Rational Canonical Form

Aryaman Maithani

June 18, 2022

§1. Introduction

This report is simply an application of the structure theorem of finitely generated modules over a PID. We state this theorem without proof. We require the reader to be familiar with the language of modules (the structure theorem can be taken as a blackbox, but the reader must know the terminology to understand what it is saying).
We do *not* assume any result about minimal polynomials or even the Cayley-Hamilton theorem. We derive these as consequences. Another pleasant consequence is discussed in Section 5.1 where we conclude that similarity of two matrices does not change under field extensions.

The theory here is standard, one reference is [DF04].

§2. Structure Theorem of a Module over a PID

In this section, R will denote a principal ideal domain (PID).

Theorem 2.1 (Structure Theorem)**.** Let R be a PID, and M be a finitely generated R-module. Then, there exists $r \geqslant 0$, $s \geqslant 0$, and nonzero nonunits $a_1, \ldots, a_s \in R$ satisfying $a_1 \mid a_2 \mid \cdots \mid a_s$ such that

$$M \cong R^r \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_s).$$

The integers r and s are uniquely determined. Moreover, the a_i are uniquely determined up to units. Equivalently, the ideals (a_i) are uniquely determined.

Proof. See page 462 of [DF04]. □

Note that R and $R/(a_i)$ are all cyclic modules. Thus, the above says that every finitely generated module (over a PID) is a sum of cyclic submodules.

Exercise 2.2. For an R-module M, define $\mathrm{Tor}(M) := \{x \in M : rx = 0 \text{ for some } r \neq 0\}$. Show that $\mathrm{Tor}(M)$ is a submodule of M (you will use the fact that R is an integral domain).

Now, assume that M is finitely generated. With notation as in Structure Theorem 2.1, show that $M/\mathrm{Tor}(M) \cong R^r$. This shows that r is uniquely determined.[1] Moreover,

$$\mathrm{Tor}(M) \cong R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_s).$$

§3. Vector spaces as $\mathbb{F}[x]$-modules

In what follows, $\mathbb{F}$ will denote a field, V an n-dimensional vector space ($n \geqslant 1$), and $T : V \to V$ an $\mathbb{F}$-linear transformation. As usual, $\mathbb{F}[x]$ will denote the polynomial ring over $\mathbb{F}$.

We quickly set up some notations: $\mathbb{F}^{n\times n}$ is the ring of $n \times n$ matrices, and $\mathrm{End}(V)$ is the ring of all $\mathbb{F}$-linear transformations from V to V.
We will use the notation $B = (v_1, \ldots, v_n)$ to denote an ordered basis of V.

Observation 3.1. V can be viewed as $\mathbb{F}[x]$-module by defining $x \cdot v := T(v)$ for $v \in V$.

More generally, any polynomial $p(x)$ acts on v via $p(T)$.

We shall refer to this module structure as "$\mathbb{F}[x]$ acts on V via T".

Exercise 3.2. Note that $\mathbb{F}$ is a subring of $\mathbb{F}[x]$. Thus, V inherits an $\mathbb{F}$-module structure by restriction of the $\mathbb{F}[x]$-module structure. Check that this is the usual vector space structure.

Show that a subset $W \subseteq V$ is an $\mathbb{F}[x]$-submodule iff W is an $\mathbb{F}$-submodule and W is T-invariant.[2]

Recall that $\mathbb{F}[x]$ is a PID. Since V is finite-dimensional over $\mathbb{F}$, it is finitely generated as an $\mathbb{F}$-module and hence, as an $\mathbb{F}[x]$-module. Now, by the Structure Theorem 2.1, we have an isomorphism $V \cong F \oplus \mathrm{Tor}(V)$, where F is a free $\mathbb{F}[x]$-module. Note that this isomorphism is also as $\mathbb{F}$-vector spaces. Since $\mathbb{F}[x]$ has infinite dimension over $\mathbb{F}$ and V has finite, it follows that $F = 0$. Thus, we obtain the following.

Theorem 3.3. We have an isomorphism of $\mathbb{F}[x]$-modules given as

$$V \cong \mathbb{F}[x]/(p_1) \oplus \cdots \oplus \mathbb{F}[x]/(p_s), \tag{3.1}$$

[1] We are using the fact that a nonzero commutative ring has the invariant basis number property: if $R^n \cong R^m$ for integers $n, m \geqslant 0$, then $m = n$.

[2] Recall that $W \subseteq V$ is said to be T-invariant if $T(w) \in W$ for all $w \in W$.

where $p_1, \dots, p_s \in \mathbb{F}[x]$ are nonconstant monic polynomials such that $p_i \mid p_{i+1}$. Moreover, each p_i is uniquely determined.

In the above, we have used the fact that each nonzero ideal in $\mathbb{F}[x]$ has a unique monic generator. Moreover, the generator being a nonunit is characterised by the generator being nonconstant.

We first look at the module $\mathbb{F}[x]/(p)$ for some fixed $p \in \mathbb{F}[x]$.

Definition 3.4. Given a monic polynomial

$$p = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{F}[x],$$

we define the companion matrix of p as

$$C_p := \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}.$$

The matrix above has 1s on the subdiagonal, the coefficients of p in the last column (in the manner shown), and 0 everywhere else.

Note that the lower left $(n-1) \times (n-1)$ block is simply the identity matrix.

> **Exercise 3.5.** Let B be an ordered basis of V. Show that the following are equivalent:
>
> 1. B is of the form $(v, Tv, \dots, T^{n-1}v)$ for some $v \in V$.
> 2. $[T]_B$ is of the form C_p for some monic $p \in \mathbb{F}[x]$ (necessarily of degree n).
>
> More explicitly, show that the correspondence is given as:
> $T^n v = -(a_0 v + \cdots + a_{n-1}T^{n-1}v)$ corresponds to $p = a_0 + \cdots + a_{n-1}x^{n-1} + x^n$. In particular, $p(T)v = 0$.

Theorem 3.6. Suppose that $V \cong \mathbb{F}[x]/(p)$ for some nonconstant monic $p \in \mathbb{F}[x]$. Write

$$p = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n.$$

Then, there exists $v \in V$ such that $B = (v, Tv, \dots, T^{n-1})$ is an ordered basis and the matrix of T with respect to B is a companion matrix.

Proof. Fix an isomorphism $\varphi : \mathbb{F}[x]/(p) \to V$ and let $v \in V$ be the image of $\bar{1}$. We claim that

the set $B = (v, Tv, \ldots, T^{n-1}v)$ is linearly independent. Note that $\mathbb{F}[x]/(p)$ (and hence, V) has dimension n over $\mathbb{F}$ (exercise). Thus, B will then turn out to be a basis. By Exercise 3.5, the statement about $[T]_B$ will follow.

Thus, all is needed to show is that B is actually linearly independent. To this end, suppose $c_0, \ldots, c_{n-1} \in \mathbb{F}$ are such that

$$c_0 v + c_1 Tv + \cdots + c_{n-1} T^{n-1} v = 0.$$

Using the definition of the module structure, we get

$$(c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}) v = 0.$$

Recall that $v = \varphi(\bar{1})$. Since φ is injective, we get that $(c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}) \cdot \bar{1} = 0$ in $\mathbb{F}[x]/(p)$. Using this, conclude that $p(x)$ divides $\sum c_i x^{i-1}$. But comparing degrees forces $c_i = 0$ for all i, as desired. □

The converse is true as well.

Theorem 3.7. Suppose that V has an ordered basis B such that $[T]_B = C_p$ for some $p \in \mathbb{F}[x]$. Then, $V \cong \mathbb{F}[x]/(p)$ as $\mathbb{F}[x]$-modules.

Sketch. By Exercise 3.5, we know that B is of the form $(v, Tv, \ldots, T^{n-1}v)$ for some $v \in V$. Define $\varphi : \mathbb{F}[x] \to V$ by $1 \mapsto v$. Since B is a basis, φ is surjective. Moreover, $p \in \ker(\varphi)$[3]. This induces an $\mathbb{F}[x]$-linear surjection $\mathbb{F}[x]/(p) \to V$. Note that this is also $\mathbb{F}$-linear and comparing dimensions shows that the induced map is actually an isomorphism. □

§4. Obtaining the Rational Canonical Form

Definition 4.1. The characteristic polynomial of an $n \times n$ matrix A is defined as $c_A(x) := \det(xI - A)$.

Note that as per our definition, $c_A(x)$ is a monic polynomial of degree n. (Some places define it to be $\det(A - xI)$, in which case the two definitions differ by a sign of $(-1)^n$.)

> **Exercise 4.2.** Let $p(x) \in \mathbb{F}[x]$ be a monic polynomial. Show that $p(x)$ is the equal to the characteristic polynomial of the companion matrix of $p(x)$.
>
> (Suggestion: use induction and expand the determinant along the first column.)

[3] Use the correspondence stated in Exercise 4.2.

Theorem 4.3. Let V be an n-dimensional $\mathbb{F}$-vector space, and $T : V \to V$ be a linear transformation. Then, there exists an ordered basis B of V such that $[T]_B$ is a block diagonal matrix of the form

$$\begin{bmatrix} C_{p_1} & & & \\ & C_{p_2} & & \\ & & \ddots & \\ & & & C_{p_s} \end{bmatrix}, \tag{4.1}$$

for some nonconstant monic polynomial $p_1, \ldots, p_s$ satisfying $p_i \mid p_{i+1}$.

Moreover, this sequence of polynomials does not depend on the choice of B.

Note that the statement of the theorem makes no reference to the structure of V as an $\mathbb{F}[x]$-module.

Notation. For convenience, we denote the block matrix in (4.1) as $\operatorname{diag}(C_{p_1}, \ldots, C_{p_s})$.

Proof. Treat V as an $\mathbb{F}[x]$-module as earlier. To prove existence of such a basis, we use Theorem 3.6 and the existence statement of the Structure Theorem 2.1.

Write V as a sum of cyclic $\mathbb{F}[x]$-modules as in (3.1). Note that under this isomorphism, each $\mathbb{F}[x]/(p_i)$ corresponds to some T-invariant subspace (Exercise 3.2), call this subspace V_i. Using Theorem 3.6, we get an ordered basis B_i of V_i such that $[T|_{V_i}]_{B_i}$ is the companion matrix C_{p_i}.

Now, consider the (ordered) union $B := B_1 \cup \cdots \cup B_s$ to get a basis of V such that $[T]_B$ has the form mentioned.

For uniqueness, we use Theorem 3.7 and the uniqueness statement of Structure Theorem 2.1.

Suppose that B' is an ordered basis such that $[T]_{B'} = \operatorname{diag}(C_{q_1}, \ldots, C_{q_r})$, with the q_j being nonconstant monic polynomials such that $q_j \mid q_{j+1}$.

The above naturally decomposes V as a direct sum $W_1 \oplus \cdots \oplus W_r$ of T-invariant subspaces, i.e., $\mathbb{F}[x]$-submodules. By Theorem 3.7, each W_j is isomorphic to $\mathbb{F}[x]/(q_j)$. In turn, we see that $V = \bigoplus_{j=1}^{r} \mathbb{F}[x]/(q_j)$. Now, Structure Theorem 2.1 forces $r = s$ and $p_i = q_i$ for all i. □

Recall that two matrices $A, B \in \mathbb{F}^{n \times n}$ are said to be similar if there exists an invertible matrix P such that $P^{-1}AP = B$. This is denoted by $A \sim B$.

The dictionary between linear transformations (and change of basis) and matrices (and similarity) gives us the following theorem.

Theorem 4.4. Let $A \in \mathbb{F}^{n\times n}$ be a square matrix. Then, there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ and nonconstant monic polynomials $p_1, \ldots, p_s$ such that

$$P^{-1}AP = \operatorname{diag}(C_{p_1}, \ldots, C_{p_s}) \tag{4.2}$$

with $p_i \mid p_{i+1}$. Moreover, $p_1, \ldots, p_s$ are uniquely determined.

The matrix on the right is called the rational canonical form of A, denoted $\mathrm{RCF}(A)$. The polynomials p_i are called the invariant factors of A.

We have theoretically shown the existence of the RCF. See page 480 of [DF04] to see an algorithm for actually computing the RCF and finding the invariant factors.

§5. Applications

§§5.1. Similarity over field extensions

Let $A, B \in \mathbb{F}^{n\times n}$. If $\mathrm{RCF}(A) = \mathrm{RCF}(B)$, then $A \sim B$ since each matrix is similar to its RCF. Conversely, if A is similar to B, then B is similar to $\mathrm{RCF}(A)$. The uniqueness of RCF then forces $\mathrm{RCF}(A) = \mathrm{RCF}(B)$. Thus, we have obtained:

Theorem 5.1. Two square matrices are similar iff they have the same rational canonical form.

The word "rational" refers to the fact that this form does not require one to go to a field extension (unlike Jordan canonical form, which requires the field to be algebraically closed in general).

Another feature of the RCF is the following: Let $A \in \mathbb{F}^{n\times n}$, and let $\mathbb{E}$ be a field extension of $\mathbb{F}$. Then, $\mathrm{RCF}(A)$ (over $\mathbb{F}$) is similar to A over $\mathbb{E}$ as well. Moreover, it continues to satisfy the definition of being a rational form over $\mathbb{E}$ (that is, it is a block diagonal matrix consisting of companion matrices of polynomials that divide the next one). Thus, we see that the rational canonical form is independent of $\mathbb{F}$. In particular, this has the following consequence:

Corollary 5.2. Let $A, B \in \mathbb{F}^{n\times n}$, and $\mathbb{F} \subseteq \mathbb{E}$ be a field extension. A and B are similar over $\mathbb{F}$ iff A and B are similar over $\mathbb{E}$.

Note that the above is useful because similarity over $\mathbb{E}$ may possibly be easier to check (using say, Jordan canonical form). At least for theoretical purposes, it shows that two matrices are similar if they have the same Jordan form (the existence of which may require passing to an extension).

§§5.2. Minimal polynomial

Definition 5.3. The minimal polynomial of a matrix $A \in \mathbb{F}^{n\times n}$ is the monic polynomial $m_A(x) \in \mathbb{F}[x]$ of lowest degree such that $m_A(A)$ is the zero matrix.

Similarly, the minimal polynomial of a linear transformation $T : V \to V$ is the monic polynomial $m_T(x) \in \mathbb{F}[x]$ of lowest degree such that $m_T(T)$ is the zero map.

To recall: given a polynomial $p(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{F}[x]$ and $A \in \mathbb{F}^{n\times n}$, one defines

$$p(A) := a_0I + a_1A + \cdots + a_nA^n \in \mathbb{F}^{n\times n}.$$

$p(T)$ is similarly defined by "replacing x with T" (and the constant is multiplied with the identity map). Note that these notions make sense because $\mathbb{F}^{n\times n}$ and $\mathrm{End}(V)$ are rings.

Exercise 5.4. Let T be a linear transformation and A be its matrix with respect to some ordered basis. Show that the minimal polynomials of T and A coincide.

Exercise 5.5. Let $A \in \mathbb{F}^{n\times n}$. Let $q(x) \in \mathbb{F}[x]$ be a polynomial such that $q(A) = O$. Show that $m_A(x) \mid q(x)$.
Derive a similar result for the minimal polynomial of a linear transformation.

Given a matrix $A \in \mathbb{F}^{n\times n}$, there is a ring homomorphism $\mathbb{F}[x] \to \mathbb{F}^{n\times n}$ given by $p(x) \mapsto p(A)$. The above shows that the kernel is $(m_A(x))$.

Exercise 5.6. Recall that given an R-module M, the annihilator of M is the ideal of R defined as $\{r \in R : rx = 0 \text{ for all } x \in M\}$.

Show that when $\mathbb{F}[x]$ acts on V via T, then the annihilator of V is precisely the ideal generated by $m_T(x)$.

Theorem 5.7. The minimal polynomial of $A \in \mathbb{F}^{n\times n}$ is its largest invariant factor.

Sketch. We leave the details to the reader. Using Exercise 5.4, it suffices to prove the statement for a linear transformation $T \in \mathrm{End}(V)$.

As usual, write

$$V \cong \mathbb{F}[x]/(p_1) \oplus \cdots \oplus \mathbb{F}[x]/(p_s),$$

with $p_i \mid p_{i+1}$.

Note that the annihilator of the module on the right is precisely (p_s). □

Theorem 5.8. The characteristic polynomial of $A \in \mathbb{F}^{n\times n}$ is the product of the invariant factors.

Sketch. Note that the objects appearing in the statement do not change if A is replaced by a matrix similar to A. Thus, we may assume that A is in RCF. Recall that for a block diagonal matrix $M = \operatorname{diag}(A_1, \ldots, A_k)$ with square blocks, one has $\det(M) = \det(A_1) \cdots \det(A_k)$.

Now, use Exercise 4.2 to conclude the result. □

Corollary 5.9 (Cayley-Hamilton)**.** The minimal polynomial divides the characteristic polynomial. Equivalently, every square matrix satisfies its characteristic polynomial.

References

[DF04] David S. Dummit and Richard M. Foote. *Abstract algebra*. 3rd ed. Wiley, 2004.