Chapter 2

Linear Algebra I

Let V be a finite dimensional K-vector space and $f : V \to V$ a linear transformation. We would like to find a basis $\mathcal{B}$ such that $[f]_{\mathcal{B}}$ is an 'simple' as possible. We know that not all linear transformations are diagonalisable. We will show that the matrix can be chosen to be in a slightly generalised form known as 'Jordan normal form'. Along the way we will state and prove the Cayley-Hamilton Theorem.

1 Revision on vector spaces

We list here some important definitions and results that will be used. More are given (with some overlap with the present chapter) in an appendix.

1.1 Bases and dimension

Theorem 2.1

Let V be a vector space.

1. Every spanning set for V contains a basis.
2. Every linearly independent subset of V can be extended to a basis.
3. Any two bases of V have the same cardinality.

Lemma 2.2

Let V be a vector space and $U, W \leqslant V$ two subspaces of V. Suppose that $U + V$ is finite dimensional. Then

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

1.2 Matrix representation of a linear transformation

Let V and W be two finite dimensional K-vector spaces. Fix bases $\mathcal{B} = \{v_1, \ldots, v_m\}$ and $\mathcal{C} = \{w_1, \ldots, w_n\}$ for V and W respectively.

Lemma 2.3

Let $f : V \to W$ be a linear transformation. There exists a unique matrix $[f]_{\mathcal{C},\mathcal{B}} \in M_{n\times m}(K)$ with the property that

$$\forall v \in V \quad [f(v)]_{\mathcal{C}} = [f]_{\mathcal{C},\mathcal{B}} \times [v]_{\mathcal{B}}$$

Definition 2.4. The matrix given by the above lemma is called the **matrix representation** of f with respect to $\mathcal{B}$ and $\mathcal{C}$

Remark (on notation). In the case in which $V = W$ and $\mathcal{B} = \mathcal{C}$ we sometimes write $[f]_{\mathcal{B}}$ in place of $[f]_{\mathcal{B},\mathcal{B}}$.

Lemma 2.5

The entries a_{ij} of $[f]_{\mathcal{C},\mathcal{B}}$ are given by the equation $f(v_j) = \sum_{i=1}^{n} a_{ij} w_i$

Remark. The lemma says that the j-th column of $[f]_{\mathcal{C},\mathcal{B}}$ is exactly the coordinate matrix $[f(v_j)]_{\mathcal{C}}$.

Lemma 2.6

The matrix $[f]_{\mathcal{C},\mathcal{B}}$ is invertible if and only if f is a bijection.

Notice that for all $v \in V$ we have

$$[\mathrm{Id}_W]_{\mathcal{C}',\mathcal{C}}[f]_{\mathcal{C},\mathcal{B}}[v]_{\mathcal{B}} = [f(v)]_{\mathcal{C}'} = [f]_{\mathcal{C}',\mathcal{B}'}[\mathrm{Id}_V]_{\mathcal{B}',\mathcal{B}}[v]_{\mathcal{B}}$$

$$\begin{array}{ccc} [v]_{\mathcal{B}} & \xrightarrow{[f]_{\mathcal{C},\mathcal{B}}\times} & [f(v)]_{\mathcal{C}} \\ {\scriptstyle [\mathrm{Id}_V]_{\mathcal{B}',\mathcal{B}}\times}\downarrow & & \downarrow{\scriptstyle [\mathrm{Id}_W]_{\mathcal{C}',\mathcal{C}}\times} \\ [v]_{\mathcal{B}'} & \xrightarrow[[f]_{\mathcal{C}',\mathcal{B}'}\times]{} & [f(v)]_{\mathcal{C}'} \end{array}$$

Suppose we have a linear transformation $f : V \to V$ and two bases $\mathcal{B}$ and $\mathcal{B}'$ for V. Letting $P = [\mathrm{Id}_V]_{\mathcal{B}',\mathcal{B}}$ we have

$$P[f]_{\mathcal{B}}[v]_{\mathcal{B}} = [f]_{\mathcal{B}'}P[v]_{\mathcal{B}}$$

Since this holds for all $v \in V$ we have that

$$P[f]_{\mathcal{B}} = [f]_{\mathcal{B}'}P$$

$$\begin{array}{ccc} [v]_{\mathcal{B}} & \xrightarrow{[f]_{\mathcal{B}}\times} & [f(v)]_{\mathcal{B}} \\ {\scriptstyle P\times}\downarrow & & \downarrow{\scriptstyle P\times} \\ [v]_{\mathcal{B}'} & \xrightarrow[[f]_{\mathcal{B}'}\times]{} & [f(v)]_{\mathcal{B}'} \end{array}$$

Definition 2.7. Let $A, B \in M_n(K)$. We say that A and B are **similar** if there exists an invertible matrix $P \in \mathrm{GL}_n(K)$ such that $B = P^{-1}AP$. It is denoted $A \sim B$.

From the above observations we see that, if $\mathcal{B}$ and $\mathcal{B}'$ are bases of V, then $[f]_{\mathcal{B}}$ and $[f]_{\mathcal{B}'}$ are similar.
The next lemma says that, if we fix a basis for V, there is a correspondence between linear transformations and matrices.

Lemma 2.8

Let V be an n-dimensional K-vector space and $\mathcal{B}$ a basis for V. The map $\mathrm{End}_K(V) \to M_n(K)$, $f \mapsto [f]_{\mathcal{B}}$ is an isomorphism of K-vector spaces.

1.3 Exercises

Exercise 37. Show that the relation of similarity is an equivalence relation on $M_n(K)$. That is, show that the relation is reflexive, symmetric and transitive.

Exercise 38. Let V be an n-dimensional K-vector space. Show that every element of $\mathrm{GL}_n(K)$ is a change of basis matrix for V. That is, show that for all $P \in \mathrm{GL}_(K)$ there exist bases $\mathcal{B}$ and $\mathcal{B}'$ for V such that $P = [\mathrm{Id}_V]_{\mathcal{B}',\mathcal{B}}$.

Exercise 39. Let's note first, that if we allow different bases for domain and codomain, then f does have a diagonal matrix representation. Given a linear transformation $f : V \to V$, show that there exist bases $\mathcal{B}$ and $\mathcal{B}'$ for V such that $[f]_{\mathcal{B}',\mathcal{B}}$ is diagonal and all entries are either 0 or 1. (Hint: start with a basis for the image of f.)

2 Invariant decompositions

Our approach to finding a 'simple' matrix representation of a linear transformation $f : V \to V$ will be to decompose V into smaller pieces that are each preserved by f.

Definition 2.9. Let V be a K-vector space (not necessarily finite dimensional). Let $f : V \to V$ be a linear transformation. An **eigenvalue** of f is an element $\lambda \in K$ such that there exists $v \in V \backslash \{0\}$ with $f(v) = \lambda v$. Given an eigenvalue λ, the corresponding **eigenspace** is given by $V_\lambda = \{v \in V \mid f(v) = \lambda v\}$. The non-zero elements of V_λ are called **eigenvectors** of f.

Exercise 40. Show that V_λ is a subspace of V and has dimension at least 1.

Similarly, we define eigenvalues (etc) for a square matrix $A \in M_n(K)$.

Definition 2.10. Let $A \in M_n(K)$. Consider the linear transformation $f : M_{n,1}(K) \to M_{n,1}(K)$ defined by $f(v) = Av$. We say that $\lambda \in K$ is an **eigenvalue** of A is if is an eigenvalue of f. Smilarly, the eigenspaces and eigenvectors of A are defined to be those of f.

Exercise 41. Let $A, B \in M_n(K)$ be similar matrices. Show that $\lambda \in K$ is an eigenvalue of A if and only of λ is an eigenvalue of B.

Example 2.11. Some linear transformations with their eigenvalues:

1. $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by orthogonal reflection across the plane Π given by $ax + by + cz = 0$. There are two eigenvalues: -1 and 1. The eigenspaces are $V_1 = \Pi$ and $V_{-1} = \text{span}\{(a, b, c)\}$
2. $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by rotation through an angle $\theta \in (0, \pi)$ about the line $\ell = \text{span}\{(a, b, c)\}$. There is only one eigenvalue: 1. The eigenspace is $V_1 = \ell$
3. $A = \begin{pmatrix} 2 & 1 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 7 \end{pmatrix}$ has eigenvalues 2,3 and 7.
4. $V = \{\varphi \in \mathbb{R}^{\mathbb{R}} \mid \varphi \text{ is smooth }\}$, $f : V \to V$ given by $f(\varphi) = \frac{d\varphi}{dx}$. Every $\lambda \in \mathbb{R}$ is an eigenvalue of f! Given any $\lambda \in \mathbb{R}$, the function given by $\varphi(x) = e^{\lambda x}$ is an element of V that is an eigenvector with eigenvalue λ.

Exercise 42. Let $f : V \to V$ be a linear transformation and λ an eigenvalue of f. Show that if $v \in V_\lambda$, then $f(v) \in V_\lambda$.

Definition 2.12. Let $f : V \to V$ be a linear transformation. A subspace $W \leqslant V$ is called an **invariant subspace** if $\forall w \in W$, $f(w) \in W$. Given an invariant subspace W, the **restriction** of f to W is the linear transformation

$$f|_W : W \to W \quad \text{given by} \quad f|_W(w) = f(w)$$

Exercise 43. Let V be a finite dimensional K-vector space and $f : V \to V$ a linear transformation. Suppose that $W \leqslant V$ is an f-invariant subspace. Fix a basis $\{w_1, \ldots, w_m\}$ for W and extend to a basis $\mathcal{B} = \{w_1, \ldots, w_m, v_1, \ldots, v_n\}$ for V. Show that

$$[f]_\mathcal{B} = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

for some $A \in M_m(K)$, $B \in M_{m,n}(K)$, $D \in M_n(K)$.

Example 2.13. Consider the linear transformation $f : \mathbb{Q}^3 \to \mathbb{Q}^3$ determined by

$$f(1, 0, 0) = (1, -1, 0), \quad f(0, 1, 0) = (2, 1, 0), \quad f(0, 0, 1) = (1, 0, 1)$$

Then $W = \text{span}\{(1, 0, 0), (0, 1, 0)\}$ is invariant and $[f]_\mathcal{S} = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Definition 2.14. Let V be a vector space and $W \leqslant V$ a subspace. A **complement** of W is a subspace $U \leqslant V$ such that: $U \cap W = \{0\}$ and $U + W = V$. We write $V = U \oplus W$ and say that V is a **direct sum** of U and W.

Remark. Clearly, if U is a complement of W, then W is a complement of U.

Lemma 2.15

Let V be a finite dimensional vector space and U, W two subspaces of V. Then the following are equivalent:

1. $V = U \oplus W$
2. Given any bases $\mathcal{B}$ and $\mathcal{C}$ for U and V (respectively), $\mathcal{B} \cup \mathcal{C}$ is a basis for V and $\mathcal{B} \cap \mathcal{C} = \varnothing$
3. There exist bases $\mathcal{B}$ and $\mathcal{C}$ for U and V (respectively) such that $\mathcal{B} \cup \mathcal{C}$ is a basis for V and $\mathcal{B} \cap \mathcal{C} = \varnothing$
4. $U \cap W = \{0\}$ and $\dim V = \dim U + \dim W$
5. $U + W = V$ and $\dim V = \dim U + \dim W$

Proof. $(1 \Rightarrow 2)$ Assume that the first holds. That is, assume that we have $U \cap W = \{0\}$ and $U + W = V$. Fix bases $\mathcal{B} = \{u_1, \ldots, u_m\}$ and $\mathcal{C} = \{w_1, \ldots, w_n\}$ for U and W. It is clear that $\mathcal{B} \cup \mathcal{C}$ is a spanning set for V since $V = U + W$. That the set $\mathcal{B} \cup \mathcal{C}$ is linearly independent follows from the following

$$\begin{aligned}
&\sum_{i=1}^{m} \alpha_i u_i + \sum_{i=1}^{n} \beta_i w_i = 0 && \text{(where } \alpha_i, \beta_i \in K\text{)} \\
\implies &\sum_{i=1}^{m} \alpha_i u_i = \sum_{i=1}^{n} (-\beta_i) w_i \in U \cap W \\
\implies &\sum_{i=1}^{m} \alpha_i u_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} (-\beta_i) w_i = 0 \\
\implies &\alpha_i = 0 \quad \text{and} \quad \beta_i = 0 \quad \text{for all } i
\end{aligned}$$

Therefore $\mathcal{B} \cup \mathcal{C}$ is a basis for V. Also

$$v \in \mathcal{B} \cap \mathcal{C} \implies v \in U \cap W \implies v = 0$$

Since any set containing the zero vector is linearly dependent, we conclude that $\mathcal{B} \cap \mathcal{C} = \varnothing$.

$(2 \Rightarrow 3)$ Is immediate.

$(3 \Rightarrow 4)$ Assume now that $\mathcal{B}$ and $\mathcal{C}$ are as in 2. Then

$$\begin{aligned}
\dim V &= |\mathcal{B} \cup \mathcal{C}| && (\mathcal{B} \cup \mathcal{C} \text{ is a basis for } V) \\
&= |\mathcal{B}| + |\mathcal{C}| && (\mathcal{B} \cap \mathcal{C} = \varnothing) \\
&= \dim U + \dim W
\end{aligned}$$

Also

$$\begin{aligned}
&\dim(U + W) + \dim(U \cap W) = \dim U + \dim W && \text{(Lemma 2.2)} \\
\implies &\dim V + \dim(U \cap W) = \dim V \\
\implies &\dim(U \cap W) = 0 \\
\implies &U \cap W = \{0\}
\end{aligned}$$

$(4 \Rightarrow 5)$ Assume that $U \cap W = \{0\}$ and $\dim V = \dim U + \dim W$. We have

$$\begin{aligned}
\dim V = \dim U + \dim W \implies &\dim V = \dim(U + W) + \dim(U \cap W) && \text{(using Lemma 2.2)} \\
\implies &\dim V = \dim(U + W) && (U \cap W = \{0\}) \\
\implies &V = U + W && \text{(since } U + W \leqslant V\text{)}
\end{aligned}$$

$(5 \Rightarrow 1)$ Assume that $V = U + W$ and $\dim V = \dim U + \dim W$. We have

$$\begin{aligned}
\dim V = \dim U + \dim W \implies &\dim V = \dim(U + W) + \dim(U \cap W) && \text{(Lemma 2.2)} \\
\implies &\dim V = \dim V + \dim(U + W) && \text{(since } V = U + W\text{)} \\
\implies &\dim(U + V) = 0 \\
\implies &U + V = \{0\}
\end{aligned}$$

□

Exercise 44. Show that the first three conditions in the above lemma remain equivalent even without the hypothesis that V be finite dimensional.

The following observations will be an important tool in finding a simple matrix representation.

Lemma 2.16

Let V be a finite dimensional vector space and $f : V \to V$ a linear transformation. Suppose that U and W are f-invariant subspaces of V such that $V = U \oplus W$. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for U and W respectively. Then

$$[f]_{\mathcal{B}\cup\mathcal{C}} = \begin{bmatrix} [f|_U]_{\mathcal{B}} & 0 \\ 0 & [f|_W]_{\mathcal{C}} \end{bmatrix}$$

Proof. Let $\mathcal{B} = \{u_1, \dots, u_m\}$ and $\mathcal{C} = \{w_1, \dots, w_n\}$. For $j \in \{1, \dots, m\}$ the j-th column of $[f]_{\mathcal{B}\cup\mathcal{C}}$ is equal to $[f(u_j)]_{\mathcal{B}\cup\mathcal{C}}$. Since U is f-invariant, we have $f(u_j) = \sum_{i=1}^{m} a_{ij}u_i$ for some $a_{ij} \in K$. Therefore

$$[f|_U(u_j)]_{\mathcal{B}} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad \text{and} \quad [f(u_j)]_{\mathcal{B}\cup\mathcal{C}} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Similar considerations apply to the last n columns of the matrix $[f]_{\mathcal{B}\cup\mathcal{C}}$ □

Given a linear transformation f we want to find complementary f-invariant subspaces. To help do this we consider the minimal polynomial of f.

3 Minimal polynomial

The minimal polynomial (to be defined below) is the monic polynomial of lowest degree that is satisfied by a linear transformation. We will see that it divides the characteristic polynomial.

Let V be a finite dimensional K-vector space and let $p(X) = a_0 + a_1X + \cdots + a_nX^n \in K[X]$ be a polynomial. Given a a linear transformation $f : V \to V$, define $p(f) : V \to V$ by $p(f) = a_0 \operatorname{Id}_V + a_i f + \cdots + a_n f^n$. Similarly, given a square matrix $A \in M_m(K)$ define $p(A) \in M_m(K)$ by $p(A) = a_0 I_m + a_1 A + \cdots + a_n A^n$.

Remark. If $\mathcal{B}$ a basis for V and $p(X) \in K[X]$, we have

$$[p(f)]_{\mathcal{B}} = p([f]_{\mathcal{B}})$$

Now consider the set of polynomials

$$S = \{p(X) \in K[X] \mid p(f) = 0, p(X) \neq 0\} \subseteq K[X]$$

To see that $S \neq \varnothing$. Let $n = \dim V$. We know from Lemma 2.8 that $\dim(\operatorname{End}_K(V)) = n^2$. Therefore the set $\{\operatorname{Id}_V, f, f^2, \dots, f^{n^2}\} \subseteq \operatorname{End}_K(V)$ is linearly dependent, that is, there exist $a_i \in K$ (not all equal to zero) such that $\sum_{i=0}^{n^2} a_i f^i = 0$. It follows that $\sum_{i=0}^{n^2} a_i X^i \in S$.

Since the set $\{\deg(p(X)) \mid p(X) \in S\} \subseteq \mathbb{N}$ is non-empty, there is (by the Well Ordering property of $\mathbb{N}$) a polynomial $m(X) \in S$ such that $\deg(m(X)) \leqslant \deg(p(X))$ for all $p(X) \in S$. Multiplying by an element of K if necessary, we can assume that $m(X)$ is monic.

Definition 2.17. Let $f : V \to V$. The **minimal polynomial** of f is the element of S that is of lowest degree and is monic.

Remark. If is clear from the above discussion that $m(X)$ is uniquely determined by f and that $\deg(m(X)) \geqslant 1$.

Exercise 45. Let f be a linear transformation on a vector space V with minimal polynomial $X^2 - 1$ and suppose that $2 \neq 0$ in the field of scalars. (Thus, for example, $\mathbb{F}_2$ is not allowed as the field of scalars.) Show directly that the subspaces $\{v \in V : f(v) = v\}$ and $\{v \in V : f(v) = -v\}$ are complementary subspaces of V. Find a diagonal matrix representing f.

To show that $m(X)$ divides all elements of S, we use the polynomial versions of Theorems 1.1 and 1.6.

Theorem 2.18

Let K b a field and $a(X), d(X) \in K[X]$ with $d(X) \neq 0$. There there exist $q(X), r(X) \in K[X]$ such that

$$a(X) = q(x)d(X) + r(X) \quad \text{and either } r = 0 \text{ or } \deg(r(X)) < \deg(d(X))$$

Moreover, $q(X)$ and $r(X)$ are uniquely determined by $a(X)$ and $d(X)$.

Proof. The proof follows exactly the same reasoning as in the version for $\mathbb{Z}$ (Theorem 1.1). □

Exercise 46. Modify the proof of Theorem 1.1 to produce a proof of Theorem 2.18.

Exercise 47. Use Theorem 2.18 to show that $\forall p(X) \in K[X]\ \forall k \in K, \quad p(k) = 0 \implies (X - k) \mid p(X)$

Definition 2.19. Let $a(X), b(X) \in K[X]$. A **greatest common divisor** (gcd) of $a(X)$ and $b(X)$ is an element $d(X) \in K[X]$ such that

1) $(d(X) \mid a(X)) \wedge (d(X) \mid b(X))$

2) $\forall c(X) \in K[X], \quad (c(X) \mid a(X)) \wedge (c(X) \mid b(X)) \implies c(X) \mid d(X)$

We say that $a(X)$ and $b(X)$ are **relatively prime** if 1 is a gcd of $a(X)$ and $b(X)$.

Remark. The gcd of two polynomials is not unique, but any two gcd's differ only up to multiplication by an element of K.

Example 2.20.
1. $X^2 + X + 1$ is a gcd of $X^3 - X^2 - X - 2$ and $X^4 + 2X^3 + 2X^2 + X$ in $\mathbb{R}[X]$
2. $X^2 + X + 1$ is *not* a gcd of $X^3 - X^2 - X - 2$ and $X^4 + 2X^3 + 2X^2 + X$ in $\mathbb{F}_2[X]$. A gcd is $X^3 + X^2 + X$.

Exercise 48. Let $a, b \in K$ with $a \neq b$.

a) Show that $(X - a)$ and $(X - b)$ are relatively prime.

b) Show that if $d(X)$ is monic and divides $(X - a)^m$, then $d(X) = (X - a)^k$ for some $0 \leqslant k \leqslant n$.

c) Let $m, n \in \mathbb{N}$. Show that $(X - a)^m$ and $(X - b)^n$ are relatively prime.

Theorem 2.21

Let $a(X), b(X) \in K[X]$ be two polynomials at least one of which in non-zero. Then there exists a gcd $d(X)$ of $a(X)$ and $b(X)$. Moreover, for any gcd $d(X)$ there exist $\alpha(X), \beta(X) \in K[X]$ such that

$$d(X) = \alpha(X)a(X) + \beta(X)b(X)$$

Proof. The proof follows exactly the same reasoning as in the version for $\mathbb{Z}$ (Theorem 1.6). □

Exercise 49. Modify the proof of Theorem 1.6 to produce a proof of Theorem 2.21.

We can now show that the minimal polynomial divides any polynomial that has f as a root.

Proposition 2.22

Let $m(X) \in K[X]$ be the minimal polynomial of a linear transformation f. Then

$$\forall\, p(X) \in K[X], \quad p(f) = 0 \implies m(X) \mid p(X)$$

Proof. Let $p(X) \in K[X]$ be such that $p(f) = 0$. By Theorem 2.18, there exist $q(X), r(X) \in K[X]$ such that $p(X) =$

$q(X)m(X) + r(X)$ and either $r(X) = 0$ or $\deg(r(X)) < \deg(m(X))$.

$$\begin{aligned} & r(X) = p(X) - m(X)q(X) \\ \implies & r(f) = p(f) - m(f) \circ q(f) \\ & = 0 && \text{(since } p(f) = m(f) = 0\text{)} \end{aligned}$$

We must therefore have that $r(X) = 0$, since otherwise $r(X) \in S$ and has lower degree than $m(X)$. Therefore $p(X) = q(X)m(X)$. □

Examples 2.23.

1. Let V be an n-dimensional K-vector space, let $\lambda \in K$, and consider the linear transformation $f = \lambda \operatorname{Id}_V$. The minimal polynomial is $m(X) = (X - \lambda)$. The characteristic polynomial is $(X - \lambda)^n$.
2. Consider a reflection (across a line through the origin) $f : \mathbb{R}^2 \to \mathbb{R}^2$. Since $f^2 = \operatorname{Id}_V$, we know that $m(X) \mid (X^2 - 1)$. Therefore $m(X)$ is one of $X + 1$, $X - 1$ or $X^2 - 1$ (as these are the only monic divisors of $X^2 - 1$). But if $m(X) = X + 1$, then $f = -\operatorname{Id}_V$. Similarly, if $m(X) = X - 1$, then $f = \operatorname{Id}_V$. Therefore $m(X) = X^2 - 1$. The characteristic polynomial is also $X^2 - 1$.
3. Fix a basis $\mathcal{B}$ of $\mathbb{R}^4$ and let $f : \mathbb{R}^4 \to \mathbb{R}^4$ be given by $[f]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$. Calculating the characteristic polynomial of the matrix gives $c(X) = (X - 1)(X - 2)^3$. Therefore

   $$[c(f)]_{\mathcal{B}} = c([f]_{\mathcal{B}}) = 0 \qquad \text{(by the Cayley-Hamilton theorem)}$$

   Therefore $c(f) = 0$ and hence $m(X)$ must be one of the monic divisors of $c(X)$:

   $$X - 1, X - 2, (X - 1)(X - 2), (X - 1)(X - 2)^2, (X - 1)(X - 2)^3$$

   Since

   $$[f]_{\mathcal{B}} - I \neq 0 \qquad [f]_{\mathcal{B}} - 2I \neq 0 \qquad ([f]_{\mathcal{B}} - I)([f]_{\mathcal{B}} - 2I) \neq 0 \qquad ([f]_{\mathcal{B}} - I)([f]_{\mathcal{B}} - 2I)^2 = 0$$

   we conclude that $m(X) = (X - 1)(X - 2)^2$.

Since the minimal polynomial divides the characteristic polynomial, any root of the minimal polynomial is also a root of the characteristic polynomial and therefore an eigenvalue. The next result is that, conversely, all eigenvalues are roots of the minimal polynomial.

Lemma 2.24

Let V be a finite dimensional K-vector space, $f : V \to V$ a linear transformation, and $m(X) \in K[X]$ the minimal polynomial of f. Then

$$\forall \lambda \in K, \quad m(\lambda) = 0 \iff \lambda \text{ is an eigenvalue of } f$$

Proof. Suppose that $\lambda \in K$ is an eigenvalue of f. Let $v \in V \backslash \{0\}$ be such that $f(v) = \lambda v$. For any $p(X) \in K[X]$ we have $p(f)(v) = p(\lambda)v$ and therefore

$$\begin{aligned} & m(f)(v) = m(\lambda)v \\ \implies & 0(v) = m(\lambda)v && (m(f) = 0 \in \operatorname{End}_K(V)) \\ \implies & 0_V = m(\lambda)v && (0_{\operatorname{End}_K(V)}(v) = 0_V) \\ \implies & m(\lambda) = 0_K && (v \neq 0_V) \end{aligned}$$

Now for the converse. Suppose that $m(\lambda) = 0$. We will show, without appealing to the Cayley-Hamilton theorem, that λ is an eigenvalue of f. From Exercise 47 we know that $m(\lambda) = 0$ implies that $(X - \lambda) \mid m(X)$. Let $t \in \mathbb{N}$ be given by

$$t = \max\{n \in \mathbb{N} \mid (X - \lambda)^n \mid m(X)\}$$

Then $m(X) = (X - \lambda)^t q(X)$ with $\deg(q(X)) < \deg(m(X))$ and $q(\lambda) \neq 0$. Since $\deg(q(X)) < \deg(m(X))$ we have $q(f) \neq 0_{\operatorname{End}_K(V)}$. Let $v \in V$ be such that $q(f)(v) \neq 0_V$. Letting $w = q(f)(v)$ we have

$$(f - \lambda \operatorname{Id}_V)^t(w) = (f - \lambda \operatorname{Id}_V)^t q(f)(v) = m(f)(v) = 0_{\operatorname{End}_K(V)}(v) = 0_V$$

Now define $s \in \mathbb{Z}$, with $0 \leqslant s < t$ to be maximal with the property that $(f - \lambda \mathrm{Id}_V)^s(w) \neq 0$. Then letting $u = (f - \lambda \mathrm{Id}_V)^s(w)$ we have $u \neq 0$ and

$$(f - \lambda \mathrm{Id}_V)u = (f - \lambda \mathrm{Id}_V)^{s+1}(w) = 0$$

Therefore u is an eigenvector with eigenvalue λ.

□

Exercise 50. Find the minimal polynomials of the matrices:

$$\begin{bmatrix} 2 & 0 \\ 3 & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}.$$

Exercise 51. Show that the matrices

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

have the same minimal polynomial. Do they have the same characteristic polynomial?

Exercise 52. Show that the matrix

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$$

has minimal polynomial $X^2 - 2X - 8$. Use this to determine the inverse of A.

Exercise 53. Show that a linear transformation f is invertible if and only if its minimal polynomial has non-zero constant term. Assuming f is invertible, how can the inverse be calculated if the minimal polynomial is known?

Exercise 54. Suppose that A is an $n \times n$ upper triangular matrix with zeroes on the diagonal. Prove that $A^n = 0$.

Lemma 2.25

Let V be a (not necessarily finite dimensional) K-vector space and let $p(X) \in K[X]$. Then $\ker(p(f))$ is an f-invariant subspace of V.

Proof. Let $p(X) = \sum_{i=0}^{n} a_i X^i$. Let $v \in \ker(p(f))$. Then

$$\begin{aligned} p(f)(f(v)) &= (\sum_{i=0}^{n} a_i f^i)(f(v)) = \sum_{i=0}^{n} a_i f^{i+1}(v) = f(\sum_{i=0}^{n} a_i f^i(v)) \\ &= f(p(f)(v)) = f(0) = 0 \end{aligned}$$

Therefore $f(v) \in \ker(p(f))$. □

Remark. Notice that the point in the above proof is the f and $p(f)$ commute.

The following will be a crucial tool in developing Jordan normal form.

Lemma 2.26

Let V be a finite dimensional K-vector space. Let $f : V \to V$ be a linear transformation and $m(X) \in K[X]$ its minimal polynomial. Suppose that $m(X) = p(X)q(X)$ where $p(X), q(X) \in K[X]$ are monic polynomials that are relatively prime. Then

1. $V = \ker(p(f)) \oplus \ker(q(f))$
2. the minimal polynomial of $f|_{\ker(p(f))}$ is $p(X)$

3. the minimal polynomial of $f|_{\ker(q(f))}$ is $q(X)$

Proof.

$$\begin{aligned}
&\exists\, a(X), b(X) \in K[X], \quad a(X)p(X) + b(X)q(X) = 1 && \text{Theorem 2.21}\\
&a(f)p(f) + b(f)q(f) = \mathrm{Id}_V\\
&\forall v \in V, \quad v = a(f)p(f)(v) + b(f)q(f)(v) && (*)
\end{aligned}$$

sdfsdf

$$V = \ker(q(f)) + \ker(p(f))$$

and

$$v \in \ker(q(f)) \cap \ker(p(f)) \implies v = a(f)(0) + b(f)(0) \qquad \text{(by } *\text{)}$$

Therefore $V = \ker(q(f)) \oplus \ker(p(f))$
To see that $p(X)$ is the minimal polynomial of $g = f|_{\ker(p(f))}$ note first that

$$p(g) = p(f|_{\ker(p(f))}) = p(f)|_{\ker(p(f))} = 0$$

Let $p'(X)$ be any non-zero polynomial such that $p'(g) = 0$. Then, for all $v \in V$ we have

$$\begin{aligned}
p'(f)q(f)(v) &= p'(f)q(f)(u + w) && \text{(for some } u \in \ker(q(f)) \text{ and } w \in \ker(p(f)))\\
&= p'(f)q(f)(u) + p'(f)q(f)(w)\\
&= p'(f)q(f)(w) && (u \in \ker(q(f)))\\
&= p'(f)(z) && (\text{letting } z = q(f)(w))\\
&= p'(g)(z) && (\text{since } z \in \ker(p(f)))\\
&= 0(z) = 0
\end{aligned}$$

Therefore $(p(X)q(X)) \mid (p'(X)q(X))$ since $p(X)q(X)$ is the minimal polynomial of f. This implies that $\deg(p(X)) \leqslant \deg(p'(X))$.
The same argument can be used to show that $q(X)$ is the minimal polynomial of $f|_{\ker(q(f))}$. □

Proposition 2.27

Let V be a finite dimensional K-vector space and $f : V \to V$ a linear transformation. Suppose that the minimal polynomial of f can be factorised as a product of pairwise relatively prime monic polynomials $q_i \in K[X]$:

$$m(X) = q_1(X)q_2(X)\dots q_N(X)$$

Let $\mathcal{B}_i$ be a basis for $\ker(q_i(f))$ and $A_i = [f|_{\ker(q_i(f))}]_{\mathcal{B}_i}$
Then $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_N$ is a basis for V and

$$[f]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_N \end{bmatrix}$$

Proof. We consider the case $N = 2$. Extension to the general case is then an easy induction argument.
Suppose $m(X) = q_1(X)q_2(X)$. Let $K_i = \ker(q_i(f))$. By Lemma 2.25 K_i is f-invariant and by Lemma 2.26 $V = K_1 \oplus K_2$. Applying Lemma 2.16 we have

$$[f]_{\mathcal{B}_1 \cup \mathcal{B}_2} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

□

Example 2.28. As an illustration of the above proposition, consider $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by $[f]_{\mathcal{S}} = A = \begin{bmatrix} -13 & 4 & 7 \\ -18 & 6 & 9 \\ -14 & 4 & 8 \end{bmatrix}$
A quick calculation gives the characteristic polynomial as $X^2(X-1)$ and so the eigenvalues as 0 and 1. Multiplication gives that $A(A-I) \neq 0$ and $A^2(A-I) = 0$. Therefore the minimal polynomial of f is $m(X) = X^2(X-1)$. Let $q_1(X) = X^2$ and $q_2(X) = X - 1$. Note that they are relatively prime. Using that

$$q_1(A) = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -2 & 0 & 2 \end{bmatrix} \qquad q_2(A) = \begin{bmatrix} -14 & 4 & 7 \\ -18 & 5 & 9 \\ -14 & 4 & 7 \end{bmatrix}$$

we obtain bases $\mathcal{B}_1 = \{(1,0,1),(0,1,0)\}$ and $\mathcal{B}_2 = \{(1,0,2)\}$ for $\ker(q_1(f))$ and $\ker(q_2(f))$ respectively. Letting $\mathcal{B} = \{(1,0,1),(0,1,0),(1,0,2)\}$, we have

$$[f]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}^{-1} A \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} -6 & 4 & 0 \\ -9 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We want to analyze the blocks A_i that arise in the Proposition 2.27 from factors of the form $(X-\lambda)^m$.

Lemma 2.29

Let V be an n-dimensional K-vector space and let $f : V \to V$ be a linear transformation. Suppose that the minimal polynomial of f if $(X-\lambda)^m$ for some $\lambda \in K$ and $m \in \mathbb{N}$. Then $m \leqslant n$ and there exists a basis $\mathcal{B}$ of V such that $[f]_{\mathcal{B}}$ is in upper-triangular form with all entries on the diagonal equal to λ. That is,

$$[f]_{\mathcal{B}} = \begin{bmatrix} \lambda & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda \end{bmatrix}$$

Proof. Since λ is a root of the minimal polynomial, it is an eigenvalue (Lemma 2.24). For $0 \leqslant i \leqslant m$ define $V_i = \ker(f - \lambda \operatorname{Id}_V)^i$. Note that

$$\{0\} = V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_m = V$$

We now show that $V_{i-1} \neq V_i$. Suppose, for a contradiction, that $V_{i-1} = V_i$ for some i. Then for all $v \in V$, we have

$$\begin{aligned} & (f - \lambda \operatorname{Id}_V)^m(v) = 0 \\ \implies & (f - \lambda \operatorname{Id}_V)^i \circ (f - \lambda \operatorname{Id}_V)^{m-i}(v) = 0 \\ \implies & (f - \lambda \operatorname{Id}_V)^{i-1} \circ (f - \lambda \operatorname{Id}_V)^{m-i}(v) = 0 \qquad \text{(since } V_{i-1} = V_i\text{)} \\ \implies & (f - \lambda \operatorname{Id}_V)^{m-1}(v) = 0 \end{aligned}$$

Since this holds for all $v \in V$ we have that $(f - \lambda \operatorname{Id}_V)^{m-1} = 0$, contradicting the minimality of $m(X)$.
Now choose $\mathcal{B}_i \subset V_i$ such that $\mathcal{B}_i$ is a basis for V_i and $\mathcal{B}_1 \subsetneq \mathcal{B}_2 \subsetneq \cdots \subsetneq \mathcal{B}_m$. Since $|B_{i+1}| \geqslant |\mathcal{B}_i| + 1$ we have that $n = |\mathcal{B}_m| \geqslant m$.
We also have

$$\begin{aligned} v \in V_i \implies & (f - \lambda \operatorname{Id}_V)^i(v) = 0 \\ \implies & (f - \lambda \operatorname{Id}_V)^{i-1}(f - \lambda \operatorname{Id}_V)(v) = 0 \\ \implies & (f - \lambda \operatorname{Id}_V)(v) \in V_{i-1} \\ \implies & f(v) - \lambda v = w \quad \text{for some } w \in V_{i-1} \\ \implies & f(v) = \lambda v + w \end{aligned}$$

That $[f]_{\mathcal{B}_m}$ has the desired form then follows. □

Exercise 55. Let $f : (\mathbb{F}_5)^3 \to (\mathbb{F}_5)^3$ be given by $[f]_{\mathcal{S}} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$

(a) Calculate the eigenvalues of f.

(b) Find the minimal polynomial of f.

(c) Find a basis $\mathcal{B}$ of $(\mathbb{F}_5)^3$ such that $[f]_{\mathcal{B}}$ is in upper triangular form.

Proposition 2.30

Let V be an finite dimensional K-vector space and let $f : V \to V$ be a linear transformation. Suppose that the minimal polynomial of f is of the form

$$m(X) = (X - \lambda_1)^{m_1}(X - \lambda_2)^{m_2} \dots (X - \lambda_N)^{m_N}$$

for some $N \in \mathbb{N}$, $m_i \in \mathbb{N}$ and $\lambda_i \in K$ with $\lambda_i \neq \lambda_j$ if $i \neq j$.
Then there exists a basis $\mathcal{B}$ of V such that $[f]_{\mathcal{B}}$ is in upper triangular form.

Proof. Note that for $i \neq j$ the polynomials $(X - \lambda_i)^{m_i}$ and $(X - \lambda_j)^{m_j}$ are relatively prime.
Let $q_i(X) = (X - \lambda_i)^{m_i} \in K[X]$, $V_i = \ker(q_i(f)) \leqslant V$, and $f_i = f|_{V_i}$. Choose a basis $\mathcal{B}_i$ for V_i and let $A_i = [f_i]_{\mathcal{B}_i}$.
By Lemma 2.26, $V = V_1 \oplus \dots \oplus V_N$ and the minimal polynomial of f_i is $q_i(X)$.
By Lemma 2.29

$$A_i = \begin{bmatrix} \lambda_i & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_i \end{bmatrix}$$

Let $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_N$. By Proposition 2.27, $\mathcal{B}$ is a basis for V and

$$[f]_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_N \end{bmatrix} = \begin{bmatrix} \lambda_1 & * & * & & & 0 \\ 0 & \ddots & * & & & \\ 0 & 0 & \lambda_1 & & & \\ & & & \ddots & & \\ & 0 & & & \lambda_N & * & * \\ & & & & 0 & \ddots & * \\ & & & & 0 & 0 & \lambda_N \end{bmatrix}$$

□

Corollary 2.31

Let V be an finite dimensional K-vector space and let $f : V \to V$ be a linear transformation. If K is algebraically closed, then there exists a basis $\mathcal{B}$ of V such that $[f]_{\mathcal{B}}$ is in upper triangular form.

Proof. Since K is algebraically closed, any polynomial in $K[X]$ can be written as a product of linear terms. In particular, for the minimal polynomial of f we have

$$m(X) = (X - \lambda_1)^{m_1}(X - \lambda_2)^{m_2} \dots (X - \lambda_N)^{m_N}$$

for some $N \in \mathbb{N}$, $m_i \in \mathbb{N}$ and $\lambda_i \in K$ with $\lambda_i \neq \lambda_j$ if $i \neq j$. Apply the preceding result. □

4 The Cayley-Hamilton theorem

We first recall here the definition of the characteristic polynomial.

Definition 2.32. Let K be a field and $A \in M_n(K)$. The **characteristic polynomial** of A is the polynomial $c_A(X) \in K[X]$ given by

$$c_A(X) = \det(XI_n - A)$$

Remark. The polynomial $c_A(X)$ is always monic and of degree n. The constant term of $c(X)$ is equal to $(-1)^n \det(A)$.

Exercise 56. Show that if $A, B \in M_n(K)$ are similar, then $c_A(X) = c_B(X)$.

Definition 2.33. Let V be an finite dimensional K-vector space and let $f : V \to V$ be a linear transformation. The **characteristic polynomial** of f is denoted $c_f(X)$ and given by $c_f(X) = c_A(X)$ for some matrix representation $A = [f]_{\mathcal{B}}$ of f.

Remark. The above exercise shows that $c_f(X)$ does not depend on the choice of A.

Theorem 2.34: Cayley-Hamilton Theorem

Let V be an finite dimensional K-vector space and let $f : V \to V$ be a linear transformation. Let $c_f(X) \in K[X]$ be the characteristic polynomial of f. Then $c_f(f) = 0$.
(That is, 'f satisfies its own characteristic equation'.)

Proof. Assume first that K is algebraically closed. Then the minimal polynomial has the form

$$m(X) = (X - \lambda_1)^{m_1}(X - \lambda_2)^{m_2} \dots (X - \lambda_N)^{m_N}$$

for some $N \in \mathbb{N}$, $m_i \in \mathbb{N}$ and $\lambda_i \in K$ with $\lambda_i \neq \lambda_j$ if $i \neq j$. As in the proof of Proposition 2.30 define $V_i = \ker(f - \lambda_i)^{m_i}$. By Lemma 2.25, V_i is f-invariant. Let $f_i = f|_{V_i}$. By Lemma 2.26, $V = V_1 \oplus \dots \oplus V_N$ andt he minimal polynomial of f_i is $(X - \lambda_i)^{m_i}$. Let $n_i = \dim(V_i)$. By Lemma 2.29, $m_i \leqslant n_i$ and there exists a basis $\mathcal{B}_i$ for V_i such that

$$[f_i]_{\mathcal{B}_i} = \begin{bmatrix} \lambda_i & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_i \end{bmatrix} \in M_{n_i}(K)$$

By Lemma 2.16 we have

$$[f]_{\mathcal{B}} = \begin{bmatrix} [f_i]_{\mathcal{B}_i} & & 0 \\ & \ddots & \\ 0 & & [f_i]_{\mathcal{B}_i} \end{bmatrix} = \begin{bmatrix} \begin{smallmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_1 \end{smallmatrix} & & 0 \\ & \ddots & \\ 0 & & \begin{smallmatrix} \lambda_N & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_N \end{smallmatrix} \end{bmatrix}$$

Letting $A = [f]_{\mathcal{B}}$, we have We have

$$\begin{aligned} c_f(X) = c_A(X) &= \det(XI_n - A) \\ &= \Pi_{i=1}^N \det(XI_{n_i} - A_i) \\ &= \Pi_{i=1}^N (X - \lambda_i)^{n_i} \end{aligned}$$

Since $m_i \leqslant n_i$, we have

$$m(X) = (X - \lambda_1)^{m_1} \dots (X - \lambda_N)^{m_N} \mid (X - \lambda_1)^{n_1} \dots (X - \lambda_N)^{n_N} = c(X)$$

Therefore $c(f) = 0$ and we're done in the case in which K is algebraically closed.
Now consider the case in which K is not algebraically closed. Let L be an algebraically closed field with $K \subseteq L$ (see Theorem 1.34). Let $\mathcal{B}$ be any basis of V and let $A = [f]_{\mathcal{B}}$. Then $A \in M_n(K) \subseteq M_n(L)$. Applying the result already obtained for algebraically closed fields, we have that A satisfies its characteristic equation and therefore

$$[c(f)]_{\mathcal{B}} = c(A) = 0$$

□

Recall that a linear transformation $f : V \to V$ is called **diagonalisable** if there exists a basis $\mathcal{B}$ of V such that $[f]_{\mathcal{B}}$ is a diagonal matrix.

Proposition 2.35

Let K be an algebraically closed field and let V be a finite dimensional K-vector space. A linear transformation $f : V \to V$ is diagonalisable if and only if its minimal polynomial can be written as a product of distinct linear factors (in $K[X]$).

Proof. Denote the minimal polynomial by $m(X) \in K[X]$. Since K is algebraically closed, $m(X)$ can be written as a product of linear factors

$$m(X) = (X - \lambda_1)^{m_1} \dots (X - \lambda_N)^{m_N}$$

with $\lambda_i \neq \lambda_j$ if $i \neq j$. Define V_i, f_i and $\mathcal{B}_i$ as above. Then

$$\begin{aligned} f \text{ is diagonalisable} &\iff \forall i, \quad f_i \text{ is diagonalisable} \\ &\iff \forall i, \quad f_i = \lambda_i \operatorname{Id}_{V_i} && (\lambda_i \text{ is the only eigenvaluie of } f_i) \\ &\iff \forall i, \quad \text{the minimal polynomial of } f_i \text{ is } (X-\lambda_i) \\ &\iff \forall i, \quad m_i = 1 && (\text{the minimal polynomial of } f_i \text{ is } (X-\lambda_i)^{m_i}) \end{aligned}$$

□

5 Jordan normal form

Linear transformations that are not diagonalisable do nonetheless have a nice matrix representation which is block diagonal with each block of a simple form.

Definition 2.36. Let $\lambda \in K$ and $n \in \mathbb{N}$. Define a matrix $J(\lambda, n) \in M_n(K)$ to be the matrix with λ at all entries on the main diagonal, 1 at all entries directly above the main diagonal, and 0 elsewhere.

$$J(\lambda, n) = \begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & & & & \lambda \end{bmatrix}$$

A marix of this form is called a **Jordan block**.

Example 2.37. $J(4,3) = \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}$

Exercise 57.

(a) Show that the characteristic polynomial of $J(\lambda, n)$ is $(X-\lambda)^n$

(b) Show that the minimal polynomial of $J(\lambda, n)$ is $(X-\lambda)^n$

(c) Show that the eigenspace has dimension 1.

Definition 2.38. A square matrix is said to be in **Jordan normal form** (JNF) if it is block diagonal and each of the blocks is a Jordan block.

Example 2.39. The matrix shown is in JNF. Note that the characteristic and minimal polynomials are

$$c(X) = (X-2)^3(X-i)^2 \qquad m(X) = (X-2)^2(X-i)^2$$

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & i & 1 & 0 & 0 \\ 0 & 0 & i & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Definition 2.40. A linear transformation $T : V \to V$ is called **nilpotent** if there exists $n \in \mathbb{N}$ such that $T^n = 0$. Similarly, a square matrix $A \in M_m(K)$ is called nilpotent if $A^n = 0$ for some $n \in \mathbb{N}$.

Example 2.41.

1. The linear transformation $D : \mathcal{P}_d(K) \to \mathcal{P}_d(K)$ given by differentiation is nilpotent.
2. The matrix $\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ is nilpotent.

Exercise 58. Show that if a linear transformation $T : V \to V$ is nilpotent, then it is not injective.

The following technical looking lemma will be used in the proof of the existence of a Jordan normal form matrix representative.

Lemma 2.42

Let K be a field and V a finite dimensional K-vector space. Suppose that $N : V \to V$ is a nilpotent linear transformation. Then there exist $v_1, \dots, v_k \in V$ and $m_1, \dots, m_k \in \mathbb{N}$ such that

$$\forall i \in \{1, \dots, k\}, \quad N^{m_i+1} v_i = 0$$

and

$$\begin{aligned} \{\, & N^{m_1}u_1, \dots, N^2u_1, Nu_1, u_1, \\ & N^{m_2}u_2, \dots, N^2u_2, Nu_2, u_2, \\ & \vdots \\ & N^{m_k}u_k, \dots, N^2u_k, Nu_k, u_k \,\} \end{aligned}$$

is a basis for V.

Remark. If we let $\mathcal{B}$ denote the (ordered) basis given in the above lemma, then $[N]_{\mathcal{B}}$ is in Jordan normal form. The blocks are of the form $J(0, m_i + 1)$.

Proof of Lemma 2.42. We use (strong) induction on the dimension of V. For the base case ($\dim(V) = 1$), let $u \in \ker(N) \backslash \{0\}$. Then $\{u\}$ is a basis for V and $Nu = 0$.

Now suppose that $\dim(V) \geqslant 2$ and that the lemma holds in all cases with lower dimension. Note that, by the rank-nullity theorem, we have

$$\dim(\mathrm{Im}(N)) = \dim(V) - \dim(\ker(N)) < \dim(V)$$

By the induction hypothesis, the result holds for the transformation $N|_{\mathrm{Im}(N)} : \mathrm{Im}(N) \to \mathrm{Im}(N)$. Let $u_1, \dots, u_k \in \mathrm{Im}(N)$ and $m_1, \dots, m_k \in \mathbb{N}$ be such that

$$\mathcal{A} = \{u_1, Nu_1, \dots, N^{m_1}u_1, u_2, Nu_2, \dots, N^{m_2}u_2, \dots, u_k, Nu_k, \dots, N^{m_k}u_k\}$$

is a basis for $\mathrm{Im}(N)$ and $N^{m_i+1}u_i = 0$ for all i. Choose $v_i \in V$ such that $Nv_i = u_i$ and define

$$\mathcal{B} = \mathcal{A} \cup \{v_1, \dots, v_k\}$$

The set $\mathcal{B}$ is linearly independent since

$$\begin{aligned} \sum_{i=1}^{k} \sum_{j=0}^{m_i+1} \alpha_{ij} N^j v_i = 0 &\implies \sum_{i=1}^{k} \sum_{j=0}^{m_i+1} \alpha_{ij} N^{j+1} v_i = 0 && \text{(applying } N \text{ to both sides)} \\ &\implies \sum_{i=1}^{k} \sum_{j=0}^{m_i+1} \alpha_{ij} N^j u_i = 0 && (u_i = Nv_i) \\ &\implies \sum_{i=1}^{k} \sum_{j=0}^{m_i} \alpha_{ij} N^j u_i = 0 && (N^{m_i+1}u_i = 0) \\ &\implies \alpha_{ij} = 0 \text{ for all } i \in \{1, \dots, k\} \text{ and } j \in \{0, \dots, m_i\} && (\mathcal{A} \text{ is linear independent}) \end{aligned}$$

which then also gives

$$\begin{aligned} & \sum_{i=1}^{k} \alpha_{i,m_i+1} N^{m+1} v_i = 0 \\ \implies & \sum_{i=1}^{k} \alpha_{i,m_i+1} N^m u_i = 0 \\ \implies & \alpha_{i,m+1} = 0 \text{ for all } i \in \{1, \dots, k\} && (\mathcal{A} \text{ is linear independent}) \end{aligned}$$

Having shown that $\mathcal{B}$ is linearly independent, we know that it can be extended to a basis of V. Let $\tilde{w}_i \in V$ be such

that $\mathcal{B} \cup \{\tilde{w}_1, \ldots, \tilde{w}_\ell\}$ is a basis of V. We have (for all i)

$$\begin{aligned} & N\tilde{w}_i \in \mathrm{Im}(V) \\ \implies & N\tilde{w}_i \in \mathrm{span}(\mathcal{A}) \\ \implies & N\tilde{w}_i \in \mathrm{span}(N(\mathcal{B})) && (N(\mathcal{B}) = \mathcal{A} \cup \{0\}) \\ \implies & N\tilde{w}_i = N\hat{w}_i && (\text{for some } \hat{w}_i \in \mathrm{span}(\mathcal{B})) \\ \implies & \tilde{w}_i - \hat{w}_i \in \ker(N) \end{aligned}$$

Letting $w_i = \tilde{w}_i - \hat{w}_i$ we have that

$$\mathcal{C} = \mathcal{B} \cup \{w_1, \ldots, w_\ell\}$$

is a basis for V and is of the desired form. □

Theorem 2.43: Jordan normal form

Let K be an algebraically closed field, V a finite dimensional K-vector space, and $f : V \to V$ a linear transformation. There exists a basis $\mathcal{B}$ of V such that

$$[f]_{\mathcal{B}} = \begin{bmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_N \end{bmatrix}$$

where each J_i is a Jordan block.

Proof. By Lemmas 2.25, 2.26, 2.27 it is enough to consider the case in which f has only one eigenvalue. (In the notation of the previous proof, it is enough to consider the transformation f_i.) So assume that we have a linear transformation $f : V \to V$ with minimal polynomial $(X - \lambda)^m$. Define $N : V \to V$ by $N = f - \lambda \,\mathrm{Id}$. Then N is nilpotent since $N^m = (f - \lambda \,\mathrm{Id})^m = 0$. Applying Lemma 2.42, there is a basis $\mathcal{B}$ of V such that $[N]_{\mathcal{B}}$ is in JNF and each Jordan block is of the form $J(0, n_i)$ for some $n_i \in \mathbb{N}$. Since $f = N + \lambda \,\mathrm{Id}$ we have that $[f]_{\mathcal{B}} = [N]_{\mathcal{B}} + \lambda I$ where I is the identity matrix of size $\dim(V)$. Therefore $[f]_{\mathcal{B}}$ is in JNF and has blocks of the form $J(\lambda, n_i)$. □

Remark. The JNF is unique up to rearrangement of the Jordan blocks. See Exercise 69. Two matrices in JNF are similar if and only if one can be obtained from the other by permuting the Jordan blocks.

The minimal and characteristic equation can be read from the Jordan normal form.

Exercise 59. Let K be an algebraically closed field, V a finite dimensional K-vector space, and $f : V \to V$ a linear transformation. Suppose that $\lambda \in K$ is an eigenvalue of f and let $m, n \in \mathbb{N}$ be maximal with the property that $(X - \lambda)^m$ divides that minimal polynomial and $(X - \lambda)^n$ divides the characteristic polynomial. Show that

(a) $m =$ the size of the largest Jordan block having λ on the diagonal

(b) $n =$ is the sum of the sizes of all Jordan blocks having λ on the diagonal

(c) the dimension of the λ-eigenspace is equal to the number of Jordan blocks having λ on the diagonal.

Example 2.44. Suppose that $A \in M_7(\mathbb{C})$ is similar to the matrix shown (which is in JNF). Then the characteristic and minimal polynomials of A are

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & i & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$c(X) = (X-2)^5(X-i)^2 \qquad m(X) = (X-2)^2(X-i)^2$$

The eigenspaces have dimensions

$$\dim(V_2) = 3 \qquad \dim(V_i) = 1$$

See Example 2.39 for another example.

Example 2.45. Find the Jordan normal form for

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

The characteristic polynomial is $c(X) = (X-1)^3$ so there is only one eigenvalue, $\lambda = 1$. Using row reduction, we find the corresponding eigenspace Nullspace$(A - I)$ has dimension 2. Thus the Jordan normal form J has 2 blocks, hence

$$J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Remark. For square matrices of size 2 or 3, the JNF can be determined from the minimal and characteristic polynomials. However, this is not true for larger matrices

Example 2.46. The following two matrices (both in JNF) are not similar.

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

However both have $m(X) = (X-2)^3$, $c(X) = (X-2)^7$ and $\dim(V_2) = 3$. One way to see that they are not similar is to note that $\dim(\ker(A-2I)^2) = 5$ but $\dim(\ker(B-2I)^2) = 6$

5.1 Exercises

Exercise 60. Show that the linear transformation $\mathcal{P}_n(\mathbb{R}) \to \mathcal{P}_n(\mathbb{R})$ given by differentiation cannot be represented by a diagonal matrix.

Exercise 61. If f is a linear transformation on a finite dimensional vector space V satisfying $f^2 = f$, explain how to find a diagonal matrix representing f.

Exercise 62. Suppose that linear transformations f and g on a vector space V commute; that is, that $fg = gf$. Show that an eigenspace of f will be g-invariant. If the field F of scalars is algebraically closed and V is finite dimensional, deduce that f and g have a common eigenvector.

Exercise 63. Find the Jordan normal form of the following matrices:

$$\begin{bmatrix} -1 & 1 \\ -1 & -3 \end{bmatrix}, \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 0 \\ 2 & 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}.$$

Exercise 64. For each of the following pairs of minimal and characteristic polynomials, find all possibilities for the Jordan normal form:

Minimal polynomial	Characteristic polynomial
$X^2(X+1)^2$	$X^2(X+1)^4$
$(X-3)^2$	$(X-3)^5$
X^3	X^7
$(X-1)^2(X+1)^2$	$(X-1)^4(X+1)^4$.

Exercise 65. Which of the following pairs of matrices (over $\mathbb{C}$) are similar?

(a) $\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 5 \\ 0 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

Exercise 66. Given a 4×4 matrix A over $\mathbb{C}$ and given the minimal and characteristic polynomials of A, describe the possibilities for the JNF of A. (There will be one case where there are two possibilities.)

Exercise 67. Show that any JNF matrix J is a sum $J = D + N$ where D is diagonal and N is nilpotent; that is $N^k = 0$ for some k. Deduce that any linear transformation f of a finite dimensional complex vector space can be written in the form $f = d + n$ where d is diagonalisable and n is nilpotent.

Exercise 68. In the language of the previous question, show that $JN = NJ$ and $JD = DJ$. Deduce that $fd = df$ and $fn = nf$.

Exercise 69. (Harder) Show that the Jordan normal form of a complex matrix A is completely determined by the dimensions of the nullspaces of $(A - \lambda I)^i$, $i = 1, 2, 3, \ldots$ for all the eigenvalues λ of A.