11. Rational canonical form (continued).

Let F be a field, $a(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0 \in F[x]$. Recall that the matrix

$$C_{a(x)} = \begin{pmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

is called the companion matrix of $a(x)$. Last time we proved the following:

Theorem 10.2'. *Any matrix $A \in Mat_n(F)$ is similar to a matrix of the form*

$$\begin{pmatrix} C_{a_1(x)} & 0 & \ldots & 0 \\ 0 & C_{a_2(x)} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & C_{a_m(x)} \end{pmatrix}$$

for uniquely defined monic non-constant polynomials $a_1(x) \mid \ldots \mid a_m(x)$. This matrix is called the rational canonical form of A and denoted by $RCF(A)$.

Definition. The polynomials $a_1(x), \ldots, a_m(x)$ are called the invariant factors of A.

11.1. **Computation of RCF. Method I.**

Theorem 11.1. *Let $A = (a_{ij}) \in Mat_n(F)$, and consider the matrix*

$$xI - A = \begin{pmatrix} x - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & x - a_{22} & \ldots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \ldots & x - a_{nn} \end{pmatrix} \in Mat_n(F[x]).$$

Let $a_1(x), \ldots, a_m(x)$ be the invariant factors of A. Then the diagonal matrix

$$diag(\underbrace{1, \ldots, 1}_{n-m \text{ times}}, a_1(x), \ldots, a_m(x)) = \begin{pmatrix} 1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & 1 & & & \\ \vdots & & & a_1(x) & & \\ \vdots & & & & \ddots & \\ 0 & 0 & \ldots & \ldots & \ldots & a_m(x) \end{pmatrix} \in Mat_n(F[x])$$

is a Smith normal form of $xI - A$. Thus, to compute $RCF(A)$ it is enough to compute $SNF(xI - A)$.

Proof. Step 1: Let $V = F^n$, $\Omega = \{e_1, \ldots, e_n\}$ the standard basis of V and let $T \in \mathfrak{gl}(V)$ be the linear transformation such that $[T]_\Omega = A$. As before let V_T denote V considered as $F[x]$-module with x acting as T.

As we established last time $a_1(x), \ldots, a_m(x)$ are invariant factors of V_T, that is,

$$V_T \cong F[x]/(a_1(x)) \oplus \ldots \oplus F[x]/(a_m(x)).$$

Step 2: To compute $a_1(x), \ldots, a_m(x)$ we can use the algorithm from the proof of the classification theorem for modules over PID. Let $M = F[x]^n$ be the standard free $F[x]$-module of rank n with basis $\varepsilon_1, \ldots, \varepsilon_n$. Let $\varphi : M \to V_T$ be the unique $F[x]$-module homomorphism s.t. $\varphi(\varepsilon_i) = e_i$, and let $N = \operatorname{Ker} \varphi$. Then $V_T \cong M/N$, and to find $a_1(x), \ldots, a_m(x)$ it is enough to find compatible bases for M and N.

Step 3: To find compatible bases for M and N we choose generators $u_1, \ldots, u_l$ of N and express them in terms of $\varepsilon_1, \ldots, \varepsilon_n$:

$$\begin{pmatrix} u_1 \\ \vdots \\ u_l \end{pmatrix} = B \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \text{ for some } B \in Mat_{l\times n}(F[x]).$$

If we now put B into Smith normal form, its nonzero diagonal entries will be precisely $\underbrace{1, \ldots, 1}_{n-m \text{ times}}, a_1(x), \ldots, a_m(x)$ (in this order). We claim that $B = xI - A$ for a suitable choice of $u_1, \ldots, u_l$ (this will finish the proof).

Step 4: To find generators of N (as $F[x]$-module) we observe that the following relations hold in V_T:

$$xe_j = \sum_{i=1}^{n} a_{ij} e_j \text{ for } 1 \le j \le n \text{ (since } x \text{ acts as } T).$$

Furthermore, these are defining relations as any additional relation (which does not follow from these) would force $e_1, \ldots, e_n$ to be linearly dependent over F (check details).

Thus, N is generated as $F[x]$-module by the elements $u_j = (x\varepsilon_j - \sum\limits_{i=1}^{n} a_{ij}\varepsilon_i)$ where $1 \le j \le n$. The matrix expressing $u_1, \ldots, u_n$ in terms of $\varepsilon_1, \ldots, \varepsilon_n$ is precisely $xI - A$, which is what we wanted. □

11.2. **Minimal and characteristic polynomial of a matrix and their relation to RCF.**

Definition (Definition-Claim)**.** Let $A \in Mat_n(F)$.

1. Let $Ann(A) = \{p(x) \in F[x] : p(A) = 0\}$. Then $Ann(A)$ is a nonzero ideal of $F[x]$ (why nonzero?) Since $F[x]$ is a PID, there exists unique

monic polynomial $\mu_A(x)$ s.t. $Ann(A) = (\mu_A(x))$. Then $\mu_A(x)$ is called the minimal polynomial of A.

2. The polynomial $\chi_A(x) = \det(xI - A)$ is called the characteristic polynomial of A.

Theorem 11.2. *Let $A \in Mat_n(F)$ and $a_1(x), \ldots, a_m(x)$ the invariant factors of A. Then*

(1) $\mu_A(x) = a_m(x)$
(2) $\chi_A(x) = a_1(x) \ldots a_m(x)$.

Corollary 11.3 (Cayley-Hamilton Theorem)**.** *The minimal polynomial $\mu_A(x)$ divides the characteristic polynomial $\chi_A(x)$. Equivalently, $\chi_A(A) = 0$.*

Proof of Theorem 11.2. Let $V = F^n$ and let $V_A = V$ considered as $F[x]$-module with $xv = Av$. Then as before

$$V_A \cong F[x]/(a_1(x)) \oplus \ldots \oplus F[x]/(a_m(x)).$$

Given $p(x) \in F[x]$ we have $p(A) = 0 \iff p(x)$ acts trivially on $V_A \iff a_i(x) \mid p(x)$ for each $i \iff a_m(x) \mid p(x)$ (since $a_1(x) \mid \ldots \mid a_m(x)$).

Thus $p(A) = 0 \iff a_m(x) \mid p(x)$. Since $a_m(x)$ is monic, by definition $\mu_A(x) = a_m(x)$. This proves (1).

Recall that the matrix $diag(\underbrace{1, \ldots, 1}_{n-m \text{ times}}, a_1(x), \ldots, a_m(x))$ is a Smith Normal Form of $xI - A$. When a matrix is being reduced to Smith Normal Form, its determinant remains unchanged up to sign. Thus,

$$\det(xI - A) = \pm \det(diag(\underbrace{1, \ldots, 1}_{n-m \text{ times}}, a_1(x), \ldots, a_m(x))) = \pm a_1(x) \ldots a_m(x).$$

Since both $\det(xI - A)$ and $a_1(x) \ldots a_m(x)$ are monic, the sign must be $+$. Thus proves (2). □

11.3. **Second method of computing RCF.** Theorem 11.2 provides an alternative procedure for computing RCF of a matrix. Indeed, if we know the characteristic polynomial $\chi_A(x)$ of A, there remains only finitely many possibilities for the invariant factors of A since any polynomial in $F[x]$ has finitely many monic divisors. Moreover, using other information, we can often determine the invariant factors without further computation. The same is true if we know the minimal polynomial $\mu_A(x)$. If we happen to know both $\chi_A(x)$ and $\mu_A(x)$, the number of possibilities will be even smaller.

The following example illustrtates this general technique.

Example 11.1: Classify similarity classes of all $A \in Mat_7(\mathbb{R})$ s.t. $A^3 = I$ (here I is the identity matrix and $\mathbb{R}$ stands for reals).

Solution: By Theorem 10.2' two matrices are similar if and only if they have the same RCF, so the problem is equivalent to classifying possible RCFs of matrices A with $A^3 = I$.

Note that $A^3 = I \iff \mu_A(x)|(x^3 - 1)$. The polynomial $x^3 - 1$ factors into irreducibles over $\mathbb{R}$ as $x^3 - 1 = (x-1)(x^2 + x + 1)$, and thus we have three possibilities: $\mu_A(x) = x - 1$, $x^2 + x + 1$ or $x^3 - 1$.

Recall that $\mu_A(x) = a_m(x)$, the last invariant factor of A. The only other restriction comes from the fact that A is a 7×7 matrix, so the sum of degrees of all invariant factors must be equal to 7.

Case 1: $a_m(x) = x - 1$. Since all other invariant factors divide $a_m(x)$ and are non-constant, they must all be equal to $x - 1$, and there must be 7 of them. In other words, $a_1(x) = \ldots = a_7(x) = x - 1$, so $A = I$. Of course, we could have concluded the same just from the fact that the polynomial $x - 1$ vanishes at A.

Case 2: $a_m(x) = x^2 + x + 1$. Since $x^2 + x + 1$ is irreducible, as in case 1 we conclude that all invariant factors must equal $x^2 + x + 1$. But then the sum of their degrees is $2m$ which cannot equal 7. Thus, this case cannot occur.

Case 3: $a_m(x) = x^3 - 1$. The other invariant factors may equal to $x^3 - 1$, $x - 1$ and $x^2 + x + 1$, and we cannot have both $x - 1$ and $x^2 + x + 1$ (since none of them divides the other). The only other restriction (in this case) is that the sum of the degrees equals 7. Thus, we are reduced to classifying partitions of 7 where the largest part is 3, and either 1 or 2 does not occur. Clearly, there are three such partitions: $7 = 3 + 3 + 1$, $7 = 3 + 2 + 2$ and $7 = 3 + 1 + 1 + 1 + 1$. They corrrespond to the following sets of invariant factors:

(i) $a_1(x) = x - 1$, $a_2(x) = a_3(x) = x^3 - 1$

(ii) $a_1(x) = a_2(x) = x^2 + x + 1$, $a_3(x) = x^3 - 1$

(iii) $a_1(x) = a_2(x) = a_3(x) = a_4(x) = x - 1$, $a_5(x) = x^3 - 1$.

Thus, we found that there are four similarity classes of 7×7 matrices A over $\mathbb{R}$ for which $A^3 = I$.