4 Linear operators and matrices

Given a basis $\{e_1, e_2, \dots, e_n\}$ of a finite dimensional vector space V, we recall from Section 3.5 that the **matrix of components** $\mathsf{T} = [T^i{}_j]$ of a linear operator $T : V \to V$ with respect to this basis is defined by Eq. (3.6) as:

$$T e_j = T^i{}_j e_i, \tag{4.1}$$

and under a transformation of basis,

$$e_i = A^j{}_i e'_j, \qquad e'_i = A'^k{}_i e_k, \tag{4.2}$$

where

$$A'^k{}_j A^j{}_i = \delta^k_i, \qquad A^i{}_k A'^k{}_j = \delta^i{}_j, \tag{4.3}$$

the components of any linear operator T transform by

$$T'^j{}_i = A^j{}_m T^m_k A'^k{}_i. \tag{4.4}$$

The matrices $\mathsf{A} = [A^j{}_i]$ and $\mathsf{A}' = [A'^k{}_i]$ are inverse to each other, $\mathsf{A}' = \mathsf{A}^{-1}$, and (4.4) can be written in matrix notation as a *similarity transformation*

$$\mathsf{T}' = \mathsf{A}\,\mathsf{T}\,\mathsf{A}^{-1}. \tag{4.5}$$

The main task of this chapter will be to find a basis that provides a standard representation of any given linear operator, called the *Jordan canonical form*. This representation is uniquely determined by the operator and encapsulates all its essential properties. The proof given in Section 4.2 is rather technical and may be skipped on first reading. It would, however, be worthwhile to understand its appearance, summarized at the end of that section, as it has frequent applications in mathematical physics. Good references for linear operators and matrices in general are [1–3], while a detailed discussion of the Jordan canonical form can be found in [4].

It is important to realize that we are dealing with linear operators on *free vector spaces*. This concept will be defined rigorously in Chapter 6, but essentially it means that the vector spaces have no further structure imposed on them. A number of concepts such as 'symmetric', 'hermitian' and 'unitary', which often appear in matrix theory, have no place in free vector spaces. For example, the requirement that T be a symmetric matrix would read $T^i{}_j = T^j{}_i$ in components, an awkward-looking relation that violates the rules given in Section 3.6. In Chapter 5 we will find a proper context for notions such as 'symmetric transformations' and 'hermitian transformations'.

4.1 Eigenspaces and characteristic equations

Invariant subspaces

A subspace U of V is said to be **invariant** under a linear operator $S : V \to V$ if

$$SU = \{Su \mid u \in U\} \subseteq U.$$

In this case, the action of S restricted to the subspace U, $S\big|_U$, gives rise to a linear operator on U.

Example 4.1 Let V be a three-dimensional vector space with basis $\{e_1, e_2, e_3\}$, and S the operator defined by

$$\begin{aligned} Se_1 &= e_2 + e_3, \\ Se_2 &= e_1 + e_3, \\ Se_3 &= e_1 + e_2. \end{aligned}$$

Let U be the subspace of all vectors of the form $(a + b)e_1 + be_2 + (-a + b)e_3$, where a and b are arbitrary scalars. This subspace is spanned by $f_1 = e_1 - e_3$ and $f_2 = e_1 + e_2 + e_3$ and is invariant under S, since

$$Sf_1 = -f_1,\ \ Sf_2 = 2f_2 \implies S(af_1 + bf_2) = -af_1 + 2bf_2 \in U.$$

Exercise: Show that if both U and W are invariant subspaces of V under an operator S then so is their intersection $U \cap W$ and their sum $U + W = \{v = u + w \mid u \in U, w \in W\}$.

Suppose $\dim U = m < n = \dim V$ and let $\{e_1, \dots, e_m\}$ be a basis of U. By Theorem 3.7 this basis can be extended to a basis $\{e_1, \dots, e_n\}$ spanning all of V. The invariance of U under S implies that the first m basis vectors are transformed among themselves,

$$Se_a = \sum_{b=1}^{m} S^b_{\ a} e_b \quad (a \le m).$$

In such a basis, the components $S^k_{\ i}$ of the operator S vanish for $i \le m, k > m$, and the $n \times n$ matrix $\mathsf{S} = [S^k_{\ i}]$ has the *upper block diagonal* form

$$\mathsf{S} = \begin{pmatrix} \mathsf{S}_1 & \mathsf{S}_3 \\ \mathsf{O} & \mathsf{S}_2 \end{pmatrix}.$$

The submatrix S_1 is the $m \times m$ matrix of components of $S|_U$ expressed in the basis $\{e_1, \dots, e_m\}$, while S_3 and S_2 are submatrices of orders $m \times p$ and $p \times p$, respectively, where $p = n - m$, and O is the zero $p \times m$ matrix.

If $V = U \oplus W$ is a decomposition with *both* U and W invariant under S, then choose a basis $\{e_1, \dots, e_m, e_{m+1}, \dots, e_n\}$ of V such that the first m vectors span U while the last $p = n - m$ vectors span W. Then $S^k_{\ i} = 0$ whenever $i > m$ and $k \le m$, and the matrix of the operator S has *block diagonal* form

$$\mathsf{S} = \begin{pmatrix} \mathsf{S}_1 & \mathsf{O} \\ \mathsf{O} & \mathsf{S}_2 \end{pmatrix}.$$

Example 4.2 In Example 4.1 set $f_3 = e_3$. The vectors $f_1,\ f_2,\ f_3$ form a basis adapted to the invariant subspace spanned by f_1 and f_2,

$$Sf_1 = -f_1, \quad Sf_2 = 2f_2, \quad Sf_3 = f_2 - f_3,$$

and the matrix of S has the upper block diagonal form

$$\mathsf{S} = \left(\begin{array}{cc|c} -1 & 0 & 0 \\ 0 & 2 & 1 \\ \hline 0 & 0 & -1 \end{array}\right).$$

On the other hand, the one-dimensional subspace W spanned by $f_3' = e_3 - \frac{1}{2}e_1 - \frac{1}{2}e_2$ is invariant since $Sf_3' = -f_3'$, and in the basis $\{f_1,\ f_2,\ f_3'\}$ adapted to the invariant decomposition $V = U \oplus W$ the matrix S takes on block diagonal form

$$\mathsf{S} = \left(\begin{array}{cc|c} -1 & 0 & 0 \\ 0 & 2 & 1 \\ \hline 0 & 0 & -1 \end{array}\right).$$

Eigenvectors and eigenvalues

Given an operator $S : V \to V$, a scalar $\lambda \in \mathbb{K}$ is said to be an **eigenvalue** of S if there exists a non-zero vector v such that

$$Sv = \lambda v \quad (v \neq 0) \tag{4.6}$$

and v is called an **eigenvector** of S corresponding to the eigenvalue λ. Eigenvectors are those non-zero vectors that are 'stretched' by an amount λ on application of the operator S. It is important to stipulate $v \neq 0$ since the equation (4.6) *always* holds for the zero vector, $S0 = 0 = \lambda 0$.

For any scalar $\lambda \in \mathbb{K}$, let

$$V_\lambda = \{u \mid Su = \lambda u\}. \tag{4.7}$$

The set V_λ is a vector subspace, for

$$\begin{aligned} Su = \lambda u \text{ and } Sv = \lambda v &\implies S(u + av) = Su + aSv \\ &\implies S(u + av) = \lambda u + a\lambda v = \lambda(u + av) \qquad \text{for all } a \in \mathbb{K}. \end{aligned}$$

For every λ, the subspace V_λ is invariant under S,

$$u \in V_\lambda \implies Su = \lambda u \implies S(Su) = \lambda Su \implies Su \in V_\lambda.$$

V_λ consists of the set of all eigenvectors having eigenvalue λ, supplemented with the zero vector $\{0\}$. If λ is not an eigenvalue of S, then $V_\lambda = \{0\}$.

If $\{e_1, e_2, \dots, e_n\}$ is any basis of the vector space V, and $v = v^i e_i$ any vector of V, let $\mathbf{v}$ be the column vector of components

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}.$$

By Eq. (3.8) the matrix equivalent of (4.6) is

$$\mathsf{S}\mathbf{v} = \lambda\mathbf{v}, \tag{4.8}$$

where S is the matrix of components of S. Under a change of basis (4.2) we have from Eqs. (3.18) and (4.5),

$$\mathbf{v}' = \mathsf{A}\mathbf{v}, \qquad \mathsf{S}' = \mathsf{A}\mathsf{S}\mathsf{A}^{-1}.$$

Hence, if $\mathbf{v}$ satisfies (4.8) then $\mathbf{v}'$ is an eigenvector of S' with the same eigenvalue λ,

$$\mathsf{S}'\mathbf{v}' = \mathsf{A}\mathsf{S}\mathsf{A}^{-1}\mathsf{A}\mathbf{v} = \mathsf{A}\mathsf{S}\mathbf{v} = \lambda\mathsf{A}\mathbf{v} = \lambda\mathbf{v}'.$$

This result is not unexpected, since Eq. (4.8) and its primed version are simply representations with respect to different bases of the same basis-independent equation (4.6).

Define the ***n*th power** S^n of an operator inductively, by setting $S^0 = \mathrm{id}_V$ and

$$S^n = S \circ S^{n-1} = SS^{n-1}.$$

Thus $S^1 = S$ and $S^2 = SS$, etc. If $p(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$ is any polynomial with coefficients $a_i \in \mathbb{K}$, the operator polynomial $p(S)$ is defined in the obvious way,

$$p(S) = a_0 + a_1 S + a_2 S^2 + \dots + a_n S^n.$$

If λ is an eigenvalue of S and v a corresponding eigenvector, then v is an eigenvector of any power S^n corresponding to eigenvalue λ^n. For $n = 0$,

$$S^0 v = \mathrm{id}_V v = \lambda^0 v \quad \text{since} \quad \lambda^0 = 1,$$

and the proof follows by induction: assume $S^{n-1}v = \lambda^{n-1}v$ then by linearity

$$S^n v = SS^{n-1}v = S\lambda^{n-1}v = \lambda^{n-1}Sv = \lambda^{n-1}\lambda v = \lambda^n v.$$

For a polynomial $p(x)$, it follows immediately that v is an eigenvector of the operator $p(S)$ with eigenvalue $p(\lambda)$,

$$p(S)v = p(\lambda)v. \tag{4.9}$$

Characteristic equation

The matrix equation (4.8) can be written in the form

$$(\mathsf{S} - \lambda\mathsf{I})\mathbf{v} = \mathbf{0}. \tag{4.10}$$

A necessary and sufficient condition for this equation to have a non-trivial solution $\mathbf{v} \neq \mathbf{0}$ is

$$f(\lambda) = \det(\mathsf{S} - \lambda \mathsf{I}) = \begin{vmatrix} S^1_1 - \lambda & S^1_2 & \dots & S^1_n \\ S^2_1 & S^2_2 - \lambda & \dots & S^2_n \\ \vdots & \vdots & & \vdots \\ S^n_1 & S^n_2 & \dots & S^n_n - \lambda \end{vmatrix} = 0, \tag{4.11}$$

called the **characteristic equation** of S. The function $f(\lambda)$ is a polynomial of degree n in λ,

$$f(\lambda) = (-1)^n \big(\lambda^n - S^k_k \lambda^{n-1} + \dots + (-1)^n \det \mathsf{S}\big), \tag{4.12}$$

known as the **characteristic polynomial** of S.

If the field of scalars is the complex numbers, $\mathbb{K} = \mathbb{C}$, then the fundamental theorem of algebra implies that there exist complex numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$f(\lambda) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n).$$

As some of these roots of the characteristic equation may appear repeatedly, we can write the characteristic polynomial in the form

$$\begin{aligned} f(z) &= (-1)^n (z - \lambda_1)^{p_1} (z - \lambda_2)^{p_2} \dots (z - \lambda_m)^{p_m} \\ &\text{where} \quad p_1 + p_2 + \dots + p_m = n. \end{aligned} \tag{4.13}$$

Since for each $\lambda = \lambda_i$ there exists a non-zero complex vector solution $\mathbf{v}$ to the linear set of equations given by (4.10), the eigenvalues of S must all come from the set of roots $\{\lambda_1, \dots, \lambda_m\}$. The positive integer p_i is known as the **multiplicity** of the eigenvalue λ_i.

Example 4.3 When the field of scalars is the real numbers $\mathbb{R}$ there will not in general be real eigenvectors corresponding to complex roots of the characteristic equation. For example, let A be the operator on $\mathbb{R}^2$ defined by the following action on the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$,

$$Ae_1 = e_2, \quad Ae_2 = -e_1 \implies \mathsf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The characteristic polynomial is

$$f(z) = \begin{vmatrix} -z & -1 \\ 1 & -z \end{vmatrix} = z^2 + 1,$$

whose roots are $z = \pm i$. The operator A thus has no real eigenvalues and eigenvectors. However, if we regard the field of scalars as being $\mathbb{C}$ and treat A as operating on $\mathbb{C}^2$, then it has complex eigenvectors

$$\begin{aligned} u &= e_1 - ie_2, & Au &= iu, \\ v &= e_1 + ie_2, & Av &= -iv. \end{aligned}$$

It is worth noting that, since $A^2e_1, = Ae_2 = -e_1$ and $A^2e_2 = -Ae_1 = -e_2$, the operator A satisfies its own characteristic equation

$$A^2 + \mathrm{id}_{\mathbb{R}^2} = 0 \Longrightarrow \mathsf{A}^2 + \mathsf{I} = 0.$$

This is a simple example of the important *Cayley–Hamilton theorem* – see Theorem 4.3 below.

Example 4.4 Let V be a three-dimensional complex vector space with basis e_1, e_2, e_3, and $S : V \to V$ the operator whose matrix with respect to this basis is

$$\mathsf{S} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The characteristic polynomial is

$$f(z) = \begin{vmatrix} 1-z & 1 & 0 \\ 0 & 1-z & 0 \\ 0 & 0 & 2-z \end{vmatrix} = -(z-1)^2(z-2).$$

Hence the eigenvalues are 1 and 2, and it is trivial to check that the eigenvector corresponding to 2 is e_3. Let $u = xe_1 + ye_2 + ze_3$ be an eigenvector with eigenvalue 1,

$$\mathsf{Su} = \mathbf{u} \quad \text{where} \quad \mathbf{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

then

$$x + y = x, \qquad y = y, \qquad 2z = z.$$

Hence $y = z = 0$ and $u = xe_1$. Thus, even though the eigenvalue $\lambda = 1$ has multiplicity 2, all corresponding eigenvectors are multiples of e_1.

Note that while e_2 is not an eigenvector, it is annihilated by $(S - \mathrm{id}_V)^2$, for

$$\begin{aligned} Se_2 = e_1 + e_2 &\Longrightarrow (S - \mathrm{id}_V)e_2 = e_1 \\ &\Longrightarrow (S - \mathrm{id}_V)^2 e_2 = (S - \mathrm{id}_V)e_1 = 0. \end{aligned}$$

Operators of the form $S - \lambda_i \mathrm{id}_V$ and their powers $(S - \lambda \mathrm{id}_V)^m$, where λ_i are eigenvalues of S, will make regular appearances in what follows. These operators evidently commute with each other and there is no ambiguity in writing them as $(S - \lambda_i)^m$.

Theorem 4.1 *Any set of eigenvectors corresponding to distinct eigenvalues of an operator S is linearly independent.*

Proof: Let $\{f_1, f_2, \dots, f_k\}$ be a set of eigenvectors of S corresponding to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, no pair of which are equal,

$$Sf_i = \lambda_i f_i \quad (i = 1, \dots, k),$$

and let $c_1, c_2, \dots, c_k$ be scalars such that

$$c_1 f_1 + c_2 f_2 + \dots + c_k f_k = 0.$$

If we apply the polynomial $P_1(S) = (S - \lambda_2)(S - \lambda_3)\dots(S - \lambda_k)$ to this equation, then all terms except the first are annihilated, leaving

$$c_1 P_1(\lambda_1) f_1 = 0.$$

Hence

$$c_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\dots(\lambda_1 - \lambda_k) f_1 = 0,$$

and since $f_1 \neq 0$ and all the factors $(\lambda_1 - \lambda_i) \neq 0$ for $i = 2, \dots, k$, it follows that $c_1 = 0$. Similarly, $c_2 = \dots = c_k = 0$, proving linear independence of $f_1, \dots, f_k$. ■

If the operator $S : V \to V$ has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$ where $n = \dim V$, then Theorem 4.1 shows the eigenvectors $f_1, f_2, \dots, f_n$ are l.i. and form a basis of V. With respect to this basis the matrix of S is diagonal and its eigenvalues lie along the diagonal,

$$\mathsf{S} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Conversely, any operator whose matrix is diagonalizable has a basis of eigenvectors (the eigenvalues need not be distinct for the converse). The more difficult task lies in the classification of those cases such as Example 4.4, where an eigenvalue λ has multiplicity $p > 1$ but there are less than p independent eigenvectors corresponding to it.

Minimal annihilating polynomial

The space of linear operators $L(V, V)$ is a vector space of dimension n^2 since it can be put into one-to-one correspondence with the space of $n \times n$ matrices. Hence the first n^2 powers $I \equiv \mathrm{id}_V = S^0,\ S = S^1,\ S^2, \dots,\ S^{n^2}$ of any linear operator S on V cannot be linearly independent since there are $n^2 + 1$ operators in all. Thus S must satisfy a polynomial equation,

$$P(S) = c_0 I + c_1 S + c_2 S^2 + \dots + c_{n^2} S^{n^2} = 0,$$

not all of whose coefficients $c_0, c_1, \dots, c_n$ vanish.

Exercise: Show that the matrix equivalent of any such polynomial equation is basis-independent by showing that any similarity transform $\mathsf{S}' = \mathsf{A}\mathsf{S}\mathsf{A}^{-1}$ of S satisfies the same polynomial equation, $P(\mathsf{S}') = 0$.

Let

$$\Delta(S) = S^k + c_1 S^{k-1} + \dots + c_k I = 0$$

be the polynomial equation with leading coefficient 1 of *lowest* degree $k \leq n^2$, satisfied by S. The polynomial $\Delta(S)$ is unique, for if

$$\Delta'(S) = S^k + c'_1 S^{k-1} + \dots + c'_k I = 0$$

is another such polynomial equation, then on subtracting these two equations we have

$$(\Delta - \Delta')(S) = (c_1 - c_1')S^{k-1} + (c_2 - c_2')S^{k-2} + \cdots + (c_k - c_k') = 0,$$

which is a polynomial equation of degree $< k$ satisfied by S. Hence $c_1 = c_1', c_2 = c_2', \ldots, c_k = c_k'$. The unique polynomial $\Delta(z) = z^k + c_1 z^{k-1} + \cdots + c_k$ is called the **minimal annihilating polynomial of** S.

Theorem 4.2 *A scalar λ is an eigenvalue of an operator S over a vector space V if and only if it is a root of the minimal annihilating polynomial $\Delta(z)$.*

Proof: If λ is an eigenvalue of S, let $u \neq 0$ be any corresponding eigenvector, $Su = \lambda u$. Since $0 = \Delta(S)u = \Delta(\lambda)u$ it follows that $\Delta(\lambda) = 0$.

Conversely, if λ is a root of $\Delta(z) = 0$ then there exists a polynomial $\Delta'(z)$ such that

$$\Delta(z) = (z - \lambda)\Delta'(z),$$

and since $\Delta'(z)$ has lower degree than $\Delta(z)$ it cannot annihilate S,

$$\Delta'(S) \neq 0.$$

Therefore, there exists a vector $u \in V$ such that $\Delta'(S)u = v \neq 0$, and

$$0 = \Delta(S)u = (S - \lambda)\Delta'(S)u = (S - \lambda)v.$$

Hence $Sv = \lambda v$, and λ is an eigenvalue of S with eigenvector v. ■

It follows from this theorem that the minimal annihilating polynomial of an operator S on a complex vector space can be written in the form

$$\Delta(z) = (z - \lambda_1)^{k_1}(z - \lambda_2)^{k_2} \ldots (z - \lambda_m)^{k_m} \quad (k_1 + k_2 + \cdots + k_m = k) \tag{4.14}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ run through all the distinct eigenvalues of S. The various factors $\big(z - \lambda_i\big)^{k_i}$ are called the **elementary divisors** of S. The following theorem shows that the characteristic polynomial is always divisible by the minimal annihilating polynomial; that is, for each $i = 1, 2, \ldots, m$ the coefficient $k_i \leq p_i$ where p_i is the multiplicity of the ith eigenvalue.

Theorem 4.3 (Cayley–Hamilton) *Every linear operator S over a finite dimensional vector space V satisfies its own characteristic equation*

$$f(S) = (S - \lambda_1)^{p_1}(S - \lambda_2)^{p_2} \ldots (S - \lambda_m)^{p_m} = 0.$$

Equivalently, every $n \times n$ matrix S satisfies its own characteristic equation

$$f(\mathsf{S}) = (\mathsf{S} - \lambda_1 \mathsf{I})^{p_1}(\mathsf{S} - \lambda_2 \mathsf{I})^{p_2} \ldots (\mathsf{S} - \lambda_m \mathsf{I})^{p_m} = 0.$$

Proof: Let $e_1, e_2, \ldots, e_n$ be any basis of V, and let $\mathsf{S} = [S^k_{\ j}]$ be the matrix of components of S with respect to this basis,

$$Se_j = S^k_{\ j} e_k.$$

This equation can be written as

$$(S^k_{\ j}\,\mathrm{id}_V - \delta^k_{\ j} S)e_k = 0,$$

or alternatively as

$$T^k_{\ j}(S)e_k = 0, \tag{4.15}$$

where

$$T^k_{\ j}(z) = S^k_{\ j} - \delta^k_{\ j} z.$$

Set $\mathsf{R}(z) = [R^k_{\ j}(z)]$ to be the matrix of cofactors of $\mathsf{T}(z) = [T^k_{\ j}(z)]$, such that

$$R^j_{\ i}(z) T^k_{\ j}(z) = \delta^k_i \det \mathsf{T}(z) = \delta^k_{\ i} f(z).$$

The components $R^k_{\ j}(z)$ are polynomials of degree $\le (n-1)$ in z, and multiplying both sides of Eq. (4.15) by $R^j_{\ i}(S)$ gives

$$R^j_{\ i}(S) T^k_{\ j}(S) e_k = \delta^k_{\ i} f(S) e_k = f(S) e_i = 0.$$

Since the e_i span V we have the desired result, $f(S) = 0$. The matrix version is simply the component version of this equation. ■

Example 4.5 Let A be the matrix operator on the space of complex 4×1 column vectors given by

$$\mathsf{A} = \begin{pmatrix} i & \alpha & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Successive powers of A are

$$\mathsf{A}^2 = \begin{pmatrix} -1 & 2i\alpha & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{A}^3 = \begin{pmatrix} -i & -3\alpha & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and it is straightforward to verify that the matrices $\mathsf{I}, \mathsf{A}, \mathsf{A}^2$ are linearly independent, while

$$\mathsf{A}^3 = (-1 + 2i)\mathsf{A}^2 + (1 + 2i)\mathsf{A} + \mathsf{I}.$$

Hence the minimal annihilating polynomial of A is

$$\Delta(z) = z^3 + (1 - 2i)z^2 - (1 + 2i)z - 1 = (z + 1)(z - i)^2.$$

The elementary divisors of A are thus $z + 1$ and $(z - i)^2$ and the eigenvalues are -1 and i. Computation of the characteristic polynomial reveals that

$$f(z) = \det(\mathsf{A} - z\mathsf{I}) = (z + 1)(z - i)^3,$$

which is divisible by $\Delta(z)$ in agreement with Theorem 4.3.

Problem

Problem 4.1 The **trace** of an $n \times n$ matrix $\mathsf{T} = [T^i_{\ j}]$ is defined as the sum of its diagonal elements,

$$\operatorname{tr}\mathsf{T} = T^i_i = T^1_1 + T^2_2 + \cdots + T^n_n.$$

Show that

(a) $\operatorname{tr}(\mathsf{ST}) = \operatorname{tr}(\mathsf{TS})$.
(b) $\operatorname{tr}(\mathsf{ATA}^{-1}) = \operatorname{tr} T$.
(c) If $T : V \to V$ is any operator define its trace to be the trace of its matrix with respect to a basis $\{e_i\}$. Show that this definition is independent of the choice of basis, so that there is no ambiguity in writing $\operatorname{tr} T$.
(d) If $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + (-1)^n z^n$ is the characteristic polynomial of the operator T, show that $\operatorname{tr} T = (-1)^{n-1} a_{n-1}$.
(e) If T has eigenvalues $\lambda_1, \ldots, \lambda_m$ with multiplicities $p_1, \ldots, p_m$, show that

$$\operatorname{tr} T = \sum_{i=1}^{m} p_i \lambda_i .$$

4.2 Jordan canonical form

Block diagonal form

Let S be a linear operator over a complex vector space V, with characteristic polynomial and minimal annihilating polynomial given by (4.13) and (4.14), respectively. The restriction to complex vector spaces ensures that all roots of the polynomials $f(z)$ and $\Delta(z)$ are eigenvalues. A canonical form can also be derived for operators on real vector spaces, but it relies heavily on the complex version.

If $(z - \lambda_i)^{k_i}$ is the ith elementary divisor of S, define the subspace

$$V_i = \{u \mid (S - \lambda_i)^{k_i} u = 0\}.$$

This subspace is invariant under S, for

$$\begin{aligned} u \in V_i &\Longrightarrow (S - \lambda_i)^{k_i} u = 0 \\ &\Longrightarrow (S - \lambda_i)^{k_i} Su = S(S - \lambda_i)^{k_i} u = 0 \\ &\Longrightarrow Su \in V_i . \end{aligned}$$

Our first task will be to show that V is a direct sum of these invariant subspaces,

$$V = V_1 \oplus V_2 \oplus \ldots$$

Lemma 4.4 *If $(z - \lambda_i)^{k_i}$ is an elementary divisor of the operator S and $(S - \lambda_i)^{k_i + r} u = 0$ for some $r > 0$, then $(S - \lambda_i)^{k_i} u = 0$.*

Proof: There is clearly no loss of generality in setting $i = 1$ in the proof of this result. The proof proceeds by induction on r.

Case $r = 1$: Let u be any vector such that $(S - \lambda_1)^{k_1 + 1} u = 0$ and set

$$v = (S - \lambda_1)^{k_1} u.$$

Since $(S - \lambda_1)v = 0$, this vector satisfies the eigenvector equation, $Sv = \lambda_1 v$. If $\Delta(S)$ is the minimal annihilating polynomial of S, then

$$\begin{aligned} 0 = \Delta(S)u &= (S - \lambda_2)^{k_2} \dots (S - \lambda_m)^{k_m} v \\ &= (\lambda_1 - \lambda_2)^{k_2} \dots (\lambda_1 - \lambda_m)^{k_m} v. \end{aligned}$$

As all $\lambda_i \neq \lambda_j$ for $i \neq j$ it follows that $v = 0$, which proves the case $r = 1$.

Case $r > 1$: Suppose the lemma has been proved for $r - 1$. Then

$$\begin{aligned} (S - \lambda_1)^{k_1+r} u = 0 &\implies (S - \lambda_1)^{k_1+r-1}(S - \lambda_1)u = 0 \\ &\implies (S - \lambda_1)^{k_1}(S - \lambda_1)u = 0 \qquad \text{by induction hypothesis} \\ &\implies (S - \lambda_1)^{k_1} u = 0 \qquad \text{by the case } r = 1, \end{aligned}$$

which concludes the proof of the lemma. ■

If $\dim(V_1) = p$ let $h_1, \dots, h_p$ be a basis of V_1 and extend to a basis of V using Theorem 3.7:

$$h_1,\ h_2,\ \dots,\ h_p,\ h_{p+1}, \dots,\ h_n. \tag{4.16}$$

Of course if $(z - \lambda_1)^{k_1}$ is the only elementary divisor of S then $p = n$, since $V_1 = V$ as every vector is annihilated by $(S - \lambda_1)^{k_1}$. If, however, $p < n$ we will show that the vectors

$$h_1,\ h_2,\ \dots,\ h_p,\ \hat{h}_1,\ \hat{h}_2,\ \dots,\ \hat{h}_{n-p} \tag{4.17}$$

form a basis of V, where

$$\hat{h}_a = (S - \lambda_1)^{k_1} h_{p+a} \quad (a = 1, \dots, n - p). \tag{4.18}$$

Since the vectors listed in (4.17) are n in number it is only necessary to show that they are linearly independent. Suppose that for some constants $c^1, \dots, c^p, \hat{c}^1, \dots, \hat{c}^{n-p}$

$$\sum_{i=1}^{p} c^i h_i + \sum_{a=1}^{n-p} \hat{c}^a \hat{h}_a = 0. \tag{4.19}$$

Apply $(S - \lambda_1)^{k_1}$ to this equation. The first sum on the left is annihilated since it belongs to V_1, resulting in

$$(S - \lambda_1)^{k_1} \sum_{a=1}^{n-p} \hat{c}^a \hat{h}_a = (S - \lambda_1)^{2k_1} \sum_{a=1}^{n-p} \hat{c}^a h_{p+a} = 0,$$

and since $k_1 > 0$ we conclude from Lemma 4.4 that

$$(S - \lambda_1)^{k_1} \sum_{a=1}^{n-p} \hat{c}^a h_{p+a} = 0.$$

Hence $\sum_{a=1}^{n-p} \hat{c}^a h_{p+a} \in V_1$, and there exist constants $d^1,\ d^2, \dots,\ d^p$ such that

$$\sum_{a=1}^{n-p} \hat{c}^a h_{p+a} = \sum_{i=1}^{p} d^i h_i.$$

As the set $\{h_1, \dots, h_n\}$ is by definition a basis of V, these constants must vanish: $\hat{c}^a = d^i = 0$ for all $a = 1, \dots, n - p$ and $i = 1, \dots, p$. Substituting into (4.19), it follows from the linear independence of $h_1, \dots, h_p$ that $c^1, \dots, c^p$ all vanish as well. This proves the linear independence of the vectors in (4.17).

Let $W_1 = L(\hat{h}_1, \hat{h}_2, \dots, \hat{h}_{n-p})$. By Eq. (4.18) every vector $x \in W_1$ is of the form

$$x = (S - \lambda_1)^{k_1} y$$

since this is true of each of the vectors spanning W_1. Conversely, suppose $x = (S - \lambda_1)^{k_1} y$ and let $\{y^1, \dots, y^n\}$ be the components of y with respect to the original basis (4.16); then

$$x = (S - \lambda_1)^{k_1} \Big(\sum_i y^i h_i + \sum_a y^{p+a} h_{p+a} \Big) = \sum_a y^{p+a} \hat{h}_a \in W_1.$$

Hence W_1 consists precisely of all vectors of the form $x = (S - \lambda_1)^{k_1} y$ where y is an arbitrary vector of V. Furthermore W_1 is an invariant subspace of V, for if $x \in W_1$ then

$$x = (S - \lambda_1)^{k_1} y \implies Sx = (S - \lambda_1)^{k_1} Sy \implies Sx \in W_1.$$

Hence W_1 and V_1 are complementary invariant subspaces, $V = V_1 \oplus W_1$, and the matrix of S with respect to the basis (4.17) has block diagonal form

$$\mathsf{S} = \begin{pmatrix} \mathsf{S}_1 & \mathsf{O} \\ \mathsf{O} & \mathsf{T}_1 \end{pmatrix}$$

where S_1 is the matrix of $S_1 = S\big|_{V_1}$ and T_1 is the matrix of $T_1 = S\big|_{W_1}$.

Now on the subspace V_1 we have, by definition,

$$(S_1 - \lambda_1)^{k_1} = (S - \lambda_1)^{k_1}\Big|_{V_1} = 0.$$

Hence λ_1 is the only eigenvalue of S_1, for if $u \in V_1$ is an eigenvector of S_1 corresponding to an eigenvalue σ,

$$S_1 u = \sigma u,$$

then

$$(S_1 - \lambda_1)^{k_1} u = (\sigma - \lambda_1)^{k_1} u = 0,$$

from which it follows that $\sigma = \lambda_1$ since $u \neq 0$. The characteristic equation of S_1 is therefore

$$\det(\mathsf{S}_1 - z\mathsf{I}) = (-1)^p (\lambda_1 - z)^p. \tag{4.20}$$

Furthermore, the operator

$$(T_1 - \lambda_1)^{k_1} = (S - \lambda_1)^{k_1}\Big|_{W_1} : W_1 \to W_1$$

is invertible. For, let x be an arbitrary vector in W_1 and set $x = (S - \lambda_1)^{k_1} y$ where $y \in V$. Let $y = y_1 + y_2$ be the unique decomposition such that $y_1 \in V_1$ and $y_2 \in W_1$; then

$$x = (S - \lambda_1)^{k_1} y_2 = (T_1 - \lambda_1)^{k_1} y_2,$$

and since any surjective (onto) linear operator on a finite dimensional vector space is bijective (one-to-one), the map $(T_1 - \lambda_1)^{k_1}$ must be invertible on W_1. Hence

$$\det(\mathsf{T}_1 - \lambda_1 \mathsf{I}) \neq 0 \tag{4.21}$$

and λ_1 cannot be an eigenvalue of T_1.

The characteristic equation of S is

$$\det(\mathsf{S} - z\mathsf{I}) = \det(\mathsf{S}_1 - z\mathsf{I}) \det(\mathsf{T}_1 - z\mathsf{I})$$

and from (4.20) and (4.21) the only way the right-hand side can equal the expression in Eq. (4.13) is if

$$p_1 = p \quad \text{and} \quad \det(\mathsf{T}_1 - z\mathsf{I}) = (-1)^{n-p}(z - \lambda_2)^{p_2} \dots (z - \lambda_m)^{p_m}.$$

Hence, the dimension p_i of each space V_i is equal to the multiplicity of the eigenvalue λ_i and from the Cayley–Hamilton Theorem 4.3 it follows that $p_i \geq k_i$.

Repeating this process on T_1, and proceeding inductively, it follows that

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_m,$$

and setting

$$h_{11}, h_{12}, \dots, h_{1p_1}, h_{21}, \dots, h_{2p_2}, \dots, h_{m1}, \dots, h_{mp_m}$$

to be a basis adapted to this decomposition, the matrix of S has block diagonal form

$$\mathsf{S} = \begin{pmatrix} \mathsf{S}_1 & 0 & \dots & 0 \\ 0 & \mathsf{S}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathsf{S}_m \end{pmatrix}. \tag{4.22}$$

The restricted operators $S_i = S\big|_{V_i}$ each have a minimal polynomial equation of the form

$$(S_i - \lambda_i)^{k_i} = 0,$$

so that

$$S_i = \lambda_i \,\mathrm{id}_i + N_i \quad \text{where} \quad N_i^{k_i} = 0. \tag{4.23}$$

Nilpotent operators

Any operator N satisfying an equation of the form $N^k = 0$ is called a **nilpotent operator**. The matrix N of any nilpotent operator satisfies $\mathsf{N}^k = 0$, and is called a **nilpotent matrix**. From Eq. (4.23), each $p_i \times p_i$ matrix S_i in the decomposition (4.22) is a multiple of the unit matrix plus a nilpotent matrix N_i,

$$\mathsf{S}_i = \lambda_i \mathsf{I} + \mathsf{N}_i \quad \text{where} \quad (\mathsf{N}_i)^{k_i} = 0.$$

We next find a basis that expresses the matrix of any nilpotent operator in a standard (canonical) form.

Let U be a finite dimensional space, not necessarily complex, $\dim U = p$, and N a nilpotent operator on U. Set k to be the smallest positive integer such that $N^k = 0$. Evidently $k = 1$ if $N = 0$, while if $N \neq 0$ then $k > 1$ and $N^{k-1} \neq 0$. Define the subspaces X_i of U by

$$X_i = \{x \in U \mid N^i x = 0\} \quad (i = 0, 1, \dots, k).$$

These subspaces form an increasing sequence,

$$\{0\} = X_0 \subset X_1 \subset X_2 \subset \cdots \subset X_k = U$$

and all are invariant under N, for if $u \in X_i$ then Nu also belongs to X_i since $N^i Nu = NN^i u = N0 = 0$. The set inclusions are strict inclusion in every case, for suppose that $X_i = X_{i+1}$ for some $i \leq k - 1$. Then for any vector $x \in X_{i+2}$ we have

$$N^{i+2}x = N^{i+1}Nx = 0 \implies N^i Nx = 0 \implies x \in X_{i+1}.$$

Hence $X_{i+1} = X_{i+2}$, and continuing inductively we find that $X_i = \cdots = X_{k-1} = X_k = U$. This leads to the conclusion that $N^{k-1}x = 0$ for all $x \in U$, which contradicts the assumption that $N^{k-1} \neq 0$. Hence none of the subspaces X_i can be equal to each other.

We call a set of vectors $v_1, v_2, \dots, v_s$ belonging to X_i *linearly independent with respect to X_{i-1}* if

$$a^1 v_1 + \cdots + a^s v_s \in X_{i-1} \implies a^1 = \cdots = a^s = 0.$$

Lemma 4.5 *Set $r_i = \dim X_i - \dim X_{i-1} > 0$ so that $p = r_1 + r_2 + \cdots + r_k$. Then r_i is the maximum number of vectors in X_i that can form a set that is linearly independent with respect to X_{i-1}.*

Proof: Let $\dim X_{i-1} = q$, $\dim X_i = q' > q$ and let $u_1, \dots, u_q$ be a basis of X_{i-1}. Suppose $\{v_1, \dots, v_r\}$ is a maximal set of vectors l.i. with respect to X_{i-1}; that is, a set that cannot be extended to a larger such set. Such a maximal set must exist since any set of vectors that is l.i. with respect to X_{i-1} is linearly independent and therefore cannot exceed q in number. We show that $S = \{u_1, \dots, u_q, v_1, \dots, v_r\}$ is a basis of X_i:
(a) S is a l.i. set since

$$\sum_{i=1}^{q} a^i u_i + \sum_{a=1}^{r} b^a v_a = 0$$

implies firstly that all b^a vanish by the requirement that the vectors v_a are l.i. with respect to X_{i-1}, and secondly all the $a^i = 0$ because the u_i are l.i.
(b) S spans X_i else there would exist a vector x that cannot be expressed as a linear combination of vectors of S, and $S \cup \{x\}$ would be linearly independent. In that case, the set of vectors $\{v_1, \dots, v_r, x\}$ would be l.i. with respect to X_{i-1}, for if $\sum_{a=1}^{r} b^a v_a + bx \in X_{i-1}$ then from the linear independence of $S \cup \{x\}$ all $b^a = 0$ and $b = 0$. This contradicts the maximality of $v_1, \dots, v_r$, and S must span the whole of X_i.
This proves the lemma. ■

Let $\{h_1, \dots, h_{r_k}\}$ be a maximal set of vectors in X_k that is l.i. with respect to X_{k-1}. From Lemma 4.5 we have $r_k = \dim X_k - \dim X_{k-1}$. The vectors

$$h'_1 = Nh_1,\ h'_2 = Nh_2, \dots,\ h'_{r_k} = Nh_{r_k}$$

all belong to X_{k-1} and are l.i. with respect to X_{k-2}, for if

$$a^1 h'_1 + a^2 h'_2 + \dots + a^{r_k} h'_{r_k} \ \in X_{k-2}$$

then

$$N^{k-1}(a^1 h_1 + a^2 h_2 + \dots + a^{r_k} h_{r_k}) = 0,$$

from which it follows that

$$a^1 h_1 + a^2 h_2 + \dots + a^{r_k} h_{r_k} \ \in \ X_{k-1}.$$

Since $\{h_1, \dots, h_{r_k}\}$ are l.i. with respect to X_{k-1} we must have

$$a^1 = a^2 = \dots = a^{r_k} = 0.$$

Hence $r_{k-1} \geq r_k$. Applying the same argument to all other X_i gives

$$r_k \leq r_{k-1} \leq \dots \leq r_2 \leq r_1. \tag{4.24}$$

Now complete the set $\{h'_1, \dots, h'_{r_k}\}$ to a maximal system of vectors in X_{k-1} that is l.i. with respect to X_{k-2},

$$h'_1, \dots, h'_{r_k}, h'_{r_{k+1}}, \dots, h'_{r_{k-1}}.$$

Similarly, define the vectors $h''_i = Nh'_i$ $(i = 1, \dots, r_{k-1})$ and extend to a maximal system $\{h''_1, h''_2, \dots, h''_{r_{k-2}}\}$ in X_{k-2}. Continuing in this way, form a series of $r_1 + r_2 + \dots + r_k = p = \dim U$ vectors that are linearly independent and form a basis of U and may be displayed in the following scheme:

$$\begin{array}{cccccccccc}
h_1 & \dots & h_{r_k} & & & & & & \\
h'_1 & \dots & h'_{r_k} & \dots & h'_{r_{k-1}} & & & & \\
h''_1 & \dots & h''_{r_k} & \dots & h''_{r_{k-1}} & \dots & h''_{r_{k-2}} & & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\
h^{(k-1)}_1 & \dots & h^{(k-1)}_{r_k} & \dots & h^{(k-1)}_{r_{k-1}} & \dots & \dots & \dots & h^{(k-1)}_{r_1}.
\end{array}$$

Let U_a be the subspace generated by the ath column where $a = 1, \dots, r_1$. These subspaces are all invariant under N, since $Nh^{(j)}_a = h^{(j+1)}_a$, and the bottom elements $h^{(k-1)}_a \in X_1$ are annihilated by N,

$$Nh^{(k-1)}_a \in X_0 = \{0\}.$$

Since the vectors $h^{(i)}_a$ are linearly independent and form a basis for U, the subspaces U_a are non-intersecting, and

$$U = U_1 \oplus U_2 \oplus \dots \oplus U_{r_1},$$

where the dimension $d(a) = \dim U_a$ of the ath subspace is given by height of the ath column. In particular, $d(1) = k$ and

$$\sum_{a=1}^{r_1} d(a) = p.$$

If a basis is chosen in U_a by proceeding up the ath column starting from the vector in the bottom row,

$$f_{a1} = h_a^{(k-1)},\ f_{a2} = h_a^{(k-2)},\ \dots,\ f_{ad(a)} = h_a^{(k-d(a))},$$

then the matrix of $N_a = N\big|_{U_a}$ has all components zero except for 1's in the superdiagonal,

$$\mathsf{N}_a = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ & & \ddots & \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix}. \tag{4.25}$$

Exercise: Check this matrix representation by remembering that the components of the matrix of an operator M with respect to a basis $\{u_i\}$ are given by

$$Mu_i = M^j_{\ i} u_j.$$

Now set $M = N_a$, $u_1 = f_{a1}$, $u_2 = f_{a2}, \dots, u_{d(a)} = f_{ad(a)}$ and note that $Mu_1 = 0$, $Mu_2 = u_1$, etc.

Selecting a basis for U that runs through the subspaces $U_1, \dots, U_{r_1}$ in order,

$$e_1 = f_{11},\ e_2 = f_{12},\ \dots,\ e_k = f_{1k},\ e_{k+1} = f_{21},\ \dots,\ e_p = f_{r_1 d(r_1)},$$

the matrix of N appears in block diagonal form

$$\mathsf{N} = \begin{pmatrix} \mathsf{N}_1 & & & \\ & \mathsf{N}_2 & & \\ & & \ddots & \\ & & & \mathsf{N}_{r_1} \end{pmatrix} \tag{4.26}$$

where each submatrix N_a has the form (4.25).

Jordan canonical form

Let V be a complex vector space and $S : V \to V$ a linear operator on V. To summarize the above conclusions: there exists a basis of V such that the operator S has matrix S in block diagonal form (4.22), and each S_i has the form $\mathsf{S}_i = \lambda_i \mathsf{I} + \mathsf{N}_i$, where N_i is a nilpotent matrix. The basis can then be further specialized such that each nilpotent matrix is in turn decomposed into a block diagonal form (4.26) such that the submatrices along the diagonal all have the form (4.25). This is called the **Jordan canonical form** of the matrix S.

In other words, if S is an arbitrary $n \times n$ complex matrix, then there exists a non-singular complex matrix A such that ASA^{-1} is in Jordan form. The essential features of the matrix S can be summarized by the following **Segré characteristics**:

Eigenvalues	λ_1	...	λ_m
Multiplicities	p_1	...	p_m
	$(d_{11} \dots d_{1r_1})$	...	$(d_{m1} \dots d_{mr_m})$

where r_i is the number of eigenvectors corresponding to the eigenvalue λ_i and

$$\sum_{a=1}^{r_i} d_{ia} = p_i, \qquad \sum_{i=1}^{m} p_i = n = \dim V.$$

The Segré characteristics are determined entirely by properties of the operator S such as its eigenvalues and its elementary divisors. It is important, however, to realize that the Jordan canonical form only applies in the context of a complex vector space since it depends critically on the fundamental theorem of algebra. For a real matrix there is no guarantee that a *real* similarity transformation will convert it to the Jordan form.

Example 4.6 Let S be a transformation on a four-dimensional vector space having matrix with respect to a basis $\{e_1, e_2, e_3, e_4\}$ whose components are

$$\mathsf{S} = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

The characteristic equation can be written in the form

$$\det(\mathsf{S} - \lambda\mathsf{I}) = ((\lambda - 1)^2 + 1)^2 = 0,$$

which has two roots $\lambda = 1 \pm i$, both of which are repeated roots. Each root corresponds to just a single eigenvector, written in column vector form as

$$\mathbf{f}_1 = \begin{pmatrix} 1 \\ 0 \\ -i \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{f}_3 = \begin{pmatrix} 1 \\ 0 \\ i \\ 0 \end{pmatrix},$$

satisfying

$$\mathsf{S}\mathbf{f}_1 = (1 + i)\mathbf{f}_1 \quad \text{and} \quad \mathsf{S}\mathbf{f}_3 = (1 - i)\mathbf{f}_3.$$

Let $\mathbf{f}_2$ and $\mathbf{f}_4$ be the vectors

$$\mathbf{f}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -i \end{pmatrix} \quad \text{and} \quad \mathbf{f}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ i \end{pmatrix},$$

and we find that

$$\mathsf{S}\mathbf{f}_2 = \mathbf{f}_1 + (1+i)\mathbf{f}_2 \quad \text{and} \quad \mathsf{S}\mathbf{f}_4 = \mathbf{f}_3 + (1-i)\mathbf{f}_4.$$

Expressing these column vectors in terms of the original basis

$$f_1 = e_1 - ie_3,\ \ f_2 = e_2 - ie_4,\ \ f_3 = e_1 + ie_3,\ \ f_4 = e_2 + ie_4$$

provides a new basis with respect to which the matrix of operator S has block diagonal Jordan form

$$\mathsf{S}' = \begin{pmatrix} 1+i & 1 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 1 \\ 0 & 0 & 0 & 1-i \end{pmatrix}.$$

The matrix A needed to accomplish this form by the similarity transformation $\mathsf{S}' = \mathsf{A}\mathsf{S}\mathsf{A}^{-1}$ is found by solving for the e_i in terms of the f_j,

$$e_1 = \tfrac{1}{2}(f_1 + f_3), \quad e_2 = \tfrac{1}{2}(f_2 + f_4), \quad e_3 = \tfrac{1}{2}i(f_1 - f_3), \quad e_4 = \tfrac{1}{2}i(f_2 - f_4),$$

which can be written

$$e_j = A^i{}_j f_i \quad \text{where} \quad \mathsf{A} = [A^i{}_j] = \frac{1}{2}\begin{pmatrix} 1 & 0 & i & 0 \\ 0 & 1 & 0 & i \\ 1 & 0 & -i & 0 \\ 0 & 1 & 0 & -i \end{pmatrix}.$$

The matrix S' is summarized by the Segré characteristics:

$1+i$	$1-i$
2	2
(2)	(2)

Exercise: Verify that $\mathsf{S}' = \mathsf{A}\mathsf{S}\mathsf{A}^{-1}$ in Example 4.6.

Problems

Problem 4.2 On a vector space V let S and T be two commuting operators, $ST = TS$.

(a) Show that if v is an eigenvector of T then so is Sv.
(b) Show that a basis for V can be found such that the matrices of both S and T with respect to this basis are in upper triangular form.

Problem 4.3 For the operator $T : V \to V$ on a four-dimensional vector space given in Problem 3.10, show that no basis exists such that the matrix of T is diagonal. Find a basis in which the matrix of T has the Jordan form

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$$

for some λ, and calculate the value of λ.

Problem 4.4 Let S be the matrix

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

Find the minimal annihilating polynomial and the characteristic polynomial of this matrix, its eigenvalues and eigenvectors, and find a basis that reduces it to its Jordan canonical form.

4.3 Linear ordinary differential equations

While no techniques exist for solving general differential equations, systems of linear ordinary differential equations with constant coefficients are completely solvable with the help of the Jordan form. Such systems can be written in the form

$$\dot{\mathbf{x}} \equiv \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathsf{A}\mathbf{x}, \tag{4.27}$$

where $\mathbf{x}(t)$ is an $n \times 1$ column vector and A an $n \times n$ matrix of real constants. Initially it is best to consider this as an equation in complex variables $\mathbf{x}$, even though we may only be seeking real solutions. Greater details of the following discussion, as well as applications to non-linear differential equations, can be found in [4, 5].

Try for a solution of (4.27) in exponential form

$$\mathbf{x}(t) = \mathrm{e}^{\mathsf{A}t}\mathbf{x}_0, \tag{4.28}$$

where $\mathbf{x}_0$ is an arbitrary constant vector, and the exponential of a matrix is defined by the convergent series

$$\mathrm{e}^{\mathsf{S}} = \mathsf{I} + \mathsf{S} + \frac{\mathsf{S}^2}{2!} + \frac{\mathsf{S}^3}{3!} + \dots \tag{4.29}$$

If S and T are two commuting matrices, $\mathsf{ST} = \mathsf{TS}$, it then follows just as for real or complex scalars that $\mathrm{e}^{\mathsf{ST}} = \mathrm{e}^{\mathsf{S}}\mathrm{e}^{\mathsf{T}}$.

The initial value at $t = 0$ of the solution given by Eq. (4.28) is clearly $\mathbf{x}(0) = \mathbf{x}_0$. If P is any invertible $n \times n$ matrix, then $\mathbf{y} = \mathsf{P}\mathbf{x}$ satisfies the differential equation

$$\dot{\mathbf{y}} = \mathsf{A}'\mathbf{y} \quad \text{where} \quad \mathsf{A}' = \mathsf{PAP}^{-1},$$

and gives rise to the solution

$$\mathbf{y} = \mathrm{e}^{\mathsf{A}'t}\mathbf{y}_0 \quad \text{where} \quad \mathbf{y}_0 = \mathsf{P}\mathbf{x}_0.$$

If P is chosen such that A' has the Jordan form

$$\mathsf{A}' = \begin{pmatrix} \lambda_1 \mathsf{I} + \mathsf{N}_{11} & & \\ & \ddots & \\ & & \lambda_m \mathsf{I} + \mathsf{N}_{mr_m} \end{pmatrix}$$

where the N_{ij} are nilpotent matrices then, since $\lambda_i \mathsf{I}$ commutes with N_{ij} for every $i,\ j$, the exponential term has the form

$$e^{\mathsf{A}'t} = \begin{pmatrix} e^{\lambda_1 t} e^{\mathsf{N}_{11} t} & & \\ & \ddots & \\ & & e^{\lambda_m t} e^{\mathsf{N}_{mr_m} t} \end{pmatrix}.$$

If N is a $k \times k$ Jordan matrix having 1's along the superdiagonal as in (4.25), then N^2 has 1's in the next diagonal out and each successive power of N pushes this diagonal of 1's one place further until N^k vanishes altogether,

$$\mathsf{N}^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \dots, \quad \mathsf{N}^{k-1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ & & & \ddots & \\ & & & \dots & 0 \end{pmatrix}, \quad \mathsf{N}^k = \mathsf{O}.$$

Hence

$$e^{\mathsf{N}t} = \begin{pmatrix} 1 & t & \frac{t^2}{2} & \dots & \frac{t^{k-1}}{(k-1)!} \\ 0 & 1 & t & \dots & \\ & & & \ddots & \\ 0 & 0 & & \dots & 1 \end{pmatrix},$$

and the solution (4.28) can be expressed as a linear supposition of solutions of the form

$$\mathbf{x}_r(t) = \mathbf{w}_r(t) e^{\lambda_i t} \tag{4.30}$$

where

$$\mathbf{w}_r(t) = \frac{t^{r-1}}{(r-1)!}\mathbf{h}_1 + \frac{t^{r-2}}{(r-2)!}\mathbf{h}_2 + \dots + t\mathbf{h}_{r-1} + \mathbf{h}_r. \tag{4.31}$$

If A is a *real* matrix then the matrices P and A' are in general complex, but given real initial values $\mathbf{x}_0$ the solution having these values at $t = 0$ is

$$\mathbf{x} = \mathsf{P}^{-1}\mathbf{y} = \mathsf{P}^{-1} e^{\mathsf{A}'t} \mathsf{P} \mathbf{x}_0,$$

which must necessarily be real by the existence and uniqueness theorem of ordinary differential equations. Alternatively, for A real, both the real and imaginary parts of any complex solution $\mathbf{x}(t)$ are solutions of the linear differential equation (4.27), which may be separated by the identity

$$e^{\lambda t} = \cos \lambda t + i \sin \lambda t.$$

Two-dimensional autonomous systems

Consider the special case of a planar (two-dimensional) system (4.27) having constant coefficients, known as an **autonomous system**,

$$\dot{\mathbf{x}} = \mathsf{A}\mathbf{x} \quad \text{where} \quad \mathsf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Both the matrix A and vector $\mathbf{x}$ are assumed to be real. A **critical point** $\mathbf{x}_0$ refers to any constant solution $\mathbf{x} = \mathbf{x}_0$ of (4.27). The analysis of autonomous systems breaks up into a veritable zoo of cases and subcases. We consider the case where the matrix A is non-singular, for which the only critical point is $\mathbf{x}_0 = \mathbf{0}$. Both eigenvalues λ_1 and λ_2 are $\neq 0$, and the following possibilities arise.

(1) $\lambda_1 \neq \lambda_2$ and both eigenvalues are real. In this case the eigenvectors $\mathbf{h}_1$ and $\mathbf{h}_2$ form a basis of $\mathbb{R}^2$ and the general solution is

$$\mathbf{x} = c_1\, e^{\lambda_1 t}\mathbf{h}_1 + c_2\, e^{\lambda_2 t}\mathbf{h}_2.$$

(1a) If $\lambda_2 < \lambda_1 < 0$ the critical point is called a **stable node**.
(1b) If $\lambda_2 > \lambda_1 > 0$ the critical point is called an **unstable node**.
(1c) If $\lambda_2 < 0 < \lambda_1$ the critical point is called a **saddle point**.

These three cases are shown in Fig. 4.1, after the basis of the vector space axes has been transformed to lie along the vectors $\mathbf{h}_i$.

(2) $\lambda_1 = \lambda$, $\lambda_2 = \overline{\lambda}$ where λ is complex. The eigenvectors are then complex conjugate to each other since A is a real matrix,

$$\mathsf{A}\mathbf{h} = \lambda\mathbf{h} \;\Rightarrow\; \mathsf{A}\overline{\mathbf{h}} = \overline{\lambda}\overline{\mathbf{h}},$$

and the arbitrary real solution is

$$\mathbf{x} = c\, e^{\lambda t}\mathbf{h} + \overline{c}\, e^{\overline{\lambda} t}\overline{\mathbf{h}}.$$

If we set

$$\mathbf{h} = \tfrac{1}{2}(\mathbf{h}_1 - i\mathbf{h}_2), \qquad \lambda = \mu + i\nu, \qquad c = Re^{i\alpha}$$

where $\mathbf{h}_1$, $\mathbf{h}_2$, μ, ν, $R > 0$ and α are all real quantities, then the solution $\mathbf{x}$ has the form

$$\mathbf{x} = R\, e^{\mu t}\big(\cos(\nu t + \alpha)\mathbf{h}_1 + \sin(\nu t + \alpha)\mathbf{h}_2\big).$$

(2a) $\mu < 0$: This is a logarithmic spiral approaching the critical point $\mathbf{x} = \mathbf{0}$ as $t \to \infty$, and is called a **stable focus**.
(2b) $\mu > 0$: Again the solution is a logarithmic spiral but arising from the critical point $\mathbf{x} = \mathbf{0}$ as $t \to -\infty$, called an **unstable focus**.
(2c) $\mu = 0$: With respect to the basis $\mathbf{h}_1$, $\mathbf{h}_2$, the solution is a set of circles about the origin. When the original basis $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is used, the solutions are a set of ellipses and the critical point is called a **vortex point**.

These solutions are depicted in Fig. 4.2.

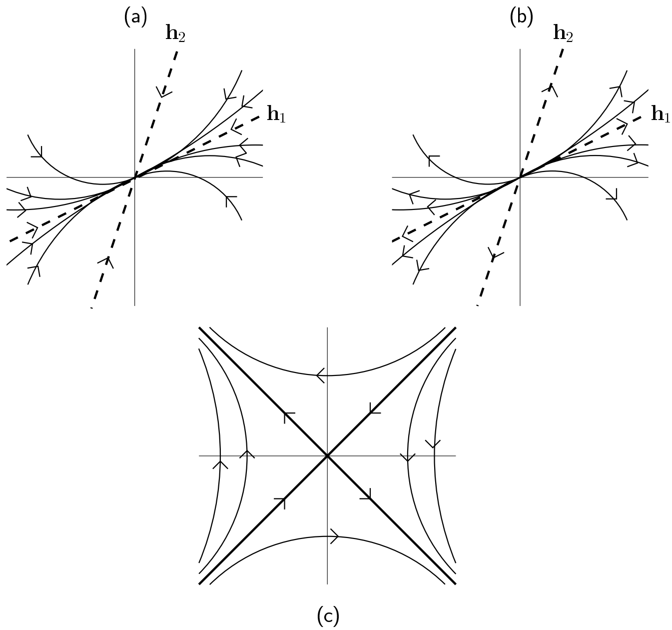

Figure 4.1 (a) Stable node, (b) unstable node, (c) saddle point

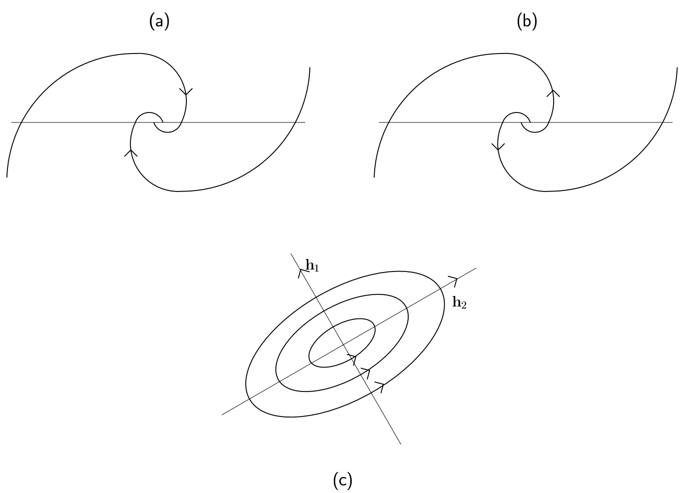

Figure 4.2 (a) Stable focus, (b) unstable focus, (c) vortex point

Problems

Problem 4.5 Verify that (4.30), (4.31) is a solution of $\dot{\mathbf{x}}_r = \mathsf{A}\mathbf{x}_r(t)$ provided

$$\begin{aligned}\mathsf{A}\mathbf{h}_1 &= \lambda_i \mathbf{h}_1\\ \mathsf{A}\mathbf{h}_2 &= \lambda_i \mathbf{h}_2 + \mathbf{h}_1\\ &\vdots\\ \mathsf{A}\mathbf{h}_r &= \lambda_i \mathbf{h}_r + \mathbf{h}_{r-1}\end{aligned}$$

where λ_i is an eigenvalue of A.

Problem 4.6 Discuss the remaining cases for two-dimensional autonomous systems: (a) $\lambda_1 = \lambda_2 = \lambda \neq 0$ and (i) two distinct eigenvectors $\mathbf{h}_1$ and $\mathbf{h}_2$, (ii) only one eigenvector $\mathbf{h}_1$; (b) A a singular matrix. Sketch the solutions in all instances.

Problem 4.7 Classify all three-dimensional autonomous systems of linear differential equations having constant coefficients.

4.4 Introduction to group representation theory

Groups appear most frequently in physics through their actions on vector spaces, known as *representations*. More specifically, a **representation** of any group G on a vector space V is a homomorphism T of G into the group of linear automorphisms of V,

$$T : G \to GL(V).$$

For every group element g we then have a corresponding linear transformation $T(g) : V \to V$ such that

$$T(g)T(h)v = T(gh)v \qquad \text{for all } g, h \in G,\ v \in V.$$

Essentially, a representation of an abstract group is a way of providing a concrete model of the elements of the group as linear transformations of a vector space. The representation is said to be **faithful** if it is one-to-one; that is, if $\ker T = \{e\}$. While in principal V could be either a real or complex vector space, we will mostly consider representations on complex vector spaces. If V is finite-dimensional its dimension n is called the **degree** of the representation. We will restrict attention almost entirely to representations of finite degree. Group representation theory is developed in much greater detail in [6–8].

Exercise: Show that any representation T induces a faithful representation of the factor group $G/\ker T$ on V.

Two representations $T_1 : G \to V_1$ and $T_2 : G \to V_2$ are said to be **equivalent**, written $T_1 \sim T_2$, if there exists a vector space isomorphism $A : V_1 \to V_2$ such that

$$T_2(g)A = AT_1(g), \quad \text{for all } g \in G. \tag{4.32}$$

If $V_1 = V_2 = V$ then $T_2(g) = AT_1(g)A^{-1}$. For finite dimensional representations the matrices representing T_2 are then derived from those representing T_1 by a similarity transformation. In this case the two representations can be thought of as essentially identical, since they are related simply by a change of basis.

Any operator A, even if it is singular, which satisfies Eq. (4.32) is called an **intertwining operator** for the two representations. This condition is frequently depicted by a commutative diagram

$$\begin{array}{ccc} V_1 & \xrightarrow{A} & V_2 \\ {\scriptstyle T_1(g)}\downarrow & & \downarrow{\scriptstyle T_2(g)} \\ V_1 & \xrightarrow[A]{} & V_2 \end{array}$$

Irreducible representations

A subspace W of V is said to be **invariant** under the action G, or G**-invariant**, if it is invariant under each linear transformation $T(g)$,

$$T(g)W \subseteq W \qquad \text{for all } g \in G.$$

For every $g \in G$ the map $T(g)$ is surjective, $T(g)W = W$, since $w = T(g)(T(g^{-1}w))$ for every vector $w \in W$. Hence the restriction of $T(g)$ to W is an automorphism of W and provides another representation of G, called a **subrepresentation** of G, denoted $T_W : G \to GL(W)$. The whole space V and the trivial subspace $\{0\}$ are clearly G-invariant for any representation T on V. If these are the only invariant subspaces the representation is said to be **irreducible**.

If W is an invariant subspace of a representation T on V then a representation is induced on the quotient space V/W, defined by

$$T_{V/W}(g)(v + W) = T(g)v + W \qquad \text{for all } g \in G, v \in V.$$

Exercise: Verify that this definition is independent of the choice of representative from the coset $v + W$, and that it is indeed a representation.

Let V be finite dimensional, $\dim V = n$, and $W' \cong V/W$ be a complementary subspace to W, such that $V = W \oplus W'$. From Theorem 3.7 and Example 3.22 there exists a basis whose first $r = \dim W$ vectors span W while the remaining $n - r$ span W'. The matrices of the representing transformations with respect to such a basis will have the form

$$\mathsf{T}(g) = \begin{pmatrix} \mathsf{T}_W(g) & \mathsf{S}(g) \\ \mathsf{O} & \mathsf{T}_{W'}(g) \end{pmatrix}.$$

The submatrices $\mathsf{T}_{W'}(g)$ form a representation on the subspace W' that is equivalent to the quotient space representation, but W' is not in general G-invariant because of the existence of the off-block diagonal matrices G. If $\mathsf{S}(g) \neq \mathsf{O}$ then it is essentially impossible to recover the original representation purely from the subrepresentations on W and W'. Matters are much improved, however, if the complementary subspace W' is G-invariant as well as W. In this case the representing matrices have the block diagonal form in a basis adapted to W

and W',

$$\mathsf{T}(g) = \begin{pmatrix} \mathsf{T}_W(g) & \mathsf{O} \\ \mathsf{O} & \mathsf{T}_{W'}(g) \end{pmatrix}$$

and the representation is said to be **completely reducible**.

Example 4.7 Let the map $T : \mathbb{R} \to GL(\mathbb{R}^2)$ be defined by

$$T : a \mapsto \mathsf{T}(a) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}.$$

This is a representation since $\mathsf{T}(a)\mathsf{T}(b) = \mathsf{T}(a+b)$. The subspace of vectors of the form $\begin{pmatrix} x \\ 0 \end{pmatrix}$ is invariant, but there is no complementary invariant subspace – for example, vectors of the form $\begin{pmatrix} 0 \\ y \end{pmatrix}$ are not invariant under the matrices $\mathsf{T}(a)$. Equivalently, it follows from the Jordan canonical form that no matrix A exists such that $\mathsf{A}\mathsf{T}(1)\mathsf{A}^{-1}$ is diagonal. The representation T is thus an example of a representation that is reducible but not completely reducible.

Example 4.8 The symmetric group of permutations on three objects, denoted S_3, has a representation T on a three-dimensional vector space V spanned by vectors e_1, e_2 and e_3, defined by

$$T(\pi)e_i = e_{\pi(i)}.$$

In this basis the matrix of the transformation $T(\pi)$ is $\mathsf{T} = [T^j_i(\pi)]$, where

$$T(\pi)e_i = T^j_i(\pi)e_j.$$

Using cyclic notation for permutations the elements of S_3 are $e = \text{id}$, $\pi_1 = (1\ 2\ 3)$, $\pi_2 = (1\ 3\ 2)$, $\pi_3 = (1\ 2)$, $\pi_4 = (2\ 3)$, $\pi_5 = (1\ 3)$. Then $T(e)e_i = e_i$, so that $\mathsf{T}(e)$ is the identity matrix I, while $T(\pi_1)e_1 = e_2$, $T(\pi_1)e_2 = e_3$, $T(\pi_1)e_3 = e_1$, etc. The matrix representations of all permutations of S_3 are

$$\mathsf{T}(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{T}(\pi_1) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathsf{T}(\pi_2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\mathsf{T}(\pi_3) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{T}(\pi_4) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathsf{T}(\pi_5) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Let $v = v^i e_i$ be any vector, then $T(\pi)v = T^j_i(\pi)v^i e_j$ and the action of the matrix $\mathsf{T}(\pi)$ is left multiplication on the column vector

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix}.$$

We now find the invariant subspaces of this representation. In the first place any one-dimensional invariant subspace must be spanned by a vector v that is an eigenvector of each operator $T(\pi_i)$. In matrices,

$$\mathsf{T}(\pi_1)\mathbf{v} = \alpha\mathbf{v} \implies v^1 = \alpha v^2, \quad v^2 = \alpha v^3, \quad v^3 = \alpha v^1,$$

whence

$$v^1 = \alpha v^2 = \alpha^2 v^3 = \alpha^3 v^1.$$

Similarly $v^2 = \alpha^3 v^2$ and $v^3 = \alpha^3 v^3$, and since $\mathbf{v} \neq \mathbf{0}$ we must have that $\alpha^3 = 1$. Since $\alpha \neq 0$ it follows that all three components v^1, v^2 and v^3 are non-vanishing. A similar argument gives

$$\mathsf{T}(\pi_3)v = \beta v \implies v^1 = \beta v^2, \quad v^2 = \beta v^1, \quad v^3 = \beta v^3,$$

from which $\beta^2 = 1$, and $\alpha\beta = 1$ since $v^1 = \alpha v^2 = \alpha\beta v^1$. The only pair of complex numbers α and β satisfying these relations is $\alpha = \beta = 1$. Hence $v^1 = v^2 = v^3$ and the only one-dimensional invariant subspace is that spanned by $v = e_1 + e_2 + e_3$.

We shall now show that this representation is completely reducible by choosing the basis

$$f_1 = e_1 + e_2 + e_3, \qquad f_2 = e_1 - e_2, \qquad f_3 = e_1 + e_2 - 2e_3.$$

The inverse transformation is

$$e_1 = \frac{1}{3}f_1 + \frac{1}{2}f_2 + \frac{1}{6}f_3, \qquad e_2 = \frac{1}{3}f_1 - \frac{1}{2}f_2 + \frac{1}{6}f_3, \qquad e_3 = \frac{1}{3}f_1 - \frac{1}{3}f_3,$$

and the matrices representing the elements of S_3 are found by calculating the effect of the various transformations on the basis elements f_i. For example,

$$\begin{aligned}
T(e)f_1 = f_1, \qquad T(e)f_2 &= f_2, \qquad T(e)f_3 = f_3,\\
T(\pi_1)f_1 &= T(\pi_1)(e_1 + e_2 + e_3)\\
&= e_2 + e_3 + e_1 = f_1,\\
T(\pi_1)f_2 &= T(\pi_1)(e_1 - e_2)\\
&= e_2 - e_3 = \frac{1}{2}f_2 + \frac{1}{2}f_3,\\
T(\pi_1)f_3 &= T(\pi_1)(e_1 + e_2 - 2e_3)\\
&= e_2 + e_3 - 2e_1 = -\frac{3}{2}f_2 - \tfrac{1}{2}f_3, \text{ etc.}
\end{aligned}$$

Continuing in this way for all $T(\pi_i)$ we arrive at the following matrices:

$$\mathsf{T}(e) = \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{T}(\pi_1) = \begin{pmatrix} 1 & 0 & 0\\ 0 & -\frac{1}{2} & -\frac{3}{2}\\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \qquad \mathsf{T}(\pi_2) = \begin{pmatrix} 1 & 0 & 0\\ 0 & -\frac{1}{2} & \frac{3}{2}\\ 0 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix},$$

$$\mathsf{T}(\pi_3) = \begin{pmatrix} 1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{T}(\pi_4) = \begin{pmatrix} 1 & 0 & 0\\ 0 & \frac{1}{2} & \frac{3}{2}\\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \qquad \mathsf{T}(\pi_5) = \begin{pmatrix} 1 & 0 & 0\\ 0 & \frac{1}{2} & -\frac{3}{2}\\ 0 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

The two-dimensional subspace spanned by f_2 and f_3 is thus invariant under the action of S_3 and the representation T is completely reducible.

Exercise: Show that the representation T restricted to the subspace spanned by f_2 and f_3 is irreducible, by showing that there is no invariant one-dimensional subspace spanned by f_2 and f_3.

Schur's lemma

The following key result and its corollary are useful in the classification of irreducible representations of groups.

Theorem 4.6 (Schur's lemma) *Let $T_1 : G \to GL(V_1)$ and $T_2 : G \to GL(V_2)$ be two irreducible representations of a group G, and $A : V_1 \to V_2$ an intertwining operator such that*

$$T_2(g)A = AT_1(g) \quad \textit{for all} \quad g \in G.$$

Then either $A = 0$ or A is an isomorphism, in which case the two representations are equivalent, $T_1 \sim T_2$.

Proof: Let $v \in \ker A \subseteq V_1$. Then

$$AT_1(g)v = T_2(g)Av = 0$$

so that $T_1(g)v \in \ker A$. Hence $\ker A$ is an invariant subspace of the representation T_1. As T_1 is an irreducible representation we have that either $\ker A = V_1$, in which case $A = 0$, or $\ker A = \{0\}$. In the latter case A is one-to-one. To show it is an isomorphism it is only necessary to show that it is onto. This follows from the fact that $\operatorname{im} A \subset V_2$ is an invariant subspace of the representation T_2,

$$T_2(g)(\operatorname{im} A) = T_2(g)A(V_1) = AT_1(g)(V_1) \subseteq A(V_1) = \operatorname{im} A.$$

Since T_2 is an irreducible representation we have either $\operatorname{im} A = \{0\}$ or $\operatorname{im} A = V_2$. In the first case $A = 0$, while in the second A is onto. Schur's lemma is proved. ■

Corollary 4.7 *Let $T : G \to GL(V)$ be a representation of a finite group G on a complex vector space V and $A : V \to V$ an operator that commutes with all $T(g)$; that is, $AT(g) = T(g)A$ for all $g \in G$. Then $A = \alpha \mathrm{id}_V$ for some complex scalar α.*

Proof: Set $V_1 = V_2 = V$ and $T_1 = T_2 = T$ in Schur's lemma. Since $AT(g) = T(g)A$ we have

$$(A - \alpha \mathrm{id}_V)T(g) = T(g)(A - \alpha \mathrm{id}_V)$$

since id_V commutes with all linear operators on V. By Theorem 4.6 either $A - \alpha \mathrm{id}_V$ is invertible or it is zero. Let α be an eigenvalue of A – for operators on a complex vector space this is always possible. The operator $A - \alpha \mathrm{id}_V$ is not invertible, for if it is applied to a corresponding eigenvector the result is the zero vector. Hence $A - \alpha \mathrm{id}_V = 0$, which is the desired result. ■

It should be observed that the proof of this corollary only holds for complex representations since real matrices do not necessarily have any real eigenvalues.

Example 4.9 If G is a finite abelian group then all its irreducible representations are one-dimensional. This follows from Corollary 4.7, for if $T : G \to GL(V)$ is any representation of G then any $T(h)$ ($h \in G$) commutes with all $T(g)$ and is therefore a multiple of the identity,

$$T(h) = \alpha(h)\mathrm{id}_V.$$

Hence any vector $v \in V$ is an eigenvector of $T(h)$ for all $h \in G$ and spans an invariant one-dimensional subspace of V. Thus, if $\dim V > 1$ the representation T cannot be irreducible.

References

[1] P. R. Halmos. *Finite-dimensional Vector Spaces*. New York, D. Van Nostrand Company, 1958.
[2] S. Hassani. *Foundations of Mathematical Physics*. Boston, Allyn and Bacon, 1991.
[3] F. P. Hildebrand. *Methods of Applied Mathematics*. Englewood Cliffs, N. J., Prentice-Hall, 1965.
[4] L. S. Pontryagin. *Ordinary Differential Equations*. New York, Addison-Wesley, 1962.
[5] D. A. Sánchez. *Ordinary Differential Equations and Stability Theory: An Introduction*. San Francisco, W. H. Freeman and Co., 1968.
[6] S. Lang. *Algebra*. Reading, Mass., Addison-Wesley, 1965.
[7] M. Hammermesh. *Group Theory and its Applications to Physical Problems*. Reading, Mass., Addison-Wesley, 1962.
[8] S. Sternberg. *Group Theory and Physics*. Cambridge, Cambridge University Press, 1994.