcal{C}^{(3)}, (A-\sigma\mathbb{I})\mathcal{C}^{(3)}, (A-\sigma\mathbb{I})^2\mathcal{C}^{(3)}\,;\, \text{etc.} \right\}$$

(where for instance $(A-\sigma\mathbb{I})\mathcal{C}^{(2)}$ means: apply $(A-\sigma\mathbb{I})$ to each of the vectors of $\mathcal{C}^{(2)}$). Each element of $\mathcal{C}^{(i)}$ will correspond to a small Jordan block of the form $\mathcal{N}^i_\sigma$, and this is in fact all the information you need to write down the big Jordan block for σ_k.

In terms of our picture: the whole diagram corresponds to a big Jordan block for *one eigenvalue* σ; each column corresponds to a little Jordan (i.e., $(A-\sigma\mathbb{I})$-cyclic) block; and the dimension of the little block is the height of the corresponding column.

This concludes our "algorithm"; let's now apply it.

VI.E.9. EXAMPLE. Consider the matrix

$$A = \begin{pmatrix} 2 & -1 & 1 & -1 \\ -1 & 2 & -2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 \end{pmatrix} \quad \longrightarrow \quad \begin{aligned} f_A(\lambda) &= \det(\lambda\mathbb{I} - A) \\ &= (\lambda-1)^3(\lambda-2). \end{aligned}$$

We will work on $\sigma = 1$ first.

Since

$$\ker(A-\mathbb{I}) = \mathit{span}\begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

is 1-dimensional, Corollary VI.E.7 $\implies$ the big Jordan block for $\sigma = 1$ contains 1 little Jordan block! That is, $\widetilde{\ker}(A-\mathbb{I})$ is *cyclic* under $A - \mathbb{I} = N$. Therefore we know immediately the Jordan normal form

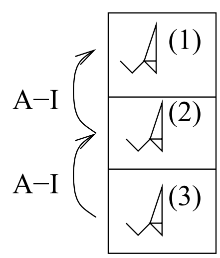
$$\mathcal{J} = \begin{pmatrix} \boxed{\begin{matrix} 1 & & \\ 1 & 1 & \\ & 1 & 1 \end{matrix}} & \\ & \boxed{2} \end{pmatrix}$$

of A but not the basis change S (yet). Continue to compute:

$$\ker(A - \mathbb{I})^2 = span \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\},$$

$$\ker(A - \mathbb{I})^3 = span \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

The picture of $W = \widetilde{\ker}(A - \mathbb{I})$ is



where

$$\mathcal{A}^{(1)} = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \ \mathcal{A}^{(2)} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \ \mathcal{A}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \hat{e}_4.$$

The beginning of “Step II” is now $\mathcal{C}^3 = \mathcal{A}^{(3)}$. In fact, there is no $\mathcal{C}^{(2)}$ or $\mathcal{C}^{(1)}$ (our picture of W has no “steps”, just one column), and so this is also the end. Throwing out $\mathcal{A}^{(1)}, \mathcal{A}^{(2)}$ and replacing them with the cyclic images of $\mathcal{C}^3 = \hat{e}_4$ under $(A - \mathbb{I})$, the desired basis is

$$\mathcal{C}_1 = \left\{ \mathcal{C}^{(3)},\, (A-\mathbb{I})\mathcal{C}^{(3)},\, (A-\mathbb{I})^2\mathcal{C}^{(3)} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

(and *not* $\mathcal{A}^{(1)}, \mathcal{A}^{(2)}, \mathcal{A}^{(3)}$).

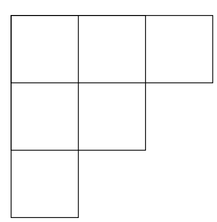
For the eigenvalue $\sigma = 2$ we find

$$\ker(A - 2\mathbb{I}) = \mathit{span} \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}; \quad \text{so } \mathcal{C}_2 = \left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Setting $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2\}$, we may now write $A = P_{\mathcal{C}}\, \mathcal{J}\, P_{\mathcal{C}}^{-1} =$

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

For a more "general" example, giving rise to a diagram of the form



you might look at

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

Application to differential equations. So suppose we had a system of equations

$$\left.\begin{aligned} \tfrac{dy_1}{dt} &= 2y_1 - y_2 + y_3 - y_4 \\ \tfrac{dy_2}{dt} &= -y_1 + 2y_2 - 2y_3 + y_4 \\ \tfrac{dy_3}{dt} &= y_2 + y_3 + y_4 \\ \tfrac{dy_4}{dt} &= -y_2 + y_3 \end{aligned}\right\} \quad \longrightarrow \quad \frac{d\vec{y}}{dt} = A\vec{y}$$

(where A is the same as in the last example). Setting $\vec{c} = P_{\mathcal{C}}^{-1}\vec{y}$ — that is, changing coordinates as usual — we obtain the new system

$$\frac{dc_1}{dt} = c_1\,, \;\; \frac{dc_2}{dt} = c_1 + c_2\,, \;\; \frac{dc_3}{dt} = c_2 + c_3\,, \;\; \frac{dc_4}{dt} = 2c_4.$$

Clearly $c_4(t) = Ce^{2t}$, while $(c_1(t), c_2(t), c_3(t))$ is a linear combination of

$$\left(e^t,\, te^t,\, \frac{t^2}{2}e^t\right)\,, \;\; \left(0,\, e^t,\, te^t\right)\,, \text{ and } \left(0,\, 0,\, e^t\right).$$

Applying $P_{\mathcal{C}}$ to $\vec{c}(t)$, as usual, recovers $\vec{y}(t)$. More generally, for a block equation in Jordan form

$$\frac{d}{dt}\begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ c_r \end{pmatrix} = \begin{pmatrix} \lambda & & & \\ 1 & \lambda & & \\ & 1 & \lambda & \\ & & 1 & \lambda \end{pmatrix}\begin{pmatrix} c_1 \\ \vdots \\ \vdots \\ c_r \end{pmatrix},$$

one finds that any solution $(c_1(t), \ldots, c_r(t))$ is a linear combination of

$$\left(e^{\lambda t},\, te^{\lambda t},\, \ldots\ldots,\, \frac{t^{r-1}}{(r-1)!}e^{\lambda t}\right)$$

and its "shifts" to the right (as above).

The same functions arise in a slightly different fashion (as basis vectors rather than time-dependent coefficients). We are going to find all solutions of the ODE

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1}f}{dx^{n-1}} + \ldots + a_1\frac{df}{dx} + a_0 f = 0.$$

Working over $\mathbb{C}$, we write

$$q(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = \prod_{i=1}^{s}(\lambda - \sigma_i)^{k_i}$$

(where the σ_i are distinct), define

$$V \ := \ \text{space of solutions to} \ \left\{ q\left(\frac{d}{dx}\right) \ f \ = 0 \right\}$$

(assume it's finite-dimensional); and let

$$D\colon V \to V \ \text{be the restriction of} \ \frac{d}{dx} \ \text{to} \ V.$$

Clearly $q(D) = 0$. Let λ_0 be an eigenvalue of D; then there exists a nonzero $f \in V$ such that $Df = \lambda_0 f$. But then

$$0 = q(D)f = q(\lambda_0)f \implies q(\lambda_0) = 0,$$

and so λ_0 is a root ($\implies \lambda_0 = \sigma_i$). Therefore V decomposes into stable eigenspaces

$$V = \widetilde{\ker}(D - \sigma_1\mathbb{I}) \oplus \cdots \oplus \widetilde{\ker}(D - \sigma_s\mathbb{I}).$$

Moreover, $\dim\ker(D - \sigma_i\mathbb{I}) = 1$ (because the only solution to $Df = \sigma_i f$ is $Ce^{\sigma_i x}$), and so by Corollary VI.E.7 each of the stable kernels is *cyclic* nilpotent, spanned by the iterates of some generator

$$g_i\,,\ (D - \sigma_i\mathbb{I})g_i\,,\ (D - \sigma_i\mathbb{I})^2 g_i\,, \ldots\ldots.$$

Suppose this doesn't terminate at k_i iterations: more precisely say (for $m \geq 1$) that

$$(D - \sigma_i\mathbb{I})^{k_i+m-1}g_i \ \neq \ 0 \qquad \text{but} \qquad (D - \sigma_i\mathbb{I})^{k_i+m}g_i \ = \ 0$$

(this must happen for some m by the definition of stable kernel). Now on the one hand (since $k_i + m - 1 \geq k_i$)

(VI.E.10)

$$q(D)g_i = 0 \implies \left(\prod_{j\neq i}(D - \sigma_j\mathbb{I})^{k_j}\right)(D - \sigma_i\mathbb{I})^{k_i+m-1}g_i = 0.$$

On the other, $(D - \sigma_i\mathbb{I})^{k_i+m-1}g_i$ is an eigenvector for D with eigenvalue σ_i (because it is killed by $D - \sigma_i\mathbb{I}$) and therefore must be $Ce^{\sigma_i x}$

(C a nonzero constant). Plugging this in to (VI.E.10), we have

$$0 = \prod_{j \neq i} (D - \sigma_j \mathbb{I}) e^{\sigma_i x} = \left(\prod_{j \neq i} (\sigma_i - \sigma_j) \right) e^{\sigma_i x},$$

which is impossible because the $\sigma_1, \ldots, \sigma_s$ are *distinct*!

Therefore $\widetilde{\ker}(D - \sigma_i \mathbb{I})$ is cyclic nilpotent (under $D - \sigma_i \mathbb{I}$) with height k_i, as you might expect. The generator g_i is just $x^{k_i - 1} e^{\sigma_i x}$. Thus the solutions to the above ODE are the functions of the form

$$f(x) = \sum_{i=1}^{s} \sum_{j=0}^{k_i - 1} C_{ij} x^j e^{\sigma_i x}$$

where C_{ij} are arbitrary constants.

* * *

Uniqueness. We conclude this section with the general statement about matrices and Jordan form. By a *Jordan matrix*, we shall mean simply a block diagonal matrix with blocks of the form $\mathcal{N}_\sigma^h$ (i.e., little Jordan blocks). For instance, all diagonal matrices are Jordan; so is a single $\mathcal{N}_\sigma^h$.

VI.E.11. THEOREM. *Let $A \in M_n(\mathbb{C})$ be an arbitrary matrix. Up to reordering blocks, there is exactly one Jordan matrix $\mathcal{J}(A)$ similar to A.*

PROOF. Clearly we have established existence: there is an invertible $P \in M_n(\mathbb{C})$ (*far* from unique!) such that

$$P^{-1} A P = \mathcal{J} = \mathrm{diag}\{\mathcal{N}_{\sigma_1}^{h_1^{(1)}}, \ldots, \mathcal{N}_{\sigma_1}^{h_1^{(m_1)}}; \ldots; \mathcal{N}_{\sigma_s}^{h_s^{(1)}}, \ldots, \mathcal{N}_{\sigma_s}^{h_s^{(m_s)}}\},$$

with $\sigma_1, \ldots, \sigma_s$ the eigenvalues of $\mathcal{J}$, hence A. Furthermore, for each k, the dimension of the big Jordan block is

$$\sum_{j=1}^{m_k} h_k^{(j)} = \dim(\widetilde{E}_{\sigma_k}(\mathcal{J})) = \dim(\widetilde{E}_{\sigma_k}(A)) = \tilde{d}_k$$

since A and $\mathcal{J}$ are similar. In addition, we have seen that $m_k = \dim(E_{\sigma_k}(A)) = d_k$.

To show that the entire list $h_k^{(1)}, \ldots, h_k^{(d_k)}$ (for each k) is determined by A as well, write $c_{k\ell}$ for the number of times ℓ appears in this list. We have

$$\begin{aligned} a_{k\ell} &:= \text{nullity}\{(A - \sigma_k \mathbb{I})^{\ell}\} - \text{nullity}\{(A - \sigma_k \mathbb{I})^{\ell-1}\} \\ &= \text{nullity}\{(\mathcal{J} - \sigma_k \mathbb{I})^{\ell}\} - \text{nullity}\{(\mathcal{J} - \sigma_k \mathbb{I})^{\ell-1}\} \\ &= \sum_{h \geq \ell} c_{kh}. \end{aligned}$$

(You may check this by hand, or sort it out from the discussion of bases $\mathcal{A}_i^{(\ell)}$ and $\mathcal{C}_k^{(\ell)}$ above; $a_{k\ell}$ and $c_{k\ell}$ are just the numbers of vectors in these sets.) Hence

$$c_{k\ell} = a_{k\ell} - a_{k,\ell+1}$$

is completely determined and the uniqueness is proved. $\square$

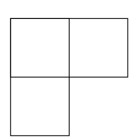
VI.E.12. REMARK. Of course, a matrix A with real entries belongs to $M_n(\mathbb{C})$, and the Theorem applies. However, it should be understood that, as with diagonalization, one has to allow S and $\mathcal{J}$ to have complex entries if f_A does not split into linear factors over $\mathbb{R}$.

Exercises

(1) Both of the following matrices have only *one* eigenvalue:

$$\begin{pmatrix} 3 & -5 & -35 \\ 0 & -4 & -49 \\ 0 & 1 & 10 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ -1 & 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

Put them in Jordan normal form (= generalized "diagonal" form, if you will); that is, write the entire decomposition $A = S\mathcal{J}S^{-1}$. It may help you to draw out the diagrams, and write the vectors you find where they belong. For the first one it should look like



(2) Find all solutions of the system

$$\frac{dx}{dt} = 2x + z\,, \;\; \frac{dy}{dt} = y + 2z\,, \;\; \frac{dz}{dt} = 2z.$$

[Hint: use Jordan.]

(3) (a) Let N be cyclic nilpotent on W, $\dim W = d$.

(i) What must the minimal polynomial of N be?

(ii) So what is the minimal polynomial of a little Jordan block $\mathcal{N}_\sigma^d$?

(iii) How about a big Jordan block $\mathrm{diag}\{\mathcal{N}_\sigma^{d_1(\sigma)}, \ldots, \mathcal{N}_\sigma^{d_r(\sigma)}\}$? (Assume $d_1(\sigma) \geq \cdots \geq d_r(\sigma)$.)

(iv) Finally, how about an arbitrary Jordan-form matrix $\mathcal{J}$?

(b) A matrix A having this Jordan form ($A = S\mathcal{J}S^{-1}$) will have the same minimal polynomial (why?). Also the Jordan form is *unique*, up to reordering of the blocks along the diagonal. Use these facts to explain why A is diagonalizable iff its minimal polynomial has no repeated root.

(c) Find the minimal polynomials of the matrices in problem (1).

(4) Use Jordan form to show that if $f_A(\lambda) = \prod_i (\lambda - \sigma_i)^{n_i}$, then we have $\det(A) = \prod_i \sigma_i^{n_i}$. (This can be seen without Jordan — cf. the beginning of §V.C — but this proof is simpler.)

(5) Consider the transformation $T\colon P^3 \to P^3$ on polynomials defined by $T(p(t)) = p(t+1)$. What must the Jordan form for T be? (You can do this without writing out the matrix $[T]_{\{1,t,t^2,t^3\}}$.)

(6) In this problem you will show that the Jordan normal form $\mathcal{J}(A)$ of $A \in M_n(\mathbb{C})$ may be regarded as a "refinement" of its rational canonical form,[24] and see how to calculate it directly from the nontrivial invariant factors $\Delta_r, \ldots, \Delta_n$ of $\lambda\mathbb{I} - A$.

(a) Given $B \in M_m(\mathbb{C})$, show that the power k to which $\lambda - \sigma$ appears in $m_B(\lambda)$ is the size of the biggest "little Jordan block"

[24]In the theory of "modules over a principal ideal domain", the main result is something called the *structure theorem*, which decomposes the module into its constituent factors. The rational canonical form is the direct manifestation of this theorem when applied to $T\colon V \to V$ (which makes V into a "module over $\mathsf{F}[\lambda]$"). Further decomposing these constituent factors into their "primary components" is what leads to the Jordan form.

with eigenvalue σ (i.e. $\mathcal{N}_\sigma^k$) in $\mathcal{J}(B)$. [Hint: what is the "height" of multiplication by $B - \sigma\mathbb{I}_m$ on $\widetilde{E}_\sigma(B)$?]

(b) If $B = M(p)$ is the companion matrix of a monic polynomial $p(\lambda) = \prod_{i=1}^t (\lambda - \sigma_i)^{\ell_i}$ of degree m, use (a) to calculate $\mathcal{J}(B)$. (In particular, you should find that each *eigenspace* $E_{\sigma_j}(B)$ is only 1-dimensional.)

(c) The rational form $\mathrm{diag}\{M(\Delta_r), \ldots, M(\Delta_n)\}$ of A is similar to $\mathrm{diag}\{\mathcal{J}(M(\Delta_r)), \ldots, \mathcal{J}(M(\Delta_n))\}$. Conclude that the latter matrix *is* $\mathcal{J}(A)$, and that the powers of linear factors in the Δ_k's are the sizes of the small Jordan blocks of $\mathcal{J}(A)$.

(d) If you did Exercises VI.B.3 and/or VI.C.7, use this result to quickly calculate $\mathcal{J}(A)$ in each case. (Note that Ex. VI.C.7 acually has 6 different cases.)

(7) The (multiplicative) *Jordan Decomposition Theorem* states that any invertible matrix A can be written uniquely as a product SU of a commuting pair of semisimple and unipotent[25] matrices.

(a) Show that a Jordan matrix can be written as SU. Then do it for an arbitrary matrix, proving the existence part of JDT. (This is the only part of this problem that uses the Jordan normal form *per se*.)

(b) Given $\vec{v} \in \widetilde{E}_\sigma(A)$, and S a semisimple matrix commuting with A with $A - S$ nilpotent, prove that $\vec{v} \in E_\sigma(S)$. [Hint: show that $(S - \sigma\mathbb{I})^{2n}\vec{v} = \vec{0}$ by writing $S - \sigma\mathbb{I} = (S - A) + (A - \sigma\mathbb{I})$.]

(c) Prove the uniqueness part of the JDT. [Hint: use (b) to show that $A = SU = US$ completely determines S. It may help to use the Jordan Structure Theorem from the last section and think of S as a transformation.]

[25] We say U is unipotent if $U - \mathbb{I}$ is nilpotent. Also recall that semisimple means diagonalizable. So the theorem means that we can write $A = SU = US$ and if also $A = S'U' = U'S'$ (with S' semisimple and U' unipotent) then $S' = S$ and $U' = U$.