Chapter 22

Annihilating Polynomials and the Primary Decomposition

In this chapter all vector spaces are defined over an arbitrary field K.

In Section 6.7 we explained that if $f\colon E \to E$ is a linear map on a K-vector space E, then for any polynomial $p(X) = a_0X^d + a_1X^{d-1} + \cdots + a_d$ with coefficients in the field K, we can define the *linear map* $p(f)\colon E \to E$ by

$$p(f) = a_0f^d + a_1f^{d-1} + \cdots + a_d\mathrm{id},$$

where $f^k = f \circ \cdots \circ f$, the k-fold composition of f with itself. Note that

$$p(f)(u) = a_0f^d(u) + a_1f^{d-1}(u) + \cdots + a_du,$$

for every vector $u \in E$. Then we showed that if E is finite-dimensional and if $\chi_f(X) = \det(XI - f)$ is the characteristic polynomial of f, by the Cayley–Hamilton theorem, we have

$$\chi_f(f) = 0.$$

This fact suggests looking at the set of all polynomials $p(X)$ such that

$$p(f) = 0.$$

Such polynomials are called *annihilating polynomials* of f, the set of all these polynomials, denoted $\mathrm{Ann}(f)$, is called the *annihilator* of f, and the Cayley-Hamilton theorem shows that it is nontrivial since it contains a polynomial of positive degree. It turns out that $\mathrm{Ann}(f)$ contains a polynomial m_f of smallest degree that generates $\mathrm{Ann}(f)$, and this polynomial divides the characteristic polynomial. Furthermore, the polynomial m_f encapsulates a lot of information about f, in particular whether f can be diagonalized. One of the main reasons for this is that a scalar $\lambda \in K$ is a zero of the minimal polynomial m_f if and only if λ is an eigenvalue of f.

The first main result is Theorem 22.12 which states that if $f\colon E \to E$ is a linear map on a finite-dimensional space E, then f is diagonalizable iff its minimal polynomial m is of the form

$$m = (X - \lambda_1) \cdots (X - \lambda_k),$$

where $\lambda_1, \ldots, \lambda_k$ are distinct elements of K.

One of the technical tools used to prove this result is the notion of *f-conductor*; see Definition 22.7. As a corollary of Theorem 22.12 we obtain results about finite commuting families of diagonalizable or triangulable linear maps.

If $f\colon E \to E$ is a linear map and $\lambda \in K$ is an eigenvalue of f, recall that the eigenspace E_λ associated with λ is the kernel of the linear map $\lambda \mathrm{id} - f$. If all the eigenvalues $\lambda_1 \ldots, \lambda_k$ of f are in K and if f is diagonalizable, then

$$E = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k},$$

but in general there are not enough eigenvectors to span E. A remedy is to generalize the notion of eigenvector and look for (nonzero) vectors u (called generalized eigenvectors) such that

$$(\lambda \mathrm{id} - f)^r(u) = 0, \quad \text{for some } r \geq 1.$$

Then it turns out that if the minimal polynomial of f is of the form

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

then $r = r_i$ does the job for λ_i; that is, if we let

$$W_i = \mathrm{Ker}\,(\lambda_i \mathrm{id} - f)^{r_i},$$

then

$$E = W_1 \oplus \cdots \oplus W_k.$$

The above facts are parts of the *primary decomposition theorem* (Theorem 22.17). It is a special case of a more general result involving the factorization of the minimal polynomial m into its irreducible monic factors; see Theorem 22.16.

Theorem 22.17 implies that every linear map f that has all its eigenvalues in K can be written as $f = D + N$, where D is diagonalizable and N is nilpotent (which means that $N^r = 0$ for some positive integer r). Furthermore D and N commute and are unique. This is the *Jordan decomposition*, Theorem 22.18.

The Jordan decomposition suggests taking a closer look at nilpotent maps. We prove that for any nilpotent linear map $f\colon E \to E$ on a finite-dimensional vector space E of dimension n over a field K, there is a basis of E such that the matrix N of f is of the form

$$N = \begin{pmatrix} 0 & \nu_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \nu_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \nu_n \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where $\nu_i = 1$ or $\nu_i = 0$; see Theorem 22.22. As a corollary we obtain the *Jordan form*; which involves matrices of the form

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

called *Jordan blocks*; see Theorem 22.23.

22.1 Basic Properties of Polynomials; Ideals, GCD's

In order to understand the structure of $\mathrm{Ann}(f)$, we need to review three basic properties of polynomials. We refer the reader to Hoffman and Kunze, [38], Artin [3], Dummit and Foote [19], and Godement [26] for comprehensive discussions of polynomials and their properties.

We begin by recalling some basic nomenclature. Given a field K, any nonzero polynomial $p(X) \in K[X]$ has some monomial of highest degree a_0X^n with $a_0 \neq 0$, and the integer $n = \deg(p) \geq 0$ is called the *degree* of p. It is convenient to set the degree of the zero polynomial (denoted by 0) to be

$$\deg(0) = -\infty.$$

A polynomial $p(X)$ such that the coefficient a_0 of its monomial of highest degree is 1 is called a *monic* polynomial. For example, let $K = \mathbb{R}$. The polynomial $p(X) = 4X^7 + 2X^5$ is of degree 7 but is not monic since $a_0 = 4$. On the other hand, the polynomial $p(X) = X^3 - 3X + 1$ is a monic polynomial of degree 3.

We now discuss three key concepts of polynomial algebra:

1. Ideals

2. Greatest common divisors and the Bezout identity.

3. Irreducible polynomials and prime factorization.

Recall the definition a of ring (see Definition 2.2).

Definition 22.1. A *ring* is a set A equipped with two operations $+\colon A \times A \to A$ (called *addition*) and $*\colon A \times A \to A$ (called *multiplication*) having the following properties:

(R1) A is an abelian group w.r.t. $+$;

(R2) $*$ is associative and has an identity element $1 \in A$;

(R3) $*$ is distributive w.r.t. $+$.

The identity element for addition is denoted 0, and the additive inverse of $a \in A$ is denoted by $-a$. More explicitly, the axioms of a ring are the following equations which hold for all $a, b, c \in A$:

$$a + (b + c) = (a + b) + c \qquad \text{(associativity of } +\text{)} \tag{22.1}$$
$$a + b = b + a \qquad \text{(commutativity of } +\text{)} \tag{22.2}$$
$$a + 0 = 0 + a = a \qquad \text{(zero)} \tag{22.3}$$
$$a + (-a) = (-a) + a = 0 \qquad \text{(additive inverse)} \tag{22.4}$$
$$a * (b * c) = (a * b) * c \qquad \text{(associativity of } *\text{)} \tag{22.5}$$
$$a * 1 = 1 * a = a \qquad \text{(identity for } *\text{)} \tag{22.6}$$
$$(a + b) * c = (a * c) + (b * c) \qquad \text{(distributivity)} \tag{22.7}$$
$$a * (b + c) = (a * b) + (a * c) \qquad \text{(distributivity)} \tag{22.8}$$

The ring A is *commutative* if

$$a * b = b * a \quad \text{for all } a, b \in A.$$

From (22.7) and (22.8), we easily obtain

$$a * 0 = 0 * a = 0 \tag{22.9}$$
$$a * (-b) = (-a) * b = -(a * b). \tag{22.10}$$

The first crucial notion is that of an ideal.

Definition 22.2. Given a commutative ring A with unit 1, *an ideal of* A is a nonempty subset $\mathfrak{I}$ of A satisfying the following properties:

(ID1) If $a, b \in \mathfrak{I}$, then $b - a \in \mathfrak{I}$.

(ID2) If $a \in \mathfrak{I}$, then $ax \in \mathfrak{I}$ for all $x \in A$.

An ideal $\mathfrak{I}$ is a *principal ideal* if there is some $a \in \mathfrak{I}$, called a *generator*, such that

$$\mathfrak{I} = \{ax \mid x \in A\}.$$

In this case we usually write $\mathfrak{I} = aA$ or $\mathfrak{I} = (a)$. The ideal $\mathfrak{I} = (0) = \{0\}$ is called the *null ideal* (or *zero ideal*).

The following proposition is a fundamental result about polynomials over a field.

Proposition 22.1. *If K is a field, then every polynomial ideal $\mathfrak{I} \subseteq K[X]$ is a principal ideal. As a consequence, if $\mathfrak{I}$ is not the zero ideal, then there is a unique monic polynomial*

$$p(X) = X^n + a_1 X^{n-1} + \cdots + a_{n-1} X + a_n$$

in $\mathfrak{I}$ such that $\mathfrak{I} = (p)$.

Proof. This result is not hard to prove if we recall that polynomials can divided. Given any two nonzero polynomials $f, g \in K[X]$, there are unique polynomials q, r such that

$$f = qg + r, \quad \text{and} \quad \deg(r) < \deg(g). \tag{*}$$

If $\mathfrak{I}$ is not the zero ideal, there is some polynomial of smallest degree in $\mathfrak{I}$, and since K is a field, by suitable multiplication by a scalar, we can make sure that this polynomial is monic. Thus, let f be a monic polynomial of smallest degree in $\mathfrak{I}$. By (ID2), it is clear that $(f) \subseteq \mathfrak{I}$. Now let $g \in \mathfrak{I}$. Using $(*)$, there exist unique $q, r \in K[X]$ such that

$$g = qf + r \quad \text{and} \quad \deg(r) < \deg(f).$$

If $r \neq 0$, there is some $\lambda \neq 0$ in K such that λr is a monic polynomial, and since $\lambda r = \lambda g - \lambda q f$, with $f, g \in \mathfrak{I}$, by (ID1) and (ID2), we have $\lambda r \in \mathfrak{I}$, where $\deg(\lambda r) < \deg(f)$ and λr is a monic polynomial, contradicting the minimality of the degree of f. Thus, $r = 0$, and $g \in (f)$. The uniqueness of the monic polynomial f is left as an exercise. □

We will also need to know that the greatest common divisor of polynomials exist. Given any two nonzero polynomials $f, g \in K[X]$, recall that f divides g if $g = qf$ for some $q \in K[X]$.

Definition 22.3. Given any two nonzero polynomials $f, g \in K[X]$, a polynomial $d \in K[X]$ is a *greatest common divisor of f and g* (for short, a *gcd of f and g*) if d divides f and g and whenever $h \in K[X]$ divides f and g, then h divides d. We say that f and g are *relatively prime* if 1 is a gcd of f and g.

Note that f and g are relatively prime iff all of their gcd's are constants (scalars in K), or equivalently, if f, g have no common divisor q of degree $\deg(q) \geq 1$. For example, over $\mathbb{R}$, $\gcd(X^2 - 1, X^3 + X^2 - X - 1) = (X - 1)(X + 1)$ since $X^3 + X^2 - X - 1 = (X - 1)(X + 1)^2$, while $\gcd(X^3 + 1, X - 1) = 1$.

We can characterize gcd's of polynomials as follows.

Proposition 22.2. *Let K be a field and let $f, g \in K[X]$ be any two nonzero polynomials. For every polynomial $d \in K[X]$, the following properties are equivalent:*

(1) The polynomial d is a gcd of f and g.

(2) The polynomial d divides f and g and there exist $u, v \in K[X]$ such that

$$d = uf + vg.$$

(3) The ideals $(f), (g)$, and (d) satisfy the equation

$$(d) = (f) + (g).$$

In addition, $d \neq 0$, and d is unique up to multiplication by a nonzero scalar in K.

As a consequence of Proposition 22.2, two nonzero polynomials $f, g \in K[X]$ are relatively prime iff there exist $u, v \in K[X]$ such that

$$uf + vg = 1.$$

The identity

$$d = uf + vg$$

of Part (2) of Lemma 22.2 is often called the *Bezout identity*. For an example of Bezout's identity, take $K = \mathbb{R}$. Since $X^3 + 1$ and $X - 1$ are relatively prime, we have
$1 = 1/2(X^3 + 1) - 1/2(X^2 + X + 1)(X - 1)$.

An important consequence of the Bezout identity is the following result.

Proposition 22.3. *(Euclid's proposition) Let K be a field and let $f, g, h \in K[X]$ be any nonzero polynomials. If f divides gh and f is relatively prime to g, then f divides h.*

Proposition 22.3 can be generalized to any number of polynomials.

Proposition 22.4. *Let K be a field and let $f, g_1, \ldots, g_m \in K[X]$ be some nonzero polynomials. If f and g_i are relatively prime for all i, $1 \leq i \leq m$, then f and $g_1 \cdots g_m$ are relatively prime.*

Definition 22.3 is generalized to any finite number of polynomials as follows.

Definition 22.4. Given any nonzero polynomials $f_1, \ldots, f_n \in K[X]$, where $n \geq 2$, a polynomial $d \in K[X]$ is a *greatest common divisor of $f_1, \ldots, f_n$* (for short, a *gcd of $f_1, \ldots, f_n$*) if d divides each f_i and whenever $h \in K[X]$ divides each f_i, then h divides d. We say that $f_1, \ldots, f_n$ are *relatively prime* if 1 is a gcd of $f_1, \ldots, f_n$.

It is easily shown that Proposition 22.2 can be generalized to any finite number of polynomials.

Proposition 22.5. *Let K be a field and let $f_1, \ldots, f_n \in K[X]$ be any $n \geq 2$ nonzero polynomials. For every polynomial $d \in K[X]$, the following properties are equivalent:*

(1) The polynomial d is a gcd of $f_1, \ldots, f_n$.

(2) The polynomial d divides each f_i and there exist $u_1, \ldots, u_n \in K[X]$ such that

$$d = u_1 f_1 + \cdots + u_n f_n.$$

(3) The ideals (f_i), and (d) satisfy the equation

$$(d) = (f_1) + \cdots + (f_n).$$

In addition, $d \neq 0$, and d is unique up to multiplication by a nonzero scalar in K.

As a consequence of Proposition 22.5, any $n \geq 2$ nonzero polynomials $f_1, \ldots, f_n \in K[X]$ are relatively prime iff there exist $u_1, \ldots, u_n \in K[X]$ such that

$$u_1 f_1 + \cdots + u_n f_n = 1,$$

the *Bezout identity*.

We will also need to know that every nonzero polynomial (over a field) can be factored into irreducible polynomials, which is the generalization of the prime numbers to polynomials.

Definition 22.5. Given a field K, a polynomial $p \in K[X]$ is *irreducible or indecomposable or prime* if $\deg(p) \geq 1$ and if p is not divisible by any polynomial $q \in K[X]$ such that $1 \leq \deg(q) < \deg(p)$. Equivalently, p is irreducible if $\deg(p) \geq 1$ and if $p = q_1 q_2$, then either $q_1 \in K$ or $q_2 \in K$ (and of course, $q_1 \neq 0$, $q_2 \neq 0$).

Every polynomial $aX + b$ of degree 1 is irreducible. Over the field $\mathbb{R}$, the polynomial $X^2 + 1$ is irreducible (why?), but $X^3 + 1$ is not irreducible, since

$$X^3 + 1 = (X + 1)(X^2 - X + 1).$$

The polynomial $X^2 - X + 1$ is irreducible over $\mathbb{R}$ (why?). It would seem that $X^4 + 1$ is irreducible over $\mathbb{R}$, but in fact,

$$X^4 + 1 = (X^2 - \sqrt{2}X + 1)(X^2 + \sqrt{2}X + 1).$$

However, in view of the above factorization, $X^4 + 1$ is irreducible over $\mathbb{Q}$.

It can be shown that the irreducible polynomials over $\mathbb{R}$ are the polynomials of degree 1 or the polynomials of degree 2 of the form $aX^2 + bX + c$, for which $b^2 - 4ac < 0$ (i.e., those having no real roots). This is not easy to prove! Over the complex numbers $\mathbb{C}$, the only irreducible polynomials are those of degree 1. This is a version of a fact often referred to as the "Fundamental Theorem of Algebra."

Observe that the definition of irreducibilty implies that any finite number of distinct irreducible polynomials are relatively prime.

The following fundamental result can be shown

Theorem 22.6. *Given any field K, for every nonzero polynomial*

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_0$$

of degree $d = \deg(f) \geq 1$ in $K[X]$, there exists a unique set $\{\langle p_1, k_1 \rangle, \ldots, \langle p_m, k_m \rangle\}$ such that

$$f = a_d p_1^{k_1} \cdots p_m^{k_m},$$

where the $p_i \in K[X]$ are distinct irreducible monic polynomials, the k_i are (not necessarily distinct) integers, and with $m \geq 1$, $k_i \geq 1$.

We can now return to minimal polynomials.

22.2 Annihilating Polynomials and the Minimal Polynomial

Given a linear map $f \colon E \to E$, it is easy to check that the set $\mathrm{Ann}(f)$ of polynomials that annihilate f is an ideal. Furthermore, when E is finite-dimensional, the Cayley–Hamilton theorem implies that $\mathrm{Ann}(f)$ is not the zero ideal. Therefore, by Proposition 22.1, there is a unique monic polynomial m_f that generates $\mathrm{Ann}(f)$.

Definition 22.6. If $f \colon E \to E$ is a linear map on a finite-dimensional vector space E, the unique monic polynomial $m_f(X)$ that generates the ideal $\mathrm{Ann}(f)$ of polynomials which annihilate f (the *annihilator* of f) is called the *minimal polynomial* of f.

The minimal polynomial m_f of f is the monic polynomial of smallest degree that annihilates f. Thus, the minimal polynomial divides the characteristic polynomial χ_f, and $\deg(m_f) \geq 1$. For simplicity of notation, we often write m instead of m_f.

If A is any $n \times n$ matrix, the set $\mathrm{Ann}(A)$ of polynomials that annihilate A is the set of polynomials

$$p(X) = a_0 X^d + a_1 X^{d-1} + \cdots + a_{d-1} X + a_d$$

such that

$$a_0 A^d + a_1 A^{d-1} + \cdots + a_{d-1} A + a_d I = 0.$$

It is clear that $\mathrm{Ann}(A)$ is a nonzero ideal and its unique monic generator is called the *minimal polynomial* of A. We check immediately that if Q is an invertible matrix, then A and $Q^{-1}AQ$ have the same minimal polynomial. Also, if A is the matrix of f with respect to some basis, then f and A have the same minimal polynomial.

The zeros (in K) of the minimal polynomial of f and the eigenvalues of f (in K) are intimately related.

Proposition 22.7. *Let $f \colon E \to E$ be a linear map on some finite-dimensional vector space E. Then $\lambda \in K$ is a zero of the minimal polynomial $m_f(X)$ of f iff λ is an eigenvalue of f iff λ is a zero of $\chi_f(X)$. Therefore, the minimal and the characteristic polynomials have the same zeros (in K), except for multiplicities.*

Proof. First assume that $m(\lambda) = 0$ (with $\lambda \in K$, and writing m instead of m_f). If so, using polynomial division, m can be factored as

$$m = (X - \lambda)q,$$

with $\deg(q) < \deg(m)$. Since m is the minimal polynomial, $q(f) \neq 0$, so there is some nonzero vector $v \in E$ such that $u = q(f)(v) \neq 0$. But then, because m is the minimal polynomial,

$$\begin{aligned} 0 &= m(f)(v) \\ &= (f - \lambda \mathrm{id})(q(f)(v)) \\ &= (f - \lambda \mathrm{id})(u), \end{aligned}$$

which shows that λ is an eigenvalue of f.

Conversely, assume that $\lambda \in K$ is an eigenvalue of f. This means that for some $u \neq 0$, we have $f(u) = \lambda u$. Now it is easy to show that

$$m(f)(u) = m(\lambda)u,$$

and since m is the minimal polynomial of f, we have $m(f)(u) = 0$, so $m(\lambda)u = 0$, and since $u \neq 0$, we must have $m(\lambda) = 0$. □

Proposition 22.8. *Let $f\colon E \to E$ be a linear map on some finite-dimensional vector space E. If f diagonalizable, then its minimal polynomial is a product of distinct factors of degree 1.*

Proof. If we assume that f is diagonalizable, then its eigenvalues are all in K, and if $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of f, and then by Proposition 22.7, the minimal polynomial m of f must be a product of powers of the polynomials $(X - \lambda_i)$. Actually, we claim that

$$m = (X - \lambda_1) \cdots (X - \lambda_k).$$

For this we just have to show that m annihilates f. However, for any eigenvector u of f, one of the linear maps $f - \lambda_i \mathrm{id}$ sends u to 0, so

$$m(f)(u) = (f - \lambda_1 \mathrm{id}) \circ \cdots \circ (f - \lambda_k \mathrm{id})(u) = 0.$$

Since E is spanned by the eigenvectors of f, we conclude that

$$m(f) = 0.$$ □

It turns out that the converse of Proposition 22.8 is true, but this will take a little work to establish it.

22.3 Minimal Polynomials of Diagonalizable Linear Maps

In this section we prove that if the minimal polynomial m_f of a linear map f is of the form

$$m_f = (X - \lambda_1) \cdots (X - \lambda_k)$$

for distinct scalars $\lambda_1, \ldots, \lambda_k \in K$, then f is diagonalizable. This is a powerful result that has a number of implications. But first we need of few properties of invariant subspaces.

Given a linear map $f\colon E \to E$, recall that a subspace W of E is *invariant under* f if $f(u) \in W$ for all $u \in W$. For example, if $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is $f(x, y) = (-x, y)$, the y-axis is invariant under f.

Proposition 22.9. *Let W be a subspace of E invariant under the linear map $f\colon E \to E$ (where E is finite-dimensional). Then the minimal polynomial of the restriction $f \mid W$ of f to W divides the minimal polynomial of f, and the characteristic polynomial of $f \mid W$ divides the characteristic polynomial of f.*

Sketch of proof. The key ingredient is that we can pick a basis $(e_1, \ldots, e_n)$ of E in which $(e_1, \ldots, e_k)$ is a basis of W. The matrix of f over this basis is a block matrix of the form

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix},$$

where B is a $k \times k$ matrix, D is an $(n-k) \times (n-k)$ matrix, and C is a $k \times (n-k)$ matrix. Then

$$\det(XI - A) = \det(XI - B)\det(XI - D),$$

which implies the statement about the characteristic polynomials. Furthermore,

$$A^i = \begin{pmatrix} B^i & C_i \\ 0 & D^i \end{pmatrix},$$

for some $k \times (n-k)$ matrix C_i. It follows that any polynomial which annihilates A also annihilates B and D. So the minimal polynomial of B divides the minimal polynomial of A. □

For the next step, there are at least two ways to proceed. We can use an old-fashion argument using Lagrange interpolants, or we can use a slight generalization of the notion of annihilator. We pick the second method because it illustrates nicely the power of principal ideals.

What we need is the notion of conductor (also called transporter).

Definition 22.7. Let $f\colon E \to E$ be a linear map on a finite-dimensional vector space E, let W be an invariant subspace of f, and let u be any vector in E. The set $S_f(u, W)$ consisting of all polynomials $q \in K[X]$ such that $q(f)(u) \in W$ is called the *f-conductor of u into W*.

Observe that the minimal polynomial m_f of f always belongs to $S_f(u, W)$, so this is a nontrivial set. Also, if $W = (0)$, then $S_f(u, (0))$ is just the annihilator of f. The crucial property of $S_f(u, W)$ is that it is an ideal.

Proposition 22.10. *If W is an invariant subspace for f, then for each $u \in E$, the f-conductor $S_f(u, W)$ is an ideal in $K[X]$.*

We leave the proof as a simple exercise, using the fact that if W invariant under f, then W is invariant under every polynomial $q(f)$ in $S_f(u, W)$.

Since $S_f(u, W)$ is an ideal, it is generated by a unique monic polynomial q of smallest degree, and because the minimal polynomial m_f of f is in $S_f(u, W)$, the polynomial q divides m.

Definition 22.8. The unique monic polynomial which generates $S_f(u, W)$ is called the *conductor of u into W*.

Example 22.1. For example, suppose $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ where $f(x, y) = (x, 0)$. Observe that $W = \{(x, 0) \in \mathbb{R}^2\}$ is invariant under f. By representing f as $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, we see that $m_f(X) = \chi_f(X) = X^2 - X$. Let $u = (0, y)$. Then $S_f(u, W) = (X)$, and we say X is the conductor of u into W.

Proposition 22.11. *Let $f\colon E \to E$ be a linear map on a finite-dimensional space E and assume that the minimal polynomial m of f is of the form*

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

where the eigenvalues $\lambda_1, \ldots, \lambda_k$ of f belong to K. If W is a proper subspace of E which is invariant under f, then there is a vector $u \in E$ with the following properties:

(a) $u \notin W$;

(b) $(f - \lambda \mathrm{id})(u) \in W$, for some eigenvalue λ of f.

Proof. Observe that (a) and (b) together assert that the conductor of u into W is a polynomial of the form $X - \lambda_i$. Pick any vector $v \in E$ not in W, and let g be the conductor of v into W, i.e. $g(f)(v) \in W$. Since g divides m and $v \notin W$, the polynomial g is not a constant, and thus it is of the form

$$g = (X - \lambda_1)^{s_1} \cdots (X - \lambda_k)^{s_k},$$

with at least some $s_i > 0$. Choose some index j such that $s_j > 0$. Then $X - \lambda_j$ is a factor of g, so we can write

$$g = (X - \lambda_j)q. \tag{*}$$

By definition of g, the vector $u = q(f)(v)$ cannot be in W, since otherwise g would not be of minimal degree. However, $(*)$ implies that

$$\begin{aligned} (f - \lambda_j \mathrm{id})(u) &= (f - \lambda_j \mathrm{id})(q(f)(v)) \\ &= g(f)(v) \end{aligned}$$

is in W, which concludes the proof. □

We can now prove the main result of this section.

Theorem 22.12. *Let $f\colon E \to E$ be a linear map on a finite-dimensional space E. Then f is diagonalizable iff its minimal polynomial m is of the form*

$$m = (X - \lambda_1) \cdots (X - \lambda_k),$$

where $\lambda_1, \ldots, \lambda_k$ are distinct elements of K.

Proof. We already showed in Proposition 22.8 that if f is diagonalizable, then its minimal polynomial is of the above form (where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of f).

For the converse, let W be the subspace spanned by all the eigenvectors of f. If $W \neq E$, since W is invariant under f, by Proposition 22.11, there is some vector $u \notin W$ such that for some λ_j, we have

$$(f - \lambda_j \mathrm{id})(u) \in W.$$

Let $v = (f - \lambda_j \mathrm{id})(u) \in W$. Since $v \in W$, we can write

$$v = w_1 + \cdots + w_k$$

where $f(w_i) = \lambda_i w_i$ (either $w_i = 0$ or w_i is an eigenvector for λ_i), and so for every polynomial h, we have

$$h(f)(v) = h(\lambda_1)w_1 + \cdots + h(\lambda_k)w_k,$$

which shows that $h(f)(v) \in W$ for every polynomial h. We can write

$$m = (X - \lambda_j)q$$

for some polynomial q, and also

$$q - q(\lambda_j) = p(X - \lambda_j)$$

for some polynomial p. We know that $p(f)(v) \in W$, and since m is the minimal polynomial of f, we have

$$0 = m(f)(u) = (f - \lambda_j \mathrm{id})(q(f)(u)),$$

which implies that $q(f)(u) \in W$ (either $q(f)(u) = 0$, or it is an eigenvector associated with λ_j). However,

$$q(f)(u) - q(\lambda_j)u = p(f)((f - \lambda_j \mathrm{id})(u)) = p(f)(v),$$

and since $p(f)(v) \in W$ and $q(f)(u) \in W$, we conclude that $q(\lambda_j)u \in W$. But, $u \notin W$, which implies that $q(\lambda_j) = 0$, so λ_j is a double root of m, a contradiction. Therefore, we must have $W = E$. □

Remark: Proposition 22.11 can be used to give a quick proof of Theorem 14.5.

22.4 Commuting Families of Diagonalizable and Triangulable Maps

Using Theorem 22.12, we can give a short proof about commuting diagonalizable linear maps.

Definition 22.9. If $\mathcal{F}$ is a family of linear maps on a vector space E, we say that $\mathcal{F}$ is a *commuting family* iff $f \circ g = g \circ f$ for all $f, g \in \mathcal{F}$.

Proposition 22.13. *Let $\mathcal{F}$ be a commuting family of diagonalizable linear maps on a vector space E. There exists a basis of E such that every linear map in $\mathcal{F}$ is represented in that basis by a diagonal matrix.*

Proof. We proceed by induction on $n = \dim(E)$. If $n = 1$, there is nothing to prove. If $n > 1$, there are two cases. If all linear maps in $\mathcal{F}$ are of the form $\lambda\mathrm{id}$ for some $\lambda \in K$, then the proposition holds trivially. In the second case, let $f \in \mathcal{F}$ be some linear map in $\mathcal{F}$ which is not a scalar multiple of the identity. In this case, f has at least two distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, and because f is diagonalizable, E is the direct sum of the corresponding eigenspaces $E_{\lambda_1}, \ldots, E_{\lambda_k}$. For every index i, the eigenspace E_{λ_i} is invariant under f and under every other linear map g in $\mathcal{F}$, since for any $g \in \mathcal{F}$ and any $u \in E_{\lambda_i}$, because f and g commute, we have

$$f(g(u)) = g(f(u)) = g(\lambda_i u) = \lambda_i g(u)$$

so $g(u) \in E_{\lambda_i}$. Let $\mathcal{F}_i$ be the family obtained by restricting each $f \in \mathcal{F}$ to E_{λ_i}. By Proposition 22.9, the minimal polynomial of every linear map $f \mid E_{\lambda_i}$ in $\mathcal{F}_i$ divides the minimal polynomial m_f of f, and since f is diagonalizable, m_f is a product of distinct linear factors, so the minimal polynomial of $f \mid E_{\lambda_i}$ is also a product of distinct linear factors. By Theorem 22.12, the linear map $f \mid E_{\lambda_i}$ is diagonalizable. Since $k > 1$, we have $\dim(E_{\lambda_i}) < \dim(E)$ for $i = 1, \ldots, k$, and by the induction hypothesis, for each i there is a basis of E_{λ_i} over which $f \mid E_{\lambda_i}$ is represented by a diagonal matrix. Since the above argument holds for all i, by combining the bases of the E_{λ_i}, we obtain a basis of E such that the matrix of every linear map $f \in \mathcal{F}$ is represented by a diagonal matrix. □

There is also an analogous result for commuting families of linear maps represented by upper triangular matrices. To prove this we need the following proposition.

Proposition 22.14. *Let $\mathcal{F}$ be a nonempty commuting family of triangulable linear maps on a finite-dimensional vector space E. Let W be a proper subspace of E which is invariant under $\mathcal{F}$. Then there exists a vector $u \in E$ such that:*

1. *$u \notin W$.*

2. *For every $f \in \mathcal{F}$, the vector $f(u)$ belongs to the subspace $W \oplus Ku$ spanned by W and u.*

Proof. By renaming the elements of $\mathcal{F}$ if necessary, we may assume that $(f_1, \ldots, f_r)$ is a basis of the subspace of $\mathrm{End}(E)$ spanned by $\mathcal{F}$. We prove by induction on r that there exists some vector $u \in E$ such that

1. $u \notin W$.

2. $(f_i - \alpha_i\mathrm{id})(u) \in W$ for $i = 1, \ldots, r$, for some scalars $\alpha_i \in K$.

Consider the base case $r = 1$. Since f_1 is triangulable, its eigenvalues all belong to K since they are the diagonal entries of the triangular matrix associated with f_1 (this is the easy direction of Theorem 14.5), so the minimal polynomial of f_1 is of the form

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

where the eigenvalues $\lambda_1, \ldots, \lambda_k$ of f_1 belong to K. We conclude by applying Proposition 22.11.

Next assume that $r \geq 2$ and that the induction hypothesis holds for $f_1, \ldots, f_{r-1}$. Thus, there is a vector $u_{r-1} \in E$ such that

1. $u_{r-1} \notin W$.
2. $(f_i - \alpha_i \mathrm{id})(u_{r-1}) \in W$ for $i = 1, \ldots, r-1$, for some scalars $\alpha_i \in K$.

Let

$$V_{r-1} = \{w \in E \mid (f_i - \alpha_i \mathrm{id})(w) \in W,\ i = 1, \ldots, r-1\}.$$

Clearly, $W \subseteq V_{r-1}$ and $u_{r-1} \in V_{r-1}$. We claim that V_{r-1} is invariant under $\mathcal{F}$. This is because, for any $v \in V_{r-1}$ and any $f \in \mathcal{F}$, since f and f_i commute, we have

$$(f_i - \alpha_i \mathrm{id})(f(v)) = f((f_i - \alpha_i \mathrm{id})(v)), \quad 1 \leq i \leq r-1.$$

Now $(f_i - \alpha_i \mathrm{id})(v) \in W$ because $v \in V_{r-1}$, and W is invariant under $\mathcal{F}$, so $f(f_i - \alpha_i \mathrm{id})(v)) \in W$, that is, $(f_i - \alpha_i \mathrm{id})(f(v)) \in W$.

Consider the restriction g_r of f_r to V_{r-1}. The minimal polynomial of g_r divides the minimal polynomial of f_r, and since f_r is triangulable, just as we saw for f_1, the minimal polynomial of f_r is of the form

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

where the eigenvalues $\lambda_1, \ldots, \lambda_k$ of f_r belong to K, so the minimal polynomial of g_r is of the same form. By Proposition 22.11, there is some vector $u_r \in V_{r-1}$ such that

1. $u_r \notin W$.
2. $(g_r - \alpha_r \mathrm{id})(u_r) \in W$ for some scalars $\alpha_r \in K$.

Now since $u_r \in V_{r-1}$, we have $(f_i - \alpha_i \mathrm{id})(u_r) \in W$ for $i = 1, \ldots, r-1$, so $(f_i - \alpha_i \mathrm{id})(u_r) \in W$ for $i = 1, \ldots, r$ (since g_r is the restriction of f_r), which concludes the proof of the induction step. Finally, since every $f \in \mathcal{F}$ is the linear combination of $(f_1, \ldots, f_r)$, Condition (2) of the inductive claim implies Condition (2) of the proposition. □

We can now prove the following result.

Proposition 22.15. *Let $\mathcal{F}$ be a nonempty commuting family of triangulable linear maps on a finite-dimensional vector space E. There exists a basis of E such that every linear map in $\mathcal{F}$ is represented in that basis by an upper triangular matrix.*

Proof. Let $n = \dim(E)$. We construct inductively a basis $(u_1, \ldots, u_n)$ of E such that if W_i is the subspace spanned by $(u_1 \ldots, u_i)$, then for every $f \in \mathcal{F}$,

$$f(u_i) = a^f_{1i} u_1 + \cdots + a^f_{ii} u_i,$$

for some $a^f_{ij} \in K$; that is, $f(u_i)$ belongs to the subspace W_i.

We begin by applying Proposition 22.14 to the subspace $W_0 = (0)$ to get u_1 so that for all $f \in \mathcal{F}$,

$$f(u_1) = \alpha^f_1 u_1.$$

For the induction step, since W_i invariant under $\mathcal{F}$, we apply Proposition 22.14 to the subspace W_i, to get $u_{i+1} \in E$ such that

1. $u_{i+1} \notin W_i$.
2. For every $f \in \mathcal{F}$, the vector $f(u_{i+1})$ belong to the subspace spanned by W_i and u_{i+1}.

Condition (1) implies that $(u_1, \ldots, u_i, u_{i+1})$ is linearly independent, and Condition (2) means that for every $f \in \mathcal{F}$,

$$f(u_{i+1}) = a^f_{1i+1} u_1 + \cdots + a^f_{i+1i+1} u_{i+1},$$

for some $a^f_{i+1j} \in K$, establishing the induction step. After n steps, each $f \in \mathcal{F}$ is represented by an upper triangular matrix. □

Observe that if $\mathcal{F}$ consists of a single linear map f and if the minimal polynomial of f is of the form

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

with all $\lambda_i \in K$, using Proposition 22.11 instead of Proposition 22.14, the proof of Proposition 22.15 yields another proof of Theorem 14.5.

22.5 The Primary Decomposition Theorem

If $f \colon E \to E$ is a linear map and $\lambda \in K$ is an eigenvalue of f, recall that the eigenspace E_λ associated with λ is the kernel of the linear map $\lambda \mathrm{id} - f$. If all the eigenvalues $\lambda_1 \ldots, \lambda_k$ of f are in K, it may happen that

$$E = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k},$$

but in general there are not enough eigenvectors to span E. What if we generalize the notion of eigenvector and look for (nonzero) vectors u such that

$$(\lambda \mathrm{id} - f)^r(u) = 0, \quad \text{for some } r \geq 1?$$

It turns out that if the minimal polynomial of f is of the form

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k},$$

then $r = r_i$ does the job for λ_i; that is, if we let

$$W_i = \mathrm{Ker}\,(\lambda_i \mathrm{id} - f)^{r_i},$$

then

$$E = W_1 \oplus \cdots \oplus W_k.$$

This result is very nice but seems to require that the eigenvalues of f all belong to K. Actually, it is a special case of a more general result involving the factorization of the minimal polynomial m into its irreducible monic factors (see Theorem 22.6),

$$m = p_1^{r_1} \cdots p_k^{r_k},$$

where the p_i are distinct irreducible monic polynomials over K.

Theorem 22.16. *(Primary Decomposition Theorem) Let $f\colon E \to E$ be a linear map on the finite-dimensional vector space E over the field K. Write the minimal polynomial m of f as*

$$m = p_1^{r_1} \cdots p_k^{r_k},$$

where the p_i are distinct irreducible monic polynomials over K, and the r_i are positive integers. Let

$$W_i = \mathrm{Ker}\,(p_i^{r_i}(f)), \quad i = 1, \ldots, k.$$

Then

(a) $E = W_1 \oplus \cdots \oplus W_k$.

(b) Each W_i is invariant under f.

(c) The minimal polynomial of the restriction $f \mid W_i$ of f to W_i is $p_i^{r_i}$.

Proof. The trick is to construct projections π_i using the polynomials $p_j^{r_j}$ so that the range of π_i is equal to W_i. Let

$$g_i = m/p_i^{r_i} = \prod_{j \neq i} p_j^{r_j}.$$

Note that

$$p_i^{r_i} g_i = m.$$

Since $p_1, \ldots, p_k$ are irreducible and distinct, they are relatively prime. Then using Proposition 22.4, it is easy to show that $g_1, \ldots, g_k$ are relatively prime. Otherwise, some irreducible polynomial p would divide all of $g_1, \ldots, g_k$, so by Proposition 22.4 it would be equal to one of the irreducible factors p_i. But that p_i is missing from g_i, a contradiction. Therefore, by Proposition 22.5, there exist some polynomials $h_1, \ldots, h_k$ such that

$$g_1 h_1 + \cdots + g_k h_k = 1.$$

Let $q_i = g_i h_i$ and let $\pi_i = q_i(f) = g_i(f)h_i(f)$. We have

$$q_1 + \cdots + q_k = 1,$$

and since m divides $q_i q_j$ for $i \neq j$, we get

$$\begin{aligned} \pi_1 + \cdots + \pi_k &= \mathrm{id} \\ \pi_i \pi_j &= 0, \quad i \neq j. \end{aligned}$$

(We implicitly used the fact that if p, q are two polynomials, the linear maps $p(f) \circ q(f)$ and $q(f) \circ p(f)$ are the same since $p(f)$ and $q(f)$ are polynomials in the powers of f, which commute.) Composing the first equation with π_i and using the second equation, we get

$$\pi_i^2 = \pi_i.$$

Therefore, the π_i are projections, and E is the direct sum of the images of the π_i. Indeed, every $u \in E$ can be expressed as

$$u = \pi_1(u) + \cdots + \pi_k(u).$$

Also, if

$$\pi_1(u) + \cdots + \pi_k(u) = 0,$$

then by applying π_i we get

$$0 = \pi_i^2(u) = \pi_i(u), \quad i = 1, \ldots k.$$

To finish proving (a), we need to show that

$$W_i = \mathrm{Ker}\,(p_i^{r_i}(f)) = \pi_i(E).$$

If $v \in \pi_i(E)$, then $v = \pi_i(u)$ for some $u \in E$, so

$$\begin{aligned} p_i^{r_i}(f)(v) &= p_i^{r_i}(f)(\pi_i(u)) \\ &= p_i^{r_i}(f)g_i(f)h_i(f)(u) \\ &= h_i(f)p_i^{r_i}(f)g_i(f)(u) \\ &= h_i(f)m(f)(u) = 0, \end{aligned}$$

because m is the minimal polynomial of f. Therefore, $v \in W_i$.

Conversely, assume that $v \in W_i = \mathrm{Ker}\,(p_i^{r_i}(f))$. If $j \neq i$, then $g_j h_j$ is divisible by $p_i^{r_i}$, so

$$g_j(f)h_j(f)(v) = \pi_j(v) = 0, \quad j \neq i.$$

Then since $\pi_1 + \cdots + \pi_k = \mathrm{id}$, we have $v = \pi_i v$, which shows that v is in the range of π_i. Therefore, $W_i = \mathrm{Im}(\pi_i)$, and this finishes the proof of (a).

If $p_i^{r_i}(f)(u) = 0$, then $p_i^{r_i}(f)(f(u)) = f(p_i^{r_i}(f)(u)) = 0$, so (b) holds.

If we write $f_i = f \mid W_i$, then $p_i^{r_i}(f_i) = 0$, because $p_i^{r_i}(f) = 0$ on W_i (its kernel). Therefore, the minimal polynomial of f_i divides $p_i^{r_i}$. Conversely, let q be any polynomial such that $q(f_i) = 0$ (on W_i). Since $m = p_i^{r_i} g_i$, the fact that $m(f)(u) = 0$ for all $u \in E$ shows that

$$p_i^{r_i}(f)(g_i(f)(u)) = 0, \quad u \in E,$$

and thus $\mathrm{Im}(g_i(f)) \subseteq \mathrm{Ker}\,(p_i^{r_i}(f)) = W_i$. Consequently, since $q(f)$ is zero on W_i,

$$q(f)g_i(f) = 0 \quad \text{for all } u \in E.$$

But then qg_i is divisible by the minimal polynomial $m = p_i^{r_i} g_i$ of f, and since $p_i^{r_i}$ and g_i are relatively prime, by Euclid's proposition, $p_i^{r_i}$ must divide q. This finishes the proof that the minimal polynomial of f_i is $p_i^{r_i}$, which is (c). □

To best understand the projection constructions of Theorem 22.16, we provide the following two explicit examples of the primary decomposition theorem.

Example 22.2. First let $f \colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined as $f(x, y, z) = (y, -x, z)$. In terms of the standard basis f is represented by the 3×3 matrix $X_f := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then a simple calculation shows that $m_f(x) = \chi_f(x) = (x^2+1)(x-1)$. Using the notation of the preceding proof set

$$m = p_1 p_2, \qquad p_1 = x^2 + 1, \qquad p_2 = x - 1.$$

Then

$$g_1 = \frac{m}{p_1} = x - 1, \qquad g_2 = \frac{m}{p_2} = x^2 + 1.$$

We must find $h_1, h_2 \in \mathbb{R}[x]$ such that $g_1 h_1 + g_2 h_2 = 1$. In general this is the hard part of the projection construction. But since we are only working with two relatively prime polynomials g_1, g_2, we may apply the Euclidean algorithm to discover that

$$-\frac{x+1}{2}(x-1) + \frac{1}{2}(x^2+1) = 1,$$

where $h_1 = -\frac{x+1}{2}$ while $h_2 = \frac{1}{2}$. By definition

$$\pi_1 = g_1(f)h_1(f) = -\frac{1}{2}(X_f - \mathrm{id})(X_f + \mathrm{id}) = -\frac{1}{2}(X_f^2 - \mathrm{id}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$\pi_2 = g_2(f)h_2(f) = \frac{1}{2}(X_f^2 + \mathrm{id}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then $\mathbb{R}^3 = W_1 \oplus W_2$, where

$$\begin{aligned} W_1 = \pi_1(\mathbb{R}^3) = \operatorname{Ker}(p_1(X_f)) = \operatorname{Ker}(X_f^2 + \mathrm{id}) \\ = \operatorname{Ker}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \{(x, y, 0) \in \mathbb{R}^3\}, \end{aligned}$$

$$\begin{aligned} W_2 = \pi_2(\mathbb{R}^3) = \operatorname{Ker}(p_2(X_f)) = \operatorname{Ker}(X_f - \mathrm{id}) \\ = \operatorname{Ker}\begin{pmatrix} -1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \{(0, 0, z) \in \mathbb{R}^3\}. \end{aligned}$$

Example 22.3. For our second example of the primary decomposition theorem let $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined as $f(x, y, z) = (y, -x + z, -y)$, with standard matrix representation $X_f = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. A simple calculation shows that $m_f(x) = \chi_f(x) = x(x^2 + 2)$. Set

$$p_1 = x^2 + 2, \qquad p_2 = x, \qquad g_1 = \frac{m_f}{p_1} = x, \qquad g_2 = \frac{m_f}{p_2} = x^2 + 2.$$

Since $\gcd(g_1, g_2) = 1$, we use the Euclidean algorithm to find

$$h_1 = -\frac{1}{2}x, \qquad h_2 = \frac{1}{2},$$

such that $g_1h_1 + g_2h_2 = 1$. Then

$$\pi_1 = g_1(f)h_1(f) = -\frac{1}{2}X_f^2 = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

while

$$\pi_2 = g_2(f)h_2(f) = \frac{1}{2}(X_f^2 + 2\mathrm{id}) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Although it is not entirely obvious, π_1 and π_2 are indeed projections since

$$\pi_1^2 = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \pi_1,$$

and

$$\pi_2^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \pi_2.$$

Furthermore observe that $\pi_1 + \pi_2 = \mathrm{id}$. The primary decomposition theorem implies that $\mathbb{R}^3 = W_1 \oplus W_2$ where

$$\begin{aligned} W_1 = \pi_1(\mathbb{R}^3) = \mathrm{Ker}\,(p_1(f)) &= \mathrm{Ker}\,(X^2 + 2) \\ &= \mathrm{Ker}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \mathrm{span}\{(0,1,0), (1,0,-1)\}, \end{aligned}$$

$$W_2 = \pi_2(\mathbb{R}^3) = \mathrm{Ker}\,(p_2(f)) = \mathrm{Ker}\,(X) = \mathrm{span}\{(1,0,1)\}.$$

See Figure 22.1.

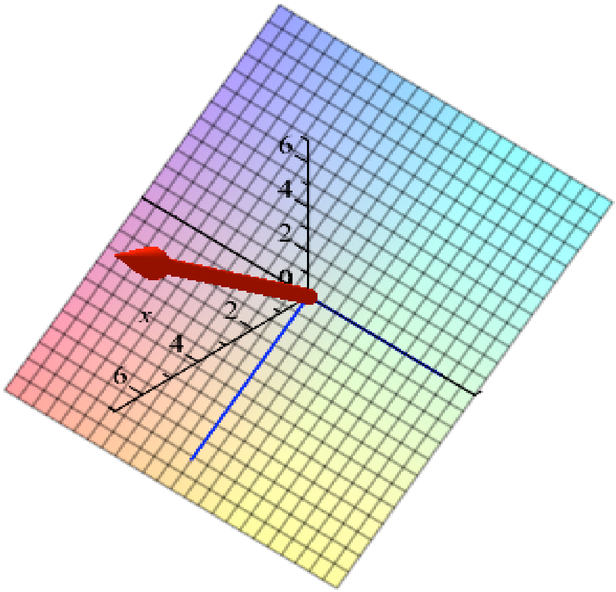

Figure 22.1: The direct sum decomposition of $\mathbb{R}^3 = W_1 \oplus W_2$ where W_1 is the plane $x + z = 0$ and W_2 is line $t(1,0,1)$. The spanning vectors of W_1 are in blue.

If all the eigenvalues of f belong to the field K, we obtain the following result.

Theorem 22.17. *(Primary Decomposition Theorem, Version 2) Let $f\colon E \to E$ be a linear map on the finite-dimensional vector space E over the field K. If all the eigenvalues $\lambda_1, \ldots, \lambda_k$ of f belong to K, write*

$$m = (X - \lambda_1)^{r_1} \cdots (X - \lambda_k)^{r_k}$$

for the minimal polynomial of f,

$$\chi_f = (X - \lambda_1)^{n_1} \cdots (X - \lambda_k)^{n_k}$$

for the characteristic polynomial of f, with $1 \leq r_i \leq n_i$, and let

$$W_i = \mathrm{Ker}\,(\lambda_i \mathrm{id} - f)^{r_i}, \quad i = 1, \ldots, k.$$

Then

(a) $E = W_1 \oplus \cdots \oplus W_k$.

(b) Each W_i is invariant under f.

(c) $\dim(W_i) = n_i$.

(d) The minimal polynomial of the restriction $f \mid W_i$ of f to W_i is $(X - \lambda_i)^{r_i}$.

Proof. Parts (a), (b) and (d) have already been proven in Theorem 22.16, so it remains to prove (c). Since W_i is invariant under f, let f_i be the restriction of f to W_i. The characteristic polynomial χ_{f_i} of f_i divides $\chi(f)$, and since $\chi(f)$ has all its roots in K, so does $\chi_i(f)$. By Theorem 14.5, there is a basis of W_i in which f_i is represented by an upper triangular matrix, and since $(\lambda_i \mathrm{id} - f)^{r_i} = 0$, the diagonal entries of this matrix are equal to λ_i. Consequently,

$$\chi_{f_i} = (X - \lambda_i)^{\dim(W_i)},$$

and since χ_{f_i} divides $\chi(f)$, we conclude hat

$$\dim(W_i) \leq n_i, \quad i = 1, \ldots, k.$$

Because E is the direct sum of the W_i, we have $\dim(W_1) + \cdots + \dim(W_k) = n$, and since $n_1 + \cdots + n_k = n$, we must have

$$\dim(W_i) = n_i, \quad i = 1, \ldots, k,$$

proving (c). □

Definition 22.10. If $\lambda \in K$ is an eigenvalue of f, we define a *generalized eigenvector* of f as a nonzero vector $u \in E$ such that

$$(\lambda \mathrm{id} - f)^r(u) = 0, \quad \text{for some } r \geq 1.$$

The *index* of λ is defined as the smallest $r \geq 1$ such that

$$\mathrm{Ker}\,(\lambda \mathrm{id} - f)^r = \mathrm{Ker}\,(\lambda \mathrm{id} - f)^{r+1}.$$

It is clear that $\mathrm{Ker}\,(\lambda \mathrm{id} - f)^i \subseteq \mathrm{Ker}\,(\lambda \mathrm{id} - f)^{i+1}$ for all $i \geq 1$. By Theorem 22.17(d), if $\lambda = \lambda_i$, the index of λ_i is equal to r_i.

22.6 Jordan Decomposition

Recall that a linear map $g\colon E \to E$ is said to be *nilpotent* if there is some positive integer r such that $g^r = 0$. Another important consequence of Theorem 22.17 is that f can be written as the sum of a diagonalizable and a nilpotent linear map (which commute). For example $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the $\mathbb{R}$-linear map $f(x, y) = (x, x + y)$ with standard matrix representation $X_f = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. A basic calculation shows that $m_f(x) = \chi_f(x) = (x - 1)^2$. By Theorem 22.12 we know that f is not diagonalizable over $\mathbb{R}$. But since the eigenvalue $\lambda_1 = 1$ of f does belong to $\mathbb{R}$, we may use the projection construction inherent within Theorem 22.17 to write $f = D + N$, where D is a diagonalizable linear map and N is a nilpotent linear map. The proof of Theorem 22.16 implies that

$$p_1^{r_1} = (x - 1)^2, \qquad g_1 = 1 = h_1, \qquad \pi_1 = g_1(f)h_1(f) = \mathrm{id}.$$

Then

$$\begin{aligned} D &= \lambda_1\pi_1 = \mathrm{id}, \\ N &= f - D = f(x, y) - \mathrm{id}(x, y) = (x, x + y) - (x, y) = (0, y), \end{aligned}$$

which is equivalent to the matrix decomposition

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

This example suggests that the diagonal summand of f is related to the projection constructions associated with the proof of the primary decomposition theorem. If we write

$$D = \lambda_1\pi_1 + \cdots + \lambda_k\pi_k,$$

where π_i is the projection from E onto the subspace W_i defined in the proof of Theorem 22.16, since

$$\pi_1 + \cdots + \pi_k = \mathrm{id},$$

we have

$$f = f\pi_1 + \cdots + f\pi_k,$$

and so we get

$$N = f - D = (f - \lambda_1\mathrm{id})\pi_1 + \cdots + (f - \lambda_k\mathrm{id})\pi_k.$$

We claim that $N = f - D$ is a nilpotent operator. Since by construction the π_i are polynomials in f, they commute with f, using the properties of the π_i, we get

$$N^r = (f - \lambda_1\mathrm{id})^r\pi_1 + \cdots + (f - \lambda_k\mathrm{id})^r\pi_k.$$

Therefore, if $r = \max\{r_i\}$, we have $(f - \lambda_k\mathrm{id})^r = 0$ for $i = 1, \ldots, k$, which implies that

$$N^r = 0.$$

It remains to show that D is diagonalizable. Since N is a polynomial in f, it commutes with f, and thus with D. From

$$D = \lambda_1 \pi_1 + \cdots + \lambda_k \pi_k,$$

and

$$\pi_1 + \cdots + \pi_k = \mathrm{id},$$

we see that

$$\begin{aligned} D - \lambda_i \mathrm{id} &= \lambda_1 \pi_1 + \cdots + \lambda_k \pi_k - \lambda_i(\pi_1 + \cdots + \pi_k) \\ &= (\lambda_1 - \lambda_i)\pi_1 + \cdots + (\lambda_{i-1} - \lambda_i)\pi_{i-1} + (\lambda_{i+1} - \lambda_i)\pi_{i+1} \\ &\quad + \cdots + (\lambda_k - \lambda_i)\pi_k. \end{aligned}$$

Since the projections π_j with $j \neq i$ vanish on W_i, the above equation implies that $D - \lambda_i \mathrm{id}$ vanishes on W_i and that $(D - \lambda_j \mathrm{id})(W_i) \subseteq W_i$, and thus that the minimal polynomial of D is

$$(X - \lambda_1) \cdots (X - \lambda_k).$$

Since the λ_i are distinct, by Theorem 22.12, the linear map D is diagonalizable.

In summary we have shown that when all the eigenvalues of f belong to K, there exist a diagonalizable linear map D and a nilpotent linear map N such that

$$\begin{aligned} f &= D + N \\ DN &= ND, \end{aligned}$$

and N and D are polynomials in f.

Definition 22.11. A decomposition of f as $f = D + N$ as above is called a *Jordan decomposition*.

In fact, we can prove more: the maps D and N are uniquely determined by f.

Theorem 22.18. *(Jordan Decomposition) Let $f \colon E \to E$ be a linear map on the finite-dimensional vector space E over the field K. If all the eigenvalues $\lambda_1, \ldots, \lambda_k$ of f belong to K, then there exist a diagonalizable linear map D and a nilpotent linear map N such that*

$$\begin{aligned} f &= D + N \\ DN &= ND. \end{aligned}$$

Furthermore, D and N are uniquely determined by the above equations and they are polynomials in f.

Proof. We already proved the existence part. Suppose we also have $f = D' + N'$, with $D'N' = N'D'$, where D' is diagonalizable, N' is nilpotent, and both are polynomials in f. We need to prove that $D = D'$ and $N = N'$.

Since D' and N' commute with one another and $f = D' + N'$, we see that D' and N' commute with f. Then D' and N' commute with any polynomial in f; hence they commute with D and N. From

$$D + N = D' + N',$$

we get

$$D - D' = N' - N,$$

and D, D', N, N' commute with one another. Since D and D' are both diagonalizable and commute, by Proposition 22.13, they are simultaneousy diagonalizable, so $D - D'$ is diagonalizable. Since N and N' commute, by the binomial formula, for any $r \geq 1$,

$$(N' - N)^r = \sum_{j=0}^{r} (-1)^j \binom{r}{j} (N')^{r-j} N^j.$$

Since both N and N' are nilpotent, we have $N^{r_1} = 0$ and $(N')^{r_2} = 0$, for some $r_1, r_2 > 0$, so for $r \geq r_1 + r_2$, the right-hand side of the above expression is zero, which shows that $N' - N$ is nilpotent. (In fact, it is easy that $r_1 = r_2 = n$ works). It follows that $D - D' = N' - N$ is both diagonalizable and nilpotent. Clearly, the minimal polynomial of a nilpotent linear map is of the form X^r for some $r > 0$ (and $r \leq \dim(E)$). But $D - D'$ is diagonalizable, so its minimal polynomial has simple roots, which means that $r = 1$. Therefore, the minimal polynomial of $D - D'$ is X, which says that $D - D' = 0$, and then $N = N'$. □

If K is an algebraically closed field, then Theorem 22.18 holds. This is the case when $K = \mathbb{C}$. This theorem reduces the study of linear maps (from E to itself) to the study of nilpotent operators. There is a special normal form for such operators which is discussed in the next section.

22.7 Nilpotent Linear Maps and Jordan Form

This section is devoted to a normal form for nilpotent maps. We follow Godement's exposition [27]. Let $f : E \to E$ be a nilpotent linear map on a finite-dimensional vector space over a field K, and assume that f is not the zero map. There is a smallest positive integer $r \geq 1$ such $f^r \neq 0$ and $f^{r+1} = 0$. Clearly, the polynomial X^{r+1} annihilates f, and it is the minimal polynomial of f since $f^r \neq 0$. It follows that $r + 1 \leq n = \dim(E)$. Let us define the subspaces N_i by

$$N_i = \mathrm{Ker}\,(f^i), \quad i \geq 0.$$

Note that $N_0 = (0)$, $N_1 = \mathrm{Ker}\,(f)$, and $N_{r+1} = E$. Also, it is obvious that

$$N_i \subseteq N_{i+1}, \quad i \geq 0.$$

Proposition 22.19. *Given a nilpotent linear map f with $f^r \neq 0$ and $f^{r+1} = 0$ as above, the inclusions in the following sequence are strict:*

$$(0) = N_0 \subset N_1 \subset \cdots \subset N_r \subset N_{r+1} = E.$$

Proof. We proceed by contradiction. Assume that $N_i = N_{i+1}$ for some i with $0 \leq i \leq r$. Since $f^{r+1} = 0$, for every $u \in E$, we have

$$0 = f^{r+1}(u) = f^{i+1}(f^{r-i}(u)),$$

which shows that $f^{r-i}(u) \in N_{i+1}$. Since $N_i = N_{i+1}$, we get $f^{r-i}(u) \in N_i$, and thus $f^r(u) = 0$. Since this holds for all $u \in E$, we see that $f^r = 0$, a contradiction. □

Proposition 22.20. *Given a nilpotent linear map f with $f^r \neq 0$ and $f^{r+1} = 0$, for any integer i with $1 \leq i \leq r$, for any subspace U of E, if $U \cap N_i = (0)$, then $f(U) \cap N_{i-1} = (0)$, and the restriction of f to U is an isomorphism onto $f(U)$.*

Proof. Pick $v \in f(U) \cap N_{i-1}$. We have $v = f(u)$ for some $u \in U$ and $f^{i-1}(v) = 0$, which means that $f^i(u) = 0$. Then $u \in U \cap N_i$, so $u = 0$ since $U \cap N_i = (0)$, and $v = f(u) = 0$. Therefore, $f(U) \cap N_{i-1} = (0)$. The restriction of f to U is obviously surjective on $f(U)$. Suppose that $f(u) = 0$ for some $u \in U$. Then $u \in U \cap N_1 \subseteq U \cap N_i = (0)$ (since $i \geq 1$), so $u = 0$, which proves that f is also injective on U. □

Proposition 22.21. *Given a nilpotent linear map f with $f^r \neq 0$ and $f^{r+1} = 0$, there exists a sequence of subspace $U_1, \ldots, U_{r+1}$ of E with the following properties:*

(1) $N_i = N_{i-1} \oplus U_i$, for $i = 1, \ldots, r + 1$.

(2) We have $f(U_i) \subseteq U_{i-1}$, and the restriction of f to U_i is an injection, for $i = 2, \ldots, r+1$.

See Figure 22.2.

Proof. We proceed inductively, by defining the sequence $U_{r+1}, U_r, \ldots, U_1$. We pick U_{r+1} to be any supplement of N_r in $N_{r+1} = E$, so that

$$E = N_{r+1} = N_r \oplus U_{r+1}.$$

Since $f^{r+1} = 0$ and $N_r = \mathrm{Ker}\,(f^r)$, we have $f(U_{r+1}) \subseteq N_r$, and by Proposition 22.20, as $U_{r+1} \cap N_r = (0)$, we have $f(U_{r+1}) \cap N_{r-1} = (0)$. As a consequence, we can pick a supplement U_r of N_{r-1} in N_r so that $f(U_{r+1}) \subseteq U_r$. We have

$$N_r = N_{r-1} \oplus U_r \quad \text{and} \quad f(U_{r+1}) \subseteq U_r.$$

By Proposition 22.20, f is an injection from U_{r+1} to U_r. Assume inductively that $U_{r+1}, \ldots, U_i$ have been defined for $i \geq 2$ and that they satisfy (1) and (2). Since

$$N_i = N_{i-1} \oplus U_i,$$

we have $U_i \subseteq N_i$, so $f^{i-1}(f(U_i)) = f^i(U_i) = (0)$, which implies that $f(U_i) \subseteq N_{i-1}$. Also, since $U_i \cap N_{i-1} = (0)$, by Proposition 22.20, we have $f(U_i) \cap N_{i-2} = (0)$. It follows that there is a supplement U_{i-1} of N_{i-2} in N_{i-1} that contains $f(U_i)$. We have

$$N_{i-1} = N_{i-2} \oplus U_{i-1} \quad \text{and} \quad f(U_i) \subseteq U_{i-1}.$$

The fact that f is an injection from U_i into U_{i-1} follows from Proposition 22.20. Therefore, the induction step is proven. The construction stops when $i = 1$. □

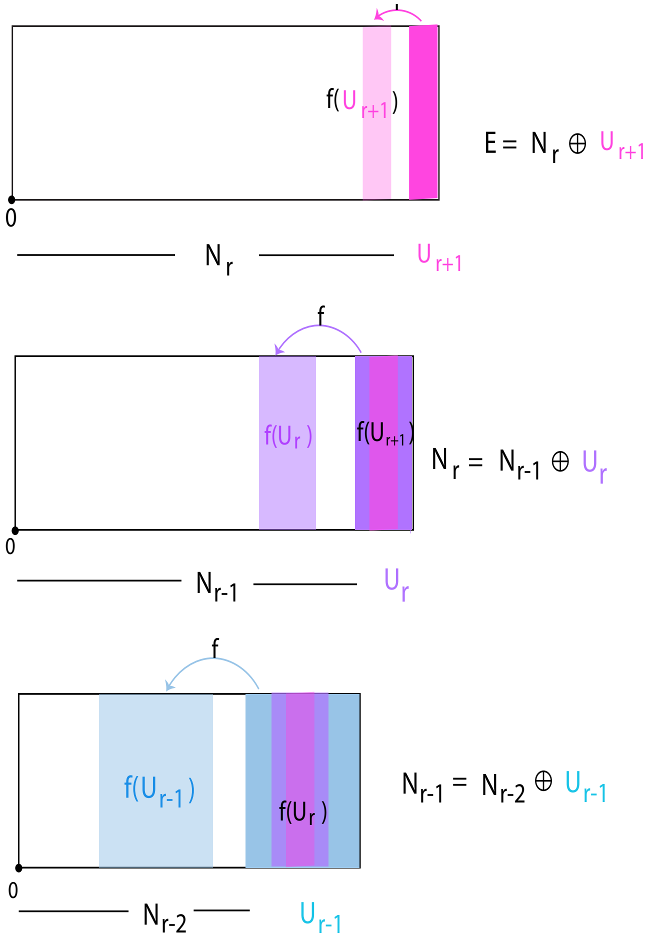

Figure 22.2: A schematic illustration of $N_i = N_{i-1} \oplus U_i$ with $f(U_i) \subseteq U_{i-1}$ for $i = r+1, r, r-1$.

Because $N_0 = (0)$ and $N_{r+1} = E$, we see that E is the direct sum of the U_i:

$$E = U_1 \oplus \cdots \oplus U_{r+1},$$

with $f(U_i) \subseteq U_{i-1}$, and f an injection from U_i to U_{i-1}, for $i = r+1, \ldots, 2$. By a clever choice of bases in the U_i, we obtain the following nice theorem.

Theorem 22.22. *For any nilpotent linear map $f \colon E \to E$ on a finite-dimensional vector space E of dimension n over a field K, there is a basis of E such that the matrix N of f is of the form*

$$N = \begin{pmatrix} 0 & \nu_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \nu_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \nu_n \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where $\nu_i = 1$ *or* $\nu_i = 0$.

Proof. First apply Proposition 22.21 to obtain a direct sum $E = \bigoplus_{i=1}^{r+1} U_i$. Then we define a basis of E inductively as follows. First we choose a basis

$$e_1^{r+1}, \ldots, e_{n_{r+1}}^{r+1}$$

of U_{r+1}. Next, for $i = r+1, \ldots, 2$, given the basis

$$e_1^i, \ldots, e_{n_i}^i$$

of U_i, since f is injective on U_i and $f(U_i) \subseteq U_{i-1}$, the vectors $f(e_1^i), \ldots, f(e_{n_i}^i)$ are linearly independent, so we define a basis of U_{i-1} by completing $f(e_1^i), \ldots, f(e_{n_i}^i)$ to a basis in U_{i-1}:

$$e_1^{i-1}, \ldots, e_{n_i}^{i-1}, e_{n_i+1}^{i-1}, \ldots, e_{n_{i-1}}^{i-1}$$

with

$$e_j^{i-1} = f(e_j^i), \quad j = 1 \ldots, n_i.$$

Since $U_1 = N_1 = \mathrm{Ker}\,(f)$, we have

$$f(e_j^1) = 0, \quad j = 1, \ldots, n_1.$$

These basis vectors can be arranged as the rows of the following matrix:

$$\left(\begin{array}{cccccccccccc}
e_1^{r+1} & \cdots & e_{n_{r+1}}^{r+1} & & & & & & & & & \\
\vdots & & \vdots & & & & & & & & & \\
e_1^{r} & \cdots & e_{n_{r+1}}^{r} & e_{n_{r+1}+1}^{r} & \cdots & e_{n_r}^{r} & & & & & & \\
\vdots & & \vdots & \vdots & & \vdots & & & & & & \\
e_1^{r-1} & \cdots & e_{n_{r+1}}^{r-1} & e_{n_{r+1}+1}^{r-1} & \cdots & e_{n_r}^{r-1} & e_{n_r+1}^{r-1} & \cdots & e_{n_{r-1}}^{r-1} & & & \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & & \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & & \\
e_1^{1} & \cdots & e_{n_{r+1}}^{1} & e_{n_{r+1}+1}^{1} & \cdots & e_{n_r}^{1} & e_{n_r+1}^{1} & \cdots & e_{n_{r-1}}^{1} & \cdots & \cdots & e_{n_1}^{1}
\end{array}\right)$$

Finally, we define the basis $(e_1, \ldots, e_n)$ by listing each column of the above matrix from the bottom-up, starting with column one, then column two, *etc*. This means that we list the vectors e_j^i in the following order:

For $j = 1, \ldots, n_{r+1}$, list $e_j^1, \ldots, e_j^{r+1}$;

In general, for $i = r, \ldots, 1$,

for $j = n_{i+1} + 1, \ldots, n_i$, list $e_j^1, \ldots, e_j^i$.

Then because $f(e_j^1) = 0$ and $e_j^{i-1} = f(e_j^i)$ for $i \geq 2$, either

$$f(e_i) = 0 \quad \text{or} \quad f(e_i) = e_{i-1},$$

which proves the theorem. □

As an application of Theorem 22.22, we obtain the *Jordan form* of a linear map.

Definition 22.12. A *Jordan block* is an $r \times r$ matrix $J_r(\lambda)$, of the form

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

where $\lambda \in K$, with $J_1(\lambda) = (\lambda)$ if $r = 1$. A *Jordan matrix*, J, is an $n \times n$ block diagonal matrix of the form

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_{r_m}(\lambda_m) \end{pmatrix},$$

where each $J_{r_k}(\lambda_k)$ is a Jordan block associated with some $\lambda_k \in K$, and with $r_1 + \cdots + r_m = n$.

To simplify notation, we often write $J(\lambda)$ for $J_r(\lambda)$. Here is an example of a Jordan matrix with four blocks:

$$J = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}.$$

Theorem 22.23. *(Jordan form) Let E be a vector space of dimension n over a field K and let $f\colon E \to E$ be a linear map. The following properties are equivalent:*

(1) The eigenvalues of f all belong to K (i.e. the roots of the characteristic polynomial χ_f all belong to K).

(2) There is a basis of E in which the matrix of f is a Jordan matrix.

Proof. Assume (1). First we apply Theorem 22.17, and we get a direct sum $E = \bigoplus_{j=1}^{k} W_k$, such that the restriction of $g_i = f - \lambda_j \mathrm{id}$ to W_i is nilpotent. By Theorem 22.22, there is a basis of W_i such that the matrix of the restriction of g_i is of the form

$$G_i = \begin{pmatrix} 0 & \nu_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \nu_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \nu_{n_i} \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where $\nu_i = 1$ or $\nu_i = 0$. Furthermore, over any basis, $\lambda_i \mathrm{id}$ is represented by the diagonal matrix D_i with λ_i on the diagonal. Then it is clear that we can split $D_i + G_i$ into Jordan blocks by forming a Jordan block for every uninterrupted chain of 1s. By putting the bases of the W_i together, we obtain a matrix in Jordan form for f.

Now assume (2). If f can be represented by a Jordan matrix, it is obvious that the diagonal entries are the eigenvalues of f, so they all belong to K. □

Observe that Theorem 22.23 applies if $K = \mathbb{C}$. It turns out that there are uniqueness properties of the Jordan blocks but more machinery is needed to prove this result.

If a complex $n \times n$ matrix A is expressed in terms of its Jordan decomposition as $A = D + N$, since D and N commute, by Proposition 8.21, the exponential of A is given by

$$e^A = e^D e^N,$$

and since N is an $n \times n$ nilpotent matrix, $N^{n-1} = 0$, so we obtain

$$e^A = e^D \left(I + \frac{N}{1!} + \frac{N^2}{2!} + \cdots + \frac{N^{n-1}}{(n-1)!} \right).$$

In particular, the above applies if A is a Jordan matrix. This fact can be used to solve (at least in theory) systems of first-order linear differential equations. Such systems are of the form

$$\frac{dX}{dt} = AX, \tag{$*$}$$

where A is an $n \times n$ matrix and X is an n-dimensional vector of functions of the parameter t.

It can be shown that the columns of the matrix e^{tA} form a basis of the vector space of solutions of the system of linear differential equations $(*)$; see Artin [3] (Chapter 4). Furthermore, for any matrix B and any invertible matrix P, if $A = PBP^{-1}$, then the system $(*)$ is equivalent to

$$P^{-1}\frac{dX}{dt} = BP^{-1}X,$$

so if we make the change of variable $Y = P^{-1}X$, we obtain the system

$$\frac{dY}{dt} = BY. \tag{$**$}$$

Consequently, if B is such that the exponential e^{tB} can be easily computed, we obtain an explicit solution Y of $(**)$, and $X = PY$ is an explicit solution of $(*)$. This is the case when B is a Jordan form of A. In this case, it suffices to consider the Jordan blocks of B. Then we have

$$J_r(\lambda) = \lambda I_r + \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} = \lambda I_r + N,$$

and the powers N^k are easily computed.

For example, if

$$B = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} = 3I_3 + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

we obtain

$$tB = t\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} = 3tI_3 + \begin{pmatrix} 0 & t & 0 \\ 0 & 0 & t \\ 0 & 0 & 0 \end{pmatrix}$$

and so

$$e^{tB} = \begin{pmatrix} e^{3t} & 0 & 0 \\ 0 & e^{3t} & 0 \\ 0 & 0 & e^{3t} \end{pmatrix} \begin{pmatrix} 1 & t & (1/2)t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{3t} & te^{3t} & (1/2)t^2e^{3t} \\ 0 & e^{3t} & te^{3t} \\ 0 & 0 & e^{3t} \end{pmatrix}.$$

The columns of e^{tB} form a basis of the space of solutions of the system of linear differential equations

$$\begin{aligned} \frac{dY_1}{dt} &= 3Y_1 + Y_2 \\ \frac{dY_2}{dt} &= 3Y_2 + Y_3 \\ \frac{dY_3}{dt} &= 3Y_3, \end{aligned}$$

in matrix form,

$$\begin{pmatrix} \frac{dY_1}{dt} \\ \frac{dY_2}{dt} \\ \frac{dY_3}{dt} \end{pmatrix} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}.$$

Explicitly, the general solution of the above system is

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = c_1 \begin{pmatrix} e^{3t} \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} te^{3t} \\ e^{3t} \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} (1/2)t^2e^{3t} \\ te^{3t} \\ e^{3t} \end{pmatrix},$$

with $c_1, c_2, c_3 \in \mathbb{R}$.

Solving systems of first-order linear differential equations is discussed in Artin [3] and more extensively in Hirsh and Smale [35].

22.8 Summary

The main concepts and results of this chapter are listed below:

- Ideals, principal ideals, greatest common divisors.

- Monic polynomial, irreducible polynomial, relatively prime polynomials.
- Annihilator of a linear map.
- Minimal polynomial of a linear map.
- Invariant subspace.
- f-conductor of u into W; conductor of u into W.
- Diagonalizable linear maps.
- Commuting families of linear maps.
- Primary decomposition.
- Generalized eigenvectors.
- Nilpotent linear map.
- Normal form of a nilpotent linear map.
- Jordan decomposition.
- Jordan block.
- Jordan matrix.
- Jordan normal form.
- Systems of first-order linear differential equations.

22.9 Problems

Problem 22.1. Prove that the minimal monic polynomial of Proposition 22.1 is unique.

Problem 22.2. Given a linear map $f\colon E \to E$, prove that the set $\mathrm{Ann}(f)$ of polynomials that annihilate f is an ideal.

Problem 22.3. Provide the details of Proposition 22.9.

Problem 22.4. Prove that the f-conductor $S_f(u, W)$ is an ideal in $K[X]$ (Proposition 22.10).

Problem 22.5. Prove that the polynomials $g_1, \ldots, g_k$ used in the proof of Theorem 22.16 are relatively prime.

Problem 22.6. Find the minimal polynomial of the matrix

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

Problem 22.7. Find the Jordan decomposition of the matrix

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}.$$

Problem 22.8. Let $f\colon E \to E$ be a linear map on a finite-dimensional vector space. Prove that if f has rank 1, then either f is diagonalizable or f is nilpotent but not both.

Problem 22.9. Find the Jordan form of the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Problem 22.10. Let N be a 3×3 nilpotent matrix over $\mathbb{C}$. Prove that the matrix $A = I + (1/2)N - (1/8)N^2$ satisfies the equation

$$A^2 = I + N.$$

In other words, A is a square root of $I + N$.

Generalize the above fact to any $n \times n$ nilpotent matrix N over $\mathbb{C}$ using the binomial series for $(1 + t)^{1/2}$.

Problem 22.11. Let K be an algebraically closed field (for example, $K = \mathbb{C}$). Prove that every 4×4 matrix is similar to a Jordan matrix of the following form:

$$\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}, \qquad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}, \qquad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix},$$

$$\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \mu & 1 \\ 0 & 0 & 0 & \mu \end{pmatrix}, \qquad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

Problem 22.12. In this problem the field K is of characteristic 0. Consider an $(r \times r)$ Jordan block

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

Prove that for any polynomial $f(X)$, we have

$$f(J_r(\lambda)) = \begin{pmatrix} f(\lambda) & f_1(\lambda) & f_2(\lambda) & \cdots & f_{r-1}(\lambda) \\ 0 & f(\lambda) & f_1(\lambda) & \cdots & f_{r-2}(\lambda) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & f_1(\lambda) \\ 0 & 0 & 0 & \cdots & f(\lambda) \end{pmatrix},$$

where

$$f_k(X) = \frac{f^{(k)}(X)}{k!},$$

and $f^{(k)}(X)$ is the kth derivative of $f(X)$.