Further linear algebra. Chapter IV. Jordan normal form.

Andrei Yafaev

In what follows V is a vector space of dimension n and B is a basis of V. In this chapter we are concerned with linear maps $T\colon V \longrightarrow V$.

Let A be the matrix representing T in the basis B. Because A is an $n \times n$ matrix, we can form powers A^k for any k with the convention that $A^0 = I_n$. Note that A^k represents the transformation T composed $k$times.

Notice for example that when the matrix is diagonal with coefficients λ_i on the diagonal, then A^n is diagonal with coefficients λ_i^n. Notice also that such a matrix is invertible if and only if all λ_is are non-zero, then A^{-1} is the diagonal matrix with coefficients λ_i^{-1} on the diagonal.

Definition 0.1 *Let $f(X) = \sum a_i X^i \in k[X]$. We define*

$$f(T) = \sum a_i T^i.$$

where we define $T^0 = \mathrm{Id}$. This is a linear transformation.

If $A \in M_n(k)$ is a matrix then we define

$$f(A) = \sum a_i A^i,$$

This matrix $f(A)$ represents $f(T)$ in the basis B.

What is means is that we can 'evaluate' a polynomial at a $n \times n$ matrix and get another $n \times n$ matrix. We write $[f(T)]_B$ for this matrix in the basis B, obviously it is the same as $f([T]_B)$.

Let's look at an example.

Take $A = \begin{pmatrix} -1 & 3 \\ 4 & 7 \end{pmatrix}$ and $f(x) = x^2 - 5x + 3$. Then

$$f(A) = A^2 - 5A + 3 = \begin{pmatrix} 21 & 3 \\ 4 & 29 \end{pmatrix}$$

Another example: $V = M_n(k)$ and T sends M to M^t. Consider $f(x) = x^2 - 1$. As $T^2 = I$, we see that $f(T) = 0$.

Notice that

$$[f(T)]_B = f([T]_B)$$

It follows that if T is a linear map, $f(T) = 0$ if and only if $f(A) = 0$ where A is the matrix of T is some (equivalently any) basis.

Another property worth noting is that if $f, g \in k[x]$, then

$$f(T)g(T) = (fg)(T) = (gf)(T) = g(T)f(T)$$

Definition 0.2 *Recall that the characteristic polynomial of an $n \times n$ matrix A is defined by*

$$\mathrm{ch}_A(x) = \det(x \cdot I_n - A) = (-1)^n \det(A - x \cdot I_n).$$

This is a monic polynomial of degree n over k. Now suppose $T : V \to V$ is a linear map. We can define ch_T to be $\mathrm{ch}_{[T]_B}$ but we need to check that this does not depend on the basis B. If C is another basis with transition matrix M then we have:

$$\begin{aligned} \mathrm{ch}_{[T]_C}(X) &= \det(X \cdot I_n - M^{-1}[T]_B M) \\ &= \det(M^{-1}(X \cdot I_n - [T]_B)M) \\ &= \det(M)^{-1} \det(X \cdot I_n - [T]_B) \det(M) \\ &= \det(X \cdot I_n - [T]_B) \\ &= \mathrm{ch}_{[T]_B}(X) \end{aligned}$$

In other words, the characteristic polynomial does not depend on the choice of the basis in which we write our matrix.

The following (important !) theorem was proved in the first year courses.

Theorem 0.1 (Cayley–Hamilton Theorem) *For any A be an $n \times n$ matrix. We have $\mathrm{ch}_A(A) = 0$.*

We therefore have:

Theorem 0.2 (Cayley–Hamilton Theorem) *For any $T : V \longrightarrow V$ linear map, we have $\mathrm{ch}_T(T) = 0$.*

Example 0.3 *Take* $A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ *Then* $ch_A(x) = (x - \lambda_1)(x - \lambda_2)$ *and clearly* $ch_A(A) = 0$.

Take $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. *Calculate* $\mathrm{ch}_A(x)$ *and check that* $\mathrm{ch}_A(A) = 0$.

There are plenty of polynomials f such that $f(A) = 0$, all the multiples of ch_A for example.

What can also happen is that some divisor g of ch_A is already such that $g(A) = 0$. Take the identity I_n for example. Its characteristic polynomial is $(x - 1)^n$ but in fact $g = x - 1$ is already such that $g(I_n) = 0$. This leads us to the notion of **minimal polynomial**.

0.1 Minimal polynomials.

Definition 0.3 *Let V be a finite dimensional vector space over a field k and $T : V \to V$ a linear map. A* minimal polynomial *of T is the* ***monic*** *polynomial $m \in k[X]$ such that*

- $m(T) = 0$*;*
- *if* $f(T) = 0$ *and* $f \neq 0$ *then* $\deg f \geq \deg m$.

Notice that if m is a polynomial as in the definition, and f a polynomial such that $f(T) = 0$ and $\deg(f) < \deg(m)$, then $f = 0$.

Indeed, if f was not zero, then dividing f by the coefficient of the leading term, one would obtain a monic polynomial with degree strictly less than m which contradicts the definition of m.

Theorem 0.4 *Every linear map $T : V \to V$ has a unique minimal polynomial m_T. Furthermore if $f \in k[x]$ is such that $f(T) = 0$ iff $m_T | f$.*

Proof. Firstly the Cayley-Hamilton theorem implies that there exists a polynomial f satisfying $f(T) = 0$, namely $f = \mathrm{ch}_T$. Among monic polynomials satisfying $f(T) = 0$ we choose one of smallest degree, call it m_T. This shows that the minimal polynomial exists.

Suppose that m_T is not unique then there exists another monic polynomial $n(x) \in k[X]$ satisfying the conditions of the definition. Because $n(T) = 0$, $\deg(n) \geq \deg(m_T)$ and because $m_T(T) = 0$, $\deg(m_T) \geq \deg(n)$ hence $\deg(m) = \deg(n)$.

If $f(x) = m_T(x) - n(x)$ then

$$f(T) = m_T(T) - n(T) = 0,$$

also $\deg(f) < \deg(m_T) = \deg(n)$. By the remark following the definition of the minimal polynomial, we see that $f = 0$ i.e. $m_T = n$. This proves the uniqueness.

Suppose $f \in k[X]$ and $f(T) = 0$. By the Division Algorithm for polynomials there exist unique $q, r \in k[X]$ with $\deg(r) < \deg(m)$ and

$$f = qm + r.$$

Then

$$r(T) = f(T) - q(T)m(T) = 0 - q(T) \cdot 0 = 0.$$

So r is the zero polynomial (by the remark following the definition.) Hence $f = qm$ and so $m|f$.

Conversely if $f \in k[X]$ and $m|f$ then $f = qm$ for some $q \in k[X]$ and so $f(T) = q(T)m(T) = q(T) \cdot 0 = 0$. $\square$

Corollary 0.5 *If $T : V \to V$ is a linear map then $m_T|\mathrm{ch}_T$.*

Proof. By the Cayley-Hamilton Theorem $\mathrm{ch}_T(T) = 0$. $\square$

Using the corollary we can calculate the minimal polynomial as follows:

- Calculate ch_T and factorise it into irreducibles.
- Make a list of all the factors.
- Find the monic factor m of smallest degree such that $m(T) = 0$.

Example 0.6 *Suppose T is represented by the matrix $\begin{pmatrix} 2 & 1 & \\ & 2 & \\ & & 2 \end{pmatrix}$. The characteristic polynomial is*

$$\mathrm{ch}_T(X) = (X - 2)^3.$$

The factors of this are:

$$1, \quad (X-2), \quad (X-2)^2, \quad (X-2)^3.$$

The minimal polynomial is $(X - 2)^2$.

In fact this method can be speeded up: there are certain factors of the characteristic polynomial which cannot arise. To explain this we recall the definition of an *eigenvalue*

Definition 0.4 *Recall that a number $\lambda \in k$ is called an* eigenvalue *of T if there is a* ***non-zero*** *vector v satisfying*

$$T(v) = \lambda \cdot v.$$

The non-zero vector v is called an eigenvector

Remark 0.7 *It is important that an eigenvector be non-zero. If you allow zero to be an eigenvector, then* ***any*** *λ would be an eigenvalue.*

Proposition 0.8 *Let v be an eigenvector of T with eigenvalue $\lambda \in k$. Then for any polynomial $f \in k[X]$,*

$$(f(T))(v) = f(\lambda) \cdot v.$$

Proof. Just use that $T(v) = \lambda v$. $\square$

Theorem 0.9 *If $T : V \to V$ is linear and $\lambda \in k$ then the following are equivalent:*

(i) *λ is an eigenvalue of T.*

(ii) $m_T(\lambda) = 0$.

(iii) $\mathrm{ch}_T(\lambda) = 0$.

Proof. (i) $\Rightarrow$ (ii): Assume $T(v) = \lambda v$ with $v \neq 0$. Then by the proposition,

$$(m_T(T))(v) = m_T(\lambda) \cdot v.$$

But $m_T(T) = 0$ so we have $m_T(\lambda) \cdot v = 0$. Since $v \neq 0$ this implies $m_T(\lambda) = 0$.

(ii) $\Rightarrow$ (iii): This is trivial since we have already shown that m_T is a factor of ch_T.

(iii) $\Rightarrow$ (i): Suppose $\mathrm{ch}_T(\lambda) = 0$. Therefore $\det(\lambda \cdot \mathrm{Id} - T) = 0$. It follows that $(\lambda \cdot \mathrm{Id} - T)$ is not invertible hence $\lambda \cdot \mathrm{Id} - T$ has a non-zero kernel. Therefore there exists $v \in V$ such that $(\lambda \cdot \mathrm{Id} - T)(v) = 0$. But then $T(v) = \lambda \cdot v$. $\square$

Now suppose the characteristic polynomial of T factorises into irreducibles as

$$\mathrm{ch}_T(X) = \prod_{i=1}^{r} (X - \lambda_i)^{a_1}.$$

By fundamental theorem of algebra, if $k = \mathbb{C}$, we can always factorise it like this.

Then the minimal polynomial has the form

$$m_T(X) = \prod_{i=1}^{r} (X - \lambda_i)^{b_1}, \quad 1 \leq b_i \leq a_i.$$

This makes it much quicker to calculate the minimal polynomial. Indeed, in practice, the number of factors and the a_is are quite small.

Example 0.10 *Suppose T is represented by the matrix* $\mathrm{diag}(2, 2, 3)$*. The characteristic polynomial is*

$$\mathrm{ch}_T(X) = (X - 2)^2(X - 3).$$

The possibilities for the minimal polynomial are:

$$(X - 2)(X - 3), \quad (X - 3)^2(X - 3).$$

The minimal polynomial is $(X - 2)(X - 3)$*.*

0.2 Generalised Eigenspaces

Definition 0.5 *Let V be a finite dimensional vector space over a field k, and let $\lambda \in k$ be an eigenvalue of a linear map $T : V \to V$. We define for $t \in \mathbb{N}$ the t-th generalised eigenspace by:*

$$V_t(\lambda) = \ker((\lambda \cdot \mathrm{Id} - T)^t).$$

Note that $V_1(\lambda)$ is the usual eigenspace (i.e. the set of eigenvectors together with zero).

Remark 0.11 *We obviously have*

$$V_1(\lambda) \subseteq V_2(\lambda) \subseteq \dots$$

and by definition,

$$\dim V_t(\lambda) = Nullity\left((\lambda \cdot \mathrm{Id} - T)^t\right).$$

Another property is that generalised eigenspaces are T invariant:

$$T(V_i(\lambda)) \subset V_i(\lambda)$$

To see this, let $v \in V_i(\lambda)$. Then $T(T - \lambda\mathrm{Id})^i v = 0 = (T - \lambda\mathrm{Id})^i T v$, therefore $T(v) \in V_i(\lambda)$.

Example 0.12 *Let*

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

We have $\mathrm{ch}_A(X) = (X-2)^3$ so 2 is the only eigenvalue. We'll now calculate the generalised eigenspaces $V_t(2)$:

$$V_1(2) = \ker \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

We calculate the kernel by row-reducing the matrix:

$$V_1(2) = \ker \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Similarly

$$V_2(2) = \ker \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

$$V_3(2) = \ker \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Example 0.13 *Let*

$$A = \begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \\ 1 & 1 & -2 \end{pmatrix}.$$

Let V be a vector space and U and W be two subspaces. Then (exercise on Sheet 4), $U + W$ is a subspace of V. If furthermore $U \cap V = \{0\}$, then we call this subspace the **direct sum** of U and W and denote it by $U \oplus W$. In this case

$$\dim(U \oplus W) = \dim U + \dim W$$

Theorem 0.14 (Primary Decomposition Theorem) *If V is a finite dimensional vector space over k and $T : V \to V$ is linear, with distinct eigenvalues $\lambda_1, \dots, \lambda_r \in k$.*

Let f be a monic polynomial such that $f(T) = 0$ and suppose that f factorises in $k[x]$ into a product of degree one polynomials

$$f(X) = \prod_{i=1}^{r} (X - \lambda_i)^{b_i},$$

then

$$V = V_{b_1}(\lambda_1) \oplus \cdots \oplus V_{b_r}(\lambda_r).$$

Lemma 0.15 *Let k be a field. If $f, g \in k[x]$ satisfy* $\gcd(f, g) = 1$ *and T is as above then*

$$\ker(fg(T)) = \ker(f(T)) \oplus \ker(g(T)).$$

Proof. (of the theorem) We have $f(T) = 0$, so $\ker(f(T)) = V$. We have a factorization of f into pairwise coprime factors of the form $(x - \lambda_i)^{b_i}$ so the lemma implies that

$$V = \ker(f(\alpha)) = \ker\left(\prod_{i=1}^{r} (T - \lambda_i 1)^{b_i}\right) = \ker(T - \lambda_1)^{b_1} \oplus \cdots \oplus \ker(T - \lambda_t)^{b_t}$$
$$= V_{b_1}(\lambda_1) \oplus \cdots \oplus V_{b_t}(\lambda_t).$$

□

Proof. (of the lemma) Let $f, g \in k[x]$ satisfy $gcd(f, g) = 1$.

Firstly if $v \in \ker f(T) + \ker g(T)$, say $v = w_1 + w_2$, with $w_1 \in \ker f(T)$ and $w_2 \in \ker g(T)$ then

$$fg(T)v = fg(T)(w_1 + w_2) = f(g(T)w_1) + f(g(T)w_2) = f(g(T)w_1)$$

Now, f and g are polynomials in $k[x]$, hence $fg = gf$, therefore

$$f(g(T)w_1) = g(f(T)w_1) = 0$$

because $w_2 \in \ker(f(T))$.
Therefore $\ker(f(\alpha)) + \ker(g(\alpha)) \subseteq \ker(fg(\alpha))$.
We need to prove the eqaulity, here we will use that $gcd(f, g) = 1$.
Now since $gcd(f, g) = 1$ there exist $a, b \in k[x]$ such that

$$af + bg = 1.$$

So

$$a(T)f(T) + b(T)g(T) = 1 \qquad \text{(the identity map).}$$

Let $v \in \ker(fg(T))$. If

$$v_1 = a(T)f(T)v, \qquad v_2 = b(T)g(T)v$$

then $v = v_1 + v_2$ and

$$g(T)v_1 = (gaf)(T)v = a(fg(T)v) = a(T)0 = 0.$$

So $v_1 \in \ker(g(T))$. Similarly $v_2 \in \ker(f(T))$ since

$$f(T)v_2 = (fbg)(T)v = b(fg(T)v) = b(T)0 = 0.$$

Hence $\ker fg(T) = \ker f(T) + \ker g(T)$. Moreover, if $v \in \ker f(T) \cap \ker g(T)$ then $v_1 = 0 = v_2$ so $v = 0$. Hence

$$\ker fg(T) = \ker f(T) \oplus \ker g(T).$$

□

Notice that, by fundamental theorem of algebra, in $\mathbb{C}[x]$ every polynomial factorises into a product of degree one ones. The theorem applies to $\mathrm{ch}_T(x)$ and $m_T(x)$.

Definition 0.6 *Recall that a linear map $T : V \to V$ is* diagonalisable *if there is a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is a diagonal matrix. This is equivalent to saying that the basis vectors in $\mathcal{B}$ are all eigenvectors.*

Theorem 0.16 *Let V is a finite dimensional vector space over a field k and let $T : V \to V$ be a linear map with* **distinct** *eigenvalues $\lambda_1, \dots, \lambda_r \in k$. Then T is diagonalizable iff we have (in $k[X]$):*

$$m_T(X) = (X - \lambda_1) \dots (X - \lambda_r).$$

Proof. First suppose that T is diagonalisable and let $\mathcal{B}$ be a basis of eigenvectors. Let $f(X) = (X - \lambda_1) \dots (X - \lambda_r)$. We already know that $f | m_T$, so to prove that $f = m_T$ we just have to check that $f(T) = 0$. To show this, it is sufficient to check that $f(T)(v) = 0$ for each basis vector $v \in \mathcal{B}$. Suppose $v \in \mathcal{B}$, so v is an eigenvector with some eigenvalue λ_i. Then we have

$$f(T)(v) = f(\lambda) \cdot v = 0 \cdot v = 0.$$

Therefore $m_T = f$.

Conversely if $m_T = f$ then by the primary decomposition theorem we have

$$V = V_1(\lambda_1) \oplus \dots \oplus V_1(\lambda_r).$$

Let $\mathcal{B}_i$ be a basis for $V_1(\lambda_i)$. Then obviously the elements of $\mathcal{B}_i$ are eigenvectors and $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_r$ is a basis of V. Therefore T is diagonalisable. $\square$

Remark 0.17 *Observe that if the characteristic polynomial splits as $\mathrm{ch}_T(x) = (x - \lambda_1) \cdots (x - \lambda_r)$ where λ_is are distinct eigenvalues, then $m_T = \mathrm{ch}_T$ and the matrix is diagonalisable.*

The converse, of course, does not hold.

Example 0.18 *Let $k = \mathbb{C}$ and let*

$$A = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}.$$

The characteristic polynomials is $(x - 1)(x - 6)$. The minimal polynomial is the same. The matrix is diagonalisable.

One finds that the basis of eigenvectors is

$$\begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

.

In fact this matrix is diagonalisable over $\mathbb{R}$ or even $\mathbb{Q}$.

Example 0.19 *Let $k = \mathbb{R}$ and let*

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

The characteristic polynomial is x^2+1. It is irreducible over $\mathbb{R}$. The minimal polynomial is the same. The matrix is not diagonalisable over $\mathbb{R}$, however over $\mathbb{C}$ $x^2+1 = (x-i)(x+i)$ and the matrix is diagonalisable.

Example 0.20 *Let $k = \mathbb{C}$ and let*

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

The characteristic polynomials is X^2. Since $A \neq 0$ the minimal polynomial is also X^2. Since this is not a product of distinct linear factors, A is not diagonalizable over $\mathbb{C}$.

Example 0.21 *Let $k = \mathbb{C}$ and let*

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

The minimal polynomial is $(x-1)^2$. Not diagonalisable.

0.3 Jordan Bases in the one eigenvalue case

Let $T : V \to V$ be a linear map. Fix a basis for V and let A be the matrix of T in this basis. As we have seen above, it is not always the case that T can be diagonalised; i.e. there is not always a basis of V consisting of eigenvalues of T. This the case that there is no basis of eigenvalues, the best kind of basis is a Jordan basis. We shall define a Jordan basis in several steps.

Suppose $\lambda \in k$ is the only eigenvalue of a linear map $T : V \to V$. We have defined generalized eigenspaces:

$$V_1(\lambda) \subseteq V_2(\lambda) \subseteq \ldots \subseteq V_b(\lambda),$$

where b is the power of $X - \lambda$ in the minimal polynomial m_T.

We can choose a basis $\mathcal{B}_1$ for $V_1(\lambda)$. Then we can choose $\mathcal{B}_2$ so that $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for $V_2(\lambda)$ etc. Eventually we end up with a basis $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_b$ for $V_b(\lambda)$. We'll call such a basis a *pre-Jordan* basis.

Consider
$$A = \begin{pmatrix} 3 & -2 \\ 8 & -5 \end{pmatrix}.$$
One calculates the characteristic polynomial and finds $(x+1)^2$ hence $\lambda = -1$ is the only eigenvalue. The unique eigenvector is $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Hence $V_1(-1) = Span(v)$. Of course we have $(A - \lambda I_2)^2 = 0$ hence $V_2(-1) = \mathbb{C}^2$ and we complete v to a basis of $\mathbb{C}^2 = V_2(-1)$, by $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for example. We have $Av_2 = -2v_1 - v_2$ and hence in the basis $\{v_1, v_2\}$ the matrix of A
$$\begin{pmatrix} -1 & -2 \\ 0 & -1 \end{pmatrix}$$
The basis $\{v_1, v_2\}$ is a pre-Jordan basis for A.

Example 0.22
$$A = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix}$$
We have $\mathrm{ch}_A(X) = (X-1)^3$ *and* $m_A(X) = (X-1)^2$. *There is only one eigenvalue* $\lambda = 1$, *and we have generalized eigenspaces*
$$V_1(1) = \ker \begin{pmatrix} 1 & 1 & -2 \end{pmatrix}, \quad V_2(1) = \ker(0) = \mathbb{C}^3.$$
So we can choose a pre-Jordan basis as follows:
$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$
This in fact also works over $\mathbb{R}$.

Now note the following:

Lemma 0.23 *If* $v \in V_t(\lambda)$ *with* $t > 1$ *then*
$$(T - \lambda \cdot \mathrm{Id})(v) \in V_{t-1}(\lambda).$$

Proof. Clear from the definition of the generalised eigenspaces. □

Now suppose we have a pre-Jordan basis $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_b$. We call this a *Jordan basis* if in addition we have the condition:

$$(T - \lambda \cdot \mathrm{Id})\mathcal{B}_t \subset \mathcal{B}_{t-1}, \quad t = 2, 3, \ldots, b.$$

If we have a pre-Jordan basis $\mathcal{B}_1 \cup \ldots \cup \mathcal{B}_b$, then to find a Jordan basis, we do the following:

- For each basis vector $v \in \mathcal{B}_b$, replace one of the vectors in $\mathcal{B}_{b-1}$ by $(T - \lambda \cdot \mathrm{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.
- For each basis vector $v \in \mathcal{B}_{b-1}$, replace one of the vectors in $\mathcal{B}_{b-2}$ by $(T - \lambda \cdot \mathrm{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.
- etc.
- For each basis vector $v \in \mathcal{B}_2$, replace one of the vectors in $\mathcal{B}_1$ by $(T - \lambda \cdot \mathrm{Id})(v)$. When choosing which vector to replace, we just need to take care that we still have a basis at the end.

  Once finished, order the vectors of the basis approriately.

We'll prove later that this method always works.

Example 0.24 *Let's look again at*

$$A = \begin{pmatrix} 3 & -2 \\ 8 & -5 \end{pmatrix}.$$

We have seen that $\{v_1, v_2\}$ is a pre-Jordan basis, here v_2 is the second vector in the standard basis.

Replace v_1 by the vector $(A + I_2)v_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$. Then $\{v_1, v_2\}$ still forms a basis of $\mathbb{C}^2$. This is the Jordan basis for A.

We have $Av_1 = -v_1$ and $Av_2 = v_1 - v_2$ (you do not need to calculate, just use $(A + I_2)v_2 = v_1$). Hence in the new basis the matrix is

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Example 0.25 *In the example above, we replace one of the vectors in* $\mathcal{B}_1$ *by*

$$(A - I_3)\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}.$$

So we can choose a Jordan basis as follows:

$$\mathcal{B}_1 = \left\{\begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right\}, \quad \mathcal{B}_2 = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}.$$

Example 0.26 *Take* $k = \mathbb{R}$*.*

$$A = \begin{pmatrix}1 & -1\\1 & -1\end{pmatrix}.$$

Here, we have seen that the characteristic and minimal polynomials are x^2*. Therefore,* 0 *is the only eigenvalue.*

Clearly $v_1 = \begin{pmatrix}1\\1\end{pmatrix}$ *generates the eigenspace and* $V_2(0) = \mathbb{R}^2$*. We complete the basis by taking* $v_2 = \begin{pmatrix}0\\1\end{pmatrix}$*. We get a pre-Jordan basis.*

Let's construct a Jordan basis. Replace v_1 *by* $Av_2 = \begin{pmatrix}-1\\-1\end{pmatrix}$*. This is a Jordan basis. The matrix of* A *in the new basis is*

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

Example 0.27

$$A = \begin{pmatrix}2 & 2 & 2\\0 & 2 & 2\\0 & 0 & 2\end{pmatrix}.$$

Clearly, the characteristic polynomial is $(x-2)^3$ *and it is equal to the minimal polynomial,* 2 *is the only eigenvalue.*

$V_1(2)$ *has equations* $y = z = 0$*, hence it's spanned by* $v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$

$V_2(2)$ is $z = 0$ and $V_3(2)$ is $\mathbb{R}^3$. Therefore, the standard basis is a pre-Jordan basis.

We have

$$(A - 2I_3)^2 = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

hence we get

Now,

$$A - 2I_3 = \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

We have

$$(A - 2I_3)v_3 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$$

and we replace v_2 by this vector.

Now $(A - 2I_3)v_2 = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$

We get:

$$v_1 = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

We have

$$Av_1 = 2v_1, \quad Av_2 = v_1 + 2v_2, \quad Av_3 = v_2 + 2v_3$$

In this basis the matrix of A is:

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

This is a Jordan basis.

0.4 Jordan Canonical (or Normal) Form in the one eigenvalue case

The Jordan canonical form of a linear map $T : V \to V$ is essentially the matrix of T with respect to a Jordan basis. We just need to order the vecors appropriately. Everything is over a field k, often k will be $\mathbb{C}$.

Suppose for the moment that $m_T = (x - \lambda)^b$, in particular T has only one eigenvalue λ. Choose a Jordan basis:

$$\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_b,$$

Of course, as $(A - \lambda id)^b = 0$, we have $V_b(\lambda) = V$.

We have a chain of subspaces

$$V_1(\lambda) \subset V_2(\lambda) \subset \cdots \subset V_b(\lambda) = V$$

and the **pre-Jordan** basis was constructed by starting with a basis of $V_1(\lambda)$ and completing it successefully to get a basis of $V_b(\lambda) = V$. WE then altered this basis so that

$$(T - \lambda id)\mathcal{B}_i \subset \mathcal{B}_{i-1}$$

Notice that we can arrange the basis elements in **chains** : starting with a vector $v \in \mathcal{B}_b$ we get a chain

$$v, (T - \lambda id)v, (T - \lambda id)^2 v, \ldots, (T - \lambda id)^{b-1} v$$

This last vector $w = (T - \lambda id)^{b-1} v$ is in $V_1(\lambda)$. Indeed

$$(T - \lambda id)^b v = 0$$

hence

$$(T - \lambda id)w = 0$$

therefore

$$Tw = \lambda w$$

therefore $w \in V_1(\lambda)$.

We have the following

Lemma 0.28 *For any $v \in \mathcal{B}_b$ (in particular $v \notin \mathcal{B}_i$ for $i < b$!), the vectors*

$$v, (T - \lambda id)v, (T - \lambda id)^2 v, \ldots, (T - \lambda id)^{b-1} v$$

are linearly independent.

Proof. Suppose that

$$\sum_i \mu_i (T - \lambda id)^i v = 0$$

Then

$$\mu_0 v + (T - \lambda id) w = 0$$

where w is a lineat combination of $(T - \lambda id)^k v$. Multiplying by $(T - \lambda id)^{b-1}$, we get

$$\mu_0 (T - \lambda id)^{b-1} v = 0$$

but, as $v \notin V_{b-1}(\lambda)$, we see that

$$(T - \lambda id)^{b-1} v \neq 0$$

hence $\mu_0 = 0$.

Repeating the process inductively, we get that $\mu_i = 0$ for all i and the vectors we consider are linearly independent. □

Let us number the vectors in this chain as $v_b = (T - \lambda id)^{b-1} v, \ldots, v_0 = v$. In other words

$$v_i = (T - \lambda id)^{b-i} v$$

Then

$$(T - \lambda id) v_i = v_{i-1}$$

i.e.

$$T v_i = \lambda v_i + v_{i-1}$$

In other words,

$$T(v_i) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

in the basis formed by this chain.

This gives a **Jordan block** i.e. $b \times b$ matrix:

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}.$$

In this way, we arrange our Jordan basis in chains starting with $\mathcal{B}_i$ (for $i = b, b-1, \ldots, 1$) and terminating at $V_1(\lambda)$.

By putting the chains together, we get that in the Jordan basis, the matrix is of the following form:

$$\begin{pmatrix} \lambda & 1 & & & & & & & & & & \\ & \lambda & 1 & & & & & & & & & \\ & & \lambda & 1 & & & & & & & & \\ & & & \lambda & & & & & & & & \\ & & & & \lambda & 1 & & & & & & \\ & & & & & \lambda & 1 & & & & & \\ & & & & & & \lambda & 1 & & & & \\ & & & & & & & \lambda & & & & \\ & & & & & & & & \lambda & 1 & & \\ & & & & & & & & & \lambda & & \\ & & & & & & & & & & \lambda & \\ & & & & & & & & & & & \lambda \end{pmatrix}.$$

We nad write it as

$$[T]_{\mathcal{B}} = \operatorname{diag}(J_{h_1}(\lambda), \ldots, J_{h_w}(\lambda)).$$

where the J_{h_i}s are blocks corresponding to a chain of length h_i.

We can actually say more; in fact the following results determines the number of blocks:

Lemma 0.29 *The number of blocks is the dimension of the eigenspace* $V_1(\lambda)$*.*

Proof. Let $(v_1, \ldots, v_k)$ be the Jordan basis of the subspace U corresponding to one block. It is a chain, we have

$$Tv_1 = \lambda v_1$$

and

$$Tv_i = \lambda v_i + v_{i-1}$$

for $2 \le i \le k$.

Let $v \in U$ be an eigenvector : $Tv = \lambda v$. Write $v = \sum_{i=1}^{k} c_i v_i$. Then

$$Tv = c_1 \lambda v_1 + \sum_{i \ge 2} c_i(\lambda v_i + v_{i-1}) = \lambda v + \sum_{i \ge 2} c_i v_i$$

It follows that $T_v = \lambda v$ if and only if $\sum_{i \ge 2} c_i v_i = 0$ which implies that $c_2 = \cdots = c_n = 0$ and hence v is in the subspace generated by v_1. Therefore, each block determines exactly one eigenvector for eigenvalue λ. As eigenvectors from different blocks are linearly independent : they are members of a basis, the number of blocks is exactly the dimension of the eigenspace $V_1(\lambda)$. $\square$

SUMMARY :

To summarise what we have seen so far. Suppose T has one eigenvalue λ, let $m_T(x) = (x - \lambda id)^b$ be its minimal polynomial. We construct a **pre-Jordan** basis by choosing a basis $\mathcal{B}_1$ for the eigenspace $V_1(\lambda)$ and then complete by $\mathcal{B}_2$ (a certain number of vectors in $V_2(\lambda)$) and then to $\mathcal{B}_3, \ldots, \mathcal{B}_b$. Note that $V_b(\lambda) = V$. We get a pre-Jordan basis $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_b$.

Now we **alter** the pre-Jordan basis by doing the following. Start with a vector $v_b \in \mathcal{B}$, replace one of the vectors in $\mathcal{B}_{b-1}$ by $v_{b-1} = (T - \lambda id)v_b$ making sure that this v_{b-1} is linearly independent of the other vectors in $\mathcal{B}_{b-1}$. Then replace a vector in $\mathcal{B}_{b-2}$ by $v_{b-2} = (T - \lambda id)v_{b-1}$ (again choose a vector to replace by choosing one such that v_{b-2} is linearly independent of the others)... continue until you get to $V_1(\lambda)$. The last vector will be $v_1 \in V_1(\lambda)$ i.e. v_1 is an eigenvector.

We obtain a chain of vectors

$$v_1 = (T - \lambda id)v_2, v_2 = (T - \lambda id)v_3, \ldots, v_{b-1} = (T - \lambda id)v_b, v_b$$

Hence in particular

$$Tv_k = v_{k-1} + \lambda v_k$$

The subspace U spanned by this chain is T-stable (because $Tv_k = v_{k-1} + \lambda v_k$) and this chain is **linearly independent** hence the chain forms a basis of U.

In restriction to U and with respect to this basis the matrix of T is

$$J(b)(\lambda) = \begin{pmatrix} \lambda & 1 & & & & & & & & & & \\ & \lambda & 1 & & & & & & & & & \\ & & \lambda & 1 & & & & & & & & \\ & & & \lambda & & & & & & & & \\ & & & & \lambda & 1 & & & & & & \\ & & & & & \lambda & 1 & & & & & \\ & & & & & & \lambda & 1 & & & & \\ & & & & & & & \lambda & & & & \\ & & & & & & & & \lambda & 1 & & \\ & & & & & & & & & \lambda & & \\ & & & & & & & & & & \lambda & \\ & & & & & & & & & & & \lambda \end{pmatrix}.$$

One constructs such chains with all elements of $\mathcal{B}_b$. Once done, one looks for elements in $\mathcal{B}_{b-1}$ **which are not in the previously constructed chains starting at** $\mathcal{B}_b$ and constructs chains with them. Then with $\mathcal{B}_{b-2}$, etc...

In the end, the union of chains will be a **Jordan basis** and in it the matrix of T is of the form :

$$\mathrm{diag}(J_{h_1}(\lambda), \ldots, J_{h_w}(\lambda)).$$

Notice the following two observations :

1. There is always a block of size $b \times b$. Hence by knowing the degree of the minimal polynomial, in some cases it is possible to determine the shape of Jordan normal form.

2. The number of blocks is the dimension of the eigenspace $V_1(\lambda)$

Here are some examples:

Suppose you have a matrix such that

$$ch_A = (x - \lambda)^5$$

and

$$m_A(x) = (x - \lambda)^4$$

There is always a block of size 4×4, hence the Jordan normal form has one 4×4 block and one 1×1 block.

Suppose ch_A is the same but $m_A(x) = (x - \lambda)^3$. Here you need to know more. There is one 3×3 block and then either two 1×1 blocks or one 2×2

block. This is determined by the diomension of $V_1(\lambda)$. If it's three then the first possibility, if it's two then the second.

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x-1)^4$. We have

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

Clearly the rank of $A - I$ is 1, hence $\dim V_1(\lambda) = 3$.

This means that the Jordan normal form will have three blocks. Therefore there will be two blocks of size 1×1 and one of size 2×2. The Jordan normal form is

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

Another example:

$$A = \begin{pmatrix} -2 & 0 & -1 & 1 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x+2)^4$. We have

$$A + 2I = \begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We see that $(A+2I)^2 = 0$ and therefore $m_A(x) = (x+2)^2$. As there is always a block of size two, there are two possibilities : either two 2×2 blocks or one 2×2 and two 1×1.

To decide which one it is, we see that the rank of $A + 2I$ is 2 hence the dimension of the kernel is 2. There are therefore 2 blocks and the Jordan normal form is

$$\begin{pmatrix} -2 & 1 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

0.5 Jordan canonical form in general.

Once we know how to determine Jordan canonical form in one eigenvalue case, the general case is easy. Let T be a linear transformation and $\lambda_1, \dots, \lambda_r$ its eigenvalues. Suppose that the minimal polynomial decomposes as

$$m_T(x) = \prod_{i=1}^{r} (x - \lambda_i)^{b_i}$$

(recall again that this is always true over $\mathbb{C}$.)

The we have seen that

$$V = V_{b_1}(\lambda_1) \oplus \cdots \oplus V_{b_r}(\lambda_r)$$

and each $V_{b_i}(\lambda_i)$ is stable by T. Therefore in restriction to each $V_{b_i}(\lambda_i)$, T is a linear transformation with one eigenvelue, namely λ_i and the minimal polynomial of T restricted to $V_{b_i}(\lambda_i)$ is $(x - \lambda_i)^{b_i}$.

One gets the Jordan basis by taking the union of Jordan bases for each $V_{b_i}(\lambda_i)$ which are constructed as previously.

Let's look at an example.

$$A = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 2 & 1 \\ -1 & 1 & 1 \end{pmatrix}$$

One calculates that $ch_A(x) = x(x-1)^2$. Then 0 and 1 are the only eigenvalues and

$$V = V_1(0) \oplus V_2(1)$$

One finds that $V_1(0)$ is generated by

$$v_0 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

That will be the first vector of the basis.

For $\lambda = 1$.

We have

$$A - I = \begin{pmatrix} -2 & 1 & 1 \\ -2 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

We find that $V_1(\lambda)$ is spanned by

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Then

$$(A - I)^2 = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

and $V_2(\lambda)$ is spanned by

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Notice that $(A - I)v_2 = v_1$ and therefore this is already a Jordan basis. The matrix in this basis is

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

Another example :

$$A = \begin{pmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

One calculates that $ch_A(x) = (x - 1)^2(x - 10)$. Then 1 and 10 are the only eigenvalues.

One finds

$$V_1(1) = Span(u_1 = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, u_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix})$$

The dimension is two, therefore there will be two blocks of size 1×1 corresponding to the eigenvalue 1.

For $V_1(10)$, one finds

$$V_1(10) = Span(u_3 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix})$$

In the basis (u_1, u_2, u_3), the matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

It is diagonal, the matrix is diagonalisable, in fact $m_A = (x-1)(x-10)$.

And a last example : find Jordan basis and normal form of :

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

One finds that the characteristic polynomial is $ch_A(x) = (x-2)^2(x-3)^2$. Hence 2 and 3 are eigenvalues and we have

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

Clearly the dimension of the kernel is 2 and spanned by e_2 and e_4.
The eigenspace has dimension two.
So we will have two blocks of size 1×1 corresponding to eigenvalue 2.
For the eigenvalue 3:

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

We see that rows one and three are identical, others are linearly independent. It follows that the eigenspace is one-dimensional and spanned by $u = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 3 \end{pmatrix}$)

We will have one block.

Let us calculate:

$$(A - 3I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -3 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 1 \end{pmatrix}$$

We take the vector $v = \begin{pmatrix} 0 \\ 4 \\ 1 \\ -2 \end{pmatrix}$) to complete the basis of $\ker(A - 3I)^2$.

Now, we have $(A - 3I)v = u$ hence we already have a Jordan basis.

The basis (e_2, e_4, u, v) is a Jordan basis and in this basis the matrix is

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$