The Jordan canonical form

Attila Máté
Brooklyn College of the City University of New York

November 14, 2014

Contents

1 Preliminaries

1.1 Rank-nullity theorem

Let U and V be vector spaces over a field F, and let $T : U \to V$ be a linear transformation. Then the domain of T, denoted as $\mathrm{dom}(T)$, is the set U itself. The range of T is the set $\mathrm{ra}(T) = \{T\mathbf{u} : \mathbf{u} \in U\}$, and it is a subspace of V; one often writes TU instead of $\mathrm{ra}(T)$. The kernel, or null space, of T is the set $\ker(T) = \{\mathbf{u} \in U : T\mathbf{u} = \mathbf{0}\}$; the kernel is a subspace of U. A key theorem is the following, often called the Rank-Nullity Theorem:

Theorem 1. *Given vector spaces U and V over F and a linear transformation $T : U \to V$, we have*

$$\dim(U) = \dim(\ker(T)) + \dim(\mathrm{ra}(T)). \tag{1}$$

Proof. Let $\mathbf{u}_1$, $\mathbf{u}_2$, …, $\mathbf{u}_k$, and $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_l$, be vectors in U such that $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ is a basis of $\ker(T)$, and $(T\mathbf{v}_1, T\mathbf{v}_2, \dots, T\mathbf{v}_l)$ is a basis of $\mathrm{ra}(T)$. We claim that then

$$(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l) \tag{2}$$

is a basis of U. This claim will establish the result, since according to this claim, if we have $\dim(\ker(T)) = k$ and $\dim(\mathrm{ra}(T)) = l$ then we have $\dim(U) = k + l$.

First we will show that the system in (2) is linearly independent. To this end, assume that with some scalars α_i and β_j for i with $1 \le i \le k$ and for j with $1 \le j \le l$ we have

$$\sum_{i=1}^{k} \alpha_i \mathbf{u}_i + \sum_{j=1}^{l} \beta_j \mathbf{v}_j = \mathbf{0}.$$

Applying T to both sides of equation and noting that $T\mathbf{u}_i = \mathbf{0}$ we have[1]

$$\sum_{j=1}^{l} \beta_j T\mathbf{v}_j = \mathbf{0}.$$

As the system $(T\mathbf{v}_1, T\mathbf{v}_2, \dots, T\mathbf{v}_l)$ is linearly independent, it follows that $\beta_j = 0$ for all j with $1 \le j \le l$. Hence the previous displayed equation becomes

$$\sum_{i=1}^{k} \alpha_i \mathbf{u}_i = \mathbf{0}.$$

As the system $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ is a basis of $\ker(U)$, it follows that $\alpha_i = 0$ for all i with $1 \le i \le k$ This shows that the system given in (2) is linearly independent.

Next we will show that this system also spans U. To this end, let $\mathbf{v} \in U$ be arbitrary. Then $T\mathbf{v} \in \mathrm{ra}(T)$, and so there are scalars β_j for j with $1 \le j \le l$ such that

$$T\mathbf{v} = \sum_{j=1}^{l} \beta_j T\mathbf{v}_j,$$

[1] We use $\mathbf{0}$ for the zero vector to distinguish it from the zero scalar; however, it is difficult to maintain consistent notation. For example, the last $\mathbf{0}$ is the zero vector of the space V, while the $\mathbf{0}$ at the end of the preceding displayed equation is the zero vector of the space U. If one is too fastidious, one may write $\mathbf{0}_U$ and $\mathbf{0}_V$ for these different zeros.

because the system $(T\mathbf{v}_1, T\mathbf{v}_2, \dots, T\mathbf{v}_l)$ spans $\mathrm{ra}(T)$. That is, we have

$$\mathbf{0} = T\mathbf{v} - \sum_{j=1}^{l} \beta_j T\mathbf{v}_j = T\Big(\mathbf{v} - \sum_{j=1}^{l} \beta_j \mathbf{v}_j\Big).$$

Hence the vector $\mathbf{v} - \sum_{j=1}^{l} \beta_j \mathbf{v}_j$ is in the kernel of T. As the system $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$ spans $\ker(T)$, we have

$$\mathbf{v} - \sum_{j=1}^{l} \beta_j \mathbf{v}_j = \sum_{i=1}^{k} \alpha_i \mathbf{u}_i$$

for some scalars α_i for i with $1 \le i \le k$. That is,

$$\mathbf{v} = \sum_{j=1}^{l} \beta_j \mathbf{v}_j + \sum_{i=1}^{k} \alpha_i \mathbf{u}_i$$

for some scalars α_i for i with $1 \le i \le k$ and β_j for j with $1 \le j \le l$. This shows that the system in (2) also spans U. This establishes the claim, completing the proof of the theorem. □

1.2 Existence of eigenvalues and eigenvectors

Let V be a vector space over a field F, and $T : V \to V$ a linear transformation. The elements of F are called scalars. The vector $\mathbf{u}$ is called an *eigenvector* of T if $\mathbf{u} \ne 0$ and $T\mathbf{u} = \lambda\mathbf{u}$ for some scalar λ; the scalar λ is called the eigenvalue associated with $\mathbf{u}$. We have the following

Theorem 2. *Let $V \ne \{\mathbf{0}\}$ be a finite dimensional vector space over an algebraically closed field F, and let $T : V \to V$ be a linear transformation. Then T has an eigenvalue.*

Proof. Let $\mathbf{v} \in V$ be a nonzero vector, and form the vectors $T^k\mathbf{v}$ for nonnegative integers k. Let $n \ge 0$ be the smallest integer for which the system $(T^k\mathbf{v} : 0 \le k \le n)$ is linearly dependent. Clearly, $n > 0$, since $\mathbf{v} \ne \mathbf{0}$; further, n cannot be greater than the dimension of V. Since this system is linearly dependent, there are scalars α_k for k with $0 \le k \le n$ such that

$$\sum_{k=0}^{n} \alpha_k T^k \mathbf{v} = \mathbf{0};$$

here $\alpha_n \ne 0$ in view of the minimality of n. Write $P(x) = \sum_{k=0}^{n} \alpha_k x^k$, where x is an unknown; $P(x)$ is a polynomial of degree n over F.

As F is algebraically closed, there is a $\lambda \in F$ such that $P(\lambda) = 0$. Let $Q(x)$ be the polynomial of degree $n-1$ over F such that $P(x) = (x - \lambda)Q(x)$. In view of the minimality of n, we have $Q(T)\mathbf{v} \ne \mathbf{0}$ and $P(T)\mathbf{v} = (T - \lambda)Q(T)\mathbf{v} = \mathbf{0}$. That is, $T\ Q(T)\mathbf{v} = \lambda\ Q(T)\mathbf{v}$, where $Q(T)\mathbf{v}$ is a nonzero vector. Thus $Q(T)\mathbf{v}$ is an eigenvector of T with λ as the associated eigenvalue. □

2 Existence of a Jordan decomposition

Let F be a field; in what follows, by *vector space* we will mean a vector space over F. An element of F will also be referred to as a scalar. If X and Y are subspaces of V such that $X \cap Y = \{\mathbf{0}\}$ and $X \cup Y$ span V, then we will write $V = X \oplus Y$, and we will say that V is the *direct sum* of X and Y. If $T : V \to V$ is a linear transformation and X is a subspace of V, we call X *invariant* for T if $TX \overset{def}{=} \{T\mathbf{x} : \mathbf{x} \in X\} \subset X$. We have:

Lemma 1. *Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation such that 0 is an eigenvalue of T. Then there is a vector $\mathbf{u} \in V$, a positive integer k, and subspaces U and W of V such that $U \oplus W = V$, U and W are invariant subspaces for T, the system*

$$(T^i\mathbf{u} : 0 \le i < k)$$

is a basis of U, and $T^k\mathbf{u} = \mathbf{0}$.

Proof. We use induction on the dimension of V; so assume that the statement is true for any vector space having a smaller dimension than V. Since 0 is an eigenvalue of T, the kernel of T contains a nonzero vector, and so TV has lower dimension than V by Theorem 1.

That is, the statement of the lemma is true for TV replacing V provided that TV satisfies the assumptions. If the assumptions of the lemma are true for TV replacing V, let U' and W' be the subspaces of TV satisfying the assumptions of the lemma with TV replacing V. Further, let $\mathbf{u}' \in U'$ and $k' > 0$ be such that

$$S_{U'} = (T^i\mathbf{u}' : 0 \le i < k') \tag{3}$$

is a basis of U' and $T^{k'}\mathbf{u}' = \mathbf{0}$. The only way for T restricted to TV not to satisfy the assumptions is that 0 is not an eigenvalue of T restricted to TV (this includes also the special case when $TV = \{\mathbf{0}\}$). This case will start the induction; in this case, we will put $U' = \{0\}$, $W' = TV$, $k' = 0$, and $\mathbf{u}'$ will not be defined.

If $U' = \{\mathbf{0}\}$, then let $\mathbf{u} \in V \setminus TV$ be such that $T\mathbf{u} = \mathbf{0}$; there is such a $\mathbf{u}$ since 0 is an eigenvalue of T. If $U' \ne \{\mathbf{0}\}$, then let $\mathbf{u}$ be such that $T\mathbf{u} = \mathbf{u}'$; there is such a $\mathbf{u}$ since $\mathbf{u}' \in U' \subset TV$. Let $k = k' + 1$, and let U be the subspace of V spanned by

$$S_U = (T^i\mathbf{u} : 0 \le i \le k) = (\mathbf{u}, S_{U'}). \tag{4}$$

Since $\mathbf{u} \notin U'$ and $S_{U'}$ is a basis of U' (cf. (3), the system S_U is linearly independent; hence S_U is a basis of U. It is also clear that for any $\mathbf{u}'_0 \in U'$ there is a $\mathbf{u}_0 \in U$ such that

$$T\mathbf{u}_0 = \mathbf{u}'_0. \tag{5}$$

Let $l \ge 0$ be an integer and let $\mathbf{w}_j$ for $1 \le j \le l$ be vectors with such that $T\mathbf{w}_j \in W'$, the span of the system

$$S_W = (\mathbf{w}_j : 1 \le j \le l) \tag{6}$$

includes W', and, further,

$$S_V = (T^i\mathbf{u}, \mathbf{w}_j : 0 \le i \le k \text{ and } 0 \le j \le l) \tag{7}$$

is a linearly independent system of vectors in V, and, finally, such that the system S_V is maximal in the sense that no new vector $\mathbf{w} \in V$ with $T\mathbf{w} \in W'$ can be added while preserving linear independence. (The case $l = 0$ is allowed; in this case no vectors $\mathbf{w}_j$ can be found because the system without them is already maximal.) We claim that there is an S_V satisfying these conditions, and that such an S_V is a basis of V. To show this, assume that $\mathbf{v} \in V$ is not in the span of S_V. If $\mathbf{v} \in W'$ then we also have $T\mathbf{v} \in W'$ since the subspace W' is invariant and $W' \cap U = \{\mathbf{0}\}$. Thus, first adding a basis of W' will ensure that the span of S_W included W'. In any case, since we have $TV = W' \oplus U'$, and $T\mathbf{v} \in TV$, we have

$$T\mathbf{v} = \mathbf{w}' + \mathbf{u}'_0 \tag{8}$$

for some $\mathbf{w}' \in W'$ and $\mathbf{u}_0' \in U'$. According to (5) there is a $\mathbf{u}_0 \in U$ such that $T\mathbf{u}_0 = \mathbf{u}_0'$. Let $\mathbf{w} = \mathbf{v} - \mathbf{u}_0$; then $T\mathbf{w} = T\mathbf{v} - T\mathbf{u}_0 = \mathbf{w}' \in W'$. Given that $\mathbf{u}_0 \in U$ is in the span of S_V and $\mathbf{v}$ is not, $\mathbf{w}$ is not in the span of S_V. Thus $\mathbf{w}$ can be added to S_V, contradicting the maximality of the latter.

Let W be the subspace spanned by the system system by S_W just defined. As $T\mathbf{w}_j \in W'$, the subspace is invariant for T. S_W is linearly independent, it is a basis of W.

To complete the proof of the lemma, we need to show that $U \oplus W = V$; the rest of the assertions of the lemma are obviously satisfied. For this we need to show that $U \cup W$ span V and $U \cap W = \{\mathbf{0}\}$. To show the former, it is enough to note that the basis of S_V of V can be obtained by merging the bases S_U of U and S_W of W – cf (4), (7), and (6).

We need yet to show that $U \cap W = \{\mathbf{0}\}$. To this end, assume that $\mathbf{x} \in U \cap W$; we then need to show that $\mathbf{x} = \mathbf{0}$. We have

$$\mathbf{x} = \sum_{i=0}^{k-1} \alpha_i T^i \mathbf{u} = \sum_{j=1}^{l} \beta_j \mathbf{w}_j$$

for some scalars α_i and β_j – cf. (4) and (6). Therefore,

$$\sum_{i=0}^{k-1} \alpha_i T^i \mathbf{u} - \sum_{j=1}^{l} \beta_j \mathbf{w}_j = \mathbf{0}.$$

Since the system S_V in (7) is linearly independent, it follows that $\alpha_i = \beta_j = 0$ for $0 \le i < k$ and $1 \le j \le l$. Thus $\mathbf{x} = \mathbf{0}$, as we wanted to show. □

Corollary 1. *Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation such that λ is an eigenvalue of T. Then there is a vector $\mathbf{u} \in V$, a positive integer k, and subspaces U and W of V such that $U \oplus W = V$, U and W are invariant subspaces for T, the system*

$$\big((T - \lambda)^i \mathbf{u} : 0 \le i < k\big)$$

is a basis of U, and $(T - \lambda)^k \mathbf{u} = \mathbf{0}$.

Proof. The result follows immediately follows by using Lemma 1 with $T - \lambda$ replacing T if one observes that a subspace of V is invariant for T if and only if it is invariant for $T - \lambda$. □

It will be helpful to set down the type of subspace U described by the above corollary:

Definition 1. Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation and let λ be a scalar. A subspace U of V is called a *Jordan subspace* of T for (or associated with) the eigenvalue λ if there is a subspace W of V such that $U \oplus W = V$, U and W are invariant for T, and there is a vector $\mathbf{u} \in U$, a positive integer k such that the system

$$\big((T - \lambda)^i \mathbf{u} : 0 \le i < k\big)$$

is a basis of U, and $(T - \lambda)^k \mathbf{u} = \mathbf{0}$. The vector $\mathbf{u}$ is called the *cyclic generator* of the subspace U.

Observe that calling λ and eigenvalue is justified, since λ is indeed an eigenvalue with $(T-\lambda)^{k-1}\mathbf{u}$ being the corresponding eigenvector. We are now ready to state

Theorem 3. *Let V be a finite-dimensional vector space over an algebraically closed field, and let $T : V \to V$ be a linear transformation. There are Jordan subspaces J_1, J_2, …, J_n of T such that*

$$V = \bigoplus_{i=1}^{n} J_n.$$

The decomposition of V described in this theorem is called a *Jordan decomposition.*

Proof. The result follows by repeated applications of Corollary 1. Noting that every linear transformation of a vector field over an algebraically closed field into itself has an eigenvalue according to Theorem 2, using an eigenvalue of T we can obtain invariant subspaces J_1 and W_1 such that $V = J_1 \oplus W_1$ and J_1 is a Jordan subspace of T. Next, restricting T to T_1, we can obtain a Jordan subspace $J_2 \subset W_1$ of T restricted to W_1. We can continue this until the whole space V is used up. □

3 Uniqueness of the Jordan decomposition

The Jordan decomposition of a vector space V for a linear transformation is unique except for the order in which the Jordan subspaces are listed. The eigenvalue associated with each Jordan subspace is also uniquely determined. We will show this through a number of lemmas. In the lemmas, the vector space V and the linear transformation $T : V \to V$ may not be explicitly mentioned. By an invariant subspace we will mean a subspace invariant for T.

Lemma 2. *A Jordan subspace cannot be split as a nontrivial direct sum of two invariant subspaces.*

Proof. Without loss of generality, we may assume that the Jordan subspace is the whole space V, and the eigenvalue associated with this subspace is 0; this latter because if the eigenvalue is λ, we may replace the linear transformation T with $T - \lambda$. Let u and k be such that $(T^i\mathbf{u} : 0 \leq i < k)$ is a basis of V, and $T^k\mathbf{u} = \mathbf{0}$. Assume, on the contrary, that $V = A \oplus B$, $A \neq \{\mathbf{0}\}$, and $B \neq \{\mathbf{0}\}$.

First, note that each vector $\mathbf{v} \in V$ can be represented as $\mathbf{v} = P(T)\mathbf{u}$ for some polynomial P over the scalar field F of degree less than k. As V is the direct sum of A and B, we have $\mathbf{u} = (P(T) + Q(T))\mathbf{u}$ such that $P(T)\mathbf{u} \in A$ and $Q(T)\mathbf{u} \in B$. One of the polynomials P or Q must have a nonzero constant term; without loss of generality, we may assume that the constant term of P is $\alpha_0 \neq 0$. It is then easy to see that the vector $\mathbf{u}$ can be expressed as a linear combination of the vectors $T^iP(T)\mathbf{u}$ for i with $0 \leq i < k$. In fact, writing $P(x) = \sum_{j=0}^{k-1} \alpha_j x^j$, we have

$$T^iP(T)\mathbf{u} = \sum_{j=i}^{k-1} \alpha_{j-i}T^j\mathbf{u};$$

this is because $T^i\mathbf{u} = \mathbf{0}$ for $i \geq k$. That is a triangular system of equations that we can easily solve for $T^i\mathbf{u}$ by back-substitution. That is, the equation for $i = k-1$ gives $T^{k-1}P(T)\mathbf{u} = \alpha_0 T^{k-1}\mathbf{u}$, easily solved for $T^{k-1}\mathbf{u}$. For $i = k-2$ we have $T^{k-2}P(T)\mathbf{u} = \alpha_0 T^{k-2}\mathbf{u} + \alpha_1 T^{k-1}\mathbf{u}$; substituting the just obtained expression for $T^{k-1}\mathbf{u}$, we can now solve this for $T^{k-2}\mathbf{u}$, and so on.

As A is an invariant subspace for T, it follows that $\mathbf{u} \in A$. Hence, each of the basis vectors $T^i\mathbf{u}$ of V $(0 \leq i < k)$ belongs to A, and so $A = V$. This shows that $V = A \oplus B$ is a trivial direct sum. □

Corollary 2. *If U_1 and U_2 are two Jordan subspaces, and $U_1 \cap U_2 \neq \{\mathbf{0}\}$ then $U_1 = U_2$.*

It is not assumed that the subspaces U_1 and U_2 are associated with the same eigenvalue.

Proof. Let W_1 and W_2 be such that $V = U_1 \oplus W_1$ and $V = U_2 \oplus W_2$. Then the subspaces $U_1 \cap U_2$ and $U_1 \cap W_2$ are invariant, and so the direct sum decomposition $U_1 = (U_1 \cap U_2) \oplus (U_1 \cap W_2)$ must be trivial according to Lemma 2. That is, $U_1 \subset U_2$ or $U_1 \subset W_2$. We can exclude the latter, since it implies that $U_1 \cap U_2 = \{\mathbf{0}\}$; thus $U_1 \subset U_2$ follows.

We can establish $U_2 \subset U_1$ in a similar way; thus we can conclude that $U_1 = U_2$. □

Lemma 3. *A Jordan subspace can be associated with only one eigenvalue.*

Proof. Without loss of generality, we may assume that the Jordan subspace is V itself and one of the eigenvalues is $\mathbf{0}$; let the other eigenvalue be $\lambda \neq 0$. Assume that the bases associated with these eigenvalues are $(T^i\mathbf{u} : 0 \leq i < k)$ and $\big((T-\lambda)^i\mathbf{v} : 0 \leq i < k\big)$; note that k is the dimension of the space V, and so we used the same k for both bases, though this will not play an important role in our considerations. The vector $\mathbf{w} = (T-\lambda)^{k-1}\mathbf{v}$ is an eigenvector with eigenvalue λ, that is $T\mathbf{w} = \lambda\mathbf{w}$; hence

$$T^k\mathbf{w} = \lambda^k\mathbf{w} \neq \mathbf{0}.$$

On the other hand, $\mathbf{w} = P(T)\mathbf{u}$ for some polynomial P of degree less than k. As $T^k\mathbf{u} = \mathbf{0}$, is follows that

$$T^k\mathbf{w} = T^kP(T)\mathbf{u} = \mathbf{0}.$$

This is a contradiction, completing the proof. □

Theorem 4. *The Jordan decomposition of a vector space is essentially unique; that is, two Jordan decompositions must agree except for the order in which the Jordan subspaces are listed.*

The basis $(T^k\mathbf{u} : 0 \leq i < k)$ associated with a Jordan subspace is certainly unique, since instead of $\mathbf{u}$ we can take any other vector $P(T)\mathbf{u}$ for a polynomial P with nonzero constant term. For each J_i of the Jordan subspaces J_1, J_2, …, J_n the pair (λ_i, k_i) is associated in a unique way, and the sequence $\big((\lambda_i, k_i) : 1 \leq k \leq n\big)$ is called the *Jordan signature* of T.

4 The minimal polynomial of a linear transformation

In this section we will consider finite dimensional vector spaces V over a field F that does not need to be algebraically closed unless otherwise mentioned. By a *monic* polynomial over F we will mean a polynomial of leading coefficient 1 (the unit element of the field F).

Definition 2. Given a vector field V over a field F, the minimal polynomial of a linear transformation $T : V \to V$ the monic polynomial P over F of the smallest degree such that $P(T) = 0$.[2]

4.1 Existence of the minimal polynomial

We have

Lemma 4. *Let V be a finite dimensional vector space over a field F, and let $T : V \to V$ be a linear tranformation. Then T has a minimal polynomial. Furthermore, the minimal polynomial of T is unique.*

Proof. Let $S = (\mathbf{v}_i : 1 \leq i \leq n)$ be a basis of V; then n is the dimension of V. Let i be an integer with $1 \leq i \leq n$. Then the system $(T^j\mathbf{v}_i : 0 \leq j \leq n)$ consists of $n+1$ vectors; thus it is liearly dependent. Therefore, we have

$$\sum_{j=0}^{n} \alpha_{ij}T^j\mathbf{v}_i = \mathbf{0}$$

[2] Observe that the linear transformation 0 is different from the scalar 0. While we use the notation $\mathbf{0}$ for the zero vector to distinguish it from the scalar 0, it would be difficult to maintain such a distinction between all the different kinds of zero elements that may occur in the discussion.

for some scalars α_{ij} $(0 \le j \le n)$, not all of which are zero. Writing

$$Q_i(x) = \sum_{j=0}^{n} \alpha_{ij} \mathbf{x}^j = \mathbf{0},$$

$Q_i(x)$ is a nonzero polynomial such that $Q_i(T)\mathbf{v}_i = \mathbf{0}$.

Let $Q(x) = \prod_{i=1}^{n} Q_i(x)$. Then $Q(x)\mathbf{v}_i = \mathbf{0}$ for all i with $1 \le i \le n$ Since every vector V can be expressed as a linear combination of the vectors $\mathbf{v}_i$, we have $Q(T)\mathbf{v} = \mathbf{0}$ for every $\mathbf{v} \in V$. Thus, there is a polynomial $Q(x)$ that annihilates every vector in V. Then a monic polynomial $P(x)$ of the smallest degree that annihilates every vector in V is a minimal polynomial of T.

To show the uniqueness of the minimal polynomial, assume, on the contrary, that $P_1(x)$ and $P_2(x)$ are two different monic polynomials that are minimal polynomials of T; then $P_1(x)$ and $P_2(x)$ have the same degree. Let $D(x)$ be the greatest common divisor of $P_1(x)$ and $P_2(x)$; then the degree of $D(x)$ is lower than the common degree of $P_1(x)$ and $P_2(x)$. Furthermore, we have

$$P(x) = R_1(x)P_1(x) + R_2(x)P_2(x)$$

for some polynomials $R_1(x)$ and $R_2(x)$, according to the Euclidean algorithm. Thus

$$P(T)\mathbf{v} = R_1(T)P_1(T)\mathbf{v} + R_2(T)P_2(T)\mathbf{v} = \mathbf{0}$$

for every vector $\mathbf{v} \in V$. This is a contradiction, since the degree of $D(x)$ is lower than that of $P_1(x)$ and $P_2(x)$. □

4.2 The minimal polynomial for algebraically closed fields

Assume V is a vector space over an algebraically closed field F, and let $T : V \to V$ be a linear transformation Then T has a Jordan decomposition according to Theorem 3. The minimal polynomial T can be described with the help of this Jordan decomposition. Indeed, let $\{J_{ij} : 1 \le i \le m$ and $1 \le j \le m_i\}$ be the Jordan subspaces of T with the corresponding Jordan signature

$$\big((\lambda_i, k_{ij}) : 1 \le i \le m \text{ and } 1 \le j \le m_i\big); \tag{9}$$

that is, the Jordan subspace J_{ij} is associated with the eigenvalue λ_i and has dimension k_{ij}. Assume that the subspaces are so arranged that $k_{i1} \ge k_{ij}$ for $1 < j \le m_i$. Then the minimal polynomial of T is

$$P(x) = P_{\min,T}(x) \overset{def}{=} \prod_{i=1}^{m} (x - \lambda_i)^{k_{i1}}. \tag{10}$$

To see that $P(T) = 0$, we need to show that $P(T)\mathbf{v} = \mathbf{0}$ for every vector $\mathbf{v}$ in a basis of V. So consider the Jordan basis associated with the above Jordan decomposition

$$\big((T - \lambda_i)^l \mathbf{u}_{ij} : 1 \le i \le m,\ 1 \le j \le m_i \text{ and } 0 \le l < k_{ij}\big).$$

Let i, j, and l be an integers with $1 \le i \le m$ and $0 \le j \le m_j$. Then

$$(T - \lambda_i)^{k_{ij}} (T - \lambda_i)^l \mathbf{u}_{ij} = (T - \lambda_i)^l (T - \lambda_i)^{k_{ij}} \mathbf{u}_{ij} = 0,$$

where the latter equation holds since $(T-\lambda_i)^{k_{ij}}\mathbf{u}_{ij}=0$; this shows that the polynomial $P(T)$ annihilates each of the basis vectors $(T-\lambda_i)^l\mathbf{u}_{ij}$. On the other hand, if $n\ge 0$ is any integer $\eta\in F$ with $\eta\ne\lambda_i$, then

$$\begin{aligned}(T-\eta)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}&=\big((T-\lambda_i)-(\eta-\lambda_i)\big)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}\\&=(T-\lambda_i)^{k_{ij}}\mathbf{u}_{ij}+(\lambda_i-\eta)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}=(\lambda_i-\eta)(T-\lambda_i)^{k-1}\mathbf{u}_{ij}\ne\mathbf{0};\end{aligned}$$

the second equation holds since $(T-\lambda_i)^{k_{ij}}=\mathbf{0}$. So, if

$$Q(x)=\prod_{r=1}^{n}(x-\eta_r)$$

is an arbitrary polynomial over F, then

$$\begin{aligned}Q(T)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}&=\Big(\prod_{r=1}^{n}(T-\eta_r)\Big)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}\\&=\Big(\prod_{r=1}^{n}(\lambda_i-\eta_r)\Big)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}=Q(\lambda_i)(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}.\end{aligned}$$

Since we have $(T-\lambda_i)^{k_{ij}-1}\mathbf{u}_{ij}\ne\mathbf{0}$, the right-hand side is not zero unless $Q(\lambda_i)=0$. This shows that a polynomial $R(T)$ does not annihilate the basis vector $\mathbf{u}_{ij}$ unless $R(t)$ is divisible by $(x-\lambda_i)^{k_{ij}}$. This shows that $P(x)$ given in (10) is indeed the minimal polynomial of T.

4.3 The characteristic polynomial and the Cayley–Hamilton theorem

Consider a linear transformation $T:V\to V$ over an algebraically closed vector field F. with the Jordan signature given in T, and let $k_i=\sum_{j=1}^{m_i}$ for $1\le i\le m$. The integer k_i is called the *multiplicity* of the eigenvalue λ. The characteristic polynomial of T is defined as

$$P_{\mathrm{char},T}(x)\overset{def}{=}\prod_{i=1}^{m}(\lambda_i-x)^{k_i}. \tag{11}$$

Observe that the minimal polynomial of T defined in (10) is a divisor of the characteristic polynomial; hence, $P_{\mathrm{char},T}(T)=0$. This result is the *Cayley–Hamilton theorem.*

Below we will give a description of the characteristic polynomial that does not rely on the Jordan decomposition of V. Such a description is important, since if the scalar field F is not algebraically closed, the Jordan decomposition of V for T may not even exist.

4.4 Finding the minimal polynomial

Let V be a finite dimensional vector space over a field F, and let $T:V\to V$ be a linear transformation. There is a simple algorithm that allows us to find the minimal polynomial of T. Finding the minimal polynomial is much simpler than finding the Jordan decomposition.

We need to find a polynomial P such that $P(T)\mathbf{v}=\mathbf{0}$ for all vectors in a basis of V. Starting with an arbitrary nonzero vector $\mathbf{v}_1$, we first find a polynomial P_1 such that $P_1(T)\mathbf{v}_1=\mathbf{0}$ as follows. Form the vectors $\mathbf{v}_1$, $T\mathbf{v}_1$, $T^2\mathbf{v}_1$, $\dots$ until we arrive at a system $(T^i\mathbf{v}_1:0\le i\le k_1)$ that is linearly dependent; i.e.,

$$\sum_{i=0}^{k_1}\alpha_{1i}T^i\mathbf{v}_1=\mathbf{0};$$

such that not all coefficients α_i are zero. In fact, we cannot have $\alpha_{1k_1} = 0$ since $(T^i\mathbf{v}_1 : 0 \le i \le k_1 - 1)$ is linearly independent according to our construction. It is easy to see that

$$P_1(x) = \sum_{i=0}^{k_1} \alpha_i x^i$$

is the polynomial of the lowest degree $P(x)$ for which $P(T)\mathbf{v}_1 = 0$. Observe that we then also have $P_1(T)T^i\mathbf{v}_1 = T^iP_1(T)\mathbf{v}_1 = \mathbf{0}$. If the system $(T^i\mathbf{v}_1 : 0 \le i < k_1)$ span V, then P_1 is the minimal polynomial of T. If they do not span V, then take a vector $\mathbf{v}_2$ that is not in the span of $(T^i\mathbf{v}_1 : 0 \le i < k_1)$, and form the vectors $P_1(T)\mathbf{v}_2$, $TP_1(T)\mathbf{v}_2$, $T^2P_1(T)\mathbf{v}_2$, $\dots$ until we arrive at a system $(T^i\mathbf{P}_1(T)v_2 : 0 \le i \le k_2)$ is linearly dependent; observe that we will have $k_2 = 0$ in case $P_1(T)\mathbf{v}_2 = \mathbf{0}$, whereas we cannot have $k_1 = 0$. The linear dependence of this system means

$$\sum_{i=0}^{k_2} \alpha_{2i}T^iP_1(T)\mathbf{v}_2 = \mathbf{0};$$

Write

$$P_2(x) = \sum_{i=0}^{k_2} \alpha_{2i}x^iP_1(x).$$

It is easy to show that $P_2(x)$ is the lowest degree polynomial $P(x)$ divisible by $P_1(x)$ for which $P(T)\mathbf{v}_2 = \mathbf{0}$. Observe again that we have $P_2(T)T^i\mathbf{v}_2 = T^iP_2(T)\mathbf{v}_2 = \mathbf{0}$. The vectors $P_2(T)T^i\mathbf{v}_2$ have, however, not been calculated, so we only make use of the consequence of this that

$$P_2(T)T^iP_1(T)\mathbf{v}_2 = 0$$

for all i with $0 \le i \le k_2 - 1$. If the system system

$$(T^{i_1}\mathbf{v}_1, \mathbf{v}_2, T^{i_2}P_1(T)\mathbf{v}_2 : 0 \le i_1 < k_1 \text{ and } 0 \le i_2 < k_2)$$

span V, then P_2 is the minimal polynomial of T. If they do not span V, then take a vector $\mathbf{v}_3$ that is not in the of this system, and continue the procedure.

In general, assume that the vectors that for some $m > 1$ $\mathbf{v}_r$, polynomials P_r, and the integers $k_r \ge 0$ have been constructed for $1 \le r < m$. If the system

$$(T^{i_1}\mathbf{v}_1, \mathbf{v}_r, T^{i_r}P_{r-1}(T)\mathbf{v}_r : 0 \le i_1 < k_1, 2 \le r \le r < m \text{ and } 0 \le i_r < k_r) \tag{12}$$

span V, then then P_{m-1} is the minimal polynomial of T. If not, then take a vector $\mathbf{v}_m$ that is not in the span of this system, and form the vectors and form the vectors $P_{m-1}(T)\mathbf{v}_m$, $TP_{m-1}(T)\mathbf{v}_m$, $T^2P_{m-1}(T)\mathbf{v}_m$, $\dots$ until we arrive at a system $(T^iP_{m-1}(T)\mathbf{v}_m : 0 \le i \le k_m)$ is linearly dependent; observe that we will have $k_m = 0$ in case $P_{m-1}(T)\mathbf{v}_m = \mathbf{0}$. The linear dependence of this system means

$$\sum_{i=0}^{k_m} \alpha_{mi}T^iP_{m-1}(T)\mathbf{v}_m = \mathbf{0};$$

Write

$$P_m(x) = \sum_{i=0}^{k_m} \alpha_{mi}x^iP_{m-1}(x).$$

It is easy to show that $P_m(x)$ is the lowest degree polynomial $P(x)$ divisible by $P_1(x)$ for which $P(T)\mathbf{v}_m = \mathbf{0}$. Observe again that we have $P_m(T)T^i\mathbf{v}_m = T^iP_m(T)\mathbf{v}_m = \mathbf{0}$. The vectors $P_m(T)T^i\mathbf{v}_m$ have, however, not been calculated, so, as before we only make use of the consequence of this that $P_m(T)T^iP_{m-1}(T)\mathbf{v}_2 = 0$ for all i with $0 \le i \le k_m - 1$.

Continuing in this manner, we will arrive at a point when the system given in (12) spans V, since V is finite dimensional. At that point, P_{m-1} is the minimal polynomial of T.

We will give a numerical example for finding the minimal polynomial in in Section 6 below, where these results will be discussed in terms of matrices.

5 Consequences for matrices

5.1 Representation of vector spaces and linear transformations

Finite dimensional vector spaces over a field F and linear transformations between them can be represented by column vectors (matrices consisting of a single column) and matrices over F. We recall the basic definitions. The set of $m \times n$ matrices over F will be denoted by $F_{m,n}$; here m, n are nonnegative integers. The cases $m = 0$ or $n = 0$ usually have no uses, but they are occasionally helpful in proofs to support induction. Row vectors are $1 \times n$ matrices and column vectors are $m \times 1$ matrices. The transpose of a matrix A will be denoted by A^T. Given a vector space V over F with a basis $\mathcal{X} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ we say that the column vector $\mathbf{c} = (c_1, c_2, \dots, c_n)^T$ represents the vector if $\mathbf{v} = \sum_{i=1}^n c_i\mathbf{v}_i$; it will be convenient to extend the usual matrix multiplication rules and use the abbreviated notation $\mathbf{v} = \mathcal{X}\mathbf{c}$, as if $\mathcal{X}$ were a row vector, even though it is not (since it is not a matrix over F). In this case, we say that $\mathbf{c} = \mathcal{R}_{\mathcal{X}}\mathbf{v}$ – see [1, p. 133–138]. If $\mathcal{Y}$ is another basis of V then $\mathcal{X} = \mathcal{Y}P$ for a nonsingular $n \times n$ matrix over F, and $\mathbf{v} = \mathcal{X}\mathbf{c} = (\mathcal{Y}P)\mathbf{c} = \mathcal{Y}(P\mathbf{c})$, and so $\mathcal{R}_{\mathcal{Y}}\mathbf{v} = P\mathcal{R}_{\mathcal{X}}\mathbf{v}$ – see [1, (3.5.5) Theorem, p. 137].

If X and Y are vector spaces of F with dimensions with dimensions n and m and bases $\mathcal{X}$ and $\mathcal{Y}$ and $T : X \to Y$ is a linear transformation, then there is a matrix $A \in F_{m,n}$ such that for every column vector $\mathbf{c} = F_{n,1}$ we have $T(\mathcal{X}\mathbf{c}) = \mathcal{Y}(Ac)$ (the parentheses are for emphasis only; the formal matrix multiplication rules being associative, the parentheses can be dropped). The shortened version of this equation, $T\mathcal{X} = \mathcal{Y}A$ is also used. We call the matrix A the representation of T with respect to $\mathcal{X}$ and $\mathcal{Y}$, and we write $A = \mathcal{R}_{\mathcal{Y}\mathcal{X}}T$. If $\mathcal{V}$ is another basis of X then $\mathcal{V} = \mathcal{X}P$ for an $n \times n$ nonsingular matrix P, and if $\mathcal{W}$ is another basis of Y then $\mathcal{W} = \mathcal{Y}Q$ for an $m \times m$ nonsingular matrix. We have $T\mathcal{V}P^{-1} = T\mathcal{X} = \mathcal{Y}A = \mathcal{W}Q^{-1}A$; omitting the middle members and multiplying the sides by P on the right, we obtain $T\mathcal{V} = \mathcal{W}Q^{-1}AP$, i.e., $Q^{-1}AP = \mathcal{R}_{\mathcal{W}\mathcal{V}}T$. That is,

$$\mathcal{R}_{\mathcal{W}\mathcal{V}}T = Q^{-1}(\mathcal{R}_{\mathcal{Y}\mathcal{X}}T)P \tag{13}$$

(see [1, (5.3.1) Theorem, p. 232].

Matrix multiplication and the composition of linear transformations are closely related. Let U, V, W be vector spaces with bases $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$. If $S : U \to V$ and $T : V \to W$ are linear transformations, then the composition $T \circ S : U \to W$ is usually written as TS and is referred to as multiplication of the linear transformations. If $A = \mathcal{R}_{\mathcal{W}\mathcal{V}}T$ and $B = \mathcal{R}_{\mathcal{V}\mathcal{U}}S$ then $T\mathcal{V} = \mathcal{W}A$ and $S\mathcal{U} = \mathcal{V}B$. Hence $TS\mathcal{U} = T\mathcal{V}B = \mathcal{W}AB$. Hence $AB = \mathcal{R}_{\mathcal{W}\mathcal{U}}(TS)$, and so

$$\mathcal{R}_{\mathcal{W}\mathcal{U}}TS = (\mathcal{R}_{\mathcal{W}\mathcal{V}}T)(\mathcal{R}_{\mathcal{V}\mathcal{U}}S), \tag{14}$$

where we deliberately dropped the parentheses around TS for easy readability, since no other placement of the parentheses would be meaningful – see [1, (5.2.5) Theorem, p. 223].

5.2 Similarity transformations

Given an arbitrary $n \times n$ matrix A over the field F, $T : F_{n,1} \to F_{n,1}$ be the linear transformation defined by $T\mathbf{x} = A\mathbf{x}$ for $\mathbf{x} \in F_{n,1}$, let $\mathbf{e}_k$ be the kth *unit column vector*, that is, $\mathbf{e}_k = (\delta_{ik} : 1 \le i \le n)^T$, and let $\mathcal{E} = (\mathbf{e}_k : 1 \le k \le n)$. Then $\mathcal{E}$ is a basis of $F_{n,1}$; it is called the *canonical basis* of $F_{n,1}$. We have

$$A = \mathcal{R}_{\mathcal{E}\mathcal{E}}T. \tag{15}$$

In such a situation, it is convenient, though perhaps not literally correct, to consider the matrix A and the linear transformation T to be the same object.

Let P be a nonsingular matrix; then $\mathcal{X} = \mathcal{E}P$ is a basis of $F_{n,1}$. According to (13), for the above A and T we have

$$\mathcal{R}_{\mathcal{X}\mathcal{X}}T = P^{-1}(\mathcal{R}_{\mathcal{E}\mathcal{E}}T)P = P^{-1}AP. \tag{16}$$

The transformation $A \mapsto P^{-1}AP$ is called a *similarity transformation.* According to the last displayed equation, a similarity transformation amounts to a change of basis in the space $F_{n,1}$. If $B = P^{-1}AP$ for some nonsingular matrix P, we say that the matrices A and B are *similar.*

5.3 Direct sums of matrices

If A is an $m \times m$ matrix and B is an $n \times n$ matrix, $0_{m\times n}$ is the $m \times n$ zero matrix (a matrix with all its entries 0), and $0_{n\times m}$ is the $n \times m$ zero matrix, then the matrix

$$\begin{pmatrix} A & 0_{m\times n} \\ 0_{n\times m} & B \end{pmatrix}$$

is called the direct sum of the matrices A and B and is denoted as $A \oplus B$. Let V be a vector space, $T : V \to V$ a linear transformation, X and Y invariant subspaces for T such that $V = X \oplus Y$, $\mathcal{X}$ a basis of X, and $\mathcal{Y}$ a basis of Y. Let T_X be the restriction of T to X, and T_Y be its restriction to Y. Finally, let $A = \mathcal{R}_{\mathcal{X}\mathcal{X}}T_X$ and $B = \mathcal{R}_{\mathcal{Y}\mathcal{Y}}T_Y$. Then it is easy to see that

$$\mathcal{R}_{(\mathcal{X},\mathcal{Y})(\mathcal{X},\mathcal{Y})}T = A \oplus B. \tag{17}$$

5.4 Jordan block matrices

An $n \times n$ matrix $A = (a_{ij})$ is called an *auxiliary unit matrix* if $a_{ij} = \delta_{i\,j-1}$. A *Jordan block* is a matrix $\lambda I + A$, where λ is a scalar, I is the $n \times n$ identity matrix, and A is an auxiliary unit matrix. That is, a Jordan block is a matrix of form

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \lambda & \dots & 0 & 0 \\ \multicolumn{7}{c}{\dots\dots\dots\dots\dots\dots\dots\dots} \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

Theorem 5. *Let n be a positive integer, let V be an n-dimensional vector space, $T : V \to V$ a linear transformation, and assume that V is a Jordan subspace of itself for T with eigenvalue λ and Jordan generator $\mathbf{u}$. Let $\mathbf{x}_k = (T - \lambda)^{n-k}\mathbf{u}$ for $1 \le k \le n$, and $\mathcal{X} = (\mathbf{x}_k : 1 \le k \le n)$. Then the matrix $\mathcal{R}_{\mathcal{X}\mathcal{X}}T$ is a Jordan block.*

The proof amounts to a routine calculation of the representation $\mathcal{R}_{\mathcal{X}\mathcal{X}}T$. A matrix said to be in *Jordan canonical form* if it is a direct sum of Jordan block matrices.

5.5 The Jordan canonical form of a matrix

Let A be an arbitrary $n \times n$ matrix, and consider A as a linear transformation of $F_{n,1}$ into $F_{n,1}$ as in Subsection 5.2. According to Theorem 3 $F_{n,1}$ slits up into a direct sum of Jordan subspaces J_i of T for i with $1 \le i \le m$ for some $m \ge 0$. Choosing an appropriate basis $\mathcal{X}_i$ on J_i, the linear transformation A restricted to J_i can be represented by a Jordan block matrix. Putting $\mathcal{X} = (\mathcal{X}_i : 1 \le i \le m)$, the linear transformation A is represented in the basis $\mathcal{X}$ as a direct sum of these Jordan block matrices, i.e., as a matrix in Jordan canonical form. Since the representation of A in the basis $\mathcal{X}$ is similar to the matrix A according to (16), we proved the following

Theorem 6. *Every square matrix A is similar to a matrix J in Jordan canonical form. Each eigenvalue of multiplicity k occurs as a diagonal element of J exactly k times, and each diagonal element of J is an eigenvalue of A.*

The sentence about the eigenvalues of A as diagonal elements if J is clear from the structure of Jordan subspaces (see Subsection 4.3 for the definition of the multiplicity of an eigenvalue).

5.6 The characteristic polynomial as a determinant

Let A be an $n \times n$ matrix, and let P be a nonsingular $n \times n$ matrix such that $P^{-1}AP$ is in Jordan canonical form. The diagonal elements of $P^{-1}AP$ are the eigenvalues of A occurring with their multiplicities; thus, the diagonal elements of $P^{-1}AP - I\lambda$, where I is the $n \times n$ identity matrix, are $\lambda_i - \lambda$ for each eigenvalue of A (occurring a number of times according to its multiplicity). The product of these diagonal elements is the characteristic polynomial of the matrix A according to (10), where we replaced the variable x with λ, as is customary. Since $P^{-1}AP - I\lambda$ is an upper triangular matrix, its determinant is equal to the product of its diagonal elements. That is

$$\begin{aligned} P_{\mathrm{char},A}(\lambda) &= \det(P^{-1}AP - \lambda I) = \det\bigl(P^{-1}(A - \lambda I)P\bigr) \\ &= \det(P^{-1})\det(A - \lambda I)\det(P) = \det(P^{-1})\det(P)\det(A - \lambda I) \\ &= \det(P^{-1}P)\det(A - \lambda I) = \det(I)\det(A - \lambda I) = \det(A - \lambda I). \end{aligned}$$

That is the characteristic polynomial of A is equal to the determinant $\det(A - \lambda I)$.

6 An example for finding the minimal polynomial of a matrix

Here we will use the method described in Subsection 4.4 to find the minimal polynomial of the matrix[3]

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 & -1 \\ -3 & 8 & 5 & 5 & 2 & -2 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 4 & -10 & -7 & -6 & -3 & 3 \\ -1 & 3 & 2 & 2 & 2 & -1 \\ -2 & 6 & 4 & 4 & 2 & -1 \end{pmatrix} \tag{18}$$

[3] The computer algebra system Maxima was used in creating this example.

over the field $\mathbb{R}$ of real numbers. Let $\mathbf{e}_1 = (1, 0, 0, 0, 0, 0)^T$, and form the vectors

$$\begin{array}{rcrrrrrrrr}
\mathbf{u}_0 & = & \mathbf{e}_1 = & (& 1, & 0, & 0, & 0, & 0, & 0)^T, \\
\mathbf{u}_1 & = & A\mathbf{u}_0 = & (& 0, & -3, & 1, & 4, & -1, & -2)^T, \\
\mathbf{u}_2 & = & A^2\mathbf{u}_0 = & (& 0, & 3, & 0, & -4, & 1, & 2)^T, \\
\mathbf{u}_3 & = & A^3\mathbf{u}_0 = & (& 0, & 2, & -1, & -3, & 1, & 2)^T, \\
\mathbf{u}_4 & = & A^4\mathbf{u}_0 = & (& -1, & -6, & 0, & 8, & -2, & -4)^T.
\end{array}$$

One can ascertain that the system $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ is linearly independent, while the system $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$ is linearly dependent;[4] in fact, we have $\mathbf{u}_4 + 2\mathbf{u}_2 + \mathbf{u}_0 = (A^4 + 2A_2 + I)\mathbf{u}_0 = \mathbf{0}$. That is we take $P_1(x) = x^4 + 2x^2 + 1$.

The vector $\mathbf{e}_2 = (0, 1, 0, 0, 0, 0)$ is not in the span of the system $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, so we form the vectors

$$\begin{array}{rcrrrrrrr}
 & & \mathbf{e}_2 = & (\; 0, & 1, & 0, & 0, & 0, & 0)^T, \\
\mathbf{v}_0 & = & P_1(A)\mathbf{e}_2 = & (-4, & 8, & -4, & -8, & 4, & 8)^T, \\
\mathbf{v}_1 & = & AP_1(A)\mathbf{e}_2 = & (-4, & 8, & -4, & -8, & 4, & 8)^T.
\end{array}$$

The system $(\mathbf{v}_0, \mathbf{v}_1)$ is linearly dependent; in fact, we have $\mathbf{v}_1 = \mathbf{v}2$. That is, $\mathbf{v}_2 - \mathbf{v}_1 = (A - I)P_1(A)\mathbf{e}_2 = \mathbf{0}$, so we put $P_2(x) = (x - 1)P_1(x) = (x - 1)(x^4 + 2x^2 + 1)$.

The system $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{e}_2)$, is linearly independent, and $\mathbf{v}_0$ is in the span of this system. We find that the vector $\mathbf{e}_5 = (0, 0, 0, 0, 1, 0)^T$ transpose is not in the span of this system, but $P_2(A)\mathbf{e}5 = \mathbf{0}$. Next, we find that the system $(\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{e}_2, \mathbf{e}5)$ spans the whole space of $\mathbb{R}_{6,1}$. As we have seen, $P_2(A)$ annihilates all vectors in this system. Hence $P(x) = P_2(x) = (x-1)(x^4+2x^2+1)$ is the minimal polynomial of the matrix A. Having determined the minimal polynomial of A, we will no longer need the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, and these letters will be re-used below to denote different vectors.

7 The example continued: finding the Jordan canonical form

The proof we gave for Theorem 3 was basically algorithmic, and the proof described there can be adapted to finding the Jordan subspaces of the matrix A. We found that the minimal polynomial of A is $P_{\min,A}(x) = (x - 1)(x^4 + 2x^2 + 1) = (x - 1)(x^2 + 1)^2$. The eigenvalues of A are the roots of this polynomial, i.e., they are 1, i, and $-i$, where $i = \sqrt{-1}$. While A is a real matrix, i.e., it is a matrix over the real numbers $\mathbb{R}$, we need to extend out scalar field to the field $\mathbb{C}$ of complex numbers, so that the eigenvalues be included in the scalar field.

7.1 The first two Jordan subspaces

First we deal with the eigenvalue 1. We will mimic the proof of Lemma 1 with $V = \mathbb{C}_{6,1}$ and $T = A - I$, where I is the identity matrix in $C_{6,6}$. The space V is spanned by the columns of I, and the space $TV = (A - I)V$ are spanned by the columns of the matrix $(A - I)I = A - I$. The dimension of the column space of I is the rank of I, i.e., it is 6. The dimension of the column space of $A - I$ is the rank of this matrix, which happens to be 4. The space $T(TV)$ is spanned by the columns of $(A - I)^2$; this dimension also happens to be 4, which means that $T(TV) = TV$. That is, 0 is not an eigenvalue of T restricted to TV, i.e., TV is the space where we start the induction. For

[4] To see this one may want to use variants of Gaussian elimination or find the row-echelon form of the matrices whose rows are the transposes of these vectors. We did not include these details so as to focus on our main goal of finding the minimal polynomial of A.

the decomposition $TV = U' \oplus W'$ of Lemma 1 in this case we have $U' = \{\mathbf{0}\}$ and $W' = TV$. That is, W' is the column space of the matrix

$$A - I = \begin{pmatrix} -1 & 1 & 0 & 0 & -1 & -1 \\ -3 & 7 & 5 & 5 & 2 & -2 \\ 1 & 0 & -2 & 0 & -1 & 0 \\ 4 & -10 & -7 & -7 & -3 & 3 \\ -1 & 3 & 2 & 2 & 1 & -1 \\ -2 & 6 & 4 & 4 & 2 & -2 \end{pmatrix}.$$

We find that this matrix has rank 4; in fact, its first four columns are linearly independent, and the rest of the columns linearly depend on them. Next, we need to find a vector $\mathbf{u} \in TV \setminus V$ for which $T\mathbf{u} = (A - I)\mathbf{0} = \mathbf{0}$.

To find such a $\mathbf{u}$, recall that $P_{\min,A}(A) = (A - I)(A^2 + I)^2 = (A - I)(A^2 + I)^2\ I$; hence the matrix $A - I$ annihilates all the columns of the matrix

$$(A^2 + I)^2 = \begin{pmatrix} 0 & -4 & 0 & -4 & -4 & 0 \\ 0 & 8 & 0 & 8 & 0 & 4 \\ 0 & -4 & 0 & -4 & -4 & 0 \\ 0 & -8 & 0 & -8 & 0 & -4 \\ 0 & 4 & 0 & 4 & 4 & 0 \\ 0 & 8 & 0 & 8 & 0 & 4 \end{pmatrix}$$

In fact, it is easy to see that the columns of this matrix span the space of vectors $\mathbf{v} \in V$ for which $(A - I)\mathbf{v} = \mathbf{0}$, but we do not need this fact for our present purposes.[5] The rank of this matrix is 2, so its column space is spanned by the vectors $\mathbf{u}_1 = (1, 0, 1, 0, -1, 0)^T$, the fifth column multiplied by $-1/4$, and $\mathbf{u}_2 = (0, 1, 0, -1, 0, 1)^T$, the sixth column divided by 4.

We will define the space U_1 as the space spanned by the by $\mathbf{u} = \mathbf{u}_1$. Note that we have $(A - I)\mathbf{u}_1 = 0$, so U_1 is an invariant subspace (for A, or, what is the same, for A). Further, u_1 is not in $TV = (A - I)V$, since 1 is not an eigenvalue of A restricted to $(A - I)V$.

Next we need to pick the vectors $\mathbf{w}_j$ to form the system (7). These vectors must be picked in such a way that $\mathbf{w}_j$ is not in the span of $(\mathbf{u}_1)$ and $W' = (A - I)V$ is in the span of S_W (see (6). The and the condition $T\mathbf{w}_j \in W'$ is automatically satisfied, since $W' = TV$. Thus we can pick the vectors $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, and $\mathbf{w}_4$ as the first four columns of the matrix $A - I$.

We need to put one more vector $\mathbf{w}_5$, and the most expedient choice $\mathbf{w}_5 = \mathbf{u}_2$, since as we will see, this will lead to a further reduction of the matrix A. We change to the new basis

$$\mathcal{X}_1 = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4),$$

where we listed $\mathbf{u}_2 = \mathbf{u}_5$ second. That is, $\mathcal{X}_1$ is formed by the columns of the matrix

$$S_1 = \begin{pmatrix} -1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -3 & 7 & 5 & 5 \\ -1 & 0 & 1 & 0 & -2 & 0 \\ 0 & -1 & 4 & -10 & -7 & -7 \\ 1 & 0 & -1 & 3 & 2 & 2 \\ 0 & 1 & -2 & 6 & 4 & 4 \end{pmatrix} \tag{19}$$

[5] That is, we can continue with the example of finding the Jordan Canonical Form of A; but to show that this procedure will work all the time, we would need to prove this fact.

Changing to this new basis, we have

$$A_1 = S_1^{-1} A S_1 = \mathcal{R}_{\mathcal{X}\mathcal{X}} T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 & 2 & 2 \\ 0 & 0 & -1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & 2 & -4 & -3 & -2 \end{pmatrix}. \tag{20}$$

This matrix A_1 can be written as a direct sum $A_1 = I_1 \oplus I_1 \oplus A_2$, where I_1 is the 1×1 identity matrix, and

$$A_2 = \mathcal{R}_{\mathcal{X}\mathcal{X}} T = \begin{pmatrix} -1 & 4 & 2 & 2 \\ -1 & 2 & 1 & 1 \\ 0 & 3 & 1 & 2 \\ 2 & -4 & -3 & -2 \end{pmatrix}.$$

This means that the spaces $\langle \mathbf{u}_1 \rangle$, $\langle \mathbf{u}_2 \rangle$, and $\langle \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4 \rangle$, are invariant subspaces of T spanning V. The first two of these are already Jordan subspaces.[6] Thus, we need to focus on the Jordan decomposition of the third subspace; this amounts to determining the Jordan canonical form of the matrix A_2. Occasionally, we will write T_1 for the transformation from $\mathbb{C}_{4,1}$ into $\mathbb{C}_{4,1}$ represented by A_2 with respect to the canonical basis of this space.

7.2 The next to Jordan subspaces

Since we removed two Jordan subspaces belonging to the eigenvalue 1 in forming the matrix A_2, 1 may no longer be an eigenvalue of A_2; indeed, we can ascertain of this by checking that $A_2 - I$ has rank 4, and so 1 is in fact no longer an eigenvalue of A_2. This removes the factor $x - 1$ from the minimal polynomial of A and so the minimal polynomial of A_2 is $P_{\min,A_2}(x) = (x^2+1)^2$. That is,

$$A_3 = A_2^2 + I = \begin{pmatrix} 2 & 2 & -2 & 2 \\ 1 & 0 & -2 & 0 \\ 1 & 1 & -1 & 1 \\ -2 & -1 & 3 & -1 \end{pmatrix}. \tag{21}$$

This matrix has rank 2, and we will use elementary row operations to simplify it. The kind of elementary row operations we need adds λ times row j to row i ($i \neq j$); this row operation can be performed by left-multiplication with the elementary matrix $E_{ij}(\lambda) = E_{\mathrm{III},ij}(\lambda)$, that is, a type III elementary matrix having 1s in the main diagonal, having λ as the entry at the intersection of row i and column j, with all other entries being zero. The inverse of this matrix is $E_{\mathrm{III},ij}(-\lambda)$. (See [1, pp. 47–57] for a discussion of elementary matrices.)

When choosing the elementary operations to be performed, we will try to make as many entries in the matrix 0 as possible. With

$$EA_3 = E = E_{23}(-1)E_{34}(1)E_{42}(2)E_{13}(-2) = \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 2 & 1 & 1 \\ 0 & 2 & 0 & 1 \end{pmatrix},$$

[6] The first of these we found by searching for it deliberately; the second one "accidentally," though it could have been anticipated on theoretical grounds that by making the choice of the eigenvector $\mathbf{u}_2$ as $\mathbf{w}_1$ we would obtain a second Jordan subspace.

we have

$$EA_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & -1 & -1 & -1 \end{pmatrix}.$$

Using elementary row operations simplified the matrix A_3 but it did not preserve its eigenvalues. We want to preserve these eigenvalues, and so we need a similarity transformation, which represents a change of basis in the vector underlying vector space (cf. (16). That is, instead of EA_3, our real interests are the matrices

$$B_2 = EA_2E^{-1} = \begin{pmatrix} -1 & 2 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 2 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

and

$$B_3 = EA_3E^{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Observe that we have $B_3 = B_2^2 + I$; indeed, $B_2^2 + I = (EA_2E^{-1})^2 + I = E(A_2^2 + I)E^{-1} = EA_3E^{-1} = B_3$. The minimal polynomial of A_2 is $(x^2 + 1)^2$; the minimal polynomial of B_2 is the same. Hence, we have $0 = (B_2^2 + I)^2 = (B_2 - iI)(B_2 + iI)B_3$. That is, the columns of the matrix

$$(B_2 + iI)B_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1+i & 2 & 0 & 0 \\ -1 & 1+i & 0 & 0 \end{pmatrix}$$

are eigenvectors of B_2 associated with eigenvalue i. Actually, the second column of this matrix is $-1 - i$ times its first column, so this matrix has (essentially, i.e., up to a scalar factor) only one eigenvector. Similarly, since $0 = (B_2^2 + I)^2 = (B_2 + iI)(B_2 - iI)B_3$, the columns of the matrix

$$(B_2 - iI)B_3 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1-i & 2 & 0 & 0 \\ -1 & 1-i & 0 & 0 \end{pmatrix}$$

are eigenvectors of B_2 associated with eigenvalue $-i$. Again, we only get essentially one eigenvector, since the second column of this matrix is $-1 + i$ times its first column.

Pick the second columns of these matrices, and write $\mathbf{z}'_3 = (0, 0, 2, 1+i)^T$ and $\mathbf{z}'_4 = (0, 0, 2, 1-i)^T$ for these eigenvectors of B_2 associated with the eigenvalues i and $-i$, respectively. We will proceed along the lines of the proof of Lemma 1. We will first deal with the eigenvector $\tilde{\mathbf{u}}'_3 = (0, 0, 2, 1+i)^T$.

As in the proof of Lemma 1, we need to find vector $\mathbf{z}_3$ such that $(B_2 - iI)\mathbf{z}_3 = \mathbf{z}'_3$; here $B_2 - iI$, $\mathbf{z}_3$, and $\mathbf{z}'_3$ are replacing T, $\mathbf{u}$, and $\mathbf{u}'$ in that proof. Writing $\mathbf{z}_3 = (\alpha, \beta, \gamma, \delta)^T$, the equation $(B_2 - iI)\mathbf{z}_3 = \mathbf{z}'_3$ can be written as

$$\begin{array}{rcl} (-1-i)\alpha + 2\beta & = & 0, \\ -\alpha - i\beta & = & 0, \\ -\beta - (1+i)\gamma + 2\delta & = & 2, \\ -\gamma + (1-i)\delta & = & 1+i. \end{array}$$

The matrix augmented by the right-hand side of this system of equation brought into (unreduced)[7] row echelon form is

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

This shows that in solving these equation, one can choose δ arbitrarily. After choosing δ is is a simple matter to find the other unknowns: $\alpha = 4i$, $\beta = -2+2i$, and $\gamma = -1-i+(1-i)\delta$. That is

$$\mathbf{z}_3 = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} 4i \\ -2+2i \\ -1-i \\ 0 \end{pmatrix} + \delta \begin{pmatrix} 0 \\ 0 \\ 1-i \\ 1 \end{pmatrix} = \begin{pmatrix} 4i \\ -2+2i \\ -1-i \\ 0 \end{pmatrix} + \frac{\delta}{1+i}\mathbf{z}'_3;$$

recall that $\mathbf{z}'_3$ was found to be an eigenvector of B_2 associated with the eigenvalue i; that is $(B_2 - iI)\mathbf{z}'_3 = \mathbf{0}$. Hence, the presence of $\mathbf{z}'_3$ on the right-hand side of the equation z_3 is to be expected; after all, $\mathbf{z}_3$ is the solution of the equation $(B_2 - iI)\mathbf{z}_3 = \mathbf{z}'_3$. The simplest choice is to take $\delta = 0$; then $\mathbf{z}_3 = (4i, -2+2i, -1-i, 0)$. The vectors $\mathbf{z}_3$ and $(B_2 - iI)\mathbf{z}_3$ span a Jordan subspace of B_2 for the eigenvalue i.

Similarly, for the eigenvector $\mathbf{z}'_4$ associated with the eigenvalue $-i$ we need to find a vector $\mathbf{z}_4$ for which $(B_2 + iI)\mathbf{z}_4 = \mathbf{z}'_4$. As the components of the matrix B_2 are real, we can find $\mathbf{z}_4$ simply by complex conjugation. Indeed $-i$ is the complex conjugate of i and the components of $\mathbf{z}'_4$ are the complex conjugates of $\mathbf{z}_3$; hence the components of $\mathbf{z}_4$ will be the complex conjugates of the components of $\mathbf{z}_3$; that is, we have $\mathbf{z}_4 = (-4i, -2-2i, -1+i, 0)$. Further, $\mathbf{z}_4$ and $(B_2 + iI)\mathbf{z}_4$ span a Jordan subspace of B_2 for the eigenvalue $-i$.

7.3 The Jordan canonical form

Next we need to revert back to the matrix A_2. Since we have $B_2 = EA_2E_1$, we the vectors $\mathbf{z}_3$, $\mathbf{z}'_3$, $\mathbf{z}_4$, and $\mathbf{z}'_4$ will be transformed into $\mathbf{y}_3 = E^{-1}\mathbf{z}_3$, $\mathbf{y}'_3 = E^{-1}\mathbf{z}'_3$, $\mathbf{y}_4 = E^{-1}\mathbf{z}_4$, and $\mathbf{y}'_4 = E^{-1}\mathbf{z}'_3$. Since we have $E(A_2 - iI)\mathbf{y}_3 = E(A_2 - iI)E^{-1}\mathbf{z}_3 = (EA_2E^{-1} - iI)\mathbf{z}_3 = (B_2 - iI)\mathbf{z}_3 = \mathbf{z}'_3$, we have $(A_2 - iI)\mathbf{y}_3 = E^{-1}\mathbf{z}'_3 = \mathbf{y}'_3$; similarly for the other equations. That is,

$$(22) \qquad (A_2 - iI)\mathbf{y}_3 = \mathbf{y}'_3, \quad (A_2 - iI)\mathbf{y}'_3 = \mathbf{0}, \quad (A_2 - iI)\mathbf{y}_4 = \mathbf{y}'_4, \quad (A_2 - iI)\mathbf{y}'_4 = \mathbf{0},$$

where

$$\mathbf{y}_3 = \begin{pmatrix} -2+2i \\ -3+i \\ -1-i \\ 6-2i \end{pmatrix}, \quad \mathbf{y}'_3 = \begin{pmatrix} 2-2i \\ 2 \\ 1-i \\ -3+i \end{pmatrix}, \quad \mathbf{y}_4 = \begin{pmatrix} -2-2i \\ -3-i \\ -1+i \\ 6+2i \end{pmatrix}, \quad \mathbf{y}'_3 = \begin{pmatrix} 2+2i \\ 2 \\ 1+i \\ -3-i \end{pmatrix}.$$

We have $A_1 = I_1 \oplus I_1 \oplus A_2$, so the Jordan subspaces of A_1 are spanned by $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}'_1$, $\mathbf{x}_4$, and $\mathbf{x}'_4$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are unit vectors, and $\mathbf{x}_3$, $\mathbf{x}'_1$, $\mathbf{x}_4$, and $\mathbf{x}'_4$ were obtained from $\mathbf{y}_3$, $\mathbf{y}'_1$, $\mathbf{y}_4$, and $\mathbf{y}'_4$

[7] The row echelon form discussed row echelon form discussed in [1, pp. 57–62] is often called *reduced* row echelon form, where the first nonzero entry in the row echelon form, called the leading entry, of the matrix is 1, and this is the only nonzero entry in the column. In the unreduced form it is not required that each leading entry be 1, and that the entries above the leading entry to be nonzero. The relevant facts about the solution of systems of linear equation discussed in [1] in the mentioned pages remain valid for this unreduced echelon form. In our case, the leading entries are 1, but they are not the only nonzero entries in their column.

by adding two 0 components on top:

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ -2+2i \\ -3+i \\ -1-i \\ 6-2i \end{pmatrix},$$

$$\mathbf{x}'_3 = \begin{pmatrix} 0 \\ 0 \\ 2-2i \\ 2 \\ 1-i \\ -3+i \end{pmatrix}, \quad \mathbf{x}_4 = \begin{pmatrix} 0 \\ 0 \\ -2-2i \\ -3-i \\ -1+i \\ 6+2i \end{pmatrix}, \quad \mathbf{x}'_3 = \begin{pmatrix} 0 \\ 0 \\ 2+2i \\ 2 \\ 1+i \\ -3-i \end{pmatrix}.$$

We have

$$A_1\mathbf{x}_1 = \mathbf{x}_1, \quad A_1\mathbf{x}_2 = \mathbf{x}_2, \quad A_1\mathbf{x}'_3 = i\mathbf{x}'_3,$$
$$A_1\mathbf{x}_3 = \mathbf{x}'_3 + i\mathbf{x}_3, \quad A_1\mathbf{x}'_4 = -i\mathbf{x}'_4, \quad A_1\mathbf{x}_4 = \mathbf{x}'_4 - i\mathbf{x}_4;$$

the first two of these equations can be verified immediately, while the rest of the equations, and the rest of them are immediate consequence of equations (21) – the third one follows from the second one above, the fourth from the first, the fifth from the fourth, and the sixth from the third. These equations show that the matrix A_1 in the basis $\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}'_3, \mathbf{x}_3, \mathbf{x}'_4, \mathbf{x}_4)$ is the Jordan canonical form of A_1 (note the *order* the vectors are listed). That is, for the matrix

$$X = \mathcal{R}_{\mathcal{X}\mathcal{X}}T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2-2i & -2+2i & 2+2i & -2-2i \\ 0 & 0 & 2 & -3+i & 2 & -3-i \\ 0 & 0 & 1-i & -1-i & 1+i & -1+i \\ 0 & 0 & -3+i & 6-2i & -3-i & 6+2i \end{pmatrix}$$

whose columns are the basis vectors in $\mathcal{X}$, we have

$$J = X^{-1}A_1X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & i & 1 & 0 & 0 \\ 0 & 0 & 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 & -i & 1 \\ 0 & 0 & 0 & 0 & 0 & -i \end{pmatrix}$$

That is, with the matrix S_1 defined in (19), writing

$$\begin{aligned} S &= XS_1 \\ &= \begin{pmatrix} -1 & 0 & 2i & -1-i & -2i & -1+i \\ 0 & 1 & -2+6i & 10-14i & -2-6i & 10+14i \\ -1 & 0 & 0 & 4i & 0 & -4i \\ 0 & -1 & 2-8i & -13+19i & 2+8i & -13-19i \\ 1 & 0 & 2i & 3-5i & -2i & 3+5i \\ 0 & 1 & 4i & 6-10i & -4i & 6+10i \end{pmatrix}, \end{aligned}$$

we have $S^{-1}AS$, where A is the matrix given in (18) (cf. (7.1)).

8 Matrices and field extensions

The example described in Sections 6 and 7 gives a matrix over the field $\mathbb{R}$ of real numbers for which the Jordan Canonical Form was over the field $\mathbb{C}$ of complex numbers; on the other hand, the minimal polynomial was in the field of real numbers. The question then arises:

Let F be a subfield of F' and let A be a square matrix over the field F. Clearly, A can also be considered as a matrix over F'. Is the minimal polynomial of A considered as a matrix over F the same as it is when A is considered as a matrix over F'? Similarly, is the characteristic polynomial of A considered as a matrix over F the same as it is when A is considered as a matrix over F'?

The answer for the characteristic polynomial is an obvious yes, since the characteristic polynomial is $\det(A - xI)$, where I is the identity matrix of the same size as A. The answer is also yes for the minimal polynomial, but for less obvious reasons. In Subsection 4.4 we described how to find the minimal polynomial of a linear transformation. In case the vector space is a space of column vectors, the steps in this procedure do not change if we change the field from F to F'. The reason is that the linear independence of a system of column vectors does not change when we change the field F to F'; or, what amounts to the same, the rank of a matrix A is the same whether this matrix is considered as a matrix over F or over F'. This is because the rank of A can be determined by the triangulating A, and this triangulation can be carried out in the field F.

As far as the Cayley–Hamilton theorem is concerned in Subsection 4.3, we needed to assume that the underlying field F is algebraically closed; however, the determination of the minimal polynomial did not depend on this fact. As for the characteristic polynomial, even its definition depended on the algebraic closedness of F. However, we can define the characteristic polynomial by first representing T as a matrix (see Subsection 5.1), assuming that V is finite dimensional, and then taking the characteristic polynomial of the matrix representing T.

Even if F is not algebraically closed, it still remains to be true that the minimal polynomial $P_{\mathrm{min},T}(x)$ of T is a divisor of the characteristic polynomial $P_{\mathrm{char},T}(x)$; hence the Cayley–Hamilton theorem remains valid.

8.1 A further decomposition theorem

We have

Lemma 5. *Let V be a finite dimensional vector space over a field F, let $T : V \to V$ be a linear transformation. Further let $P(x)$ be a polynomial such that $P(T) = 0$, and assume that $P(x) = P_1(x)P_2(x)$ such that the greatest common divisor of $P_1(x)$ and $P_2(x)$ is 1. Let*

$$V_i = \{\mathbf{v} : P_i(T)\mathbf{v} = \mathbf{0}\}$$

for $i = 1$ or $i = 2$. Then $V = V_1 \oplus V_2$.

It is clear that V_1 and V_2 are invariant subspaces for T.

Proof. We need to prove that $V_1 \cap V_2 = \{\mathbf{0}\}$ and that V_1 and V_2 span V. Since the greatest common divisor of $P_1(x)$ and $P_2(x)$ is 1, we have polynomials $Q_1(x)$ and $Q_2(x)$ such that

$$P_1(x)Q_1(x) + Q_2(x)R_2(x) = 1.$$

Hence, for any vector $\mathbf{v} \in V$ we have

$$P_1(T)Q_1(T)\mathbf{v} + Q_2(T)R_2(T)\mathbf{v} = (P_1(T)Q_1(T) + Q_2(T)R_2(T)) = I\mathbf{v} = \mathbf{v}, \tag{23}$$

where $I : V \to V$ is the identity transformation. Now, if $\mathbf{v} \in V_1 \cap V_2$ then $P_1(T)Q_1(T)\mathbf{v} = \mathbf{0}$, since $\mathbf{v} \in V_1$, and $P_2(T)Q_2(T)\mathbf{v} = \mathbf{0}$ since $\mathbf{v} \in V_2$; thus the vector on the left-hand side of (23) is $\mathbf{0}$; hence $\mathbf{v} = \mathbf{0}$. This shows that $V_1 \cap V_2 = \{\mathbf{0}\}$.

If $\mathbf{v} \in V$ is an arbitrary vector, then we have $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ according to (23) with $\mathbf{v}_1 = P_2(T)Q_2(T)\mathbf{v}$ and $\mathbf{v}_2 = P_1(T)Q_1(T)\mathbf{v}$. As $P(T) = P_1(T)P_2(T) = 0$ we have

$$P_1(T)\mathbf{v}_1 = P_1(T)P_2(T)\mathbf{v}_1 = P(T)\mathbf{v}_1 = \mathbf{0},$$

and so $\mathbf{v}_1 \in V_1$. Similarly,

$$P_2(T)\mathbf{v}_2 = P_2(T)P_1(T)\mathbf{v}_2 = P(T)\mathbf{v}_1 = \mathbf{0},$$

and so $\mathbf{v}_2 \in V_2$. This shows that $\mathbf{v}$ is in the span of V_1 and V_2, completing the proof. □

Corollary 3. *Let V be a finite dimensional vector space over a field F, let $T : V \to V$ be a linear transformation. Further let $P(x)$ be the minimal polynomial of T, and assume that $P(x) = \prod_{i=1}^{n} P_i(x)$ such that the greatest common divisor of $P_i(x)$ and $P_j(x)$ is 1 for any i and j with $1 \le i < j \le n$. Let*

$$V_i = \{\mathbf{v} : P_i(T)\mathbf{v} = \mathbf{0}\}$$

for i with $1 \le i \le n$. Then $V = \bigoplus_{i=1}^{n} V_i$.

This can be proved by repeated application of Lemma 5 by making use of the fact that V_1 and V_2 are invariant subspaces for T of V.

References

[1] Hans Schneider and George Philip Barker. *Matrices and Linear Algebra, 2nd ed.* Dover Publications, New York, 1973.