M2P2 Notes on the Jordan Canonical Form

First, here is the proof of Theorem 16.12 in the lectures (in lectures I just gave an example illustrating the idea of the proof).

Theorem 16.12 *Let W be a vector space and $S : W \to W$ a linear transformation such that $S^a = 0$ for some positive integer a. Then there is a basis B of W such that $[S]_B$ is a JCF matrix of the form $J_{n_1}(0) \oplus \cdots \oplus J_{n_k}(0)$.*

Proof As in lecs, for $w \in W$, let r be the least positive integer such that $S^r(w) = 0$, and define $w^{(S)}$ to be the set of vectors

$$w^{(S)} = \{w, S(w), S^2(w), \dots, S^{r-1}(w)\}.$$

If we can find a basis for W of the form

$$B = w_1^{(S)} \cup w_2^{(S)} \cup \cdots \cup w_k^{(S)}, \qquad (1)$$

then $[S]_B$ will be the JCF matrix in the conclusion of the theorem.

To show that there is a basis of the form (1), start with a spanning set of this form: take a spanning set $v_1, \dots, v_n$ of W and define

$$C = v_1^{(S)} \cup v_2^{(S)} \cup \cdots \cup v_n^{(S)}.$$

Then certainly C spans W. For each i let r_i be such that

$$v_i^{(S)} = \{v_i, S(v_i), \dots, S^{r_i-1}(v_i)\}$$

(so $S^{r_i}(v_i) = 0$).

If C is linearly independent then it is a basis of the required form and we are done. So suppose C is linearly dependent. We show how to replace one of the sets $v_i^{(S)}$ by a *smaller* set $v_i'^{(S)}$ in such a way that we still have a spanning set.

Here's how. Since C is linearly dependent, there is a linear relation of the form

$$\sum_{j=1}^{r_1-1} \alpha_j S^j(v_1) + \cdots + \sum_{j=1}^{r_n-1} \lambda_j S^j(v_n) = 0 \qquad (2)$$

where not all the coefficients are zero. Now apply to both sides the largest power of S which does not kill (i.e. send to 0) the LHS. This gives an equation

$$\beta_1 S^{r_1-1}(v_1) + \cdots + \beta_n S^{r_n-1}(v_n) = 0,$$

for some scalars β_i, where not all the β_i are zero. By ordering the original v_i, we may take it that $r_1 \leq r_2 \leq \cdots \leq r_n$. So if β_i is the first nonzero coefficient in the above equation, then

$$S^{r_i-1}(\beta_i v_i + \cdots + \beta_n S^{r_n - r_i}(v_n)) = 0.$$

Define $v_i' = \beta_i v_i + \cdots + \beta_n S^{r_n - r_i}(v_n)$, so that $v_i' \neq 0$ and $S^{r_i-1}(v_i') = 0$. Then the set $v_i'^{(S)}$ has size at most $r_i - 1$, so is smaller than $v_i^{(S)}$. In the spanning set C, replace $v_i^{(S)}$ by $v_i'^{(S)}$ to get

$$C' = v_1^{(S)} \cup \cdots v_i'^{(S)} \cup \cdots \cup v_n^{(S)}.$$

Then C' is smaller than C, but still spans W since the span of C' contains v_i (a linear combination of v_i' and $S^j(v_l)$'s for $l \neq i$) and similarly contains all $S^j(v_i)$'s.

Hence we have managed to replace C by a smaller spanning set of the same form (i.e. of the form (1)). We continue like this, replacing our spanning sets by smaller and smaller spanning sets of this form, until we end up with a basis of the form (1). This completes the proof.

Example Call a basis of the form (1) a *Jordan basis* of W. Here's an example of how to find a Jordan basis of $\mathbb{R}^4$ for the matrix

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

Let $e_1, \ldots, e_4$ be the standard basis. Observe that

$$Ae_1 = 0, \; Ae_2 = e_1, \; Ae_3 = e_1, \; Ae_4 = e_2 + e_3.$$

We have $e_4^{(A)} = \{e_4, e_2 + e_3, 2e_1\}$ and $e_3^{(A)} = \{e_3, e_1\}$ and the union of these two sets spans $\mathbb{R}^4$. So let's start with the spanning set

$$C = e_3^{(A)} \cup e_4^{(A)}.$$

This is not a basis, so let's use the method of the above proof to replace one of the sets. Tho obvious relation between the vectors $e_1 \in e_3^{(A)}$ and $2e_1 \in e_4^{(A)}$ gives

$$2Ae_3 - A^2 e_4 = 0.$$

As in the above proof the next step is to hit this with the largest power of A which does not kill it. But A kills the LHS, so we do not have to do this step.

The above eqn says $A(2e_3 - Ae_4) = 0$. So in the set C we replace $e_3^{(A)}$ by $e_3^{\prime(A)} = e_3'$, where $e_3' = 2e_3 - Ae_4 = e_3 - e_2$. So we now have a new spanning set

$$C' = e_3 - e_2,\ e_4^{(A)} = e_3 - e_2, e_4, e_2 + e_3, 2e_1.$$

This is a Jordan basis, and we see that the JCF of A is $J_1(0) \oplus J_3(0)$.

Now here are some notes on the proof of Theorem 16.11 of lecs, which I omitted in the lectures.

Theorem 16.11 *Let $T : V \to V$ be a linear transformation with characteristic polynomial*

$$p(x) = \prod_{i=1}^{k} (x - \lambda_i)^{a_i},$$

where λ_i are the distinct eigenvalues of T. For each i define $V_i = \ker(T - \lambda_i I)^{a_i}$. Then

$$V = V_1 \oplus \cdots \oplus V_k.$$

For the proof we need some basic facts about polynomials. Let $F = \mathbb{R}$ or $\mathbb{C}$, and let $F[x]$ be the set of all polynomials in x over F, i.e. the set of all polys $f(x) = a_n x^n + \cdots + a_1 x + a_0$ with $a_i \in F$. Define the degree $\deg(f)$ to be the highest power of x which appears in $f(x)$ with a nonzero coefficient. I will often write just f rather than $f(x)$ to save space.

Just as for $\mathbb{Z}$, there is a Euclidean algorithm for polynomials:

Euclidean Algorithm *Let f, g be polynomials in $F[x]$ with $\deg(g) \geq 1$. Then there are polynomials $q, r \in F[x]$ such that*

$$f = qg + r \quad \textit{and } \deg(r) < \deg(g).$$

We call q the quotient and r the remainder. For example, taking $f = x^3 + x$, $g = x^2 - x + 1$ we have

$$x^3 + x = (x + 1)(x^2 - x + 1) + x - 1$$

so here $q = x + 1$ and $r = x - 1$.

Not too surprisingly, if f, g are polynomials we say f divides g if $g = qf$ for some polynomial q. And we say a polynomial d is a highest common factor of f and g if d is of maximum possible degree among all polynomials dividing both f and g. Just as you saw for $\mathbb{Z}$ in the days of M1F, we can use the Euclidean algorithm to calculate hcf's of polys. For example taking f and g as above, we do the Euc alg:

$$\begin{aligned} f &= (x+1)g + x - 1 \\ g &= x(x-1) + 1 \end{aligned}$$

Hence we see that $hcf(f, g) = 1$.

Just as for $\mathbb{Z}$, we see that

(*) *If* $d = hcf(f, g)$*, then there are polynomials* $s, t \in F[x]$ *such that* $d = sf + tg$*.*

For example with $f = x^3 + x$, $g = x^2 - x + 1$ as above, we have

$$\begin{aligned} 1 &= g - x(x-1) \\ &= g - x(f - (x+1)g) \\ &= -xf + (x^2 + x + 1)g. \end{aligned}$$

Just as for $\mathbb{Z}$ in M1F, all this leads quickly to a unique prime factorisation theorem for polys. For $\mathbb{C}[x]$ this just says that every poly is a unique product of linear factors.

Now we are ready to start the proof of Theorem 16.11. Here is the key prop:

Proposition *Let* $T : V \to V$ *be a linear transformation, and suppose* $f(x)$ *and* $g(x)$ *are polynomials such that* $hcf(f, g) = 1$ *and* $f(T)g(T) = 0$*. Then*

$$V = \ker(f(T)) \oplus \ker(g(T)).$$

Proof By the fact (*) above, there are polys $s, t \in F[x]$ such that $1 = sf + tg$. Applying this to T we get $s(T)f(T) + t(T)g(T) = I$, the identity map. Hence for any $v \in V$,

$$v = I(v) = s(T)f(T)(v) + t(T)g(T)(v).$$

The first vector in this sum is in $\ker(g(T))$, since $g(T)s(T)f(T)(v) = s(T)f(T)g(T)(v) = 0$ (by the assumption that $f(T)g(T) = 0$). Similarly the second vector in the sum is in $\ker(f(T))$. Hence we have shown that any v is the sum of vectors in these kernels, in other words

$$V = \ker(f(T)) + \ker(g(T)).$$

To show that this is a direct sum, we need to show that the intersection of the two kernels is 0. But if $v \in \ker(f(T)) \cap \ker(g(T))$, then $v = s(T)f(T)(v) + t(T)g(T)(v) = 0$, so the intersection is indeed 0. This completes the proof of the prop.

Since $hcf((x-\lambda_1)^{a_1}, (x-\lambda_2)^{a_2}) = 1$ if $\lambda_1 \neq \lambda_2$ (this follows by unique factorisation in $F[x]$), we deduce immediately from the proposition:

Corollary *If $T : V \to V$ has char poly $(x-\lambda_1)^{a_1}(x-\lambda_2)^{a_2}$ then $V = \ker(T-\lambda_1 I)^{a_1}) \oplus \ker(T-\lambda_2 I)^{a_2})$.*

This is the case $k = 2$ of Theorem 16.11, and the general case follows by a straightforward induction on k. Here is the argument. Assume the hypotheses of Thm 16.11. Define $V_i = \ker(T-\lambda_i I)^{a_i}$.

Let $f(x) = (x-\lambda_1)^{a_1} \cdots (x-\lambda_{k-1})^{a_{k-1}}$ and $g(x) = (x-\lambda_k)^{a_k}$. Then $hcf(f, g) = 1$ (this follows by unique factorisation in $F[x]$). Of course $f(T)g(T) = 0$ by Cayley-Hamilton, so by the prop

$$V = \ker(f(T)) \oplus \ker(g(T)) = \ker(f(T)) \oplus V_k.$$

By induction we have

$$\ker(f(T)) = V_1 \oplus \cdots \oplus V_{k-1}.$$

Putting these together, we get $V = V_1 \oplus \cdots \oplus V_{k-1} \oplus V_k$. This completes the proof of Theorem 16.11.

Have a great vacation! Don't work too hard! Well, at least give yourself a break on Boxing Day....