

# Jordan normal form

Sebastian Ørsted

December 16, 2017

## Abstract

In these notes, we expand upon the coverage of linear algebra as presented in Thomsen (2016). Namely, we introduce some concepts and results of fundamental importance in the field, but which were not available to us without the greater level of abstraction developed in the current course. Specifically, we introduce concepts like the *Jordan normal form* and *generalized eigenspaces*.

**G**IVEN A FINITE-DIMENSIONAL VECTOR SPACE  $V$  over a field  $K$ , recall that a linear operator  $\theta: V \rightarrow V$  is called **diagonalizable** if  $V$  admits a basis consisting of eigenvectors of  $\theta$ . Thus each element of this basis lies in some eigenspace

$$E(\lambda) = \{v \in V \mid \theta(v) = \lambda v\}$$

corresponding to an eigenvalue  $\lambda$  of  $\theta$  (but all basis elements need not belong to the same  $E(\lambda)$ !).

We note the following alternative formulation of being diagonalizable. It is considered standard knowledge from linear algebra, and we omit the proof.

**Proposition J.1.** *If  $\lambda_1, \lambda_2, \dots, \lambda_r$  are the distinct eigenvalues of  $\theta$ , then  $\theta$  is diagonalizable if and only if  $V$  can be written as the direct sum*

$$V = E(\lambda_1) \oplus E(\lambda_2) \oplus \dots \oplus E(\lambda_r)$$

*of the eigenspaces of  $\theta$ . Furthermore, we may choose a basis for each space  $E(\lambda_i)$  and combine them to a basis for  $V$  in which  $\theta$  has the block diagonal form*

$$\theta = \begin{pmatrix} D_{n_1}(\lambda_1) & & & 0 \\ & D_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & D_{n_r}(\lambda_r) \end{pmatrix},$$

*where  $D_{n_i}(\lambda_i) = \lambda_i \text{id}_{n_i}$  is the  $n_i \times n_i$  diagonal matrix with all diagonal entries equal to  $\lambda_i$ .*

Recall that  $V$  being the **direct sum** of subspaces  $V_1, V_2, \dots, V_k$ , means that any  $v \in V$  can be written *uniquely* as a sum

$$v = v_1 + v_2 + \dots + v_n$$

of vectors with  $v_i \in V_i$ . In this case we write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k = \bigoplus_{i=1}^k V_i.$$

In general, not all operators are diagonalizable; indeed, many matrices, like  $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$  over  $K = \mathbb{R}$ , do not have eigenvalues at all, since its characteristic polynomial  $x^2 + 1 \in \mathbb{R}[x]$  has no roots in  $\mathbb{R}$ . Nevertheless, one of the many consequences of the theory we develop here is that when the field  $K$  is, for instance, the complex numbers, *all* operators  $\theta$  satisfy a slightly weaker form of Proposition J.1: In the definition of  $E(\lambda)$ , we replace the condition that  $\theta(v) = \lambda v$  (or, equivalently,  $(\theta - \lambda \text{id})v = 0$ ) by the requirement that  $(\theta - \lambda \text{id})^n v = 0$  for some  $n \geq 1$  (which is allowed to depend on  $v$ ). In other words, define

$$\bar{E}(\lambda) = \{v \in V \mid (\theta - \lambda \text{id})^n v = 0 \text{ for some } n \geq 1\},$$

the **generalized eigenspace** of  $\theta$  with respect to  $\lambda$ . We note that  $E(\lambda)$  is a vector subspace of  $\bar{E}(\lambda)$  (see Exercise J.1), and that this inclusion can be strict. Note also that we have  $\theta(\bar{E}(\lambda)) \subset \bar{E}(\lambda)$  (see Exercise J.2). It would seem natural to extend the well-known terminology further and refer to  $\lambda$  as a ‘generalized eigenvalue’ if  $\bar{E}(\lambda) \neq 0$ ; however, this terminology turns out to be redundant, because

**Proposition J.2.** *A ‘generalized eigenvalue’ is automatically an eigenvalue,*

hence we simply refer to  $\lambda$  as the corresponding ‘eigenvalue’.

**Proof.** If  $(\theta - \lambda \text{id})^n v = 0$  for some  $v \neq 0$ , then  $0 = \det((\theta - \lambda \text{id})^n) = (\det(\theta - \lambda \text{id}))^n$ , so  $\det(\theta - \lambda \text{id}) = 0$ , hence  $\lambda$  is an eigenvalue.  $\square$

**Example J.3.** The matrix

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

is not diagonalizable over  $\mathbb{C}$ ; the only root of its characteristic polynomial  $(1 - x)^2$  is  $x = 1$ . If it were diagonalizable, we would get from Proposition J.1 that  $\mathbb{C}^2 = E(1)$ , which would imply  $\theta v = v$  for all  $v \in \mathbb{C}^2$ ; this is absurd since  $\theta$  is not the identity matrix. However, the difference

$$\theta - \text{id} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

satisfies  $(\theta - \text{id})^2 = 0$ , hence  $(\theta - \text{id})^2 v = 0$  for all  $v \in \mathbb{C}^2$ . This implies that  $\bar{E}(1) = \mathbb{C}^2$ .  $\circ$

**Example J.4.** We have only defined generalized eigenspaces for vector spaces of finite dimension and will only deal with those outside of this example. However, the definition obviously makes sense in the infinite-dimensional setting as well, and we shall give a particularly pretty example. Let  $V = C^\infty(\mathbb{R})$  be the real vector space of smooth (i.e., infinitely often differentiable) functions  $\mathbb{R} \rightarrow \mathbb{R}$ , and let  $\theta = d/dx: V \rightarrow V$  be the differential operator. The generalized eigenspace  $\bar{E}(0)$  consists of all smooth functions  $f: \mathbb{R} \rightarrow \mathbb{R}$  such

that  $d^n f/dx^n = 0$  for some  $n$ . Calculating antiderivatives one by one, this means that  $d^{n-1} f/dx^{n-1}$  is a constant, hence  $d^{n-2} f/dx^{n-2}$  is a polynomial of degree at most 1, and in general,  $d^{n-k} f/dx^{n-k}$  is a polynomial of degree  $\leq k - 1$ . In particular,  $f = d^0 f/dx^0$  is a polynomial of degree at most  $n - 1$ . We conclude that  $\bar{E}(0) = \mathbb{R}[x]$  consists exactly of all polynomial functions on  $\mathbb{R}$ . This is probably one of the many things that make polynomials interesting for an analyst as well as an algebraist.  $\circ$

In order to state our main result, we need one more definition. Recall the result known by the misleading name, ‘the Fundamental Theorem of Algebra’; it is not fundamental to modern abstract algebra, but received its name from an earlier discipline known simply as ‘algebra’, which was mainly concerned with finding roots of real and complex polynomials. The ‘abstract’ in ‘abstract algebra’ is there for a reason.

**Theorem J.5.** *Every non-constant polynomial in  $\mathbb{C}[x]$  has a root in  $\mathbb{C}$ .*

**Proof.** See, for instance, Exercise 1.54 in Stetkær (2012), Theorem 4.23 in Berg (2013), or Theorem 1.8 in Hatcher (2002).  $\square$

This motivates the following definition: The field  $K$  is called **algebraically closed** if it satisfies the same theorem, that is, if every non-constant polynomial in  $K[x]$  has a root in  $K$ . Thus  $\mathbb{C}$  is algebraically closed, while  $\mathbb{R}$  and  $\mathbb{Q}$  are not. No finite field is algebraically closed (see Exercise J.3). There are many fields satisfying this definition, but very few can be constructed concretely, and most of our results here will be developed with  $\mathbb{C}$  in mind.

Our main theorem states that for algebraically closed fields, a basis for the vector space  $V$  can be chosen in which  $\theta$  has a representation given in terms of blocks of the form

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ 0 & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix},$$

the  $n \times n$  square matrix with  $\lambda$ s along the diagonal and ones immediately above it. This block is known as the **Jordan block** of order  $n$  with respect to  $\lambda$ .

**Theorem J.6.** *If the field  $K$  is algebraically closed and  $\theta: V \rightarrow V$  is a linear operator on a finite-dimensional  $K$ -vector space, and if  $\lambda_1, \lambda_2, \dots, \lambda_r$  are its distinct eigenvalues, then*

$$V = \bar{E}(\lambda_1) \oplus \bar{E}(\lambda_2) \oplus \dots \oplus \bar{E}(\lambda_r)$$

*is the direct sum of the generalized eigenspaces. Restricting  $\theta$  to an operator  $\theta: \bar{E}(\lambda_i) \rightarrow \bar{E}(\lambda_i)$  on one of these spaces, we may choose a basis for  $\bar{E}(\lambda_i)$  in which  $\theta$  has the block diagonal form*

$$B(\lambda_i) = \begin{pmatrix} J_{n_1}(\lambda_i) & & 0 \\ & J_{n_2}(\lambda_i) & \\ & & \ddots \\ 0 & & & J_{n_k}(\lambda_i) \end{pmatrix}$$

consisting of Jordan blocks, where  $n_1 \geq n_2 \geq \dots \geq n_k \geq 1$  are integers (depending on  $i$ ). Combining these bases to a basis for all of  $V$ ,  $\theta$  thus has the matrix representation

$$\theta = \begin{pmatrix} B(\lambda_1) & & & 0 \\ & B(\lambda_2) & & \\ & & \ddots & \\ 0 & & & B(\lambda_r) \end{pmatrix}.$$

This representation, called the **Jordan normal form** of  $\theta$ , is unique up to reordering of the blocks.

The Jordan normal form is named after the French mathematician Camille Jordan (1838–1922). The proof of this theorem will occupy the next two sections.

Since we have  $E(\lambda_i) \subset \bar{E}(\lambda_i)$ , we see that  $\theta$  is diagonalizable if and only if equality holds. Also recall from linear algebra that two different matrix representations  $A$  and  $B$  of the same linear map are *similar*, meaning that  $S^{-1}AS = B$  for some invertible matrix  $S$ . In other words, the theorem shows that *any* square matrix over an algebraically closed field is similar to some matrix on Jordan normal form.

Looking at the definition of  $\bar{E}(\lambda)$ , we see it is not immediately obvious how to calculate it; it is equal to the union of the kernels  $\text{Ker}((\theta - \lambda \text{id})^k)$  for all  $k$ , but do we have to calculate infinitely many powers of  $\theta - \lambda \text{id}$  and their kernels? Fortunately, one is enough:

**Corollary J.7.** *Let  $N$  be the algebraic multiplicity of  $\lambda$  as an eigenvalue of  $\theta$ . Then*

$$\bar{E}(\lambda) = \{v \in V \mid (\theta - \lambda \text{id})^N v = 0\} = \text{Ker}((\theta - \lambda \text{id})^N).$$

**Proof.** Exercise J.8(iii). □

The next example provides the general algorithm for computing the Jordan normal form.

**Example J.8.** Let us consider the complex matrix

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

Its characteristic polynomial is given by

$$\chi(x) = x^4 - 2x^3 + x^2 = x^2(x-1)^2,$$

hence the eigenvalues are  $\lambda = 0$  and  $\lambda = 1$ , both with algebraic multiplicity 2. According to Corollary J.7, the generalized eigenspaces are therefore given by  $\bar{E}(0) = \text{Ker}(\theta^2)$  and  $\bar{E}(1) = \text{Ker}((\theta - \text{id})^2)$ . As the reader can verify using their existing knowledge of linear algebra, we therefore have

$$\bar{E}(0) = \mathbb{C}(1, 0, 0, 0) + \mathbb{C}(0, 1, 0, 0)$$

and

$$\bar{E}(1) = \mathbb{C}(1, 0, 1, 0) + \mathbb{C}(0, 1, 0, 1).$$

These four vectors are definitely a basis for  $\mathbb{C}^4$ . The matrix representation of  $\theta$  in this basis is given by  $S^{-1}\theta S$ , where

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

This yields

$$S^{-1}\theta S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which has Jordan normal form, consisting of the two Jordan blocks  $J_2(0)$  and  $J_2(1)$ .  $\circ$

### Proof of existence of the Jordan normal form

**F**OR ANY PRINCIPAL IDEAL DOMAIN  $R$  (like  $K[x]$ !), recall that if  $d$  is a greatest common divisor of  $x, y \in R$ , then there exist  $\lambda, \mu \in R$  with  $d = \lambda x + \mu y$ . This statement can be generalized to any finite collection of elements. Given  $x_1, x_2, \dots, x_n$  in any commutative ring  $R$  (not necessarily a principal ideal domain), an element  $d \in R$  is called a **greatest common divisor** if it is a common divisor and if any other common divisor divides  $d$ . In a unique factorization domain, any finite collection of elements has a greatest common divisor (see Exercise J.4). A finite collection of elements is called **coprime** if 1 is a greatest common divisor.

**Lemma J.9.** *Given elements  $x_1, x_2, \dots, x_n$  in a principal ideal domain  $R$  with greatest common divisor  $d$ , there exist  $\mu_1, \mu_2, \dots, \mu_n \in R$  such that*

$$d = \mu_1 x_1 + \mu_2 x_2 + \dots + \mu_n x_n.$$

*Proof.* Exercise J.5.  $\square$

**Warning J.10.** We should warn the reader that we use the convention that the characteristic polynomial of a matrix  $\theta$  is given by  $\chi(x) = \det(x\text{id} - \theta)$ , in contrast to Thomsen (2016) where it is given by  $\det(\theta - x\text{id})$ . The difference is a factor of  $(-1)^n$ , where  $n$  is the dimension of the vector space. Our convention has the advantage that  $\chi$  becomes a monic polynomial.  $\triangle$

**Proof that  $V$  is the direct sum of the generalized eigenspaces.**

Let  $\chi(x) = \det(x\text{id} - \theta)$  denote the characteristic polynomial of  $\theta$ . The Cayley–Hamilton theorem (see, for instance, *ibid.*, Sætning 15.11) tells us that  $\chi(\theta) = 0$ . Because the base field  $K$  is algebraically closed, we may factorize  $\chi$  in the form

$$\chi(x) = (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \dots (x - \lambda_r)^{n_r},$$

where the  $\lambda_i$  are the distinct eigenvalues of  $\theta$  and all  $n_i \geq 1$ . We first claim that in fact  $\bar{E}(\lambda_i) = V_i$ , where

$$V_i = \{v \in V \mid (\theta - \lambda_i \text{id})^{n_i} v = 0\},$$

so that in the definition of  $\overline{E}(\lambda_i)$ , the multiplicity  $n_i$  always works in the place of  $n$  (note that we cannot appeal to Corollary J.7, since this result relies on the existence of the Jordan normal form). It is clear that  $V_i \subset \overline{E}(\lambda_i)$ , so let us prove the other inclusion. If  $v \in \overline{E}(\lambda_i)$ , then  $(\theta - \lambda_i)^n v = 0$  for some  $n$ , and we might as well assume that  $n \geq n_i$ . Now  $(x - \lambda_i)^{n_i}$  is a greatest common divisor of  $\chi(x)$  and  $(x - \lambda_i)^n$ , thus we may find  $p(x), q(x) \in K[x]$  such that

$$(x - \lambda_i)^{n_i} = p(x)\chi(x) + q(x)(x - \lambda_i)^n.$$

Substituting  $\theta$  for  $t$  and using  $\chi(\theta) = 0$ , we find that

$$(\theta - \lambda_i \text{id})^{n_i} = p(\theta)\chi(\theta) + q(\theta)(\theta - \lambda_i \text{id})^n = q(\theta)(\theta - \lambda_i \text{id})^n.$$

Applying this to  $v$ , we get  $(\theta - \lambda_i \text{id})^{n_i} v = 0$ , so  $v \in V_i$ .

For  $i = 1, 2, \dots, n$ , we now define

$$f_i(x) = \frac{\chi(x)}{(x - \lambda_i)^{n_i}} = (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \cdots (x - \lambda_{i-1})^{n_{i-1}} (x - \lambda_{i+1})^{n_{i+1}} \cdots (x - \lambda_r)^{n_r}$$

and note that  $f_1, f_2, \dots, f_r$  are coprime polynomials (why?). Hence Lemma J.9 implies that there exist  $\mu_1, \mu_2, \dots, \mu_r \in K[x]$  such that

$$1 = f_1 \mu_1 + f_2 \mu_2 + \cdots + f_r \mu_r.$$

Substituting  $\theta$  for  $x$ , we get

$$\text{id} = f_1(\theta)\mu_1(\theta) + f_2(\theta)\mu_2(\theta) + \cdots + f_r(\theta)\mu_r(\theta).$$

Applying this to any  $v \in V$ , we find that

$$v = f_1(\theta)\mu_1(\theta)v + f_2(\theta)\mu_2(\theta)v + \cdots + f_r(\theta)\mu_r(\theta)v.$$

In particular, any  $v \in V$  lies in the sum  $\text{Im}(f_1(\theta)) + \cdots + \text{Im}(f_r(\theta))$  of the images of the  $f_i(\theta)$ , hence

$$V = \text{Im}(f_1(\theta)) + \text{Im}(f_2(\theta)) + \cdots + \text{Im}(f_r(\theta)).$$

Also, note that  $0 = \chi(\theta) = (\theta - \lambda_i \text{id})^{n_i} f_i(\theta)$ , which implies  $\text{Im}(f_i(\theta)) \subset V_i$ . Thus we also have

$$V = V_1 + V_2 + \cdots + V_r.$$

To prove that this sum is direct, it suffices by Exercise J.6 to show that

$$V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_r) = 0$$

for all  $i = 1, 2, \dots, r$ . So let  $v$  lie in this intersection. Because  $v \in V_i$ , we have  $(\theta - \lambda_i \text{id})^{n_i} v = 0$ , while the fact that  $v$  lies in the set in parentheses implies that  $f_i(\theta)v = 0$ . Now  $f_i(x)$  and  $(x - \lambda_i)^{n_i}$  are coprime polynomials, so we may find  $u(x), v(x) \in K[x]$  such that  $1 = u(x)f_i(x) + v(x)(x - \lambda_i)^{n_i}$ . Substituting  $\theta$  for  $x$ , we have

$$\text{id} = u(\theta)f_i(\theta) + v(\theta)(\theta - \lambda_i \text{id})^{n_i}.$$

Now applying this to  $v$ , both terms become zero, and we have  $v = 0$ .  $\square$

In order to prove the existence of a basis for  $\bar{E}(\lambda_i)$  in which the matrix has Jordan normal form, we note that  $\theta - \lambda_i \text{id}$  is a nilpotent operator on  $\bar{E}(\lambda_i)$ , since, according to the above proof,  $(\theta - \lambda_i \text{id})^{n_i} = 0$  on  $\bar{E}(\lambda_i)$ . Hence to finish the proof, all we need to do is to apply the following proposition to  $\theta - \lambda_i \text{id}$  on each subspace  $\bar{E}(\lambda_i)$ .

**Proposition J.11.** *Given a nilpotent operator  $\theta: V \rightarrow V$  on a finite-dimensional  $K$ -vector space  $V$ , there exists a basis for  $V$  in which  $\theta$  has block diagonal form*

$$\theta = \begin{pmatrix} J_{n_1}(0) & & & 0 \\ & J_{n_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(0) \end{pmatrix}$$

for suitable  $n_1 \geq n_2 \geq \dots \geq n_k \geq 1$ .

For the proof, we define a **Jordan chain** to be a chain of elements of the form

$$v, \theta(v), \theta^2(v), \dots, \theta^{p-1}(v)$$

where  $\theta^i(v) \neq 0$  for  $i = 0, 1, 2, \dots, p-1$ , but  $\theta^p(v) = 0$ . Certainly, Jordan chains exist in  $V$  because  $\theta$  is nilpotent (those worried about allowing empty Jordan chains may assume  $V \neq 0$ ). Let us note that a Jordan chain is automatically linearly independent: If  $\sum_{i=0}^{p-1} \alpha_i \theta^i(v) = 0$ , suppose that there is some  $j$  with  $\alpha_j \neq 0$ , and assume that  $j$  is minimal with this property, so that  $\alpha_0 = \alpha_1 = \dots = \alpha_{j-1} = 0$ . Then we have

$$0 = \theta^{p-1-j} \sum_{i=0}^p \alpha_i \theta^i(v) = \alpha_0 \theta^{p-1-j}(v) + \alpha_1 \theta^{p-1-j+1}(v) + \dots + \alpha_j \theta^{p-1}(v) = \alpha_j \theta^{p-1}(v),$$

showing that  $\alpha_j = 0$ , a contradiction. If  $U = \text{span}\{\theta^{p-1}(v), \theta^{p-2}(v), \dots, \theta(v), v\}$  is the span of some Jordan chain, then the chain is a basis for  $U$ , and the matrix representation of  $\theta$  on  $U$  becomes

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

which just happens to be  $J_p(0)$ .

A Jordan chain as above is called **maximal** if  $v$  does not lie in the image  $\theta(V)$  of  $\theta$ ; this means that we cannot extend the chain backwards. Note that any  $v \in V$  is contained in some maximal Jordan chain (why?). For the proof, because of the above matrix representation, it suffices to prove that  $V$  is the direct sum  $V = U_1 \oplus \dots \oplus U_k$  of subspaces  $U_k$  each spanned by some maximal Jordan chain.

**Proof of Proposition J.11.** We shall argue by induction on  $\dim V$ , the case  $\dim V = 0$  being empty. Because  $\theta$  is nilpotent,  $\theta(V)$  must be a proper subspace of  $V$  (why?). By induction, we may write  $\theta(V) = W_1 \oplus \cdots \oplus W_r$  as a direct sum of subspaces each spanned by some Jordan chain which is maximal in  $\theta(V)$ . Choose a generator in  $W_i$  for each such chain and write it as  $\theta(u_i)$  because it lies in the image  $\theta(V)$ . Then the chain has the form  $\theta(u_i), \theta^2(u_i), \dots, \theta^{p_i-1}(u_i)$ . We let  $U_i = W_i \oplus Ku_i$  for  $i = 1, 2, \dots, r$ .

Now  $\theta^{p_1}(u_1), \dots, \theta^{p_r}(u_r)$  form a basis for  $\theta(V) \cap \text{Ker } \theta$  (why?), and we may extend this to a basis for all of  $\text{Ker } \theta$  by adding elements which we denote  $u_{r+1}, u_{r+2}, \dots, u_k$ . Each of these forms a maximal Jordan chain consisting of one element. We let  $U_i = Ku_i$  for  $i = r+1, \dots, k$ .

We claim that  $V = U_1 \oplus \cdots \oplus U_k$ , or, in other words, that

$$u_1, \theta(u_1), \dots, \theta^{p_1-1}(u_1), \dots, u_k, \theta(u_k), \dots, \theta^{p_k-1}(u_k),$$

is a basis for  $V$ , where we define  $p_{r+1} = p_{r+2} = \cdots = p_k = 1$ . To check that  $V = U_1 + \cdots + U_k$ , let  $v \in V$  be arbitrary. Then  $\theta(v) \in \theta(V) = W_1 \oplus \cdots \oplus W_r$ , and because of how the  $u_i$  were chosen, we can find a  $u \in U_1 + \cdots + U_r$  such that  $\theta(u) = \theta(v)$ . Thus  $\theta(v - u) = 0$ , so that  $v - u \in \text{Ker } \theta \subset U_1 + \cdots + U_k$ . Therefore,  $v = u + (v - u)$  lies in  $U_1 + \cdots + U_k$  as well. Now to check that the above vectors constitute a basis, it is enough to count that their numbers are equal to the dimension of  $V$ . From linear algebra,

$$\dim V = \dim \theta(V) + \dim \text{Ker } \theta.$$

But  $\theta(V)$  had a basis consisting of  $\theta(u_i), \theta^2(u_i), \dots, \theta^{p_i-1}(u_i)$  for  $i = 1, 2, \dots, r$ , so

$$\dim \theta(V) = (p_1 - 1) + (p_2 - 1) + \cdots + (p_r - 1).$$

By choice,  $\theta^{p_1-1}(u_1), \dots, \theta^{p_r-1}(u_r)$  together with  $u_{r+1}, \dots, u_k$  are a basis for  $\text{Ker } \theta$ , hence  $\dim \text{Ker } \theta = k$ . Thus

$$\begin{aligned} \dim V &= (p_1 - 1) + (p_2 - 1) + \cdots + (p_r - 1) + k \\ &= p_1 + p_2 + \cdots + p_r + (k - r) \\ &= p_1 + p_2 + \cdots + p_r + p_{r+1} + \cdots + p_k. \end{aligned}$$

This is exactly the number of elements in the claimed basis, which proves that  $V = U_1 \oplus \cdots \oplus U_k$ . Reorganizing the  $U_i$  according to dimension  $n_i = \dim U_i$ , we can get the desired form of  $\theta$  where  $n_1 \geq \cdots \geq n_k \geq 1$ .  $\square$

**Remark J.12.** Notice that the assumption that  $K$  was algebraically closed was only used to argue that the characteristic polynomial of  $\theta$  could be factorized as the product

$$\chi(x) = (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \cdots (x - \lambda_r)^{n_r}$$

of linear polynomials. So in other words, the conclusion of the theorem holds for operators over any field  $K$  as long as the characteristic polynomial can be factorized this way. In particular, this is the case for  $K = \mathbb{R}$  if  $\chi$  has no non-real roots. Note that the uniqueness proof of the next section does not rely on the assumption of algebraic closure, either.  $\triangle$

## Proof of uniqueness of the Jordan normal form

TO PROVE THE UNIQUENESS STATEMENT of Theorem J.6, note first that if  $\theta$  is represented by a matrix of this form, then the size of the block  $B(\lambda_i)$  measures the dimension of  $\bar{E}(\lambda_i)$ . Thus the size of each such block is uniquely determined. It thus suffices to prove that the internal structure of the block  $B(\lambda_i)$  is uniquely determined, or, in other words, that the matrix representation of  $\theta$  on  $\bar{E}(\lambda_i)$  is uniquely determined for each  $i$ . Therefore, in proving uniqueness, we may as well assume that  $V = \bar{E}(\lambda)$  for some  $\lambda$ , so that  $\theta - \lambda \text{id}$  is nilpotent on all of  $V$ . Also, replacing  $\theta$  by  $\theta - \lambda \text{id}$ , we may as well assume that  $\lambda = 0$ , so that  $\theta$  itself is a nilpotent operator on  $V$ .

With these assumptions, suppose that

$$\theta = \begin{pmatrix} J_{n_1}(0) & & & 0 \\ & J_{n_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(0) \end{pmatrix}$$

where  $n_1 \geq n_2 \geq \dots \geq n_k \geq 1$ . All we need is to prove that the integers  $n_i$  are uniquely determined. We claim that

$$\dim \text{Ker } \theta^k - \dim \text{Ker } \theta^{k-1} = \#\{i \mid k \leq n_i\}. \quad (1)$$

for all  $k \geq 0$ . We leave it to the reader to verify that this uniquely determines the sequence  $n_1 \geq n_2 \geq \dots \geq n_k \geq 1$ .

To prove the claim, note that

$$\theta^k = \begin{pmatrix} J_{n_1}(0)^k & & & 0 \\ & J_{n_2}(0)^k & & \\ & & \ddots & \\ 0 & & & J_{n_k}(0)^k \end{pmatrix},$$

and that each block  $J_{n_i}(0)^k$  is given by

$$J_{n_i}(0)^k = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ & & & \ddots & \vdots \\ & & \ddots & & 1 \\ 0 & & & & \vdots \\ & & & & 0 \end{pmatrix},$$

which is the  $n_i \times n_i$  matrix which is everywhere zero except for a diagonal line of ones from the  $(1, k)$ th to the  $(k, 1)$ th cell. In particular,  $J_{n_i}(0)^k = 0$  if and only if  $k \geq n_i$ . For each such matrix  $J_{n_i}(0)^k$ , the kernel has a basis consisting of the basis vectors corresponding to the first  $k$  columns. So the difference in dimension between  $\text{Ker}(J_{n_i}(0))^k$  and  $\text{Ker}(J_{n_i}(0))^{k-1}$  is 1 if  $k \leq n_i$  and 0 otherwise. Adding the dimensions of the kernels of each block, we arrive at the claim (1).

## References

- Berg, C. (2013). *Complex Analysis*. Matematisk afdeling, Københavns Universitet.
- Hatcher, A. (2002). *Algebraic Topology*. Cambridge University Press. URL: [www.math.cornell.edu/~hatcher/AT/AT.pdf](http://www.math.cornell.edu/~hatcher/AT/AT.pdf).
- Stetkær, H. (2012). *Følger og rækker af funktioner*. Lecture notes for the course *Mathematical analysis 2*.
- Thomsen, J. F. (2016). *Lineær algebra*. Lecture notes for Linear Algebra at Aarhus University.

## Exercises

**J.1.** Prove that for any  $\lambda \in K$ ,  $E(\lambda)$  is a vector subspace of  $\bar{E}(\lambda)$ .

**J.2.** Prove that  $\theta(\bar{E}(\lambda)) \subset \bar{E}(\lambda)$ . We usually formulate this by saying that  $\bar{E}(\lambda)$  is **invariant** under  $\theta$ .

**J.3.** Prove that no finite field  $\mathbb{F}_q$  is algebraically closed. (*Hint:* Recall that  $x^q = x$  for all  $x \in \mathbb{F}_q$ .)

**J.4.** Prove that in a unique factorization domain, a greatest common divisor exists between any finite number of elements.

**J.5.** Prove Lemma J.9.

**J.6.** Prove that a vector space  $V$  is the direct sum  $V = V_1 \oplus \cdots \oplus V_n$  of subspaces if and only if  $V = V_1 + \cdots + V_n$  and

$$V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_n) = 0$$

for all  $i = 1, 2, \dots, n$ .

**J.7.** Calculate the Jordan normal form of the following matrices:

$$(a) \begin{pmatrix} 4 & 0 & 0 \\ 2 & 1 & 3 \\ 5 & 0 & 4 \end{pmatrix}, \quad (b) \begin{pmatrix} 5 & 4 & 3 \\ -1 & 0 & -3 \\ 1 & -2 & 1 \end{pmatrix}, \quad (c) \begin{pmatrix} 9 & 7 & 3 \\ -9 & -7 & -4 \\ 4 & 4 & 4 \end{pmatrix}.$$

**J.8.** If  $\theta$  is given on Jordan normal form as in Theorem J.6, verify the following:

- (i) The algebraic multiplicity of  $\lambda_i$  is the size of the block  $B(\lambda_i)$ .
- (ii) The geometric multiplicity of  $\lambda_i$  is the number of Jordan blocks in  $B(\lambda_i)$ .
- (iii) The generalized eigenspace is given by

$$\bar{E}(\lambda_i) = \{v \in V \mid (\theta - \lambda_i \text{id})^{n_i} = 0\} = \text{Ker}((\theta - \lambda_i \text{id})^{n_i}),$$

where  $n_i$  is the algebraic multiplicity of  $\lambda_i$ ; in other words, in the definition of  $\bar{E}(\lambda_i)$ ,  $n = n_i$  always suffices.

**J.9.** Generalize the results of Example J.4 by finding  $\bar{E}(\lambda)$  for any  $\lambda \in \mathbb{R}$ . Also try replacing  $\theta = d/dx$  by  $\theta = d^2/dx^2$ .

**J.10.** The Cayley–Hamilton theorem states that a linear operator  $\theta: V \rightarrow V$  on a finite-dimensional vector space is annihilated by its characteristic polynomial  $\chi$ , meaning that  $\chi(\theta) = 0$ . However, there may very well be non-zero polynomials of smaller degree with the same property.

- (i) Verify that  $I = \{f \in K[x] \mid f(\theta) = 0\}$  is an ideal of  $K[x]$ . Deduce that  $I$  has a unique monic generator  $\mu$ , called the **minimal polynomial** of  $\theta$ . Note that  $\mu$  divides  $\chi$ .
- (ii) What are the roots of  $\mu$ ?

- (iii) Can you deduce  $\mu$  from the Jordan normal form of  $\theta$ ?
- (iv) Under what circumstances do we have  $\chi = \mu$ ?
- (v) What is the minimal polynomial of a diagonalizable operator?

**J.11.** Prove that the determinant and trace of a linear map on a finite-dimensional vector space are well-defined in the sense that they do not depend on the choice of matrix representation. (Recall that the **trace**  $\text{tr}(A)$  of a square matrix is the sum of its diagonal entries and that  $\text{tr}(AB) = \text{tr}(BA)$  for all  $A, B$ .)

**J.12.** Let  $\theta: V \rightarrow V$  be an operator on a finite-dimensional vector space over the algebraically closed field  $K$ . Let  $\lambda_1, \lambda_2, \dots, \lambda_r$  be the distinct eigenvalues of  $\theta$  with algebraic multiplicities  $n_1, n_2, \dots, n_r$ , respectively. Prove that the determinant and trace of  $\theta$  are given by (cf. Exercise J.11)

$$\det(\theta) = \lambda_1^{n_1} \lambda_2^{n_2} \cdots \lambda_r^{n_r}$$

and

$$\text{tr}(\theta) = n_1 \lambda_1 + n_2 \lambda_2 + \cdots + n_r \lambda_r.$$

In other words, the determinant and trace are the product and sum, respectively, of the eigenvalues, *counted with multiplicity*.

**J.13.** Let  $\theta: V \rightarrow V$  be a linear operator on a finite-dimensional vector space over the algebraically closed field  $K$ .

- (i) Prove that there exists a *unique* decomposition  $\theta = D + P$  of  $\theta$  as a sum of a diagonal matrix  $D$  and a nilpotent matrix  $N$ . This is called the **additive Jordan decomposition**.
- (ii) Prove that if  $\theta$  is nonsingular, it can be written uniquely as a product  $\theta = DU$  of a diagonalizable matrix  $D$  and a **unipotent** matrix  $U$ , meaning that  $U$  is the sum of the identity and a nilpotent matrix. This is called the **multiplicative Jordan decomposition**.

**J.14.**

- (i) Prove that over an algebraically closed field of characteristic  $p > 0$ , some positive power of any matrix is diagonalizable.
- (ii) Prove the same statement over finite fields.

**J.15.** The **exponential** of a linear operator  $\theta: V \rightarrow V$  on a finite-dimensional real or complex vector space is defined by

$$\exp(\theta) = \sum_{n=0}^{\infty} \frac{1}{n!} \theta^n.$$

It is not immediately obvious that the sum always converges, and in this exercise, we give a proof using the Jordan normal form. Because of this, we assume that the base field is  $\mathbb{C}$ .

- (i) Argue that the sum converges for diagonalizable and nilpotent matrices.
- (ii) Prove that if  $\exp(A)$  and  $\exp(B)$  converge for two matrices  $A$  and  $B$  satisfying  $AB = BA$ , then  $\exp(A + B)$  converges to  $\exp(A)\exp(B)$ .

- (iii) Combine the two above statements with the existence of the Jordan normal form to prove that  $\exp(\theta)$  exists for all operators  $\theta$ , and write down an explicit expression for it.

**J.16.** In this note, we proved the existence of the Jordan normal form as a corollary to Cayley–Hamilton, but it is also possible to give independent proofs. Show how one can then derive Cayley–Hamilton as a corollary to the existence of the Jordan normal form. (The proof becomes universal once we know that *any* field can be embedded inside an algebraically closed field. The smallest algebraically closed field containing a given field  $K$  is called the **algebraic closure** of  $K$  and is written  $\bar{K}$ . For instance,  $\mathbb{C}$  is the algebraic closure of  $\mathbb{R}$ . Existence and uniqueness (up to isomorphism) of the algebraic closure can be proved using the Axiom of Choice.)