6. Canonical Forms

Let $\phi\colon V \longrightarrow V$ be an F-linear endomorphism, where V is a finite dimensional vector space over the field F. Our goal is to understand ϕ. If we do the usual thing, which is to pick a basis for V, $v_1, v_2, \dots, v_n$ and expand ϕ in this basis, then we get a matrix $A = (a_{ij})$ and if we choose a different basis then we get a similar matrix, BAB^{-1}, where B is the matrix giving the change of basis.

The trick is to look for invariant subspaces:

Definition 6.1. *Let $\phi\colon V \longrightarrow V$ be a linear map of finite dimensional vector spaces.*

We say that a subspace W is ***invariant*** *if $\phi(W) \subset W$.*

Suppose that W is a one dimensional invariant subspace. If $w \in W$ is a non-zero vector then

$$\phi(w) = \lambda w,$$

so that w is an eigenvector with eigenvalue λ. Now if V decomposes as a direct sum of invariant one dimensonal subspaces then we can pick a basis for these one dimensonal subspaces and ϕ is a diagonal matrix in this basis. In general this is too much to hope for but this suggests we should look for the minimal non-trivial invariant subspaces.

Suppose that we pick a non-zero vector v and and look at its image under iterates of ϕ,

$$v_i = \begin{cases} v & \text{if } i = 0 \\ \phi(v_{i-1}) & \text{otherwise.} \end{cases}$$

At some point v_i is a linear combination of the previous vectors,

$$v_i = \sum a_j v_j.$$

Put differently, if we let

$$V_i = \langle v_1, v_2, \dots, v_i \rangle,$$

then

$$V_0 \subset V_1 \subset V_2 \dots.$$

Either $V_{i+1} \neq V_i$, in which case $\dim V_{i+1} = \dim V_i + 1$ or $V_{i+1} = V_i$. This gives us a sequence of increasing linear subspaces and the last one is an invariant subspace.

Note that if W is invariant under the action of ϕ, it also invariant under the action of any polynomial in ϕ

$$f(\phi)(W) \subset W.$$

If we think of V as an $F[x]$-module, via the action of ϕ, the invariant subspaces are precisely the R-submodules and the minimal invariant subspaces are the cyclic subspaces, the R-submodules generated by a single vector. Thus the examples above are the minimal invariant subspaces.

Definition-Lemma 6.2. *Let $\phi\colon V \longrightarrow V$ be a linear endomorphism of a finite dimensonal vector space. The* ***minimal polynomial*** *$m(x) \in F[x]$ of ϕ is a monic polynomial of smallest degree such that $m(\phi) = 0$.*

$$\operatorname{Ann}(V) = \langle m \rangle.$$

Proof. Since R is a PID, the ideal $\operatorname{Ann}(V)$ is principal and a generator has minimal degree. □

There are now two cases. If F is not algebraically closed then we don't try to be too clever. We just pick a vector v, look at its iterates. The matrix of this linear transformation looks like:

$$\begin{pmatrix} 0 & 0 & 0 & \dots & -a_0 \\ 1 & 0 & 0 & \dots & -a_1 \\ 0 & 1 & 0 & \dots & -a_2 \\ 0 & 0 & 1 & \dots & -a_3 \\ \vdots & \vdots & \vdots & & -a_{n-1} \end{pmatrix}$$

Here the last entries are $-a_0$, $-a_1$, $\dots$.

Definition 6.3. *Let $m(x)$ be the monic polynomial*

$$m(x) = \sum a_i x^i,$$

of degree n. The ***companion matrix*** *of $m(x)$ is the matrix*

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

Here the last row consists of the coefficients of $-m(x)$, not including the leading term.

If F is algebraically closed we can do much better.

Definition 6.4. *Let λ be a scalar. A* ***Jordan block*** *is a matrix of the form*

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

The entries containing the ones above the main diagonal is called the super diagonal. Somewhat fancifully $J = A - \lambda I_n$ is the companion matrix associated to the zero polynomial.

Definition 6.5. *Let $\phi\colon V \longrightarrow V$ be a linear endomorphism of a finite dimensonal vector space. The* ***characteristic polynomial***

$$g(x) = \det(\phi - xI),$$

where I is the identity transformation.

Lemma 6.6. *Let A be an $n \times n$ matrix with entries in a field F.*

(1) If A is the companion matrix of a monic polynomial $g(x)$ of degree n then the minimal polynomial $m(x)$ of A is $g(x)$ and the characteristic polynomial is $(-1)^n m(x)$.
(2) If J is a Jordan block then the minimal polynomial is $(x - \lambda)^n$ and the characteristic polynomial is $(-1)^n (x - \lambda)^n$.

In particular the characteristic and minimal polynomial differ only by sign.

Proof. Suppose first that A is the companion matrix of $g(x)$. Let

$$v_i = \begin{cases} e_1 & \text{if } i = 0 \\ A^t v_{i-1} & \text{if } i > 0. \end{cases}$$

Then $v_i = e_{i+1}$, for $i = 0, 1, \dots n - 1$. In particular $v_0, v_1, \dots, v_{n-1}$ are independent.

Suppose that

$$f(x) = \sum b_i x^i \in F[x].$$

If $f(\phi) = 0$ then $f(\phi)(v) = 0$, so that

$$\sum b_i v_i = 0.$$

In particular if $f(x)$ is monic of degree k then $k \geq n$ as $v_0, v_1, \dots, v_{n-1}$ are independent.

On the other hand $m(\phi)(v) = 0$, so that $m(\phi) = 0$, as the images of v span V. It follows easily that $m(x)$ is the minimal polynomial.

One can prove that the characteristic polynomial is $-(1)^n m(x)$ by expanding about the top row and induction on n.

Now suppose that A is a Jordan block. Let $B = A - \lambda I$. Note that by direct computation that the image of B^k is the span of the first k vectors. Thus

$$B^k = 0,$$

if and only if $k \geq n$.

Thus the minimal polynomial of B divides $(x-\lambda)^n$. It follows that the minimal polynomial is equal to $(x-\lambda)^d$, for some $d \leq n$. As $B^d \neq 0$ it follows that the minimal polynomial is $(x-\lambda)^n$.

It easy to see that the characteristic polynomial is $(-1)^n(x-\lambda)^n$. □

Definition 6.7. *Let A be a matrix. We say that A is in* ***rational canonical form*** *if A is a block matrix, with zero matrices everywhere, except a bunch of square matrices containing the diagonal which are companion matrices of polynomials, $g_1, g_2, \dots, g_k$, where $g_i(x)$ divides $g_{i+1}(x)$.*

Definition 6.8. *Let A be a matrix. We say that A is in* ***Jordan canonical form*** *if A is a block matrix, with zero matrices everywhere, except a bunch of square matrices containing the diagonal which are Jordan blocks.*

Theorem 6.9 (Rational Canonical Form)**.** *Let $\phi \colon V \longrightarrow V$ be a linear map between finite dimensional vector spaces, over a field F.*

Then there is a basis $e_1, e_2, \dots, e_n$ such that the corresponding matrix is in rational canonical form. Equivalently every matrix A is similar to a matrix in rational canonical form. This decomposition is unique, if we order the blocks so that d_i divides d_{i+1}.

The minimal polynomial of $m(x)$ is the last polynomial $d_k(x)$ and the characteristic polynomial is equal to the product of $d_1, d_2, \dots, d_k$.

Proof. As $R = F[x]$ is a PID we may apply the classification of modules over a PID to conclude that V is isomorphic to the direct sum $R^r \oplus T$. As R is an infinite dimensional vector space, it follows that $r = 0$. We can present T as

$$F[x]/\langle d_1(x)\rangle \oplus F[x]/\langle d_2(x)\rangle \oplus \cdots \oplus F[x]/\langle d_k(x)\rangle,$$

where d_i divides d_{i+1}. Now each direct summand corresponds to a block of our matrix. The action is given by multiplication by x. It follows that the action of ϕ preserves this decomposition, so that in block form we only get zero matrices off the main diagonal. So we might as well assume that there is only one summand (and then only one block).

Consider the action of ϕ with respect to the basis 1, x, x^2, $\dots$, x^{n-1}, where n is the degree of $d(x)$. ϕ sends 1 to x, x to x^2 and so on. Now

$$x^k = \sum -a_i x^i,$$

where the entries $a_1, a_2, \dots, a_k$ are in the last column and

$$d(x) = x^n + \sum a_i x^i.$$

Taking transposes we get the companion matrix of $d(x)$. $\square$

Corollary 6.10 (Cayley-Hamilton)**.** *Let $\phi\colon V \longrightarrow V$ be a linear map between finite dimensional vector spaces over a field F.*

Then ϕ satisfies its own characteristic polynomial. Equivalently, the minimal polynomial divides the characteristic polynomial.

Proof. Applying the classification above, we see that the minimum polynomial is $m(x) = d_n(x)$ and the characteristic polynomial is $d_1(x)d_2(x)\dots d_n(x)$. $\square$

Theorem 6.11 (Jordan Canonical Form)**.** *Let $\phi\colon V \longrightarrow V$ be a linear map between finite dimensional vector spaces, over an algebraically closed field F.*

Then there is a basis $e_1, e_2, \dots, e_n$ such that the corresponding matrix is in Jordan canonical form. Equivalently every matrix A is similar to a matrix in Jordan canonical form.

Proof. V is is isomorphic to

$$F[x]/\langle p_1^{m_1}(x)\rangle \oplus F[x]/\langle p_2^{m_2}(x)\rangle \oplus \dots \oplus F[x]/\langle p_k^{m_k}(x)\rangle,$$

where each $p_i(x)$ is a prime (equivalently irreducible) polynomial. As before we might as well assume that there is only one summand (and then only one block).

Since F is algebraically closed, the only irreducible polynomials are in fact linear polynomials. Thus

$$p(x) = x - \lambda$$

for some $\lambda \in F$. Note that $m_1 = n$, so that V is isomorphic to

$$F[x]/\langle (x-\lambda)^n\rangle.$$

Consider the linear map $\psi = \phi - \lambda I$. For this action V is isomorphic to

$$F[y]/\langle y^n\rangle.$$

It is easy to see that if we put ψ into rational canonical form then the corresponding matrix for ϕ is a Jordan block. $\square$