JORDAN AND RATIONAL CANONICAL FORMS

MATH 551

Throughout this note, let V be a n-dimensional vector space over a field k, and let $\phi\colon V \to V$ be a linear map. Let $\mathcal{B} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V, and let A be the matrix for ϕ with respect to the basis $\mathcal{B}$. Thus $\phi(\mathbf{e}_j) = \sum_{i=1}^n a_{ij}\mathbf{e}_j$ (so the jth column of A records $\phi(\mathbf{e}_j)$). This is the standard convention for talking about vector spaces over a field k. To make these conventions coincide with Hungerford, consider V as a *right* module over k. Recall that if $\mathcal{B}'$ is another basis for V, then the matrix for ϕ with respect to the basis $\mathcal{B}'$ is CAC^{-1}, where C is the matrix whose ith column is the description of the ith element of $\mathcal{B}$ in the basis $\mathcal{B}'$.

1. Rational Canonical Form

We give a $k[x]$-module structure to V by setting $x \cdot v = \phi(v) = Av$. Recall that the structure theorem for modules over a PID (such as $k[x]$) guarantees that $V \cong k[x]^r \oplus_{i=1}^l k[x]/f_i$ as a $k[x]$-module. Since V is a finite-dimensional k-module (vector space!), and $k[x]$ is an infinite-dimensional k-module, we must have $r = 0$, so $V \cong \oplus_{i=1}^l k[x]/f_i$. By the classification theorem we may assume that $f_1|f_2|\dots|f_l$. We may also assume that each f_i is monic (has leading coefficient one). To see this, let $f = \lambda x^s + \sum_{i=0}^{s-1} a_i x^i$. Then $\lambda^{s-1} f = (\lambda x)^s + \sum_{i=0}^{s-1} a_i \lambda^{s-1-i}(\lambda x)^i$. Now $k[x]/f \cong k[x]/\lambda^{s-1} f$, since the two polynomials generate the same ideal, and $k[x]/\lambda^{s-1} f \cong k[y]/f'$, where $f'(y) = y^s + \sum_{i=0}^{s-1} a_i \lambda^{s-1-i} y^i$. This transformation can be done preserving the relationship that f_i divides f_{i+1}.

Let $\psi : \oplus_{i=1}^l k[x]/f_i \to V$ be the isomorphism, and let V_i be $\psi(k[x]/f_i)$ (the image of this term of the direct sum). If we choose a basis for V consisting of the unions of bases for each V_i, then the matrix for ϕ will be in block form, since $\phi(v) \in V_i$ for each $v \in V_i$. Thus we can restrict our attention to $\phi|_{V_i}$.

Let $f_i = x^m + \sum_{j=1}^{m-1} a_{ij} x^j$. Notice that $\{1, x, x^2, \dots, x^{m-1}\}$ is a basis for $k[x]/f_i$. Let $v = \psi(1) \in V_i$. Then $\{v, Av, A^2v, \dots, A^{m-1}v\}$ is a basis for V_i. The matrix for $\phi|_{V_i}$ in this basis is:

$$\begin{pmatrix} 0 & 0 & \dots & 0 & -a_{i0} \\ 1 & 0 & \dots & 0 & -a_{i1} \\ 0 & 1 & \dots & 0 & -a_{i2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -a_{i(m-2)} \\ 0 & 0 & \dots & 1 & -a_{i(m-1)} \end{pmatrix}.$$

Thus if we take as our basis for V the union of these bases for V_i we have proved the existence of *Rational Canonical Form.*

Definition 1. Let $f = x^n + \sum_{i=1}^m a_i x^i$. Then the companion matrix of f is

$$\begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & -a_{n-2} \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}.$$

Theorem 2. *Every $n \times n$ matrix A is similar to a matrix B which is block-diagonal, with the ith block the companion matrix of a monic polynomial f_i, with $f_1|f_2|\dots|f_l$.*

2. Jordan Canonical Form

For this section we assume that the field k is algebraically closed.

Definition 3. A field k is algebraically closed if for every polynomial $f \in k[x]$ there is $a \in k$ with $f(a) = 0$.

Recall the alternative statement of the classification of modules over a PID: instead of having $f_1|f_2|\dots|f_l$, we can choose to have each $f_i = p_i^{n_i}$, where p_i is a prime in $k[x]$. If k is algebraically closed, then the primes in $k[x]$ are all of the form $x - a$ for $a \in k$, so when $V = \oplus_i k[x]/f_i = \oplus_i V_i$, V_i is isomorphic to $k[x]/(x - \lambda_i)^{n_i}$ for some $\lambda_i \in k$, $n_i \in \mathbb{N}$.

Let $B = A - \lambda_i I$, and consider the $k[y]$-module structure on V given by $y \cdot v = Bv$. Then for $v \in V_i$, $y \cdot v = x \cdot v - \lambda v \in V_i$, so we also have $V \cong \oplus_i V_i$ as a $k[y]$-module. Note that $y^{n_i} \cdot V_i = 0$, but $y^{n_i - 1} \cdot V_i \neq 0$, so the rational canonical form of $B|_{V_i}$ is

$$\begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

The matrix of $\phi|_{V_i}$ with respect to this basis is thus

$$\begin{pmatrix} \lambda_i & 0 & \dots & 0 & 0 \\ 1 & \lambda_i & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_i & 0 \\ 0 & 0 & \dots & 1 & \lambda_i \end{pmatrix}.$$

Reversing the order of the basis, we get the matrix of $\phi|_{V_i}$ is

$$\begin{pmatrix} \lambda_i & 1 & \dots & 0 & 0 \\ 0 & \lambda_i & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & \dots & 0 & \lambda_i \end{pmatrix}. \tag{1}$$

We thus have:

Theorem 4. *Any $n \times n$ matrix A is similar to a matrix J which is in block-diagonal form, where every block is of the form (1) for some λ_i.*

3. Computing the Jordan Canonical Form

Recall first the definition of eigenvalues of a matrix A.

Definition 5. If A is a $n \times n$ matrix over k, then $\lambda \in k$ is an eigenvalue for A if there is $v \neq 0$ in V with $Av = \lambda v$. If $\lambda \in k$ is an eigenvalue, then $v \in V$ is an eigenvector for A if $Av = \lambda v$. The characteristic polynomial of A is $p_A(x) = \det(A - xI) \in k[x]$. An element $\lambda \in k$ is an eigenvalue for A if and only $p_A(\lambda) = 0$.

Definition 6. For $\lambda \in k$, and $m \in \mathbb{N}$, let $E_\lambda^m = \{v \in V : (A - \lambda I)^m v = 0\}$. Since E_λ^m is the kernel of a matrix, it is a subspace of V.

Lemma 7. *The subspace $E_\lambda^m \neq 0$ for some m if and only if λ is an eigenvalue of A, and $E_\lambda^m \cap E_\mu^n \neq \{0\}$ for some $m, n > 0$ implies that $\lambda = \mu$.*

Proof. Suppose first that λ is an eigenvalue of A. Then E_λ^1 is the eigenspace corresponding to λ, which is thus nonempty. Conversely, suppose that E_λ^m is nonempty for some m. We will show that E_λ^1 is nonempty, so λ is an eigenvalue of A. To see this, consider $v \in E_\lambda^m$ with $v \neq 0$. We may assume that $v \notin E_\lambda^{m-1}$ (otherwise replace m by $m-1$ until this is possible or until $v \in E_\lambda^1$). Consider $w = (A - \lambda I)^{m-1} v$. Since $v \notin E_\lambda^{m-1}$, $w \neq 0$, and $(A - \lambda I) w = (A - \lambda I)^m v = 0$, so $v \in E_\lambda^1 \setminus \{0\}$, and thus λ is an eigenvalue.

Suppose that $v \in E_\lambda^m \cap E_\mu^n$ with $v \neq 0$. As above we may assume that m, n have been chosen minimally. Then consider $w = (A - \lambda I)^{m-1}$. Now $w \in E_\lambda^1 \cap E_\mu^n$ and $w \neq 0$. Replace n by a smaller integer if necessary so that $w \notin E_\mu^{n-1}$. Then $w' = (A - \mu I)^{n-1} w \neq 0$, and $w' \in E_\lambda^1 \cap E_\mu^1$. But this means $Aw' = \lambda w' = \mu w'$, so $\lambda = \mu$. □

Proposition 8. *If the characteristic polynomial of A is $p_A(x) = \prod_\lambda (x - \lambda)^{n_\lambda}$, then $E_\lambda^m \subseteq E_\lambda^{n_\lambda}$ for all m, and $\dim E_\lambda^{n_\lambda} = n_\lambda$. Furthermore, $V = \oplus_\lambda E_\lambda^{n_\lambda}$.*

Proof. Since the characteristic polynomial is the same for similar matrices (since $\det(A - xI) = \det(C(A - xI)C^{-1}) = \det(CAC^{-1} - xI)$), we can compute the characteristic polynomial from the Jordan canonical form. We thus see that n_λ is the sum of the sizes of all λ Jordan blocks. Also, note that if J is a λ Jordan block, then the corresponding standard basis vectors all lie in E_λ^m for some $m \leq n_\lambda$, and are linearly independent, and by the Lemma E_λ^m for different eigenvalues do not intersect, so we see that $V \cong \oplus E_\lambda^{n_\lambda}$. Since $\sum_\lambda n_\lambda = n$, and $\dim E_\lambda^{n_\lambda} \geq n_\lambda$, we must thus have $\dim E_\lambda^{n_\lambda} = n_\lambda$, and $E_\lambda^m = E_\lambda^{n_\lambda}$ for $m > n_\lambda$. □

Thus we have the following algorithm to compute the Jordan Canonical Form of A:

Algorithm 9.
1. Compute and factor the characteristic polynomial of A.
2. For each λ, compute a basis $\mathcal{B} = \{v_1, \ldots, v_k\}$ for $E_\lambda^{n_\lambda} / E_\lambda^{n_\lambda - 1}$, and lift to elements of $E_\lambda^{n_\lambda}$. Add the elements $(A - \lambda I)^m v_i$ to $\mathcal{B}$ for $1 \leq m < n_\lambda$.
3. Set $i = n_\lambda - 1$.
4. Complete $\mathcal{B} \cap E_\lambda^i$ to a basis for $E_\lambda^i / E_\lambda^{i-1}$. Add the element $(A - \lambda I)^m v$ to $\mathcal{B}$ for all m and $v \in \mathcal{B}$.

(5) If $i \geq 1$, set $i = i - 1$, and return to the previous step.
(6) Output $\mathcal{B}$ - the matrix for A with respect to a suitable ordering of $\mathcal{B}$ is in Jordan Canonical Form.

Proof of correctness. To show that this algorithm works we need to check that it is always possible to complete $\mathcal{B} \cap E_\lambda^k$ to a basis for $E_\lambda^k / E_\lambda^{k-1}$. Suppose $\mathcal{B} \cap E_\lambda^k$ is linearly dependent. Then there are $v_1, \ldots, v_s \in \mathcal{B} \cap E_\lambda^k$ with $\sum_i c_i v_i = 0$, with not all $c_i = 0$. By the construction of $\mathcal{B}$ we know that $v_i = (A - \lambda I) w_i$ for some $w_i \in \mathcal{B}$, so consider $w = \sum_i c_i w_i$. Then $w \neq 0$, since the w_i are linearly independent, and not all c_i are zero. In fact, by the construction of the w_i, we know $w \notin E_\lambda^k$. But $(A - \lambda I) w = 0$, so $w \in E_\lambda^1$, which is a contradiction, since $k \geq 1$. □

Example 10. Consider the matrix

$$A = \begin{pmatrix} 2 & -4 & 2 & 2 \\ -2 & 0 & 1 & 3 \\ -2 & -2 & 3 & 3 \\ -2 & -6 & 3 & 7 \end{pmatrix}.$$

Then the characteristic polynomial of A is $(x-2)^2(x-4)^2$. A basis for E_2^1 is $\{(2,1,0,2), (0,1,2,0)\}$, so since there is a two-dimensional eigenspace for 2, the Jordan canonical form will have two distinct 2 blocks, each of size one. To confirm this, check that $E_2^m = E_2^1$ for all $m > 1$. A basis for E_4^1 is $\{(0,1,1,1)\}$, while a basis for E_4^2 is $\{(0,1,1,1), (1,0,0,1)\}$, so we can take $\{(1,0,0,1)\}$ as a basis for E_4^2 / E_4^1. Then $(A - 4I)(1,0,0,1)^T = (0,1,1,1)^T$, so our basis is then $\{(2,1,0,2), (0,1,2,0), (0,1,1,1), (1,0,0,1)\}$. The matrix of the transformation with respect to this basis is:

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

4. The minimal and characteristic polynomials

Definition 11. Let $I = \{f \in k[x] : f \cdot v = 0 \text{ for all } v \in V\} = \{f \in k[x] : f(A) = 0\}$. Then I is an ideal of $k[x]$, so since $k[x]$ is a PID, $I = \langle g \rangle$ for some polynomial $g \in k[x]$. We can choose g to be monic (have leading coefficient one). The polynomial g is called the minimal polynomial of the matrix A (or linear transformation ϕ).

Lemma 12. *The minimal polynomial of a nonzero matrix A is nonzero.*

Proof. Let $\{v_1, \dots, v_n\}$ be a basis for V. Then for each i $\mathcal{O}_{v_i} = \{f \in k[x] : f(A)v_i = 0\}$ is nonzero, since $\{v_i, Av_i, A^2v_i, \dots, A^nv_i\}$ is linearly dependent. Pick nonzero $f_i \in \mathcal{O}_{v_i}$ for each i. Then $\prod_{i=1}^n f_i \in \cap_{i=1}^n \mathcal{O}_{v_i} = I$, so $I \neq 0$, and thus the generator is nonzero. □

Proposition 13. *If A is the companion matrix of a monic polynomial f, then f is the minimal polynomial of A.*

Proof. First note that $\mathbf{e}_i = A^{i-1}\mathbf{e}_1$, so $\{\mathbf{e}_1, A\mathbf{e}_1, \dots, A^{n-1}\mathbf{e}_1\}$ is linearly independent. Thus the minimal polynomial of A has degree at least n. If $f = x^n + \sum_{i=0}^{n-1} c_i x^i$, then $f(A)\mathbf{e}_1 = \sum_{i=0}^{n-1} -c_i\mathbf{e}_i + \sum_{i=0}^{n-1} c_i\mathbf{e}_i = 0$ by the construction of the companion matrix. Also $f(A)\mathbf{e}_i = f(A)A^{i-1}\mathbf{e}_1 = A^{i-1}f(A)\mathbf{e}_1 = 0$, so $f(A) = 0$, and thus $f \in I$. If f were not the minimal polynomial, then there would be a monic $g \in I$ with g dividing f. But since f is itself monic g would have to have degree less than n, which we showed above is impossible, so f is the minimal polynomial of its companion matrix. □

Corollary 14. *The minimal polynomial of A is f_l, if its rational canonical form has blocks the companion matrices of $f_1, \dots, f_l$ with $f_1|f_2|\dots|f_l$. This is $\prod_\lambda (x-\lambda)^{m(\lambda)}$, where $m(\lambda)$ is the size of the largest Jordan block corresponding to the eigenvalue λ.*

Proof. Since applying a polynomial to a matrix in block-diagonal form applies it to each block, we know that $f_l(A) = 0$, and thus the minimal polynomial of A divides f_l. Conversely, if $f(A) = 0$, then f_l divides f, since f applied to the last block of the rational canonical form is zero. Thus f_l is the minimal polynomial of A.

The second description of the minimal polynomial follows from the method to convert between the two different descriptions of modules over a PID. □

Theorem 15. *If $p_A(x)$ is the characteristic polynomial of A, then $p_A(A) = 0$.*

Proof. By Proposition 8 we know that the characteristic polynomial of A is $\prod_i (x-\lambda_i)^{n_i}$ where the ith Jordan block has eigenvalue λ_i and size n_i. Thus by Corollary 14 the minimal polynomial of A divides the characteristic polynomial of A, and thus $p_A(A) = 0$. □