CHAPTER 3

The Jordan normal form

3.1. The normal form problem

A central problem in linear algebra is the normal form problem, which can be stated as follows.

Let \( f : V \to W \) be a linear map. Find ordered bases \( A \) of \( V \) and \( B \) of \( W \) in which the matrix of \( f \) is "as simple as possible".

This problem has three important special cases.

(i) \( V \) and \( W \) are arbitrary vector spaces and \( A \) and \( B \) are arbitrary ordered bases.
(ii) \( W = V \) and \( B = A \).
(iii) \( W = V^* \) and \( B = A^* \).

Case (i) is the general case. Case (ii) is that of a map from a vector space to itself. Case (iii) is that of a bilinear form on a vector space. (See Theorem 1.21.) In case (i) there is no relationship between the two bases \( A \) and \( B \). In cases (ii) and (iii) we insist on choosing a basis of the target space of \( f \) which is related to the basis of the source space in a reasonable way, namely the same basis if \( W = V \) and the dual basis if \( W = V^* \). In the finite-dimensional case, \( \dim V = n < \infty \), \( \dim W = m < \infty \), these three problems are closely related to the three equivalence relations among matrices introduced in Exercise 1.25. Indeed, the base-change formula \( B = Q^{-1}AP \) allows us to translate problems (i)-(iii) into matrix language as follows:

For each of the three equivalence relations of Exercise 1.25 and for each matrix \( A \), find a matrix \( B \) which is equivalent to \( A \) and which is "as simple as possible".

The matrix \( B \) is a normal form of \( A \) (or of \( f \), if \( A \) is the matrix of \( f \) with respect to some bases). To solve the normal form problem, we need practical criteria for matrices to be equivalent, but we must also explain what we mean by "as simple as possible". This turns out to be fairly easy in case (i), which is deferred to Exercise 3.1. Cases (ii) and (iii) are referred to as the similarity problem and the congruence problem, respectively. These cases are much harder and the answers turn out to depend on the ground field \( F \). In this chapter we will solve case (ii) for algebraically closed fields, which will lead us into the theory of Jordan normal forms.

A Jordan normal form is the next best thing to a diagonal form. Not every linear map from a vector space to itself is diagonalizable, but it has a Jordan normal
form, provided that it has "enough" eigenvalues (more precisely, if its characteristic polynomial splits into linear factors). If a linear map is diagonalizable, its Jordan normal form is equal to its diagonal form. A Jordan normal form is unique up to permutations of its building blocks, which are known as the Jordan blocks.

Jordan normal forms have applications in many areas of pure and applied mathematics, such as difference and differential equations, Markov chains, invariant theory, Lie groups and algebraic groups, etc.

3.2. Eigenvalues and eigenvectors

In this chapter $V$ denotes a fixed vector space of dimension $n$ over a field $F$, and $f$ denotes a fixed linear map from $V$ to itself. Usually we shall assume $n$ to be finite, although a few of the more basic results hold also if $n$ is infinite. (Such results are labelled $n \leq \infty$.) The first step in any type of normal form problem is to try to break objects into smaller pieces. A (linear) subspace $W$ of $V$ is called invariant (under $f$), or $f$-invariant, if $f(W) \subseteq W$. Examples of $f$-invariant subspaces are $\{0\}$, $V$, $\ker f$ and $\im f$. A one-dimensional subspace $W$ is the span of a single nonzero vector $v$ and therefore is invariant if and only if $f(v) = \lambda v$ for some $\lambda \in F$. Thus the notion of a one-dimensional invariant subspace is equivalent to that of an eigenspace of $f$: a nonzero vector $v$ satisfying $f(v) = \lambda v$ for some $\lambda \in F$, called the eigenvalue of $\lambda$. The concept of an eigenspace gives rise to several more examples of invariant subspaces. For $\lambda \in F$ we denote by $V_\lambda$ the corresponding eigenspace

$$V_\lambda = \{ v \in V \mid f(v) = \lambda v \}.$$  

(Note that $V_\lambda = \{0\}$ if $\lambda$ is not an eigenvalue.)

3.1. Lemma. $V_\lambda$ is an invariant subspace.

Proof. This is easy to prove directly, but we will give a slightly indirect proof in order to introduce some helpful notation. Define $f_\lambda$ to be the linear map $f - \lambda \cdot \id_V$, i.e. the map defined by $f_\lambda(v) = f(v) - \lambda v$. Then $V_\lambda = \ker f_\lambda$, so $V_\lambda$ is a subspace. Moreover, note that

$$f = f_0 \quad \text{and} \quad f_\lambda \circ f_\mu = f_\mu \circ f_\lambda$$

for all $\lambda, \mu \in F$. In particular, $f \circ f_\lambda = f_\lambda \circ f$, from which it easily follows that $f(\ker f_\lambda) \subseteq \ker f_\lambda$. QED

Observe also that $f_\lambda(v) = (\mu - \lambda)v$ for $v \in V_\mu$.

The spectrum of $f$, denoted $\spec f$, is the collection of eigenvalues of $f$, i.e. the set of $\lambda \in F$ such that $V_\lambda \neq \{0\}$. The geometric multiplicity $\mult_g(f, \lambda) = \dim V_\lambda$ of any $\lambda \in F$ is the dimension of $V_\lambda$, or equivalently the nullity of $f_\lambda$. The following result is quite easy but incredibly useful. The idea of the proof will appear again several times.

3.2. Theorem ($n \leq \infty$). Let $\lambda_1, \lambda_2, \ldots, \lambda_s \in F$ be distinct eigenvalues of $f$ (i.e. $\lambda_p \neq \lambda_q$ if $p \neq q$). Let $v_i \in V_{\lambda_i}$ and suppose $v_1 + v_2 + \cdots + v_s = 0$. Then $v_1 = v_2 = \cdots = v_s = 0$.  


3.2. EIGENVALUES AND EIGENVECTORS

PROOF. Applying \( f_{\lambda_1} \) to both sides of the identity \( \sum_{r=1}^{s} v_r = 0 \) we get
\[
0 = \sum_{r=1}^{s} f_{\lambda_1}(v_r) = \sum_{r=1}^{s} (\lambda_r - \lambda_1)v_r = \sum_{r=2}^{s} (\lambda_r - \lambda_1)v_r.
\]
Successively applying \( f_{\lambda_2}, f_{\lambda_3}, \ldots, f_{\lambda_{s-1}} \), to this identity we find
\[
0 = (\lambda_3 - \lambda_2-1)(\lambda_3 - \lambda_2-2) \cdots (\lambda_3 - \lambda_1)v_3.
\]
Because the \( \lambda_r \) are distinct, the product \( \prod_{r=1}^{s} (\lambda_3 - \lambda_r) \) is nonzero, and hence \( v_3 = 0 \). Arguing by induction on \( s \) we get \( v_1 = v_2 = \cdots = v_s = 0 \). QED

3.3. COROLLARY \((n \leq \infty)\).

(i) The map \( s: \bigoplus_{\lambda \in \text{spec}} V_\lambda \to V \) sending a tuple \( (v_\lambda)_{\lambda \in \text{spec}} \) to its sum \( \sum_{\lambda \in \text{spec}} v_\lambda \) is injective.

(ii) If \( w_1, w_2, \ldots, w_s \) are eigenvectors of \( f \) with distinct eigenvalues, then \( \{w_1, w_2, \ldots, w_s\} \) is independent.

(iii) \( |\text{spec}| \leq n \).

In particular, if \( V \) is finite-dimensional the spectrum of \( f \) is finite. If \( V \) is infinite-dimensional, the spectrum may be infinite, in which case the space \( \bigoplus_{\lambda \in \text{spec}} V_\lambda \) is a direct sum of infinitely many subspaces. See Exercise 3.2 for an explanation of infinite direct sums.

The map \( f \) is diagonalizable (over \( F \)) if the sum map \( s \) is surjective, i.e. an isomorphism.

3.4. COROLLARY \((n \leq \infty)\). The following assertions are equivalent.

(i) \( f \) is diagonalizable.

(ii) \( V \) has a basis consisting of eigenvectors of \( f \).

(iii) The matrix of \( f \) with respect to a suitable basis is diagonal.

In the finite-dimensional case this has the following corollary, which is a useful sufficient condition for a matrix to be diagonalizable. See Exercise 3.10 for some commentary on this corollary.

3.5. COROLLARY. If \( |\text{spec}| = n < \infty \), then \( f \) is diagonalizable.

PROOF. Let \( \text{spec} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\} \); choose eigenvectors \( v_i \) for each \( \lambda_i \). By Corollary 3.3(ii) the set \( B = \{v_1, v_2, \ldots, v_n\} \) is independent. Hence \( B \) is a basis because \( |B| = n = \dim V \). So \( f \) is diagonalizable by Corollary 3.4. QED

Diagonal matrices stand out as desirable candidates for normal forms. The reason is that many operations of linear algebra, such as solving systems \( (A_x = b) \), raising to a power \( (A^k) \) and exponentiating \( (e^A) \), are easy to perform on diagonal matrices. Sadly, not all linear maps are diagonalizable.

3.6. EXAMPLE \((n = \infty)\). Let \( V = C^\infty(\mathbb{R}) \), the space of smooth (infinitely differentiable) functions \( \phi: \mathbb{R} \to \mathbb{R} \). The differentiation operator \( D \) is defined by \( D(\phi) = \phi' \). For \( \lambda \in \mathbb{R} \) put \( \phi_\lambda(x) = e^{\lambda x} \). From calculus you know that \( D(\phi_\lambda) = \lambda \phi_\lambda \), so \( \phi_\lambda \in V_\lambda \). From Corollary 3.3 we now immediately draw the following conclusions.

(i) \( \{\phi_\lambda \mid \lambda \in \mathbb{R}\} \) is independent.
(ii) \( \dim V \) is uncountable.
(iii) The sum map \( s: \bigoplus_{\lambda \in \mathbb{R}} V_\lambda \to V \) is injective.

What are the eigenspace and the geometric multiplicity of each eigenvalue? To determine \( V_\lambda \) we must find the general solution of \( \phi' = \lambda \phi \). If \( \phi \) is a solution, then putting \( \psi(x) = \phi(x)/\phi(\lambda) = e^{-\lambda x} \phi(x) \) gives \( \psi' = 0 \), and so \( \psi = c \), a constant. Hence \( \phi = c\phi_\lambda \). This just means \( V_\lambda = \text{span}(\phi_\lambda) \) and therefore \( \text{mult}_\lambda(D, \lambda) = 1 \) for all \( \lambda \in \mathbb{R} \). Finally, is \( D \) diagonalizable? If it were, we could write any smooth function, say \( \phi(x) = x^2 \), as a (finite!) linear combination \( x = \sum_{r=1}^s c_r e^{\lambda_r x} \). A little fiddling shows that this cannot be done. (For instance, divide by \( x \) and let \( x \to \infty \) to get a contradiction.) We conclude that \( D \) is not diagonalizable.

3.7. EXAMPLE. Let \( F = \mathbb{R} \) and \( V = \mathbb{R}^2 \). The matrix

\[
A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}
\]

defines a linear map \( V \to V \), namely the rotation through 90° about the origin. No line in the plane is mapped into itself, so there are no one-dimensional invariant subspaces, and hence no eigenvectors. Hence \( A \) is not diagonalizable.

3.8. EXAMPLE. Let \( F = \mathbb{R} \), \( V = \mathbb{R}^2 \) and \( \mu \in \mathbb{R} \). The matrix

\[
A = \begin{pmatrix} \mu & 1 \\ 0 & \mu \end{pmatrix}
\]

defines a linear map \( V \to V \), namely a shear transformation. The only line in the plane mapped into itself is the horizontal axis, and \( A e_1 = \mu e_1 \). So the only eigenvalue is \( \mu \) and the eigenspace is \( V_\mu = \text{span}(e_1) \). Hence \( A \) is not diagonalizable.

These examples show there are at least three obstacles to a map being diagonalizable.

(i) \( f \) has plenty of eigenvalues and eigenvectors, but \( V \) is infinite-dimensional.
(ii) \( f \) has not enough eigenvalues.
(iii) \( f \) has not enough eigenvectors.

Obstacle (i) is surmounted by replacing direct sums with direct integrals. This involves functional analysis, which is beyond the scope of this course. Obstacle (ii) can be cleared by enlarging the field to a field which contains all the eigenvalues of \( f \) and correspondingly enlarging the vector space \( V \), as explained in Section 1.10. One drastic option is to replace \( F \) with its algebraic closure \( \overline{F} \). (E.g. in Example 3.7 replace \( \mathbb{R} \) with \( \mathbb{C} \) and \( \mathbb{R}^2 \) with \( \mathbb{C}^2 \).) Obstacle (iii) cannot be circumvented. We must extend our notion of normal form to include so-called Jordan normal forms. We will do this in the next section, after proving a final useful fact concerning eigenvalues. The characteristic polynomial of \( f \) is the polynomial \( \chi_f(x) \in F[x] \) defined by

\[
\chi_f(x) = \det(A - xI),
\]

where \( A \) is the matrix of \( f \) with respect to any ordered basis of \( V \). Note \( A - xI \) is a matrix whose entries are polynomials (of degree \( \leq 1 \)), i.e. an element of \( M_n[F[x]] \), so \( \det(A - xI) \) is indeed a polynomial. Also, it follows from the base-change formula
and the product formula for determinants that \( \chi_f \) does not depend on the choice of the basis of \( V \).

3.9. THEOREM. An element of \( F \) is an eigenvalue of \( f \) if and only if it is a root of \( \chi_f \).

PROOF. \( \lambda \in \text{spec } f \iff \text{nullity } f_{\lambda} > 0 \iff \det f_{\lambda} = 0 \iff \det A_{\lambda} = 0 \iff \chi_f(\lambda) = 0 \). Here \( A \) is the matrix of \( f \) with respect to any ordered basis of \( V \), and \( A_{\lambda} = A - \lambda I \).

A more general statement is the following.

3.10. THEOREM. Let \( W \) be an invariant subspace of \( V \). Then the characteristic polynomial of \( f|W \) divides the characteristic polynomial of \( f \).

PROOF. See Exercise 3.11.

QED

3.3. Generalized eigenvectors

The vector \( e_2 \) in Example 3.8 is not an eigenvector, because \( f_{\mu}(e_2) = e_1 \neq 0 \). However, it satisfies \( f_{\mu}^2(e_2) = f_{\mu}(e_1) = 0 \), because \( e_1 \) is an eigenvector. Convention: we will use \( f_{\mu}^k \) as a shorthand for

\[
(f_{\mu})^2 = (f - \mu \text{id}_V)^2 = f^2 - 2\mu f + \mu^2 \text{id}_V.
\]

Don't confuse this with \( f^2 \).

A generalized eigenvector with eigenvalue \( \lambda \in F \) is a nonzero vector \( v \in V \) satisfying \( f_{\lambda}^k(v) = 0 \) for some \( k \geq 1 \). The generalized eigenspace corresponding to \( \lambda \) is

\[
V(\lambda) = \{ v \in V \mid f_{\lambda}^k(v) = 0 \text{ for some } k \geq 1 \}.
\]

The following result is a generalization of Theorem 3.2.

3.11. THEOREM \((n \leq \infty)\). (i) \( V(\lambda) \) is an \( f_{\mu} \)-invariant subspace for every \( \lambda \) and \( \mu \) in \( F \).

(ii) \( f_{\mu}|V(\lambda) \) is injective if \( \mu \neq \lambda \).

(iii) Let \( \lambda_1, \lambda_2, \ldots, \lambda_s \) be distinct elements of \( F \). Let \( v_r \in V(\lambda_r) \) and suppose \( v_1 + v_2 + \cdots + v_s = 0 \). Then \( v_1 = v_2 = \cdots = v_s = 0 \).

PROOF. For \( k \geq 1 \) let \( V_k \) be the subspace ker \( f_{\lambda}^k \). By definition \( V(\lambda) \) is the union \( \bigcup_{k=1}^{\infty} V_k \). Moreover, the subspaces \( V_k \) are increasing in the sense that \( V_k \subseteq V_l \) if \( k \leq l \). By Exercise 3.6 this implies \( V(\lambda) \) is a subspace. The fact that \( f_{\lambda} \circ f_{\mu} = f_{\mu} \circ f_{\lambda} \) implies that \( f_{\mu}(V_k) \subseteq V_k \) for all \( k \) and hence \( f_{\mu}|V(\lambda) \subseteq V(\lambda) \).

Assume \( \mu \neq \lambda \). Let \( v \in V(\lambda) \) be a nonzero vector satisfying \( f_{\mu}(v) = 0 \). Let \( k \) be the minimal positive integer \( l \) such that \( f_{\lambda}^l(v) = 0 \). Then \( k < \infty \) because \( v \in V(\lambda) \), and \( w = f_{\lambda}^{k-1}(v) \neq 0 \) because \( k \) is minimal. Moreover, \( f_{\lambda}(w) = f_{\lambda}^{k}(v) = 0 \), so \( w \in V_{\lambda} \). Hence \( f_{\mu}(w) = (\lambda - \mu)w \), wherefore

\[
0 \neq (\lambda - \mu)w = f_{\mu}(w) = f_{\mu}(f_{\lambda}^{k-1}(v)) = f_{\lambda}^{k-1}(f_{\mu}(v)) = f_{\lambda}^{k-1}(0) = 0,
\]

which is a contradiction. We conclude that \( f_{\mu}|V(\lambda) \) has kernel \( \{0\} \); hence it is injective.
3. THE JORDAN NORMAL FORM

For \( v_r \in V(\lambda_r) \) let \( k_r \) be a positive integer such that \( f_{\lambda_r}^{k_r}(v_r) = 0 \). Suppose 
\[ \sum_{r=1}^s v_r = 0; \]
apply successively \( f_{\lambda_1}^{k_1}, f_{\lambda_2}^{k_2}, \ldots, f_{\lambda_s}^{k_s-1} \) to both sides of this identity to get
\[ f_{\lambda_1}^{k_1-1} \circ f_{\lambda_2}^{k_2-1} \circ \cdots \circ f_{\lambda_s}^{k_s-1}[v_s] = 0. \]
It follows from part (i) that \( f_{\lambda_i}^{k_i} \circ f_{\lambda_i}^{k_i-1} \circ \cdots \circ f_{\lambda_i}^{k_i-1}[v_s] \in V(\lambda_i) \) for all \( r \leq s - 1 \).
Applying part (ii) \( s - 1 \) times and using \( \lambda_r \neq \lambda_s \) for \( r < s \) we conclude \( v_s = 0 \). By 
induction on \( s \) we get \( v_1 = v_2 = \ldots = v_s = 0 \). QED

And here is the analogue of Corollary 3.3.

3.12. COROLLARY \( (n \leq \infty) \).

(i) The map \( s: \bigoplus_{\lambda \in \text{spec} f} V(\lambda) \to V \) sending a tuple \( (v_\lambda)_{\lambda \in \text{spec} f} \) to its sum \( \sum_{\lambda \in \text{spec} f} v_\lambda \) is injective.

(ii) If \( v_1, v_2, \ldots, v_s \) are generalized eigenvectors of \( f \) with distinct 
eigenvalues, then \( \{v_1, v_2, \ldots, v_s\} \) is independent.

Now that we know that the generalized eigenspaces are independent of each 
other, we will look for collections of independent vectors inside each individual 
generalized eigenspace. A string in \( V \) is an \([\nu]-\text{tuple of vectors}\)
\[ S = (v_1, v_2, \ldots, v_\nu) \]
(where \( \nu \) is allowed to be countably infinite) subject to the following requirements:

(i) \( v_1 \) is an eigenvector with eigenvalue \( \lambda \), i.e. \( f_\lambda[v_1] = 0 \) and \( v_1 \neq 0 \);
(ii) \( f_\lambda(v_{k+1}) = v_k \) for \( 1 \leq k \leq \nu - 1 \).

We call \( v_1 \) the leading vector, \( \nu \) the length, and \( \lambda \) the eigenvalue of \( S \).

3.13. LEMMA. span \( S \) is \( f_\mu \)-invariant for every \( \mu \in F \).

PROOF. For simplicity let us put \( v_0 = 0 \). Then for \( 0 \leq k \leq \nu - 1 \) we have 
\[ f_\lambda(v_{k+1}) = v_k \], so \( f_\lambda(v_{k+1}) = \lambda v_{k+1} + v_k \) and therefore \( f_\mu(v_{k+1}) = (\lambda - \mu)v_{k+1} + v_k \). 
Thus \( f_\mu \) maps \( S \) into span\( S \). Hence, by linearity, \( f_\mu(\text{span} S) \subseteq \text{span} S \). QED

Conditions (i) and (ii) show that the string \( S \) is contained in the generalized 
eigenspace \( V(\lambda) \). One finds a string starting at an eigenvector \( v_1 \in V_\lambda \) by successively 
solving the inhomogeneous linear equations \( f_\lambda(v_{k+1}) = v_k \) for \( k = 1, 2, \ldots \). 
When this equation is no longer solvable, the string must end. There may exist 
many different strings through a given \( v_1 \), and these strings may have different 
lengths. It is possible for a string to have length \( \nu \), in which case it consists of \( \lambda \) eigenvectors. In an infinite-dimensional space, strings may be infinitely long, 
but note that if \( \nu \) is finite and \( w = v_\nu \) is the last vector in the string, then
\[ S = (f_\lambda^{-1}(w), f_\lambda^{-2}(w), \ldots, f_\lambda^0(w), w). \] (3.1)

It is a pleasant surprise that strings are always independent (and hence finitely 
long if \( n < \infty \)). Indeed, the following variant of Theorem 3.11 shows that the 
union of any collection of strings is independent as soon as their leading vectors 
are independent.
3.14. **Theorem** \((n \leq \infty)\). For each \(i \in \text{index set I}\) let \(S(i)\) be a string in \(V\). Let \(l(i)\) be the length of \(S(i)\), \(\lambda(i)\) its eigenvalue, and \(v_1(i)\) its leading vector. If the collection of leading vectors \(\{v_1(i) \mid i \in I\}\) is independent, then the collection \(\bigcup_{i \in I} S(i)\) is independent.

**Proof.** It follows from Theorem 3.11 that strings with different eigenvalues are independent of each other. We may therefore assume without loss of generality that all eigenvalues are the same, \(\lambda(i) - \lambda\) for all \(i \in I\).

Suppose

\[
\sum_{i \in I} \sum_{k=1}^{l(i)} \mu_k(i) v_k(i) = 0
\]

(3.2)

is a linear relation among elements of all the strings \(S(i)\), where \(\mu_k(i) \in F\). We are neither assuming that \(I\) is finite nor that \(l(i)\) is finite, but for the combination (3.2) to be finite we assume that \(\mu_k(i) = 0\) for all but finitely many \(i \in I\) and \(k \geq 1\). Let us suppose that \(\mu_k(i) \neq 0\) for some \(i\) and some \(k\), and try to arrive at a contradiction. For each \(i \in I\) let \(m(i)\) be the maximal \(k \geq 1\) such that \(\mu_k(i) \neq 0\). Also let \(m\) be the maximum of \(m(i)\) over all \(i \in I\). Then \(m(i)\) and \(m\) are both finite because only finitely many \(\mu_k(i)\) are nonzero. If we apply the map \(f_{\lambda}^{m-1}\) to both sides of (3.2) and use that \(f_{\lambda}^{m-1}(v_k(i)) = 0\) for \(l \geq k\), we see that all terms on the left are killed, except if \(i\) is such that \(m(i) = m\) and if \(k = m\). Thus we get

\[
0 = \sum_{i \in I} \mu_m(i) f_{\lambda}^{m-1}(v_m(i)) - \sum_{i \in I} \mu_m(i) v_1(i),
\]

where each \(\mu_m(i)\) is nonzero. But this contradicts the assumption that the \(v_1(i)\)'s are independent. \(\quad\) QED

Applying the theorem to a single string we immediately get the following.

3.15. **Corollary** \((n \leq \infty)\). Every string \(S\) in \(V\) is linearly independent. Hence \(S\) is a basis of \(\text{span}\overline{S}\) and the length of \(S\) is at most \(n\). In particular, the length is finite if \(n < \infty\).

Because of Lemma 3.13 and Corollary 3.15 it makes sense to write the matrix of \(f\) (or strictly speaking the restriction of \(f\) to \(\text{span}\overline{S}\)) relative to any string \(S\). It is the Jordan block with eigenvalue \(\lambda\) of size \(1\),

\[
J_{\lambda,1} = \begin{pmatrix}
\lambda & 1 \\
& \ddots & \ddots \\
& & \lambda \\
& & & 1
\end{pmatrix},
\]

(3.3)

(The word “eigenvalue” is used here in a generalized sense; it just means the scalar \(\lambda\) on the diagonal. The dots indicate uninterrupted sequences of \(\lambda\)'s, resp. \(1\)'s, and all vacant slots are equal to 0.) For future reference we also write the matrix of \(f_{\lambda}\) relative to \(S\), which is found by subtracting \(\lambda I\) from \(J_{\lambda,1}\), and the matrices of the
powers of $f_{\lambda}$:

$$J_{0,1} = \begin{pmatrix} 0 & 1 \\ 0 & \ddots & 1 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots & 1 \\ 0 & \ddots & \ddots & \ddots & 0 \end{pmatrix}, \quad J_{0,2}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \end{pmatrix}, \quad J_{0,1}^3 = 0. \quad (3.4)$$

A Jordan basis for $f$ is an ordered basis $B$ of $V$ such that the matrix of $f$ with respect to $B$ is a Jordan matrix

$$J = \begin{pmatrix} J_1 \\ J_2 \\ \vdots \\ J_q \end{pmatrix}, \quad (3.5)$$

where $J_p = J_{x_p,1_p}$ is the Jordan block with eigenvalue $\lambda_p$ of size $l_p$, both of which may depend on $p$. A Jordan basis is nothing other than a basis $B$ which can be written as a union $S_1 \cup S_2 \cup \cdots \cup S_q$ of mutually disjoint strings $S_p$. The Jordan matrix $J$ is called a Jordan normal form of $f$. Theorem 3.14 suggests a strategy for finding a Jordan basis: find all the eigenvalues of $f$, for each eigenvalue determine as many independent eigenvectors as possible, and calculate strings starting at each of the eigenvectors. The resulting strings will form an independent set. The problem is that this set will not span $V$ unless the strings are well chosen. We will see how this is done in Section 3.4.

Note that by changing the order of the strings $S_p$ in a Jordan basis $B$ we can always permute the Jordan blocks in $J$. Usually we group the Jordan blocks with the same eigenvalues together. Then, if $\lambda_1, \lambda_2, \ldots, \lambda_s$ are the distinct eigenvalues of $f$, the Jordan matrix looks like

$$J = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_s \end{pmatrix}, \quad (3.6)$$

where each block $M_r$ contains all of the Jordan blocks with eigenvalue equal to $\lambda_r$. Within each block $M_r$ the Jordan blocks are commonly arranged in non-increasing order according to size. (It may happen that $M_r$ contains several Jordan blocks of the same size.) If the field $F$ is ordered (i.e. $F$ is a subfield of $\mathbb{R}$), the eigenvalues are often numbered in decreasing order: $\lambda_1 > \lambda_2 > \cdots > \lambda_s$, which completely removes the ambiguity in the order of the Jordan blocks.

### 3.4. Existence and uniqueness

The following statement imposes a constraint on the existence of a Jordan normal form for $f$.
3.16. Lemma. Suppose $f$ has a Jordan normal form $J$ as in (3.6). Then
\[ \chi_f(x) = \prod_{i=1}^{g} (x - \lambda_i)^{n_i}, \]
where $n_r$ is the size of $M_r$. In particular, the characteristic polynomial $\chi_f$ is a product of linear factors.

Proof. See Exercise 3.16. QED

In other words, if $f$ has a Jordan normal form the polynomial $\chi_f$ has $n$ roots in the field $F$ (counted with multiplicities). So if $\chi_f$ does not split into linear factors (which may well happen if $F$ is not algebraically closed), $f$ cannot have a Jordan normal form. The Jordan normal form theorem is the converse of this assertion.

3.17. Theorem (Jordan normal form theorem). Assume that the characteristic polynomial $\chi_f$ splits into linear factors. Then there exists a Jordan normal form for $f$. The Jordan normal form is unique up to the order of the Jordan blocks.

Proof. We prove the existence of a Jordan basis by induction on $n$. For $n = 0$ the empty set is a Jordan basis. Now let $n \geq 1$ and assume the result is proved for all vector spaces of dimension $< n$.

The assumption that $\chi_f$ splits into linear factors implies that $f$ has at least one eigenvalue. Let $\lambda$ be any eigenvalue and let $W = \text{im} f_{\lambda}$ be the image of $f_{\lambda}$. Then $W$ is an $f$-invariant subspace of dimension $r = \text{rank} f_{\lambda}$. Note that $r = n - \text{nullity} f_{\lambda} < n$, because $f_{\lambda}$ is not injective. Let $g = f|W; W \rightarrow W$ be the restriction of $f$ to $W$. By Theorem 3.10, the characteristic polynomial of $g$ divides that of $f$, and therefore $\chi_g$ splits into linear factors. Hence, by the induction hypothesis, there exists a Jordan basis $C$ for $g$. We can write $C$ as a disjoint union $T_1 \cup T_2 \cup \cdots \cup T_q$ of strings for $g$. We are going to extend $C$ to a Jordan basis for $f$. This involves two steps: 1. some of the strings in $C$ have to be extended by one element in $V$; 2. we have to add a few strings of length 1 to $C$.

Step 1. Consider the $\lambda$-eigenspace of $g$, $W_{\lambda} = \ker g_{\lambda}$. Let $m = \text{mult}_{\lambda}(g, \lambda)$ be its dimension. Since elements of $W_{\lambda}$ are just eigenvectors of $f$ that happen to lie in $W$, we have $W_{\lambda} = W \cap V_{\lambda}$. Let us find a basis of $W_{\lambda}$. By Exercise 3.15 (applied to the linear map $g$), a basis of $W_{\lambda}$ is given by the leading vectors of those strings in $C$ which have eigenvalue equal to $\lambda$. Because $m = \dim W_{\lambda}$, there have to be $m$ such strings. After renumbering the strings in $C$ if necessary, we may assume that the first $m$ strings $T_1, T_2, \ldots, T_m$ have eigenvalue $\lambda$. Let \( v_1, v_2, \ldots, v_m \) be the last vectors in these strings. As these vectors are in $W = \text{im} f_{\lambda}$, there exist $x_1, x_2, \ldots, x_m \in V$ such that $f_{\lambda}(x_k) = v_k$ for $1 \leq k \leq m$. For $1 \leq k \leq m$ define $S_k = [T_k, x_k]$, i.e. append $x_k$ to the string $T_k$ to create a string which is longer by one element.

Step 2. Consider the $\lambda$-eigenspace of $f$, $V_{\lambda} = \ker f_{\lambda}$. Its dimension is $n - r$ and it contains the $m$-dimensional space $W_{\lambda}$. Let $v_1, v_2, \ldots, v_m$ be the leading vectors of the strings $T_1, T_2, \ldots, T_m$. As we saw under step 1, $\{v_1, v_2, \ldots, v_m\}$ is a basis of $W_{\lambda}$. Extend this to a basis
\[ \{v_1, v_2, \ldots, v_m, u_1, u_2, \ldots, u_{n-r-m}\} \]
of $V_{\lambda}$. For $1 \leq k \leq n - r - m$ define $R_k = \{u_k\}$. 
We assert that
\[ B = R_1 \cup R_2 \cup \cdots \cup R_{n-r} \cup S_1 \cup S_2 \cup \cdots \cup S_m \cup T_{m+1} \cup \cdots \cup T_q \]
is a Jordan basis for \( f \). Indeed, the \( R \)'s, \( S \)'s and \( T \)'s are strings and their leading vectors are by construction independent. Therefore \( B \) is independent by Theorem 3.14. Since \( B \) is obtained from \( C \) by adding \( n - r - m = n - r \) vectors, the cardinality of \( B \) is \( |B| = n - r + |C| = n - r + r = n \), so \( B \) is a basis of \( V \). We have now found a Jordan basis for \( f \) and thereby completed the inductive step.

Next we prove the uniqueness. Let \( J \) be the matrix of \( f \) with respect to a Jordan basis \( B \). For each \( \mu \in F \) and for each integer \( m \geq 1 \) let \( d_{\mu,m} \) be the number of Jordan blocks in \( J \) with eigenvalue \( \mu \) and of size \( m \geq 1 \). (If \( \mu \) is not an eigenvalue of \( f \), we put \( d_{\mu,m} = 0 \).) The indices \( d_{\mu,m} \) determine the matrix \( J \) up to the order of the blocks (because the number of blocks with eigenvalue \( \mu \) and size equal to \( m \) is \( d_{\mu,m} - d_{\mu,m-1} \)). To show that \( J \) is unique up to the order of the blocks, it therefore suffices to show that the numbers \( d_{\mu,m} \) depend only on the linear map \( f \), not on the choice of the Jordan basis \( B \). We do this by proving that for all \( \mu \in F \) and all \( m \geq 1 \) we have
\[ d_{\mu,m} = c_{\mu,m}, \tag{3.7} \]
where \( c_{\mu,m} = \dim(\ker f_{\mu} \cap \text{im} f_{\mu}^{m-1}) \).

Let us first show that (3.7) holds under the assumption that \( B \) consists of a single string with eigenvalue \( \lambda \) of length \( l \). Then \( J \) is the single block \( I_{\lambda,1}, \) so \( d_{\lambda,1} = 1 \) and \( d_{\lambda,m} = 0 \) if \( \mu \neq \lambda \) or \( m > 1 \). The matrix of \( f_{\mu} \) is \( J_{\lambda,1} - \mu I_{1} = J_{\lambda,\mu,1} \). The nullspace of \( J_{\lambda,\mu,1} \) is \( [0] \) if \( \lambda \neq \mu \) and it is spanned by the standard basis vector \( e_{1} \) if \( \lambda = \mu \). In particular, \( c_{\mu,m} = 0 = d_{\mu,m} \) if \( \mu \neq \lambda \). Now assume \( \mu = \lambda \). Then the matrix of \( f_{\mu} \) is \( J_{0,1} \) and we have for all \( l \geq 1 \)
\[ \text{im} J_{0,1}^{m-1} = \begin{cases} \text{span}(e_{1}, e_{2}, \ldots, e_{1-m+1}) & \text{if } m \leq l, \\ \{0\} & \text{if } m > l. \end{cases} \]
(This follows from (3.4).) Thus \( \ker J_{0} \cap \text{im} J_{0}^{m-1} = \text{span}(e_{1}) \) for \( m \leq l \) and \( \ker J_{0} \cap \text{im} J_{0}^{m-1} = \{0\} \) for \( m > l \). Consequently \( c_{\lambda,m} = 1 \) for \( m \leq l \) and \( c_{\lambda,m} = 0 \) for \( m > l \), which proves (3.7) in this case.

To establish the general case, where \( B \) consists of a number of strings \( S_1, S_2, \ldots, S_q \), we will write \( h = f_{\mu}, V_{p} = \text{span} S_{p}, \) and \( h_{p} = h|V_{p} \), and show that
\[ \ker h \cap \text{im} h^{m-1} = \bigoplus_{p=1}^{q} \ker h_{p} \cap \text{im} h_{p}^{m-1}. \tag{3.8} \]
Because of the additivity of dimension over direct sums (see Exercise 1.10) the general case then follows from the case \( q = 1 \). Let \( v \in \ker h \cap \text{im} h^{m-1} \). Since \( V = \bigoplus_{p=1}^{q} V_{p} \), \( v \) can be written uniquely as a sum \( v = \sum_{p=1}^{q} v_{p} \) with \( v_{p} \in V_{p} \). We need to show that \( v_{p} \in \ker h_{p} \cap \text{im} h_{p}^{m-1} \). Since \( h(v) = 0 \) and \( V_{p} \) is \( h \)-invariant, we have \( h(v_{p}) = 0 \) for each \( p \). Since \( v \in \text{im} h^{m-1} \), we have \( v = h^{m-1}(u) \) for some \( u \in V \). Write \( u = \sum_{p=1}^{q} u_{p} \) with \( u_{p} \in V_{p} \). Then \( v = \sum_{p=1}^{q} h^{m-1}(u_{p}) \). Since \( u_{p} \in V_{p} \), we have \( h^{m-1}(u_{p}) = h_{p}^{m-1}(u_{p}) \in V_{p} \), so, by the uniqueness of the \( v_{p} \), \( v_{p} = h_{p}^{m-1}(u_{p}) \). The conclusion is that \( v_{p} \in \ker h_{p} \cap \text{im} h_{p}^{m-1} \). QED
3.18. **Corollary.** Assume that \( \chi_f \) splits into linear factors. Then the map 
\( s: \bigoplus_{\lambda \in \text{Spec} \ F} V(\lambda) \to V \) defined in Corollary 3.12(i) is an isomorphism.

**Proof.** Corollary 3.12(i) says that \( s \) is injective, so it remains to show that \( s \) is surjective. Let \( v \in V \). By Theorem 3.17 there exists a Jordan basis \( B \) for \( f \). Each element of \( B \) is a generalized eigenvector, i.e. an element of \( V(\lambda) \) for some \( \lambda \in \text{Spec} \ F \). Therefore \( v \) can be written as a sum of elements of the \( V(\lambda) \)'s. We conclude that \( v \) is in the image of \( s \).

QED

3.19. **Corollary.** If \( F \) is algebraically closed, every linear map \( f: V \to V \) has a Jordan normal form.

**Proof.** If \( F \) is algebraically closed, every polynomial in \( F[x] \), in particular \( \chi_f \), splits into linear factors. Now apply Theorem 3.17.

QED

The Jordan normal form theorem is a splendid result, but as always in linear algebra it is good to have a coordinate-free version, i.e. a version that does not refer to a choice of basis. Luckily, such a version is near at hand. Recall that \( f \) is diagonalizable if its matrix with respect to a suitable basis is diagonal. Note that a diagonal matrix is a Jordan matrix in which every Jordan block has size 1. We say \( f \) is nilpotent if \( f^k = 0 \) for some \( k \geq 1 \).

3.20. **Theorem** (Jordan decomposition theorem). Assume that \( \chi_f \) splits into linear factors. Then there exist unique linear maps \( f_{\text{diag}} \) and \( f_{\text{nil}} \) satisfying the following conditions:

(i) \( f = f_{\text{diag}} + f_{\text{nil}} \) and \( f_{\text{diag}} \circ f_{\text{nil}} = f_{\text{nil}} \circ f_{\text{diag}} \).

(ii) \( f_{\text{diag}} \) is diagonalizable and \( f_{\text{nil}} \) is nilpotent.

**Proof.** See Exercise 3.22.

QED

A comment on what happens if \( \chi_f \) does not split into linear factors. This can happen if \( F \) is not algebraically closed (e.g. if \( F = \mathbb{F}_p, \mathbb{Q} \), or \( \mathbb{R} \)). But in some ways this is not a very serious drawback, because it is possible to enlarge the field \( F \) so that \( \chi_f \) splits into linear factors in the larger field, and Theorem 3.17 then says that \( f \) has a Jordan normal form over this larger field. For example, if \( F = \mathbb{R} \), it may happen that not all the eigenvalues of \( f \) are real. In this case \( f \) does not have a Jordan normal form over \( \mathbb{R} \), but it always has a Jordan normal form over \( \mathbb{C} \). However, in some applications extending the scalars is not a good thing to do, and in such cases one is forced to look for other kinds of normal forms. See for example Exercise 3.31 for a real version of the Jordan normal form theorem.

The Jordan normal form theorem leads to a criterion for similarity of matrices. For \( \lambda \in \text{Spec} \ F \) and \( t \geq 1 \) let \( d_{\lambda, t}(f) = \dim(\ker f_{\lambda} \cap \ker f_{\lambda}^{t-1}) \) be the number of Jordan blocks with eigenvalue \( \lambda \) and size \( \geq t \). If \( A \) is an \( n \times n \)-matrix, we put \( d_{\lambda, 1}(A) = d_{\lambda, 1}(f_A) \), where \( f_A: F^n \to F^n \) is the linear map given by \( f_A(x) = Ax \).

3.21. **Theorem.** Assume that \( F \) is algebraically closed. Let \( A \) and \( B \) be in \( M_n[F] \). Then the following statements are equivalent:

(i) \( A \) and \( B \) are similar.

(ii) \( d_{\lambda, 1}(A) = d_{\lambda, 1}(B) \) for all \( \lambda \in \text{Spec} \ F \) and \( t \geq 1 \).
3. The Jordan Normal Form

(iii) $A$ and $B$ have the same Jordan normal form up to the order of the blocks.

**Proof.** If $A$ and $B$ are similar, then $B = P^{-1}AP$ for an invertible $P$ and so $B_\lambda = P^{-1}A_\lambda P$ for all $\lambda$. Thus we have, for all vectors $v$ and $w \in V$, $B_\lambda v = 0$ if and only if $A_\lambda v = 0$ and $A_\lambda v^k = PB_\lambda v^k P^{-1}w$. These identities imply $\ker A_\lambda = P(\ker B_\lambda)$ and $\im A_\lambda^k = P(\im B_\lambda^k)$ for all $\lambda \in F$ and $k \geq 0$. Hence $d_{\lambda,1}(A) = d_{\lambda,1}(B)$. So $i$ implies $ii$.

Since $F$ is algebraically closed, for every square matrix $A$ there exist a Jordan basis $B_A$ of $F^n$ and a Jordan normal form $J_A$. The implication $ii \implies iii$ now follows from the fact that the numbers $d_{\lambda,1}(A)$ determine $J_A$ up to the order of the blocks.

The base-change formula says that $A = P_A^{-1}J_A P_A$, where $P_A$ is the base-change matrix from the standard basis to $B_A$. If $J_A$ is the same as $J_B$ up to the order of the blocks, it is easy to see that $J_A$ and $J_B$ are similar (via a permutation matrix).

From the fact that similarity is an equivalence relation (see Exercise 1.25) we now infer that $A$ and $B$ are similar. Thus $iii$ implies $i$. QED

The *algebraic multiplicity* $\mult_{a}(f, \lambda)$ of any $\lambda \in F$ is $\mult(\chi_f, \lambda)$, i.e. the multiplicity of $\lambda$ considered as a root of the characteristic polynomial. (See Exercise 1.29.)

3.22. **Theorem.**

(i) $\mult_{a}(f, \lambda) \leq \mult_{g}(f, \lambda)$ for all $\lambda \in F$.

(ii) Suppose $f$ has a Jordan normal form. Then $f$ is diagonalizable if and only if $\mult_{g}(f, \lambda) = \mult_{a}(f, \lambda)$ for all $\lambda \in F$.

**Proof.** If $\mult_{g}(f, \lambda) = m$, then the characteristic polynomial of $f|_{V_\lambda}$ is $(\lambda - \mu)^m$ and therefore, by Theorem 3.10, $\lambda$ is a root of $\chi_f$ of multiplicity at least $m$.

By Lemma 3.16, $\mult_{a}(f, \lambda)$ is the sum of the sizes of all Jordan blocks in the Jordan normal form of $f$ with eigenvalue $\lambda$. On the other hand, $\mult_{g}(f, \lambda)$ is the number of all such Jordan blocks. Hence $\mult_{g}(f, \lambda) = \mult_{a}(f, \lambda)$ if and only if each Jordan block with eigenvalue $\lambda$ has size 1. Therefore $\mult_{g}(f, \lambda) = \mult_{a}(f, \lambda)$ for all $\lambda$ is equivalent to $f$ being diagonalizable. QED

3.5. The minimal polynomial

We now discuss the Cayley–Hamilton theorem as an application of the Jordan normal form theorem. The proof uses the minimal (or minimum) polynomial of $f$, which has many striking properties in its own right.

The *annihilator* of $f$ is the set of polynomials $p \in F[x]$ that satisfy $p(f) = 0$ (where the zero indicates the zero linear map from $V$ to $V$). It is denoted by $I_f$. The *minimal polynomial* $m_f$ of $f$ is the element of the annihilator characterized as follows: $m_f$ is monic (i.e. has leading coefficient 1) and $\deg m_f \leq \deg p$ for all nonzero $p \in I_f$.

3.23. **Lemma.**

(i) $I_f$ contains the zero polynomial and is closed under addition. If $p \in I_f$, then $pq \in I_f$ for all polynomials $q \in F[x]$.

(ii) $I_{f^{-1} \circ g} = I_f$ for every invertible linear mapping $g : V \to V$. 


(iii) Every element of \( I_f \) is divisible by \( m_f \). Hence \( I_f = \{ qm_f \mid q \in F[x] \} \).

**Proof.** (i) follows from \( 0(f) = 0, (p + q)(f) = p(f) + q(f) \) and \( (pq)(f) = p(f)q(f) \).

(ii) follows from the fact that \( p(g^{-1} \circ f) = g^{-1} \circ p(f) \circ g \) for every \( p \in F[x] \).

Let \( p \in I_f \) and write \( p = qm_f + r \) with \( \text{deg} r < \text{deg} m_f \) (Euclidean algorithm). Then \( 0 = p(f) = q(f)m_f(f) + r(f) = r(f) \) because \( p \) and \( m_f \) annihilate \( f \). Hence \( r \in I_f \). Since \( m_f \) is the nonzero polynomial of minimal degree in \( I_f \) we conclude \( r = 0 \). So \( p \) is divisible by \( m_f \).

QED

For those who know a little ring theory: part (i) says that \( I_f \) is an ideal in the ring \( F[x] \) and part (iii) says that \( m_f \) generates \( I_f \). The next result says that the minimal and characteristic polynomials of a map do not change if we extend the scalars.

3.24. **Theorem.** Let \( F' \) be a subfield of \( F \), \( W \) a vector space over \( F' \) and \( \varphi : W \to W \) a linear map. Suppose that \( V = W_f \) and that \( f = g \varphi \) is the \( F \)-linear extension of \( g \) to \( V \). Then \( m_f = m_g \) and \( \chi_f = \chi_g \).

**Proof.** Let \( A \) be the matrix of \( \varphi \) relative to any ordered basis \( B \) of \( W \). By Lemmas 1.27 and 1.28, \( B \) is also a basis (over \( F \)) of \( V \) and the matrix of \( f \) is also equal to \( A \). Hence \( \chi_f(x) = \det(A - xl) = \chi_g(x) \). Let \( p \in I_g \). Then \( p(g) = 0 \) and so \( p(A) = 0 \). Hence \( p \in I_f \) because \( f \) also has the matrix \( A \). Thus we see that \( I_g \subseteq I_f \) and in particular \( m_g \in I_f \). This implies \( m_f \mid m_g \) by Lemma 3.23(iii). Thus, since \( m_g \) and \( m_f \) are both monic, to show that they are equal it suffices to show that \( m_g \mid m_f \). Write \( m_f(x) = x^k + \sum_{i=0}^{k-1} c_i x^i \), where \( c_i \in F \) and \( k < \text{deg} m_g \). Then

\[
A^k + \sum_{i=0}^{k-1} c_i A^i = A^k + c_{k-1} A^{k-1} + \cdots + c_1 A + c_0 I = 0.
\]

This identity can be viewed as a system of \( n^2 \) linear equations in the unknowns \( c_0, c_1, \ldots, c_{k-1} \), whose coefficients are the various entries of the powers of \( A \). Since \( A \) has entries in \( F' \), the system has coefficients in \( F' \). By definition of the minimal polynomial, the solution \( \langle c_0, c_1, \ldots, c_{k-1} \rangle \in F^k \) is unique. Hence the solution must be in \( \langle F' \rangle^k \). In other words, \( m_f \in I_g \) and therefore \( m_f \mid m_g \). QED

This proof is based on the following elementary principle.

3.25. **Lemma.** Let \( F' \) be a subfield of \( F \). Let \( A \in M_{n \times n}(F') \) and \( b \in (F')^m \). Then the system \( Ax = b \) has solutions in \( (F')^n \) if and only if it has solutions in \( F^n \). Hence, if it has a unique solution in \( F^n \), that solution must be in \( (F')^n \).

**Proof.** Exercise 3.24. QED

The minimal polynomial is easy to find from the Jordan normal form.

3.26. **Theorem.** Assume that \( \varphi \) has a Jordan normal form. Let \( \lambda_1, \lambda_2, \ldots, \lambda_s \) be the distinct eigenvalues of \( \varphi \), and for \( 1 \leq r \leq s \) let \( c_r \) be the size of the largest Jordan block with eigenvalue \( \lambda_r \). Then the minimal polynomial \( m_{\varphi} \) is equal to \( \prod_{r=1}^{s} (x - \lambda_r)^{c_r} \).
3. THE JORDAN NORMAL FORM

Proof. Put \( p(x) = \prod_{r=1}^k (x - \lambda_r)^{e_r} \). We want to show \( p = m_f \). As both \( p \) and \( m_f \) are monic, it suffices to show that \( m_f | p \) and \( p | m_f \). To prove that \( m_f | p \) it is enough to show that \( p(f) = 0 \). You are asked to check this in Exercise 3.25.

Next we show that \( m_f | p \). Because we already know \( m_f | p \) and that \( p \) is a product of factors \( x - \lambda_r \), we must have \( m_f(x) = \prod_{r=1}^k (x - \lambda_r)^{d_r} \) for certain \( d_r \leq e_r \). Suppose \( d_t < e_t \) for some \( t \). Because of the definition of \( e_t \) we can find \( v \in V \) such that \( f_{\lambda_t}^{-1} v = 0 \) but \( f_{\lambda_t}^{-1} v \neq 0 \). In other words, \( f_{\lambda_t}^{-1} v \neq 0 \), and therefore

\[
m_f(f)(v) = \left( \prod_{r=1}^k (f - \lambda_r)^{d_r} \right)(v) = \left( \prod_{r=1}^k f_{\lambda_r}^{d_r} \right)(v) = \left( \prod_{r \neq t} f_{\lambda_r}^{d_r} \right) f_{\lambda_t}^{d_t} v \neq 0
\]

because \( f_{\lambda_t} \) is injective on \( V(\lambda_t) \) for \( t \neq r \) by Theorem 3.11(ii). This contradicts the fact that \( m_f(f) = 0 \). QED

In particular, the roots of \( m_f \) are exactly the eigenvalues of \( f \).

3.27. COROLLARY. An element of \( F \) is an eigenvalue of \( f \) if and only if it is a root of \( m_f \).

Proof. Because of Theorem 3.24 we may enlarge the field \( F \) so that it contains all the roots of \( \chi_f \). Then \( f \) has a Jordan normal form over \( F \). The result is now immediate from Theorem 3.26. QED

3.28. THEOREM (Cayley-Hamilton). The minimal polynomial \( m_f \) divides the characteristic polynomial \( \chi_f \). Hence \( \chi_f(f) = 0 \), i.e. \( f \) satisfies its own characteristic equation.

Proof. Because of Theorem 3.24 we may just as well assume that \( F \) contains all the roots of \( \chi_f \), so that \( f \) has a Jordan normal form over \( F \). That \( m_f \) divides \( \chi_f \) then follows immediately from Lemma 3.16 and Theorem 3.26. Lemma 3.23(iii) then shows \( \chi_f \in F \), i.e. \( \chi_f(f) = 0 \). QED

3.29. THEOREM. \( f \) is diagonalizable over \( F \) if and only if its minimal polynomial \( m_f \) splits into distinct linear factors: \( m_f(x) = \prod_{r=1}^k (x - \lambda_r) \), where \( \lambda_1, \lambda_2, \ldots, \lambda_k \) are distinct elements of \( F \).

Proof. Observe that \( f \) is diagonalizable over \( F \) if and only if it has a Jordan normal form over \( F \) such that the largest Jordan block for each eigenvalue has size 1. Hence, if \( f \) is diagonalizable, by Theorem 3.26 its minimal polynomial \( m_f \) splits into distinct linear factors. Conversely, by Corollary 3.27, if \( m_f \) splits into distinct linear factors over \( F \), then in particular its eigenvalues all lie in \( F \), so it has a Jordan normal form. Because all the linear factors are distinct, \( e_r = 1 \) for all \( r \) by Theorem 3.26, so \( f \) is diagonalizable. QED

This leads to an alternative (though hardly easier) proof of Corollary 3.5: if \( f \) has \( n \) distinct eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_n \in F \), then \( \chi_f \) splits into distinct linear factors, and therefore so does \( m_f \) by Cayley-Hamilton, so \( f \) is diagonalizable by Theorem 3.29.
3.6. Examples

Given complete knowledge of the eigenvalues of \( f \), the proof of Theorem 3.17 we have presented amounts to an efficient algorithm for finding a Jordan basis. Finding the eigenvalues can be problematic. One can try to solve the characteristic equation \( \chi_f(x) = 0 \), although that may be a dumb idea if the dimension is big. Once an eigenvalue \( \lambda \) is known, however, one finds a complete set of strings with that eigenvalue by grinding through the following \( k + 1 \)-step algorithm.

**Step 0.** For \( l = 0, 1, 2, \ldots \) determine the subspace \( W_l = \text{im} f^l \). Then \( W_{l+1} = f_\lambda(W_l) \), so \( V = W_0 \supseteq W_1 \supseteq W_2 \supseteq \cdots \). Stop at the least value of \( l \) satisfying \( W_l = W_{l+1} \). Call this value \( k \).

**Step 1.** Find a basis of \( V_\lambda \cap W_{k-1} \). Regard the elements \( u \) of this basis as strings \( [u] \) of length \( 1 \) with eigenvalue \( \lambda \).

For \( j = 2 \) to \( k \) perform the following step.

**Step j.** For each string \( \delta \) with eigenvalue \( \lambda \) obtained so far, let \( v \) be the last vector of \( \delta \) and find a \( w \in W_{k-1} \) such that \( f_\lambda(w) \). (This is possible because \( W_{k-j+1} = f_\lambda(W_{k-j}) \).) Append \( w \) to the string \( \delta \) to obtain a longer string \( [\delta, w] \). The leading vectors of the strings found so far constitute a basis of \( V_\lambda \cap W_{k-1} \). Extend this basis to a basis of \( V_\lambda \cap W_{k-j} \). Each of the new vectors \( x \) found in this base-extension is an eigenvector and so can be regarded as a string \( [x] \) of length \( 1 \) with eigenvalue \( \lambda \). Add these strings \( [x] \) to the existing collection of strings.

When this algorithm is done, repeat it for the eigenvalues \( \mu \) other than \( \lambda \). The strings with eigenvalue \( \mu \) are located inside \( W_k \), which is the sum of the generalized eigenspaces for the eigenvalues other than \( \lambda \). The entire process can be implemented on a computer and is easily done by hand for small \( n \). A useful check on your work is that \( \lambda \) should appear \( \dim_{\lambda}(f, \lambda) \) times in the Jordan matrix. If you need to know the Jordan basis for only the Jordan matrix, you can proceed as follows.

Pick an eigenvalue \( \lambda \). Find \( k \) as in step 0 above. For all \( l \) such that \( 1 \leq l \leq k \) determine \( d_{\lambda, l} = \dim(V_\lambda \cap W_{k-l}) \). Then the Jordan matrix has \( d_{\lambda, 1} - d_{\lambda, 1+1} \) blocks of size \( 1 \) with eigenvalue \( \lambda \). Repeat for all remaining eigenvalues.

3.30. EXAMPLE. Let \( V = F^2 \), where \( F \) is any field of characteristic \( \neq 2 \). Let

\[
A = \begin{pmatrix} 8 & 8 \\ 0 & 8 \end{pmatrix}
\]

and let us find a Jordan basis for \( A \). The characteristic polynomial is \( \chi_A(x) = (x - 8)^2 \), so \( A \) has eigenvalue \( \lambda = 8 \).

**Step 0.** Let \( W_1 = \text{im} A_8 \). Since

\[
A_8 = A - 8I = \begin{pmatrix} 0 & 8 \\ 0 & 0 \end{pmatrix},
\]

we have \( W_1 = \text{span}(e_1) \) and \( W_1 = \{0\} \) for \( l \geq 2 \). So \( k = 2 \).

**Step 1.** The eigenspace is \( V_8 = \text{span}(e_1) \), so a basis of \( V_8 \cap W_{k-1} = V_8 \cap W_1 \) is \( \{e_1\} \), so we start with a string \( \{e_1\} \).

**Step 2.** To the string \( \{e_1\} \) append a vector \( w \in W_0 = V \) such that \( A_8w = e_1 \), e.g. \( w = 8^{-1}e_2 \). The string \( \{e_1, 8^{-1}e_2\} \) is already a basis of \( V_8 \), so we are done. The
Jordan matrix is a single block
\[ J = \begin{pmatrix} 8 & 1 \\ 0 & 8 \end{pmatrix}. \]

The base-change matrix \( P \) (i.e. the matrix that gives \( J = P^{-1}AP \)) is found by lining up the Jordan basis vectors in columns:
\[ P = \begin{pmatrix} 1 & 0 \\ 0 & 8^{-1} \end{pmatrix}. \]

3.31. Example. Let \( V = F^6 \), where \( F \) is any field of characteristic \( \neq 2 \). Let
\[ A = \begin{pmatrix} 2 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \]
and let us determine a Jordan basis for \( A \). The characteristic polynomial is \( \chi_A(x) = x^5(x - 2)^2 \), so the eigenvalues are 0 and 2 with algebraic multiplicities 3, resp. 2. First find the strings with eigenvalue 0.

\textbf{Step 0.} Let \( W_1 = \text{im} A^3 \) = \( \text{im} A^4 \). Since
\[ A^3 = \begin{pmatrix} 8 & 0 & 0 & 8 & 20 \\ 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{pmatrix}, \]
we have \( W_1 = \text{span} \{ e_1, e_2, e_5 \} \) and \( W_2 = W_3 = \text{span} \{ e_1, e_2 + e_5 \} \). So \( k = 2 \).

\textbf{Step 1.} The eigenspace is \( V_0 = \text{span} \{ e_2, e_3 \} \), so a basis of \( V_0 \cap W_{k-1} = V_0 \cap W_1 \) is \( \{ e_2 \} \), so we start with a string \( \{ e_2 \} \).

\textbf{Step 2.} To the string \( \{ e_2 \} \) append a vector \( w \in W_0 = V \) such that \( Aw = e_2 \), e.g. \( w = -2^{-1} e_1 + 2^{-1} e_4 \), to get a string \( \{ e_2, -2^{-1} e_1 + 2^{-1} e_4 \} \). Next extend \( \{ e_2 \} \) to a basis of \( V_0 \cap W_0 = V_0 = \text{span} \{ e_2, e_3 \} \), e.g. by picking \( e_3 \). New string: \( \{ e_3 \} \).

Repeat the process for the eigenvalue 2.

\textbf{Step 0.} Let \( W_1 = \text{im} A^2 \). Using \( A_2 = A - 2I \) we get
\[ A_2 = \begin{pmatrix} 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \]
\[ A_2^2 = \begin{pmatrix} 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & -8 & -4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \]
so \( W_1 = \text{span} \{ e_1, e_2, e_3, -e_4 \} \), \( W_2 = W_3 = \text{span} \{ e_1 - e_4, e_2, e_3 \} \). So again \( k = 2 \).

\textbf{Step 1.} The eigenspace is \( V_2 = \text{span} \{ e_1 \} \), so a basis of \( V_2 \cap W_{k-1} = V_2 \cap W_1 \) is \( \{ e_1 \} \), so we start with a string \( \{ e_1 \} \).

\textbf{Step 2.} To the string \( \{ e_1 \} \) append a vector \( w \in W_0 = V \) such that \( A_2 w = e_1 \), e.g. \( w = 2^{-1} e_2 + 2^{-1} e_4 \), to get a string \( \{ e_1, 2^{-1} e_2 + 2^{-1} e_4 \} \). All strings taken together now have \( 5 = \dim V \) vectors, so we are done.

The Jordan basis consists of the strings
\[ \{ e_2, -2^{-1} e_1 + 2^{-1} e_4 \}, \{ e_3 \}, \{ e_1, 2^{-1} e_2 + 2^{-1} e_3 \}. \]
so the Jordan matrix and the base-change matrix are

\[
J = \begin{pmatrix}
0 & 1 \\
0 & 1 \\
0 & 2 \\
2 & 2
\end{pmatrix}, \quad P = \begin{pmatrix}
0 & 2^{-1} & 0 & 1 \\
1 & 0 & 0 & 2^{-1} \\
0 & 0 & 1 & 0 \\
0 & 2^{-1} & 0 & 0
\end{pmatrix}.
\]

### 3.7. Applications

Many applications of linear algebra (population dynamics etc.) require exponentiating a matrix or raising it to a power. Both are straightforward once a Jordan basis for the matrix is known.

**Powers of a matrix.** Let \( A \) be an \( n \times n \)-matrix over a field \( F \). For \( k = 0, 1, 2, \ldots \) consider the recursive linear equation

\[
v(k+1) = Av(k),
\]

where the \( v(k) \)'s are vectors in \( F^n \). This is shorthand for a system of \( n \) recursive linear equations in \( n \) variables,

\[
\begin{align*}
v_1(k+1) &= a_{11}v_1(k) + a_{12}v_2(k) + \cdots + a_{1n}v_n(k), \\
v_2(k+1) &= a_{21}v_1(k) + a_{22}v_2(k) + \cdots + a_{2n}v_n(k), \\
& \vdots \\
v_n(k+1) &= a_{n1}v_1(k) + a_{n2}v_2(k) + \cdots + a_{nn}v_n(k).
\end{align*}
\]

A solution to (3.9) is by definition a sequence of vectors

\[
(v(k))_{k \geq 0} = (v(0), v(1), v(2), \ldots)
\]

such that (3.9) is satisfied for all \( k \). The following result is obvious by induction on \( k \).

#### 3.32. Lemma. For every vector \( v \in F^n \) there exists a unique solution \( (v(k))_{k \geq 0} \) to the system (3.9) satisfying \( v(0) = v \). This solution is given by \( (v, Av, A^2v, \ldots) \), i.e. \( v(k) = A^kv \).

How to find \( A^k \) in an efficient way? Suppose we have a Jordan basis \( \mathcal{B} \) for \( A \). Then the Jordan normal form \( J \) of \( A \) is given by \( J = P^{-1}AP \), where the transformation matrix \( P \) is the matrix whose column vectors are the elements of \( \mathcal{B} \). Then so \( A = PJP^{-1} \), so

\[
A^k = \underbrace{A \cdots A}_{k \text{ times}} = PJ^{-1}P^{-1}P^{-1} \cdots P^{-1}P^{-1} = PJ^kP^{-1}.
\]

Now \( J \) consists of blocks \( J_1, J_2, \ldots, J_q \) arranged along the diagonal, so

\[
J^k = \begin{pmatrix}
J_1^k \\
J_2^k \\
\vdots \\
J_q^k
\end{pmatrix}.
\]
3. THE JORDAN NORMAL FORM

To calculate the k-th power of a single Jordan block $J_{\lambda,1}$ we write $J_{\lambda,1} = \lambda I + N$, where $N = J_{0,1}$ is the Jordan block with eigenvalue 0.

3.33. LEMMA.

$$J_{\lambda,1}^k = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} N^j = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \frac{k(k-1)}{2!} \lambda^{k-2} & \cdots & \frac{k!}{(k-j)!} \lambda^{k-j} \\ & \lambda^k & k\lambda^{k-1} & \cdots & \frac{k!}{(k-j)!} \lambda^{k-j} \\ & & \lambda^k & \cdots & \frac{k!}{(k-j)!} \lambda^{k-j} \\ & & & \ddots & \vdots \\ & & & & \lambda^k \end{pmatrix}.$$ 

PROOF. Use the binomial formula $(A + B)^k = \sum_{j=0}^{k} \binom{k}{j} A^{k-j} B^j$, which applies to any pair of commuting matrices $A$ and $B$, and use $|\lambda|^{k-j} |N^j| = \lambda^{k-j} |N^j|$. QED

(Note that $N^1 = 0$, so the terms for $k \geq 1$ in the sum vanish.) Substituting this formula back into $J^k$ we obtain a formula for $A^k = P J^k P^{-1}$.

3.34. EXAMPLE. Let us find a formula for $A^k$, where

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

The characteristic polynomial is $\chi_A(x) = (x - 2)^2$, so $A$ has a double eigenvalue 2. There is one independent eigenvector, $v_1 = e_1 + e_2$. So the Jordan form has one block:

$$J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \quad \text{so} \quad J^k = \begin{pmatrix} 2^k & k2^{k-1} \\ 0 & 2^k \end{pmatrix}. $$

The second Jordan basis vector $v_2$ must satisfy $A_2 v_2 = (A - 2I) v_2 = e_1 + e_2$. Take e.g. $v_2 = e_2$. Then

$$P = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

so we get

$$A^k = P J^k P^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2^k & k2^{k-1} \\ 0 & 2^k \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 2^{k-1} \begin{pmatrix} 2 - k & k \\ -k & 2 + k \end{pmatrix}.$$

Exponentiating a matrix. Let $F = \mathbb{R}$ or $\mathbb{C}$. For $A \in M_n(F)$ and $t \in F$ the exponential of $tA$ is defined by

$$\exp(tA) = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = I + tA + \frac{t^2}{2} A^2 + \frac{t^3}{6} A^3 + \cdots + \frac{t^k}{k!} A^k + \cdots.$$ 

This is to be regarded as a power series in $t$ with coefficients in $M_n(F)$. By the root test,

$$0 \leq \lim_{k \to \infty} \left( \frac{1}{k!} ||A^k|| \right)^{1/k} = \lim_{k \to \infty} \frac{||A||}{\sqrt[k]{k!}} = 0,$$

its radius of convergence is infinite, so $\exp(tA)$ is well-defined for all $t \in F$ and $A \in M_n(F)$, and we can find the derivative of $\exp(tA)$ with respect to $t$ by differentiating
term by term:
\[
\frac{d}{dt} \exp tA = \sum_{k=0}^{\infty} \frac{d}{dt} \frac{t^k}{k!} A^k = \sum_{k=1}^{\infty} k \frac{t^{k-1}}{k!} A^{k-1} = \sum_{l=0}^{\infty} \frac{t^1}{l!} A^{l+1} = A \exp tA.
\]

The usual proof of the exponential law, \( e^{a+b} = e^a e^b \), applies to give
\[
\exp(A + B) = \exp A \exp B,
\]
provided that \( A \) and \( B \) commute. The following result is one of the main justifications of matrix exponentials. Let \( V = C^\infty[F] \) be the vector space of smooth functions from \( F \) to \( F \). Let \( \phi = (\phi_1, \phi_2, \ldots, \phi_n)^T \in V \oplus V \oplus \cdots \oplus V \) be an \( n \)-tuple of smooth functions. For each \( t \in F \) the \( n \)-tuple \( \phi(t) \) is a vector in \( F^n \), so by interpreting \( t \) as “time” we can think of \( \phi \) as a “time-dependent” vector in \( F^n \). For this reason we call \( \phi \) a smooth curve in \( F^n \). Consider the linear differential equation
\[
\phi' = A\phi.
\]

This is to be viewed as a “continuous” analogue of the recursive equation (3.9). It is shorthand for a system of \( n \) linear differential equations in \( n \) functions,
\[
\begin{align*}
\phi_1' &= a_{11} \phi_1 + a_{12} \phi_2 + \cdots + a_{1n} \phi_n, \\
\phi_2' &= a_{21} \phi_1 + a_{22} \phi_2 + \cdots + a_{2n} \phi_n, \\
&\vdots \\
\phi_n' &= a_{n1} \phi_1 + a_{n2} \phi_2 + \cdots + a_{nn} \phi_n.
\end{align*}
\]

Think of \( \phi'(t) \) as the velocity vector of the curve \( \phi \) at time \( t \). Then (3.10) is an equation that prescribes the velocity vector along the curve. It is an important fact that for each point \( v \in F^n \) there exists a unique curve \( \phi \) that at time \( t = 0 \) passes through \( v \) and satisfies (3.10).

3.35. THEOREM. For every vector \( v \in F^n \) there exists a unique solution \( \phi \) to the system (3.10) satisfying \( \phi(0) = v \). This solution is given by \( \phi(t) = (\exp tA)v \).

PROOF. Uniqueness of the solution is guaranteed by a theorem in analysis. That \( \phi(t) = (\exp tA)v \) is a solution follows from
\[
\frac{d\phi}{dt}(t) = \frac{d}{dt}(\exp tA)v = A(\exp tA)v = A\phi(t).
\]

QED

To compute \( \exp tA \) explicitly, we proceed as when we raised \( A \) to a power:
\[
\exp tA = \exp t(PP^{-1}) = \sum_{k=0}^{\infty} \frac{t^k}{k!} (PP^{-1})^k = \sum_{k=0}^{\infty} \frac{t^k}{k!} P^k P^{-1} = P \left( \sum_{k=0}^{\infty} \frac{t^k}{k!} \right) P^{-1} = P(\exp t)P^{-1}.
\]
To exponentiate $t$ note that

$$
\exp tJ = \begin{pmatrix}
\exp tl_1 \\
\exp tl_2 \\
\vdots \\
\exp tl_q
\end{pmatrix},
$$

where $J_p$ is the $p$-th Jordan block of $J$. A formula for the exponential of $t$ times a Jordan block $J_{1,1}$ is given in Exercise 3.28.

If $F = \mathbb{R}$ the Jordan normal form of $A$ may be a complex matrix and so the resulting formula for $\phi(t) = (\exp tA)v$ may involve complex numbers, even though $\phi(t)$ is a real vector for all $t$. For practical applications one sometimes prefers formulae that do not involve complex numbers; to achieve this one uses the real version of the Jordan normal form explained in Exercise 3.31. (A similar comment applies to finding the powers of $A$.)

**Exercises**

3.1. Consider the equivalence relation on $m \times n$-matrices of Exercise 1.25(i).

(i) Each $m \times n$-matrix $A$ is equivalent to a unique matrix of the form

$$
A \sim \begin{pmatrix}
I_r & 0 \\
0 & 0
\end{pmatrix},
$$

where $r = \text{rank } A$ and where the three 0's stand for blocks of size $r \times (n - r)$, $(m - r) \times r$ and $(m - r) \times (n - r)$, respectively. (Use either row and column reduction or the fundamental theorem for linear maps, Theorem 1.7.)

(ii) $A \sim B$ if and only if $\text{rank } A = \text{rank } B$.

3.2. This problem generalizes the notions of direct product and direct sum to infinite collections of vector spaces. (For infinite collections, these two notions no longer coincide!) Let $S$ be an arbitrary set and let $F$ be a field. A family of vector spaces parametrized or indexed by $S$ is a function $V$ from $S$ into the collection of all vector spaces over $F$. This means that for each $s \in S$ we are given a vector space $V(s)$. Instead of $V(s)$ we will write $V_s$ and note the family as $\{V_s\}_{s \in S}$. A $S$-tuple with values in $V$ is a function $v$ defined on $S$ such that $v(s) \in V_s$ for all $s \in S$. Instead of $v(s)$ we write $v_s$. We note the tuple $v$ as $[v_s]_{s \in S}$ and call $v_s \in V_s$ the $s$-th coordinate of $v$, to emphasize the analogy with finite $n$-tuples of scalars. The direct product of the family $V$, denoted by $\prod_{s \in S} V_s$, is the collection consisting of all $V$-valued $S$-tuples. The support of a tuple $[v_s]_{s \in S}$ is the set $\text{supp}(v_s) = \{s \in S \mid v_s \neq 0\}$. The direct sum of the family $V$, denoted by $\bigoplus_{s \in S} V_s$, is the subset of the direct product consisting of all $V$-valued $S$-tuples which have finite support.

(i) Define appropriate notions of addition and scalar multiplication in $\prod_{s \in S} V_s$ and check that it is a vector space over $F$. (How would you multiply $[v_s]_{s \in S}$ by $\lambda \in F$ and how would you add $[v_s]_{s \in S}$ to $[w_s]_{s \in S}$?)

(ii) $\bigoplus_{s \in S} V_s$ is a subspace of $\prod_{s \in S} V_s$.

(iii) Discuss how the notions of a finite direct sum $V_1 + V_2 + \cdots + V_n$ (see Section 1.5) and of the function spaces $F^S$, $F^{(S)}$ (see Section 1.4) and $W^S$ (see Exercise 1.28) are special kinds of direct products or sums.
3.3 (universal property of the direct product). Let \((V_s)_{s \in S}\) be a family of vector spaces over \(F\) parametrized by a set \(S\). For each \(s \in S\) the projection map

\[ p_s : \prod_{t \in S} V_t \to V_s \]

is defined by \(p_s((v_t)_{t \in S}) = v_s\). ("Select the \(s\)-th coordinate of the tuple"). \(p_s\) is linear. For every vector space \(W\) and every collection of linear maps \(\{f_s : W \to V_s \mid s \in S\}\) there exists a unique linear map \(f : W \to \prod_{s \in S} V_s\) satisfying \(p_s \circ f = f_s\) for all \(s \in S\). Here is a picture for a two-element set \(S = \{a, b\}\).

![Diagram](https://via.placeholder.com/150)

(This sounds complicated, but it is really a follow-your-nose exercise. For any \(w \in W\) you must define \(f(w)\). The condition \(p_s \circ f = f_s\) means that for each \(s \in S\) you want the \(s\)-th coordinate of \(f(w)\) to be equal to \(f_s(w)\). So what is \(f(w)\)?)

3.4 (universal property of the direct sum). Let \((V_s)_{s \in S}\) be a family of vector spaces over \(F\) parametrized by a set \(S\). For each \(s \in S\) the injection map

\[ i_s : V_s \to \bigoplus_{t \in S} V_t \]

is defined by

\[ (i_s(w))_t = \begin{cases} w & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases} \]

for \(w \in V_s\). ("Place \(w \in V_s\) in the \(s\)-th slot and set all other coordinates equal to 0") \(i_s\) is linear. For every vector space \(W\) and every collection of linear maps \(\{f_s : V_s \to W \mid s \in S\}\) there exists a unique linear map \(f : \bigoplus_{s \in S} V_s \to W\) satisfying \(f \circ i_s = f_s\) for all \(s \in S\). Here is a picture for a two-element set \(S = \{a, b\}\).

![Diagram](https://via.placeholder.com/150)

(Similar comment as for Exercise 3.3.)

3.5. Let \((V_s)_{s \in S}\) be a family of vector spaces over \(F\) parametrized by a set \(S\). The dual of \(\bigoplus_{s \in S} V_s\) is isomorphic to \(\prod_{s \in S} V_s^*\).

3.6. Let \((V_s)_{s \in S}\) be a family of vector spaces indexed by a set \(S\). We call \((V_s)_{s \in S}\) totally ordered if for all \(s, t \in S\) we have either \(V_s \subseteq V_t\) or \(V_t \subseteq V_s\). Suppose the \(V_s\)'s are all subspaces of one single vector space \(V\). If \((V_s)_{s \in S}\) is totally ordered, then the union \(\bigcup_{s \in S} V_s\) is a subspace of \(V\).

3.7. A real division algebra with unit is a real vector space \(V\) equipped with an operation \(\cdot : V \times V \to V\) called multiplication (usually written as \(xy\) instead of \(m(x, y)\)) and an element \(e\) called the unit such that the following axioms hold: (a) \(m\) is bilinear; (b) \(cx = xc\) for all \(c \in V\); (c) \(e \neq 0\) and \(xy = yx\) then \(x = y\); (d) if \(x \neq 0\) and \(xy = y\) then \(x = y\). Prove that if \(V\) is a real division algebra with unit and the dimension of \(V\) is finite and odd, then \(V = \text{span}\{e\}\). (Use that the linear map \(1_x : V \to V\) defined by \(1_x(y) = xy\) has an eigenvector.)
3.8. Find an invertible complex $4 \times 4$-matrix $P$ such that $P^{-1}AP$ is a diagonal complex $4 \times 4$-matrix, where
\[
A = \begin{pmatrix}
2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
1 & -2 & 0 & -1 \\
2 & -4 & 1 & 0
\end{pmatrix}.
\]

In the following exercises $V$ denotes a vector space of dimension $n$ ($n < \infty$ unless otherwise specified) over a field $F$ and $f : V \rightarrow V$ a linear map.

3.9. Let $g : V \rightarrow V$ be another linear map. Suppose that both $f$ and $g$ are diagonalizable and that $f \circ g = g \circ f$. Then there exists a basis of $V$ consisting of common eigenvectors of $f$ and $g$.

3.10. Prove or disprove the following assertions.
(i) $[\text{spec } f] = n$ implies $\text{mult}_g[f, \lambda] \leq 1$ for all $\lambda \in F$.
(ii) If $f$ is diagonalizable then $[\text{spec } f] = n$.
(iii) If $[\text{spec } f] = n \leq \infty$, then $f$ is diagonalizable.

3.11. (i) Prove Theorem 3.10. (Choose a basis of $W$; extend it to a basis of $V$; write the matrix $A$ of $f$ relative to this basis; note that $A$ has a number of zeroes so that the determinant of $A - xI$ simplifies.)
(ii) Explain why Theorem 3.9 is a special case of Theorem 3.10.

3.12. (i) $V_\lambda$ is an $f_\mu$-invariant subspace for every $\lambda$ and $\mu$ in $F$.
(ii) $f_\mu$ maps $V_\lambda$ bijectively onto itself if $\mu \neq \lambda$.

3.13. Let $v$ be a nonzero vector in $V$.
(i) $W = \text{span}\{v, f(v), f^2(v), f^3(v), \ldots\}$ is an $f$-invariant subspace.
(ii) If $W'$ is an $f$-invariant subspace and $v \in W'$, then $W \subseteq W'$.
(iii) Let $k = \dim W - 1$. Then $\{v, f(v), f^2(v), f^3(v), \ldots, f^k(v)\}$ is a basis for $W$.

3.14. Suppose $B$ is a Jordan basis for $f$. Let $S = \{v_1, v_2, \ldots, v_l\}$ be one of the strings contained in $B$. Let $\lambda$ be the eigenvalue of $S$. Then $S$ is as long as possible in the sense that there exists no $v_{l+1} \in V$ such that $v_{l+1} = v_l$.

3.15. Suppose that $B = S_1 \cup S_2 \cup \ldots \cup S_n$ is a Jordan basis for $f$. Let $\lambda \in \text{spec } f$ be an eigenvalue. Suppose $S_1, S_2, \ldots, S_n$ are those strings among the $S_i$ that have eigenvalue $\lambda$, and let $v_1, v_2, \ldots, v_n$ be their leading vectors. Then the set $\{v_1, v_2, \ldots, v_n\}$ is a basis for $V_\lambda$.


3.17. Let $n = 2$ and suppose that $\chi_f(x) = (x - \lambda_1)(x - \lambda_2)$. What are all the possible Jordan normal forms of $f$ if $\lambda_1 \neq \lambda_2$? And if $\lambda_1 = \lambda_2$?

3.18. Let $\lambda$, $\mu \in F$ and let
\[
A = \begin{pmatrix}
\lambda & 1 \\
0 & \mu
\end{pmatrix}.
\]

Determine a Jordan basis and Jordan normal form $J$ for $A$. Find a matrix $P$ such that $J = P^{-1}AP$. (The answers depend on the value of $\mu$.)

3.19. The following conditions are equivalent:
(i) $X_f(x) = x^n$. 

(ii) \( f^n = 0 \).
(Use the Jordan normal form theorem, Theorem 3.17.)

3.20. For \( n = 4 \), write down a complete list of all possible Jordan normal forms (up to permutations of the blocks) of maps \( f \) satisfying \( f^n = 0 \).

3.21. Let \( g : V \to V \) be another linear map. Suppose that both \( \chi_f \) and \( \chi_g \) split into linear factors and that \( f \circ g = g \circ f \). Then there exists a basis of \( V \) which is a Jordan basis for \( f \) and \( g \) simultaneously.

3.22. Prove Theorem 3.20.

3.23. Assume \( \text{char} F = 0 \). Let \( \mathcal{D} : F[x] \to F[x] \) be the differentiation operator. (See Exercise 1.20.) Find a Jordan basis and a Jordan normal form for \( \mathcal{D} \). (The monomials \( 1, x, x^2, \ldots \), form “almost” a Jordan basis. Try to find a Jordan basis by taking suitable multiples of the \( x^n \), as in Example 3.30.)


3.25. Finish the proof of Theorem 3.26 by showing that \( p(f) = 0 \).

3.26. Find a formula for \( A^k \), where \( A \) is as in Example 3.31. Be as explicit as possible.

3.27. Same problem for
\[
A = \begin{pmatrix} 4 & -1 & -1 \\ 3 & 1 & -2 \\ 1 & -1 & 2 \end{pmatrix}.
\]

3.28. Let \( F = \mathbb{C} \) and let \( J_{\lambda, 1} \) be the Jordan block with eigenvalue \( \lambda \) of size \( 1 \). Then
\[
\exp t J_{\lambda, 1} = e^{\lambda t} \sum_{k=0}^{\infty} \frac{t^k}{k!} = e^{\lambda t} \begin{pmatrix} 1 & t & \frac{t^2}{2} & \frac{t^3}{6} & \cdots & \frac{t^{i-1}}{(i-1)!} \\ 1 & t & \frac{t^2}{2} & \frac{t^3}{6} & \cdots & \frac{t^{i-1}}{(i-1)!} \\ 1 & t & \frac{t^2}{2} & \frac{t^3}{6} & \cdots & \frac{t^{i-1}}{(i-1)!} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t & \frac{t^2}{2} & \frac{t^3}{6} & \cdots & \frac{t^{i-1}}{(i-1)!} \\ 1 & t & \frac{t^2}{2} & \frac{t^3}{6} & \cdots & \frac{t^{i-1}}{(i-1)!} \end{pmatrix}.
\]
(Use the method of Lemma 3.33.)

3.29. Find the general solution \( \phi = (\phi_1, \phi_2)^T \) of the following system of differential equations. (Use Theorem 3.35 and Exercise 3.28. To save work, observe that the matrix \( A \) involved made a previous appearance in Section 3.7.)
\[
\begin{align*}
\phi'_1 &= \phi_1 + \phi_2, \\
\phi'_2 &= -\phi_1 + 3\phi_2.
\end{align*}
\]

3.30. Find the general solution \( \phi = (\phi_1, \phi_2, \phi_3, \phi_4)^T \) of the following system of differential equations.
\[
\begin{align*}
\phi'_1 &= \phi_4, \\
\phi'_2 &= -\phi_3 + \phi_4, \\
\phi'_3 &= \phi_1 + \phi_2, \\
\phi'_4 &= -\phi_1.
\end{align*}
\]
3.31 (real Jordan normal forms). Let \( F = \mathbb{R}, V = \mathbb{R}^n \), and let \( A \) be the matrix of \( f \) relative to the standard basis of \( V \). Let \( V_C = \mathbb{C}^n \) and let \( f_C \) be the complex linear extension of \( f \) to \( V_C \) which is given by the same matrix \( A \). The complex conjugate of a vector \( x = (x_1, x_2, \ldots, x_n)^T \in V_C \) is given by \( \bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)^T \). The complex conjugate of any subset \( \Lambda \) of \( V_C \) is the collection \( \bar{\Lambda} \) consisting of all \( \bar{x} \) such that \( x \in \Lambda \), i.e. the image of \( \Lambda \) under complex conjugation.

(i) If \( \Sigma = (\nu_1, \nu_2, \ldots, \nu_l) \) is a string with eigenvalue \( \lambda \) for the map \( f_C \), then \( \bar{\Sigma} = (\bar{\nu}_1, \bar{\nu}_2, \ldots, \bar{\nu}_l) \) is a string with eigenvalue \( \bar{\lambda} \).

(ii) \( \Sigma \) and \( \bar{\Sigma} \) are independent if \( \lambda \) is not real.

(iii) Suppose \( \lambda \) is not real and write \( \lambda = \alpha + i\beta \) with \( \beta \neq 0 \). Let \( u_k = \nu_k + \bar{\nu}_k \) and \( v_k = i(\nu_k - \bar{\nu}_k) \) for \( 1 \leq k \leq l \). Then \( \{u_1, v_1, u_2, v_2, \ldots, u_l, v_l\} \) is an ordered basis of \( W = \text{span}(\Sigma \cup \bar{\Sigma}) \) and the matrix of the restriction of \( f_C \) to \( W \) is a real Jordan block

\[
J^R_{\lambda, 1} = \begin{pmatrix}
\lambda & 1 & & & \\
& \lambda & & & \\
& & \ddots & & \\
& & & \lambda & 1 \\
& & & & \lambda
\end{pmatrix}, \text{ where } \lambda = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \text{ and } I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\]

(iv) There exists a Jordan basis for \( f_C \) of the form

\[
B = R_1 \cup R_2 \cup \ldots \cup R_p \cup S_1 \cup S_2 \cup \ldots \cup S_q \cup \bar{S}_q.
\]

Here \( R_1, R_2, \ldots, R_p \) are strings with real eigenvalues and are contained in \( V \), and \( S_1, S_2, \ldots, S_q, \bar{S}_q \) are strings with nonreal eigenvalues.

(v) State and prove a real version of the Jordan normal form theorem for the real linear map \( f : V \to V \). (Your statement is not supposed to mention the complexified space \( V_C \) or the complexified map \( f_C \), although your proof probably will.)

3.32. Let \( J^R_{\lambda, 1} \) be the real Jordan block of Exercise 3.31(iii). Then

\[
\exp J^R_{\lambda, 1} = e^{\alpha t} \begin{pmatrix}
R_{\beta_1} & & & & \\
& R_{\beta_1} & & & \\
& & \ddots & & \\
& & & R_{\beta_1} & \\
& & & & R_{\beta_1}
\end{pmatrix} \begin{pmatrix}
1 & \frac{t}{1!} & \frac{t^2}{2!} & \ldots & \frac{t^{l-1}}{(l-1)!} \\
1 & \frac{t}{1!} & \frac{t^2}{2!} & \ldots & \frac{t^{l-1}}{(l-1)!} \\
& \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
& & & \ddots & \ddots \\
& & & & 1
\end{pmatrix},
\]

where

\[
R_{\beta} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \quad \text{and} \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\]

(Write \( J^R_{\lambda, 1} = D + N \) with

\[
D = \begin{pmatrix}
\lambda & & & \\
& \lambda & & \\
& & \ddots & \\
& & & \lambda
\end{pmatrix}, \quad N = J^R_{0, 1} = \begin{pmatrix} 0 & 1 & & \\
& 0 & 1 & \\
& & \ddots & 1 \\
& & & 0
\end{pmatrix}.
\]
Show that $\exp(t(D + N)) = \exp(tD) \exp(tN)$ and

$$
\exp(tD) = \begin{pmatrix}
\exp(tA) \\
\exp(tA) \\
\ddots \\
\exp(tA)
\end{pmatrix}.
$$

Find $\exp(t\Lambda)$ by writing $\Lambda = \alpha J + \beta I$, where $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and by showing that $\exp(t\Lambda) = \exp(\alpha t) |\exp(\beta t)| e^{\alpha t} R_{\beta t}$.