Supplement on diagonalization, triangularization and the Jordan normal form

Michael Müger (University of Nijmegen)

June 1, 2026

Abstract

We give (i) an alternative proof to Theorem 6.16 in [4], (ii) a self-contained discussion of triangularizability of linear maps (with a nice proof using dual spaces and transposes of maps) and (iii) give a construction of the Jordan normal form that emphasizes the role of nilpotence and assumes no non-trivial results from the algebra of polynomials. (Yet, in an appendix, we give a version of the approach using polynomial algebra with efficient proofs of the needed prerequisites.)

Contents

1 Alternative proof of Theorem 6.16

Recall that a linear map $T : V \to V$ on a finite dimensional vector space V is called diagonalizable if there exists an ordered basis β for V such that the matrix $[T]_\beta$ is diagonal. This is equivalent to β consisting of eigenvectors of T.

If the vector space is an inner product space, it is natural to want the basis β diagonalizing T to be orthonormal. Whether this is possible turns out to have to do with the adjoint T^*. Recall that T is self-adjoint if $T = T^*$ and normal if $TT^* = T^*T$.

Theorem 1 *Let T be a linear map on a finite-dimensional inner product space V. Then the following are equivalent:*

(i) V admits an orthonormal basis consisting of eigenvectors of T.

(ii) T is normal (in case $\mathbb{F} = \mathbb{C}$) or self-adjoint (when $\mathbb{F} = \mathbb{R}$).

The proofs given in [4] for the implication (ii)⇒(i) use Schur's triangularization Theorem 6.14. I find the proof given below much more natural since it only uses the easy results of Theorem 6.15 (and straightforwardly extends to infinitely many dimensions, provided one makes some additional assumptions). Also the treatment of triangularization in [4] is unsatisfactory: In the 5th edition it is reduced to an exercise in Chapter 5 that was not assigned in LinAlgA. In the 4th edition one finds a self-contained proof of Theorem 6.14, but only for inner product spaces, while an analogous result holds in general. See Section 4 below.

Proof. (i)⇒(ii) We restate the argument from the book: Let β be an orthonormal basis such that the matrix $A = [T]_\beta$ is diagonal. Then also the adjoint matrix A^* (defined by $(A^*)_{ij} = \overline{A_{ji}}$) is diagonal and therefore commutes with A: $AA^* = A^*A$. Thus

$$[TT^*]_\beta = [T]_\beta[T^*]_\beta = [T]_\beta([T]_\beta)^* = AA^* = A^*A = ([T]_\beta)^*[T]_\beta = [T^*]_\beta[T]_\beta = [T^*T]_\beta.$$

(The third and sixth equalities are from Theorem 6.10 in the book, which requires β to be orthonormal.) Since linear maps with the same matrix representation are equal, we have $TT^* = T^*T$, thus T is normal. If $\mathbb{F} = \mathbb{R}$, all eigenvalues of T are real, thus the diagonal matrix A is real-valued, implying $A^* = A$, so that $T^* = T$, thus T is self-adjoint.

(ii)⇒(i) Case $\mathbb{F} = \mathbb{C}$. Let T be normal. Let $\{\lambda_1, \dots, \lambda_n\}$ be the eigenvalues of T without repetition. (Thus $\lambda_i \neq \lambda_j$ if $i \neq j$.) For $i \in \{1, \dots, n\}$ let

$$E_{\lambda_i} = N(T - \lambda_i \mathbf{1}) = \{x \in V \mid Tx = \lambda_i\, x\}$$

be the eigenspace corresponding to λ_i. Each E_{λ_i} is a linear subspace of V, thus an inner product space. Thus we can find orthonormal bases (ONB) β_i for all E_{λ_i}. If $i \neq j$ and $x \in \beta_i$, $y \in \beta_j$ then normality of T and Theorem 6.15(d) imply $\langle x, y\rangle = 0$. Thus $\beta = \beta_1 \cup \dots \cup \beta_n$ is an orthonormal set, and it is an ONB for the linear subspace

$$W = E_{\lambda_1} + \dots + E_{\lambda_n} = \{x_1 + \dots + x_n \mid x_i \in E_{\lambda_i}\ \forall i\} \subseteq V.$$

Thus if we prove $W = V$, we have found an orthonormal basis β for V consisting of eigenvectors of T.

Assume $W \subsetneq V$. If $x \in E_{\lambda_i}$ then $Tx = \lambda_i x$, so that [4, Theorem 6.15(c)] gives $T^*x = \overline{\lambda_i}\, x$. Thus each subspace E_{λ_i} is T^*-invariant, and the same holds for the sum W of these spaces, i.e. $T^*W \subseteq W$.

Let now $x \in W^\perp$, $y \in W$. Then

$$\langle Tx, y \rangle = \langle x, T^*y \rangle = 0, \tag{1}$$

since $x \in W^\perp$ and, as just proven, $T^*y \in W$. Since (1) holds for all $y \in W$, we have proven $Tx \in W^\perp$ for all $x \in W^\perp$. Thus $W^\perp$ is T-invariant: $TW^\perp \subseteq W^\perp$.

Since we assume $W \subsetneq V$, we have $\dim W < \dim V$, thus $\dim W^\perp = \dim V - \dim W \geq 1$. Since $\mathbb{F} = \mathbb{C}$, the restriction $T_{W^\perp}$ has an eigenvector $x \in W^\perp \backslash \{0\}$ with eigenvalue λ. Thus $Tx = \lambda x$. Since the set $\{\lambda_1, \dots, \lambda_n\}$ contains all eigenvalues of T, we must have $\lambda = \lambda_i$ for some $i \in \{1, \dots, n\}$. This implies $x \in E_{\lambda_i}$ and therefore $x \in W$. Since also $x \in W^\perp$, we conclude $x \in W \cap W^\perp$, but we have seen that this intersection is $\{0\}$. Thus $x = 0$, but this contradicts $x \in W^\perp \backslash \{0\}$. This contradiction proves that $W \subsetneq V$ is impossible, so that $W = V$. As explained above, this gives the diagonalizability of T.

Case $\mathbb{F} = \mathbb{R}$. Let $T : V \to V$ be self-adjoint: $T = T^*$. Again, let $\{\lambda_1, \dots, \lambda_n\}$ be the eigenvalues of $\mathbb{R}$, clearly contained in $\mathbb{R}$ since we are working over this field. By (b) of [4, Lemma in Section 6.4], there really are real eigenvalues (which need not be true if T is not self-adjoint), and the sum of their multiplicities equals $\dim V$. Now exactly the same proof as above proves $E_{\lambda_1} + \dots + E_{\lambda_n} = V$, from which diagonalizability of V follows as above. ■

2 Beyond diagonalizability: Triangularizations

We know from [4, Theorem 5.1] that a linear map $T : V \to V$, where V is finite-dimensional, is diagonalizable if and only if V has a basis consisting of eigenvectors of V. Since the eigenspaces $E_\lambda = N(T - \lambda \mathbf{1})$ are linearly independent[1], this is equivalent to $\sum_\lambda \dim E_\lambda = \dim V$. Since $E_\lambda \subseteq V$ is a T-invariant subspace, and since T acts on E_λ by multiplication by λ, the characteristic polynomial of T_{E_λ} is just $t \mapsto (\lambda - t)^{\dim E_\lambda}$. By [4, Theorem 5.20], the characteristic polynomial of T_{E_λ} divides the characteristic polynomial P_T of T (in the sense that $P_T(t) = (\lambda - t)^{\dim E_\lambda} \cdot g$, where g is a certain polynomial). Thus λ is a zero of P_T of multiplicity at least $\dim E_\lambda$ or, in other words $\dim E_\lambda \leq m(\lambda)$, where $m(\lambda)$ is the multiplicity of λ as a zero of the characteristic polynomial of T, i.e. the largest m for which $(t - \lambda)^m$ divides P_T. (Recall that $\dim E_\lambda$ and $m(\lambda)$ are called the geometric and algebraic, respectively, multiplicities of the eigenvalue λ).

On the other hand, we have the inequality $\sum_\lambda m(\lambda) \leq \deg P_T = \dim V$. Thus in order for T to be diagonalizable, it must satisfy $\sum_\lambda m(\lambda) = \dim V$ and $\dim E_\lambda = m(\lambda)$ for each λ. The first condition means that P_T has as many zeros, taking their multiplicities into account, as permitted by its degree $\dim V$. In other words, P_T splits into linear factors:

$$P_T(t) = (-1)^{\dim V} \prod_\lambda (t - \lambda)^{m(\lambda)}. \tag{2}$$

The simplest example where this fails is the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, whose characteristic

[1] By this we mean that if $\{\lambda_1, \dots, \lambda_n\}$ are mutually distinct eigenvalues and $x_1 + \dots + x_n = 0$, where $x_i \in E_{\lambda_i}$ for each i, then all x_i are zero. We don't prove this here since it is a special case of Proposition 20 proven below.

polynomial is t^2+1. Thus over the real field $\mathbb{R}$, A has no eigenvalues. On the other hand P_T always factors into linear factors if the field $\mathbb{F}$ is algebraically closed.

But even if P_T factorizes into linear factors, thus $\sum_\lambda m(\lambda) = \dim V$ holds, diagonalizability of T fails if $\dim E_\lambda < m(\lambda)$ for some λ. Again there is a standard example:

Example 2 Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $P_A(t) = (1-t)^2$, so that A has eigenvalue $\lambda = 1$ with multiplicity two. Now $A - \lambda\mathbf{1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, whose kernel is the one-dimensional space spanned by $(1,0)$. Thus A is not diagonalizable.

Exercise 3 Consider the linear map defined by the matrix $A = \begin{pmatrix} a & c \\ 0 & b \end{pmatrix}$, where a, b, c are complex numbers. Determine precisely for which triples (a, b, d) the matrix A is diagonalizable.

Thus even over an algebraically closed field there are non-diagonalizable linear maps. One way out is to try to prove that every linear map admits a weaker 'normal form', generalizing the notion of diagonal matrix:

Definition 4 *Let V be a finite-dimensional vector space and $T : V \to V$ a linear map. We say T is triangularizable if there exists an ordered basis β of V such that the matrix $A = [T]_\beta$ is upper triangular (i.e. $A_{ij} = 0$ whenever $i > j$).*

The matrix from Example 2 is already upper triangular, thus it (or rather the linear map L_A) is trivially triangularizable, taking β to be the standard basis of $\mathbb{F}^2$. This example already shows that there are more triangularizable linear maps than diagonalizable ones.

Lemma 5 *Let V be a finite dimensional vector space and $T : V \to V$ a linear map that is triangularizable. Then the characteristic polynomial P_T of T splits into linear factors.*

Proof. Exercise. ■

We see that splitting of the characteristic polynomial P_T is a necessary condition also for triangularizability. In Section 4 we will see that it also is sufficient. (In particular, if $\mathbb{F}$ is an algebraically closed field, every linear map on a finite-dimensional vector space over $\mathbb{F}$ is triangularizable.)

Before we prove this (and the more specific Jordan normal form) we need some facts about polynomials, where we may limit ourselves to those that split into linear factors, which is a big simplification.

3 Some facts about splitting polynomials

The constructions of a (Schur) triangularization and of the Jordan normal (or canonical) form of a linear map require some facts about polynomials. We will adopt approaches that limit the needed prerequisites to the absolute minimum, avoiding even division with remainder of polynomials. This is possible since Lemma 5 tells us that we need to work only with polynomials that split into linear factors. While the results below may seem obvious or trivial at first sight this is not quite true, and proofs are necessary.

PROPOSITION 6 *Let $\mathbb{F}$ be a field and all polynomials considered are over $\mathbb{F}$.*

(a) A polynomial $p(x)$ is divisible by $x - c$ (i.e. $p(x) = (x - c)q(x)$ for some polynomial q) if and only if $p(c) = 0$.

(b) If $x - c$ divides a product $p_1 \cdots p_n$ of polynomials, $x - c$ must divide p_i for some i.[2]

(c) If $(x - c)^n$ $(n \in \mathbb{N})$ divides a product gh of polynomials and $x - c$ does not divide h then $(x - c)^n$ divides g.

(d) If $c_1, \ldots, c_I \in \mathbb{F}$ are mutually distinct and $m, m_1, \ldots, m_I \in \mathbb{N} = \{1, 2, \ldots\}$ then $(x - c)^m$ divides $p(x) = \prod_{i=1}^{I}(x - c_i)^{m_i}$ if and only if $c = c_i$ for some i and $m \leq m_i$.

(e) Assume p splits, i.e. $p(x) = a\prod_i(x - c_i)^{m_i}$, where $a \neq 0$ and we may assume the c_i to be mutually distinct. Then a and the pairs (c_i, m_i) are uniquely determined up to permutation.

(f) Assume p splits as in (e). Let $p = gh$ for polynomials g, h. Then g and h split

$$g(x) = b\prod_i(x - c_i)^{r_i}, \qquad h(x) = c\prod_i(x - c_i)^{s_i}, \tag{3}$$

where $bc = a$ and $r_i, s_i \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ with $r_i + s_i = n_i$ for each i.

Proof. (a) For $c = 0$ this equivalence is quite clear since $p(0) = 0$ is equivalent to the vanishing of the constant term of p, which in turn is equivalent to $p = xq$ for some polynomial q. If $c \neq 0$, define $q(x) = p(x+c)$. By the $c = 0$ case, $p(c) = 0 \Leftrightarrow q(0) = 0 \Leftrightarrow q$ is divisible by x, thus $q(x) = xr(x)$. Replacing x by $x-c$, this is equivalent to $p(x) = q(x-c) = (x-c)r(x-c)$. Since $x \mapsto r(x - c)$ is a polynomial, the last statement is equivalent to $x - c$ dividing p.

(b) If $x - c$ divides $p_1 \cdots p_n$ then (a) gives $p_1(c) \cdots p_n(c) = (p_1 \cdots p_n)(c) = 0$. Since $\mathbb{F}$ is a field, this implies $p_i(c) = 0$ for some i. Using (a) again, this gives that $x - c$ divides p_i.

(c) For $n = 1$ this is immediate by (b). If $n = 2$, in particular $x - c$ divides gh, thus g. Thus $g(x) = (x - c)\tilde{g}(x)$. Now $(x - c)^2$ divides $(x - c)\tilde{g}(x)$, thus $x - c$ divides $\tilde{g}$, which in turn means that $(x - c)^2$ divides g. This is easily turned into an inductive argument for all n, which we leave as an exercise.

(d) It is quite obvious that $x - c$ does not divide $x - d$ if $c \neq d$. Thus if $(x - c)^m$ divides the given p then by (b), c must equal one of the c_i. Then $p(x) = (x - c_i)^{m_i}\prod_{j\neq i}(x - c_j)^{m_j}$. Since $x - c$ does not divide $\prod_{j\neq i}(x - c_j)^{m_j}$, (c) gives that $(x - c)^m$ divides $(x - c_i)^{m_i}$, implying $m \leq m_i$.

(e) Let

$$p = a\prod_{i=1}^{I}(x - c_i)^{m_i} = b\prod_{j=1}^{J}(x - d_j)^{n_j} \tag{4}$$

be factorizations of p into linear factors, where the c_i are mutually distinct and the $m_i > 0$, and similarly for the d_j, n_j. Since the products without the factors a, b give monic polynomials (i.e. the coefficients of the highest power is one), we must have $a = b$, so that we assume this from now on. Since $x - c$ divides p if and only if $p(c) = 0$, it follows that the sets $\{c_1, \ldots, c_I\}$ and $\{d_1, \ldots, d_J\}$ both equal $p^{-1}(0)$. In particular $I = J$, so that permuting the d_i if necessary, we can assume $d_i = c_i$ for each i. Then $\prod_{i=1}^{I}(x - c_i)^{m_i} = \prod_{i=1}^{I}(x - c_i)^{n_i}$. Applying (d) gives $m_i \leq n_i$ and $n_i \leq m_i$, thus $m_i = n_i$, for all i.

(f) Since p splits, by (c) it has a representation $p(x) = a\prod_{i=1}^{I}(x - c_i)^{m_i}$, unique up to permutation. This means that p is divisible by $x - c_i$ exactly m_i-times. Applying (c) to

[2] In the language of algebra: Polynomials of the form $x - c$ are prime (among the polynomials).

$p = gh$ shows that if $x - c_i$ divides g r_i times and h s_i times, we must have $r_i + s_i \geq m_i$. Since this holds for all i and $\sum_i m_i = \deg p = \deg g + \deg h = \sum_i (r_i + s_i)$, we see that actually $r_i + s_i = m_i$ for all i, and g, h have no other polynomial factors. Thus g is a scalar multiple of $\prod_i (x - c_i)^{r_i}$ and h is a scalar multiple of $\prod_i (x - c_i)^{s_i}$. The claim follows readily. ■

4 Triangularizability of linear maps

THEOREM 7 *Let V be a finite-dimensional vector space of dimension d and $T : V \to V$ a linear map. Then the following are equivalent:*

(i) The characteristic polynomial P_T splits into linear factors.

(ii) There exist T-invariant linear subspaces $W_1 \subset W_2 \subset \cdots \subset W_{d-1} \subset V$ with $\dim W_i = i$ for all $i \in \{1, \ldots, d-1\}$.[3]

(iii) T is triangularizable.

If V is an inner product space over $\mathbb{R}$ or $\mathbb{C}$, we can add:

(iv) There exists an orthonormal basis β for V such that $[T]_\beta$ is upper triangular.

The proof of the theorem will use the following Proposition, whose proof we postpone.

PROPOSITION 8 *Let V be a finite dimensional vector space and $T : V \to V$ a linear map with splitting P_T. Then there exists a T-invariant subspace $W \subset V$ with $\dim W = \dim V - 1$.*

Proof of Theorem 7. (iv)⇒(iii) is trivial.

(iii)⇒(i) was proven in Lemma 5.

(iii)⇒(ii) Let $\beta = \{e_1, \ldots, e_d\}$ be a basis for V such that $A = [T]_\beta$ is diagonal. Define $W_i = \operatorname{span}(e_1, \ldots, e_i)$ for all $i \in \{1, \ldots, d-1\}$. Since β is a basis, the e_i are linearly independent, implying $\dim W_i = i$ for all i. Now,

$$Te_j = \sum_{i=1}^{d} A_{ij} e_i = \sum_{i=1}^{j} A_{ij} e_i \in W_j$$

where the first equality holds by definition of $A = [T]_\beta$, the second by upper triangularity of A and the "$\in$" by definition of W_j. Thus all W_i are T-invariant.

(ii)⇒(iii)+(iv) Choose a non-zero $e_1 \in W_1$. Since $\dim W_2 = 2 > 1 = \dim W_1$, we can choose $e_2 \in W_2$ such that $\{e_1, e_2\}$ is a basis for W_2. Next, choose $e_3 \in W_3$ such that $\{e_1, e_2, e_3\}$ is a basis for W_3, and so on including $i = d$. Now $\beta = \{e_1, \ldots, e_d\}$ is a basis for V. Since all W_i are T-invariant, each Te_j is a linear combination of $e_1, \ldots, e_j$, implying for $A = [T]_\beta$ that $A_{ij} = 0$ if $i > j$. Thus β 'triangularizes' T.

If V is an inner product space, we can choose each e_i in the one-dimensional space $W_i \cap W_{i-1}^\perp$ (e.g. by Gram-Schmidt). Doing this, each e_i is orthogonal to $e_1, \ldots, e_{i-1}$, and this clearly implies that $\beta = \{e_1, \ldots, e_d\}$ is an orthogonal set. We may as well normalize all e_i, obtaining an ONB. This proves (iv).

It remains to prove (i)⇒(ii). Given Proposition 8, this is immediate: Apply the proposition to V and call the resulting T-invariant subspace W_{d-1}. It has the desired dimension $d - 1$. By [4, Theorem 5.20] the characteristic polynomial $P_{T_{W_{d-1}}}$ of $T_{W_{d-1}}$ divides that

[3] Such a sequence of subspaces is sometimes called a *flag* or *fan* for T, e.g. in [6].

of T. Since P_T splits by assumption, Proposition 6(f) gives that also $P_{T_{W_{d-1}}}$ splits. Thus we can apply the proposition to $T : W_{d-1} \to W_{d-1}$ and call W_{d-2} the resulting $(d-2)$-dimensional subspace of W_{d-1}. Iterate this until we have obtained W_1. It is evident that we have obtained T-invariant $W_1 \subset \cdots \subset W_{d-1} \subset V$ with $\dim W_i = i$, as required in (ii). ■

Proof of Proposition 8. How would we find a T-invariant subspace $W \subseteq V$ of dimension $\dim V - 1$? The most natural approach is to take a linear map $\phi : V \to \mathbb{F}$ that is not the zero map. Then its image $R(\phi)$ has dimension one, and the null space $N(\phi) \subset V$ has the wanted dimension $\dim V - 1$ by the rank-nullity theorem (or dimension theorem, see [4, Theorem 2.3]). Putting $W = N(\phi)$, we want W to be T-invariant. This means that $\phi(v) = 0$ must imply $\phi(Tv) = 0$. This is clearly the case if there exists $\lambda \in \mathbb{F}$ such that $\phi(Tv) = \lambda\phi(v)$ for all v. We have thus reduced the problem to finding a non-zero linear map $\phi : V \to \mathbb{F}$ satisfying $\phi(Tv) = \lambda\phi(v)$ for all v (where the value of λ does not matter). This vaguely looks like an eigenvalue problem – and indeed it is!

To make this precise we need a few facts on dual spaces from the optional Section 2.6 of [4], where more facts and examples can be found. If V is a finite-dimensional vector space over the field $\mathbb{F}$, we call V^* the space $\mathcal{L}(V, \mathbb{F})$ of linear maps from V to $\mathbb{F}$ (which we encountered in [4, Theorem 6.8]). It is a vector space of the same dimension as V (since $\dim \mathcal{L}(V, W) = \dim V \cdot \dim W$ and $\dim \mathbb{F} = 1$).

If $T : V \to V$ is a linear map and $\phi \in V^*$, the composition $\phi \circ T$ is again in V^*. We call $T^t : V^* \to V^*$ (the transpose of T) the map defined by $T^t\phi = \phi \circ T$. It is easy to see that T^t is a linear map and that $(\mathbf{1}_V)^t = \mathbf{1}_{V^*}$. Now the desired identity $\phi \circ T = \lambda\phi$ is equivalent to $T^t\phi = \lambda\phi$, which is just the statement that $\phi \in V^*$ is an eigenvector of T^t. If the field $\mathbb{F}$ is algebraically closed, we know that every linear map has (non-zero) eigenvectors, thus a $\phi : V \to \mathbb{F}$ with the desired properties exists, and we are done.

If $\mathbb{F}$ is not algebraically closed, we claim that T^t has the same characteristic polynomial as T. Since P_T is assumed to split, so does P_{T^t}. Thus it has zeros, and T^t has eigenvectors, so that again we are done. It remains to prove $P_T = P_{T^t}$. This requires a little bit of effort.

If $\beta = \{e_1, \ldots, e_d\}$ is a basis for V, every $v \in V$ can be written as $a_1e_1 + \cdots + a_de_d$ for unique $a_1, \ldots, a_d$. The maps ϕ_i sending $v \in V$ to the coefficient a_i of e_i are linear, thus elements of V^*. One sees that $\phi_i(e_j) = \delta_{ij}$ for all i, j. The set $\{\phi_1, \ldots, \phi_d\} \subset V^*$ is linearly independent, thus a basis for V^*. It is called the (unique) basis of V^* dual to β.

If $\beta = \{e_1, \ldots, e_d\}$ and $\gamma = \{\phi_1, \ldots, \phi_d\}$, the linear maps T and T^t have matrix representations $A = [T]_\beta$ and $B = [T^t]_\gamma$ defined by

$$Te_j = \sum_{k=1}^{d} A_{kj}e_k, \qquad T^t\phi_i = \sum_{k=1}^{d} B_{ki}\phi_k.$$

Applying ϕ_i to the first identity and applying both sides of the second identity (which are in V^*) to e_j and using $\phi_i(e_j) = \delta_{ij}$ and the definition of T^t gives

$$A_{ij} = \sum_{k=1}^{d} A_{kj}\phi_i(e_k) = \phi_i(Te_j) = (T^t\phi_i)(e_j) = \sum_{k=1}^{d} B_{ki}\phi_k(e_j) = B_{ji},$$

thus $B = A^t$, or $[T^t]_\gamma = \big([T]_\beta\big)^t$. Now also the matrices $[T^t - s\mathbf{1}_{V^*}]_\gamma$ and $[T - s\mathbf{1}_V]_\beta$ are transposes of each other and therefore have the same determinant. This implies

$$P_{T^t}(s) = \det(T^t - s\mathbf{1}_{V^*}) = \det[T^t - s\mathbf{1}_{V^*}]_\gamma = \det[T - s\mathbf{1}_V]_\beta = \det(T - s\mathbf{1}_V) = P_T(s).$$

This finishes the proof of Proposition 8 and thereby of Theorem 7. ■

In view of Lemma 5, nothing can be done about the condition of splitting characteristic polynomial, but over an algebraically closed field every linear map can be triangularized.[4]

The triangularization result (usually attributed to I. Schur) has some applications, like for the proof of Theorem 6.16 given in [4] and Exercise 9 below. There also are applications to the numerical algorithms in linear algebra. But for many purposes, the upper triangular form is too unspecific. This will be fixed by the Jordan normal (or canonical) form studied next.

EXERCISE 9 Let V be a finite-dimensional vector space and $T : V \to V$ a linear map whose characteristic polynomial splits. Let $\{\lambda_1, \dots, \lambda_n\}$ be the eigenvalues of T with multiplicities $\{m_1, \dots, m_n\}$. Let p be a polynomial of degree ≥ 1. Use Theorem 7 to prove that the eigenvalues of $p(T)$ are $\{p(\lambda_1), \dots, p(\lambda_n)\}$ with multiplicities $\{m_1, \dots, m_n\}$.

5 Nilpotent linear maps

With an eye to our later needs, we introduce two related notions:

DEFINITION 10 *An $n \times n$ matrix A is called strictly upper triangular if $A_{ij} = 0$ whenever $i \geq j$. (This is equivalent to A being upper triangular and having all diagonal elements equal to zero.)*

DEFINITION 11 *A linear map $T : V \to V$ or a matrix A is called nilpotent if there is $n \in \mathbb{N}$ such that $T^n = 0$. The smallest such n is called the order of nilpotency of A.*

EXERCISE 12 Prove:

(i) If A is strictly upper triangular $n \times n$ matrix, prove $A^n = 0$. (Thus A is nilpotent of order of nilpotency at most n.)

(ii) Let A, B be $n \times n$ matrices that are upper triangular. Prove that AB is upper triangular and that the diagonal elements are $(AB)_{ii} = A_{ii}B_{ii}$.

(iii) An upper triangular matrix is nilpotent if and only if it is strictly upper triangular.

(iv) Give an example of a non-zero nilpotent 2×2 matrix that not strictly upper or lower triangular.

EXERCISE 13 Let V be a finite-dimensional vector space and $T : V \to V$ a nilpotent linear map. Prove directly, using only the definitions of nilpotency and eigenvalues/vectors:

(i) $N(T)$ is non-trivial, thus 0 is an eigenvalue of T.

(ii) T has no non-zero eigenvalues.

(iii) If T is diagonalizable and nilpotent then $T = 0$.

This shows that (non-zero) nilpotent linear maps are very far from being diagonalizable.

The connection between nilpotency and strict upper triangularity is even stronger than suggested by Exercise 13:

[4] And conversely: If $\mathbb{F}$ is not algebraically closed, there exists a polynomial P that does not split over $\mathbb{F}$. Since every polynomial is a scalar multiple of the characteristic polynomial of a matrix (check the matrix appearing in the proof of [4, Theorem 5.21]), we have a matrix that is not triangularizable.

Theorem 14 *Let V be a finite-dimensional vector space over the field $\mathbb{F}$ and $T : V \to V$ a linear map. Consider the following statements:*

(i) T is nilpotent, i.e. $T^n = 0$ for some $n \in \mathbb{N}$.

(ii) T^t is nilpotent.

(iii) There is a basis β for V such that $A = [T]_\beta$ is strictly upper triangular, i.e. $A_{ij} = 0$ if $i \geq j$.

(iv) $P_T(t) = (-t)^d$, where $d = \dim V$.

(v) $P_T^{-1}(0) = \{0\}$.

Then (i)⇔(ii)⇔(iii)⇔(iv)⇒(v). If P_T splits (as when $\mathbb{F}$ is algebraically closed) then also (v) implies the other statements.

Proof. (i)⇒(ii) This is immediate once we prove that $(T^t)^n \phi = \phi \circ T^n$. For $n = 1$ this is true by definition of T^t. Assuming the claim to be true for n, the computation $(T^t)^{n+1}\phi = T^t((T^t)^n\phi) = T^t(\phi \circ T^n) = \phi \circ T^{n+1}$ is the inductive step.

(ii)⇒(iii) Since T^t is nilpotent, it has non-trivial null space by the above Exercise. We can thus find $\phi \in V^* \backslash \{0\}$ such that $T^t\phi = 0$, meaning $\phi \circ T = 0$. Reconsidering the proof of Proposition 8 we find for every $v \in V$ that $\phi(Tv) = 0$, thus $Tv \in N(\phi) = W$. We therefore have the conclusion that $T(V) \subseteq W$, which is stronger than just T-invariance of W, i.e. $T(W) \subseteq W$. Now in the proof of (ii)⇒(iii) in Theorem 7 this means that T maps each W_i into W_{i-1}. In particular Te_i is a linear combination of $e_1, \ldots, e_{i-1}$, and this translates into strict upper triangularity of the matrix $[T]_\beta$.

(iii)⇒(iv) Since $[T]_\beta$ is strictly upper triangular, the matrix $[T - t\mathbf{1}_V]_\beta = [T]_\beta - t\mathbf{1}$ has all diagonal elements equal to $-t$, thus

$$P_T(t) = \det(T - t\mathbf{1}_V) = \det[T - 1\mathbf{1}_V] = (-t)^d, \quad \text{with } d = \dim V.$$

(iv)⇒(i) By Cayley-Hamilton, $P_T(T) = 0$, thus $T^d = 0$. (The actual order of nilpotency of T may be lower than d.)

(iv)⇒(v) is trivial. And if (v) holds and P_T splits then the multiplicity of the zero 0 of P_T must be d, implying (iv). ■

Exercise 15 Give an example of a $n \times n$ matrix A with $n \leq 3$ such that $P_A^{-1}(0) = \{0\}$, but A is not nilpotent.

Lemma 16 *Let V be a non-zero vector space and $T : V \to V$ a linear map. Then*

(i) $N(T) \subseteq N(T^2) \subseteq N(T^3) \subseteq \cdots$. Thus the sequence of null-spaces is non-decreasing.

(ii) If $N(T^n) = N(T^{n+1})$ for some n then $N(T^k) = N(T^n)$ for all $k \geq n$, implying $\bigcup_{k\in\mathbb{N}} N(T^k) = N(T^n)$.

(iii) If V is finite-dimensional, an n as in (ii) always exists. It satisfies $n \leq \dim V$.

(iv) If V is finite-dimensional, the following are equivalent:

(α) There exists $n \in \mathbb{N}$ such that $T^n = 0$.

(β) For each $v \in V$ there exists $n \in \mathbb{N}$ such that $T^n v = 0$.

Proof. (i) $v \in N(T^k)$ is equivalent to $T^k v = 0$. If now $l \geq k$ then $T^l v = T^{k-l}(T^k v) = T^{k-l}0 = 0$, thus $v \in N(T^l)$.

(ii) Assume $N(T^n) = N(T^{n+1})$ for some n. If $x \in N(T^{n+2})$ then $Tx \in N(T^{n+1}) = N(T^n)$, where we used the hypothesis. Thus $x \in N(T^{n+1})$, so that we have proven $N(T^{n+2}) \subseteq N(T^{n+1})$. Since (i) gives the converse inclusion, we have $N(T^{n+2}) = N(T^{n+1})$. Now the claim follows from an obvious iteration of the argument.

(iii) If T is injective then $N(T^k) = \{0\}$ for all k, so that the claim is true. If T is not injective and there was no such n, (ii) would imply $\{0\} \subsetneqq N(T) \subsetneqq N(T^2) \subsetneqq \cdots$, thus $\dim N(T^k) \geq k$ for all k, which is impossible. Thus $N(T^n) = N(T^{n+1})$ must happen for $n \leq \dim V$.

(iv) $(\alpha) \Rightarrow (\beta)$ This is quite obvious since $T^n v = 0v = 0$ for each $v \in V$. $(\beta) \Rightarrow (\alpha)$ By (iii) there exists a smallest n such that $\bigcup_{k=1}^{\infty} N(T^k) = N(T^n)$. Since the assumption (β) is equivalent to $V = \bigcup_{k=1}^{\infty} N(T^k)$ this implies $V = N(T^n)$, thus $T^n = 0$. ■

Remark 17 ⋆ If V is infinite-dimensional, the implication $(\alpha) \Rightarrow (\beta)$ still holds, but (i) and $(\beta) \Rightarrow (\alpha)$ can be false. Example: Let V be the vector space of infinite sequences $a = (a_1, a_2, a_3, \ldots)$ in $\mathbb{R}$ or $\mathbb{C}$ such that a_k is non-zero for only finitely many k. Now define a linear map $T : V \to V$ by $(Ta)_k = a_{k+1}$ for all $k \in \mathbb{N}$. Equivalently, $T(a_1, a_2, \ldots) = (a_2, a_3, \ldots)$. This implies that $N(T^k)$ consists of the sequences $(a_1, a_2, \ldots)$ satisfying $a_l = 0$ for all $l > k$. Since every $a \in V$ has only finitely many non-zero coordinates a_l, it is clear that $T^n a = 0$ for n large enough. On the other hand for each $n \in \mathbb{N}$ one can find $a \in V$ such that $T^n a \neq 0$ (for example $a \in V$ defined by $a_k = \delta_{k,n+1}$). This implies $T^n \neq 0$ for all $n \in \mathbb{N}$. □

6 Generalized eigenspaces

All vector spaces called V are assumed to be non-zero!

Definition 18 *Let V be a finite-dimensional vector space over a field $\mathbb{F}$, and let $T : V \to V$ be a linear map. For each $\lambda \in \mathbb{F}$ we define*

$$\begin{aligned} E_\lambda &= \{v \in V \mid Tv = \lambda v\} = N(T - \lambda \mathbf{1}), \\ K_\lambda &= \{v \in V \mid (T - \lambda \mathbf{1})^n v = 0 \text{ for some } n \in \mathbb{N}\} = \bigcup_{n \in \mathbb{N}} N((T - \lambda \mathbf{1})^n). \end{aligned}$$

Now E_λ is the eigenspace of V for eigenvalue λ, and K_λ is called the generalized eigenspace of V for eigenvalue λ.

- This definition is motivated by Example 2, where E_1 has dimension one, while K_1 has dimension two.
- Clearly $E_\lambda \subseteq V$ is a linear subspace. Its dimension $\dim E_\lambda$ is called the geometric multiplicity of $\lambda \in \mathbb{F}$.
- As we know, $E_\lambda \neq \{0\}$ if and only if $T - \lambda \mathbf{1}$ has non-trivial null-space, which is equivalent (only for finite-dimensional V!) to T not being invertible, which in turn is equivalent to $P_T(\lambda) = 0$, where $P_T(\lambda) = \det(T - \lambda \mathbf{1}_V)$ is the characteristic polynomial of V.

- Clearly $E_\lambda \subseteq K_\lambda$, so that $E_\lambda \neq \{0\} \Rightarrow K_\lambda \neq \{0\}$. On the other hand, $E_\lambda = \{0\}$ means that $T - \lambda\mathbf{1}$ is invertible. Then $(T - \lambda\mathbf{1})^n$ is invertible for all n, so that $K_\lambda = \{0\}$. Thus $P_T(\lambda) = 0 \Leftrightarrow E_\lambda \neq \{0\} \Leftrightarrow K_\lambda \neq \{0\}$.

LEMMA 19 *Let V be a finite-dimensional vector space and $T : V \to V$ linear. Then*

(i) $K_\lambda \subseteq V$ is a linear subspace for each λ.

(ii) K_λ is T-invariant.

(iii) If $\mu \neq \lambda$ then the restriction of $T-\mu\mathbf{1}$ to K_λ is invertible. In particular $K_\mu \cap K_\lambda = \{0\}$.

Proof. (i) If $x \in K_\lambda$ and $c \in \mathbb{F}$, then $(T - \lambda\mathbf{1})^n x = 0$ for some $n \in \mathbb{N}$, and it is obvious that this implies $(T - \lambda\mathbf{1})^n(cx) = 0$, thus $cx \in K_\lambda$. Now let $x, y \in K_\lambda$. This means that there are $n, m \in \mathbb{N}$ such that $(T - \lambda\mathbf{1})^n x = 0$ and $(T - \lambda\mathbf{1})^m y = 0$. Since $(T - \lambda\mathbf{1})^k z = 0$ implies $(T - \lambda\mathbf{1})^\ell z = 0$ for all $\ell \geq k$, it follows that $x, y \in N((T - \lambda\mathbf{1})^{\max(n,m)})$ thus also $x + y \in N((T - \lambda\mathbf{1})^{\max(n,m)})$, implying $x + y \in K_\lambda$.

(ii) If $v \in K_\lambda$, thus $(T - \lambda\mathbf{1})^n v = 0$ for some n, we have

$$(T - \lambda\mathbf{1})^n T v = T(T - \lambda\mathbf{1})^n v = T0 = 0,$$

thus $Tv \in K_\lambda$, proving the T-invariance of K_λ.

(iii) Since K_λ is T-invariant, it is also invariant under $T - \mu\mathbf{1}$, so that the statement makes sense. Since V is finite-dimensional, invertibility of $(T - \mu\mathbf{1})_{K_\lambda}$ is equivalent to injectivity, thus triviality of the null-space. Since $N(T - \mu\mathbf{1}) = E_\mu$, the first claim is equivalent to $K_\lambda \cap E_\mu = \{0\}$. Let thus $v \in K_\lambda \cap E_\mu$. Since T acts on v by multiplication with μ, we have $(T - \lambda\mathbf{1})^n x = (\mu - \lambda)^n x$ for all $n \in \mathbb{N}$. On the other hand, there is an n such that $(T - \lambda\mathbf{1})^n x = 0$. Thus $(\mu - \lambda)^n x = 0$, and since $\mu \neq \lambda$, we conclude that $x = 0$, proving the first claim.

Since $T - \mu\mathbf{1}$ acts invertibly on K_λ, the same holds for $(T - \mu\mathbf{1})^n$ for all n. Thus no non-zero element of K_λ can be in K_μ. This proves the second claim $K_\mu \cap K_\lambda = \{0\}$. ■

If a vector space V has two linear subspaces W_1, W_2 such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$ then every $v \in V$ can be written in a unique way as $w = w_1 + w_2$ with $W_i \in W_i$. (See the exercises in [4, Section 1.3].) But given more than two subspaces $W_1, \ldots, W_k$ with $\sum_{i=1}^k W_i = V$, the condition $W_i \cap W_j = \{0\}$ whenever $i \neq j$ is too weak to imply that every $x \in V$ can be written uniquely as $x = \sum_{j=1}^k w_j$ with $w_j \in W_j$. (See the (unfortunately) optional section on direct sums in [4, Section 5.2].)

PROPOSITION 20 *Let $\{\lambda_1, \ldots, \lambda_n\}$ be the distinct zeros of P_T and $x_i \in K_{\lambda_i}$ for each $i \in \{1, \ldots, n\}$. If $x_1 + x_2 + \cdots + x_n = 0$ then $x_i = 0$ for each i.*

Proof. Let $x_i \in K_{\lambda_i}$ for each i such that $x_1 + x_2 + \cdots + x_n = 0$. For each i pick n_i such that $(T - \lambda_i\mathbf{1})^{n_i} x_i = 0$. Thus applying the product $\prod_{\substack{j=1 \\ j \neq i}}^n (T - \lambda_j\mathbf{1})^{n_j}$ to $x_1 + x_2 + \cdots + x_n = 0$ sends x_j to zero for all $j \neq i$, so that

$$0 = \Big(\prod_{\substack{j=1 \\ j \neq i}}^n (T - \lambda_j\mathbf{1})^{n_j} \Big)(x_1 + x_2 + \cdots + x_n) = \Big(\prod_{\substack{j=1 \\ j \neq i}}^n (T - \lambda_j\mathbf{1})^{n_j} \Big) x_i.$$

By Lemma 19(iii), each $(T - \lambda_j\mathbf{1})^{n_j}$ with $j \neq i$ acts invertibly on K_{λ_i}. Since $x_i \in K_{\lambda_i}$, the above equation implies $x_i = 0$. Since i was arbitrary, we have proven $x_i = 0$ for all i. ■

Remark 21 The above result implies that every element of $W = \sum_{i=1}^n K_{\lambda_i}$ can be written as $\sum_i x_i$ with $x_i \in K_{\lambda_j}$ in only one way: If $x_1 + \cdots + x_n = x_1' + \cdots + x_n'$ with $x_i, x_i' \in K_{\lambda_i}$ then $\sum_i (x_i - x_i') = 0$. Since $x_i - x_i' \in K_{\lambda_i}$, the above proposition gives $x_i = x_i'$ for each i. For this reason one calls W a direct sum and denotes it by $W = \bigoplus_{i=1}^n K_{\lambda_i}$. □

Lemma 22 *Let V be a finite-dimensional vector space and $T : V \to V$ a linear map. Then*

$$\dim K_\lambda \le m(\lambda) \tag{5}$$

for each λ. (Recall that $m(\lambda)$ is the multiplicity of λ as a zero of P_T.)

Proof. As we have seen, $K_\lambda \subseteq V$ is a T-invariant subspace for each λ. Thus by [4, Theorem 5.20], the characteristic polynomial $P_{T_{K_\lambda}}$ of T_{K_λ} divides the characteristic polynomial P_T of T. Now by definition of K_λ and Lemma 16(ii), $\widetilde{T}_{K_\lambda} = T_{K_\lambda} - \lambda\mathbf{1}$ is nilpotent, so that by Theorem 14, there is a basis β for K_λ such that $[\widetilde{T}_{K_\lambda}]_\beta$ is strictly upper triangular. Then $[T_{K_\lambda} - t\mathbf{1}]_\beta = [\widetilde{T}_{K_\lambda} + (\lambda - t)\mathbf{1}]_\beta$ is upper triangular with all diagonal elements equal to $\lambda - t$, implying $P_{T_{K_\lambda}}(t) = (\lambda - t)^{\dim K_\lambda}$. Thus $(t - \lambda)^{\dim K_\lambda}$ divides P_T, implying $\dim K_\lambda \le m(\lambda)$. ■

Now the question arises whether $\sum_\lambda K_\lambda$ equals the total space V. Since the sum is a direct sum by Proposition 20, this is equivalent to $\sum_\lambda \dim K_\lambda = \dim V$. Since we know from Lemma 22 that $\dim K_\lambda \le m(\lambda)$ for each λ and already knew that $\sum_\lambda m(\lambda) \le \dim V$, it follows that the sum of the K_λ equals V if and only if $\sum_\lambda m(\lambda) = \dim V$, i.e. P_T splits into linear factors, and $\dim K_\lambda = m(\lambda)$ for each λ. The methods used so far are insufficient for proving $\dim K_\lambda = m(\lambda)$, the main problem being that we know nothing about the action of T on vectors of V that are not in $\sum_\lambda K_\lambda$ (if they exist).

There are at least two different ways to attack this problem, one being given in [4, Chapter 7] (also included in the appendices with a more self-contained exposition). But this proof requires the results from [4, Appendix E] on polynomials that go well beyond what we proved in Section 3 and typically are taught not before the second year. For this reason we follow an alternative approach (from [5]) that replaces the theory of polynomials by a clever result on linear maps, Theorem 23 below. We also put more emphasis than [4] on the notion of nilpotent linear maps, without which the Jordan normal cannot be understood properly.

7 Proof of $V = \bigoplus_\lambda K_\lambda$

For every finite-dimensional vector space V and linear map $T : V \to V$, the spaces $N(T)$ and $R(T)$ are T-invariant, the restriction $T_{N(T)}$ clearly being the zero map. But despite the fact that $\dim N(T) + \dim R(T) = \dim V$ always holds, $N(T)$ and $R(T)$ do not have to be in a simple relative position. It is not hard to prove that the following statements are equivalent: (i) $N(T) \cap R(T) = \{0\}$, (ii) $N(T) + R(T) = V$, (iii) $T_{R(T)}$ is invertible.

We now prove a decomposition that always holds (in finite dimensions):

Theorem 23 *Let V be a finite-dimensional vector space and $T : V \to V$ linear. Let $n \in \mathbb{N}$ be the smallest for which $N(T^n) = N(T^{n+1})$ (compare Lemma 16(iii)). Then $U = N(T^n)$ and $W = R(T^n)$ are T-invariant subspaces such that T_U is nilpotent of order n, T_W is invertible, and $U \cap W = \{0\}$, $U + W = V$ (thus $V = U \oplus W$).*

Proof. Let n be the smallest element of $\mathbb{N}$ such that $N(T^n) = N(T^{n+1}) = \cdots$ as obtained in Lemma 16(i). We put $U = N(T^n)$ and $W = R(T^n)$, both of which are T-invariant. (If $T^n v = 0$ then $T^n(Tv) = TT^n v = 0$, thus $Tv \in N(T^n)$. And if $x \in R(T^n)$, there exists $y \in V$ such that $x = T^n y$. Now $Tx = TT^n y = T^n(Ty)$, thus $Tx \in R(T^n)$.)

Since every $u \in U$ satisfies $T^n u = 0$, and since by the choice of n there exists $u \in U$ with $T^{n-1}u \neq 0$, the restriction T_U of T to U is nilpotent of index n. Now assume $Tw = 0$ for some $w \in W$. By the definition of W there exists $y \in Y$ such that $w = T^n y$. Inserting this into $Tw = 0$ gives $T^{n+1}y = 0$, thus $y \in N(T^{n+1}) = N(T^n)$ by our choice of n. Thus $w = T^n y = 0$, proving that T_W has trivial null-space. Now the finite-dimensionality of W implies that T_W is invertible.

Now assume $x \in U \cap W = N(T^n) \cap R(T^n)$. Then $x \in U := N(T^n)$ implies $T^n x = 0$. On the other hand, T_W is injective, thus the same holds for $(T_W)^n$ for all n, so that $T_W^n x = T^n x = 0$ implies $x = 0$. Thus $U \cap W = \{0\}$. Finally, by the rank-nullity theorem (or dimension theorem, see [4, Theorem 2.3]) we have

$$\dim U + \dim W = \dim N(T^n) + \dim R(T^n) = \dim V,$$

which implies $V = U \oplus W$. Thus every $v \in V$ can be written as $v = u + w$ with $u \in U$, $w \in W$ in a unique way. ■

Remark 24 Decompositions of algebraic objects involving the splitting off a nilpotent part play an important role in many areas of algebra, like commutative as well as non-commutative ring theory and in the theory of Lie-algebras. □

Theorem 25 *Let V be a finite dimensional vector space of dimension $d = \dim V \geq 1$ and $T : V \to V$ a linear map whose characteristic polynomial P_T factorizes into d linear factors, i.e. (2) holds. (Then of course $\sum_i m_i = d$.) Then $V = \bigoplus_{i=1}^n K_{\lambda_i}$, i.e. every $x \in V$ can uniquely be written as $x = x_1 + \cdots + x_n$, where $x_i \in K_{\lambda_i}$ for each i.*

Proof. Let λ_1 be one of the eigenvalues of T (which exist since P_T splits). Apply Theorem 23 to $\widetilde{T} = T - \lambda_1 \mathbf{1}$. We obtain $\widetilde{T}$-invariant subspaces U, W such that $\widetilde{T}_U$ is nilpotent and $\widetilde{T}_W$ is invertible. According to Theorem 23, $U = \bigcup_{k=1}^{\infty} N(\widetilde{T}^k)$, thus $U = K_{\lambda_1}$.

Clearly, W, U are also invariant under $T = \widetilde{T} + \lambda \mathbf{1}$. If $U = V$, equivalently $W = \{0\}$, we are done. If not, we notice that $\dim W < \dim V$ since $U = K_{\lambda_1} \supseteq E_{\lambda_1}$ has dimension at least one. And P_{T_W} divides P_T and therefore splits by Proposition 6(f). Thus we can apply the above reasoning to the linear map $T_W : W \to W$. Thus T_W has an eigenvalue λ_2 (distinct from λ_1 since $(T - \lambda_1 \mathbf{1})_W$ is invertible) and T_W-invariant subspaces $U' = K_{\lambda_2} \subseteq W$ and $W' \subseteq W$. Iterating this construction until we have remainder $W = \{0\}$, we obtain a direct sum decomposition $V = \bigoplus_{i=1}^n K_{\lambda_i}$ for mutually distinct eigenvalues λ_i. Since all these spaces are T-invariant, the characteristic polynomial factorizes as $P_T(t) = \prod_{i=1}^n P_{T_{K_{\lambda_i}}}(t) = \prod_{i=1}^n (\lambda_i - t)^{\dim K_{\lambda_i}}$. It is now clear that the λ_i exhaust the eigenvalues of T and that $\dim K_{\lambda_i} = m(\lambda_i)$ for each i. ■

Remark 26 The proofs of the Theorems 23 (known as Fitting's lemma) and 25 are adapted from Halmos' book [5]. Another good reference is [3]. □

8 Normal form for nilpotent linear maps

In this section, there will be no assumption of characteristic polynomials splitting into linear factors!

Definition 27 *Let V be a vector space over $\mathbb{F}$ and W a linear subspace. Then vectors $\{v_1, \ldots, v_m\} \subset V$ are called linearly independent over W if $a_1 v_1 + \cdots + a_m v_m \in W$, where $a_1, \ldots, a_m \in \mathbb{F}$, implies $a_1 = \cdots = a_m = 0$.*

This condition is somewhat more restrictive than ordinary linear independence (which corresponds to $W = \{0\}$). For example $\{v_1, \ldots, v_m\}$ is never linearly independent over W if any one of the vectors is in W.

Lemma 28 *Let V be a finite-dimensional vector space and $W \subseteq V$ a linear subspace. Then vectors $\{v_1, \ldots, v_m\} \subset V$ are linearily independent over W if and only*

$$\dim \big(W + \mathrm{span}(v_1, \ldots, v_m)\big) = \dim W + m. \tag{6}$$

Proof. We pick and fix a basis $\{w_1, \ldots, w_d\}$ for W, where $d = \dim W$.

$\Rightarrow$ Assume that $\{v_1, \ldots, v_m\} \subset V$ is linearily independent over W. If $b_1 w_1 + \cdots + b_d w_d + a_1 v_1 + \cdots + a_m v_m = 0$ then the linear independence of $\{v_1, \ldots, v_m\}$ over W implies $a_1 = \cdots = a_m = 0$, thus $b_1 w_1 + \cdots + b_d w_d = 0$. Now the linear independence of $\{w_1, \ldots, w_d\}$ implies $b_1 = \cdots = b_d = 0$. Thus the set $\{w_1, \ldots, w_d, v_1, \ldots, v_m\}$ is linearly independent, implying $\dim \big(W + \mathrm{span}(v_1, \ldots, v_m)\big) = d + m = \dim W + m$, thus (6).

$\Leftarrow$ If (6) holds, the set $\{w_1, \ldots, w_d, v_1, \ldots, v_m\}$ is linearly independent (since otherwise its span would have dimension $< d+m$). If now $a_1 v_1 + \cdots + a_m v_m = w \in W$, there are unique $b_1, \ldots, b_d$ such that $w = -(b_1 w_1 + \cdots b_d w_d)$, implying $b_1 w_1 + \cdots + b_d w_d + a_1 v_1 + \cdots + a_m v_m = 0$. Now the linear independence of $\{w_1, \ldots, w_d, v_1, \ldots, v_m\}$ implies that all a_i, b_j are zero. Thus $\{v_1, \ldots, v_m\}$ is linearly independent over W. ■

Exercise 29 Let V be finite-dimensional vector space over $\mathbb{F}$, $W \subset V$ a linear subspace and $\eta : V \to V/W$ the quotient map. Prove that a subset $\{v_1, \ldots, v_n\} \subset V$ is linearly independent over W if and only if $\{\eta(v_1), \ldots, \eta(v_n)\} \subset V/W$ is linearly independent.

(Using this one can give a somewhat simpler proof of the preceding lemma.)

Lemma 30 *Let V be a finite-dimensional vector space and $T : V \to V$ nilpotent of order n. Then*

(i) $N(T) \subseteq N(T^2) \subseteq \cdots \subseteq N(T^{n-1}) \subseteq N(T^n) = V$.

(ii) For $k \in \mathbb{N}, v \in V$, we have $v \in N(T^k) \Leftrightarrow Tv \in N(T^{k-1})$. ($N(T^0) = N(\mathbf{1}_V) = \{0\}$.)

(iii) If $\{x_1, \ldots, x_m\} \subset N(T^k)$ is linearily independent over $N(T^{k-1})$ then $\{Tx_1, \ldots, Tx_m\} \subset N(T^{k-1})$ is linearily independent over $N(T^{k-2}) \subseteq N(T^{k-1})$.

Proof. (i) We've proven this already in Lemma 16(i).

(ii) $v \in N(T^k) \Leftrightarrow T^k v = 0 \Leftrightarrow T^{k-1}(Tv) = 0 \Leftrightarrow Tv \in N(T^{k-1})$.

(iii) Assume that $\{Tx_1, \ldots, Tx_m\} \subset N(T^{k-1})$ are linearily dependent over $N(T^{k-2}) \subseteq N(T^{k-1})$, thus $a_1 T x_1 + \cdots a_m T x_m \in N(T^{k-2})$ for $a_1, \ldots, a_m \in \mathbb{F}$ not all of which are zero. Then $T(a_1 x_1 + \cdots + a_m x_m) \in N(T^{k-2})$, which is equivalent to $a_1 x_1 + \cdots + a_m x_m \in N(T^{k-1})$ by (ii). Thus $\{x_1, \ldots, x_m\}$ are linearily dependent over $N(T^{k-1})$. ■

Remark 31 Using Lemma 28 one sees that that (iii) implies

$$\dim N(T^k) - \dim N(T^{k-1}) \leq \dim N(T^{k-1}) - \dim N(T^{k-2}) \qquad \forall k.$$

This strengthens the result $\{0\} \subsetneqq N(T) \subsetneqq N(T^2) \subsetneqq \cdots \subsetneqq N(T^{n-1}) \subsetneqq N(T^n) = V$, known from Lemma 16(ii). □

Proposition 32 *Let V be a finite-dimensional vector space and $T : V \to V$ a linear map that is nilpotent of order n. Then there are natural numbers $r, \ell_1, \ldots, \ell_r$ satisfying $n = \ell_1 \geq \ell_2 \geq \cdots$ and non-zero vectors $\{x_1, \ldots, x_r\} \subset V$ such that*

$$T^{\ell_i} x_i = 0 \neq T^{\ell_i - 1} x_i \qquad \forall i \in \{1, \ldots, r\}$$

and the set (of T-orbits of the x_i)

$$\left\{ T^k x_i \mid i \in \{1, \ldots, r\},\ k \in \{0, 1, \ldots, \ell_i - 1\} \right\} \tag{7}$$

is a basis for V. (Thus $\ell_1 + \cdots + \ell_r = \dim V$.)

The numbers $r, \ell_1, \ldots, \ell_r$ depend only on the dimensions $\dim N(T^k)$ for $k = 1, \ldots, n$.

Proof. If $n = 1$, thus $T = 0$, just put $r = \dim V$, $\ell_1 = \cdots = \ell_r = 1$ and let $\{x_1, \ldots, x_r\}$ be a basis for V. One checks easily that this satisfies the requirements.

Thus assume $n \geq 2$. We know that $N(T^n) = V$. If $N(T^{n-1}) = N(T)$ was true, we would conclude $N(T^{n-1}) = V$, thus $T^{n-1} = 0$, contradicting the minimality of n. Thus $N(T^{n-1}) \subsetneqq N(T^n) = V$. Now put $r_1 = \dim V - \dim N(T^{n-1}) > 0$ and let $\{y_{n,1}, \ldots, y_{n,r_1}\} \in V$ be vectors[5] such that

$$N(T^{n-1}) + \mathrm{span}(y_{n,1}, \ldots, y_{n,r_1}) = N(T^n) = V,$$

which by Lemma 28 is equivalent to $\{y_{n,1}, \ldots, y_{n,r_1}\} \subset V = N(T^n)$ being linearly independent over $N(T^{n-1})$. Thus by Lemma 30, the set $\{Ty_{n,1}, \ldots, Ty_{n,r_1}\} \subset N(T^{n-1})$ is linearly independent over $N(T^{n-2})$. Thus the subspace

$$N(T^{n-2}) + \mathrm{span}(Ty_{n,1}, \ldots, Ty_{n,r_1}) \subset N(T^{n-1})$$

has dimension $\dim N(T^{n-2}) + r_1$, again by Lemma 28. This subspace may or may not be all of $N(T^{n-1})$. The difference of dimensions is

$$r_2 = \dim N(T^{n-1}) - \big(\dim N(T^{n-2}) + r_1\big). \tag{8}$$

If $r_2 > 0$, we pick r_2 vectors $\{y_{n-1,1}, \ldots y_{n-1,r_2}\}$ in $N(T^{n-1})$ such that

$$N(T^{n-2}) + \mathrm{span}(Ty_{n,1}, \ldots, Ty_{n,r_1}, y_{n-1,1}, \ldots y_{n-1,r_2}) = N(T^{n-1}).$$

Now, by construction the set

$$\{Ty_{n,1}, \ldots, Ty_{n,r_1}, y_{n-1,1}, \ldots y_{n-1,r_2}\} \subset N(T^{n-1})$$

is linearly independent over $N(T^{n-2})$. (Notice that (8) is equivalent to $\dim N(T^{n-1}) - \dim N(T^{n-2}) = r_1 + r_2$.) And the subset

$$\{y_{n,1}, \ldots, y_{n,r_1}, Ty_{n,1}, \ldots, Ty_{n,r_1}, y_{n-1,1}, \ldots y_{n-1,r_2}\}$$

is linearly independent over $N(T^{n-2})$ with which it spans V. By Lemma 30, the set

$$\{T^2 y_{n,1}, \ldots, T^2 y_{n,r_1}, Ty_{n-1,1}, \ldots Ty_{n-1,r_2}\} \subset N(T^{n-2})$$

is linearly independent over $N(T^{n-3})$, so that the subspace

$$N(T^{n-3}) + \mathrm{span}(T^2 y_{n,1}, \ldots, T^2 y_{n,r_1}, Ty_{n-1,1}, \ldots Ty_{n-1,r_2}) \subset N(T^{n-2})$$

[5] Here and in what follows, $y_{k,i}$ is in $N(T^k)$, but not in $N(T^{k-1})$.

has dimension $\dim N(T^{n-3}) + r_1 + r_2$. Put

$$r_3 = \dim N(T^{n-2}) - \big(\dim N(T^{n-3}) + r_1 + r_2\big),$$

equivalent to

$$\dim N(T^{n-2}) - \dim N(T^{n-3}) = r_1 + r_2 + r_3.$$

If $r_3 > 0$, pick r_3 vectors $\{y_{n-2,1}, \ldots, y_{n-2,r_3}\}$ in $N(T^{n-2})$ that are linearly independent over $N(T^{n-3})$. This implies

$$N(T^{n-3}) + \mathrm{span}(T^2 y_{n,1}, \ldots, T^2 y_{n,r_1}, T y_{n-1,1}, \ldots T y_{n-1,r_2}, y_{n-2,1}, \ldots, y_{n-2,r_3}) = N(T^{n-2}).$$

Furthermore,

$$\begin{aligned} &\{y_{n,1}, \ldots, y_{n,r_1}\} \cup \{T y_{n,1}, \ldots, T y_{n,r_1}, y_{n-1,1}, \ldots y_{n-1,r_2}\} \\ &\quad \cup \{T^2 y_{n,1}, \ldots, T^2 y_{n,r_1}, T y_{n-1,1}, \ldots T y_{n-1,r_2}, y_{n-2,1}, \ldots, y_{n-2,r_3}\} \end{aligned}$$

is linearly independent over $N(T^{n-3})$ with which it spans V.

Hopefully it will be intelligible that we can iterate this construction, producing further r's and vectors y such that

$$N(T^{n-k}) = N(T^{n-k-1}) + \mathrm{span}\{T^{k-j} y_{n-j,l} \mid 0 \le j \le k,\ 1 \le l \le r_{j+1}\}$$

and

$$V = N(T^{n-k-1}) + \mathrm{span}\{T^i y_{n-j,l} \mid 0 \le j \le k,\ 0 \le i \le k-j,\ 1 \le l \le r_{j+1}\}$$

for all $k \le n-1$. For $k = n-1$, observing $N(T^0) = N(\mathbf{1}) = \{0\}$, the second formula becomes

$$V = \mathrm{span}\{T^i y_{n-j,l} \mid 0 \le j \le n-1,\ 0 \le i \le n-j-1,\ 1 \le l \le r_{j+1}\}. \tag{9}$$

Defining

$$\begin{aligned} r &= r_1 + \cdots + r_n, & (10)\\ \{x_1, \ldots, x_r\} &= \{y_{n,1}, \ldots, y_{n,r_1}, y_{n-1,1} \ldots, y_{n-1,r_2}, \ldots, y_{1,1}, \ldots, y_{1,r_n}\}, & (11)\\ \{\ell_1, \ldots, \ell_r\} &= \{\underbrace{n, \ldots, n}_{r_1 \text{ times}}, \underbrace{n-1, \ldots, n-1}_{r_2 \text{ times}}, \ldots, \underbrace{1, \ldots, 1}_{r_n \text{ times}}\}, & (12) \end{aligned}$$

one finds that $r, \ell_1, \ldots, \ell_r \in \mathbb{N}$ and $x_1, \ldots, x_r \in V$ have the properties claimed in the proposition. In particular

$$\dim V = \sum_{i=1}^{n} r_i(n+1-i) = \ell_1 + \cdots + \ell_r.$$

It is clear from the construction that, while there is freedom in the choice of the vectors $y_{n,i}$, the numbers $r_1, \ldots, r_n$ depend only on the dimensions $\dim N(T^k)$ for $k \in \{1, \ldots, n\}$, since they are determined by

$$r_1 + \cdots + r_k = \dim N(T^{n-k+1}) - \dim N(T^{n-k}) \quad \forall k \in \{1, \ldots, n\}.$$

Thus the same holds for the numbers $r, \ell_1, \ldots, \ell_r$ as defined in (10) and (12). ■

THEOREM 33 *Let V be a finite-dimensional vector space and $T : V \to V$ a nilpotent linear map. Then there is a basis β of V such that $A = [V]_\beta$ is of the block-diagonal form*

$$A = \begin{pmatrix} B_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & B_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & B_r \end{pmatrix} \tag{13}$$

where $B_1, B_2, \ldots, B_r$ are square matrices of possibly different sizes. All consist entirely of zeros except that the elements just above the main diagonal are ones. I.e. $(B_i)_{kl} = \delta_{k+1,l}$, which for the size 4×4 looks like

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{14}$$

(Notice that this implies that A is strictly upper triangular (i.e. upper triangular with only zeros on the diagonal).) The number r of blocks is uniquely determined, and the B_i up to permutation. If we require that the sizes ℓ_i of the blocks are decreasing from top left to bottom right, A is uniquely determined.

Proof. Let $r, \ell_1, \ldots, \ell_r \in \mathbb{N}$ and $x_1, \ldots, x_r \in V$ as provided by Proposition 32. By Proposition 32, for each $i \in \{1, \ldots, r\}$ the subspace $W_i = \text{span}(x_i, Tx_i, T^2x_i, \ldots, T^{\ell_i - 1}x_i)$ is T-invariant and admits $\beta_i = \{T^{\ell_i-1}x_i, T^{\ell_i-2}x_i, \ldots, Tx_i, x_i\}$ as ordered basis. Now T sends the first element of β_i to zero, the second to the first, and so on. Thus the $\ell_i \times \ell_i$ matrix $B_i = [T_{W_i}]_{\beta_i}$ is precisely of the form considered above. Define an ordered basis β for V by collecting the bases $\beta_1, \ldots, \beta_r$ of the subspaces $W_1, \ldots, W_r$ together (in this order), i.e.

$$\{y_1, y_2, \ldots, y_{\dim V}\} = \{T^{\ell_1-1}x_1, \ldots, x_1, T^{\ell_2-1}x_2, \ldots, x_2, \ldots, T^{\ell_r-1}x_r, \ldots, x_r\}.$$

Then $[T]_\beta$ is precisely of the form (13), each block B_i having size ℓ_i.

The uniqueness statement is a direct consequence of the last sentence in the proof of Proposition 32. ■

REMARK 34 1. It is not hard to see that the statement of Theorem 33 not only follows from Proposition 32, but is equivalent to it. Namely, given $T : V \to V$, the existence of a basis β for V such that $[T]_\beta$ has the form given in Theorem 33 with decreasing sizes ℓ_i of the r blocks, there are vectors $\{x_1, \ldots, x_r\}$ satisfying the conditions in Proposition 32. (If $\beta = \{z_1, \ldots, z_{\dim V}\}$, these are $x_1 = z_{\ell_1}, x_2 = z_{\ell_1+\ell_2}$, etc.

2. The above proof is essentially the one in Appendix 15 of the excellent book [7], except that we replaced the use of quotient spaces by the notion of linear (in)dependence over a subspace, the connection being given by Exercise 29. □

9 The Jordan normal form

We are finally in a position to state and easily prove the main result of Sections 6-9:

THEOREM 35 *Let V be a finite-dimensional vector space and $T : V \to V$ a linear map whose characteristic polynomial P_T splits into linear factors (2). Let $\{\lambda_1, \ldots, \lambda_n\}$ be the zeros of P_T (without repetition). Then there are $m_1, \ldots, m_n \in \mathbb{N}$ and a basis β of V such that $A = [V]_\beta$ is block-diagonal with square matrices*

$$J_{1,1}, J_{1,2}, \ldots, J_{1,m_1}, J_{2,1}, \ldots, J_{2,m_2}, J_{3,1}, \ldots, J_{n,m_n} \tag{15}$$

on its diagonal and zeros elsewhere. Here each 'Jordan block' $J_{i,j}$ (with $1 \le i \le n$, $1 \le j \le m_i$) has only zero elements except for λ_i on the diagonal (which may of course be zero, but for at most one i) and 1's above the main diagonal, i.e. $(J_{i,j})_{kl} = \lambda_i \delta_{kl} + \delta_{k+1,l}$. For the size 4×4 this looks like

$$\begin{pmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ 0 & 0 & \lambda_i & 1 \\ 0 & 0 & 0 & \lambda_i \end{pmatrix}. \tag{16}$$

(The sizes of all Jordan blocks must clearly sum up to $\dim V$.)

Proof. As proven in Section 7 (and also Appendix A.2), V is the direct sum of the generalized eigenspaces $K_{\lambda_1}, \ldots, K_{\lambda_n}$. Since $(T - \lambda_i \mathbf{1})_{K_{\lambda_i}}$ is nilpotent, the results of Section 8 apply and there is a basis β_i of K_{λ_i} such that the matrix $[(T - \lambda_i \mathbf{1})_{K_{\lambda_i}}]_{\beta_i}$ is a block diagonal matrix having finitely many, at least one, matrices $B_{i,j}$ of of the form encountered in Theorem 33 on its diagonal. Thus $[T_{K_{\lambda_i}}]_{\beta_i}$ has the same form, except that it has all diagonal elements equal to λ_i instead of zero. Now the (ordered) union $\beta = \beta_1 \cup \cdots \cup \beta_n$ is a basis for V with respect to which $[T]_\beta$ has the claimed form. ■

REMARK 36 1. If T has a Jordan normal form (which is the case if P_T splits), for each generalized eigenspace K_λ, where λ is an eigenvalue of T, there are numbers $r, \ell_1, \ldots, \ell_r$ and vectors $x_1, \ldots, x_r$ (we suppress the dependence on λ) such that

$$(T - \lambda \mathbf{1})^{\ell_i} x_i = 0 \ne (T - \lambda \mathbf{1})^{\ell_i - 1} x_i$$

and

$$\{(T - \lambda \mathbf{1})^k x_i \mid i \in \{1, \ldots, r\},\ k \in \{0, \ldots, \ell_i - 1\}\}$$

is a basis of K_λ. This should be clear from the fact that we applied Theorem 33 to the nilpotent linear map $(T - \lambda \mathbf{1})_{K_\lambda}$. In the book [4], the sets $\{x_i, (T-\lambda)x_i, \ldots, (T-\lambda)^{\ell_i - 1} x_i\}$ are called cycles. (Perhaps 'orbits' would be better since there is no cyclicity involved.)

2. Sections 6 and 8 did not require the splitting of P_T, but in Section 7 it is essential. It had to be used somewhere since it is implied by existence of the Jordan normal form.

3. When the characteristic polynomial of T does not split into linear factors, as for $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ over the reals, T does not have a Jordan normal form. But there still exists a sort of normal form, the 'rational canonical form'. We won't discuss it and refer instead to [4, Section 7.4]. (It also has more prerequisites from the algebra of polynomials.) □

EXERCISE 37 Let V be finite-dimensional and $T : V \to V$ a linear map with splitting P_T. Prove:

(i) The number of Jordan blocks corresponding to the eigenvalue λ equals $\dim E_\lambda$.

(ii) The size of the largest Jordan block corresponding to the eigenvalue λ coincides with the order of nilpotency of $(T - \lambda \mathbf{1})_{K_\lambda}$.

10 Conjugacy of linear maps

Definition 38 *Let V, W be finite-dimensional vector spaces over the same field $\mathbb{F}$ and $T : V \to V$ and $S : W \to W$ linear maps. Then T and S are called conjugate if there is an invertible linear map $U : V \to W$ such that $S = U \circ T \circ U^{-1}$.*

Theorem 39 *Let V, W, T, S as in the definition, and assume that the characteristic polynomials P_T, P_S split. Then T and S are conjugate if and only if both*

(i) T and S have the same set of eigenvalues $\{\lambda_1, \dots, \lambda_N\}$.

(ii) $\dim N((T-\lambda_i)^k) = \dim N((S-\lambda_i)^k)$ for all $i \in \{1, \dots, N\}$ and $k \in \mathbb{N}$.

Proof. ($\Rightarrow$) Let $U : V \to W$ be such that $S = U \circ T \circ U^{-1}$. Then

$$P_S(t) = \det(S - t\mathbf{1}_W) = \det\left(U \circ (T - t\mathbf{1}_V) \circ U^{-1}\right) = \det\left(T - t\mathbf{1}_V\right) = P_T(t).$$

This clearly implies that T and S have the same eigenvalues (and multiplicities), thus (i) holds. Now (omitting the $\circ$'s)

$$(S - \lambda_i \mathbf{1}_W)^k = (UTU^{-1} - \lambda_i \mathbf{1}_W)^k = (U(T - \lambda_i \mathbf{1}_V)U^{-1})^k = U(T - \lambda_i \mathbf{1}_V)^k U^{-1}.$$

Using this and the invertibility of U, we have

$$\begin{aligned} y \in N((S - \lambda_i \mathbf{1}_W)^k) \quad &\Leftrightarrow \quad U(T - \lambda_i \mathbf{1}_V)^k U^{-1} y = 0 \Leftrightarrow (T - \lambda_i \mathbf{1}_V)^k U^{-1} y = 0 \\ &\Leftrightarrow \quad U^{-1} y \in N((T - \lambda_i \mathbf{1}_V)^k) \Leftrightarrow y \in U[N((T - \lambda_i \mathbf{1}_V)^k)], \end{aligned}$$

implying $\dim N((S - \lambda_i \mathbf{1}_W)^k) = \dim N\big((T - \lambda_i \mathbf{1}_V)^k\big)$.

($\Leftarrow$) Pick and fix (for now) λ_i an eigenvalue (of S and T). Let $K_{\lambda_i} \subseteq V$ and $K'_{\lambda_i} \subseteq W$ be the corresponding generalized eigenspaces of T and S, respectively. For k, k' large enough, $N((T - \lambda_i \mathbf{1}_V)_{K_{\lambda_i}}) = K_{\lambda_i}$ and $N((S - \lambda_i \mathbf{1}_W)_{K'_{\lambda_i}}) = K'_{\lambda_i}$, thus (ii) implies $\dim K_{\lambda_i} = \dim K'_{\lambda_i}$. (Thus also $\dim V = \dim W$, which is needed if T, S are to be conjugate.)

Now the linear maps $(T - \lambda_i \mathbf{1}_V)_{K_{\lambda_i}}$ and $(S - \lambda_i \mathbf{1}_W)_{K'_{\lambda_i}}$ are nilpotent of the same order

$$\begin{aligned} n &= \min\left\{m \mid \dim N((T - \lambda \mathbf{1}_V)^m) = \dim N((T - \lambda \mathbf{1}_V)^{m+1})\right\} \\ &= \min\left\{m \mid \dim N((S - \lambda \mathbf{1}_W)^m) = \dim N((S - \lambda \mathbf{1}_W)^{m+1})\right\}, \end{aligned}$$

where the second equality follows from assumption (ii).

Now the numbers $r, \ell_1 \geq \ell_2 \geq \cdots \geq \ell_n$ obtained by application of Proposition 32 to $(T - \lambda_i \mathbf{1}_V)_{K_{\lambda_i}}$ depend only on the dimensions $\dim N((T - \lambda_i \mathbf{1}_V)^k)_{K_{\lambda_i}}$. By assumption, the latter coincide with $\dim N((S - \lambda_i \mathbf{1}_W)^k)_{K'_{\lambda_i}}$. Thus the nilpotent linear maps $(T - \lambda_i \mathbf{1}_V)^k)_{K_{\lambda_i}}$ and $(S - \lambda_i \mathbf{1}_W)^k)_{K'_{\lambda_i}}$ have the same normal forms, compare Theorem 33, consisting of r Jordan blocks (with zeros on the diagonal) of sizes $\ell_1 \geq \ell_2 \geq \cdots$. Thus with $d_i = \dim K_{\lambda_i} = \dim K'_{\lambda_i}$, there are ordered bases β_i of K_{λ_i} and β'_i of K_{λ_i} such that $[T_{K_{\lambda_i}}]_{\beta_i} = [S_{K'_{\lambda_i}}]_{\beta'_i}$.

Since the above holds for each eigenvalue $\lambda_1, \dots, \lambda_n$, putting $\beta = \beta_1 \cup \cdots \cup \beta_n$ and $\beta' = \beta'_1 \cup \cdots \cup \beta'_n$, we find $[T]_\beta = [S]_{\beta'}$. It is an easy exercise to deduce from this that T and S are conjugate, thus there exists an invertible $U : V \to W$ such that $S = U \circ T \circ U^{-1}$. ■

Exercise 40 Prove the statement in the last sentence of the preceding proof.

11 Computations using the JNF

A more practical application of the Jordan normal form is the computation of functions $f(T)$ of a linear map T (or of a matrix $f(A)$). We begin with $f = p$, a polynomial. Of course, for a polynomial, $p(A)$ is not difficult to compute in principle, but the result will easily be generalized to power series, for which the naive computation of $f(A)$ would require an infinite amount of computation.

Let V be a finite-dimensional vector space and $T : V \to V$ a linear map. By Theorem 35 there is a basis $\beta = \{e_1, \ldots, e_d\}$ for V such that the matrix $A = [T]_\beta$ is block diagonal with blocks of the form (16), where each eigenvalue λ appears in at least one of these block. For the purpose of this section, this is all we need. Once we have made sense of the matrix $f(A) = [f(T)]_\beta$, it is clear how to obtain $f(T)$ as linear map $V \to V$ by having it send $\sum_{j=1}^{d} a_j e_j \in V$ to $\sum_{i,j=1}^{d} a_j (f(A))_{ij} e_i$.

It is clear enough that a polynomial acts by sending the Jordan blocks on the diagonal of $A = [T]_\beta$ to

$$p(J_{1,1}), p(J_{1,2}), \ldots, p(J_{1,m_1}), p(J_{2,1}), \ldots, p(J_{2,m_2}), p(J_{3,1}), \ldots, p(J_{n,m_n}).$$

It remains to compute a convenient expression for $p(J)$ for a Jordan block $J = \lambda\mathbf{1} + N$, where $N_{ij} = \delta_{i+1,j}$ as in (14). We need two easy facts:

Exercise 41 For $k \in \mathbb{N}_0$ define $D \times D$ matrices N_k by $(N_k)_{ij} = \delta_{i+k,j}$. Prove

(i) $(N_1)^k = N_k$ for each k.

(ii) $N_k = 0$ if $k \geq D$.

Exercise 42 Let V be finite-dimensional vectorspace and $T, S : V \to V$ linear maps. Prove that the following are quivalent:

(i) $TS = ST$.

(ii) $(S + T)^2 = S^2 + 2ST + T^2$ (i.e. (iii) for $n = 2$).

(iii) $(S + T)^n = \sum_{k=0}^{n} \binom{n}{k} S^k T^{n-k}$ holds for all $n \in \mathbb{N}$.

Given a polynomial $p(x) = a_0 + a_1 x + \cdots + a_M x^M$ and using the above, noting $N_1 = N$, we compute

$$\begin{aligned} p(\lambda\mathbf{1} + N_1) &= \sum_{m=0}^{M} a_m (\lambda\mathbf{1} + N_1)^m = \sum_{m=0}^{M} a_m \sum_{k=0}^{m} \binom{m}{k} \lambda^{m-k} (N_1)^k \\ &= \sum_{m=0}^{M} a_m \sum_{k=0}^{m} \binom{m}{k} \lambda^{m-k} N_k = \sum_{k=0}^{M} N_k \sum_{m=k}^{M} a_m \binom{m}{k} \lambda^{m-k}, \end{aligned}$$

where we used that the summations run over the (m, k) satisfying $0 \leq k \leq m \leq M$ to swap the order of summations. With $\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m(m-1)(m-2)\cdots(m-k+1)}{k!}$, we continue the computation

$$= \sum_{k=0}^{M} \frac{N_k}{k!} \sum_{m=k}^{M} a_m\, m(m-1)(m-2)\cdots(m-k+1)\lambda^{m-k}.$$

Noticing that $\sum_{m=k}^{M} a_m\, m(m-1)(m-2)\cdots(m-k+1)x^{m-k}$ is the k-th derivative $p^{(k)}$ of $p(x) = a_0 + a_1 x + \cdots + a_M x^M$, we have proven

$$p(\lambda \mathbf{1} + N_1) = \sum_{k=0}^{M} \frac{p^{(k)}(\lambda)}{k!} N_k, \tag{17}$$

which is the upper triangular matrix whose matrix elements with $i \leq j$ are given by

$$[p(\lambda \mathbf{1} + N_1)]_{ij} = \frac{p^{(j-i)}(\lambda)}{(j-i)!}.$$

Now it is easy to generalize the computation to the case where f is given by a power series $\sum_{m=0}^{\infty} a_m x^m$. We must of course assume that the series converges for x being any of the eigenvalues λ_i of T. One can prove (typically in complex analysis) that

$$\sum_{m=k}^{\infty} a_m\, m(m-1)(m-2)\cdots(m-k+1)x^{m-k}$$

converges to the k-th derivative of f. Thus (17) continues to hold with p replaced by f and M replaced $D-1$, where D is the size of the Jordan block under consideration since $N_k = 0$ for $k \geq D$. Note that this sum is finite even though f is defined by an infinite series!

An interesting special case is $f = \exp : x \mapsto e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!}$. Since $\exp' = \exp$, induction gives $\exp^{(k)} = \exp$, so that (17) simplifies to

$$\exp(\lambda \mathbf{1} + N) = e^{\lambda} \sum_{k=0}^{D-1} \frac{N_k}{k!},$$

where again the sum is finite since $N_k = 0$ for $k \geq D$. (Interestingly, this can also be proven using $e^{A+B} = e^A e^B$, which holds whenever $AB = BA$. But this property only holds for the exponential function.)

In the context of solving differential equations, we are more interested in the function $g(t) = e^{tA}$, which satisfies the differential equation $\frac{d}{dt} g(t) = A g(t)$. For each Jordan block of T (of size $D \times D$) it is given by

$$t \mapsto e^{t\lambda} \sum_{k=0}^{D-1} \frac{t^k}{k!} N_k = e^{t\lambda} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{D-1}}{(D-1)!} \\ & 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{D-2}}{(D-2)!} \\ & & 1 & t & \cdots & \frac{t^{D-3}}{(D-3)!} \\ & & & 1 & \cdots & \cdots \\ & & & & \cdots & t \\ & & & & & 1 \end{pmatrix}.$$

12 The minimal polynomial

My ideas concerning the minimal polynomial don't deviate much from the treatment of Friedberg/Insel/Spence. For this reason I just enumerate the most important points, including proofs. For examples, see the book.

- In order to avoid pathologies, we assume $\dim V \geq 1$ thoughout.

- Let V be a finite-dimensional vector space and $T : V \to V$ a linear map. Since the space $L(V, V)$ of linear maps $V \to V$ is finite-dimensional of dimension $(\dim V)^2$, we clearly have
$$\dim \operatorname{span}\{\mathbf{1}_V, T, T^2, T^3, \dots\} \leq (\dim V)^2.$$
Thus there exists a lowest n such that T^n is a linear combination of $\mathbf{1}, T, \dots, T^{n-1}$, which we can write as $P(T) = 0$ with $P(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. The polynomial P satisfying $P(T) = 0$, having the lowest possible degree n and having 1 as coefficient[6] of the highest power t^n is uniquely determined and is called the minimal polynomial P_T^m of T.
- By the above considerations we know that the degree of P_T^m cannot be higher than $(\dim V)^2$. But by the Cayley-Hamilton theorem we already know that $P_T(T) = 0$, where P_T is the characteristic polynomial of T, having degree exactly $\dim V$. Thus clearly $\deg P_T^m \leq \deg P_T = \dim V$.
- The degree of P_T^m clearly cannot be zero. And $\deg P_T^m = 1$ if and only if $P_T^m(t) = t - c$ for some constant c, which is equivalent to $T = c\mathbf{1}_V$. (Recall that for such T we have $P_T(t) = (c - t)^{\dim V}$.)
- Now assume $\deg P_T^m$ is maximal, thus equals $\dim V$. Then $(-1)^{\dim V} P_T$ and P_T^m both are monic polynomials vanishing on T, thus the polynomial $Q = (-1)^{\dim V} P_T - P_T^m$ satisfies $Q(T) = 0$ and has lower degree than P_T^m. Since this is impossible for $Q \neq 0$ by definition of P_T^m, in this case we have $P_T^m = (-1)^{\dim V} P_T$.
- Every eigenvalue of T (=zero of P_T) is also a zero of P_T^m.

  Proof. If λ is an eigenvalue of T, there is a non-zero vector x such that $Tx = \lambda x$. This implies $T^k x = \lambda^k x$ for all $k \in \mathbb{N}_0$ and therefore $P(T)x = P(\lambda)x$ for each polynomial P. For $P = P_T^m$, we have $P_T^m(\lambda)x = P_T^m(T)x = 0$. Since $x \neq 0$, this implies $P_T^m(\lambda) = 0$, as claimed. ■

- If T has $d = \dim V$ distinct eigenvalues $\{\lambda_1, \dots, \lambda_d\}$, the minimal polynomial is uniquely determined by having the λ_i as zeros, being monic and of degree $\leq d$:
$$P_T^m(t) = (t - \lambda_1) \cdots (t - \lambda_d) = (-1)^d P_T(t).$$
- In general the degree of P_T^m can be lower than that of P_T. In fact, P_T^m divides P_T.

  Proof. The division algorithm for polynomials (see Proposition 43) gives unique polynomials Q, R such that $P_T = QP_T^m + R$ and $\deg R < \deg P_T^m$. Evaluating this identity at T gives $P_T(T) = Q(T)P_T^m(T) + R(T)$. With $P_T^m(T) = 0$ (by definition of P_T^m) and $P_T(T) = 0$ (by the Cayley-Hamilton theorem) this reduces to $R(T) = 0$. If now R was non-zero, it would be a polynomial satisfying $R(T) = 0$ having lower degree than P_T^m, contradicting the definition of P_T^m. Thus $R = 0$, so that $P_T = QP_T^m$, which is just the statement that P_T^m divides P_T. ■

- Every zero of P_T^m also is a zero of P_T. Thus P_T and P_T^m have the same zeros.

  Proof. The first statement follows from the fact that $P_T = QP_T^m$ for some polynomial Q. Now the second statement is immediate since we already know that every zero of P_T is a zero of P_T^m. ■

[6] A polynomial in which the highest power of the variable has coefficient one is called monic.

- Now assume that P_T splits and that $\{\lambda_1, \ldots, \lambda_N\}$ are the distinct zeros of P_T with multiplicities m_i. Then P_T^m splits (by Proposition 6(f)) and (again using that P_T^m is monic)

$$P_T^m(t) = \prod_{i=1}^{N} (t - \lambda_i)^{n_i}, \tag{18}$$

with $1 \le n_i \le m_i$. It remains to determine the numbers n_i. It turns out that they have a natural description in terms of the Jordan normal form of T: For every $i \in \{1, \ldots, N\}$, the exponent n_i equals the order of nilpotency of $(T - \lambda_i \mathbf{1})_{K_{\lambda_i}}$. (Recall that the latter coincides with the size of the largest Jordan block corresponding to the eigenvalue λ_i.)

Proof. Since we already know the general form (18) of P_T^m, it only remains to find the lowest numbers $n_i \ge 1$ such that $P^m(T) = \prod_{i=1}^N (T - \lambda \mathbf{1}_V)^{n_i} = 0$. Since $V = \bigoplus_{i=1}^N K_{\lambda_i}$ and $T - \lambda_i \mathbf{1}_V$ acts invertibly on each K_{λ_j} with $j \ne i$ (see Lemma 19(iii)), we want the lowest n_i's such that $(T - \lambda_i \mathbf{1}_V)^{n_i}$ is zero on K_{λ_i}. These n_i's are precisely the orders of nilpotency of $(T - \lambda_i \mathbf{1}_V)_{K_{\lambda_i}}$. ■

- In view of the preceding result we have $n_i = 1$ if and only if $(T - \lambda_i \mathbf{1})_{K_{\lambda_i}} = 0$, which is equivalent to $K_{\lambda_i} = E_{\lambda_i}$. Since T is diagonalizable if and only if $K_{\lambda_i} = E_{\lambda_i}$ for each i, we find that diagonalizability is equivalent to $P_T^m(t) = \prod_{i=1}^N (t - \lambda_i)$ (the λ_i still being mutually distinct).
- We have $n_i < m_i$ precisely if the generalized eigenspace K_{λ_i} contains more than one cycle, which is equivalent to the JNF having more than one Jordan block corresponding to λ_i.
- The above results show that the polynomials P_T and P_T^m contain different information about the Jordan normal form: Both encode the eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$ of T. But the exponent m_i of $t - \lambda_i$ in P_T is the dimension of the generalized eigenspace K_{λ_i}, while the exponent n_i of $t - \lambda_i$ in P_T^m is the order of nilpotency of $(T - \lambda_i \mathbf{1})_{K_{\lambda_i}}$. There are other pieces of information like the exact number of cycles in K_{λ_i} that can in general not be obtained from P_T or P_T^m.
- There are authors, like [2], who for aesthetic/pedagogical/ideological reasons prefer to introduce the minimal polynomial before (or even instead of) the characteristic polynomial.

A Alternative approach to proving $V = \bigoplus_\lambda K_\lambda$

A.1 Some further results on polynomials

We begin with two results proven in Appendix E of [4]. I don't like the complicated inductive proofs given there (in particular for the Lemma preceding Theorem E.2), so here are alternative ones.

You should have seen the following result already in Calculus and/or Analysis 1.[7]

PROPOSITION 43 (POLYNOMIAL DIVISION) *Let $\mathbb{F}$ be any field and f, g polynomials with coefficients in $\mathbb{F}$ and $\deg g \ge 1$. Then there are unique polynomials q, r such that $f = qg + r$ and $\deg r < \deg g$.*

[7] It appears in [1, Section P.6], though not very clearly. A better treatment can be found in [8, Stelling 3.1].

Proof. We begin by proving the uniqueness. If q, q', r, r' are polynomials such that $f = qg + r = q'g + r'$ and $\deg r < \deg g$, $\deg r' < \deg g$ then $(q - q')g = r' - r$. The r.h.s. is a polynomial of degree less than $\deg g$, whereas the l.h.s. has degree at least $\deg g$ whenever $q - q' \neq 0$. Since this is absurd, we have $q = q'$, which in turn implies $r = r'$.

We now prove existence. If $\deg f < \deg g$ we put $q = 0, r = f$ and are done. Now assume $\deg f \geq \deg g$. Let ax^n and bx^m be the highest order monomials in f and g, respectively. Then $n = \deg f \geq \deg g = m$. Now $f_1 = f - \frac{a}{b}x^{n-m}g$ is a polynomial of degree at most $(\deg f) - 1$ since the powers x^n cancel. If $\deg f_1 < \deg g$, we are done. If not, iterate this construction by defining f_2 as f_1 minus a suitable scalar multiple of $g \cdot x^{\deg f_1 - \deg g}$ so that again the highest power of x cancels. After finitely many steps we reach $f_n = f - qg$ with $\deg f_n < \deg g$. Putting $r = f_n$, we are done. (The proof of [4, Theorem E.1] is essentially the same, except that it is stated more formally as an inductive argument.) ■

Proposition 44 *Let f, g be polynomials over the field $\mathbb{F}$ that are relatively prime. (I.e. there is no polynomial h of degree ≥ 1 that divides f and g without remainder.) Then there are polynomials a, b such that $af + bg = 1$.*

Proof. Let $S = \{af + bg \mid a, b \text{ polynomials}\}$, which is a linear subspace of the space of all polynomials over $\mathbb{F}$. Let D be the smallest degree of the non-zero elements of S, and pick $h \in S$ with $\deg h = D$. By Proposition 43 there exist polynomials q, r with $f = qh + r$ and $\deg r < \deg h$. Since f and h are in S, so is $r = f - qh$. If $r \neq 0$, we have a contradiction with the definition of D since $\deg r < \deg h$. Thus $r = 0$, so that h divides f.

In the same way one shows that h divides g. Since f and g are relatively prime, the polynomial h must have degree zero, i.e. is a non-zero constant. Since S is closed under multiplication by scalars, it contains the constant one function, and this is equivalent to the claim. ■

Remark 45 This proof is very efficient, but utterly non-constructive in that it does not explicitly give us a, b. Yet, there also is an algorithmic proof that allows to find polynomials a, b as claimed. It amounts to a polynomial version of Euclid's algorithm for computing the GCD of two integers. (As you might learn in Rings and Fields, behind this is the fact that the ring of integers $\mathbb{Z}$ and the polynomial rings $\mathbb{F}[x]$ over fields share the property of being 'Euclidean rings'. See e.g. [8].) □

Theorem 46 *Let $n \geq 2$ and let $f_1, \ldots, f_n$ be polynomials that are relatively prime. Then there are polynomials $a_1, \ldots, a_n$ such that $a_1 f_1 + \cdots + a_n f_n = 1$.*

Proof. The proof given in [4], where it is Theorem E.2 in Appendix 2, is hard to improve upon, but we include it for the sake of self-containedness of this note. It proceeds by induction over n, starting with Proposition 44, giving the case $n = 2$.

Assume the claim holds for some n and let $f_1, \ldots, f_{n+1}$ be relatively prime polynomials. In the case where $f_1, \ldots, f_n$ are relatively prime, we can apply the induction hypothesis and get $a_1, \ldots, a_n$ such that $a_1 f_1 + \cdots + a_n f_n = 1$. Putting $a_{n+1} = 0$, also $a_1 f_1 + \cdots + a_{n+1} f_{n+1} = 1$ holds, and we are done.

In the case where $f_1, \ldots, f_n$ are not relatively prime, let h be a polynomial dividing $f_1, \ldots, f_n$ of maximal degree. Then there are polynomials $g_1, \ldots, g_n$ such that $f_i = hg_i$ for $i = 1, \ldots, n$. Now the g_i are relatively prime. [Otherwise there would be a polynomial $\tilde{h}$ of degree ≥ 1 dividing all of them, but then $h\tilde{h}$ would divide all f_i and have degree larger than

h, contradicting the choice of h.] Thus by the induction hypothesis there are polynomials $\tilde{a}_1, \ldots, \tilde{a}_n$ such that $\tilde{a}_1 g_1 + \cdots + \tilde{a}_n g_n = 1$. Multiplying this identity by h we get

$$\tilde{a}_1 f_1 + \cdots + \tilde{a}_n f_n = \tilde{a}_1 h g_1 + \cdots + \tilde{a}_n h g_n = h. \tag{19}$$

Now, h and f_{n+1} are relatively prime, since a common divisor would divide all $f_1, \ldots, f_{n+1}$, contradicting the assumption of relative primality. Thus by the case $n = 2$ (Proposition 44) there are polynomials b, c such that $bh + cf_{n+1} = 1$. Plugging into this the formula (19) for h, we obtain

$$b(\tilde{a}_1 f_1 + \cdots \tilde{a}_n f_n) + cf_{n+1} = 1,$$

so that with $a_i = b\tilde{a}_i$ for $i \leq n$ and $a_{n+1} = c$, we have $\sum_{i=1}^{n+1} a_i f_i = 1$, as desired. ■

A.2 Alternative proof of $\sum_\lambda K_\lambda = V$

THEOREM 47 *Let V be a finite dimensional vector space of dimension $d = \dim V$ and $T : V \to V$ a linear map whose characteristic polynomial P_T factorizes into d linear factors, i.e. (2) holds. (Then of course $\sum_i m_i = d$.) Then $V = \bigoplus_{i=1}^n K_{\lambda_i}$, i.e. every $x \in V$ can uniquely be written as $x = x_1 + \cdots + x_n$, where $x_i \in K_{\lambda_i}$ for each i.*

Proof. For each $j \in \{1, \ldots, n\}$ define[8]

$$f_i(t) = \prod_{\substack{j=1 \\ j \neq i}}^{n} (t - \lambda_i)^{m_i}.$$

We claim that the polynomials $f_1, f_2, \ldots, f_n$ are relatively prime in the sense that there is no polynomial of degree ≥ 1 dividing all the f_i. Since all f_i split, also a polynomial g dividing all of them would split by Proposition 6(f), and every $t - c$ appearing in the factorization of g would divide all f_i. But by construction, $\lambda_i \notin f_i^{-1}(0)\ \forall i$, implying $\bigcap_{i=1}^n f_i^{-1}(0) = \emptyset$. Thus g must be constant, proving the claim. Thus by Theorem 46 there are polynomials $a_1, \ldots, a_n$ such that

$$a_1 f_1 + \cdots + a_n f_n = 1.$$

This implies the identity

$$a_1(T) f_1(T) + \cdots + a_n(T) f_n(T) = \mathbf{1}_V$$

of linear maps on V, thus for every $v \in V$, we have

$$a_1(T) f_1(T) v + \cdots + a_n(T) f_n(T) v = v.$$

Now for each i we have $(t - \lambda_i)^{m_i} f_i(t) = \prod_{j=1}^n (t - \lambda_j)^{m_j} = (-1)^d P_T(t)$. Since $P_T(T) = 0$ by the Cayley-Hamilton theorem [4, Theorem 5.22], we have $(T - \lambda_i \mathbf{1})^{m_i} f_i(T) = 0$, so that

$$(T - \lambda_i \mathbf{1})^{m_i} a_i(T) f_i(T) v = (-1)^d a_i(T) P_T(T) v = 0.$$

Thus $v_i := a_i(T) f_i(T) v$ is in K_{λ_i} for each i. This shows that every $v \in V$ can be written as a sum of elements of the K_{λ_i}. We already know from Proposition 20 that such a decomposition is unique, thus $V = \bigoplus_{i=1}^n K_{\lambda_i}$. ■

[8] Note that $f_i(T)$ equals the linear map $\prod_{\substack{j=1 \\ j \neq i}}^{n} (T - \lambda_i)^{m_i}$ appearing in the proof of Proposition 20.

Remark 48 Defining $P_i = a_i(T)f_i(T)$ for each $i \in \{1, \dots, n\}$, it follows that P_i acts on K_{λ_i} as the identity map and as the zero map on K_{λ_j} with $j \neq i$. From this it is not hard to deduce that the P_i (which all commute), satisfy $\sum_i P_i = \mathbf{1}_V$ and $P_i^2 = \mathbf{1}_V$ for each i and $P_iP_j = 0$ whenever $i \neq j$.

2. Our second proof of the completeness Theorem 25/47 is longer than the first, but it is not as unnatural as it may look. After all, every linear map $T : V \to V$ on a finite-dimensional vector space gives rise to a ring homomorphism $p \mapsto p(T)$ from the ring $k[X]$ of polynomials to the ring $\operatorname{End} V$ of linear maps from V to V. (You're not expected to understand all the terms.) So it is only natural to use some of the theory of polynomials, exploiting the nice properties of $k[X]$. And the proof of Theorem 25 also involved a (very elementary) bit of divisibility theory of $k[X]$. (The full theory uses Propositions 43 and 44.) □

References

[1] R. A. Adams, C. Essex: *Calculus*. 8th.ed. Pearson, 2013.

[2] S. Axler: *Linear algebra done right*. 4th. ed. Springer, 2024.

[3] G. Fischer: *Lernbuch Lineare Algebra und Analytische Geometrie*. 4. Aufl. Springer, 2019.

[4] S. H. Friedberg, A. J. Insel, L. E. Spence: *Linear algebra*. 4th or 5th edition. Pearson, 2003, 2019.

[5] P. R. Halmos: *Finite-dimensional vector spaces*. 2nd. ed. Springer, 1974.

[6] S. Lang: *Linear algebra*. 3rd. ed. Springer, 1987.

[7] P. D. Lax: *Linear algebra and its applications*. 2nd. ed. Wiley-Interscience, 2007.

[8] H.W. Lenstra Jr., F. Oort: *Ringen en lichamen*, 198?, 2014.