The cyclic decomposition theorem

Attila Máté
Brooklyn College of the City University of New York

December 1, 2014; revised: April 18, 2016

Contents

1 Polynomials over a field

Lemma 1.1 (Division theorem)**.** *Let $M(x)$ and $D(x)$ be polynomials over the field F. Assume $D(x)$ is not zero. Then there are polynomials $Q(x)$ and $R(x)$ such that*

$$M(x) = D(x)Q(x) + R(x) \quad \textit{and} \quad \deg R(x) < \deg D(x);$$

here we take the degree of the zero polynomial to be -1.

This an be easily proved by induction on the degree of $M(x)$. The usual algorithm of dividing polynomials can be used to find the polynomials $Q(x)$ and $R(x)$.

Definition 1.1. Given polynomials $P(x)$ and $Q(x)$, the least common multiple $\operatorname{lcm}(P(x), Q(x))$ of $P(x)$ and $Q(x)$ is defined as the monic polynomial $M(x)$ of the lowest degree such that $P(x) \mid M(x)$ and $Q(x) \mid M(x)$. If at least one of $P(x)$ and $Q(x)$ is nonzero, then the greatest common divisor $\gcd(P(x), Q(x))$ of $P(x)$ and $Q(x)$ is defined as the monic polynomial $D(x)$ of the highest degree such that $D(x) \mid P(x)$ and $D(x) \mid Q(x)$.

Lemma 1.2. *Let $P(x)$ and $Q(x)$ be polynomials over a field F at least one of which is nonzero. Then there are polynomials $M(x)$ and $N(x)$ such that*

$$\gcd(P(x), Q(x)) = P(x)M(x) + Q(x)N(x).$$

Further, if $H(x)$ is a polynomial such that $H(x) \mid P(x)$ and $H(x) \mid Q(x)$, then

$$H(x) \mid \gcd(P(x), Q(x)).$$

Proof. Let $D(x)$ be the monic polynomial with the smallest degree such that

$$D(x) = P(x)M(x) + Q(x)N(x). \tag{1.1}$$

for some polynomials $M(x)$ and $N(x)$. We claim that $D(x) = \gcd(P(x), Q(x))$.

To show this, we will first show that $D(x) \mid P(x)$. Indeed, assume that this is not the case. Then we have

$$P(x) = D(x)H(x) + R(x)$$

for some nonzero $R(x)$ with degree lower than $D(x)$, according to Lemma 1.1. Since we have

$$R(x) = P(x) - D(x)H(x) = P(x)(1 - M(x)H(x)) + Q(x)(-N(x)H(x)),$$

this contradicts the choice of $D(x)$ as the lowest degree nonzero polynomial of form (1.1), showing that indeed $D(x) \mid P(x)$. Similarly, we can see that $D(x) \mid Q(x)$.

Next, assume $H(x) \mid P(x)$ and $H(x) \mid Q(x)$. Then we have $P(x) = H(x)A(x)$ and $Q(x) = H(x)B(x)$ for some polynomials $A(x)$ and $B(x)$, and so, by (1.1), we have

$$\begin{aligned} D(x) &= P(x)M(x) + Q(x)N(x) \\ &= H(x)A(x)M(x) + H(x)B(x)N(x) = H(x)(A(x)M(x) + B(x)N(x)), \end{aligned}$$

and so $H(x) \mid D(x)$. Thus $\deg H(x) \le \deg D(x)$, which shows that $D(x)$ is indeed the greatest common divisor of $P(x)$ and $Q(x)$. The second displayed equation of the lemma is also established. □

Definition 1.2. A non-constant polynomial $P(x)$ over the field F is called irreducible if there are no non-constant polynomials $A(x)$ and $B(x)$ such that $P(x) = A(x)B(x)$.

The polynomial $x^2 + 1$ is irreducible over the field of real numbers. In fact, the only irreducible polynomials of degree > 1 over the field of real numbers are quadratic polynomials with no real zeros. In the next section we will see that there are irreducible polynomials of arbitrarily high degree over the field of rational numbers. Irreducible polynomials play a role similar to prime numbers in the integers according to the first corollary the next lemma.

Lemma 1.3. *Let $P(x)$, $A(x)$, and $B(x)$ be polynomials over the field F, and assume that*

$$\gcd(P(x), A(x)) = 1.$$

Assume, further, that $P(x) \mid A(x)B(x)$. Then $P(x) \mid B(x)$.

Proof. By Lemma 1.2 there are polynomials $M(x)$ and $N(x)$ such that

$$1 = M(x)P(x) + N(x)A(x).$$

Multiplying this by $B(x)$, we obtain that

$$B(x) = M(x)P(x)B(x) + N(x)A(x)B(x).$$

As $P(x) \mid A(x)B(x)$, the right-hand side here is divisible by $P(x)$; hence so is the left-hand side. That is, $P(x) \mid B(x)$, which is what we wanted to show. □

Corollary 1.1. *Let $P(x)$, $A(x)$, and $B(x)$ be polynomials over the field F, and assume $P(x)$ is irreducible. Assume, further, that $P(x) \mid A(x)B(x)$. Then $P(x) \mid A(x)$ or $P(x) \mid Q(x)$.*

Proof. Assume that $P(x) \nmid A(x)$; we then have to show that $P(x) \mid B(x)$. Since the only divisors of $P(x)$ are constant multiples of 1 and $P(x)$, we have $\gcd(P(x), A(x)) = 1$. Hence, by Lemma 1.3, $P(x) \mid B(x)$, which is what we wanted to show. □

An easy consequence of this is that every polynomial can be written as a product of irreducible polynomials in an essentially unique way:

Corollary 1.2. *Let $P(x)$ be a polynomial over the field F. Then there is an integer $n \geq 0$ and irreducible monic polynomials $P_i(x)$ for i with $1 \leq i \leq n$, and an $\alpha \in F$ such that*

$$P(x) = \alpha \prod_{i=1}^{n} P_i(x).$$

Further, this representation is unique aside from the order in which the irreducible monic polynomials on the right-hand side are listed.

The requirement that the irreducible polynomials be monic is not essential, except that if we allow them to be non-monic, then there are more representations in that the element α can be distributed among the polynomials on the right-hand side in an arbitrary way.

2 Polynomials over the rationals

Theorem 2.1 (Gauss)**.** *Let $P(x)$ be a polynomial with integer coefficients, let $Q(x)$ and $R(x)$ be non-constant polynomials with rational coefficients, and assume $P(x) = Q(x)R(x)$. Then there is a rational number ρ such that the polynomials $\rho Q(x)$ and $(1/\rho)R(x)$ have integer coefficients.*

For the proof, we need two lemmas.

Lemma 2.1. *Let p be a prime, let m and n be positive integers, and assume that*

$$\sum_{\nu=0}^{m+n} a_\nu x^\nu = \sum_{i=0}^{m} b_i x^i \cdot \sum_{j=0}^{n} c_j x^j$$

for all x, where a_ν, b_i, c_i are integers. Let k and l be integers with $0 \le k \le m$ and $0 \le l \le n$. Assume that $p \mid b_i$ and $p \mid c_j$ for $0 \le i < k$ and $0 \le j < l$ and $p \nmid b_k$ and $p \nmid c_l$. Then $p \nmid a_{k+l}$.

Proof. To simplify the notation, it is expedient to define b_i and c_j for all integers by keeping the original values of b_i and c_j when $0 \le i \le m$ and $0 \le j \le n$, and putting $b_i = c_j = 0$ for i and j outside these ranges. We have

$$a_{k+l} = \sum_{i=-\infty}^{+\infty} b_i c_{k+l-i}.$$

On the right-hand side, the term for $i = k$ is not divisible by p, since $p \nmid b_k$ and $p \nmid c_l$. For all other terms we have either $i < k$ or $k + l - i < l$, and all these terms are divisible by p, since $p \mid b_i$ and $p \mid c_j$ for $i < k$ and $j < l$ (note that $p \mid 0$). Thus $p \nmid a_{k+l}$, as we wanted to show. □

Corollary 2.1. *Let p be a prime, and let $P(x)$ be a polynomial with integer coefficients, and assume that each coefficient of $P(x)$ is divisible by p. Assume, further, that $P(x) = Q(x)R(x)$, where the coefficients of the non-constant polynomials $Q(x)$ and $R(x)$ are also integers. Then every coefficient of either $Q(x)$ or $R(x)$ is divisible by p.*

Proof. Using the notation of Lemma 2.1, there can be no k and l as described in that Lemma; otherwise, we would have $p \nmid a_{k+l}$, whereas we have $p \mid a_\nu$ for all ν with $0 \le \nu \le m + n$. The non-existence of a k as described means that $p \mid b_i$ for all i with $0 \le i \le m$; similarly, the non-existence of an l as described means that $p \mid c_j$ for all j with $0 \le j \le n$. In either case, the conclusions of the lemma to be proved are satisfied, as we wanted to show. □

Proof of Theorem 2.1. Choose integers b and c such that $bQ(x)$ and $cR(x)$ have integer coefficients; we have $bcP(x) = (bQ(x))(cR(x))$. Using Corollary 2.1 repeatedly, we can divide this equation by each of the prime factors p of bc, each time choosing to divide the first or the second polynomial on the right-hand side by p so as to assure that these polynomials will still have integer coefficients. When we finish, we will arrive at an equation $P(x) = (\rho_1 Q(x))(\rho_2 R(x))$ for some rational numbers ρ_1 or ρ_2, where $\rho_1 Q(x)$ and $\rho_2 R(x)$ have integer coefficients. As $P(x) = Q(x)R(x)$, we must have $\rho_1\rho_2 = 1$. The conclusions of the theorem are then satisfied by $\rho = \rho_1$. □

Theorem 2.2 (Eisenstein's criterion)**.** *Let p be a prime, and let $P(x)$ be a non-constant polynomial with integer coefficients, and assume that the leading coefficient of $P(x)$ is not divisible by p, all its other coefficients are divisible by p, and its constant term is not divisible by p^2. Then $P(x)$ is irreducible over the field of rationals.*

Proof. Assume, on the contrary, that $P(x)$ is not irreducible over the rationals. Then by Theorem 2.1 there are non-constant polynomials $Q(x)$ and $R(x)$ with integer coefficients such that $P(x) = Q(x)R(x)$. Using the notation of Lemma 2.1, the leading coefficient of $P(x)$ is $a_{m+n} = b_m c_n$. According to our assumptions, $p \nmid a_{m+n}$; hence $p \nmid b_m$ and $p \nmid c_n$. Therefore, integers k and l as described in Lemma 2.1 do exist.

We have $a_0 = b_0 c_0$. Since $p^2 \nmid a_0$, we have either $p \nmid b_0$ or $c \nmid c_0$. Without loss of generality, we may assume that $p \nmid c_0$; this means that $l = 0$. According to Lemma 2.1, $p \nmid a_{k+l}$; as $k \leq m$, $l = 0$, and $n \geq 1$, we have $k + l < m + n$ ($n > 0$ since $R(x)$ is not constant). This is a contradiction, since $p \mid a_\nu$ for all ν with $0 \leq \nu < m + n$ according to our assumptions. □

Eisenstein's criterion allows one to construct polynomials or arbitrarily high degree that are irreducible over the rationals. An interesting example is the pth *cyclotomic* polynomial

$$\frac{x^p - 1}{x - 1} = \sum_{n=0}^{p-1} x^n$$

for any prime p.[2.1] Clearly, this polynomial does not satisfy the assumptions of Eisenstein's criterion, but if one replaces x with $x + 1$, one obtains the polynomials

$$\frac{(x+1)^p - 1}{x} = \sum_{n=1}^{p} \binom{p}{n} x^{n-1},$$

where the equation follows from the binomial theorem. We have $\binom{p}{p} = 1$, $\binom{p}{1} = p$, and $p \mid \binom{p}{n}$ for n with $1 \leq n < p$. This shows that this latter polynomials satisfies Eisenstein's criterion, and so it is irreducible over the rationals. Hence the former polynomial is also irreducible over the rationals.

3 Conductor polynomial

Definition 3.1. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V. W is called *invariant* for T is $T\mathbf{w} \in W$ for all $\mathbf{w} \in W$.

Definition 3.2. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V invariant for T, and let $\mathbf{v} \in V$. The conductor polynomial $P_{\text{cond } W,T,\mathbf{v}}(x)$ for T into W of $\mathbf{v}$ is the monic polynomial $P(x)$ of the lowest degree for which $P(T)\mathbf{v} \in W$. The conductor polynomial $P_{\text{cond } W,T}(x)$ for T into W is the monic polynomial $P(x)$ of the lowest degree for which $P(T)\mathbf{v} \in W$ for all $\mathbf{v} \in V$.

Observe that the conductor polynomial $P_{\text{cond } W,T,\mathbf{v}}(x)$ always exists and is unique. As for its uniqueness, if there are two different monic polynomials $P(x)$ and $Q(x)$ of the same degree for which $P(T)\mathbf{v} \in W$ and $Q(T)\mathbf{v} \in W$, then $\big(P(T) - Q(T)\big)\mathbf{v} \in W$, and $P(x) - Q(x)$ has lower degree, and it can be made into a monic polynomial by dividing it by its leading coefficients. As for its existence, there is a positive integer n for which the system $(T^k\mathbf{v} : 0 \leq k \leq n)$ are linearly dependent, since V is finite dimensional. Then

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

[2.1] The term cyclotomic means *that which cuts the circle.* The name is explained by the location of the zeros of this polynomial on the unit circle. The nth cyclotomic polynomial is also defined if $n \geq 2$ is not a prime, but it has a more complicated form.

for some $c_k \in F$ not all of which is zero. Then $P(x) = \sum_{k=0}^{n} c_k x^k$ is a nonzero polynomial for which $P(T)\mathbf{v} = \mathbf{0} \in W$.

The existence and uniqueness of the conductor polynomial $P_{\text{cond } W,T}(x)$ of T of T is similarly easy to see. As for its uniqueness, if there are two different monic polynomials $P(x)$ and $Q(x)$ of the same degree for which $P(T)\mathbf{v} \in W$ and $Q(T)\mathbf{v} \in W$ for all $\mathbf{v} \in V$, then $\big(P(T) - Q(T)\big)\mathbf{v} \in W$, and $P(x) - Q(x)$ has lower degree, and it can be made into a monic polynomial by dividing it by its leading coefficients. As for its existence, if $(\mathbf{v}_i : 1 \le i \le n)$ span V, then the $P(x) = \prod_{i=1}^{n} P_{\text{cond } W,T,\mathbf{v}_i}(x)$ is a polynomial such that $P(T)\mathbf{v} \in W$ for all $\mathbf{v} \in V$.

Lemma 3.1. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V invariant for T, let $\mathbf{v} \in V$, $\mathbf{v} \ne \mathbf{0}$, and let $P_{\text{cond } W,T,\mathbf{v}}(x)$ be the conductor polynomial for T into W of $\mathbf{v}$. Assume, further, that $P(x)$ is a polynomial such that $P(T)\mathbf{v} \in W$. Then $P_{\text{cond } W,T,\mathbf{v}}(x) \mid P(x)$. Further, if $P(T)\mathbf{v} \in W$ for all $\mathbf{v} \in V$, then $P_{\text{cond } W,T}(x) \mid P(x)$.*

Proof. As for the first conclusion, assume that $P_{\text{cond } W,T,\mathbf{v}}(x) \nmid P(x)$, and let $R(x)$ be the remainder when $P_{\text{cond } W,T,\mathbf{v}}(x)$ is divided into $P(x)$. Then $R(x)$ is a nonzero polynomial such that $R(T)\mathbf{v} \in W$; since $R(x)$ has lower degree than $P_{\text{cond } W,T,\mathbf{v}}(x)$, this is a contradiction.

The second conclusion can be established similarly. Assuming that $P_{\text{cond } W,T}(x) \nmid P(x)$, the remainder $R(x)$ when $P_{\text{cond } W,T}(x)$ is divided into $P(x)$ is such that $R(x)\mathbf{v} \in V$ for all $\mathbf{v} \in V$. This is a contradiction, since the degree of $R(x)$ is lower than that of $P_{\text{cond } W,T}(x)$. □

Lemma 3.2. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V invariant for T, and let $P_{\text{cond } W,T}(x)$ be the conductor polynomial for T into W of V. Let $P(x)$ a polynomial irreducible over F, let n be a positive integer, and assume $\big(P(x)\big)^n \mid P_{\text{cond } W,T}(x)$. Then there is a $\mathbf{v} \in V$ such that $P_{\text{cond } W,T,\mathbf{v}}(x) = \big(P(x)\big)^n$.*

Proof. Let $Q(x)$ be the polynomial such that $P_{\text{cond } W,T}(x) = \big(P(x)\big)^n Q(x)$ and let

$$U = \{Q(T)\mathbf{v} : \mathbf{v} \in V\}.$$

Then there must be a $\mathbf{u} \in U$ for which $\big(P(T)\big)^{n-1}\mathbf{u} \notin W$; otherwise we would have $P_{\text{cond } W,T,\mathbf{v}}(x) \mid \big(P(x)\big)^{n-1} Q(x)$ for all $\mathbf{v} \in V$ according to Lemma 3.1, and so $P_{\text{cond } W,T}(x) \mid \big(P(x)\big)^{n-1} Q(x)$. On the other hand, we must have $\big(P(T)\big)^n \mathbf{u} \in W$; hence $P_{\text{cond } W,T,\mathbf{u}}(x) = \big(P(x)\big)^n$. □

Lemma 3.3. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V invariant for T, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V, and let $P(x)$ and $Q(x)$ be polynomials such that* $\gcd(P(x), Q(x)) = 1$. *Assume $P_{\text{cond } W,T,\mathbf{u}}(x) = P(x)$ and $P_{\text{cond } W,T,\mathbf{v}}(x) = Q(x)$. Then $P_{\text{cond } W,T,\mathbf{u}+\mathbf{v}}(x) = P(x)Q(x)$.*

Proof. We have

$$P(T)Q(T)(\mathbf{u} + \mathbf{v}) = Q(T)P(T)\mathbf{u} + P(T)Q(T)\mathbf{v} \in W,$$

so $P_{\text{cond } W,T,\mathbf{u}+\mathbf{v}}(x) \mid P(x)Q(x)$ according to Lemma 3.1. Hence $P_{\text{cond } W,T,\mathbf{u}+\mathbf{v}}(x) = P_1(x)Q_1(x)$ for some polynomials $P_1(x)$ and $Q_1(x)$ such that that $P_1(x)$ is a divisor of $P(x)$ and $Q_1(x)$ is a divisor of $Q(x)$. Assuming that $P_{\text{cond } W,T,\mathbf{u}+\mathbf{v}}(x) \ne P(x)Q(x)$, either $P_1(x)$ is a proper divisor of $P(x)$ or $Q_1(x)$ is a proper divisor of $Q(x)$. Without loss of generality, we may assume the former.

We have $P_1(T)Q(T)(\mathbf{u}+\mathbf{v}) = P_1(T)Q(T)\mathbf{u} + P_1(T)Q(T)\mathbf{v} = Q(T)P_1(T)\mathbf{u} + \mathbf{w}$, where the vector $\mathbf{w} = P_1(T)Q(T)\mathbf{v}$ belongs to W since $Q(x) = P_{\text{cond } W,T,\mathbf{v}}(x)$, and $P_1(T)Q(T)\mathbf{u} \notin W$ according to Lemma 3.1, since $P_{\text{cond } W,T,\mathbf{u}}(x) = P(x) \nmid P_1(x)Q(x)$. Thus $P_1(T)Q(T)(\mathbf{u} + \mathbf{v}) \notin W$, contradicting the assumption $P_{\text{cond } W,T,\mathbf{u}+\mathbf{v}}(x) = P_1(x)Q_1(x)$. □

Theorem 3.1. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V invariant for T. Then there is a $\mathbf{v} \in V$ such that $P_{\text{cond } W,T,\mathbf{v}}(x) = P_{\text{cond } W,T}(x)$.*

Proof. The result follows from Lemmas 3.2 and 3.3. □

3.1 Minimal polynomial

The polynomial $P_{\min T,\mathbf{v}}(x) \overset{def}{=} P_{\text{cond } \{\mathbf{0}\},T,\mathbf{v}}(x)$ is called the minimal polynomial for T of the vector $\mathbf{v}$, and $P_{\min T}(x) \overset{def}{=} P_{\text{cond } \{\mathbf{0}\},T}(x)$ is called the minimal polynomial of T. If U is a subspace of V invariant for T, we will write $P_{\min U,T}(x)$ for the minimal polynomial of T restricted to the subspace U. The results of this section can be directly translated into results involving minimal polynomials, and they will not be stated separately.

4 Admissible subspaces and the Cyclic Decomposition Theorem

Definition 4.1. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V. W is called *admissible* for T if W is invariant for T and for every vector $\mathbf{v} \in V$ and for every polynomial $P(x)$ such that $P(T)\mathbf{v} \in W$ there is a $\mathbf{w} \in W$ such that $P(T)\mathbf{v} = P(T)\mathbf{w}$.

Lemma 4.1. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let W be a subspace of V that is invariant for T. If for any $\mathbf{v} \in V$ there is a $\mathbf{w} \in W$ such that $P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{v} = P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{w}$ then W is admissible for T.*

Proof. If $P(T)\mathbf{v} \in W$ then $P(x) = Q(x)P_{\text{cond } W,T,\mathbf{v}}(x)$ for some polynomial $Q(x)$ according to Lemma 3.1. If there is a $\mathbf{w} \in W$ such that $P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{v} = P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{w}$ then $P(T)\mathbf{v} = Q(T)P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{v} = Q(T)P_{\text{cond } W,T,\mathbf{v}}(T)\mathbf{w} = P(T)\mathbf{w}$. □

Lemma 4.2. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let $W \neq V$ be a subspace of V that is admissible for T. Then there is a vector $\mathbf{z} \in V$ such that*

$$P_{\min T,\mathbf{z}}(x) = P_{\text{cond } W,T,\mathbf{z}}(x) = P_{\text{cond } W,T}(x) \tag{4.1}$$

and

$$P_{\text{cond } W,T,\mathbf{z}}(T)\mathbf{z} = \mathbf{0}. \tag{4.2}$$

Strictly speaking, the case $W = V$ need not be excluded, since in this case the conclusions of the lemma are satisfied by the vector $\mathbf{z} = \mathbf{0}$; of course, this case is of absolutely no interest.

Proof. According to Theorem 3.1, there is a $\mathbf{u} \in V$ such that $P_{\text{cond } W,T,\mathbf{u}}(x) = P_{\text{cond } W,T}(x)$. Let $\mathbf{w} \in W$ be such that $P_{\text{cond } W,T,\mathbf{u}}(T)\mathbf{u} = P_{\text{cond } W,T,\mathbf{u}}(T)\mathbf{w}$. Taking $\mathbf{z} = \mathbf{u} - \mathbf{w}$, we have $P_{\text{cond } W,T,\mathbf{z}}(x) = P_{\text{cond } W,T,\mathbf{u}}(x) = P_{\text{cond } W,T}(x)$ and $P_{\text{cond } W,T}(T)\mathbf{z} = P_{\text{cond } W,T,\mathbf{u}}(T)\mathbf{u} - P_{\text{cond } W,T,\mathbf{u}}(T)\mathbf{w} = \mathbf{0}$. This last equation also implies that $P_{\min T,\mathbf{z}}(x) = P_{\text{cond } W,T,\mathbf{z}}(x)$. □

Definition 4.2. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, The *cyclic subspace* $Z(\mathbf{z}, T)$ is the vector space spanned by the system $(T^k\mathbf{z} : k \geq 0)$.

While the system $(T^k\mathbf{z} : k \geq 0)$ may be infinite, it has finite subsystem that also spans $Z(\mathbf{z}, T)$ because V is finite dimensional. The letter Z is used in the notation for cyclic subspaces since the German word for cyclic is *zyklisch*. Given two subspaces U_1 and U_2 of V, we put

$$U_1 + U_2 \overset{def}{=} \{\mathbf{u}_1 + \mathbf{u}_2 : \mathbf{u}_1 \in U_1 \quad \text{and} \quad u_2 \in U_2\}.$$

Clearly, $U_1 + U_2$ is also a subspace of V. We have

Lemma 4.3. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let $W \neq V$ be a subspace of V admissible for T, and let $\mathbf{z} \in V$ be such that equation (4.1) is satisfied (and so (4.2) is also satisfied). Then $Z(\mathbf{z}, T) \cap W = \{\mathbf{0}\}$, and the subspace $Z(\mathbf{z}, T) + W$ is admissible.*

Proof. Assume $\mathbf{u} \in Z(\mathbf{z}, T) \cap W$. Every element of $Z(\mathbf{z}, T)$ can be written as $P(T)\mathbf{z}$ for some polynomial $P(x)$; let $P(x)$ be a polynomial for which $\mathbf{u} = P(T)\mathbf{z}$. Then $P(T)\mathbf{z} \in W$, and so $P(x) = Q(x)P_{\text{cond } W,T,\mathbf{z}}(x)$ for some polynomial $Q(x)$ according to Lemma 3.1. Hence $\mathbf{u} = P(T)\mathbf{z} = Q(T)P_{\text{cond } W,T,\mathbf{z}}(T)\mathbf{z} = \mathbf{0}$; the last equality holds according to (4.2). Therefore $Z(\mathbf{z}, T) \cap W = \{\mathbf{0}\}$, verifying the first assertion of the lemma.

To show the second assertion, write $Z = Z(\mathbf{z}, T)$, let $\mathbf{u} \in V$ be arbitrary. Writing

$$P(x) = P_{\text{cond } Z \oplus W,T,\mathbf{u}}(x),$$

we have

$$P(T)\mathbf{u} = \mathbf{w}' + \mathbf{z}', \tag{4.3}$$

where $\mathbf{w}' \in W$ and $\mathbf{z}' \in Z$. It will be enough to show show that there are $\mathbf{w}'' \in W$ and $\mathbf{z}'' \in Z$ such that

$$\mathbf{w}' = P(T)\mathbf{w}'' \quad \text{and} \quad \mathbf{z}' = P(T)\mathbf{z}'', \tag{4.4}$$

since then

$$P(T)\mathbf{u} = \mathbf{w}' + \mathbf{z}' = P(T)(\mathbf{w}'' + \mathbf{z}''),$$

and the admissibility of $W \oplus Z$ follows from Lemma 4.1. We have

$$P(x) = P_{\text{cond } W \oplus Z,T,\mathbf{u}}(x) \mid P_{\text{cond } W,T}(x) = P_{\min T,\mathbf{z}}(x)$$

according to Lemma 3.1, where second the equality holds according to (4.1). Let $Q(x)$ be such that $P_{\min T,\mathbf{z}}(x) = P(x)Q(x)$. Then, using equation (4.3), we obtain

$$Q(T)\mathbf{w}' + Q(T)\mathbf{z}' = Q(T)P(T)\mathbf{u} = P_{\text{cond } W,T}(T)\mathbf{u} \in W.$$

Thus, $Q(T)\mathbf{z}' \in W$; since we also have $Q(T)\mathbf{z}' \in Z$, we have $Q(T)\mathbf{z}' = \mathbf{0}$ by the already established assertion that $Z \cap W = \{\mathbf{0}\}$.

Let $R(x)$ a polynomials such that $\mathbf{z}' = R(T)\mathbf{z}$. Then we have

$$Q(T)R(T)\mathbf{z} = Q(T)\mathbf{z}' = \mathbf{0}.$$

Hence,

$$P_{\min\, T,\mathbf{z}}(x) = P(x)Q(x) \mid Q(x)R(x)$$

according to Lemma 3.1, and so $P(x) \mid R(x)$. Hence $R(x) = P(x)S(x)$ for some polynomial $S(x)$. Thus, writing $\mathbf{z}'' = S(T)\mathbf{z}$, we have $P(T)\mathbf{z}'' = P(T)S(T)\mathbf{z} = R(T)\mathbf{z} = \mathbf{z}'$. This establishes the second equation in (4.4). Therefore, according to (4.3) we have

$$P(T)(\mathbf{u} - \mathbf{z}'') = \mathbf{w}' \in W.$$

Since W is admissible, there is a $\mathbf{w}'' \in W$ such that $\mathbf{w}' = P(T)\mathbf{w}''$; i.e., the first equation in (4.4) is also established. □

Definition 4.3. Let V be a vector space over a field F and for i with $1 \le i \le n$ let W_i be a subspace of V. We write

$$\sum_{i=1}^{n} W_i = \left\{ \sum_{i=1}^{n} \mathbf{w}_i : \mathbf{w}_i \in W_i \quad \text{for} \quad 1 \le i \le n \right\}.$$

We say that the subspaces W_i are *independent* if

$$\sum_{i=1}^{n} \mathbf{w}_i = \mathbf{0}$$

holds for $\mathbf{w}_i \in W_i$ for each i with $1 \le i \le n$ only if $\mathbf{w}_i = \mathbf{0}$ for each i. If the spaces W_i are independent, we write

$$\bigoplus_{i=1}^{n} W_i \overset{def}{=} \sum_{i=1}^{n} W_i;$$

this notation is not used when the spaces are not independent.

We have

Lemma 4.4. *Let V be a vector space over a field F and for i with $1 \le i \le n$ let W_i be a subspace of V. The subspaces W_i are independent if and only if for each k with $1 < k \le n$ we have*

$$\left(\sum_{i=1}^{k-1} W_i \right) \cap W_k = \{\mathbf{0}\}.$$

Proof. Assume that the subspaces W_i are independent and

$$\mathbf{w} \in \left(\sum_{i=1}^{k-1} W_i \right) \cap W_k.$$

Then

$$\sum_{i=1}^{k-1} \mathbf{w}_i = \mathbf{w},$$

and so

$$\sum_{i=1}^{k-1} \mathbf{w}_i + (-1)\mathbf{w} = \mathbf{0}$$

for some $\mathbf{w}_i \in W_i$ for i with $1 \le i \le k-1$ and $\mathbf{w} \in W_k$. Hence $\mathbf{w} = \mathbf{w}_1 = \mathbf{w}_2 = \ldots = \mathbf{w}_{k-1} = \mathbf{0}$ by the definition of independence.

Assume, conversely, that the spaces W_i are not independent, and assume that

$$\sum_{i=1}^{n} \mathbf{w}_i = \mathbf{0}$$

where $\mathbf{w}_i \in W_i$ and not all $\mathbf{w}_i$ are equal to $\mathbf{0}$. Let k with $1 \le k \le n$ be the largest integer for which $\mathbf{w}_k \ne \mathbf{0}$. Then

$$\mathbf{w}_k = \sum_{i=1}^{k-1} (-1)\mathbf{w}_i \in \left(\sum_{i=1}^{k-1} W_i \right) \cap W_k$$

is a nonzero vector. □

Theorem 4.1 (Cyclic Decomposition Theorem)**.** *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, and let W be a subspace of V that is admissible for T. Then there is a positive integer n and there are vectors $\mathbf{z}_i \in V$ for i with $1 \le i \le n$ such that $P_{\min\ T,\mathbf{z}_1}(x) = P_{\mathrm{cond}\ W,T}(x)$,*

$$P_{\min\ T,\mathbf{z}_{i+1}}(x) \mid P_{\min\ T,\mathbf{z}_i}(x) \qquad (1 \le i < n), \tag{4.5}$$

and $V = W \oplus \left(\bigoplus_{i=1}^{n} Z(\mathbf{z}_i, T) \right)$.

Proof. We will derive the result by induction from Lemmas 4.2 and 4.3. Let $Z_0 = W$, let $k \ge 1$ be an integer, and assume for i with $1 \le i < k$ we have constructed the subspaces $Z_i = Z(\mathbf{z}_i, T)$ such that the subspaces Z_i for $0 \le i < k$ are independent. Let $W_i = \bigoplus_{j=0}^{i-1} Z_j$ (so, in particular, we have $W_0 = W$), and assume that $P_{\min\ T,z_i}(x) = P_{\mathrm{cond}\ W_i,T,z_i}(x) = P_{\mathrm{cond}\ W_i,T}(x)$ for i with $1 \le i \le k$ (cf. equation (4.1)).

Put $W_k = \bigoplus_{j=0}^{k-1} Z_j$, and assume W_k is admissible. If $W_k = V$, put $n = k-1$; otherwise, select $\mathbf{z}_k = \mathbf{z}$ as described in Lemma 4.2, and then use Lemma 4.3 to describe the properties of the cyclic subspace $Z(\mathbf{z}_k, T)$. The equation $P_{\min\ T,\mathbf{z}_{i+1}}(x) \mid P_{\min\ T,\mathbf{z}_i}(x)$ is satisfied for i with $1 \le i < k$, since $P_{\mathrm{cond}\ W_i,T}(T)\mathbf{z}_{i+1} \in W_i \subset W_{i+1}$, and so

$$P_{\min\ T,z_{i+1}}(x) = P_{\mathrm{cond}\ W_{i+1},T,z_{i+1}}(x) \mid P_{\mathrm{cond}\ W_i,T}(x) = P_{\min\ T,\mathbf{z}_i}(x)$$

by Lemma 3.1. This establishes equation (4.5), completing the proof. □

Most of the time we will use this theorem with $W = \{\mathbf{0}\}$; the general case, however, allows us to give a characterization of admissible subspaces.

Corollary 4.1. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, and let W be a subspace of V that is invariant for T. Then W is admissible if and only if there V has a subspace U invariant for T such that $V = W \oplus U$.*

Proof. **"Only if" part**. If W is admissible, then by Theorem 4.1 we have $V = W \oplus \left(\bigoplus_{i=1}^{n} Z(\mathbf{z}_i, T) \right)$, and so $V = V \oplus U$ with $U = \bigoplus_{i=1}^{n} Z(\mathbf{z}_i, T)$.

"If" part. If $V = W \oplus U$, where W and U are invariant for T, then, given any $\mathbf{v} \in V$, we have $\mathbf{v} = \mathbf{w} + \mathbf{u}$ with $\mathbf{w} \in W$ and $\mathbf{u} \in U$. Given a polynomial $P(x)$, we have $P(T)\mathbf{w} \in W$ and $P(T)\mathbf{u} \in U$ since W and U are invariant for T. Hence $P(T)\mathbf{v} \in W$ if and only if $P(T)\mathbf{u} = \mathbf{0}$, and then $P(T)\mathbf{v} = P(T)\mathbf{w}$. □

5 Decomposition of cyclic spaces

If V is a vector space over a field F, the subspaces $\{\mathbf{0}\}$ and V will be called *trivial* subspaces, and all other subspaces will be called *nontrivial.*

Lemma 5.1. *Let $P(x)$ be an irreducible polynomial over the field F, let n be a positive integer, let V be a finite-dimensional vector space, let $T : V \to V$ be a linear transformation, let $\mathbf{z} \in V$ such that $V = Z(\mathbf{z}, T)$. Assume that $P_{\min\, V,T}(x) = \big(P(x)\big)^n$. Then there are no nontrivial subspaces U and W invariant for T such that $V = U \oplus W$, $U \neq \{\mathbf{0}\}$, and $W \neq \{\mathbf{0}\}$.*

Proof. Let m be the degree of $P(x)$; then the degree of the minimal polynomial $\big(P(x)\big)^n$ of V is mn. Hence the system $(T^k\mathbf{z} : 0 \le k < mn)$ is linearly independent; otherwise the minimal polynomial of $V = Z(\mathbf{z}, T)$ would have degree lower than mn. On the other hand, the system $(T^k\mathbf{z} : 0 \le k \le mn)$ is linearly dependent, since $\big((P(T)\big)^n\mathbf{z} = 0$. Thus, the dimension of V is exactly mn. If subspaces U and V as described exist, then they each have dimension less than mn. Thus, there are be positive integers $k < n$ and $l < n$ such that $P_{\min\, U,T}(x) = \big(P(x)\big)^k$ and $P_{\min\, W,T}(x) = \big(P(x)\big)^l$ (the minimal polynomial of a subspace divides the minimal polynomial of the whole space, according to Lemma 3.1).

To show, for example, that $k < m$, use Theorem 3.1 for minimal polynomials, i.e., with $W = \{0\}$. This says that there is a $\mathbf{u} \in U$ such that $P_{\min\, T,\mathbf{u}}(x) = P_{\min\, T}(x)$. Since the degree of this polynomial is km, the system $(T^i\mathbf{u} : 0 \le i < km)$ is linearly independent, so $\dim U \ge km$. Since U is proper a subspace of V, and the latter has dimension nm, we must have $k < n$.

Since $V = U \oplus V$, we have $P_{\min\, T}(x) = \big(P(x)\big)^{\max\{k,l\}}$. Since the minimal polynomial on the left-hand side is $\big(P(x)\big)^n$, and, as we saw above, $k < n$ and $l < n$, this is a contradiction. □

Lemma 5.2. *Let $P(x)$ and $Q(x)$ be nonconstant polynomials over the field F such that $\gcd\big(P(x), Q(x)\big) = 1$. Let V be a finite-dimensional vector space, let $T : V \to V$ be a linear transformation, let $\mathbf{z} \in V$ such that $V = Z(\mathbf{z}, T)$, and assume $P_{\min\, T,V}(x) = P(x)Q(x)$. Then*

$$V = Z(P(T)\mathbf{z}, T) \oplus Z(Q(T)\mathbf{z}, T).$$

Proof. To show that $Z(P(T)\mathbf{z}, T) \cap Z(Q(T)\mathbf{z}, T) = \{\mathbf{0}\}$, assume

$$\mathbf{v} \in Z(P(T)\mathbf{z}, T) \cap Z(Q(T)\mathbf{z}, T)$$

Assuming $\mathbf{v} \neq 0$, there are nonzero polynomials $R(x)$ and $S(x)$ such that $\mathbf{v} = R(T)P(T)\mathbf{z} = S(T)Q(T)\mathbf{z}$ and $\deg R(x) < \deg Q(x)$ and $\deg S(x) < \deg P(x)$.[5.1] Then

$$\big(R(T)P(T) - S(T)Q(T)\big)\mathbf{z} = \mathbf{0}. \tag{5.1}$$

According to Lemma 1.3 we have $P(x) \nmid S(x)Q(x)$; hence the polynomial $R(x)P(x) - S(x)Q(x)$ is not divisible by $P(x)$, and so, *a fortiori*,[5.2] it is not zero. Since it is a of degree less than that of $P(x)Q(x)$, the minimal polynomial of V, equation (5.1) cannot hold, showing that $Z(P(T)\mathbf{z}, T) \cap Z(Q(T)\mathbf{z}, T) = \{\mathbf{0}\}$.

To show that

$$V = Z(P(T)\mathbf{z}, T) + Z(Q(T)\mathbf{z}, T),$$

note that the subspace on the rigth is a cyclic subspace of V; hence to show this equality, it is enough to observe that this subspac contains $\mathbf{z}$. Indeed, there are polynomials $M(x)$ and $N(x)$ such that $M(x)P(x) + N(x)Q(x) = 1$ by Lemma 1.2, and so $M(T)P(T)\mathbf{z} + N(T)Q(T)\mathbf{z} = \mathbf{z}$. □

[5.1] Indeed, if, for example, $\deg R(x) \ge \deg Q(x)$, then we can replace $R(x)$ with the remainder $R_1(x)$ when $Q(x)$ is divided into $R(x)$, since $P_{\min\, T,V}(x) = P(x)Q(x)$, and so $Q(T)P(T)\mathbf{z} = 0$.

[5.2] Latin for "with stronger reason," or, loosely, "even more so."

An immediate consequence of the last lemma is

Corollary 5.1. *Let n be a positive integer, and Let $P_i(x)$ for i with $1 \le i \le n$ be powers of polynomials irreducible over the field F such that $\gcd(P_i(x), P_j(x)) = 1$ if $1 \le i < j \le n$. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let $\mathbf{z} \in V$ such that $V = Z(\mathbf{z}, T)$. Assume $P_{\min\, V,T}(x) = \prod_{i=1}^{n} P_i(x)$. For each i with $1 \le i \le n$ let*

$$Q_i(x) = \prod_{\substack{j=1\\ j\ne i}}^{n} P_j(x).$$

Then

$$V = \bigoplus_{i=1}^{n} Z(Q_i(T)\mathbf{z}, T).$$

6 The characteristic polynomial

Definition 6.1. Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, and let $T : V \to V$ be a linear transformation. Let $k \ge 0$ be the smallest integer such that $\big(P(x)\big)^{k+1} \nmid P_{\min\, V,T}(x)$. The index $\operatorname{ind}(P(x), V, T)$ of $P(x)$ in V for T is defined as

$$\operatorname{ind}(P(x), V, T) = \frac{\dim V - \dim\left\{\big(P(x)\big)^{k}\mathbf{v} : \mathbf{v} \in V\right\}}{\deg P(x)}.$$

It is not necessary to choose k to be the smallest integer for which $\big(P(x)\big)^{k+1} \nmid P_{\min\, V,T}(x)$; any nonnegative integer satisfying this equation gives the same value in the above formula for $\operatorname{ind}\big(P(x), V, T\big)$; this is shown by the following simple

Lemma 6.1. *Let $P(x)$ be a polynomial over a field F, let V be a finite-dimensional vector space over F, let $T : V \to V$ be a linear transformation, and assume that $\gcd\big(P(x), P_{\min\, V,T}(x)\big) = 1$. Then $V = \{P(T)\mathbf{v} : \mathbf{v} \in V\}$.*

Proof. By Lemma 1.2, there are polynomials $M(x)$ and $N(x)$ such that

$$1 = M(x)P(x) + N(x)P_{\min\, V,T}(x).$$

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

$$\begin{aligned}\mathbf{v} &= \big(M(T)P(T) + N(T)P_{\min\, V,T}(T)\big)\mathbf{v} = M(T)P(T)\mathbf{v} + N(T)P_{\min\, V,T}(T)\mathbf{v}\\ &= M(T)P(T)\mathbf{v} = P(T)M(T)\mathbf{v}\end{aligned}$$

the third equation holds since $P_{\min\, V,T}(T)\mathbf{v} = \mathbf{0}$. □

Corollary 6.1. *Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, and let $T : V \to V$ be a linear transformation. Let k and l be integers with $l > k \ge 0$ such that $\big(P(x)\big)^{k+1} \nmid P_{\min\, V,T}(x)$. Then*

$$\left\{\big(P(x)\big)^{k}\mathbf{v} : \mathbf{v} \in V\right\} = \left\{\big(P(x)\big)^{l}\mathbf{v} : \mathbf{v} \in V\right\}.$$

Proof. Let $m \geq 0$ be an integer such that $P_{\min\ V,T}(x) = \big(P(x)\big)^m Q(x)$ and $P(x) \nmid Q(x)$, and write

$$U = \left\{ \big(P(x)\big)^m \mathbf{v} : \mathbf{v} \in V \right\}.$$

Then $P_{\min\ U,T}(x) = Q(x)$, and so $U = \left\{ \big(P(x)\big)^i \mathbf{v} : \mathbf{v} \in U \right\}$ for any integer $i \geq 0$, according to Lemma 6.1. □

Lemma 6.2. *Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, let $T : V \to V$ be a linear transformation, and let $V = U \oplus W$, and assume U and W are invariant for T. Then*

$$\operatorname{ind}(P(x), V, T) = \operatorname{ind}(P(x), U, T) + \operatorname{ind}(P(x), W, T).$$

Proof. Let $k \geq 0$ be an integer such that $\big(P(x)\big)^{k+1} \nmid P_{\min\ V,T}(x)$. We have

$$\left\{ \big(P(T)\big)^k \mathbf{v} : \mathbf{v} \in V \right\} = \left\{ \big(P(T)\big)^k \mathbf{v} : \mathbf{v} \in U \right\} \oplus \left\{ \big(P(T)\big)^k \mathbf{v} : \mathbf{v} \in W \right\}.$$

As U and V are invariant for T, we further have $\left\{ \big(P(T)\big)^k \mathbf{v} : \mathbf{v} \in U \right\} \subset U$ and $\left\{ \big(P(T)\big)^k \mathbf{v} : \mathbf{v} \in W \right\} \subset W$, and so the assertion follows by Corollary 6.1. □

Lemma 6.3. *Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, and let $T : V \to V$ be a linear transformation. Then $\operatorname{ind}(P(x), V, T)$ is a nonnegative integer.*

Proof. By Theorem 4.1 with $W = \{\mathbf{0}\}$ and Corollary 5.1 we may assume that $V = Z(\mathbf{z}, T)$ and $P_{\min\ V,T}(x) = \big(P(x)\big)^m$, where $P(x)$ is irreducible over F and m is a positive integer. Then $\dim V = m \cdot \deg P(x)$ and $\dim \left\{ \big(P(x)\big)^m \mathbf{v} : \mathbf{v} \in V \right\} = 0$; hence $\operatorname{ind}(P(x), V, T) = m$. □

Definition 6.2. Let V be a finite-dimensional vector space over F, and let $T : V \to V$ be a linear transformation. We define $P_{\text{char}\ V,T}(x)$ as

$$\prod_{P(x)} (P(x))^{\operatorname{ind}(P(x),V,T)}, \tag{6.1}$$

where $P(x)$ runs over all polynomials irreducible over F such that $\operatorname{ind}(P(x), V, T) > 0$.

Lemma 6.4. *Let V be a finite-dimensional vector space over F, let $T : V \to V$ be a linear transformation, and assume $V = Z(\mathbf{z}, T)$ for some $\mathbf{z} \in V$. Assume, further, that $P_{\min\ V,T}(x)$ is a power of a polynomial irreducible over F. Then $P_{\text{char}\ V,T}(x) = P_{\min\ V,T}(x)$.*

Proof. The assertion is implicit in the proof of Lemma 6.3. □

Corollary 6.2. *Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, let $T : V \to V$ be a linear transformation, and let $V = U \oplus W$, and assume U and W are invariant for T. Then*

$$P_{\text{char}\ V,T}(x) = P_{\text{char}\ T,U}(x) \cdot P_{\text{char}\ T,W}(x)$$

Proof. The result is an immediate consequence of Lemma 6.2. □

Corollary 6.3. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation. Let n, $\mathbf{z}_i$ be as described in Theorem 4.1 with $W = \{\mathbf{0}\}$. Then*

$$P_{\text{char }V,T}(x) = \prod_{i=1}^{n} P_{\min\ \mathbf{T},\mathbf{z}_i}(x). \tag{6.2}$$

As a consequence

$$\deg P_{\text{char }V,T}(x) = \dim V. \tag{6.3}$$

Proof. By Corollary 6.2, we may assume that $V = Z(\mathbf{z}, T)$, and then, also using Corollary 5.1, we may assume that $P_{\min\ V,T}(x) = \big(P(x)\big)^m$ for some polynomial $P(x)$ irreducible over F and for some positive integer m. As pointed out in the proof of Lemma 6.3, in this case $\operatorname{ind}(P(x), V, T) = m$, showing that $P_{\text{char }V,T}(x) = \big(P(x)\big)^{\operatorname{ind}(P(x),V,T)}$ in this case. We also have $P_{\min\ V,T,\mathbf{z}}(x) = P_{\min\ V,T}(x) = \big(P(x)\big)^m$ in this case, establishing equation (6.2). As for equation (6.3), it is enough to note that $\deg P_{\min\ V,T,\mathbf{z}}(x) = \dim Z(\mathbf{z}, T)$. □

Corollary 6.4. *The integer n and the polynomials $P_{\min\ T,\mathbf{z}_i}(x)$ in Theorem 4.1 with $W = \{\mathbf{0}\}$ are uniquely determined.*

Proof. We will use induction on $\dim V$. According to Theorem 4.1 with $W = \{\mathbf{0}\}$ we have $V = \bigoplus_{i=1}^{\infty} Z(\mathbf{z}_i, T)$, where $\mathbf{z}_i = \mathbf{0}$ for except for finitely many values of i. The value of n is the largest value of i for which $\mathbf{z}_i \neq \mathbf{0}$, but it is technically advantageous to avoid an explicit reference to the integer n in that theorem. We want to show that the polynomials $P_{\min\ T,\mathbf{z}_i}(x)$ are uniquely determined (if $\mathbf{z}_i = \mathbf{0}$, we take $Z(\mathbf{z}_i, T) = \{\mathbf{0}\}$ and $P_{\min\ T,\mathbf{z}_i}(x) = 1$); it is not claimed that the vectors $\mathbf{z}_i$ or the subspaces $Z(\mathbf{z}_i, T)$ themselves are uniquely determined. Let $P(x)$ be an irreducible factor of $P_{\min\ T}(x)$, and let

$$U = \bigoplus_{i=1}^{\infty} Z(\mathbf{P}(T)\mathbf{z}_i, T).$$

If we have $P(x) \mid P_{\min\ T,\mathbf{z}_i}(x)$, then $P_{\min\ T,\mathbf{z}_i}(x) = P(x) P_{\min\ T,\mathbf{P}(T)z_i}(x)$; if $P(x) \nmid P_{\min\ T,\mathbf{z}_i}(x)$, then $P_{\min\ T,\mathbf{z}_i}(x) = P_{\min\ T,\mathbf{P}(T)z_i}(x)$ according to Lemma 6.1. By equations (6.1) and (6.2) it then follows that $P(x) \mid P_{\min\ T,\mathbf{z}_i}(x)$ if and only if

$$1 \le i \le \operatorname{ind}\big(P(x), V, T\big) - \operatorname{ind}\big(P(x), U, T\big).$$

Indeed, writing

$$k = \operatorname{ind}\big(P(x), V, T\big) - \operatorname{ind}\big(P(x), U, T\big),$$

we have

$$\operatorname{ind} Z(\mathbf{z}_i, T) = \operatorname{ind} Z\big(P(T)\mathbf{z}_i, T\big) + 1$$

if $1 \le i \le k$, and

$$\operatorname{ind} Z(\mathbf{z}_i, T) = \operatorname{ind} Z\big(P(T)\mathbf{z}_i, T\big)$$

if $i > k$. Since the quantities $\operatorname{ind}\big(P(x), V, T\big)$, $\operatorname{ind}\big(P(x), U, T\big)$, and, by the induction hypothesis, the polynomials $P_{\min\ T,\mathbf{P}(T)\mathbf{z}_i}(x)$ are uniquely determined, it follows that the polynomials $P_{\min\ T,\mathbf{z}_i}(x)$ are also uniquely determined. □

Theorem 6.1 (Cayley–Hamilton theorem)**.** *Let $P(x)$ be an irreducible polynomial over a field F, let V be a finite-dimensional vector space over F, let $T : V \to V$ be a linear transformation. Then $P_{\text{char }V,T}(T) = 0$.*

Proof. The result is a consequence of the fact that the minimal polynomial of T divides its characteristic polynomial, according to equation (6.2) in Corollary 6.3 since $P_{\min T,z_1}(x) = P_{\min T}(x)$ in that equation. □

The Cayley–Hamilton theorem is a fairly simple observation given the way we arranged the material here. However, traditionally the characteristic polynomial of a matrix is defined by equation (8.8) below, and the characteristic polynomial of a linear transformation is only defined as an afterthought, since matrices can be used to describe linear transformations. With that definition, the proof of the Cayley–Hamilton theorem is not at all simple.

7 Decomposition into irreducible invariant subspaces

The uniqueness of the minimal polynomials associated with the decomposition described in Theorem 4.1 with $W = \{\mathbf{0}\}$ was shown by Corollary 6.4. The decomposition of a vector space V described by a combination of Theorem 4.1 with $W = \{\mathbf{0}\}$ and Corollary 5.1 is also unique, except for the order in which the subspaces are listed. This is described by the following

Theorem 7.1. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation. Then there is a nonnegative integer n and there are subspaces Z_i for i with $1 \le i \le n$ such that V can be decomposed as*

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

Further, for each i, $Z_i \neq \{\mathbf{0}\}$ is invariant for T, there is an $\mathbf{z}_i \in Z_i$ such that $Z_i = Z(\mathbf{z}_i, T)$, and $P_{\min Z_i,T}(x)$ is a power of a polynomial that is irreducible over F. This decomposition is unique except for the order in which the subspaces Z_i are listed. Furthermore, we have

$$P_{\mathrm{char}\ V,T}(x) = \prod_{i=1}^{n} P_{\min Z_i,T}(x). \tag{7.1}$$

The case $n = 0$ was allowed, since we did not exclude the possibility that $V = \{\mathbf{0}\}$.

Proof. The existence of such a decomposition is an immediate consequence of Theorem 4.1 with $W = \{\mathbf{0}\}$ and Corollary 5.1. As for the uniqueness, if $V = \bigoplus_{j=1}^{m} U_j$ is another such decomposition, then, for any i and j with $1 \le i \le n$ and $1 \le j \le m$, writing,

$$U'_j = \bigoplus_{\substack{k=1\\k\neq j}}^{m} U_k,$$

we have $Z_i = (Z_i \cap U_j) \oplus (Z_i \cap U'_j)$. Since Z_i cannot be properly split as the direct sum of invariant subspaces according to by Lemma 5.1, we must have $Z_i \subset U_j$ or $Z_i \cap U_j = \{\mathbf{0}\}$. A similar argument with reversing the roles of Z_i and U_j shows that either $U_j \subset Z_i$ or $U_j \cap Z_i = \{\mathbf{0}\}$. Hence the uniqueness of the decomposition follows. As for equation (7.1), this follows from Corollary 6.2 and Lemma 6.4 □

7.1 Algebraically closed fields

If F is an algebraically closed field, then the only irreducible polynomials over F are linear polynomials. Hence, Theorem 7.1 has the following consequence.

Corollary 7.1 (Jordan decomposition theorem)**.** *Let F be an algebraically closed field, and let V be a finite-dimensional vector space over F. let $T : V \to V$ be a linear transformation. Then there is a nonnegative integer n, and for each i with $1 \le i \le n$ there is a subspace Z_i, a vector $\mathbf{z}_i \in Z_i$ such that $Z_i = Z(\mathbf{z}_i, F)$, a scalar $\lambda_i \in F$, and a positive integer m such that $P_{\min\ Z_i,T}(x) = (x - \lambda_i)^{m_i}$, and*

$$V = \bigoplus_{i=1}^{n} Z_i,$$

This decomposition is unique except for the order in which the subspaces Z_i are listed. Furthermore, we have

$$P_{\text{char } V,T}(x) = \prod_{i=1}^{n} (x - \lambda_i)^{m_i}. \tag{7.2}$$

The scalars λ_i are called the eigenvalues of the linear transformation T. The *multiplicity* of an eigenvalue λ is the exponent of the factor $(x - \lambda)$ in $P_{\text{char } V,T}(x)$. Since we may have $\lambda = \lambda_i$ for several different values of i in equation (7.2), this multiplicity is the sum of all m_i's for which $\lambda_i = \lambda$. The subspaces Z_i in the above decomposition are called *Jordan subspaces* for T. That is, a Jordan subspace is a cyclic subspace $Z = Z(\mathbf{z}, V)$ such that $P_{\min\ Z,T}(x) = (x - \lambda)^m$ for some $\lambda \in F$ and some integer m.

8 Matrices

8.1 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,

$$\mathcal{R}_{\mathcal{W}\mathcal{V}}T = Q^{-1}(\mathcal{R}_{\mathcal{Y}\mathcal{X}}T)P \tag{8.1}$$

(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

$$\mathcal{R}_{\mathcal{W}\mathcal{U}}TS = (\mathcal{R}_{\mathcal{W}\mathcal{V}}T)(\mathcal{R}_{\mathcal{V}\mathcal{U}}S), \tag{8.2}$$

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].

8.2 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

$$A = \mathcal{R}_{\mathcal{E}\mathcal{E}}T. \tag{8.3}$$

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 (8.1), for the above A and T we have

$$\mathcal{R}_{\mathcal{X}\mathcal{X}}T = P^{-1}(\mathcal{R}_{\mathcal{E}\mathcal{E}}T)P = P^{-1}AP. \tag{8.4}$$

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.*

8.3 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

$$\mathcal{R}_{(\mathcal{X},\mathcal{Y})(\mathcal{X},\mathcal{Y})}T = A \oplus B. \tag{8.5}$$

8.4 The companion matrix of a polynomial

Definition 8.1. Let n be a positive integer, and let

$$P(x) = x^n + \sum_{k=0}^{n-1} a_k x^k = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

be a monic polynomial. The $n \times n$ matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

is called the *companion matrix* of the polynomial $P(x)$ (for $n = 1$ take $A = -a_0$).

Writing $\det A$ for the square matrix A, we have the following

Lemma 8.1. *Let n be a positive integer, let $P(x)$ be a monic polynomial of degree n, and let A be the companion matrix of $P(x)$. Then, writing I_n for the $n \times n$ identity matrix, we have $\det(xI_n - A) = P(x)$.*

Proof. We have to show that the determinant

$$\det(xI_n - A) = \begin{vmatrix} x & 0 & 0 & \ldots & 0 & a_0 \\ -1 & x & 0 & \ldots & 0 & a_1 \\ 0 & -1 & x & \ldots & 0 & a_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & x + a_{n-1} \end{vmatrix}$$

equals $P(x)$. To this end we will expand this determinant by its first row. We obtain that this determinant equals

$$x \begin{vmatrix} x & 0 & 0 & \ldots & 0 & a_1 \\ -1 & x & 0 & \ldots & 0 & a_2 \\ 0 & -1 & x & \ldots & 0 & a_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & x & a_{n-2} \\ 0 & 0 & 0 & \ldots & -1 & x + a_{n-1} \end{vmatrix} + (-1)^{n+1} a_0 \begin{vmatrix} -1 & x & 0 & \ldots & 0 & 0 \\ 0 & -1 & x & \ldots & 0 & 0 \\ 0 & 0 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & -1 & x \\ 0 & 0 & 0 & \ldots & 0 & -1 \end{vmatrix}.$$

The first determinant is $\det(xI_{n-1} - A_1)$, where I_{n-1} is the $(n-1) \times (n-1)$ identity matrix, and matrix A_1 is $(n-1) \times (n-1)$ matrix obtained from the matrix A by deleting its first row and its first column. So we can make the induction hypothesis that this determinant equals $P_1(x)$, where

$$P_1(x) = x^{n-1} + \sum_{k=0}^{n-2} a_{k+1} x^k = x^{n-1} + a_{n-1}x^{n-2} + \ldots + a_2 x + a_1.$$

The second determinant is $(-1)^{n-1}$, since it is the determinant of a triangular matrix, so the value of this determinant is just the product of this diagonal elements. Hence it follows that indeed

$$\det(xI_n - A) = x \cdot P_1(x) + a_0 = P(x).$$

To complete the proof by induction, one needs to check that the statement is true for $n = 1$. In case $n = 1$ we have $P(x) = x + a_0$, $A = -a_0$ (a 1×1 matrix is taken to be a scalar), and $\det(xI_1 - A) = x - A = x + a_0$, showing that the assertion is true for $n = 1$. □

Lemma 8.2. *Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation, let $\mathbf{z} \in V$ be a nonzero vector, assume $V = Z(\mathbf{z}, T)$, and let $n = \dim V$. Then $\mathcal{X} = (T^k\mathbf{z} : 0 \le k < n)$ is a basis of V, and $\mathcal{R}_{\mathcal{X}\mathcal{X}}T$ is the companion matrix of $P_{\min\ V,T}(x)$.*

Proof. The degree of $P_{\min\ V,T}(x)$ is n. Writing $P_{\min\ V,T}(x) = x^n + \sum_{k=0}^{n-1} a_k x^k$, $\mathbf{T}^k z = \mathbf{x}_k$ for k with $0 \le k < n$ the assertion is just a restatement of the equation $P_{\min\ V,T}(T)\mathbf{z} = \mathbf{0}$ in the form $T\mathbf{x}_k = \mathbf{x}_{k+1}$ for k with $0 \le k \le n-2$ and $T\mathbf{x}_{n-1} = \sum_{k=0}^{n-1}(-a_k)\mathbf{x}_k$. □

8.5 The characteristic polynomial of a matrix

Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation. Use Theorem 4.1 with $W = \{\mathbf{0}\}$ to represent V as a direct sum $V = \bigoplus_{i=1}^n Z(\mathbf{z}_i, T)$, and in each component take a basis $\mathcal{X}_i = (T^k\mathbf{z}_i : 0 \le k < n_i)$. Then, taking $A_i = \mathcal{R}_{\mathcal{X}_i\mathcal{X}_i}\big(T \restriction Z(\mathbf{z}_i, T)\big)$, where $T \restriction M$ denotes the restriction of T to the set $M \subset V$, the matrix A_i is the companion matrix of the polynomial $P_{\min\ T,\mathbf{z}_i}(x)$. Hence $P_{\min\ T,\mathbf{z}_i}(x) = I_{n_i}x - A_i$ according to Lemmas 8.1 and 8.2. Taking $\mathcal{X} = (\mathcal{X}_i : 1 \le k \le n)$ as the basis of V, we have $\mathcal{R}_{\mathcal{X}\mathcal{X}}T = \bigoplus A_i$ according to Subsection 8.3. Writing A for this matrix, we have

$$P_{\mathrm{char}\ V,T}(x) = \prod_{i=1}^n \det(xI_{n_i} - A_i) = \det(xI_n - A), \tag{8.6}$$

where the second equation holds since $\bigoplus_{i=1}^n (xI_{n_i} - A_i) = xI_n - A$, and $\det\big(\bigoplus_{i=1}^n B_i\big) = \prod_{i=1}^n \det B_i$ for any square matrices B_i.

If we choose $\mathcal{Y}$ another basis of V and $C = \mathcal{R}_{\mathcal{Y}\mathcal{Y}}$, then C is similar to A according to Subsection 8.2, i.e., $C = P^{-1}AP$ for some nonsingular $n \times n$ matrix P. Then

$$\begin{aligned}\det(xI_n - C) &= \det(I_n)\det(xI_n - C) = \det(PP^{-1})\det(xI_n - C)\\ &= \det(P)\det(P^{-1})\det(xI_n - C) = \det(P)\det(xI_n - C)\det(P^{-1})\\ &= \det\big(P(xI_n - C)P^{-1}\big) = \det(xI_n - PCP^{-1}) = \det(xI_n - A) = P_{\mathrm{char}\ V,T}(x),\end{aligned} \tag{8.7}$$

where the last equation holds according to (8.6). Therefore, for an arbitrary square matrix B, the characteristic polynomial of B is defined by the equation

$$P_{\mathrm{char}\ B}(x) = \det(xI - B), \tag{8.8}$$

where I is the identity matrix of the appropriate size.

8.6 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 $\lambda I + N$, where λ is a scalar, I is the $n \times n$ identity matrix, and N is an auxiliary unit matrix.

That is, a Jordan block is a matrix of form

$$\lambda I + N = \begin{pmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \lambda & \dots & 0 & 0 \\ \multicolumn{7}{c}{\dotfill} \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix} \tag{8.9}$$

Theorem 8.1. *Let n be a positive integer, let V be an n-dimensional vector space, $T : V \to V$ a linear transformation, let $\mathbf{z} \in V$ and assume that $V = Z(\mathbf{z}, T)$. Assume, further, that $P_{\min V,T}(x) = (x - \lambda)^n$ for some $\lambda \in F$ and some positive integer n. Writing $\mathbf{x}_k = (T - \lambda)^{n-k}\mathbf{z}$ 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.

8.7 The Jordan canonical form of a matrix

Assume F is an algebraically closed field. Let A be an arbitrary $n \times n$ matrix over F, and consider A as a linear transformation of $F_{n,1}$ into $F_{n,1}$ as in Subsection 8.2. According to Corollary 7.1, $F_{n,1}$ 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 (8.4), we proved the following

Theorem 8.2. *Every square matrix A over an algebraically closed field 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 (the definition of the multiplicity of an eigenvalue was given right after Corollary 7.1).

8.8 The characteristic polynomial of a matrix for algebraically closed fields

In Subsection 8.5, we discussed the characteristic polynomial of a matrix. This discussion is somewhat simplified when the underlying field F is algebraically closed, in which case the Jordan canonical form of a matrix is available.

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 $xI - P^{-1}AP$, where I is the $n \times n$ identity matrix, are $x - \lambda_i$ 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 linear transformation $T : V \to V$ over an n-dimensional vector space V represented by the matrix A according to Corollary 7.1.

Since $xI - P^{-1}AP$ is an upper triangular matrix, its determinant is equal to the product of its diagonal elements. That is

$$P_{\text{char }V,T}(x) = \det(xI - P^{-1}AP) = \det(xI - A)$$

as shown by the calculation in (8.7). That is the characteristic polynomial of A is equal to the characteristic polynomial T.

8.9 The invariance of the minimal polynomial of a matrix under field extensions

Let n be a positive integer, let F be a field, and let A be an $n \times n$ matrix with entries in F, that is $A \in F_{n\times n}$. If F' is an extension of the field F, then A can also be considered as a matrix with entries in F'. Equation (8.8) shows that the characteristic polynomial of A remains the same whether we consider A as a matrix in $F_{n\times n}$ or as a matrix in $F'_{n\times n}$. A similar assertion about the minimal polynomial is not immediately clear. For example, if A is a square matrix with rational entries, and the minimal polynomial of A considered as a matrix with entries in the field of the rational numbers, the question is whether we get a different minimal polynomial if we consider A as a matrix with entries in the field of the complex numbers.

To express this possible dependence on the field, write $P_{\min A,F}(x)$ for the minimal polynomial of the square matrix A when A is considered as a matrix with entries in F, and $P_{\min A,\mathbf{z},F}(x)$ for the minimal polynomial for the square matrix A of the vector $\mathbf{z}$ of the appropriate size when they are considered as having entries in the field F. The following theorem says that taking a field extension does not change the minimal polynomials.

Theorem 8.3. *Let F be a field and let F' be an extension of F. Let n be a positive integer, and let $A \in F_{n\times n}$ and $\mathbf{z} \in F_{n\times 1}$. Then $P_{\min A,F}(x) = P_{\min A,F'}(x)$ and $P_{\min A,\mathbf{z},F}(x) = P_{\min A,\mathbf{z},F'}(x)$.*

For the proof, we need the following

Lemma 8.3. *Let F be a field and let F' be an extension of F, and let m and r be a positive integers. Let $S = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r)$ be system of m-dimensional column vectors with entries in F. Then S is a linearly independent system in $F_{m\times 1}$ if and only if S is linearly independent in $F'_{m\times 1}$.*

Proof. Let B be the matrix $(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r)$; that is the columns of B are the same as the vectors in the system S. Then the assertion is that the rank of B is r in the field F if and only if its rank is r in the field F'. However, the rank of B is the same as the rank of the *row-echelon form* of B ([1, §2.3, p. 57]; some other authors use the term *reduced row echelon form* for the same matrix), and the row echelon form is uniquely determined; that is, the row echelon form of B in F is the same as its row echelon form in F'; the rank of the row echelon form is the number of its leading entries, so it is the same in F as in F'. □

The $n\times n$ matrices over a field F form an n^2-dimensional vector space over F that is isomorphic to the space of n^2-dimensional column vectors; the isomorphism can be obtained by simply rearranging the entries of a matrix into a column format. Hence Lemma 8.3 is equally applicable to matrices over F and F'.

Proof of Theorem 8.3. Let m be the degree of $P_{\min A,F}(x)$. This implies that the equation

$$\sum_{k=0}^{m-1} b_k A^k = 0$$

cannot hold for elements b_0, b_1, …, b_{m-1} of F unless $b_0 = b_1 = \ldots = b_{m-1} = 0$; in other words, the system $(I,\ A,\ A^2,\ldots,\ A^{m-1})$ is linearly independent, with the matrices considered as elements of an n^2 dimensional vector space over F. According to Lemma 8.3, this holds exactly if this system is linearly independent when the matrices considered as elements of an n^2 dimensional vector space over F'. This means that the degree of $P_{\min\ A,F'}(x)$ is at least m; so its degree is exactly equal to m, since it cannot be higher than the degree of $P_{\min\ A,F}(x)$.

Let

$$P_{\min\ A,F}(x) = \sum_{k=0}^{m} a_k x^k \qquad (a_m = 1)$$

Then

$$\sum_{k=0}^{m} a_k A^k = 0;$$

that is, the system $(I,\ A,\ A^2,\ldots,\ A^m)$ is linearly dependent, with the matrices considered as elements of an n^2 dimensional vector space over F. According to Lemma 8.3, this holds exactly if this system is linearly dependent when the matrices considered as elements of an n^2 dimensional vector space over F'. The coefficients a_0, a_1, …, a_m are uniquely determined. Indeed, if we also have

$$\sum_{k=0}^{m} a'_k A^k = 0 \qquad (a'_m = 1)$$

with coefficients a'_0, a'_1, …, a'_m in F', then we have

$$\sum_{k=0}^{m-1} (a_k - a'_k) A^k = 0,$$

and so we must have $a_0 = a'_0$, $a_1 = a'_1$, …, $a_{m-1} = a'_{m-1}$, since the system $(I,\ A,\ A^2,\ldots,\ A^{m-1})$ is linearly independent when considered over F'.

A completely analogous argument involving the linear dependence or independence over F or F' of the system $(\mathbf{z},\ A\mathbf{z},\ A^2\mathbf{z},\ldots,\ A^s\mathbf{z})$ shows the equality $P_{\min\ A,\mathbf{z},F}(x) = P_{\min\ A,\mathbf{z},F'}(x)$. □

If V is a finite dimensional vector space over a field F, and if F' is a field extension of F, it is not immediately clear what is meant by considering the vectors of V over the field F'. A meaning to this can be given really only by taking a fixed basis of V and considering the representation of the vectors in V in this basis by column vectors with entries in F, and then considering those entries as being in F'. For this reason, consideration of the dependence on the field of the minimal polynomial of a linear transformation on an abstract vector space as opposed to the minimal polynomial of a matrix does not seem to have much purpose.

9 Functions of linear transformations

Aside from addition and multiplication and inverse, one may want to define other operations for linear transformations. In doing so, one may follow the same strategy that was used in defining functions for real or complex numbers. For example, when defining the exponential function e^x for complex exponent, one may start with the formula

$$e^x = \lim_{n\to+\infty} \left(1 + \frac{x}{n}\right)^n,$$

valid for all real x, and requiring that this formula be valid also for complex values of x. This allows one to extend the function e^x to all complex exponents.

In general, for linear transformations, one can use addition, multiplication, and limiting processes to extend functions to matrices. The simplest limiting process is power series. For example, in extending the function e^x to complex exponents, instead of the approach outlined one can also use the power series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{9.1}$$

convergent for all complex values of x. In fact, this method is faster than the one outlined above, though it has less intuitive appeal. The reason for this is that while there is an intuitive connection between e^x and the right-hand side of the formula above, there is no such intuitive connection between e^x and the power series described in the second method.

When one deals with spaces of column vectors over the field $\mathbb{R}$ of real numbers or over the field of $\mathbb{C}$ of complex numbers, limiting processes are well defined, so they can be used in extending functions from $\mathbb{R}$ or $\mathbb{C}$ to linear transformations. When using power series, an immediate problem is that power series often have a limited range of convergence. For example, one can imagine a situation when $V = U \oplus W$ is a vector space over the field $\mathbb{R}$ or $\mathbb{C}$, $T : V \to V$ is a linear transformation, and U and W are invariant for T, but a certain power series has different convergence properties for T restricted to U than for T restricted to W. This may make a direct definition of a certain function f for T unusable, since somewhat different processes are needed to define $f(T \restriction U)$ and $f(T \restriction W)$. Now, for any polynomial $P(x)$ and for any vector $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{v} \in V$ we have $P(T)\mathbf{v} = P(T \restriction U)\mathbf{u} + P(T \restriction W)\mathbf{w}$, it is reasonable stipulate that such behavior is inherited by the function f. That is, we will require that

$$f(T)\mathbf{v} = f(T \restriction U)\mathbf{u} + f(T \restriction W)\mathbf{w}.$$

This allows one to take advantage of the Jordan decomposition theorem given in Corollary 7.1, which makes the definition of certain functions particularly simple, and it allows us to avoid a discussion of issues of convergence.

9.1 Nilpotent transformations

Definition 9.1. Let V be a finite-dimensional vector space over the field F, let $T : V \to V$ be a linear transformation. T is called nilpotent if $T^n = 0$ for some positive integer n. A square matrix A is called nilpotent if $A^n = 0$ for some integer n.

Lemma 9.1. *Let $V \neq \{\mathbf{0}\}$ be a finite-dimensional vector space over an algebraically closed field F, let $T : V \to V$ be a linear transformation. T is nilpotent if and only if the only eigenvalue of T is* 0.

Proof. If all eigenvalues of T are 0, then $P_{\text{char } V,T}(x) = x^m$ for some positive integer m according to equation (7.2), and so $T^m = 0$ by the Theorem 6.1 (the Cayley–Hamilton theorem).

On the other hand, if $T \neq 0$ is nilpotent then we have $P_{\text{min } V,T}(x) = x^m$ for the least positive integer m for which $T^m = 0$ by Lemma 3.1, so all eigenvalues of T are 0 (cf. Corollary 7.1). □

Lemma 9.2. *Let V be a finite-dimensional vector space over the field F, and let T_1, T_2, …, T_n be nilpotent linear transformations from V into V, and assume that any two of these transformations commute. Let $P(x_1, x_2, \ldots, x_n)$ be a polynomial without a constant term. Then the linear transformation $T = P(T_1, T_2, \ldots, T_n)$ is nilpotent.*

Proof. Let m be a positive integer such that $T_i^m = 0$ for all i with $1 \le i \le m$. Then each monomial in T^{mn} contains a factor $T_i^{m'}$ with $m' \ge m$ for some i with $1 \le i \le n$. □

Lemma 9.3. *Let V be a finite-dimensional vector space over the field F, and let $N : V \to V$ be a nilpotent linear transformation. Then, writing $I : V \to V$ for the identity transformation, $I - N$ is invertible.*

Proof. We have

$$(1-x)^{-1} = \sum_{n=0}^{\infty} x^n$$

for $-1 < x < 1$. In analogy with this series, we might try to form the inverse of $I - N$ as

$$T = \sum_{n=0}^{\infty} N^n.$$

First note that this is a finite sum, since N is nilpotent. Second, we can show that $T = (I-N)^{-1}$ by using the same argument as one can use to show that the first series represent $(1-x)^{-1}$, and when doing so, we can dispense with convergence considerations, since we are dealing with a finite sum. We have

$$(I-N)T = (I-N)\sum_{n=0}^{\infty} N^n = \sum_{n=0}^{\infty} N^n - \sum_{n=0}^{\infty} N^{n+1} = \sum_{n=0}^{\infty} N^n - \sum_{n=1}^{\infty} N^n = I,$$

and similarly $T(I-N) = I$ (in fact, T and N commute, so the latter equation follows from the former). Hence, $T = (I-N)^{-1}$. □

9.2 The square root of a linear transformation

The series used to represent $(I-N)^{-1}$ in the above proof is called *Neumann series*, named after Carl Gottfried Neumann. Nilpotent transformations occur, for example, in connection with Jordan subspaces described in Subsection 7.1. In fact, if V itself is a Jordan subspace, then $P_{\min V,T} = (x-\lambda)^m$ for some $\lambda \in F$ and some positive integer m. That is, the linear transformation $T - \lambda I$ is nilpotent. This fact has important applications.

Assuming F is either $\mathbb{R}$ or $\mathbb{C}$, and assume $\lambda \ne 0$. Writing $N = \lambda^{-1}T - I$, with the above T and λ, the linear transformation N is also nilpotent. This can be used to find a square root of $\lambda^{-1}T = I + N$. In $\mathbb{R}$, we have the binomial series

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n \tag{9.2}$$

convergent for any real α when $-1 < x < 1$, where

$$\binom{\alpha}{k} = \prod_{j=0}^{k-1} \frac{\alpha - j}{k - j} \qquad (k \ge 0).$$

Using the identity $(1+x)^\alpha (1+x)^\beta = (1+x)^{\alpha+\beta}$ and comparing the coefficients of x^n in the corresponding series, we obtain the identity

$$\sum_{k=0}^{n} \binom{\alpha}{k}\binom{\beta}{n-k} = \binom{\alpha+\beta}{n}. \tag{9.3}$$

Taking $\alpha = 1/2$, and writing

$$S = (I+N)^{1/2} = \sum_{k=0}^{\infty} \binom{1/2}{k} N^k, \tag{9.4}$$

the series on the right-hand side is a finite series, since N is a nilpotent linear transformation. and from the formal definition it appears that S is the square root of $I+N$. Using equation (9.3) with $\alpha = \beta = 1/2$, we obtain that indeed $S^2 = I + N$ (note that $\binom{1}{0} = \binom{1}{1}$, and $\binom{1}{n} = 0$ for $n > 1$, so that equation (9.2) is true also for $\alpha = 1$).

The square root of a linear transformation cannot be expected to be unique, since even the square root of a positive real number is not unique. In fact, if S is a square matrix with ± 1 is its main diagonal and 0 everywhere else, then $S^2 = I$, showing that the $n \times n$ identity matrix has at least 2^n square roots. To narrow down our possibilities, a simple observation is that if λ is an eigenvalue of the linear transformation $S : V \to V$, then λ^2 is an eigenvalue of S^2; indeed, if $S\mathbf{v} = \lambda\mathbf{v}$ then $S^2\mathbf{v} = \lambda^2\mathbf{v}$. If $N : V \to V$ is a nilpotent linear transformation, then it follows from Lemma 9.1 that the only eigenvalue of $I+N$ is 1, since if $(I+N)\mathbf{v} = \lambda\mathbf{v}$, and so $N\mathbf{v} = (\lambda - 1)\mathbf{v}$, and thus $\lambda - 1 = 0$. Therefore, the 1 or -1 can be the only eigenvalues of the square root of $I+N$. We have

Lemma 9.4. *Let V be a finite-dimensional vector space over the field $\mathbb{C}$ of complex numbers, let N be a nilpotent linear transformation, and let $I : V \to V$ be the identity transformation. Then there is exactly one linear transformation $T : V \to V$ such that $T^2 = I + N$ and all eigenvalues of T are* 1.

Proof. If the only eigenvalue of T is 1, the only eigenvalue of $T - I$ is 0; in fact, if $(T-I)\mathbf{v} = \lambda\mathbf{v}$, then $T\mathbf{v} = (\lambda+1)\mathbf{v}$, and so $\lambda + 1 = 1$, showing that $\lambda = 0$. Hence $T - I$ is nilpotent. If $T^2 = I + N$, then $T(I+N) = T^3 = T^2T = (I+N)T$, showing that T commutes with $I+N$, and so T also commutes with N.

Now, let T be a linear transformation all whose eigenvalues are 1 and assume $T^2 = I + N$, and let S be the linear transformation defined by equation (9.4). Then $S^2 = I + N$ and T commutes with S, since T commutes with N, and S is a polynomial of N. Hence

$$0 = T^2 - S^2 = T^2 + TS - ST - S^2 = (T-S)(T+S) = 2(T-S)(I+M),$$

where $M = (1/2)\big((T-I) + (S-I)\big)$ is a nilpotent linear transformation according to Lemma 9.2. Hence $(I+M)$ is invertible according Lemma 9.3. Thus $T - S = 0 \cdot (1/2)(I+M)^{-1} = 0$ according to the above equation. □

Corollary 9.1. *Let V be a finite-dimensional vector space over the field $\mathbb{C}$ of complex numbers, and let $T : V \to V$ be a linear transformation all whose eigenvalues are positive real numbers. Then there is a unique linear transformation $S : V \to V$ such that $S^2 = T$ and all eigenvalues of S are positive.*

Proof. If V itself is a Jordan subspace of T then $T = \lambda I + N$ for some nilpotent linear transformation N and for some $\lambda > 0$, and so we can find a unique $S = \sqrt{T} = \sqrt{\lambda}(I + \lambda^{-1}N)^{1/2}$ according to Lemma 9.4.

If $V = \bigoplus_{i=1}^n Z_i$ is the decomposition of V into Jordan subspaces for T given in Corollary 7.1 then we can define the square root $S_i : Z_i \to Z_i$ of $T \restriction Z_i$ as above, and we can define S by putting $S\mathbf{u} = S_i\mathbf{u}$ if $\mathbf{u} \in Z_i$. □

9.3 Normed vector spaces

Since the series (9.1) for e^x converges on the whole complex plane, the exponential function of a linear transformation on a vector space over the field of complex numbers can be discussed without invoking the Jordan decomposition, but for such a discussion one needs to invoke convergence, and for this one needs a norm:

Definition 9.2. Let V be a finite-dimensional vector space over the field F, where F is the field $\mathbb{R}$ of real numbers, or the field $\mathbb{C}$ of complex number. A norm N on V is a function $N : V \to R$ such that, for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V, and all scalars $\lambda \in F$, we have

1. $N(\mathbf{u}) \geq 0$,
2. $N\mathbf{u} = 0$ only if $\mathbf{u} = \mathbf{0}$,
3. $N(\mathbf{u} + \mathbf{v}) \leq N(\mathbf{u}) + N(\mathbf{v})$
4. $N(\lambda\mathbf{u}) \leq \ |\lambda| \ N(\mathbf{u})$.

If V is the space of column vectors $F_{n,1}$, where $F = \mathbf{R}$ or $F = \mathbf{C}$, then the simplest norm is the l^2 norm, defined as $\|\mathbf{v}\| = \sqrt{\sum_{k=1}^{n} |c_k|^2}$ for a column vector $\mathbf{v} = (c_1, c_2, \ldots, c_n)^T$ (T in superscript indicates transpose).[9.1] It takes some effort to prove that this is indeed a norm; such a proof can be found in many standard textbooks, and we omit it here.

If V is an arbitrary finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, one can define a norm on V since V can be represented as a space of column vectors – see Subsection 8.1. Let V be such a vector space with $\|\mathbf{v}\|$ denoting the norm of the vector $\mathbf{v} \in V$. For a linear transformation $T : V \to V$ write

$$\|T\| = \sup\{\|T\mathbf{v}\| : \mathbf{v} \in V \quad \text{and} \quad \|\mathbf{v}\| = 1\}.$$

It is not difficult to prove that this defines a norm on the vector space of all linear transformations from $V \to V$; it is called the norm *induced* on the space of linear transformations by the norm on the vector space V. In addition, we have

$$\|T\mathbf{v}\| \leq \|T\| \cdot \|\mathbf{v}\|$$

for any $\mathbf{v} \in V$.

If $\mathbf{v} \in V$, $\mathbf{v}_n \in V$ for all positive integers n, we say that the sequence of the vectors $\mathbf{v}_n$ converges to $\mathbf{v}$ in norm if $\lim_{n\to\infty} \|\mathbf{v}_n - \mathbf{v}\| = 0$. Similarly, if $T : V \to V$ and $T_n : V \to V$ are linear transformations for all positive integers n, we say that the sequence of the linear transformations T_n converges to T in norm if $\lim_{n\to\infty} \|T_n - T\| = 0$.[9.2] We can define the sum of an infinite series of vectors or operators as the limit of their partial sums, i.e., in the same way as they are defined for series of numbers.

[9.1] Printing column vectors takes up too much space, so it is often preferred to describe a column vector as the transpose of a row vector.

[9.2] There are other kinds of convergence that can be defined for linear transformations on normed vector spaces (i.e., for vector spaces supplied with a norm), but these other kinds of convergence are of interest only for infinite dimensional spaces.

9.4 The exponential function

If V is a finite dimensional vector space over the field F, where $F = \mathbb{R}$ or $F = \mathbb{C}$, and $T : V \to V$ is a linear transformation, then we can use the series in formula (9.1) to define e^T as

$$e^T = \sum_{n=0}^{\infty} \frac{1}{n!} T^n.$$

The exponential function of linear transformation plays an important role in the discussion of systems of linear differential equations with constant coefficients. We will not go into a detailed discussion of this function except for showing that the above definition can be used to prove that if $T : V \to V$ and $S : V \to V$ are linear transformations that commute, then $e^{T+S} = e^T e^S$; it is important to point out that this equation is not true if T and S do not commute. The proof of this equation is the same as the one for numbers; since the series (9.1) converges for all (real or complex) x, the convergence issues can be easily handled. We have

$$\begin{aligned} e^{T+S} &= \sum_{n=0}^{\infty} \frac{1}{n!}(T+S)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} T^k S^{n-k} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} T^k S^{n-k} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{k!} T^k \, \frac{1}{(n-k)!} S^{n-k} = \sum_{k=0}^{\infty} \frac{1}{k!} T^k \, \sum_{l=0}^{\infty} \frac{1}{l!} S^l = e^T e^S; \end{aligned} \tag{9.5}$$

the second equation here uses the binomial theorem, and the binomial theorem is not valid if T and S do not commute.

9.5 The square root of a matrix

When one wants to define the square root of a square matrix over the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$ with positive eigenvalues, our point of departure is to stipulate that if n is a positive integer, A is an $n \times n$ matrix with positive eigenvalues, and S is an invertible $n \times n$ matrix then $\sqrt{A} = S\sqrt{S^{-1}AS}S^{-1}$. The reason for this is not that we have $A^n = S(S^{-1}AS)^n S^{-1}$, and so

$$\sum_{n=0}^{\infty} c_n A^n = S\Big(\sum_{n=0}^{\infty} c_n (S^{-1}AS)^n\Big) S^{-1}.$$

After all, the definition of $\sqrt{T}$ above was not done by a single power series on the whole space V. The real reason is that A and $S^{-1}AS$ represent the same linear transformation with respect to different bases; see equation (8.1). That is, assuming that all eigenvalues of A are positive, to define $\sqrt{A}$, first we find by Theorem 8.2 a similar matrix $S^{-1}AS$ that is in Jordan canonical form $B = S^{-1}AS = \bigoplus_{i=1}^m B_i$, where $B_i = \lambda_i I + N_i$ is a Jordan block matrix $\lambda_i I_i + N_i$ (cf. (8.9)), where I_i is the identity matrix of the same size as N_i (that is, I_i is *not* the $n \times n$ identity matrix). Then we can define $\sqrt{B_i} = \sqrt{\lambda_i}(I_i + \lambda_i^{-1} N)^{1/2}$ using the binomial series as in Subsection 9.2. The matrix N_i is a nilpotent matrix, where a square matrix N is called *nilpotent* if $N^k = 0$ for some positive integer k, and so the binomial series used will be a finite series. We can then define $\sqrt{B} = \bigoplus_{i=1}^m \sqrt{B_i}$, and, finally, we can define the square root of A as $\sqrt{A} = S\sqrt{B}S^{-1}$.

9.6 The exponential of a matrix

Defining e^A is a simple matter for a square matrix A, since the series

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n.$$

is convergent for any reasonable definition of convergence. For example, we can use convergence in the induced l^2 norm of matrices, or simply entrywise convergence.[9.3]

9.6.1 Examples for matrix exponentials

Equation (9.5) for matrices says that $e^{A+B} = e^A e^B$ if A and B are commuting square matrices. Here we will give an example showing that this equation may not hold if A and B do not commute. To this end, let x be a real number and let

$$A = \begin{pmatrix} 0 & 0 \\ x & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & -x \\ 0 & 0 \end{pmatrix}.$$

A and B do not commute; in fact, we have

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & -x^2 \end{pmatrix} \quad \text{and} \quad BA = \begin{pmatrix} -x^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then $A^2 = B^2 = 0$, so, writing I for the 2×2 identity matrix, we have

$$e^A = I + A = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}, \quad e^B = I + B = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix},$$
$$e^A e^B = \begin{pmatrix} 1 & -x \\ x & 1 - x^2 \end{pmatrix}, \quad \text{and} \quad e^B e^A = \begin{pmatrix} 1 - x^2 & -x \\ x & 1 \end{pmatrix}.$$

Further, for any nonnegative integer n we have

$$(A+B)^{2n} = \begin{pmatrix} (-1)^n x^{2n} & 0 \\ 0 & (-1)^n x^{2n} \end{pmatrix} \quad \text{and} \quad (A+B)^{2n+1} = \begin{pmatrix} 0 & (-1)^{n+1} x^{2n+1} \\ (-1)^n x^{2n+1} & 0 \end{pmatrix}.$$

Noting that

$$\cos x = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!},$$

and

$$\sin x = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!},$$

we have

$$e^{A+B} = \sum_{n=0}^{\infty} \frac{1}{n!} (A+B)^n = \begin{pmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{pmatrix}. \tag{9.6}$$

[9.3] Entrywise convergence means that the corresponding entries of the members of a matrix sequence converge. While it is easy to understand what this means, proving entrywise convergence directly is messy, and it is easier to deal with convergence in an induced matrix norm.

9.6.2 A matrix representation of complex numbers

Given the matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

the set

$$S = \{xI + yJ : x, y \in \mathbb{R}\}$$

forms a field under matrix addition and matrix multiplication, and the mapping $f : S \to \mathbb{C}$ defined by $f(xI + yJ) = x + yi$ is an isomorphism between S and the field of complex numbers $\mathbb{C}$. In this light, equation (9.6) is equivalent to Euler's equation

$$e^{ix} = \cos x + i \sin x.$$

References

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