Operators

July 16, 2017

These notes are a work in progress, and I may modify them as we go along (the date above indicates when they were last updated). Please send me any corrections/typos by email.

Let V be a vector space over a field $\mathbb{F}$. The aim of this chapter is to study linear maps $V \to V$ - that is, elements of $\hom(V, V)$.

Definition 0.1: *An* ***operator*** *(or linear operator) on V is a linear map $f : V \to V$.*

1 A motivating example (and some definitions along the way)

1.1 Direct sums

Definition 1.1: *Vector subspaces U_1, U_2 of a vector space V are said to be* ***complementary*** *if any vector $v \in V$ can be written uniquely as a sum $v = v_1 + v_2$ with $v_1 \in U_1$ and $v_2 \in U_2$ (in particular $V = U_1 + U_2$). We then say that V is the* ***direct sum*** *of U_1 and U_2, and we write $V = U_1 \oplus U_2$.*

Example 1.2: $V = \mathbb{R}^2$ and U_1, U_2 are two distinct lines through the origin. $V = \mathbb{R}^3$, U_1 is a plane through the origin, and U_2 a line not contained in U_1.

Remark 1.3: *If $v \in U_1 \cap U_2$, then we must have $v = 0$ otherwise $v = 0 + v$ and $v = v + 0$ are two distinct ways of writing v. If V is finite dimensional we thus get $\dim V = \dim U_1 + \dim U_2$.*

Some more complicated examples.

Example 1.4:

1. Let V be the space of functions $\mathbb{R} \to \mathbb{R}$. Let U_+ be the vector subspace of even functions (i.e. functions $h : \mathbb{R} \to \mathbb{R}$ such that for all $x \in \mathbb{R}$, $h(-x) = h(x)$), and let U_- be the vector subspace of odd functions (i.e. functions $h : \mathbb{R} \to \mathbb{R}$ such that for all $x \in \mathbb{R}$, $h(-x) = -h(x)$). Then for any $h \in V$, we can define $h_+(x) = (h(x) + h(-x))/2$ and $h_-(x) = (h(x) - h(-x))/2$ and see that $h_+ \in U_+, h_- \in U_-$ and $h = h_+ + h_-$. This decomposition is in fact unique, so we have $V = U_+ \oplus U_-$.

2. Let $V = M_n(\mathbb{R})$ be the set of $n \times n$ matrices with coefficients in $\mathbb{R}$. Let

$$U_1 = \{A \in M_n(\mathbb{R}) \mid A^t = A\}$$

be the subspace of symmetric matrices, and let

$$U_2 = \{A \in M_n(\mathbb{R}) \mid A^t = -A\}$$

be the subspace of antisymmetric matrices. Then for any matrix $A \in M_n(\mathbb{R})$, we can write $A = A_1 + A_2$ with $A_1 = (A + A^t)/2 \in U_1$ and $A_2 = (A - A^t)/2 \in U_2$, and it is possible to see that this decomposition is unique.

1.2 Projections and reflections

When we have written V as a direct sum $U_1 \oplus U_2$, there are some natural examples of operators arising.

Definition 1.5: *The **projection from** V **to** U_1 **parallel to** U_2 is the map $p_1 : V \to V$ defined by $p_1(v) = v_1$, where $v = v_1 + v_2$ is the unique decomposition of v given by the direct sum. It is an operator on V. Similarly we can define the projection onto U_2 parallel to U_1, $p_2 : V \to V$.*

*The **reflection in** U_1 **parallel to** U_2 is the map $s_1 : V \to V$ defined by $s_1(v) = v_1 - v_2$, where $v = v_1 + v_2$ is the unique decomposition of v given above.*

It is not hard to see that projections and reflections are linear operators.

Example 1.6:

1. In $\mathbb{R}^2$ - see picture drawn in class.
2. If V is the space of functions $\mathbb{R} \to \mathbb{R}$, we get that $p_+(h)$ is defined by $p_+(h)(x) = (h(x)+h(-x))/2$ and $p_-(h)$ is defined by $p_-(h)(x) = (h(x) - h(-x))/2$. How about the reflection s in U_+ parallel to U_-? for any $h \in V$, we have $s(h) = h_+ - h_- = p_+(h) - p_-(h)$ so

$$s(h)(x) = \frac{h(x) + h(-x)}{2} - \frac{h(x) - h(-x)}{2} = h(-x)$$

1.3 Operations on operators

Note that if $f, g \in \hom(V, V)$ then the sum $f + g$ is also in $\hom(V, V)$, and the composition $f \circ g$ as well (this is one of the advantages when working with linear maps from the space V to itself, we can compose them easily, and even take powers $f^k = f \circ f \circ f \circ \ldots \circ f$).

Let us see what happens when we take sums and compositions of the projections and retractions defined above.

Example 1.7: Let $V = U_1 \oplus U_2$, and denote by p_1, p_2 the projections onto U_1, U_2 parallel to U_2, U_1 respectively. Let s denote the reflection in U_1 parallel to U_2.

1. $(p_1)^2 = p_1 \circ p_1 = p_1$ and $p_2 \circ p_2 = p_2$;
2. $p_1 \circ p_2$ is the zero operator which sends every vector to zero.
3. $p_1 + p_2 = \mathrm{id}$, the identity operator which sends every vector to itself;
4. $p_1 - p_2 = s$

Question 1: *The zero operator and the identity operator exist on any vector space V. What happens when you compose them with another operator f? What does this remind you of?*

1.4 Matrix representation

Now for each choice of basis for V, we can write a matrix representation for an operator $f : V \to V$.

Example 1.8: (See picture in class). Let $\mathcal{B}$ be the basis (b_1, b_2). We see that the decomposition of b_1 is $b_1 = b_1 + 0$, and that of b_2 is $b_2 = -b_1/2 + b_2'$ so we see that $s(b_1) = b_1$ and $s(b_2) = -b_1/2 - b_2' = -b_1 - b_2$. Thus the matrix of s with respect to $\mathcal{B}$ is

$$[s]_{\mathcal{B}} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$$

On the other hand, if $\mathcal{D}$ is the basis (u_1, u_2), we see that $s(u_1) = u_1$ and $s(u_2) = -u_2$ so

$$[s]_{\mathcal{D}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

In this basis it is much easier to compute the matrix.

This will be one of the main questions we will try to answer - how can I find a basis in which the matrix representing a given operator looks as simple as possible?

Another idea to take from this - the distinction between the operator - defined purely geometrically, and its matrix representation in a basis (which depends on the choice of basis).

1.5 Change of basis

We could also have obtained the matrix representing s with respect to $\mathcal{B}$ from the matrix with respect to $\mathcal{D}$ using the change of basis formula.

The matrix that helps us "translate" the coordinates of a vector v with respect to $\mathcal{B}$ to its coordinates with respect to $\mathcal{D}$ is the matrix whose columns are the coordinates of (b_1, b_2) with respect to $\mathcal{D}$, namely

$$M_{\mathcal{D}}^{\mathcal{B}} = [\mathrm{id}]_{\mathcal{D}}^{\mathcal{B}} = \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \end{bmatrix}$$

It satisfies $M_{\mathcal{D}}^{\mathcal{B}}[v]_{\mathcal{B}} = [v]_{\mathcal{D}}$. To "translate" from coordinates in $\mathcal{D}$ to coordinates in $\mathcal{B}$ we use the inverse matrix

$$M_{\mathcal{B}}^{\mathcal{D}} = [\mathrm{id}]_{\mathcal{B}}^{\mathcal{D}} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

Thus we get

$$[s]_{\mathcal{B}} = M_{\mathcal{B}}^{\mathcal{D}}\ [s]_{\mathcal{D}}\ M_{\mathcal{D}}^{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$$

2 Invariant subspaces, eigenvalues and eigenvectors

2.1 Invariant subspaces

Definition 2.1: *(Invariant subspace) Let $f : V \to V$ be a linear operator. A vector subspace $U \subseteq V$ is called an invariant subspace for f, or an f-invariant subspace, if for any $u \in U$, we have $f(u) \in U$ (equivalently, if $f(U) \subseteq U$).*

The zero subspace $\{0\}$, and the full subspace V are always f-invariant.

Example 2.2:

1. Let V be a 2-dimensional vector space over $\mathbb{R}$, U_1 and U_2 complementary one-dimensional subspaces and let s be the reflection in U_1 as in the previous section. Then U_1 is an invariant subspace of s, and so is U_2. Note that all the vectors of U_1 are fixed by s, while those of U_2 (except 0) are moved - but their image is still in U_2.
2. A rotation of $\mathbb{R}^2$ of angle θ with $0 < \theta < 2\pi$ has no nontrivial invariant subspaces. For a rotation of $\mathbb{R}^3$ along the z-axis, both the xy-plane and the z-axis are invariant subspaces.
3. $V = C^\infty(\mathbb{R})$ the space of infinitely differentiable functions defined on $\mathbb{R}$. Let $D : V \to V$ be the differentiation operator, defined by $D(f) = f'$. Then the set of all polynomial functions $\{h : \mathbb{R} \to \mathbb{R} \mid h(x) = a_0 + a_1 x + \ldots + a_k x^k \text{ for all } k \in \mathbb{N} \text{ and } a_0, \ldots, a_k \in \mathbb{R}\}$ is a vector subspace and it is D-invariant.

Remark 2.3: *If U is an f-invariant subspace, then $f\,|_U$ defines a linear operator $U \to U$. If U, W are complementary subspaces which are both f-invariant, then f is uniquely defined once we know $f\,|_U$ and $f\,|_W$: indeed, any vector $v \in V$ can be written uniquely as $v = u + w$ with $u \in U$ and $w \in W$, and then $f(v) = f(u) + f(w)$.*

Idea: to understand a map, it helps to know its invariant subspaces because then we can restrict our attention to operators on lower dimensional vector spaces.

"Divide and conquer"...in particular, when the vector space V is finite dimensional, we will be able to choose a basis in which the matrix representing f looks simple.

2.2 Cyclic subspaces

How can we find a nontrivial f-invariant vector subspace? Let v be a nonzero vector. Any f-invariant vector subspace which contains v must contain $f(v)$, thus also $f^2(v) = f \circ f(v) = f(f(v)), f^3(v) = f \circ f \circ f(v), \ldots$ as well as all their linear combinations.

Definition 2.4: *Let f be an operator on V, and let $v \in V$. The* ***cyclic subspace*** *associated to f and v is*

$$Z(f, v) = \operatorname{Span}(v, f(v), f^2(v), f^3(v), \ldots) = \{a_0 v + a_1 f(v) + \ldots + a_k f^k(v) \mid a_0, \ldots, a_k \in \mathbb{F}, k \in \mathbb{N}\}.$$

Proposition 2.5: *The cyclic subspace $Z(f, v)$ is an f-invariant subspace.*

Proof. A vector u in $Z(f, v)$ takes the form $u = a_0 v + a_1 f(v) + \ldots + a_k f^k(v)$ for some scalars $a_0, \ldots, a_k$. Thus

$$f(u) = f(a_0 v + a_1 f(v) + \ldots + a_k f^k(v)) = a_0 f(v) + a_1 f^2(v) + \ldots + a_k f^{k+1}(v)$$

which is also in $Z(f, v)$. □

The vectors $v, f(v), f^2(v), \ldots$ generate $Z(f, v)$. If V is finite dimensional, there exists an index k (which we may assume to be minimal) such that $v, f(v), \ldots, f^{k+1}(v)$ are linearly dependent. In other words, there exist scalars $a_0, \ldots, a_{k+1}$ such that

$$a_0 v + a_1 f(v) + \ldots + a_{k+1} f^{k+1}(v) = 0$$

By minimality of k, $a_{k+1} \neq 0$ so we may assume $a_{k+1} = 1$ (if not, divide everything by a_{k+1}) so we have

$$f^{k+1}(v) = -a_0 v - \ldots - a_k f^k(v)$$

Definition 2.6: *The polynomial P given by $P(X) = a_0 + a_1 X + \ldots + a_k X^k + X^{k+1}$ is called the polynomial associated to the cyclic subspace $Z(f, v)$.*

In fact we have

Proposition 2.7: *Let $f : V \to V$ be an operator on a finite dimensional vector space, and let v be a non zero vector in V. If k is the smallest integer such that $v, f(v), \ldots, f^{k+1}(v)$ are linearly dependent, then $v, f(v), \ldots, f^k(v)$ form a basis for $Z(f, v)$.*

Proof. Let $m \geq k$, we show by induction that $f^m(v) \in \operatorname{Span}(v, f(v), \ldots, f^k(v))$. We have

$$\begin{aligned} f^m(v) &= f^{m-k-1}(f^{k+1}(v)) \\ &= f^{m-k-1}(-a_0 v + \ldots + a_k f^k(v)) \\ &= -a_0 f^{m-k-1}(v) - \ldots - a_k f^{m-1}(v) \end{aligned}$$

which is in $\operatorname{Span}(v, f(v), \ldots, f^k(v))$ by induction hypothesis. □

2.3 Eigenvalues and eigenvectors

A particular case which will be extremely important is when the invariant subspace has dimension 1 - then $f\mid_U$ is a linear map on a line, that is, a multiplication by a scalar.

Definition 2.8: *A scalar $\lambda \in \mathbb{F}$ is called an* ***eigenvalue*** *of f if there exists a* ***nonzero*** *vector $v \in V$ such that $f(v) = \lambda v$. The vector v is called an* ***eigenvector*** *associated to λ.*

Remark 2.9: *If v is an eigenvector, then* $\operatorname{Span}(v)$ *is an invariant subspace for f.*

Example 2.10:
1. $\mathbb{R}^2 = U_1 \oplus U_2$: any vector on U_1 is an eigenvector with eigenvalue 1, any vector on U_2 is an eigenvector with eigenvalue -1. Same for the plane and the line for reflexion in $\mathbb{R}^3$.

2. Dilation (also called homotethy): $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(v) = \lambda v$ for all v. Then λ is an eigenvalue, any nonzero vector is an eigenvector associated to λ.

3. $V = C^\infty(\mathbb{R})$ space of infinitely differentiable functions defined on $\mathbb{R}$. Let $D : V \to V$ be the differentiation operator, defined by $D(h) = h'$. For any $\lambda \in \mathbb{R}$, the function ϕ_λ defined by $\phi_\lambda(t) = e^{\lambda t}$ is an eigenvector of D associated to the eigenvalue λ.

An especially important example

Example 2.11: $V = \mathbb{F}^n_{col}$, consider the operator f_A defined by a matrix $A = [a^i_j]$, that is

$$f_A : \mathbb{F}^n_{col} \to \mathbb{F}^n_{col}$$
$$\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} \mapsto A \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}$$

Then $x = \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}$ is an eigenvector for f_A associated to the eigenvalue λ iff we have $Ax = \lambda x$

Definition 2.12: *(eigenvalue of a matrix) If there exists a non zero vector $x \in \mathbb{F}^n_{col}$ such that $Ax = \lambda x$, we say that λ is an **eigenvalue of the matrix** A, with associated vector x.*

Proposition 2.13: *Let $A \in M_n(\mathbb{F})$ be a square matrix, and let $\lambda \in \mathbb{F}$. The following are equivalent:*

1. *λ is an eigenvalue of A;*
2. $\mathrm{Ker}(\lambda \mathrm{id} - f_A)$ *is non trivial;*
3. $\mathrm{Im}(\lambda \mathrm{id} - f_A)$ *is not all of* $\mathbb{F}^n_{col}$;
4. $(\lambda I - A)$ *is not invertible;*
5. $\det(\lambda I - A) = 0$.

Proof. ($1 \iff 2$) Note that $x \in \mathrm{Ker}(\lambda \mathrm{id} - f_A)$ iff $(\lambda I - A)x = 0$ iff $\lambda x - Ax = 0$ iff $Ax = \lambda x$. Now λ is an eigenvalue iff there exists a non trivial vector which satisfies this, that is, iff $\mathrm{Ker}(\lambda \mathrm{id} - f_A)$ is not trivial.

You saw in Linear Algebra 1 that for an operator f on a finite dimensional space represented by a matrix M, we have that $\mathrm{Ker} f$ is non trivial iff $\mathrm{Im} f$ is not all of $\mathbb{F}^n_{col}$ iff M is not invertible iff $\det(M) = 0$. This gives the equivalence of 2-3-4-5. □

An application of eigenvectors: Google's PageRank algorithm. We want to give each web page a score measuring its importance. Idea: how many links are there pointing to it? And what are the scores of the pages that link to it?

If x_i is the score of page number i, we want:

$$x_i = \sum_{j=1}^{n} \frac{\text{links } j \to i}{\text{links } j \to \text{ anywhere}} x_j$$

so if we define A to be the matrix with entries $a^j_i = \dfrac{\text{links } j \to i}{\text{links } j \to \text{ anywhere}}$ this becomes $x = Ax$, that is, the vector of "scores" is an eigenvector of the matrix A corresponding to the eigenvalue 1.

(One can show that there always exists such a vector. Problem - there might be several linearly independent such vectors, in this case how do we choose) The original PageRank algorithm is almost as simple as this - the matrix A needs to be modified a little to deal with the fact that there may be several solutions.

To go further: The $25,000,000,000$\$ eigenvector - The linear algebra behind Google, by Kurt Bryan and Tanya Leise

2.4 Eigenspaces. Linear independence of eigenspaces

If v is an eigenvector associated to λ, then for any nontrivial scalar α the vector αv is also an eigenvector for λ (Check!). But there might be other eigenvectors associated to the same eigenvalue.

Definition 2.14: *Let λ be an eigenvalue for f. The eigenspace V_λ associated to λ is the set*

$$V_\lambda = \{v \in V \mid f(v) = \lambda v\}$$

Remark 2.15:
- *Note that the zero vector lies in V_λ. If $v \neq 0$, then v is in V_λ iff it is an eigenvector associated to λ. Thus V_λ is the set of all eigenvectors associated to λ, together with the zero vector.*
- $V_\lambda = \ker(\lambda \mathrm{id} - f)$ *indeed, $v \in V_\lambda$ iff $f(v) = \lambda v$ iff $(f - \lambda \mathrm{id})(v) = 0$. In particular V_λ is a vector subspace of V. It contains at least one non-zero vector, hence it has dimension at least* 1.

Example 2.16:
1. Reflexion in a plane in $\mathbb{R}^3$ - the plane is the eigenspace associated to 1.
2. Rotation in $\mathbb{R}^3$ about the z-axis. The xy-plane is an invariant subspace but NOT an eigenspace.
3. Homothety: the whole space is an eigenspace.

Definition 2.17: *The **geometric multiplicity** of an eigenvalue λ of f is the dimension* $\dim(V_\lambda)$ *of the corresponding eigenspace.*

We now show that eigenvectors corresponding to distinct eigenvalues are linearly independent. In fact, we show that the **eigenspaces** corresponding to distinct eigenvalues are linearly independent.

Definition 2.18: *We say that a family $U_1, \ldots, U_k$ of vector subspaces of V is linearly independent if for any choice of vectors $u_1 \in U_1, \ldots, u_k \in U_k$, we have*

$$u_1 + \ldots + u_k = 0 \iff u_1 = 0, \ldots, u_k = 0$$

Remark 2.19:
- *If $U_1, \ldots, U_k$ are linearly independent, and for each i, u_i is a nonzero vector in U_i, then the vectors $u_1, \ldots, u_k$ are linearly independent. Indeed, for any scalar a_i the vector $a_i u_i$ is in U_i, hence if $a_1 u_1 + \ldots + a_k u_k = 0$ we must have $a_i u_i = 0$ for all i. Since $u_i \neq 0$, we have $a_i = 0$ for all i.*
- *A family of nonzero vectors $v_1, \ldots, v_k$ is linearly independent $\iff$ the corresponding family of subspaces* $\mathrm{Span}(v_1), \ldots, \mathrm{Span}(v_k)$ *is linearly independent.*

Example 2.20: In $\mathbb{R}^3$: two distinct lines in $\mathbb{R}^3$ are linearly independent. A line and a plane not containing it are linearly independent. Two planes are never linearly independent. Three distinct lines are linearly independent iff they are not contained in a common plane.

Proposition 2.21: *Let $f : V \to V$ be an operator over a vector space V. Let $V_{\lambda_1}, \ldots, V_{\lambda_m}$ be eigenspaces associated to eigenvalues $\lambda_1, \ldots, \lambda_m$ respectively, and assume that for any $i, j \in \{1, \ldots, m\}$ with $i \neq j$ we have $\lambda_i \neq \lambda_j$.*

Then $U_1, \ldots, U_m$ are linearly independent.

From the second part of Remark 2.19 and Proposition 2.21 we deduce the following

Corollary 2.22: *If $v_1, \ldots, v_k$ are eigenvectors associated to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then they are linearly independent.*

Remark 2.23: *This implies in particular that if V has finite dimension n, then a linear operator $f : V \to V$ has at most n distinct eigenvalues.*

Proof. (of Proposition 2.21) Let $f : V \to V$ be an operator. Let $V_{\lambda_1}, \ldots, V_{\lambda_m}$ be eigenspaces associated to eigenvalues $\lambda_1, \ldots, \lambda_m$. We prove by induction on m that they are linearly independent. For $m = 1$, this is trivial.

Suppose that $V_{\lambda_1}, \ldots, V_{\lambda_{m-1}}$, and that $u_1, \ldots, u_m$ are vectors in $V_{\lambda_1}, \ldots, V_{\lambda_m}$ respectively such that

$$u_1 + \ldots + u_m = 0 \tag{1}$$

We want to show $u_1 = \ldots = u_m = 0$.

Note that $f(u_1 + \ldots + u_m) = 0$, on the other hand

$$\begin{aligned} f(u_1 + \ldots + u_m) &= f(u_1) + \ldots + f(u_m) \\ &= \lambda_1 u_1 + \ldots + \lambda_m u_m \end{aligned}$$

so we get

$$\lambda_1 u_1 + \ldots + \lambda_m u_m = 0 \tag{2}$$

By multiplying equation (1) by λ_m, and substracting from it equation (2), the last term in u_m disappears and we get

$$(\lambda_m - \lambda_1)u_1 + \ldots + (\lambda_m - \lambda_{m-1})u_{m-1} = 0$$

By linear independence of $V_{\lambda_1}, \ldots, V_{\lambda_{m-1}}$, for each i we have $(\lambda_m - \lambda_i)u_i = 0$. Since $\lambda_i \neq \lambda_m$ for all $i < m$, we must have $u_1 = \ldots = u_{m-1} = 0$. Hence $u_m = 0$ as well. $\square$

3 Matrix of an operator

In the finite dimensional case we will want to try and see what is the "best basis" to choose to express a given linear operator as a matrix.

So first - some preliminaries about the matrices representing an operator in different bases. For example, what is the relation between two representations of a same linear operator in two different bases?

Suppose V is finite dimensional of dimension n, and let $f : V \to V$ be an operator.

Matrix of an operator with respect to a basis $\mathcal{B}$

What does it mean for a matrix A to represent f in the basis $\mathcal{B} = (b_1, \ldots, b_m)$?

For any vector $v \in V$, v can be represented in $\mathcal{B}$ by the column vector $[v]_{\mathcal{B}}$ of its coordinates in $\mathcal{B}$, that is

$$[v]_{\mathcal{B}} = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix} \text{ if and only if } v = \sum_{j=1}^{n} v^j b_j$$

Then $A = [f]_{\mathcal{B}}$ is the matrix such that for any $v \in V$ we have

$$[f(v)]_{\mathcal{B}} = A[v]_{\mathcal{B}}.$$

In particular, if $v = b_i$, then $[b_i]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ and we see that the i-th column $\begin{bmatrix} a_i^1 \\ \vdots \\ a_i^n \end{bmatrix}$ of A is exactly $[f(b_i)]_{\mathcal{B}}$.

Different matrices of a same operator - similarity of matrices

Let $\mathcal{B} = (b_1, \ldots, b_n)$ and $\mathcal{D} = (d_1, \ldots, d_n)$ be two ordered bases of V.

Recall that the change of matrix basis $M_{\mathcal{D}}^{\mathcal{B}}$ is the matrix such that for any vector $v \in V$, we have $[v]_{\mathcal{D}} = M_{\mathcal{D}}^{\mathcal{B}}[v]_{\mathcal{B}}$ (note that $M_{\mathcal{D}}^{\mathcal{B}}$ is exactly the matrix $[id]_{\mathcal{D}}^{\mathcal{B}}$ which represents the identity with respect to the bases $\mathcal{B}$ (at the source) and $\mathcal{D}$ (at the range)).

The matrix $M_{\mathcal{D}}^{\mathcal{B}}$ is regular (it represents the identity, which is invertible), and we have $M_{\mathcal{B}}^{\mathcal{D}} = (M_{\mathcal{D}}^{\mathcal{B}})^{-1}$. We then have

$$[f]_{\mathcal{D}} = M_{\mathcal{D}}^{\mathcal{B}}\, [f]_{\mathcal{B}}\, M_{\mathcal{B}}^{\mathcal{D}} = M_{\mathcal{D}}^{\mathcal{B}}\, [f]_{\mathcal{B}}\, (M_{\mathcal{D}}^{\mathcal{B}})^{-1}$$

This motivates the following definition: let $A, B \in M_n(\mathbb{F})$.

Definition 3.1: *We say that A and B are* ***similar*** *if there exists a regular (invertible) matrix $M \in M_n(\mathbb{F})$ such that $B = MAM^{-1}$.*

Thus if two matrices A, B represent the same linear operator, only with respect to different bases, they are similar.

Matrix in a basis adapted to an invariant subspace

Suppose now that U is an f-invariant subspace of V. Let $\mathcal{B}_U = (b_1, \ldots, b_m)$ be a basis for U. We extend it to a basis $\mathcal{B} = (b_1, \ldots, b_n)$ for V. What does the matrix of f look like in $\mathcal{B}$? note that if $i \leq m$, then $b_i \in U$ so $f(b_i) \in U$ so its column vector of coordinates with respect to the basis $\mathcal{B}$ is of the form:

$$[f(b_i)]_{\mathcal{B}} = \begin{bmatrix} v^1 \\ \vdots \\ v^m \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and note that } \begin{bmatrix} v^1 \\ \vdots \\ v^m \end{bmatrix} = [f(b_i)]_{\mathcal{B}_u}$$

Hence the matrix representing f is of the form:

$$\begin{bmatrix} [f(b_1)]_{\mathcal{B}} & \ldots & [f(b_m)]_{\mathcal{B}} & [f(b_{m+1})]_{\mathcal{B}} & \ldots & [f(b_n)]_{\mathcal{B}} \end{bmatrix} = \left[\begin{array}{c|c} [f\mid_U]_{\mathcal{B}_U} & A \\ \hline 0 & B \end{array}\right]$$

Suppose now that we have in addition another f-invariant subspaces W such that U, W are complementary. We will prove later in the course the following fact about complementary subspaces: if we choose a basis $\mathcal{B}_U = (b_1, \ldots, b_m)$ for U, and a basis $\mathcal{B}_W = (b_{m+1}, \ldots, b_n)$ for W, then $\mathcal{B} = (b_1, \ldots, b_m)$ is a basis for V.

Then, we know that for each $i > m$ we have $f(b_i) \in W$ so

$$[f(b_i)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ v^{m+1} \\ \vdots \\ v^n \end{bmatrix}$$

Hence the matrix representing f is of the form:

$$\begin{bmatrix} [f(b_1)]_{\mathcal{B}} & \ldots & [f(b_n)]_{\mathcal{B}} \end{bmatrix} = \left[\begin{array}{c|c} [f\mid_U]_{\mathcal{B}_U} & 0 \\ \hline 0 & [f\mid_W]_{\mathcal{B}_W} \end{array}\right]$$

We say the matrix is **diagonal by blocks**.

Diagonalization

The best possible case is if we manage to break down the space into invariant subspaces of dimension one - in that case the matrix would look like this:

$$\begin{bmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{bmatrix}$$

We say the matrix is **diagonal**. It is not always possible to find a basis in which the matrix representing f is diagonal, but we will give sufficient conditions in the sequel.

Matrix of an operator restricted to a cyclic subspace

Given an operator $f : V \to V$ let us pick a nonzero vector v and look at the corresponding cyclic subspace $Z(f, v)$. We saw it is an invariant subspace, and that if k is the largest index for which $v, f(v), \dots, f^k(v)$ are linearly independent, then $\mathcal{B} = (v, f(v), \dots, f^k(v))$ forms a basis.

[Recall that the polynomial associated to $Z(f, v)$ is $P(X) = a_0 + a_1X + \dots + a_kX^k + X^{k+1}$ such that $a_0v + a_1f(v) + \dots + a_kf^k(v) + f^{k+1}(v) = 0$.]

What is the matrix of $f\,|_{Z(f,v)}$ with respect to the basis $\mathcal{B}$?

Write $b_0 = v, b_1 = f(v), \dots, b_k = f^k(v)$. For each $i < k$ we have $f(b_i) = b_{i+1}$, and

$$f(b_k) = f(f^k(v)) = f^{k+1}(v) = -a_0v - \dots - a_kf^k(v) = -a_0b_0 - a_1b_1 - \dots - a_kb_k$$

thus the matrix of f with respect to $\mathcal{B}$ is

$$[f]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ 0 & 0 & 1 & \dots & 0 & -a_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -a_k \end{bmatrix}$$

Definition 3.2: *This matrix is called the **companion matrix** of the polynomial P.*

4 Diagonalizing

Let V be a finite dimensional vector space, of dimension n.

Definition 4.1: *A matrix $A \in M_n(\mathbb{F})$ is said to be **diagonal** if all its non diagonal entries are 0, that is, if it is of the form*

$$\operatorname{diag}(\alpha_1, \dots, \alpha_n) := \begin{bmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{bmatrix}$$

Definition 4.2: *A linear operator $f : V \to V$ is said to be **diagonalizable** if there exists a basis $\mathcal{B}$ for V such that $[f]_{\mathcal{B}}$ is diagonal.*

Example 4.3: Let A be an $n \times n$ matrix over $\mathbb{F}$. Let f_A be the linear operator on $\mathbb{F}^n_{col}$ defined by $x \mapsto Ax$. If $\mathcal{B}$ denotes the standard basis, we have precisely that $[f]_{\mathcal{B}} = A$.

By definition f_A is diagonalizable iff there exists a basis $\mathcal{D}$ such that $[f]_{\mathcal{D}}$ is diagonal. By what we saw above, this is equivalent to saying that there exists an invertible matrix M such that MAM^{-1} is diagonal.

Definition 4.4: *We say that a* ***matrix*** $A \in M_n(\mathbb{F})$ *is* ***diagonalizable*** *over* $\mathbb{F}$ *if it is similar to a diagonal matrix, that is, if there exists an invertible matrix* M *of* $M_n(\mathbb{F})$ *such that* MAM^{-1} *is diagonal.*

4.1 A first characterization

If $[f]_{\mathcal{B}}$ is a diagonal matrix $A = \operatorname{diag}(\lambda_1 \dots \lambda_n)$, what does it say about the basis $\mathcal{B} = (b_1, \dots, b_n)$? We have $[b_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, so we get $[f(b_1)]_{\mathcal{B}} = A \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = [\lambda_1 b_1]_{\mathcal{B}}$. Hence b_1 is an eigenvector associated to the eigenvalue λ_1. Similarly, each b_i is an eigenvector associated to the eigenvalue λ_i.

Proposition 4.5: *f is diagonalizable $\iff$ there exists a basis of V which consists of eigenvectors of f.*

Proof. This is essentially proved by the remark above - if $\mathcal{B} = (b_1, \dots, b_n)$ is an ordered basis such that $[f]_{\mathcal{B}} = \operatorname{diag}(\alpha_1, \dots, \alpha_n)$. Then by definition of $[f]_{\mathcal{B}}$ we have

$$[f(b_i)]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & & & & \\ & \ddots & & & \\ & & \alpha_i & & \\ & & & \ddots & \\ & & & & \alpha_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \alpha_i \\ \vdots \\ 0 \end{bmatrix} = [\alpha_i b_i]_{\mathcal{B}}$$

Thus b_i is an eigenvector associated to the eigenvalue α_i.

For the other direction, suppose that there exists eigenvectors $v_1, \dots, v_n$ associated to eigenvalues $\lambda_1, \dots, \lambda_n$ respectively such that $(v_1, \dots, v_n)$ is an ordered basis $\mathcal{D}$ for V. Then $[f]_{\mathcal{D}}$ is of the form $\begin{bmatrix} [f(v_1)]_{\mathcal{D}} & \vdots & [f(v_n)]_{\mathcal{D}} \end{bmatrix}$.

Now $[f(v_1)]_{\mathcal{D}} = [\lambda_1 v_1]_{\mathcal{D}} = \begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, and similarly for all the other indices i, so we get

$$[f]_{\mathcal{D}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

□

We deduce a useful sufficient condition for diagonalization:

Proposition 4.6: *Let f be a linear operator over an n-dimensional vector space V. If f admits n distinct eigenvalues then f is diagonalizable.*

Proof. Choose $v_1, \dots, v_n$ eigenvectors associated to $\lambda_1, \dots, \lambda_n$ respectively. Then by Proposition 2.22, the family $v_1, \dots, v_n$ is linearly independent - since $\dim V = n$, it is in fact a basis of eigenvectors. □

Note that this is not an "iff" condition. For example, the operator $f : \mathbb{R}^2_{col} \to \mathbb{R}^2_{col}$ defined by $f(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$ has only one eigenvalue, namely $\lambda = 2$. On the other hand, it is represented in the standard basis (and in fact, in any basis) by the matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Thus, f is diagonalizable, though it has only one eigenvalue.

Example 4.7: Consider the matrix $A = \begin{bmatrix} 5 & -3 \\ 6 & -4 \end{bmatrix}$. Let us try to see whether it is diagonalizable (equivalently, whether the operator $f_A : \mathbb{R}^2_{col} \to \mathbb{R}^2_{col}$ defined by $f_A(x) = Ax$ is diagonalizable).

Now λ is an eigenvalue iff there exists a nonzero vector x such that $Ax = \lambda x$, i.e $(\lambda I - A)x = 0$. In other words, λ is an eigenvalue iff the matrix $\lambda I - A$ has non trivial kernel (this is exactly what Proposition 2.13 says). Now

$$\lambda I - A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 5 & -3 \\ 6 & -4 \end{bmatrix} = \begin{bmatrix} \lambda - 5 & 3 \\ -6 & \lambda + 4 \end{bmatrix}$$

This matrix has nontrivial kernel iff its determinant is zero, that is iff

$$\det(\lambda I - A) = (\lambda - 5)(\lambda + 4) + 18 = \lambda^2 - \lambda - 2 = (\lambda + 1)(\lambda - 2)$$

Thus the eigenvalues of A are $-1, 2$. Because we have $2 = \dim(V)$ distinct eigenvalues, this means by Proposition 4.6 that f_A (and thus A) is diagonalizable.

Let us find eigenvectors: we have

$$Ax = -x \iff \begin{bmatrix} 5 & -3 \\ 6 & -4 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} -x^1 \\ -x^2 \end{bmatrix} \iff \begin{cases} 5x^1 - 3x^2 = -x^1 \\ 6x^1 - 4x^2 = -x^2 \end{cases} \iff \begin{cases} 6x^1 - 3x^2 = 0 \\ 6x^1 - 3x^2 = 0 \end{cases} \iff x \in \text{Span}(\begin{bmatrix} 1 \\ 2 \end{bmatrix})$$

$$Ay = 2y \iff \begin{bmatrix} 5 & -3 \\ 6 & -4 \end{bmatrix} \begin{bmatrix} y^1 \\ y^2 \end{bmatrix} = \begin{bmatrix} 2y^1 \\ 2y^2 \end{bmatrix} \iff \begin{cases} 5y^1 - 3y^2 = 2y^1 \\ 6y^1 - 4y^2 = 2y^2 \end{cases} \iff \begin{cases} 3y^1 - 3y^2 = 0 \\ 6y^1 - 6y^2 = 0 \end{cases} \iff y \in \text{Span}(\begin{bmatrix} 1 \\ 1 \end{bmatrix})$$

Let $\mathcal{D}$ be the basis $\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$. We thus have

$$[f_A]_{\mathcal{D}} = M^{\mathcal{B}}_{\mathcal{D}} A M^{\mathcal{D}}_{\mathcal{B}} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} \text{ where } M^{\mathcal{D}}_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

4.2 Direct sums

In fact, we can deduce from Proposition 2.21 a necessary and sufficient condition for f to be diagonalizable. But first we will need to know a bit more about linearly independent vector spaces.

Proposition 4.8: *Let $U_1, \ldots, U_k$ be a family of finite dimensional vector subspaces of a vector space V. Denote by U the sum $U = U_1 + \ldots + U_k$. The following are equivalent:*

1. *$U_1, \ldots, U_k$ are linearly independent;*
2. *for any choice of bases $\mathcal{B}_1 = (b^1_1, \ldots, b^1_{l_1}), \ldots, \mathcal{B}_k = (b^k_1, \ldots, b^k_{l_k})$ for $U_1, \ldots, U_k$ respectively, then $\mathcal{B} = (b^1_1, \ldots, b^1_{l_1}, \ldots, b^k_1, \ldots, b^k_{l_k})$ is a basis for the sum U;*
3. *$\dim U = \dim(U_1) + \ldots + \dim(U_k)$.*

Recall that in general we have $\dim(U_1 + \ldots + U_k) \leq \dim(U_1) + \ldots + \dim(U_k)$

Proof. ($1 \implies 2$) Let us see first that $\mathcal{B}$ spans U. Any vector $u \in U = U_1 + \ldots + U_k$ can be written as $u_1 + \ldots + u_k$ with $u_i \in U_i$, and each u_i can be written as a linear combination of vectors of $\mathcal{B}_i$, so u can be written as a linear combination of vectors in $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_k$: thus $\mathcal{B}$ spans U. If a linear combination of vectors of $\mathcal{B}$ is zero, by grouping the terms corresponding to each $\mathcal{B}_i$ we see that it must be a sum $u_1 + \ldots + u_k$ where each u_i is a linear combinations of the vectors of $\mathcal{B}_i$, and thus by linear independence of the $U_1, \ldots, u_k$, we get $u_1 = \ldots = u_k = 0$. Now since each $\mathcal{B}_i$ is a basis, it means all the coefficients in the linear combination forming u_i are 0.

($2 \implies 3$) is clear, since the dimension is the size of a basis.

($3 \implies 1$) Assume $U_1, \ldots, U_k$ are not linearly independent (we will show that then (3) does not hold). By definition, there exist $u_1 \in U_1, \ldots, u_k \in U_k$ not all zero such that $u_1 + \ldots + u_k = 0$. Up to renumbering, we can assume $u_k \neq 0$. We then get $u_1 + \ldots + u_{k-1} = -u_k \neq 0 \in (U_1 + \ldots + U_{k-1}) \cap U_k$. Thus $\dim(U_1 + \ldots + U_{k-1}) \cap U_k \geq 1$. But we have

$$\begin{aligned} \dim(U_1 + \ldots + U_k) &= \dim(U_1 + \ldots + U_{k-1}) + \dim(U_k) - \dim((U_1 + \ldots + U_{k-1}) \cap U_k) \\ &< \dim(U_1 + \ldots + U_{k-1}) + \dim(U_k) \\ &\leq \dim(U_1) + \ldots + \dim(U_{k+1}) + \dim(U_k) \end{aligned}$$

hence (3) does not hold. □

Definition 4.9: *If the conditions of Proposition 4.8 hold, we say that the sum U is the* ***direct sum*** *of $U_1, \ldots, U_k$, and we write $U = U_1 \oplus \ldots \oplus U_k$.*

Note that in general the sum U is not the whole space V.

Example 4.10:

- If U_1, U_2 are complementary subspaces in V, then in particular they are linearly independent (can you prove this?). In this case the sum $U_1 \oplus U_2$ is the whole of V.
- If l_1, l_2 are two distinct lines through the origin in $\mathbb{R}^3$, they are linearly independent. Then $l_1 \oplus l_2$ has dimension $\dim(l_1) + \dim(l_2) = 1 + 1 = 2$. It is the (unique) plane containing both l_1 and l_2. It is strictly contained in $\mathbb{R}^3$.

4.3 A necessary and sufficient condition for diagonalization

Proposition 4.11: *Let $f : V \to V$ be an operator. Let $\lambda_1, \ldots, \lambda_k$ be its eigenvalues. f is diagonalizable*
$\iff$ *the direct sum of the eigenspaces $V_{\lambda_1}, \ldots, V_{\lambda_k}$ is exactly V*
$\iff \dim V = \dim(V_{\lambda_1}) + \ldots + \dim(V_{\lambda_k})$

Proof. Suppose that $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$. Choose a basis $\mathcal{B}_i$ for each eigenspace V_{λ_i}: each $\mathcal{B}_i$ consists only of eigenvectors associated to λ_i. By Proposition 4.8(2), the union of the $\mathcal{B}_i$ is a basis for V, hence V admits a basis of eigenvectors of f, so f is diagonalizable by Proposition 4.5.

For the other direction, suppose that the direct sum of the eigenspaces is not all of V. Then for any family $v_1, \ldots, v_r$ of eigenvectors, $\mathrm{Span}(v_1, \ldots, v_r) \leq V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$ is a proper vector subspace of V; we cannot find a basis for V which consists exclusively of eigenvectors of f. Hence f is not diagonalizable. □

5 Finding the eigenvalues - the characteristic polynomial

5.1 Eigenvalues of an operator/ eigenvalues of a matrix

Let us clarify the relation between eigenvalues of an operator and that of a matrix. Recall that we defined as follows eigenvalues of an operator: λ is an eigenvalue for the operator $f \iff$ there exists $v \in V$ with $v \neq 0$ such that $f(v) = \lambda v$.

Now if $\mathcal{B}$ is a basis for V, we get

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

so λ is an eigenvalue for the matrix $[f]_{\mathcal{B}}$.

On the other hand, if $\lambda \in \mathbb{F}$ is an eigenvalue of a matrix $A = [f]_{\mathcal{B}} \in M_n(\mathbb{F})$ it means by definition that there exists a nonzero column vector $x \in \mathbb{F}^n_{col}$ such that $Ax = \lambda x$. Thus if v is the vector whose coordinates in $\mathcal{B}$ are exactly x, it is nonzero and we have $[f(v)]_{\mathcal{B}} = [f]_{\mathcal{B}} x = Ax = \lambda x = [\lambda v]_{\mathcal{B}}$, hence we have $f(v) = \lambda v$ so λ is an eigenvalue for f.

Summary. λ is an eigenvalue of f if and only if it is an eigenvalue for the matrix $[f]_{\mathcal{B}}$ representing f.

5.2 Definition of the characteristic polynomial

Given a linear operator $f : V \to V$, how can we find its eigenvalues?

Let $\lambda \in \mathbb{F}$. Then λ is an eigenvalue for f iff there exists a vector $v \in V$ with $v \neq 0$ such that $f(v) = \lambda v$. Now

$$\begin{aligned} f(v) = \lambda v &\iff \lambda v - f(v) = 0 \\ &\iff \lambda \mathrm{id}(v) - f(v) = 0 \\ &\iff (\lambda \mathrm{id} - f)(v) = 0 \\ &\iff v \in \mathrm{Ker}(\lambda \mathrm{id} - f) \end{aligned}$$

In other words, λ is an eigenvalue for f iff $\mathrm{Ker}(\lambda \mathrm{id} - f)$ contains a nonzero vector v.

Thus in the finite dimensional case we get

Proposition 5.1: *Let V be a finite dimensional vector space, $f : V \to V$ be a linear operator, and let $\lambda \in \mathbb{F}$. The following are equivalent:*

- *λ is an eigenvalue of f;*
- *$\mathrm{Ker}(\lambda \mathrm{id} - f)$ is non trivial;*
- *$\mathrm{Im}(\lambda \mathrm{id} - f)$ is not all of V;*
- *$(\lambda \mathrm{id} - f)$ is not invertible.*

Proposition 2.13 is the matrix version of the proposition above. We restate it here for convenience

Proposition 2.13: *Let $A \in M_n(\mathbb{F})$ be a square matrix, and let $\lambda \in \mathbb{F}$. The following are equivalent:*

- *λ is an eigenvalue of A;*
- *$\mathcal{R}^0(\lambda I - A) = \{x \in \mathbb{F}^n_{col} \mid (\lambda I - A)x = 0\}$ is non trivial;*
- *$\mathcal{C}(\lambda I - A) = \{(\lambda I - A)x \mid x \in \mathbb{F}^n_{col}\}$ is not all of $\mathbb{F}^n_{col}$;*
- *$(\lambda I - A)$ is not invertible;*
- *$\det(\lambda I - A) = 0$.*

Remark 5.2: *The eigenvalues of A are exactly the scalars $\lambda \in \mathbb{F}$ for which the matrix $A - \lambda I$ has determinant equal to zero - that is, they are the solutions in $\mathbb{F}$ of the equation $\det(tI - A) = 0$ in the variable t.*

Example 5.3: Consider $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$, we have

$$\det(tI - A) = \det \begin{bmatrix} t-1 & -2 \\ -1 & t \end{bmatrix} = (t-1)t - 2 = t^2 - t - 2 = (t+1)(t-2)$$

Thus the eigenvalues of A are exactly -1 and 2.

We formalize the previous example by giving:

Definition 5.4: *Let $A \in M_n(\mathbb{F})$ be given by*

$$A = \begin{bmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & & \ddots & \vdots \\ a_1^n & & & a_n^n \end{bmatrix} \text{ and consider } XI - A := \begin{bmatrix} (X - a_1^1) & -a_2^1 & \dots & -a_n^1 \\ -a_1^2 & (X - a_2^2) & \dots & -a_n^2 \\ \vdots & & \ddots & \vdots \\ -a_1^n & & & (X - a_n^n) \end{bmatrix}$$

Formally, $XI - A$ is a matrix in $M_n(\mathbb{F}[X])$ - its entries are polynomials with coefficients in $\mathbb{F}$. Write $P^i_j(X)$ for the entry of $XI - A$ on the i-th row and the j-th column.

*The **characteristic polynomial** of the matrix A is defined to be the polynomial*

$$\chi_A(X) = \det(XI - A) := \sum_{\sigma \in S_n} \mathrm{sg}(\sigma) P^1_{\sigma(1)}(X) \dots P^n_{\sigma(n)}(X)$$

where S_n is the group of permutations on $\{1, \dots, n\}$ and if $\sigma \in S_n$, then $\mathrm{sg}(\sigma)$ is the signature of the permutation σ.

Proposition 5.5: *The eigenvalues of $A \in M_n(\mathbb{F})$ are exactly the solutions in $\mathbb{F}$ of the polynomial equation $\chi_A(X) = 0$.*

Proof. λ is an eigenvalue of A iff $\det(\lambda I - A) = 0$, that is, iff $\chi_A(\lambda) = 0$. □

Remark 5.6: *Recall that if $P \in \mathbb{F}[X]$ is given by $P(X) = a_0 + a_1 X + \ldots + a_n X^n$ with $a_n \neq 0$, the index n is called the degree $\deg(P)$ of the polynomial P. It is not hard to show that if $A \in M_n(A)$ then χ_A has degree n.*

Example 5.7: Consider the matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Its characteristic polynomial is $\chi_A(X) = X^2 + 1$.

What are its eigenvalues? It depends!

If we consider A as a matrix in $M_2(\mathbb{R})$, it has no eigenvalues!

If we consider A as a matrix in $M_2(\mathbb{C})$, it has eigenvalues i and $-i$.

What is going on here?

In the first case, we consider A to represent a linear operator $\mathbb{R}^2 \to \mathbb{R}^2$: in fact, it is a clockwise rotation around 0 by an angle of $\pi/2$. Clearly this operator has no invariant lines, hence no eigenvectors and no eigenvalues.

In the second case, we consider A to represent a linear operator $\mathbb{C}^2 \to \mathbb{C}^2$ (harder to visualize!), for which i is an eigenvalue - for example, the vector $\begin{bmatrix} 1 \\ i \end{bmatrix}$ is an eigenvector associated to i.

IMPORTANT: remember that the set of eigenvalues of a matrix depends on the field we are working in!

5.3 Characteristic polynomial of an operator

So far we have defined the characteristic polynomial of a matrix. Given an operator $f : V \to V$ and bases $\mathcal{B}, \mathcal{D}$ for V. What is the relation between the characteristic polynomials of $[f]_\mathcal{B}$ and $[f]_\mathcal{D}$?

Lemma 5.8: *Similar matrices have the same characteristic polynomial.*

Proof. Let $A, B \in M_n(\mathbb{F})$ be similar matrices: there exists an invertible matrix M, such that $B = MAM^{-1}$. We have

$$\begin{aligned} \chi_B(X) &= \det(XI - B) = \det(XI - MAM^{-1}) \\ &= \det(XMIM^{-1} - MAM^{-1}) = \det(M(XI - A)M^{-1}) \\ &= \det(M)\det(XI - A)\det(M^{-1}) = \det(XI - A) = \chi_A(X) \end{aligned}$$

(the property that $\det(Q_1(X)Q_2(X)) = \det(Q_1(X))\det(Q_2(X))$ for matrices in $M_n(\mathbb{F}[X])$ is proved exactly as for matrices in $M_n(\mathbb{F})$). □

Since all the matrices representing a given operator are similar, we can give the following definition

Definition 5.9: *Let V be a finite dimensional vector space over $\mathbb{F}$. Let $f : V \to V$ be a linear operator. The **characteristic polynomial** $\chi_f(X)$ of f is the characteristic polynomial of a matrix A representing f with respect to some (and thus any) basis.*

Example 5.10: Let $r : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation of angle $\pi/2$ around the origin. If $\mathcal{B} = (b_1, b_2)$ denotes the standard basis, we have

$$[r]_\mathcal{B} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Hence the characteristic polynomial of r is

$$\begin{aligned} \chi_r(X) &= \chi_{[r]_\mathcal{B}} = \det(XI - [r]_\mathcal{B}) \\ &= \det \begin{bmatrix} X & -1 \\ 1 & X \end{bmatrix} = X^2 + 1 \end{aligned}$$

From the Proposition above we get

Proposition 5.11: *The eigenvalues of f are exactly the solutions in $\mathbb{F}$ of the polynomial equation $\chi_f(X) = 0$.*

Proof. Let $\mathcal{B}$ be a basis for V. Let $\lambda \in \mathbb{F}$.

Suppose λ is an eigenvalue for f: there exists $v \in V$, $v \neq 0$, such that $f(v) = \lambda v$. Now $[f(v)]_\mathcal{B} = [f]_\mathcal{B}[v]_\mathcal{B} = \lambda[v]_\mathcal{B}$ so λ is an eigenvalue for $[f]_\mathcal{B}$ - hence it is a solution of $\chi_f(X) = 0$.

Conversely if λ is a solution of $\chi_f(X) = \chi_{[f]_\mathcal{B}}(X) = 0$, it is an eigenvalue for $[f]_\mathcal{B}$ so some nonzero vector v in V satisfies $f(v) = \lambda v$ - thus λ is an eigenvalue for f. □

Remark 5.12: *If V is a vector space over $\mathbb{C}$, then it always admits an eigenvalue, because every polynomial equation over $\mathbb{C}$ admits a solution (this is the Fundamental theorem of Algebra). If V is a vector space over $\mathbb{R}$, we can think of the matrix A representing f as a matrix over $\mathbb{C}$, and find an eigenvalue for A but* ***it won't be an eigenvalue for*** *f!(recall Example 5.7).*

Note that once we know that λ is an eigenvalue for A, finding the corresponding eigenvectors amounts to finding column vectors $\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}$ such that

$$A \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = \lambda \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = \begin{bmatrix} \lambda x^1 \\ \vdots \\ \lambda x^n \end{bmatrix}$$

But this amounts to solving a (homogeneous) system of linear equations, which you learned how to do last semester.

5.4 Factors of the characteristic polynomial

If U is an f-invariant subspace, recall we can consider the restriction $g = f\ |_U \colon U \to U$ defined by $g(u) = f(u)$ for any $u \in U$. Then there is a natural relation between χ_f and χ_g, given by the following

Proposition 5.13: *Let $f : V \to V$ be an operator on a finite-dimensional vector space V. Suppose U is an f-invariant subspace. Then $\chi_{f|_U}$ divides χ_f, that is, there exists a polynomial $Q(X)$ such that $\chi_f(X) = Q(X)\chi_{f|_U}(X)$.*

Proof. Recall that if we choose a basis for U and extend it to a basis $\mathcal{B}$ of V then the matrix of f with respect to $\mathcal{B}$ is of the form

$$[f]_\mathcal{B} = \left[\begin{array}{c|c} [f\ |_U]_{\mathcal{B}_U} & A \\ \hline 0 & B \end{array}\right]$$

Hence

$$\chi_f(X) = \det(XI - [f]_{\mathcal{B}}) = \det\left[\begin{array}{c|c} XI - [f\mid_U]_{\mathcal{B}_U} & -A \\ \hline 0 & XI - B \end{array}\right]$$
$$= \det(XI - [f\mid_U]_{\mathcal{B}_U})\det(XI - B)$$
$$= \chi_{f|_U}(X)\det(XI - B)$$

□

Example 5.14: We let $V = \mathbb{R}^3_{col}$, and $f : V \to V$ be a rotation of angle $\pi/2$ around the z-axis. Denote by U the xy-plane: it is an f-invariant subspace, and the restriction of f to U is a rotation $r : U \to U$ around the origin by an angle of $\pi/2$ just as in Example 5.10. If $\mathcal{B} = (b_1, b_2, b_3)$ denotes the standard basis, we have

$$[f]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence

$$\chi_f(X) \quad = \quad \det(XI - [f]_{\mathcal{B}}) = \det\begin{bmatrix} X & -1 & 0 \\ 1 & X & 0 \\ 0 & 0 & -1 \end{bmatrix} = (X^2 - 1)(X - 1)$$

Indeed we see that the characteristic polynomial $\chi_{f|_U}(X) = X^2 - 1$ of the restriction $r = f\mid_U$ (which we computed in Example 5.10) divides the characteristic polynomial of f.

We see that the reason for this is that the matrix $[f\mid_U]_{\mathcal{B}_U} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ appears as a block in the matrix $[f]_{\mathcal{B}}$ which is diagonal by blocks.

A particular case of Proposition 5.13 is that when U is a one dimensional invariant subspace - that is, when $U = \text{Span}(v)$ with v an eigenvector. Then $[f\mid_U] = [\lambda]$ and $\chi_{f|_U} = X - \lambda$. Thus $(X - \lambda)$ divides χ_f.

5.5 Characteristic polynomial of the restriction to a cyclic subspace

Another particular case is when we take U to be the cyclic subspace $Z(v, f)$ associated to a non zero vector v.

We saw (see Section 2.2) that if k is the largest index such that the vectors $v, f(v), \ldots, f^k(v)$ are linearly independent, then $\mathcal{B} = (v, f(v), \ldots, f^k(v))$ forms a basis for $Z(f, v)$.

In the end of Section 3, we showed that if $a_0 v + a_1 f(v) + \ldots + a_k f^k(v) + f^{k+1}(v) = 0$ is the linear dependence relation that the vectors $v, f(v), \ldots, f^{k+1}(v)$ satisfy, we call $P(X) = a_0 + a_1 X + \ldots + a_k X^k + X^{k+1}$ the **polynomial associated to** $Z(f, v)$, and the matrix for the restriction $g = f\mid_{Z(f,v)}$ with respect to $\mathcal{B}$ is the **companion matrix** of P, that is it has the form

$$[g]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -a_{k-1} \\ 0 & 0 & 0 & \ldots & 1 & -a_k \end{bmatrix}$$

Proposition 5.15: *Let $f : V \to V$ be an operator on an n-dimensional vector space V, and build the cyclic subspace $Z(f, v)$ associated to f and a vector $v \in V$. The characteristic polynomial of $g = f\mid_{Z(f,v)}$ is exactly the polynomial P associated to $Z(f, v)$.*

Proof. We prove by induction on k that

$$\det\begin{bmatrix} X & 0 & 0 & \dots & 0 & a_0 \\ -1 & X & 0 & \dots & 0 & a_1 \\ 0 & -1 & X & \dots & 0 & a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & X & a_{k-1} \\ 0 & 0 & 0 & \dots & -1 & X+a_k \end{bmatrix} = a_0 + a_1X + \ldots + a_kX^k + X^{k+1} = P(X)$$

Let us compute the determinant according to the first row:

$$\begin{aligned} \chi_g(X) &= X \cdot det \begin{bmatrix} X & 0 & \dots & 0 & a_1 \\ -1 & X & \dots & 0 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & X & a_{k-1} \\ 0 & 0 & \dots & -1 & X+a_k \end{bmatrix} + (-1)^{k+2}a_0 \cdot det \begin{bmatrix} -1 & X & 0 & \dots & 0 \\ 0 & -1 & X & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & X \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix} \\ &= X(a_1 + a_2X + \ldots + a_kX^{k-1} + t^k) + (-1)^{k+2}a_0(-1)^k \text{ by applying IH for the first term} \\ &= a_0 + a_1X + \ldots + a_kX^k + X^{k+1} = P(X) \end{aligned}$$

that is, $\chi_{f|_{Z(f,v)}} = P$. □

Thus we get

Corollary 5.16: *Let f be an operator on a finite dimensional vector space V. Let $v \in V$, and consider the cyclic subspace $Z(f,v)$. Denote by P the polynomial associated to $Z(f,v)$. Then P divides χ_f, that is, there exists a polynomial Q such that $\chi_f(X) = Q(X)P(X)$.*

5.6 Cayley-Hamilton theorem

Let V be an n-dimensional vector space over a field $\mathbb{F}$.

Polynomial of operators. Recall that we know how to take powers of an operator $f : V \to V$ - we simply define $f^2 = f \circ f, f^3 = f \circ f \circ f, \ldots$ If A is the matrix representing f with some basis $\mathcal{B}$ of V, then the power f^k is represented by the matrix $A^k = A \times \ldots \times A$.

Extending this idea, if we are given a polynomial $P \in \mathbb{F}[X]$, say $P(X) = a_nX^n + \ldots + a_1X + a_0$, we can define a new operator $P(f) = a_nf^n + \ldots + a_1f + a_0\text{id}$ which sends a vector v to

$$P(f)(v) = a_nf^n(v) + \ldots + a_1f(v) + a_0v$$

Equivalently, if A is a matrix representing f in a basis $\mathcal{B}$, one can take $P(A) = a_nA^n + \ldots + a_1A + a_0I$. The matrix $P(A)$ represents the operator $P(f)$ in $\mathcal{B}$.

The following striking result says that if we plug in an operator f (respectively a matrix) in its characteristic polynomial, we get the zero operator (respectively the zero matrix).

Theorem 5.17: *(Cayley-Hamilton) Let V be a finite dimensional vector space. Let $f : V \to V$ be an operator, and let χ_f denote the characteristic polynomial of f. Then we have*

$$\chi_f(f) = 0$$

Equivalently, let A be an n-by-n matrix with characteristic polynomial χ_A. Then $\chi_A(A) = 0$.

Let us see that this is true on a few examples.

Example 5.18: 1. Consider the matrix $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right] \in M_2(\mathbb{F})$. Its characteristic polynomial is $\chi_A(X) = X^2 + 1$. Plugging A in χ_A, one gets $\chi_A(A) = A^2 + I_2 = 0 \in M_2(\mathbb{F})$. More generally, if this is done for an arbitrary $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in M_2(\mathbb{F})$, we will get $\chi_A(A) = A^2 - (\text{tr}A)A + (\det A)I_2 = A^2 - (a+d)A + (ad - bc)I_2 = 0 \in M_2(\mathbb{F})$ (do the math).

2. Let $f : v \mapsto \lambda v$. Then $\chi_f(X) = (X - \lambda)^n$ where $n = \dim V$. Then $\chi_f(f) = (f - \lambda \mathrm{id})^n = 0$ since $f - \lambda\mathrm{id}$ is already the zero operator.

3. Suppose f is diagonalizable, then there is a basis $\mathcal{B}$ of eigenvectors. Let b be a vector in $\mathcal{B}$, and λ the corresponding eigenvalue. Then we saw that $\chi_f(X) = Q(X)(X - \lambda)$, thus

$$\chi_f(f)(b) = Q(f) \circ (f - \lambda\mathrm{id})(b) = Q(f)(f(b) - \lambda b) = Q(f)(0) = 0.$$

Hence $\chi_f(f)$ sends each vector of $\mathcal{B}$ to 0: this means $\chi_f(f)$ is the zero operator.

We now give the proof in full generality.

Proof. It is enough to show that for any $v \in V$, the operator $\chi_f(f)$ sends v to 0. We thus fix a vector $v \in V$.

Consider the cyclic space $Z(f, v)$, and P the polynomial associated to $Z(f, v)$: it is the polynomial $P(X) = a_0 + a_1X + \ldots + a_kX^k + X^{k+1}$ whose coefficients correspond to those of the first linear relation

$$a_0 + a_1f(v) + \ldots + a_kf^k(v) + f^{k+1}(v) = 0 \quad (\dagger)$$

between the vectors $v, f(v), f^2(v), \ldots$ Note that the relation (†) can be rewritten as $P(f)(v) = 0$.

On the other hand, we saw in Proposition 5.16 that P divides χ_f, since it is precisely the characteristic polynomial of $f\mid_{Z(f,v)}$. In other words, $\chi_f(X) = Q(X)P(X)$ (note that this way of writing χ_f will be different for each choice of v!)

$$\chi_f(f)(v) = Q(f) \circ P(f)(v) = Q(f)(P(f)(v)) = Q(f)(0) = 0.$$

□

The following example illustrates the proof of Cayley Hamilton on a specific operator, and for three choice of the vector v.

Example 5.19: (see picture in class) Let $V = \mathbb{R}^3_{col}$, and let $f : V \to V$ be a rotation of angle $\pi/2$ around the z-axis as in Example 5.14. Recall that we showed that $\chi_f(X) = (X^2 + 1)(X - 1) = X^3 - X^2 + X - 1$. Our goal is to show that the operator $\chi_f(f) = (f^2 + \mathrm{id}) \circ (f - \mathrm{id}) = f^3 - f^2 + f - \mathrm{id}$ sends every vector to zero. Let $\mathcal{B} = (b_1, b_2, b_3)$ denotes the standard basis.

- Let $u = b_3$. The cyclic subspace $Z(f, u)$ is generated by $u, f(u), f^2(u), \ldots$ In fact, we have $f(u) = u$ so $u, f(u)$ are linearly dependent - we can rewrite this relation as

$$f(u) - u = (f - \mathrm{id})(u) = 0 \ (\dagger)$$

Hence $Z(f, u) = \mathrm{Span}(u)$ and the polynomial associated to $Z(f, u)$ is $P_u(X) = X - 1$. We see that (†) can be rewritten as $P_u(f)(u) = 0$. Now we get

$$\chi_f(f)(u) = (f^2 + \mathrm{id}) \circ (f - \mathrm{id})(u) = (f^2 + \mathrm{id})(f(u) - u) = (f^2 + \mathrm{id})(0) = 0$$

- Let $v = b_1$. The cyclic subspace $Z(f, v)$ is generated by $v, f(v), f^2(v), \ldots$ Note that $f(v) = -b_2$ is linearly independent from v, but that $f^2(v) = -v$ so $v, f(v), f^2(v)$ are linearly dependent - we can rewrite this relation as

$$f^2(v) - v = (f^2 - \mathrm{id})(v) = 0 \ (\dagger)$$

Hence $Z(f, v) = \mathrm{Span}(v, f(v))$ and the polynomial associated to $Z(f, v)$ is $P_v(X) = X^2 - 1$. We see that (†) can be rewritten as $P_v(f)(v) = 0$. Now we get

$$\chi_f(f)(v) = (f - \mathrm{id}) \circ (f^2 + \mathrm{id})(v) = (f - \mathrm{id})(f^2(v) + v) = (f - \mathrm{id})(0) = 0$$

- Let $w = b_1 + b_3$. The cyclic subspace $Z(f, w)$ is generated by $w, f(w), f^2(w), \ldots$ Note that $f(w) = -b_2 + b_3$ is linearly independent from v, that $f^2(w) = -b_1 + b_3$ so $w, f(w), f^2(w)$ are linearly independent. Now $w, f(w), f^2(w), f^3(w)$ are necessarily linearly dependent since the space in which they live has dimension 3. What is the relation (†)? Note that $w + f^2(w) = f^3(w) + f(w) = 2b_3$, so we get

$$f^3(w) - f^2(w) + f(w) - w = 0 \ (\dagger)$$

Now $Z(f, w) = \mathrm{Span}(w, f(w), f^2(w)) = \mathbb{R}^3$ and the polynomial associated to $Z(f, w)$ is $P_w(X) = X^3 - X^2 + X - 1 = \chi_f(X)$, and (†) can be rewritten as $\chi_f(f)(w) = 0$.

The following example shows things we can deduce from Cayley-Hamilton.

Example 5.20: Consider the matrix $A = \begin{bmatrix} 5 & -3 \\ 6 & -4 \end{bmatrix}$. Its characteristic polynomial is

$$\chi_A(X) = \det(XI - A) = \det \begin{bmatrix} X-5 & 3 \\ -6 & X+4 \end{bmatrix} = (X-5)(X+4) + 18 = X^2 - X - 2 = (X-2)(X+1)$$

hence we have two eigenvalues 2 and -1.

Thus we know that $\chi_A(A) = (A - 2I)(A + I) = 0$. Take any vector $x \in \mathbb{R}^2_{col}$, we have $(A - 2I)(A + I)x = 0$. There are two possibilities - either $(A + I)x = 0$, that is, x is an eigenvector for the eigenvalue -1, or $(A - 2I)y = 0$ for $y = (A + I)x$ so y is an eigenvector for the eigenvalue 2.

Starting from any vector x, this gives a way to "produce" an eigenvector (either x itself or its image under a polynomial in A).

5.7 Algebraic multiplicity - A new criterion for diagonalization

If we apply Proposition 5.13, we get the following result:

Proposition 5.21: *Let $f : V \to V$ be an operator on a finite dimensional vector space. Let λ be an eigenvalue for f, and let V_λ be the corresponding eigenspace. Then $(X - \lambda)^{\dim V_\lambda}$ divides the characteristic polynomial χ_f.*

Recall that the geometric multiplicity of λ was precisely $m_\lambda^{geom} = \dim V_\lambda$.

Proof. V_λ is an f-invariant subspace, hence $\chi_{f|_{V_\lambda}}$ divides χ_f. Now for any basis $\mathcal{B}$ of V_λ, we have

$$[f\mid_{V_\lambda}]_{\mathcal{B}} = \begin{bmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{bmatrix} \text{ so } \chi_{f|_{V_\lambda}}(X) = \det(XI - [f\mid_{V_\lambda}]) = \det \begin{bmatrix} X-\lambda & & 0 \\ & \ddots & \\ 0 & & X-\lambda \end{bmatrix} = (X-\lambda)^{\dim V_\lambda}$$

which proves the claim. □

Definition 5.22: *Let $f : V \to V$ be an operator on a finite dimensional vector space V, and let λ be an eigenvalue for f. The* ***algebraic multiplicity*** *of λ is defined to be*

$$m_\lambda^{alg} = \max\{m \mid (X - \lambda)^m \text{ divides } \chi_f\}$$

From the proposition above we get

Corollary 5.23: *Let $f : V \to V$ be an operator on a finite dimensional vector space V, and let λ be an eigenvalue for f. Then the algebraic multiplicity is at least equal to the geometric multiplicity, that is, $m_\lambda^{geom} \leq m_\lambda^{alg}$.*

Finally we show a new criterion for diagonalizability of an operator

Proposition 5.24: *(unlectured) Let V be an n-dimensional vector space. An operator $f : V \to V$ is diagonalizable over $\mathbb{F}$ if and only if its characteristic polynomial χ_f can be written as a product of linear factors over $\mathbb{F}$ and the algebraic multiplicity of each eigenvalue is equal to the geometric multiplicity.*

Linear polynomials are polynomials of degree 1, that is, they are of the form $a_0 + a_1 X$ with $a_1 \neq 0$ (in particular they can be rewritten as $a_1(X - \alpha)$ by setting $\alpha = -a_0/a_1$).

Proof. Let $\lambda_1, \dots \lambda_k$ denote the eigenvalues of f. By Proposition 4.11, f is diagonalizable iff the space V is the (direct) sum of the eigenspaces $V_{\lambda_1}, \dots, V_{\lambda_k}$.

Thus if f is diagonalizable, by Proposition 4.8 we can put together bases $\mathcal{B}_1, \dots, \mathcal{B}_k$ of $V_{\lambda_1}, \dots, \lambda_k$ to get a basis $\mathcal{B} = (b_1, \dots, b_n)$ for V. The matrix for f in $\mathcal{B}$ is diagonal with diagonal entries $\alpha_1, \dots, \alpha_n$, and for each i we have $f(b_i) = \alpha_i b_i$, where α_i is equal to λ_j for all i such that $b_i \in \mathcal{B}_j$. Computing the characteristic polynomial for A we get

$$\chi_f(X) = (X - \alpha_1) \dots (X - \alpha_n) = (X - \lambda_1)^{m_1} \dots (X - \lambda_k)^{m_k}$$

where $m_j = \dim V_{\lambda_j}$. In particular χ_f does split as a product of linear factors.

Moreover, we see that for each j, the algebraic multiplicity of λ_j is $\max\{m \mid (X-\lambda_j)^m \text{ divides } \chi_f\} = m_j = \dim V_{\lambda_j}$, so $m_{geom}(\lambda_j) = m_{alg}(\lambda_j)$.

Let us prove the other direction. Suppose that the characteristic polynomial χ_f can be written as a product of linear factors $(X - \alpha_1) \dots (X - \alpha_n)$. Each of the α_i satisfies $\chi_f(\alpha_i) = 0$ so it is equal to one of the eigenvalues $\lambda_1, \dots, \lambda_k$. Hence we can write

$$\chi_f(X) = (X - \lambda_1)^{m_1} \dots (X - \lambda_l)^{m_k}$$

The algebraic multiplicity $m_{alg}(\lambda_j)$ of the eigenvalue λ_j is precisely m_j. Note that we have $\deg \chi_f = n = m_1 + \dots + m_k$.

If the geometric multiplicity $m_{geom}(\lambda_j)$ is exactly the algebraic multiplicity m_j for each j, we have that $\dim(V_{\lambda_1}) + \dots + \dim(V_{\lambda_k}) = m_1 + \dots + m_k = n$, hence f is diagonalizable by Proposition 4.11. □

(We are using some facts about polynomials that are pretty intuitive, for a more formal exposition see the notes on polynomials from last year).

6 Jordan normal form

Not every operator is diagonalizable. One of the reason might be that it doesn't have eigenvalues. But still, an operator might have eigenvalues and yet not be diagonalizable

Example 6.1: Let $f : \mathbb{R}^2_{col} \to \mathbb{R}^2_{col}$ be represented by the matrix $\begin{bmatrix} \lambda & 0 \\ 1 & \lambda \end{bmatrix}$ in the standard basis $\mathcal{B} = \{b_1, b_2\}$ of $\mathbb{R}^2_{col}$. The characteristic polynomial of f is $(X - \lambda)^2$ so λ is an eigenvalue for f - in fact, it is its only eigenvalue.

We claim that f is not diagonalizable - if it were, the corresponding diagonal matrix would be $\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$, that is, f would be the operator which multiplies every vector by λ. But clearly $f(b_1) = \lambda b_1 + b_2$ so this is not the case.

The aim of this section is to get the "next best thing": a matrix representation for f which is "almost diagonal".

Definition 6.2: *Let $\lambda \in \mathbb{F}$. A **Jordan elementary block** $J_k(\lambda)$ is a k-by-k matrix of the form*

$$J_k(\lambda) = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 1 & \lambda & \dots & 0 \\ & \ddots & \ddots & \\ 0 & \dots & 1 & \lambda \end{bmatrix}$$

*A **Jordan block associated to** λ is a matrix which is diagonal by blocks, with Jordan elementary blocks $J_{k_1}(\lambda), \dots, J_{k_r}(\lambda)$ on the diagonal.*

*A **Jordan matrix** is a matrix which is diagonal by blocks, with Jordan blocks $J(\lambda_1), \ldots, J(\lambda_s)$ on the diagonal, , that is*

$$[f]_{\mathcal{B}} = \begin{bmatrix} \boxed{J(\lambda_1)} & & & \\ & \boxed{J(\lambda_2)} & & \\ & & \ddots & \\ & & & \boxed{J(\lambda_s)} \end{bmatrix}$$

This is the result we will prove

Theorem 6.3: *(Jordan canonical form) Let V be a finite dimensional vector space over $\mathbb{C}$. Let $f : V \to V$. There exists a basis $\mathcal{B}$ for V in which f is represented by a Jordan matrix. This matrix is unique up to the order of the blocks.*

Exercise 6.4: What are all the possible forms of Jordan blocks for $k = 1, 2, 3$? What are all the possible forms of n-by-n Jordan matrices for $n = 1, 2, 3$?

The matrix version of the above result is

Theorem 6.5: *Let $A \in M_n(\mathbb{C})$. There exists an invertible matrix $P \in M_n(\mathbb{C})$ such that PAP^{-1} is a Jordan matrix.*

Remark 6.6: *We see that the characteristic polynomial of f is of the form $\chi_f(t) = (t - \lambda_1)^{k_1} \ldots (t - \lambda_r)^{k_r}$. In particular, $\lambda_1, \ldots, \lambda_r$ are exactly the eigenvalues of f.*

Before proving the theorem, let us first try and understand the properties the basis $\mathcal{B}$ must have for the matrix of f to be of this form.

6.1 Jordan elementary blocks

What does it mean if an operator $f : V \to V$ is represented by a Jordan elementary block with respect to a basis $\mathcal{B}$? Suppose $\mathcal{B} = (b_1, \ldots, b_k)$. Then we have

$$f(b_1) = \lambda b_1 + b_2,\ f(b_2) = \lambda b_2 + b_3,\ \ldots,\ f(b_{k-1}) = \lambda b_{k-1} + b_k \text{ and } f(b_k) = \lambda b_k$$

In particular b_k is an eigenvector with eigenvalue λ.

<u>**Special case where** $\lambda = 0$**.**</u> We have

$$f(b_1) = b_2, f(b_2) = b_3, \ldots, f(b_{k-1}) = b_k \text{ and } f(b_k) = 0 \text{ that is,}$$

$$b_1 \mapsto^f b_2 \mapsto^f \ldots \mapsto^f b_k \mapsto^f 0$$

We see that this means that $V = Z(f, b_1)$, and that $\mathcal{B}$ is the standard basis for the cyclic subspace $Z(f, b_1)$, and moreover the image by f of the last vector is 0 (usually we get it as a linear combination of the previous ones).

Note also that f^k sends every vector of $\mathcal{B}$ to 0 - hence f^k is the zero map. This has a name

Definition 6.7: *An operator $f : V \to V$ is said to be **nilpotent** if there exists $m \in \mathbb{N}$ such that $f^m = 0$. If m is minimal for this property, it is called the **height** or **nilpotency index** for f. If $v \neq 0$, then the minimal integer l such that $f^l(v) = 0$ is called the **height of** v **relative to** f (note that it is at most m).*

*A matrix A is **nilpotent** if there exists $m \in \mathbb{N}$ such that $A^m = 0$.*

Exercise 6.8: If $[f]_{\mathcal{B}} = J_k(0)$, what is the matrix of $[f^2]_{\mathcal{B}}$? (without computing the product of two matrices). Of $[f^3]_{\mathcal{B}}$? etc.

Back to the general case (where λ is not necessarily 0). Then note that the operator $h = f - \lambda\text{id}$ is represented by the matrix

$$J_k(\lambda) - \lambda I = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 1 & \lambda & \dots & 0 \\ & \ddots & \ddots & \\ 0 & \dots & 1 & \lambda \end{bmatrix} - \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ & \ddots & \ddots & \\ 0 & \dots & 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ & \ddots & \ddots & \\ 0 & \dots & 1 & 0 \end{bmatrix} = J_k(0)$$

so h is nilpotent.

Note now that writing $[f]_{\mathcal{B}}$ as $D + N$ where $D = \lambda I$ and N is nilpotent is very useful for computations. Indeed,

1. it is easy to compute powers of D and of N;
2. $DN = ND$.

thus for example

$$[f^2]_{\mathcal{B}} = (D + N)^2 = D^2 + DN + ND + N^2 = D^2 + 2DN + N^2$$

Example 6.9: If $A = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix}$, compute A^2 using this technique .

6.2 Jordan blocks

What if the matrix of f with respect to $\mathcal{B}$ is a Jordan block $J(\lambda)$?

This happens exactly when the basis $\mathcal{B}$ is a union of bases of the form

$$\begin{gathered} b_1^1, b_2^1, \dots, b_{k_1}^1 \\ b_1^2, b_2^2, \dots, b_{k_2}^2 \\ \dots \\ b_1^r, b_2^r, \dots, b_{k_r}^r \end{gathered}$$

where $f(b_i^j) = \lambda b_i^j + b_{i+1}^j$ for $i < k_i$, and $f(b_{k_j}^j) = \lambda b_{k_j}^j$ the last vector of each list is an eigenvector associated to λ.

Case where $\lambda = 0$. In this case each list is the standard basis of the cyclic subspace $Z(f, b_1^j)$. The last vector $b_{k_j}^j$ is sent to 0, that is, it lies in Ker$f$.

Also here the operator is nilpotent. We can see this directly: let $k = \max\{k_1, \dots, k_r\}$. Then see that f^k sends every basis vector to 0. Or, we could note that $\chi_f = t^d$ (where d is the dimension of V) and then see that by Cayley-Hamilton, we must have $\chi_f(f) = f^d = 0$.

Definition 6.10: *Let $f : V \to V$ be an operator. A **chain for** f is a sequence $v_1, \dots, v_k$ such that $v_{i+1} = f(v_i)$ for all $i < k$, and $f(v_k) = 0$.*

Remark 6.11: *An operator f is represented by a Jordan block associated to 0 in the basis $\mathcal{B}$ iff $\mathcal{B}$ is a union of chains.*

The "only if" direction is proved in the paragraph above, the "if" direction is straightforward.

6.3 Nilpotent operators

We start by showing that a nilpotent operator can always be represented by a Jordan block $J(0)$.

Theorem 6.12: *Let $h : V \to V$ be a nilpotent operator. There exists a basis $\mathcal{B}$ for V which is a union of chains for f, equivalently, there exists a basis $\mathcal{B}$ in which the matrix representing f is a Jordan block associated to 0.*

First note that when f is nilpotent, it is easy to produce a chain: take any vector $v \in V$, if k is the height of v we have $v \mapsto f(v) \mapsto f^2(v) \mapsto \ldots \mapsto f^{k-1}(v) \mapsto 0$ so these vectors form a chain for f.

The following proposition shows that moreover the vectors of such a chain are linearly independent.

Proposition 6.13: *Let f be a nilpotent operator, and let $v \in V$ be a vector of height k - that is, $f^k(v) = 0$ but $f^{k-1}(v) \neq 0$. Then the vectors $v, f(v), \ldots, f^{k-1}(v)$ are linearly independent. They are the standard basis for $Z(f, v)$.*

Proof. Suppose not. There exist scalars $a_0, \ldots, a_{k-1}$ which are not all zero such that

$$a_0 v + a_1 f(v) + \ldots + a_{k-1} f^{k-1}(v) = 0$$

Now suppose j is the smallest index for which $a_j \neq 0$ and let us apply f^{k-j-1} on both sides of the equation. We get

$$\begin{aligned} f^{k-j-1}(0) = 0 &= f^{k-j-1}(a_j f^j(v) + \ldots + a_{k-1} f^{k-1}(v)) \\ &= a_j f^{k-1}(v) + a_{j+1} f^k(v) + \ldots + a_{k-1} f^{2k-2-j}(v) \\ &= a_j f^{k-1}(v) \end{aligned}$$

which contradicts the fact that $a_j \neq 0$ and $f^{k-1}(v) \neq 0$. □

Definition 6.14: *Let $C_1, \ldots, C_s$ be a set of chain bases. We say the chain bases $C_1, \ldots, C_s$ are **linearly independent** if the family $\{f^j(v_i) \mid 1 \leq i \leq s, 0 \leq j \leq k_i - 1\}$ is linearly independent.*

Note that the following propositions are equivalent

1. the chains $C_1, \ldots, C_s$ are linearly independent;
2. the union of the standard bases $\{f^j(v_i) \mid 0 \leq j \leq k_i - 1\}$ of the spaces $Z(f, v_i)$ form a basis of $Z(f, v_1) + \ldots + Z(f, v_s)$;
3. the spaces $Z(f, v_1), \ldots, Z(f, v_s)$ form a direct sum;
4. for any vectors $z_1, \ldots, z_s$ with $z_i \in Z(f, v_i)$, the sum $z_1 + \ldots + z_s = 0$ implies $z_1 = z_2 = \ldots = z_s = 0$.

We now show that to check whether a set of chain bases is linearly independent, it is enough to check whether the last vectors in the chains are independent.

Proposition 6.15: *Let $f : V \to V$ be a nilpotent operator. Let $v_1, \ldots, v_s$ be vectors in V, and let $C_i : v_i \mapsto f(v_i) \mapsto \ldots \mapsto f^{k_i-1}(v_i) \mapsto 0$ be the chain basis for $Z(f, v_i)$. The chains $C_1, \ldots, C_s$ are linearly independent if and only if the vectors $f^{k_1-1}(v_1), \ldots, f^{k_s-1}(v_s)$ are linearly independent.*

Proof. The "only if" direction is immediate. Suppose there exist vectors $z_1, \ldots, z_s$ with $z_i \in Z(f, v_i)$ such that $z_1 + \ldots + z_s = 0$. Let k be the maximal height of the vectors z_i which are non zero. Then $0 = f^{k-1}(z_1 + \ldots + z_s) = f^{k-1}(z_1) + \ldots + f^{k-1}(z_s)$. Note that $f^{k-1}(z_i) \in \mathrm{Span}(f^{k-1(v_i)})$, and not all the vectors $f^{k-1}(z_i)$ are zero. This proves the claim. □

To prove the Theorem about nilpotent operators, we need to find a basis for V which is a union of chains, in other words, we need to find linearly independent chains which span V. It is easy to find a family of chains which span V - take $e_1, \ldots, e_n$ a basis for V, and take the chain bases $e_i \mapsto f(e_i) \mapsto \ldots \mapsto f^{k_i}(e_i) \mapsto 0$. But this is usually not linearly independent. The following proposition enables us to build a family of linearly independent chains from a family of chains without changing the span.

Proposition 6.16: *Let $f : V \to V$ be a nilpotent operator. Let $C_1, \ldots, C_s$ be a family of chains. There exists a family $B_1, \ldots, B_r$ of linearly independent chains such that* $\mathrm{Span}(C_1 \cup \ldots \cup C_s) = \mathrm{Span}(B_1 \cup \ldots \cup B_r)$.

Proof. To produce $B_1 \cup \ldots \cup B_r$, we apply the following algorithm.

Write C_i to be $v_i \mapsto f(v_i) \mapsto \ldots \mapsto f^{k_i}(v_i)$. Wlog assume that $k_1 \geq k_2 \geq \ldots \geq k_s$. If $C_1, \ldots, C_s$ are linearly independent we are done. Assume therefore that $C_1, \ldots, C_s$ are linearly dependent.

By Lemma 6.15, we must have a linear relation of the form

$$a_1 f^{k_1-1}(v_1) + \ldots + a_s f^{k_s-1}(v_s) = 0$$

Let r be maximal such that $a_r \neq 0$. Then we get

$$f^{k_r-1}(a_1 f^{k_1-k_r}(v) + \ldots + a_{r-1} f^{k_{r-1}-1}(v_{r-1}) + a_r v_r) = 0$$

If $v' = a_1 f^{k_1-k_r}(v) + \ldots + a_{r-1} f^{k_r-1}(v_{r-1}) + a_r v_r = 0$, then we see that $v_r \in \mathrm{Span}(C_1 \cup \ldots \cup C_{r-1})$, so we replace $C_1, \ldots, C_s$ by $C_1, \ldots, C_{r-1}, C_{r+1}, \ldots, C_s$.

If $v' \neq 0$, we have $v_r \in \mathrm{Span}(C_1 \cup \ldots \cup C_{r-1} \cup \{v'\})$ so if C'_r is the chain associated to v', we get $\mathrm{Span}(C_1 \cup \ldots \cup C_r) = \mathrm{Span}(C_1 \cup \ldots \cup C'_r)$. On the other hand $f^{k_r-1}(v') = 0$ so C'_r is strictly shorter than C_r. We replace $C_1, \ldots, C_{r-1}, C_r$ by $C_1, \ldots, C_{r-1}, C'_r$.

We then repeat. Because the number of chains or the length of one of them decreases at each step, this terminates, which means we get a linearly independent set of chains with the same span. □

Finally we get

Proof. (of Theorem 6.12). Start with $e_1, \ldots, e_n$ a basis for V, and let C_i be the chain basis associated with e_i. By Proposition 6.16 above, there exists linearly independent chain bases $B_1, \ldots, B_l$ such that $\mathrm{Span}(B_1 \cup \ldots \cup B_l) = \mathrm{Span}(C_1 \cup \ldots \cup C_n) = V$. Thus V admits a basis $\mathcal{B}$ which is a union of chain bases - thus $[f]_{\mathcal{B}}$ is a of the form $J(0)$. □

6.4 General operators

We now want to prove the Jordan normal form for general operators. We begin with the following

Theorem 6.17: *(Fitting Theorem) Let $f : V \to V$ be an operator on a finite dimensional vector space V. There exist f-invariant subspaces V_N, V_I of V such that $V = V_N \oplus V_I$, the restriction $f|_{V_N}$ is nilpotent, and the restriction $f|_{V_I}$ is invertible.*

Proof. We consider the following sequences of subspaces of V

$$\{0\} \subseteq \mathrm{Ker}(f) \subseteq \mathrm{Ker}(f^2) \subseteq \ldots$$

$$V \supseteq \mathrm{Im}(f) \supseteq \mathrm{Im}(f^2) \supseteq \ldots$$

Since V is finite dimensional, these chains must stabilize, that is, there exists V_N such that for all k large enough, $\ker(f^k) = V_N$, and V_I such that for all k large enough, $\mathrm{Im}(f^k) = V_I$.

Moreover, since $f(\ker(f^{k+1})) \subseteq \ker(f^k)$ we have $f(V_N) \subseteq V_N$, so V_N is f-invariant. Also, since $V_N = \mathrm{Ker}(f^k)$ for some k, we have that $(f|_{V_N})^k = 0$, that is, $f|_{V_N}$ is nilpotent.

On the other hand, $f(\mathrm{Im}(f^k)) = \mathrm{Im}(f^{k+1})$, so $f(V_I) = V_I$ and V_I is f-invariant. Now $f|_{V_I}: V_I \to V_I$, and $\mathrm{Im}(f|_{V_I}) = f(V_I) = V_I$. By the rank-nullity theorem, this means the kernel of $f|_{V_I}$ is empty, that is, $f|_{V_I}$ is invertible.

Let us now show $V = V_N \oplus V_I$. First, we can fix k large enough so that $V_N = \ker(f^k)$ and $V_I = \mathrm{Im}(f^k)$. In particular, $\dim(V_N) + \dim(V_I) = \dim(\ker(f^k)) + \dim(\mathrm{Im}(f^k)) = V$. Thus it is enough to show that $V_I \cap V_N = \{0\}$. If $v \in V_I \cap V_N$, then $f^k(v) = 0$. On the other hand, since f^k is invertible on V_I and $v \in V_I$, we get $v = (f^k)^{-1}(0) = 0$. □

Let us now prove Theorem 6.3.

Proof. We proceed by induction on $\dim V$. If $\dim V = 1$, the matrix of f in any basis is in Jordan form. Assume now the result holds for operators on vector spaces of dimension strictly smaller than $\dim V$.

Since V is a vector space over $\mathbb{C}$, the characteristic polynomial of f has at least one root λ, which is an eigenvalue for f. Hence the operator $g = (f - \lambda \mathrm{id})$ is not invertible (if v is an eigenvector associated to λ we have $g(v) = f(v) - \lambda v = 0$ though $v \neq 0$).

Apply the Fitting decomposition Theorem to g, to get a decomposition $V = V_N \oplus V_I$ such that $g\,|_{V_N}$ is nilpotent and $g\,|_{V_I}$ is invertible. Since g is not invertible, $\dim V_I < \dim V$. By Theorem 6.12, there is a basis for V_N in which the matrix for g is of the form $J(0)$, and by induction hypothesis there is a basis for V_I in which the matrix for g is in Jordan form. Putting together these two bases, we get a basis $\mathcal{B}$ for V in which $[g]_{\mathcal{B}}$ is in Jordan normal form, with a Jordan block $J(0)$ in the top left corner. Now $[f]_{\mathcal{B}} = [g]_{\mathcal{B}} + \lambda I$ so $[f]_{\mathcal{B}}$ is in Jordan normal form (with a Jordan block $J(\lambda)$ in the top left corner). $\square$