10. Jordan canonical form

As an application of the structure theorem for PID's I explained the *Jordan canonical form* for matrices over the complex numbers. First I stated the theorem and then I proved it by reducing it to a module over a PID.

10.1. statement of the theorem.

The theorem is that any $n \times n$ matrix A with coefficients in the complex numbers is conjugate to a matrix in Jordan canonical form. This is defined to be a square matrix with the eigenvalues of A along the diagonal, 0's and 1's on the super-diagonal (right above the diagonal) and zeroes everywhere else. Also, if there is a 1 in the $i, i+1$ position then $a_{ii} = a_{i+1,i+1}$. Here is an example:

$$B^{-1}AB = \begin{pmatrix} a & 1 & 0 & 0 & 0 & 0 \\ 0 & a & 1 & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 & b & 1 \\ 0 & 0 & 0 & 0 & 0 & b \end{pmatrix}.$$

This is a matrix with three *Jordan blocks*

$$B^{-1}AB = \begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix} \oplus (a) \oplus \begin{pmatrix} b & 1 \\ 0 & b \end{pmatrix}.$$

On the second day, I used the notation $[\lambda]^k$ to denote the Jordan block of size k and eigenvalue λ. So, this decomposition would be written:

$$B^{-1}AB = [a]^3 \oplus [a]^1 \oplus [b]^2.$$

10.2. module over a PID.

The key point is something I explained earlier. Namely, any square matrix over $\mathbb{C}$ makes $\mathbb{C}^n$ into a module over $\mathbb{C}[T]$.

If A is an $n \times n$ matrix over $\mathbb{C}$ then A is an element of the matrix ring $M_n(\mathbb{C})$. We let $R = \mathbb{C}[T]$. This is a PID since $\mathbb{C}[T]$ has a Euclidean algorithm: when you divide one polynomial by another, the remainder has smaller degree. The module is $M = \mathbb{C}^n$ with the action of the ring given by the evaluation map

$$ev_A : \mathbb{C}[T] \to M_n(\mathbb{C})$$

which sends $f(T)$ to $f(A)$. In other words,

$$f(T)x = f(A)x.$$

For example, if $f(T) = aT^2 + bT + c$ then

$$f(T)x = f(A)x = aA^2x + bAx + cx.$$

Lemma 10.1. *$M = \mathbb{C}^n$ is a finitely generated torsion module over $R = \mathbb{C}[T]$.*

Proof. This uses the "restriction of scalars" idea. Since $\mathbb{C}[T]$ contains the field $\mathbb{C}$, we can restrict scalars and view M as a vector space over $\mathbb{C}$. Then M is finite dimensional. But $\mathbb{C}[T]$ is infinite dimensional over $\mathbb{C}$. So, $\mathbb{C}[T]$ cannot be a direct summand of M. So, it is torsion. Also, being finitely generated over the smaller ring means it is finitely generated over the bigger ring. □

By the structure theorem we now have:

$$M \cong \bigoplus \mathbb{C}[T]/(p_i(T)^{n_i})$$

where $p_i(T)$ is an irreducible polynomial. But $\mathbb{C}$ is *algebraically closed*. So, any polynomial has a root $a_i \in \mathbb{C}$ which means $T - a_i$ divides $p_i(T)$ which implies that

$$p_i(T) = T - a_i$$

up to multiplication by a scalar.

Lemma 10.2. *Let $p(T) = T - a$. Then the p-primary part of M is nonzero ($M_p \neq 0$) iff a is an eigenvalue of A.*

Proof. Suppose first that $x \in M = \mathbb{C}^n$ is an eigenvector of A with eigenvalue a. Then $x \neq 0$ and

$$Ax = ax$$

which means that

$$px = (A - a)x = 0.$$

This implies that x is p-primary. So, $x \in M_p$ making M_p nonzero.

Conversely, suppose that $M_p \neq 0$. Then there is an element $x \in M$ so that $ann(x) = (p^k)$. This means

$$p^k x = (A - a)^k x = 0$$

and

$$y = p^{k-1}x = (A - a)^{k-1}x \neq 0$$

But then

$$(A - a)y = 0$$

making $y \in M$ into an eigenvector of A with eigenvalue a. □

Since a is an eigenvalue, I started to write $a = \lambda$. Next I want to find a basis for the cyclic module $\mathbb{C}[T]/(p^k)$ where $p = T - \lambda$.

Lemma 10.3. *If $x \in M_p$ with $ann(x) = (p^k)$ then the cyclic module $Rx = \mathbb{C}[T]x = \mathbb{C}[A]x$ has basis*

$$y_1 = (A-\lambda)^{k-1}x, y_2 = (A-\lambda)^{k-2}x, \cdot, y_k = (A-\lambda)^0 x = x$$

as a vector space over $\mathbb{C}$.

Proof. The first point is to understand that if $ann(x) = (p^k)$ we mean that p^k is the *minimal polynomial* of A acting on x. In other words, $f(A)x = 0$ iff $p(T)^k | f(T)$. This means two things.

(1) $(A-\lambda)^k x = 0$
(2) If $f(T)$ is a nonzero polynomial of degree $< k$ then $f(A)x \neq 0$.

The second statement can be reinterpretted as saying that

$$x, Ax, A^2x, \cdots, A^{k-1}x \tag{10.1}$$

are linearly independent because any linear relation can be written as

$$\sum_{i=1}^{k} c_i A^{k-i} x = 0.$$

This says that $f(A)x = 0$ where $f(T) = c_1 T^{k-1} + \cdots + c_k$ is a polynomial of degree $k-1$ or less contradicting (2).

When (1) is expanded out it says that $A^k x$ is a linear combination of the vectors in (10.1). But then A^{k+1} is a linear combination of

$$A^k x, A^{k-1}x, \cdots, Ax$$

which is in the span of the vectors $A^{k-1}x, \cdots, x$ since A^k is in that span. So, the vectors in (10.1) for a basis for the cyclic module Rx.

This is not quite what we want. However, we can modify the basis above in the following way. Suppose that $f_i(T)$ is a polynomial of degree exactly equal to i and $f_0(T) = c$ is a nonzero constant. Then

$$f_0(A)x, f_1(A)x, \cdots, f_{k-1}(A)x$$

also forms a basis for Rx since they have the same span (by induction on k): the last vector $f_{k-1}(A)x$ is equal to a nonzero constant times $A^{k-1}x$ plus lower terms. But the lower terms are in the span of $A^i x$ for $i < k-1$ by induction. So, we get everything.

Since $f_i(T) = (T-\lambda)^i$ has degree i, we conclude that $(T-\lambda)^i x$ form a basis for Rx as claimed. □

10.3. the Jordan canonical form for A.

Lemma 10.4. *If $ann(x) = (p^k)$ where $p = T - \lambda$ then the matrix of A operating on the k-dimensional vector space Rx with respect to the basis given by the above lemma is equal to the $k \times k$ Jordan block with diagonal entries λ.*

Proof. The basis elements $y_i = (A-\lambda)^{k-i}x$ are related by the equation

$$y_{i-1} = (A-\lambda)y_i.$$

In other words,

$$Ay_i = \lambda y_i + y_{i-1}$$

and $Ay_1 = \lambda y_1$. Since the jth column of the matrix for A is the vector Ay_j written in terms of the basis $y_1, \cdots, y_k$, the matrix is

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

This is the Jordan block $[\lambda]^k$. □

Finally, I recalled how matrices behave with respect to change of basis. To make it really clear, let me first do the 2×2 case.

Suppose that $n = k = 2$ and $y_1 = (A-\lambda)x$ and $y_2 = x$. Then $Ay_1 = \lambda y_1$ and $Ay_2 = \lambda y_2 + y_1$. This can be written as the matrix formula

$$A(y_1, y_2) = (Ay_1, Ay_2) = (y_1, y_2)\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

If we write $B = (y_1, y_2. \cdots)$ for the matrix with columns the basis elements y_1, y_2, etc. we get

$$AB = BJ$$

where J is the Jordan matrix. The fundamental theorem for modules over PID's gives

$$M \cong \bigoplus \mathbb{C}[T]/((T-\lambda_i)^{k_i})$$

By the lemmas above, this implies the following.

Theorem 10.5. *After conjugating by a basis change matrix B, any $n \times n$ complex matrix A is conjugate to a matrix in Jordan canonical form:*

$$B^{-1}AB = J = \bigoplus [\lambda_i]^{k_i}.$$

We need the uniqueness part of the fundamental theorem to tell us that the pairs (λ_i, k_i) are uniquely determined.