15. Canonical Forms

Let $\phi\colon V \longrightarrow V$ be a linear map, 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, $v_1, v_2, \ldots, v_n$ 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 best one can hope for is that we can diagonalise A. But sometimes this does not work. If

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

then A has only one eigenvector and if a 2×2 matrix is diagonalisable then there must be two independent eigenvectors.

Definition 15.1. *Let F be a field. We say that F is **algebraically closed** if every polynomial has a zero.*

Example 15.2. *$\mathbb{R}$ is not algebraically closed.*

Indeed $x^2 + 1$ does not have any real zeroes.

Theorem 15.3 (Fundamental Theorem of Algebra)**.** *$\mathbb{C}$ is algebraically closed.*

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 and look at its iterates:

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

This gives us an infinite sequence of vectors

$$v_0, v_1, v_2, \ldots .$$

Eventually some iterate is a linear combination of the previous vectors,

$$v_n = -\sum_{0 \le i \le n-1} v_i \qquad \text{so that} \qquad v_n + \sum_{0 \le i \le n-1} v_i = 0.$$

The matrix of this linear transformation looks like:

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

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

Definition 15.4. *Let $m(x)$ be the monic polynomial*

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

*of degree n. The **companion matrix** of $m(x)$ is the square 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 15.5. *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 polynomial x^n.

We note that e_1 is an eigenvector of the Jordan block, with eigenvalue λ. There are no other eigenvectors.

Definition 15.6. *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, $d_1, d_2, \dots, d_k$, where $d_i(x)$ divides $d_{i+1}(x)$.*

Definition 15.7. *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.*

We are going to prove the existence of canonical forms using the classification of modules over a PID. The idea is to use the correspondence, introduced in Lecture 10, between linear maps

$$\phi \colon V \longrightarrow V$$

and $F[x]$-modules V.

Under this correspondence the action of ϕ is represented by multiplication by x. The classification of modules over a PID will give us a direct sum decomposition for $F[x]$,

$$V_1 \oplus V_2 \oplus \cdots \oplus V_k.$$

As this is a decomposition of $F[x]$-modules, multiplication by x respects this decomposition. In terms of ϕ if we choose a basis that respects the direct sum decomposition then the matrix for ϕ has a block decomposition with zeroes away from the main diagonal.

Each summand will be isomorphic to

$$\frac{F[x]}{\langle d(x) \rangle}$$

where $d(x)$ is a monic polynomial.

Lemma 15.8. *Let*

$$V = \frac{F[x]}{\langle d(x) \rangle}$$

where $d(x)$ is a monic polynomial of degree n.

Then V is a finite dimensional vector space over F with basis $v_0, v_1, \ldots, v_{n-1}$, where $v_i = \alpha^i$ and α is the image of x, that is, α is the left coset $x + \langle d(x) \rangle$.

If ϕ is the linear transformation corresponding to multiplication by x then

$$v_i = \begin{cases} 1 & \text{if } i = 0 \\ \phi(v_{i-1}) & \text{if } 0 < i < n-1 \\ -\sum_i a_i V_i & \text{if } i = n \end{cases}$$

where

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

In other words the transpose of the matrix associated to the basis $v_0, v_1, \ldots, v_{n-1}$ is the companion matrix of $d(x)$.

Proof. We just need to show that $v_0, v_1, \ldots, v_{n-1}$ is a basis for V.

1, x, x^2, $\ldots$, x^n, $\ldots$, certainly spans the F-vector space $F[x]$. Thus the powers of α span the image V. We show that α^m is a linear combination of $v_0, v_1, \ldots, v_{n-1}$, for $m \geq n$. By induction on m it suffices to show that α^m is a linear combination of lower powers of α. By assumption

$$\alpha^n + \sum_i a_i \alpha^i = 0,$$

since we quotient out by $d(x)$. Multiplying by α^{m-n} gives

$$\alpha^m + \sum_i a_i \alpha^{m-n+i} = 0.$$

Thus

$$\alpha^m = -\sum_i a_i \alpha^{m-n+i}$$

is a linear combination of lower power of α. Thus $v_0, v_1, \ldots, v_{n-1}$ span V.

Now suppose that

$$\sum b_i v_i = 0.$$

Then

$$\sum b_i \alpha^i = 0.$$

Let $f(x) = \sum_i b_i x^i$. Then $f(x) \in F[x]$ is a polynomial of degree less than n and the image of $f(x)$ is zero in the quotient V. But then $f(x)$ is a multiple of $d(x)$. But then $f(x) = 0$.

Thus $v_0, v_1, \ldots, v_{n-1}$ are independent. But then they are a basis. □

Theorem 15.9 (Rational Canonical Form)**.** *Let $\phi\colon V \longrightarrow V$ be a linear map, where V is a finite dimensional vector space over a field F.*

Then there is a basis $e_1, e_2, \ldots, e_n$ such that the corresponding matrix is in rational canonical form. If we choose any other basis such that the corresponding matrix is in rational canonical form then the two matrices are equal.

Equivalently every matrix A is similar to a unique matrix in rational canonical form.

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}. As we already observed, 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)

$$\frac{F[x]}{\langle d(x)\rangle}$$

where $d(x)$ is a monic polynomial.

By (15.8) there is a basis such that we get the transpose of the companion matrix for $d(x)$. $\square$

Theorem 15.10 (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 $e_1, e_2, \ldots, e_n$ such that the corresponding matrix is in Jordan canonical form. If there is another basis such that the corresponding matrix is in Jordan canonical form then the two matrices are the same, up to reordering the blocks.

Equivalently every matrix A is similar to a matrix in Jordan canonical form and this matrix is unique up to reordering the blocks.

Proof. Arguing as in (15.9), viewing V as an $F[x]$-module, V is torsio. Thus V is isomorphic to

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

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. $\square$

There are some other invariants one can attach to a linear transformation.

Definition-Lemma 15.11. *Let $\phi\colon V \longrightarrow V$ be a linear transformation of a finite dimensional vector space over a field F.*

*The **minimal polynomial** of ϕ is the smallest degree monic polynomial*

$$m_\phi(x) \in F[x] \qquad \text{such that} \qquad m_\phi(\phi) = 0.$$

If $g(x)$ is any other polynomial such that $g(\phi) = 0$ then m_ϕ divides g.

Proof. Let $f(x) \in F[x]$. As discussed in Lecture 10, $f(\phi)$ makes sense as a linear transformation of ϕ, that is, as an element of $\mathrm{Hom}_F(V, V)$. We claim a little bit more, we claim that we $f(\phi)$ is in fact $F[x]$-linear. Suppose we are given a scalar $g(x) \in F[x]$. We need to check that

$$f(\phi)(g(x) \cdot v) = g(x) \cdot f(\phi)(v).$$

The key point is that the action of $g(x)$ on v is also given by applying the linear transformation $g(\phi)$ to v. Thus we need to check that

$$f(\phi)(g(\phi)(v) = g(\phi) * f(\phi)(v)).$$

This follows as the LHS is the result of applying $f(x)g(x)$ to v and the RHS is the result of applying $g(x)f(x)$ to v. This is the same as multiplication of polynomials is commutative. Put differently, and equivalently, both sides are equal as the powers of ϕ commute with each other.

Let

$$E \colon F[x] \longrightarrow \mathrm{Hom}_{F[x]}(V, V)$$

be the function given by

$$E(f) = f(\phi).$$

It is easy to check that E is $F[x]$-linear, that is,

$$E(f + g) = E(f) + E(g) \qquad \text{and} \qquad E(g \cdot f) = g \cdot E(f).$$

Thus the kernel is an $F[x]$-submodule of $F[x]$. Thus the kernel is an ideal in the ring $F[x]$:

$$\mathrm{Ker}(E) = I \lhd F[x].$$

As $F[x]$ is a principal ideal domain, I is principal. Let f be a generator of I,

$$I = \langle f \rangle.$$

Note that I cannot be the zero ideal, as $F[x]$ is an infinite dimensional space and $\mathrm{Hom}_F(V, V)$ is finite dimensional (in fact $\mathrm{Hom}_F(V, V)$ has dimension n^2, where n is the dimension of V).

Thus f cannot be the zero polynomial. Possibily multiplying by a scalar, we may assume that $m = f$ is monic. Note that m has minimal degree since it generates the kernel of E. □

It is interesting to see what happens in some special cases. If we use the theory of modules then we can compute the minimal polynomial very quickly.

But this probably an instance where it might help to compute directly to gain more insight.

If A has size n (that is, A is a square $n \times n$ matrix) then A determines a linear map

$$F^n \longrightarrow F^n \qquad \text{given by} \qquad x \longrightarrow Ax$$

in the usual way. The minimal polynomial of A, denoted $m_A(x)$, is then the minimal polynomial of the associated linear transformation.

If

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

then the minimal polynomial is $m_A(x) = x^2$. Indeed it is to easy see that $A^2 = 0$, so that if $f(x) = x^2$ then $f(A) = 0$. It follows that $m_A(x)$ divides x^2 and so $m_A(x) = 1$, x or x^2. It is surely not 1, since this is never zero. But it is not x either as A is not the zero matrix. Thus $m_A(x) = x^2$. More generally we have

Lemma 15.12. *Let A be a square matrix.*

(1) If A is the companion matrix of $m(x)$ then $m(x) = m_A(x)$ is the minimal polynomial.

(2) If A is in rational canonical form then $m_A(x) = d_k(x)$ in the notation of (15.9)*.*

(3) If A is an $n \times n$ Jordan block with λ on the main diagonal then $m_A(x) = (x - \lambda)^n$.

(4) If A is a matrix in Jordan canonical form then

$$m_A(x) = \prod_{\lambda} (x - \lambda)^{n(\lambda)}$$

where λ ranges over the eigenvalues and $n(\lambda)$ is the largest size of a Jordan block with this eigenvalue.

We start with an easy:

Lemma 15.13. *Suppose A is a square matrix which has the block form*

$$\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}$$

(1) The minimal polynomial of A is the lowest common multiple of the minimal polynomials of B and C.

(2) The characteristic polynomial of A is the product of the characteristic polynomials of B and C.

Proof. We first prove (1). It is clear that the minimal polynomial of B divides the minimal polynomial of A. It is also not hard to see that the minimal polynomial divides any common multiple.

We now turn to (2). Suppose A has size n. We have to compute

$$\det(A - xI_n).$$

Note that $A - xI_n$ has block form

$$\begin{pmatrix} B - xI_u & 0 \\ 0 & C - xI_v \end{pmatrix}$$

where B has size u and C has size v (so that $n = u + v$). Then

$$\det(A - xI_n) = \det(B - xI_u)\det(C - xI_v). \qquad \square$$

Proof of (15.12). We first prove (1). We compute the minimal polynomial of the transpose A^t, which it is easy to see has the same minimal polynomial as A.

If we define a sequence $v_1, v_2, \ldots, v_n$ of vectors by the rule

$$v_i = \begin{cases} e_1 & \text{if } i = 0 \\ Av_{i-1} & \text{if } i > 1, \end{cases}$$

then

$$v_i = \begin{cases} e_i & \text{if } i < n \\ -\sum_{j<n} a_j e_j & \text{if } i = n, \end{cases}$$

where

$$m(x) = x^n + \sum_{j<n} a_j x^j,$$

by definition of the companion matrix. In particular $v_0, v_1, \ldots, v_{n-1}$ is a basis of V.

Now $m(A)$ applied to v_0 is zero, almost by definition of the companion matrix. More generally we have

$$\begin{aligned} m(A)v_i &= m(A)A^{i-1}v_0 \\ &= A^{i-1}m(A)v_0 \\ &= A^{i-1}0 \\ &= 0. \end{aligned}$$

Note that to get from the first line to the second line, we used the fact that powers of A commute, so that A commutes with $m(A)$. As $v_0, v_1, \ldots, v_{n-1}$ is spans V it follows that $m(A)$ is zero on the whole of V.

It is then easy to see that the minimal polynomial m_A of A divides m. On the other hand suppose that

$$f(x) = \sum b_i x^i$$

has degree k less than n. Consider the matrix $f(A)$. If we apply this to v_0 then we get

$$f(A) \cdot v = \sum b_i v_i.$$

As $v_1, v_2, \ldots, v_n$ are independent vectors this sum is zero if and only if $b_1, b_2, \ldots, b_k = 0$. Thus the degree of the minimal polynomial of A is at least n. Hence $m(x) = m_A(x)$. This is (1).

(2) follows from (1) and (15.13) (and an obvious induction).

We now turn to (3). Consider the matrix $B = A - \lambda I_n$. As we already pointed out this is the companion matrix associated to the matrix $n(x) = x^n$. Thus B has minimal polynomial x^n by (1). It follows that A has minimal polynomial $(x - \lambda)^n$.

(4) follows from (3) and (15.13). □

Definition-Theorem 15.14 (Cayley-Hamilton)**.** *Let $\phi\colon V \longrightarrow V$ be a linear map, where V is a vector space of dimension n over a field F.*

*The **characteristic polynomial** of ϕ is the polynomial*

$$\det(\phi - xI) \in F[x],$$

where I is the identity transformation.

The characteristic polynomial is a polynomial of degree n and the minimal polynomial divides the characteristic polynomial.

Proof. One way to compute a determinant is to pick a basis and compute

$$\det(A - xI_n)$$

where A is the matrix associated to ϕ and I_n is the identity matrix, the matrix associated to the identity transformation. It is then clear that the characteristic polynomial has degree n.

Now pick a basis so that A in rational canonical form. By (15.13) we may assume that A is the companion matrix of $d(x)$. In this case the minimal polynomial is $d(x)$. We check that the characteristic polynomial is

$$(-1)^{n-1} d(x).$$

We now have to compute. We expand the determinant about the top row. There are two terms,

$$-xa(x) - 1 \cdot b(x),$$

where $a(x)$ is the minor you get by striking out the top row and first column and $b(x)$ is the minor you get by striking out the top row and the second column.

$a(x)$ is the determinant of the companion matrix of

$$e(x) = x^{n-1} + \sum_{i>0} a_i x^{i-1}.$$

By induction on n, $a(x) = (-1)^{n-2} e(x)$.

The only way to get a non-zero term from the second minor is to take 1 from every super diagonal and then to take $-a_0$ from the last row. By induction on n we get

$$(-1)^{n-2} a_0.$$

Thus the characteristic polynomial is

$$\begin{aligned} -xa(x) - 1 \cdot b(x) &= (-1)^{n-1} x e(x) + (-1)(-1)^{n-2} a_0 \\ &= (-1)^{n-1} x (x^{n-1} + \sum_{i>0} a_i x^{i-1}) + (-1)^{n-1} a_0 \\ &= (-1)^{n-1} (x^n + \sum_{i>0} a_i x^i + a_0) \\ &= (-1)^{n-1} d(x). \end{aligned}$$

□

In general the minimal polynomial is much smaller than the characteristic polynomial. However we do have

Lemma 15.15. *Let $\phi\colon V \longrightarrow V$ be a linear map, where V is a vector space of dimension n over a field F.*

Then λ is a zero of the minimal polynomial if and only if it is a zero of the characteristic polynomial.

In particular the eigenvalues of ϕ are the zeroes of the minimal polynomial.

Proof. If λ is a zero of the minimal polynomial then λ is a zero of the characteristic polynomial, as the characteristic polynomial is a multiple of the minimal polynomial.

To show the converse we borrow a result from Math 100C, that there is an algebraically closed field $F \subset \bar{F}$. The computation of the minimal and characteristic polynomial does not change if we extend the base field, replacing V by

$$V \underset{F}{\otimes} \bar{F}$$

So we may assume that F is algebraically closed.

In this case we can pick a basis for V so that the corresponding matrix is in Jordan canonical form. To say that λ is a zero of the characteristic polynomial implies that there is a Jordan block with λ on the main diagonal. But then λ is a zero of the minimal polynomial. □