3 Vector spaces

Some algebraic structures have more than one law of composition. These must be connected by some kind of *distributive laws*, else the separate laws of composition are simply independent structures on the same set. The most elementary algebraic structures of this kind are known as *rings* and *fields*, and by combining fields and abelian groups we create *vector spaces* [1–7].

For the rest of this book, vector spaces will never be far away. For example, *Hilbert spaces* are structured vector spaces that form the basis of *quantum mechanics*. Even in non-linear theories such as *classical mechanics* and *general relativity* there exist local vector spaces known as the *tangent space* at each point, which are needed to formulate the dynamical equations. It is hard to think of a branch of physics that does not use vector spaces in some aspect of its formulation.

3.1 Rings and fields

A **ring** R is a set with two laws of composition called *addition* and *multiplication*, denoted $a+b$ and ab respectively. It is required that R is an abelian group with respect to $+$, with identity element 0 and inverses denoted $-a$. With respect to multiplication R is to be a commutative semigroup, so that the identity and inverses are not necessarily present. In detail, the requirements of a ring are:

(R1) Addition is associative, $(a+b)+c = a+(b+c)$.
(R2) Addition is commutative, $a+b = b+a$.
(R3) There is an element 0 such that $a+0 = a$ for all $a \in R$.
(R4) For each $a \in R$ there exists an element $-a$ such that $a-a \equiv a+(-a) = 0$.
(R5) Multiplication is associative, $(ab)c = a(bc)$.
(R6) Multiplication is commutative, $ab = ba$.
(R7) The **distributive law** holds, $a(b+c) = ab+ac$. By (R6) this also implies $(a+b)c = ac+bc$. It is the key relation linking the two laws of composition, addition and multiplication.

As shown in Chapter 2, the additive identity 0 is unique. From these axioms we also have that $0a = 0$ for all $a \in R$, for by (R1), R(3), (R4) and (R7)

$$0a = 0a + 0 = 0a + 0a - 0a = (0+0)a - 0a = 0a - 0a = 0.$$

Example 3.1 The integers $\mathbb{Z}$ form a ring with respect to the usual operation of addition and multiplication. This ring has a (multiplicative) identity 1, having the property $1a = a1 = a$ for all $a \in \mathbb{Z}$. The set $2\mathbb{Z}$ consisting of all even integers also forms a ring, but now there is no identity.

Example 3.2 The set M_n of all $n \times n$ real matrices forms a ring with addition of matrices $\mathsf{A} + \mathsf{B}$ and matrix product AB defined in the usual way. This is a ring with identity I, the unit matrix.

Example 3.3 The set of all real-valued functions on a set S, denoted $\mathcal{F}(S)$, forms a ring with identity. Addition and multiplication of functions $f + g$, fg are defined in the usual way,

$$(f + g)(x) = f(x) + g(x), \qquad (fg)(x) = f(x)g(x).$$

The 0 element is the zero function whose value on every $x \in S$ is the number zero, while the identity is the function having the value 1 at each $x \in S$.

These examples of rings all fail to be groups with respect to multiplication, for even when they have a multiplicative identity 1, it is almost never true that the zero element 0 has an inverse.

Exercise: Show that if 0^{-1} exists in a ring R with identity then $0 = 1$ and R must be the trivial ring consisting of just one element 0.

A **field** $\mathbb{K}$ is a ring with a multiplicative identity 1, in which every element $a \neq 0$ has an inverse $a^{-1} \in \mathbb{K}$ such that $aa^{-1} = 1$. It not totally clear why the words ‘rings’ and ‘fields’ are used to describe these algebraic entities. However, the word ‘field’ is a perhaps a little unfortunate as it has nothing whatsoever to do with expressions such as ‘electromagnetic field’, commonly used in physics.

Example 3.4 The real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$ both form fields with respect to the usual rules of addition and multiplication. These are essentially the only fields of interest in this book. We will frequently use the symbol $\mathbb{K}$ to refer to a field which could be *either* $\mathbb{R}$ or $\mathbb{C}$.

Problems

Problem 3.1 Show that the integers modulo a prime number p form a finite field.

Problem 3.2 Show that the set of all real numbers of the form $a + b\sqrt{2}$, where a and b are rational numbers, is a field. If a and b are restricted to the integers show that this set is a ring, but is not a field.

3.2 Vector spaces

A **vector space** $(V, \mathbb{K})$ consists of an additive abelian group V whose elements $u, v, \dots$ are called **vectors** together with a field $\mathbb{K}$ whose elements are termed **scalars**. The law of composition $u + v$ defining the abelian group is called **vector addition**. There is also an operation $\mathbb{K} \times V \to V$ called **scalar multiplication**, which assigns a vector $au \in V$ to any pair $a \in \mathbb{K}$, $u \in V$. The identity element 0 for vector addition, satisfying $0 + u = u$ for all vectors u, is termed the **zero vector**, and the inverse of any vector u is denoted $-u$. In principle there can be a minor confusion in the use of the same symbol $+$ for vector addition and scalar addition, and the same symbol 0 both for the zero vector and the zero scalar. It should, however, always be clear from the context which is being used. A similar remark applies to scalar multiplication au and field multiplication of scalars ab. The full list of axioms to be satisfied by a vector space is:

(VS1) For all $u, v, w \in V$ and $a, b, c \in \mathbb{K}$,

$$\begin{array}{ccc} u+(v+w)=(u+v)+w & a+(b+c)=(a+b)+c & a(bc)=(ab)c \\ u+v=v+u & a+b=b+a & ab=ba \\ u+0=0+u=u & a+0=0+a=a & a1=1a=a \\ u+(-u)=0; & a+(-a)=0; & a(b+c)=ab+ac. \end{array}$$

(VS2) $a(u + v) = au + av$.
(VS3) $(a + b)u = au + bu$.
(VS4) $a(bv) = (ab)v$.
(VS5) $1v = v$.

A vector space $(V, \mathbb{K})$ is often referred to as a **vector space V over a field $\mathbb{K}$** or simply a vector space V when the field of scalars is implied by some introductory phrase such as 'let V be a real vector space', or 'V is a complex vector space'.

Since $v = (1 + 0)v = v + 0v$ it follows that $0v = 0$ for any vector $v \in V$. Furthermore, $(-1)v$ is the additive inverse of v since, by (VS3), $(-1)v + v = (-1 + 1)v = 0v = 0$. It is also common to write $u - v$ in place of $u + (-v)$, so that $u - u = 0$. Vectors are often given distinctive notations such as $\mathbf{u}, \mathbf{v}, \dots$ or $\vec{u}, \vec{v}, \dots$, etc. to distinguish them from scalars, but we will only adopt such notations in specific instances.

Example 3.5 The set $\mathbb{K}^n$ of all n-tuples $\mathbf{x} = (x_1, x_2, \dots, x_n)$ where $x_i \in \mathbb{K}$ is a vector space, with vector addition and scalar multiplication defined by

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
$$a\mathbf{x} = (ax_1, ax_2, \dots, ax_n).$$

Specific instances are $\mathbb{R}^n$ or $\mathbb{C}^n$. Sometimes the vectors of $\mathbb{K}^n$ will be represented by $n \times 1$ column matrices and there are some advantages in denoting the components by superscripts,

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

Scalar multiplication and addition of vectors is then

$$a\mathbf{x} = \begin{pmatrix} ax^1 \\ ax^2 \\ \vdots \\ ax^n \end{pmatrix}, \qquad \mathbf{x} + \mathbf{y} = \begin{pmatrix} x^1 + y^1 \\ x^2 + y^2 \\ \vdots \\ x^n + y^n \end{pmatrix}.$$

Exercise: Verify that all axioms of a vector space are satisfied by $\mathbb{K}^n$.

Example 3.6 Let $\mathbb{K}^\infty$ denote the set of all sequences $u = (u_1, u_2, u_3, \dots)$ where $u_i \in \mathbb{K}$. This is a vector space if vector addition and scalar multiplication are defined as in Example 3.5:

$$u + v = (u_1 + v_1, u_2 + v_2, u_3 + v_3, \dots),$$
$$a\,u = (au_1, au_2, au_3, \dots).$$

Example 3.7 The set of all $m \times n$ matrices over the field $\mathbb{K}$, denoted $M^{(m,n)}(\mathbb{K})$, is a vector space. In this case vectors are denoted by $\mathsf{A} = [a_{ij}]$ where $i = 1, \dots, m$, $j = 1, \dots, n$ and $a_{ij} \in \mathbb{K}$. Addition and scalar multiplication are defined by:

$$\mathsf{A} + \mathsf{B} = [a_{ij} + b_{ij}], \qquad c\mathsf{A} = [c\,a_{ij}].$$

Although it may seem a little strange to think of a matrix as a 'vector', this example is essentially no different from Example 3.5, except that the sequence of numbers from the field $\mathbb{K}$ is arranged in a rectangular array rather than a row or column.

Example 3.8 Real-valued functions on $\mathbb{R}^n$, denoted $\mathcal{F}(\mathbb{R}^n)$, form a vector space over $\mathbb{R}$. As described in Section 1.4, the vectors in this case can be thought of as functions of n arguments,

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n),$$

and vector addition $f + g$ and scalar multiplication af are defined in the obvious way,

$$(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}), \qquad (af)(\mathbf{x}) = af(\mathbf{x}).$$

The verification of the axioms of a vector space is a straightforward exercise.

More generally, if S is an arbitrary set, then the set $\mathcal{F}(S, \mathbb{K})$ of all $\mathbb{K}$-valued functions on S forms a vector space over $\mathbb{K}$. For example, the set of complex-valued functions on $\mathbb{R}^n$, denoted $\mathcal{F}(\mathbb{R}^n, \mathbb{C})$, is a complex vector space. We usually denote $\mathcal{F}(S, \mathbb{R})$ simply by $\mathcal{F}(S)$, taking the real numbers as the default field. If S is a finite set $S = \{1, 2, \dots, n\}$, then $\mathcal{F}(S, \mathbb{K})$ is equivalent to the vector space $\mathbb{K}^n$, setting $u_i = u(i)$ for any $u \in \mathcal{F}(S, \mathbb{K})$.

When the vectors can be uniquely specified by a finite number of scalars from the field $\mathbb{K}$, as in Examples 3.5 and 3.7, the vector space is said to be *finite dimensional*. The number of independent components needed to specify an arbitrary vector is called the *dimension* of the space; e.g., $\mathbb{K}^n$ has dimension n, while $M^{m,n}(\mathbb{K})$ is of dimension mn. On the other hand, in Examples 3.6 and 3.8 it is clearly impossible to specify the vectors by a finite number of scalars and these vector spaces are said to be *infinite dimensional*. A rigorous definition of these terms will be given in Section 3.5.

Example 3.9 A set M is called a **module** over a *ring* R if it satisfies all the axioms (VS1)–(VS5) with R replacing the field $\mathbb{K}$. Axiom (VS5) is only included if the ring has an identity. This concept is particularly useful when R is a ring of real or complex-valued functions on a set S such as the rings $\mathcal{F}(S)$ or $\mathcal{F}(S, \mathbb{C})$ in Example 3.8.

A typical example of a module is the following. Let $\mathcal{C}(\mathbb{R}^n)$ be the ring of *continuous* real-valued functions on $\mathbb{R}^n$, sometimes called *scalar fields*, and let $\mathcal{V}^n$ be the set of all n-tuples of real-valued continuous functions on $\mathbb{R}^n$. A typical element of $\mathcal{V}^n$, called a *vector field* on $\mathbb{R}^n$, can be written

$$\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}),\ v_2(\mathbf{x}), \dots,\ v_n(\mathbf{x}))$$

where each $v_i(\mathbf{x})$ is a continuous real-valued function on $\mathbb{R}^n$. Vector fields can be added in the usual way and multiplied by scalar fields,

$$\mathbf{u} + \mathbf{v} = (u_1(\mathbf{x}) + v_1(\mathbf{x}), \dots, u_n(\mathbf{x}) + v_n(\mathbf{x})),$$
$$f(\mathbf{x})\mathbf{v}(\mathbf{x}) = (f(\mathbf{x})v_1(\mathbf{x}),\ f(\mathbf{x})v_2(\mathbf{x}), \dots,\ f(\mathbf{x})v_n(\mathbf{x})).$$

The axioms (VS1)–(VS5) are easily verified, showing that $\mathcal{V}^n$ is a module over $\mathcal{C}(\mathbb{R}^n)$. This *module* is finite dimensional in the sense that only a finite number of component scalar fields are needed to specify any vector field. Of course $\mathcal{V}^n$ also has the structure of a vector space over the field $\mathbb{R}$, similar to the vector space $\mathcal{F}(\mathbb{R}^n)$ in Example 3.8, but as a vector space it is *infinite dimensional*.

3.3 Vector space homomorphisms

If V and W are two vector spaces over the same field $\mathbb{K}$, a map $T : V \to W$ is called **linear**, or a **vector space homomorphism** from V into W, if

$$T(au + bv) = aTu + bTv \tag{3.1}$$

for all $a, b \in \mathbb{K}$ and all $u, v \in V$. The notation Tu on the right-hand side is commonly used in place of $T(u)$. Vector space homomorphisms play a similar role to group homomorphisms in that they preserve the basic operations of vector addition and scalar multiplication that define a vector space. They are the morphisms of the *category of vector spaces*.

Since $T(u + 0) = Tu = Tu + T0$, it follows that the zero vector of V goes to the zero vector of W under a linear map, $T0 = 0$. Note, however, that the zero vectors on the two sides of this equation lie in different spaces and are, strictly speaking, different vectors.

A linear map $T : V \to W$ that is one-to-one and onto is called a **vector space isomorphism**. In this case the inverse map $T^{-1} : W \to V$ must also be linear, for if $u, v \in W$ let

$u' = T^{-1}u,\ v' = T^{-1}v$ then

$$\begin{aligned} T^{-1}(au + bv) &= T^{-1}(aTu' + bTv') \\ &= T^{-1}(T(au' + bv')) \\ &= \mathrm{id}_V(au' + bv') \\ &= au' + bv' \\ &= aT^{-1}u + bT^{-1}v. \end{aligned}$$

Two vector spaces V and W are called **isomorphic**, written $V \cong W$, if there exists a vector space isomorphism $T : V \to W$. Two isomorphic vector spaces are essentially identical in all their properties.

Example 3.10 Consider the set $\mathcal{P}_n(x)$ of all real-valued polynomials of degree $\le n$,

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

Polynomials of degree $\le n$ can be added and multiplied by scalars in the obvious way,

$$\begin{aligned} f(x) + g(x) &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n, \\ cf(x) &= ca_0 + ca_1x + ca_2x^2 + \cdots + ca_nx^n, \end{aligned}$$

making $\mathcal{P}_n(x)$ into a vector space. The map $S : P_n(x) \to \mathbb{R}^{n+1}$ defined by

$$S(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = (a_0,\ a_1,\ a_2, \ldots,\ a_n)$$

is one-to-one and onto and clearly preserves basic vector space operations,

$$S(f(x) + g(x)) = S(f(x)) + S(g(x)), \qquad S(af(x)) = aS(f(x)).$$

Hence S is a vector space isomorphism, and $P_n(x) \cong \mathbb{R}^{n+1}$.

The set $\hat{\mathbb{R}}^\infty$ of all sequences of real numbers $(a_0, a_1, \ldots)$ having only a finite number of non-zero members $a_i \neq 0$ is a vector space, using the same rules of vector addition and scalar multiplication given for $\mathbb{R}^\infty$ in Example 3.6. The elements of $\hat{\mathbb{R}}^\infty$ are real sequences of the form $(a_0, a_1, \ldots, a_m, 0, 0, 0, \ldots)$. Let $\mathcal{P}(x)$ be the set of all real polynomials, $\mathcal{P}(x) = \mathcal{P}_0(x) \cup \mathcal{P}_1(x) \cup \mathcal{P}_2(x) \cup \ldots$ This is clearly a vector space with respect to the standard rules of addition of polynomials and scalar multiplication. The map $S : \hat{\mathbb{R}}^\infty \to \mathcal{P}$ defined by

$$S : (a_0,\ a_1, \ldots,\ a_m,\ 0,\ 0, \ldots) \mapsto a_0 + a_1x + \cdots + a_mx^m$$

is an isomorphism.

It is simple to verify that the inclusion maps defined in Section 1.4,

$$i_1 : \hat{\mathbb{R}}^\infty \to \mathbb{R}^\infty \quad \text{and} \quad i_2 : \mathcal{P}_n(x) \to \mathcal{P}(x)$$

are vector space homomorphisms.

Let $L(V, W)$ denote the set of all linear maps from V to W. If T, S are linear maps from V to W, addition $T + S$ and scalar multiplication aT are defined by

$$(T + S)(u) = Tu + Su, \qquad (aT)u = a\,Tu.$$

The set $L(V, W)$ is a vector space with respect to these operations. Other common notations for this space are $\mathrm{Hom}(V, W)$ and $\mathrm{Lin}(V, W)$.

Exercise: Verify that $L(V, W)$ satisfies all the axioms of a vector space.

If $T \in L(U, V)$ and $S \in L(V, W)$, define their **product** to be the composition map $ST = S \circ T : U \to W$,

$$(ST)u = S(Tu).$$

This map is clearly linear since

$$ST(au + bv) = S(aTu + bTv) = aSTu + bSTv.$$

If S and T are invertible linear maps then so is their product ST, and $(ST)^{-1} : W \to U$ satisfies

$$(ST)^{-1} = T^{-1}S^{-1}, \tag{3.2}$$

since

$$T^{-1}S^{-1}ST = T^{-1}\mathrm{id}_V T = T^{-1}T = \mathrm{id}_U.$$

Linear maps $S : V \to V$ are called **linear operators** on V. They form the vector space $L(V, V)$. If S is an invertible linear operator on V it is called a **linear transformation** on V. It may be thought of as a vector space isomorphism of V onto itself, or an *automorphism* of V. The linear transformations of V form a group with respect to the product law of composition, called the **general linear group on** V and denoted $GL(V)$. The group properties are easily proved:

Closure: if S and T are linear transformations of V then so is ST, since (a) it is a linear map, and (b) it is invertible by Eq. (3.2).
Associativity: this is true of all maps (see Section 1.4).
Unit: the identity map id_V is linear and invertible.
Inverse: as shown above, the inverse T^{-1} of any vector space isomorphism T is linear.

Note, however, that $GL(V)$ is *not* a vector space, since the zero operator that sends every vector in V to the zero vector 0 is not invertible and therefore does not belong to $GL(V)$.

Problems

Problem 3.3 Show that the infinite dimensional vector space $\mathbb{R}^\infty$ is isomorphic with a proper subspace of itself.

Problem 3.4 On the vector space $\mathcal{P}(x)$ of polynomials with real coefficients over a variable x, let x be the operation of multiplying by the polynomial x, and let D be the operation of differentiation,

$$x : f(x) \mapsto xf(x), \qquad D : f(x) \mapsto \frac{\mathrm{d}f(x)}{\mathrm{d}x}.$$

Show that both of these are linear operators over $\mathcal{P}(x)$ and that $Dx - xD = I$, where I is the identity operator.

3.4 Vector subspaces and quotient spaces

A **(vector) subspace** W of a vector space V is a subset that is a vector space in its own right, with respect to the operations of vector addition and scalar multiplication defined on V. There is a simple criterion for determining whether a subset is a subspace:

> A subset W is a subspace of V if and only if $u + av \in W$ for all $a \in \mathbb{K}$ and all $u, v \in W$.

For, setting $a = 1$ shows that W is closed under vector addition, while $u = 0$ implies that it is closed with respect to scalar multiplication. Closure with respect to these two operations is sufficient to demonstrate that W is a vector subspace: the zero vector $0 \in W$ since $0 = 0v \in W$ for any $v \in W$; the inverse vector $-u = (-1)u \in W$ for every $u \in W$, and the remaining vector space axioms (VS1)–(VS5) are all satisfied by W since they are inherited from V.

Example 3.11 Let $U = \{(u_1, u_2, \dots, u_m, 0, \dots, 0)\} \subseteq \mathbb{R}^n$ be the subset of n-vectors whose last $n - m$ components all vanish. U is a vector subspace of $\mathbb{R}^n$, since

$$\begin{aligned}(u_1, \dots, u_m, 0, \dots, 0) &+ a(v_1, \dots, v_m, 0, \dots, 0)\\ &= (u_1 + av_1, \dots, u_m + av_m, 0, \dots, 0) \in U.\end{aligned}$$

This subspace is isomorphic to $\mathbb{R}^m$, through the isomorphism

$$T : (u_1, \dots, u_m, 0, \dots, 0) \mapsto (u_1, \dots, u_m).$$

Exercise: Show that $\mathbb{R}^n$ is isomorphic to a subspace of $\mathbb{R}^\infty$ for every $n > 0$.

Example 3.12 Let V be a vector space over a field $\mathbb{K}$, and $u \in V$ any vector. The set $U = \{cu \mid c \in \mathbb{K}\}$ is a subspace of V, since for any pair of scalars $c, c' \in \mathbb{K}$,

$$cu + a(c'u) = cu + (ac')u = (c + ac')u \in U.$$

Exercise: Show that the set $\{(x, y, z) \mid x + 2y + 3z = 0\}$ forms a subspace of $\mathbb{R}^3$, while the subset $\{(x, y, z) \mid x + 2y + 3z = 1\}$ does not.

Example 3.13 The set of all continuous real-valued functions on $\mathbb{R}^n$, denoted $\mathcal{C}(\mathbb{R}^n)$, is a subspace of $\mathcal{F}(\mathbb{R}^n)$ defined in Example 3.8, for if f and g are any pair of continuous functions on $\mathbb{R}^n$ then so is any linear combination $f + ag$ where $a \in \mathbb{R}$.

Exercise: Show that the vector space of all real polynomials $\mathcal{P}(x)$, defined in Example 3.10, is a vector subspace of $\mathcal{C}(\mathbb{R})$.

Given two subspaces U and W of a vector space V, their set-theoretical intersection $U \cap W$ forms a vector subspace of V, for if $u, w \in U \cap W$ then any linear combination $u + aw$ belongs to each subspace U and W separately. This argument can easily be extended to show that the intersection $\bigcap_{i \in I} U_i$ of any family of subspaces is a subspace of V.

Complementary subspaces and quotient spaces

While the intersection of any pair of subspaces U and W is a vector subspace of V, this is not true of their set-theoretical union $U \cup V$ – consider, for example, the union of the two subspaces $\{(c, 0) \mid c \in \mathbb{R}\}$ and $\{(0, c) \mid c \in \mathbb{R}\}$ of $\mathbb{R}^2$. Instead, we can define the **sum** $U + W$ of any pair of subspaces to be the 'smallest' vector space that contains $U \cup W$,

$$U + W = \{u + w \mid u \in U,\ w \in W\}.$$

This is a vector subspace, for if $u = u_1 + w_1$ and $v = u_2 + w_2$ belong to $U + W$, then

$$u + av = (u_1 + w_1) + a(u_2 + w_2) = (u_1 + au_2) + (w_1 + aw_2) \in U + W.$$

Two subspaces U and W of V are said to be **complementary** if every vector $v \in V$ has a *unique* decomposition $v = u + w$ where $u \in U$ and $w \in W$. V is then said to be the **direct sum** of the subspaces U and W, written $V = U \oplus W$.

Theorem 3.1 *U and W are complementary subspaces of V if and only if (i) $V = U + W$ and (ii) $U \cap W = \{0\}$.*

Proof: If U and W are complementary subspaces then (i) is obvious, and if there exists a non-zero vector $u \in U \cap V$ then the zero vector would have alternative decompositions $0 = 0 + 0$ and $0 = u + (-u)$. Conversely, if (i) and (ii) hold then the decomposition $v = u + w$ is unique, for if $v = u' + w'$ then $u - u' = w - w' \in U \cap W$. Hence $u - u' = w - w' = 0$, so $u = u'$ and $w = w'$. ■

Example 3.14 Let $\mathbb{R}_1$ be the subspace of $\mathbb{R}^n$ consisting of vectors of the form $\{(x_1, 0, 0, \dots, 0) \mid x_1 \in \mathbb{R}\}$, and S_1 the subspace

$$S_1 = \{(0, x_2, x_3, \dots, x_n) \mid (x_i \in \mathbb{R})\}.$$

Then $\mathbb{R}^n = \mathbb{R}_1 \oplus S_1$. Continuing in a similar way S_1 may be written as a direct sum of $\mathbb{R}_2 = \{(0, x_2, 0, \dots, 0)\}$ and a subspace S_2. We eventually arrive at the direct sum decomposition

$$\mathbb{R}^n = \mathbb{R}_1 \oplus \mathbb{R}_2 \oplus \cdots \oplus \mathbb{R}_n \cong \mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}.$$

If U and W are arbitrary vector spaces it is possible to define their direct sum in a constructive way, sometimes called their **external direct sum**, by setting

$$U \oplus W = U \times W = \{(u, w) \mid u \in U, w \in W\}$$

with vector addition and scalar multiplication defined by

$$(u, w) + (u', w') = (u + u', w + w'), \qquad a(u, w) = (au, aw).$$

The map $\varphi : U \to \hat{U} = \{(u, 0) \mid u \in U\} \subset U \oplus W$ defined by $\varphi(u) = (u, 0)$ is clearly an isomorphism. Hence we may identify U with the subspace $\hat{U}$, and similarly W is identifiable with $\hat{W} = \{(0, w) \mid w \in W\}$. With these identifications the constructive notion of direct sum is equivalent to the 'internally defined' version, since $U \oplus W = \hat{U} \oplus \hat{W}$.

The real number system $\mathbb{R}$ can be regarded as a real vector space in which scalar multiplication is simply multiplication of real numbers – vectors and scalars are indistinguishable in this instance. Since the subspaces $\mathbb{R}_i$ defined in Example 3.14 are clearly isomorphic to $\mathbb{R}$ for each $i = 1, \dots, n$, the decomposition given in that example can be written

$$\mathbb{R}^n \cong \underbrace{\mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}}_{n}.$$

For any given vector subspace there always exist complementary subspaces. We give the proof here as an illustration of the use of Zorn's lemma, Theorem 1.6, but it is somewhat technical and the reader will lose little continuity by moving on if they feel so inclined.

Theorem 3.2 *Given a subspace $W \subseteq V$ there always exists a complementary subspace $U \subseteq V$ such that $V = U \oplus W$.*

Proof: Given a vector subspace W of V, let $\mathcal{U}$ be the collection of all vector subspaces $U \subseteq V$ such that $U \cap W = \{0\}$. The set $\mathcal{U}$ can be partially ordered by set inclusion as in Example 1.5. Furthermore, if $\{U_i \mid i \in I\}$ is any totally ordered subset of $\mathcal{U}$ such that for every pair $i, j \in I$ we have either $U_i \subseteq U_j$ or $U_j \subseteq U_i$, then their union is bounded above by

$$\tilde{U} = \bigcup_{i \in I} U_i.$$

The set $\tilde{U}$ is a vector subspace of V, for if $u \in \tilde{U}$ and $v \in \tilde{U}$ then there exists a member U_i of the totally ordered family such that both vectors must belong to the same member U_i – if $u \in U_j$ and $v \in U_k$ then set $i = k$ if $U_j \subseteq U_k$, else set $i = j$. Hence $u + av \in U_i \subseteq \tilde{U}$ for all $a \in \mathbb{K}$. By Zorn's lemma we conclude that there exists a maximal subspace $U \in \mathcal{U}$.

It remains to show that U is complementary to W. Suppose not; then there exists a vector $v' \in V$ that cannot be expressed in the form $v' = u + w$ where $u \in U$, $w \in W$. Let U' be the vector subspace defined by

$$U' = \{av' + u \mid u \in U\} = \{av'\} \oplus U.$$

It belongs to the family $\mathcal{U}$, for if $U' \cap W \neq \{0\}$ then there would exist a non-zero vector $w' = av' + u$ belonging to W. This implies $v' = a^{-1}(w' - u)$, in contradiction to the requirement that v' cannot be expressed as a sum of vectors from U and W. Hence we have strict inclusion $U \subset U'$, contradicting the maximality of U. Thus U is a subspace complementary to W, as required. ■

This proof has a distinctly non-constructive feel to it, which is typical of proofs invoking Zorn's lemma. A more direct way to arrive at a vector space complementary to a given subspace W is to define an equivalence relation $\equiv_W$ on V by

$$u \equiv_W v \quad \text{iff} \quad u - v \in W.$$

Checking the equivalence properties is easy:

Reflexive: $u - u = 0 \in W$ for all $u \in V$,
Symmetric: $u - v \in W \Longrightarrow v - u = -(u - v) \in W$,
Transitive: $u - v \in W$ and $v - w \in W \Longrightarrow u - w = (u - v) + (v - w) \in W$.

The equivalence class to which u belongs is written $u + W$, where

$$u + W = \{u + w \mid w \in W\},$$

and is called a **coset** of W. This definition is essentially identical to that given in Section 2.5 for the case of an abelian group. It is possible to form the sum of cosets and multiply them by scalars, by setting

$$(u + W) + (v + W) = (u + v) + W, \qquad a(u + W) = (au) + W.$$

For consistency, it is necessary to show that these definitions are independent of the choice of coset representative. For example, if $u \equiv_W u'$ and $v \equiv_W v'$ then $(u' + v') \equiv_W (u + v)$, for

$$(u' + v') - (u + v) = (u' - u) + (v' - v) \in W.$$

Hence

$$(u' + W) + (v' + W) = (u' + v') + W = (u + v) + W = (u + W) + (v + W).$$

Similarly $au' \equiv_W au$ since $au' - au = a(u' - u) \in W$ and

$$a(u' + W) = (au') + W = (au) + W = a(u + W).$$

The task of showing that the set of cosets is a vector space with respect to these operations is tedious but undemanding. For example, the distributive law (VS2) follows from

$$\begin{aligned} a((u + W) + (v + W)) &= a((u + v) + W) \\ &= a(u + v) + W \\ &= (au + av) + W \\ &= ((au) + W) + ((av) + W) \\ &= a(u + W) + a(v + W). \end{aligned}$$

The rest of the axioms follow in like manner, and are left as exercises. The vector space of cosets of W is called the **quotient space** of V by W, denoted V/W.

To picture a quotient space let U be any subspace of V that is complementary to W. Every element of V/W can be written *uniquely* as a coset $u + W$ where $u \in U$. For, if $v + W$ is any coset, let $v = u + w$ be the unique decomposition of v into vectors from U and W respectively, and it follows that $v + W = u + W$ since $v \equiv_W u$. The map $T : U \to V/W$ defined by $T(u) = u + W$ describes an isomorphism between U and V/W. For, if $u + W = u' + W$ where $u, u' \in U$ then $u - u' \in W \cap U$, whence $u = u'$ since U and W are complementary subspaces.

Exercise: Complete the details to show that the map $T : U \to V/W$ is linear, one-to-one and onto, so that $U \cong V/W$.

This argument also shows that all complementary spaces to a given subspace W are isomorphic to each other. The quotient space V/W is a method for constructing the 'canonical complement' to W.

Example 3.15 While V/W is in a sense complementary to W it is not a subspace of V and, indeed, there is no *natural* way of identifying it with any subspace complementary to W. For example, let $W = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$ be the subspace $z = 0$ of $\mathbb{R}^3$. Its cosets are planes $z = a$, parallel to the x–y plane, and it is these planes that constitute the 'vectors' of V/W. The subspace $U = \{(0, 0, z) \mid z \in \mathbb{R}\}$ is clearly complementary to W and is isomorphic to V/W using the map

$$(0, 0, a) \mapsto (0, 0, a) + W = \{(x, y, a) \mid x, y \in \mathbb{R}\}.$$

However, there is no natural way of identifying V/W with a complementary subspace such as U. For example, the space $U' = \{(0, 2z, z) \mid z \in \mathbb{R}\}$ is also complementary to W since $U' \cap W = \{0\}$ and every vector $(a, b, c) \in \mathbb{R}^3$ has the decomposition

$$(a, b, c) = (a, b - 2c, 0) + (0, 2c, c), \qquad (a, b - 2c, 0) \in W, \quad (0, 2c, c) \in U'.$$

Again, $U' \cong V/W$, under the map

$$(0, 2c, c) \mapsto (0, 2c, c) + W = (0, 0, c) + W.$$

Note how the 'W-component' of (a, b, c) depends on the choice of complementary subspace; $(a, b, 0)$ with respect to U, and $(a, b - 2c, 0)$ with respect to U'.

Images and kernels of linear maps

The **image** of a linear map $T : V \to W$ is defined to be the set

$$\operatorname{im} T = T(V) = \{w \mid w = Tv\} \subseteq W.$$

The set $\operatorname{im} T$ is a subspace of W, for if $w, w' \in \operatorname{im} T$ then

$$w + aw' = Tv + aTv' = T(v + av') \in \operatorname{im} T.$$

The **kernel** of the map T is defined as the set

$$\ker T = T^{-1}(0) = \{v \in V \mid Tv = 0\} \subseteq V.$$

This is also a subspace of V, for if $v, v' \in \ker T$ then $T(v + av') = Tv + aTv' = 0 + 0 = 0$. The two spaces are related by the identity

$$\operatorname{im} T \cong V/\ker T. \tag{3.3}$$

Proof: Define the map $\tilde{T} : V/\ker T \to \operatorname{im} T$ by

$$\tilde{T}(v + \ker T) = Tv.$$

This map is well-defined since it is independent of the choice of coset representative v,

$$\begin{aligned} v + \ker T = v' + \ker T &\implies v - v' \in \ker T \\ &\implies T(v - v') = 0 \\ &\implies Tv = Tv' \end{aligned}$$

and is clearly linear. It is onto and one-to-one, for every element of $\operatorname{im} T$ is of the form $Tv = \tilde{T}(v + \ker T)$ and

$$\begin{aligned}\tilde{T}(v + \ker T) = \tilde{T}(v' + \ker T) &\Longrightarrow Tv = Tv' \\ &\Longrightarrow v - v' \in \ker T \\ &\Longrightarrow v + \ker T = v' + \ker T.\end{aligned}$$

Hence $\tilde{T}$ is a vector space isomorphism, which proves Eq. (3.3). ■

Example 3.16 Let $V = \mathbb{R}^3$ and $W = \mathbb{R}^2$, and define the map $T : V \to W$ by

$$T(x, y, z) = (x + y + z,\ 2x + 2y + 2z).$$

The subspace $\operatorname{im} T$ of $\mathbb{R}^2$ consists of the set of all vectors of the form $(a, 2a)$, where $a \in \mathbb{R}$, while $\ker T$ is the subset of all vectors $(x, y, z) \in V$ such that $x + y + z = 0$ – check that these do form a subspace of V. If $v = (x, y, z)$ and $a = x + y + z$, then $v - ae \in \ker T$ where $e = (1, 0, 0) \in V$, since

$$T(v - ae) = T(x, y, z) - T(x + y + z, 0, 0) = (0, 0).$$

Furthermore a is the unique value having this property, for if $a' \neq a$ then $v - a'e \notin \ker T$. Hence every coset of $\ker T$ has a unique representative of the form ae and may be written uniquely in the form $ae + \ker T$. The isomorphism $\tilde{T}$ defined in the above proof is given by

$$\tilde{T}(ae + \ker T) = (a, 2a) = a(1, 2).$$

Problems

Problem 3.5 If L, M and N are vector subspaces of V show that

$$L \cap (M + (L \cap N)) = L \cap M + L \cap N$$

but it is not true in general that

$$L \cap (M + N) = L \cap M + L \cap N.$$

Problem 3.6 Let $V = U \oplus W$, and let $v = u + w$ be the unique decomposition of a vector v into a sum of vectors from $u \in U$ and $w \in W$. Define the *projection operators* $P_U : V \to U$ and $P_W : V \to W$ by

$$P_U(v) = u, \qquad P_W(v) = w.$$

Show that

(a) $P_U^2 = P_U$ and $P_W^2 = P_W$.
(b) Show that if $P : V \to V$ is an operator satisfying $P^2 = P$, said to be an *idempotent* operator, then there exists a subspace U such that $P = P_U$. [*Hint*: Set $U = \{u \mid Pu = u\}$ and $W = \{w \mid Pw = 0\}$ and show that these are complementary subspaces such that $P = P_U$ and $P_W = \mathrm{id}_V - P$.]

3.5 Bases of a vector space

Subspace spanned by a set

If A is any subset of a vector space V define the **subspace spanned or generated by** A, denoted $L(A)$, as the set of all *finite* linear combinations of elements of A,

$$L(A) = \left\{ \sum_{i=1}^{n} a^i v_i | a^i \in \mathbb{K},\ v_i \in A,\ n = 1, 2, \dots \right\}.$$

The word 'finite' is emphasized here because no meaning can be attached to infinite sums until we have available the concept of 'limit' (see Chapters 10 and 13). We may think of $L(A)$ as the intersection of all subspaces of V that contain A – essentially, it is the 'smallest' vector subspace containing A. At first sight the notation whereby the indices on the coefficients a^i of the linear combinations have been set in the superscript position may seem a little peculiar, but we will eventually see that judicious and systematic placements of indices can make many expressions much easier to manipulate.

Exercise: If M and N are subspaces of V show that their sum $M + N$ is identical with the span of their union, $M + N = L(M \cup N)$.

The vector space V is said to be **finite dimensional** [7] if it can be spanned by a finite set, $V = L(A)$, where $A = \{v_1, v_2, \dots, v_n\}$. Otherwise we say V is **infinite dimensional**. When V is finite dimensional its **dimension**, dim V, is defined to be the smallest number n such that V is spanned by a set consisting of just n vectors.

Example 3.17 $\mathbb{R}^n$ is finite dimensional, since it can be generated by the set of 'unit vectors',

$$A = \{e_1 = (1, 0, \dots, 0),\quad e_2 = (0, 1, \dots, 0),\ \dots,\ e_n = (0, 0, \dots, 1)\}.$$

Since any vector u can be written

$$u = (u_1,\ u_2,\ \dots,\ u_n) = u_1 e_1 + u_2 e_2 + \dots + u_n e_n,$$

these vectors span $\mathbb{R}^n$, and $\dim \mathbb{R}^n \leq n$. We will see directly that $\dim \mathbb{R}^n = n$, as to be expected.

Example 3.18 $\mathbb{R}^\infty$ is clearly infinite dimensional. It is not even possible to span this space with the set of vectors $A = \{e_1, e_2, \dots\}$, where

$$e_1 = (1, 0, \dots),\quad e_2 = (0, 1, \dots),\quad \dots.$$

The reason is that any finite linear combination of these vectors will only give rise to vectors having at most a finite number of non-zero components. The set of all those vectors that are finite linear combinations of vectors from A does in fact form an infinite dimensional subspace of $\mathbb{R}^\infty$, but it is certainly not the whole space. The space spanned by A is precisely the subspace $\hat{\mathbb{R}}^\infty$ defined in Example 3.10.

Exercise: If V is a vector space and $u \in V$ is any non-zero vector show that $\dim V = 1$ if and only if every vector $v \in V$ is proportional to u; i.e., $v = au$ for some $a \in \mathbb{K}$.

Exercise: Show that the set of functions on the real line, $\mathcal{F}(\mathbb{R})$, is an infinite dimensional vector space.

Basis of a vector space

A set of vectors A is said to be **linearly independent**, often written 'l.i.', if every finite subset of vectors $\{v_1, v_2, \dots, v_k\} \subseteq A$ has the property that

$$\sum_{i=1}^{k} a^i v_i = 0 \implies a^j = 0 \text{ for all } j = 1, \dots, k.$$

In other words, the zero vector 0 cannot be written as a non-trivial linear combination of these vectors. The zero vector can never be a member of a l.i. set since $a0 = 0$ for any $a \in \mathbb{K}$. If A is a finite set of vectors, $A = \{v_1, v_2, \dots, v_n\}$, it is sufficient to set $k = n$ in the above definition. A subset E of a vector space V is called a **basis** if it is linearly independent and spans the whole of V. A set of vectors is said to be **linearly dependent** if it is not l.i.

Example 3.19 The set of vectors $\{e_1 = (1, 0, \dots, 0), \dots, e_n = (0, 0, \dots 0, 1)\}$ span $\mathbb{K}^n$, since every vector $v = (v_1, v_2, \dots, v_n)$ can be expressed as a linear combination

$$v = v_1 e_1 + v_2 e_2 + \cdots + v_n e_n.$$

They are linearly independent, for if $v = 0$ then we must have $v_1 = v_2 = \cdots v_n = 0$. Hence $e_1, \dots, e_n$ is a basis of $\mathbb{K}^n$.

Exercise: Show that the vectors $f_1 = (1, 0, 0)$, $f_2 = (1, 1, -1)$ and $f_3 = (1, 1, 1)$ are l.i. and form a basis of $\mathbb{R}^3$.

It is perhaps surprising to learn that even infinite dimensional vector space such as $\mathbb{R}^\infty$ always has a basis. Just try and construct a basis! The set $A = \{e_1 = (1, 0, 0, \dots),$ $e_2 = (0, 1, 0, \dots), \dots\}$ clearly won't do, since any vector having an infinite number of non-zero components cannot be a *finite* linear combination of these vectors. We omit the proof as it is heavily dependent on Zorn's lemma and such bases are only of limited use. For the rest of this section we only consider bases in finite dimensional spaces.

Theorem 3.3 *Let V be a finite dimensional vector space of dimension n. A subset $E = \{e_1, e_2, \dots, e_n\}$ spans V if and only if it is linearly independent.*

Proof: *Only if*: Assume $V = L(E)$, so that every vector $v \in V$ is a linear combination

$$v = \sum_{i=1}^{n} v^i e_i.$$

The set E is then linearly independent, for suppose there exists a vanishing linear combination

$$\sum_{i=1}^{n} a^i e_i = 0 \quad (a^i \in \mathbb{K})$$

where, say, $a^1 \neq 0$. Replacing e_1 by $e_1 = b^2 e_2 + \cdots + b^n e_n$ where $b^i = -a^i/a^1$, we find

$$v = \sum_{j=2}^{n} \bar{v}^j e_j \quad \text{where} \quad \bar{v}^j = v^j + b^j v^1.$$

Thus $E' = E - \{e_1\}$ spans V, contradicting the initial hypothesis that V cannot be spanned by a set of fewer than n vectors on account of $\dim V = n$.

If: Assume E is linearly independent. Our aim is to show that it spans all of V. Since $\dim V = n$ there must exist a set $F = \{f_1, f_2, \ldots, f_n\}$ of exactly n vectors spanning V. By the above argument this set is l.i. Expand e_1 in terms of the vectors from F,

$$e_1 = a^1 f_1 + a^2 f_2 + \cdots + a^n f_n, \tag{3.4}$$

where, by a permutation of the vectors of the basis F, we may assume that $a^1 \neq 0$. The set $F' = \{e_1, f_2, \ldots, f_n\}$ is a basis for V:

(a) F' is linearly independent, for if there were a vanishing linear combination

$$ce_1 + c^2 f_2 + \cdots + c^n f_n = 0,$$

then substituting (3.4) gives

$$ca^1 f_1 + (ca^2 + c^2) f_2 + \cdots + (ca^n + c^n) f_n = 0.$$

By linear independence of $\{f_1, \ldots, f_n\}$ and $a^1 \neq 0$ it follows that $c = 0$, and subsequently that $c^2 = \cdots = c^n = 0$.

(b) The set F' spans the vector space V since by Eq. (3.4)

$$f_1 = \frac{1}{a^1}(e_1 - a^2 f_2 - \cdots - a^n f_n),$$

and every $v \in V$ must be a linear combination of $\{e_1, f_2, \ldots, f_n\}$ since it is spanned by $\{f_1, \ldots, f_n\}$.

Continuing, e_2 must be a unique linear combination

$$e_2 = b^1 e_1 + b^2 f_2, + \cdots + b^n f_n.$$

Since, by hypothesis, e_1 and e_2 are linearly independent, at least one of the coefficients $b^2, \ldots, b^n$ must be non-zero, say $b^2 \neq 0$. Repeating the above argument we see that the set $F'' = \{e_1, e_2, f_3, \ldots, f_n\}$ is a basis for V. Continue the process n times to prove that $E = F^{(n)} = \{e_1, e_2, \ldots, e_n\}$ is a basis for V. ■

Corollary 3.4 *If $E = \{e_1, e_2, \ldots, e_n\}$ is a basis of the vector space V then* $\dim V = n$.

Proof: Suppose that $\dim V = m < n$. The set of vectors $E' = \{e_1, e_2, \ldots, e_m\}$ is l.i., since it is a subset of the l.i. set E. Hence, by Theorem 3.3 it spans V, since it consists of exactly $m = \dim V$ vectors. But this is impossible since, for example, the vector e_n cannot be a linear combination of the vectors in E'. Hence we must have $\dim V \geq n$. However, by the definition of dimension it is impossible to have $\dim V > n$; hence $\dim V = n$. ■

Exercise: Show that if $A = \{v_1, v_2, \ldots, v_m\}$ is an l.i. set of vectors then $m \leq n = \dim V$.

Theorem 3.5 *Let V be a finite dimensional vector space, $n = \dim V$. If $\{e_1, \dots, e_n\}$ is a basis of V then each vector $v \in V$ has a unique decomposition*

$$v = \sum_{i=1}^{n} v^i e_i, \quad v^i \in \mathbb{K}. \tag{3.5}$$

The n scalars $v^i \in \mathbb{K}$ are called the **components** *of the vector v with respect to this basis.*

Proof: Since the e_i span V, every vector v has a decomposition of the form (3.5). If there were a second such decomposition,

$$v = \sum_{i=1}^{n} v^i e_i = \sum_{i=1}^{n} w^i e_i$$

then

$$\sum_{i=1}^{n} (v^i - w^i) e_i = 0.$$

Since the e_i are linearly independent, each coefficient of this sum must vanish, $v^i - w^i = 0$. Hence $v^i = w^i$, and the decomposition is unique. ■

Theorem 3.6 *If V and W are finite dimensional then they are isomorphic if and only if they have the same dimension.*

Proof: Suppose V and W have the same dimension n. Let $\{e_i\}$ be a basis of V and $\{f_i\}$ a basis of W, where $i = 1, 2, \dots, n$. Set $T : V \to W$ to be the linear map defined by $Te_1 = f_1,\ Te_2 = f_2, \dots,\ Te_n = f_n$. This map extends uniquely to all vectors in V by linearity,

$$T\left(\sum_{i=1}^{n} v^i e_i\right) = \sum_{i=1}^{n} v^i \, Te_i = \sum_{i=1}^{n} v^i f_i$$

and is clearly one-to-one and onto. Thus T is an isomorphism between V and W.

Conversely suppose V and W are isomorphic vector spaces, and let $T : V \to W$ be a linear map having inverse $T^{-1} : W \to V$. If $\{e_1, e_2, \dots, e_n\}$ is a basis of V, we show that $\{f_i = Te_i\}$ is a basis of W:

(a) The vectors $\{f_i\}$ are linearly independent, for suppose there exist scalars $a^i \in \mathbb{K}$ such that

$$\sum_{i=1}^{n} a^i f_i = 0.$$

Then,

$$0 = T^{-1}\left(\sum_{i=1}^{n} a^i f_i\right) = \sum_{i=1}^{n} a^i T^{-1} f_i = \sum_{i=1}^{n} a^i e_i$$

and from the linear independence of $\{e_i\}$ it follows that $a^1 = a^2 = \dots = a^n = 0$.

(b) To show that the vectors $\{f_i\}$ span W let w be any vector in W and set $v = T^{-1}w \in V$.

Since $\{e_i\}$ spans V there exist scalars v^i such that

$$v = \sum_{i=1}^{n} v^i e_i .$$

Applying the map T to this equation results in

$$w = \sum_{i=1}^{n} v^i T e_i = \sum_{i=1}^{n} v^i f_i ,$$

which shows that the set $\{f_1, \dots, f_n\}$ spans W.

By Corollary 3.4 it follows that $\dim W = \dim V = n$ since both vector spaces have a basis consisting of n vectors. ∎

Example 3.20 By Corollary 3.4 the space $\mathbb{K}^n$ is n-dimensional since, as shown in Example 3.19, the set $\{e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)\}$ is a basis. Using Theorem 3.6 every n-dimensional vector space V over the field $\mathbb{K}$ is isomorphic to $\mathbb{K}^n$, which may be thought of as the archetypical n-dimensional vector space over $\mathbb{K}$. Every basis $\{f_1, f_2, \dots, f_n\}$ of V establishes an isomorphism $T : V \to \mathbb{K}^n$ defined by

$$Tv = (v^1, v^2, \dots, v^n) \in \mathbb{K}^n \quad \text{where} \quad v = \sum_{i=1}^{n} v^i f_i .$$

Example 3.20 may lead the reader to wonder why we bother at all with the abstract vector space machinery of Section 3.2, when all properties of a finite dimensional vector space V could be referred to the space $\mathbb{K}^n$ by simply picking a basis. This would, however, have some unfortunate consequences. Firstly, there are infinitely many bases of the vector space V, each of which gives rise to a different isomorphism between V and $\mathbb{K}^n$. There is nothing *natural* in the correspondence between the two spaces, since there is no general way of singling out a preferred basis for the vector space V. Furthermore, any vector space concept should ideally be given a basis-independent definition, else we are always faced with the task of showing that it is independent of the choice of basis. For these reasons we will persevere with the 'invariant' approach to vector space theory.

Matrix of a linear operator

Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. Given a basis $\{e_1, e_2, \dots, e_n\}$ of V define the **components** T^a_k of the linear operator T with respect to this basis by setting

$$T e_j = \sum_{i=1}^{n} T^i{}_j \, e_i . \tag{3.6}$$

By Theorem 3.5 the components $T^i{}_j$ are uniquely defined by these equations, and the square $n \times n$ matrix $\mathsf{T} = [T^i{}_j]$ is called the **matrix of** T with respect to the basis $\{e_i\}$. It is usual to take the superscript i as the 'first' index, labelling rows, while the subscript j labels the columns, and for this reason it is generally advisable to leave some horizontal spacing between these two indices. In Section 2.3 the components of a matrix were denoted by subscripted symbols such as $\mathsf{A} = [a_{ij}]$, but in general vector spaces it is a good idea to display the components of a matrix representing a linear operator T in this 'mixed script' notation.

If $v = \sum_{k=1}^{n} v^k e_k$ is an arbitrary vector of V then its image vector $\tilde{v} = Tv = \sum_{j=1}^{m} w^j e_j$ is given by

$$\tilde{v} = Tv = T\Big(\sum_{j=1}^{n} v^j e_j\Big) = \sum_{j=1}^{n} v^j Te_j = \sum_{j=1}^{n}\sum_{i=1}^{n} v^j T^i_{\ j} e_i ,$$

and the components of $\tilde{v}$ are given by

$$\tilde{v}^i = (Tv)^i = \sum_{j=1}^{n} T^i_{\ j} v^j . \tag{3.7}$$

If we write the components of v and $\tilde{v}$ as column vectors or $n \times 1$ matrices, $\mathbf{v}$ and $\tilde{\mathbf{v}}$,

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}, \qquad \tilde{\mathbf{v}} = \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \vdots \\ \tilde{v}^n \end{pmatrix},$$

then Eq. (3.7) is the componentwise representation of the matrix equation

$$\tilde{\mathbf{v}} = \mathsf{T}\mathbf{v}. \tag{3.8}$$

The matrix of the composition of two operators $ST \equiv S \circ T$ is given by

$$\begin{aligned} ST(e_i) &= \sum_{j=1}^{n} S(T^j_{\ i} e_j) \\ &= \sum_{j=1}^{n} T^j_{\ i} Se_j \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} T^j_{\ i} S^k_{\ j} e_k \\ &= \sum_{k=1}^{n} (ST)^k_{\ i} e_k \end{aligned}$$

where

$$(ST)^k_{\ i} = \sum_{j=1}^{n} S^k_{\ j} T^j_{\ i} . \tag{3.9}$$

This can be recognized as the componentwise formula for the matrix product ST.

Example 3.21 Care should be taken when reading off the components of the matrix T from (3.6) as it is very easy to come up mistakenly with the 'transpose' array. For example, if a transformation T of a three-dimensional vector space is defined by its effect on a basis

e_1, e_2, e_3,

$$\begin{aligned} Te_1 &= e_1 - e_2 + e_3 \\ Te_2 &= e_1 - e_3 \\ Te_3 &= e_2 + 2e_3, \end{aligned}$$

then its matrix with respect to this basis is

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

The result of applying T to a vector $u = xe_1 + ye_2 + ze_3$ is

$$Tu = xTe_1 + yTe_2 + zTe_3 = (x+y)e_1 + (-x+z)e_2 + (x-y+2z)e_3,$$

which can also be obtained by multiplying the matrix T and the column vector $\mathbf{u} = (x, y, z)^T$,

$$\mathsf{T}\mathbf{u} = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y \\ -x+z \\ x-y+2z \end{pmatrix}.$$

If S is a transformation given by

$$\begin{aligned} Se_1 &= e_1 + 2e_3 \\ Se_2 &= e_2 \\ Se_3 &= e_1 - e_2 \end{aligned}$$

whose matrix with respect to this basis is

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

the product of these two transformations is found from

$$\begin{aligned} STe_1 &= Se_1 - Se_2 + Se_3 &&= 2e_1 - 2e_2 + 2e_3 \\ STe_2 &= Se_1 - Se_3 &&= e_2 + 2e_3 \\ STe_3 &= Se_2 + 2Se_3 &&= 2e_1 - e_2. \end{aligned}$$

Thus the matrix of ST is the matrix product of S and T,

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

Exercise: In Example 3.21 compute TS by calculating $T(Se_i)$ and also by evaluating the matrix product of T and S.

Exercise: If V is a finite dimensional vector space, $\dim V = n$, over the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, show that the group $GL(V)$ of linear transformations of V is isomorphic to the matrix group of invertible $n \times n$ matrices, $GL(n, \mathbb{K})$.

Basis extension theorem

While specific bases should not be used in general definitions of vector space concepts if at all possible, there are specific instances when the singling out of a basis can prove of great benefit. The following theorem is often useful, in that it allows us to extend any l.i. set to a basis. In particular, it implies that if $v \in V$ is any non-zero vector, one can always find a basis such that $e_1 = v$.

Theorem 3.7 *Let $A = \{v_1, v_2, \dots, v_m\}$ be any l.i. subset of V, where $m \leq n = \dim V$. Then there exists a basis $E = \{e_1, e_2, \dots, e_n\}$ of V such that $e_1 = v_1, e_2 = v_2, \dots, e_m = v_m$.*

Proof: If $m = n$ then by Theorem 3.3 the set E is a basis of V and there is nothing to show. Assuming $m < n$, we set $e_1 = v_1, e_2 = v_2, \dots, e_m = v_m$. By Corollary 3.4 the set A cannot span V since it consists of fewer than n elements, and there must exist a vector $e_{m+1} \in V$ that is not a linear combination of $e_1, \dots, e_m$. The set $A' = \{e_1, e_2, \dots, e_{m+1}\}$ is l.i., for if

$$a^1 e_1 + \cdots + a^m e_m + a^{m+1} e_{m+1} = 0,$$

then we must have $a^{m+1} = 0$, else e_{m+1} would be a linear combination of $e_1, \dots, e_m$. The linear independence of $e_1, \dots, e_m$ then implies that $a^1 = a^2 = \cdots = a^m = 0$. If $m + 1 < n$ continue adding vectors that are linearly independent of those going before, until we arrive at a set $E = A^{(n-m)}$, which is l.i. and has n elements. This set must be a basis and the process can be continued no further. ∎

The following examples illustrate how useful this theorem can be in applications.

Example 3.22 Let W be a k-dimensional vector subspace of a vector space V of dimension n. We will demonstrate that the dimension of the factor space V/W, known as the **codimension** of W, is $n - k$. By Theorem 3.7 it is possible to find a basis $\{e_1, e_2, \dots, e_n\}$ of V such that the first k vectors $e_1, \dots, e_k$ are a basis of W. Then $\{e_{k+1} + W, e_{k+2} + W, \dots, e_n + W\}$ forms a basis for V/W since every coset $v + W$ can be written

$$v + W = v^{k+1}(e_{k+1} + W) + v^{k+2}(e_{k+2} + W) + \cdots + v^n(e_n + W)$$

where $v = \sum_{i=1}^n v^i e_i$ is the unique expansion given by Theorem 3.5. These cosets therefore span V/W. They are also l.i., for if

$$0 + W = a^{k+1}(e_{k+1} + W) + a^{k+2}(e_{k+2} + W) + \cdots + a^n(e_n + W)$$

then $a^{k+1} e_{k+1} + a^{k+2} e_{k+2} + \cdots + a^n e_n \in W$, which implies that there exist $b^1, \dots, b^k$ such that

$$a^{k+1} e_{k+1} + a^{k+2} e_{k+2} + \cdots + a^n e_n = b^1 e_1 + \cdots + b^k e_k.$$

By the linear independence of $e_1, \dots, e_n$ we have that $a^{k+1} = a^{k+2} = \dots = a^n = 0$. The desired result now follows,

$$\operatorname{codim} W \equiv \dim(V/W) = n - k = \dim V - \dim W.$$

Example 3.23 Let $A : V \to V$ be a linear operator on a finite dimensional vector space V. Define its **rank** $\rho(A)$ to be the dimension of its image im A, and its **nullity** $\nu(A)$ to be the dimension of its kernel ker A,

$$\rho(A) = \dim \operatorname{im} A, \qquad \nu(A) = \dim \ker A.$$

By Theorem 3.7 there exists a basis $\{e_1, e_2, \dots, e_n\}$ of V such that the first ν vectors $e_1, \dots, e_\nu$ form a basis of ker A such that $Ae_1 = Ae_2 = \dots = Ae_\nu = 0$. For any vector $u = \sum_{i=1}^n u^i e_i$

$$Au = \sum_{i=\nu+1}^{n} u^i Ae_i,$$

and $\operatorname{im} A = L(\{Ae_{\nu+1}, \dots, Ae_n\})$. Furthermore the vectors $Ae_{\nu+1}, \dots, Ae_n$ are l.i., for if there were a non-trivial linear combination

$$\sum_{i=\nu+1}^{n} b^i Ae_i = A\Big(\sum_{i=\nu+1}^{n} b^i e_i\Big) = 0$$

then $\sum_{i=\nu+1}^n b^i e_i \in \ker A$, which is only possible if all $b^i = 0$. Hence $\dim \operatorname{im} A = n - \dim \ker A$, so that

$$\rho(A) = n - \nu(A) \quad \text{where} \quad n = \dim V. \tag{3.10}$$

Problems

Problem 3.7 Show that the vectors $(1, x)$ and $(1, y)$ in $\mathbb{R}^2$ are linearly dependent iff $x = y$. In $\mathbb{R}^3$, show that the vectors $(1, x, x^2)$, $(1, y, y^2)$ and $(1, z, z^2)$ are linearly dependent iff $x = y$ or $y = z$ or $x = z$.

Generalize these statements to $(n + 1)$ dimensions.

Problem 3.8 Let V and W be any vector spaces, which are possibly infinite dimensional, and $T : V \to W$ a linear map. Show that if M is a l.i. subset of W, then $T^{-1}(M) = \{v \in V \mid Tv \in M\}$ is a linearly independent subset of V.

Problem 3.9 Let V and W be finite dimensional vector spaces of dimensions n and m respectively, and $T : V \to W$ a linear map. Given a basis $\{e_1, e_2, \dots, e_n\}$ of V and a basis $\{f_1, f_2, \dots, f_m\}$ of W, show that the equations

$$Te_k = \sum_{a=1}^{m} T^a_{\ k} f_a \quad (k = 1, 2, \dots, n)$$

serve to uniquely define the $m \times n$ matrix of components $\mathsf{T} = [T^a_{\ k}]$ of the linear map T with respect to these bases.

If $v = \sum_{k=1}^{n} v^k e_k$ is an arbitrary vector of V show that the components of its image vector $w = Tv$ are given by

$$w^a = (Tv)^a = \sum_{k=1}^{n} T^a_k v^k.$$

Write this as a matrix equation.

Problem 3.10 Let V be a four-dimensional vector space and $T : V \to V$ a linear operator whose effect on a basis $e_1, \dots, e_4$ is

$$\begin{aligned} Te_1 &= 2e_1 - e_4 \\ Te_2 &= -2e_1 + e_4 \\ Te_3 &= -2e_1 + e_4 \\ Te_4 &= e_1. \end{aligned}$$

Find a basis for $\ker T$ and $\operatorname{im} T$ and calculate the rank and nullity of T.

3.6 Summation convention and transformation of bases

Summation convention

In the above formulae summation over an index such as i or j invariably occurs on a pair of equal indices that are oppositely placed in the superscript and subscript position. Of course it is not inconceivable to have a summation between indices on the same level but, as we shall see, it is unlikely to happen in a natural way. In fact, this phenomenon occurs with such regularity that it is possible to drop the summation sign $\sum_{i=1}^{n}$ whenever the same index i appears in opposing positions without running into any serious misunderstandings, a convention first proposed by Albert Einstein (1879–1955) in the theory of general relativity, where multiple summation signs of this type arise repeatedly in the use of tensor calculus (see Chapter 18). The principal rule of Einstein's **summation convention** is:

If, in any expression, a superscript is equal to a subscript then it will be assumed that these indices are summed over from 1 to n where n is the dimension of the space.

Repeated indices are called **dummy** or **bound**, while those appearing singly are called **free**. Free indices are assumed to take all values over their range, and we omit statements such as $(i, j = 1, 2, \dots, n)$. For example, for any vector u and basis $\{e_i\}$ it is acceptable to write

$$u = u^i e_i \equiv \sum_{i=1}^{n} u^i e_i.$$

The index i is a dummy index, and can be replaced by any other letter having the same range,

$$u = u^i e_i = u^j e_j = u^k e_k.$$

For example, writing out Eqs. (3.6), (3.7) and (3.9) in this convention,

$$T\,e_j = T^i_{\ j}\,e_i, \tag{3.11}$$

$$\tilde{v}^i = (T\,v)^i = T^i_{\ j}\,v^j, \tag{3.12}$$

$$(ST)^k_{\ i} = S^k_{\ j} T^j_{\ i}. \tag{3.13}$$

In more complicated expressions such as

$$T^{ijk} S_{hij} \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} T^{ijk} S_{hij} \quad (h, k = 1, \dots, n)$$

i and j are dummy indices and h and k are free. It is possible to replace the dummy indices

$$T^{ilk} S_{hil} \quad \text{or} \quad T^{mik} S_{hmi}, \text{ etc.}$$

without in any way changing the meaning of the expression. In such replacements any letter of the alphabet other than one already used as a free index in that expression can be used, but you should always stay within a specified alphabet such as Roman, Greek, upper case Roman, etc., and sometimes even within a particular range of letters.

Indices should not be repeated on the same level, and in particular no index should ever appear more than twice in any expression. This would occur in $V^j T_{ij}$ if the dummy index j were replaced by the already occurring free index i to give the non-permissible $V^i T_{ii}$. Although expressions such as $V_j T_{ij}$ should not occur, there can be exceptions; for example, in cartesian tensors all indices occur in the subscript position and the summation convention is often modified to apply to expressions such as $V_j T_{ij} \equiv \sum_{j=1}^{n} V_j T_{ij}$.

In an equation relating indexed expressions, a given free index should appear in the same position, either as a superscript or subscript, on each expression of the equation. For example, the following are examples of equations that are *not* permissible unless there are mitigating explanations:

$$T^i = S_i, \qquad T^j + U_j F^{jk} = S^j, \qquad T^{kk}_{\ \ \ k} = S^k.$$

A free index in an equation can be changed to any symbol not already used as a dummy index in any part of the equation. However, the change must be made simultaneously in all expressions appearing in the equation. For example, the equation

$$Y_j = T^k_j X_k$$

can be replaced by

$$Y_i = T^j_i X_j$$

without changing its meaning, as both equations are a shorthand for the n equations

$$\begin{aligned} Y_1 &= T^1_1 X_1 + T^2_1 X_2 + \cdots + T^n_1 X_n \\ \dots &= \dots \\ Y_n &= T^1_n X_1 + T^2_n X_2 + \cdots + T^n_n X_n. \end{aligned}$$

Among the most useful identities in the summation convention are those concerning the **Kronecker delta** $\delta^i{}_j$ defined by

$$\delta^i{}_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{3.14}$$

These are the components of the unit matrix, $\mathsf{I} = [\delta^i{}_j]$. This is the matrix of the identity operator id_V with respect to any basis $\{e_1, \dots, e_n\}$. The Kronecker delta often acts as an 'index replacement operator'; for example,

$$T^{ij}_k \delta^m_i = T^{mj}_k,$$

$$T^{ij}_k \delta^k_l = T^{ij}_l,$$

$$T^{ij}_k \delta^k_j = T^{ij}_j = T^{ik}_k.$$

To understand these rules, consider the first equation. On the left-hand side the index i is a dummy index, signifying summation from $i = 1$ to $i = n$. Whenever $i \neq m$ in this sum we have no contribution since $\delta^m_i = 0$, while the contribution from $i = m$ results in the right-hand side. The remaining equations are proved similarly.

Care should be taken with the expression $\delta^i{}_i$. If we momentarily suspend the summation convention, then obviously $\delta^i_i = 1$, but with the summation convention in operation the i is a dummy index, so that

$$\delta^i_i = \delta^1_1 + \delta^2_2 + \cdots + \delta^n_n = 1 + 1 + \cdots + 1.$$

Hence

$$\delta^i_i = n = \dim V. \tag{3.15}$$

In future, the summation convention will always be assumed to apply unless a rider like 'summation convention suspended' is imposed for some reason.

Basis transformations

Consider a change of basis

$$E = \{e_1, e_2, \dots, e_n\} \longrightarrow E' = \{e'_1, e'_2, \dots, e'_n\}.$$

By Theorem 3.5 each of the original basis vectors e_i has a unique linear expansion in terms of the new basis,

$$e_i = A^j{}_i e'_j, \tag{3.16}$$

where A^j_i represents the jth component of the vector e_i with respect to the basis E'. Of course, the summation convention has now been adopted.

What happens to the components of a typical vector v under such a change of basis? Substituting Eq. (3.16) into the component expansion of v results in

$$v = v^i e_i = v^i A^j{}_i e'_j = v'^j e'_j,$$

where

$$v'^j = A^j_{\ i} v^i. \tag{3.17}$$

This law of transformation of components of a vector v is sometimes called the **contravariant transformation law of components**, a curious and somewhat old-fashioned terminology that possibly defies common sense. Equation (3.17) should be thought of as a 'passive' transformation, since only the *components* of the vector change, not the physical vector itself. On the other hand, a linear transformation $S : V \to V$ of the vector space can be thought of as moving actual vectors around, and for this reason is referred to as an 'active' transformation.

Neverthless, it is still possible to think of (3.17) as a matrix equation if we represent the components v^i and v'^j of the vector v as *column vectors*

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}, \qquad \mathbf{v}' = \begin{pmatrix} v'^1 \\ v'^2 \\ \vdots \\ v'^n \end{pmatrix},$$

and the transformation coefficients $A^j_{\ i}$ as an $n \times n$ matrix

$$\mathsf{A} = \begin{pmatrix} A^1_{\ 1} & A^1_{\ 2} & \dots & A^1_{\ n} \\ A^2_{\ 1} & A^2_{\ 2} & \dots & A^2_{\ n} \\ \dots & \dots & \dots & \dots \\ A^n_{\ 1} & A^n_{\ 2} & \dots & A^n_{\ n} \end{pmatrix}.$$

Equation (3.17) can then be written as a matrix equation

$$\mathbf{v}' = \mathsf{A}\mathbf{v}. \tag{3.18}$$

Note, however, that $\mathsf{A} = [A^j_{\ i}]$ is a matrix of coefficients representing the old basis $\{e_i\}$ in terms of the new basis $\{e'_j\}$. It is not the matrix of components of a linear operator.

Example 3.24 Let V be a three-dimensional vector space with basis $\{e_1, e_2, e_3\}$. Vectors belonging to V can be set in correspondence with the 3×1 column vectors by

$$v = v^1 e_1 + v^2 e_2 + v^3 e_3 \longleftrightarrow \mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix}.$$

Let $\{e'_i\}$ be a new basis defined by

$$e'_1 = e_1, \qquad e'_2 = e_1 + e_2 - e_3, \qquad e'_3 = e_1 + e_2 + e_3.$$

Solving for e_i in terms of the e'_j gives

$$\begin{aligned} e_1 &= e'_1 \\ e_2 &= -e'_1 + \tfrac{1}{2}(e'_2 + e'_3) \\ e_3 &= \tfrac{1}{2}(-e'_2 + e'_3), \end{aligned}$$

and the components of the matrix $\mathsf{A} = [A^j{}_i]$ can be read off using Eq. (3.16),

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

A general vector v is written in the e'_i basis as

$$\begin{aligned} v &= v^1 e_1 + v^2 e_2 + v^3 e_3 \\ &= v^1 e'_1 + v^2 \left(-e'_1 + \tfrac{1}{2}(e'_2 + e'_3)\right) + v^3 \tfrac{1}{2}(-e'_2 + e'_3) \\ &= (v^1 - v^2) e'_1 + \tfrac{1}{2}(v^2 - v^3) e'_2 + \tfrac{1}{2}(v^2 + v^3) e'_3 \\ &= v'^1 e'_1 + v'^2 e'_2 + v'^3 e'_3, \end{aligned}$$

where

$$\mathbf{v}' \equiv \begin{pmatrix} v'^1 \\ v'^2 \\ v'^3 \end{pmatrix} = \begin{pmatrix} v^1 - v^2 \\ \frac{1}{2}(v^2 - v^3) \\ \frac{1}{2}(v^2 + v^3) \end{pmatrix} = \mathsf{A}\mathbf{v}.$$

We will denote the inverse matrix to $\mathsf{A} = [A^i{}_j]$ by $\mathsf{A}' = [A'^j{}_k] = \mathsf{A}^{-1}$. Using the summation convention, the inverse matrix relations

$$\mathsf{A}'\,\mathsf{A} = \mathsf{I}, \qquad \mathsf{A}\,\mathsf{A}' = \mathsf{I}$$

may be written componentwise as

$$A'^k{}_j A^j{}_i = \delta^k_i, \qquad A^i{}_k A'^k{}_j = \delta^i{}_j. \tag{3.19}$$

From (3.16)

$$A'^i{}_k e_i = A'^i{}_k A^j{}_i e'_j = \delta^j_k e'_j = e'_k,$$

which can be rewritten as

$$e'_j = A'^k{}_j e_k. \tag{3.20}$$

Exercise: From Eq. (3.19) or Eq. (3.20) derive the inverse transformation law of vector components

$$v^i = A'^i{}_j v^j. \tag{3.21}$$

We are now in a position to derive the transformation law of components of a linear operator $T : V \to V$. The matrix components of T with respect to the new basis, denoted $\mathsf{T}' = [T'^j{}_i]$, are given by

$$Te'_i = T'^j{}_i\, e'_j,$$

and using Eqs. (3.16) and (3.20) we have

$$\begin{aligned} Te'_i &= T(A'^k_{\ i}\, e_k) \\ &= A'^k_{\ i}\, Te_k \\ &= A'^k_{\ i}\, T^m_k\, e_m \\ &= A'^k_{\ i}\, T^m_k\, A^j_{\ m}\, e'_j. \end{aligned}$$

Hence

$$T'^j_{\ i} = A^j_{\ m}\, T^m_k\, A'^k_{\ i}, \tag{3.22}$$

or in matrix notation, since $\mathsf{A}' = \mathsf{A}^{-1}$,

$$\mathsf{T}' = \mathsf{A}\mathsf{T}\mathsf{A}' = \mathsf{A}\,\mathsf{T}\,\mathsf{A}^{-1}. \tag{3.23}$$

Equation (3.23) is the *passive view* – it represents the change in *components* of an operator under a change of basis. With a different interpretation Eq. (3.23) could however be viewed as an operator equation. If we treat the basis $\{e_1, \dots, e_n\}$ as fixed and regard A as being the matrix representing an operator whose effect on vector components is given by

$$v'^i = A^i_{\ j} v^j \iff \mathbf{v}' = \mathsf{A}\mathbf{v},$$

then Eq. (3.23) represents a *change of operator*, called a **similarity transformation**. If $x' = Ax$, $y' = Ay$, then

$$y' = ATx = ATA^{-1}Ax = T'x',$$

and $T' = ATA^{-1}$ is the operator that relates the transforms, under A, of any pair of vectors x and y that were originally related through the operator T. This is called the *active view* of Eq. (3.23). The two views are often confused in physics, mainly because operators are commonly identified with their matrices. The following example should help to clarify any lingering confusions.

Example 3.25 Consider a clockwise rotation of axes in $\mathbb{R}^2$ through an angle θ,

$$\begin{aligned} e'_1 &= \cos\theta e_1 - \sin\theta e_2 \\ e'_2 &= \sin\theta e_1 + \cos\theta e_2 \end{aligned} \iff \begin{aligned} e_1 &= \cos\theta e'_1 + \sin\theta e'_2 \\ e_2 &= -\sin\theta e'_1 + \cos\theta e'_2. \end{aligned}$$

The matrix of this basis transformation is

$$\mathsf{A} = [A^i_{\ j}] = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and the components of any position vector

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

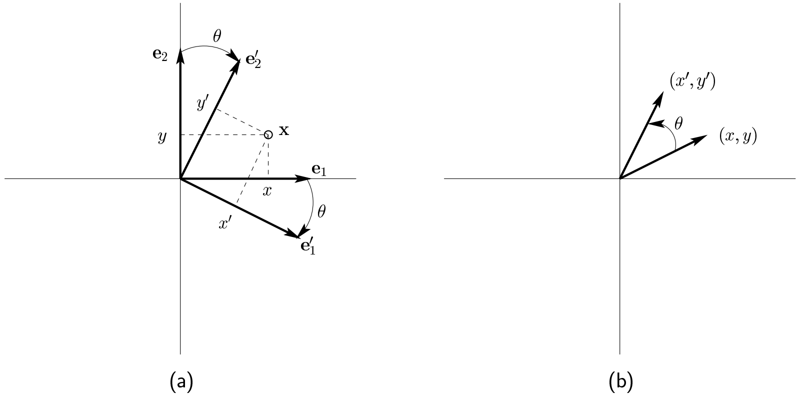

Figure 3.1 Active and passive views of a transformation

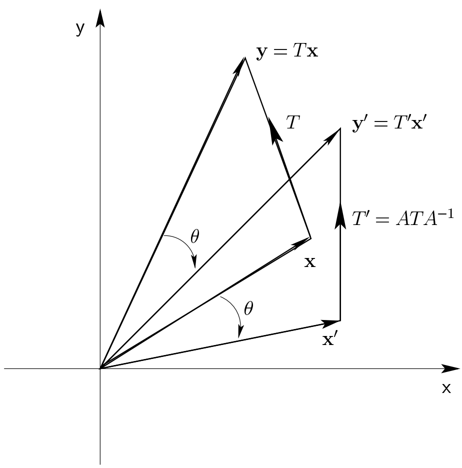

Figure 3.2 Active view of a similarity transformation

change by

$$\mathbf{x}' = \mathsf{A}\mathbf{x} \iff \begin{array}{l} x' = \cos\theta\, x - \sin\theta\, y \\ y' = \sin\theta\, x + \cos\theta\, y. \end{array}$$

This is the passive view. On the other hand, if we regard A as the matrix of components of an operator with respect to fixed axes e_1, e_2, then it represents a physical rotation of the space by an angle θ in a *counterclockwise* direction, opposite to the rotation of the axes in the passive view. Figure 3.1 demonstrates the apparent equivalence of these two views, while Fig. 3.2 illustrates the active view of a similarity transformation $T' = ATA^{-1}$ on a linear operator $T : \mathbb{R}^2 \to \mathbb{R}^2$.

Problems

Problem 3.11 Let $\{e_1, e_2, e_3\}$ be a basis of a three-dimensional vector space V. Show that the vectors $\{e'_1, e'_2, e'_3\}$ defined by

$$e'_1 = e_1 + e_3$$
$$e'_2 = 2e_1 + e_2$$
$$e'_3 = 3e_2 + e_3$$

also form a basis of V.

What are the elements of the matrix $\mathsf{A} = [A^j{}_i]$ in Eq. (3.16)? Calculate the components of the vector

$$v = e_1 - e_2 + e_3$$

with respect to the basis $\{e'_1, e'_2, e'_3\}$, and verify the column vector transformation $\mathbf{v}' = \mathsf{A}\mathbf{v}$.

Problem 3.12 Let $T : V \to W$ be a linear map between vector spaces V and W. If $\{e_i \mid i = 1, \dots, n\}$ is a basis of V and $\{f_a \mid a = 1, \dots, m\}$ a basis of W, how does the matrix T, defined in Problem 3.9, transform under a transformation of bases

$$e_i = A^j{}_i e'_j, \quad f_a = B^b{}_a f'_b \;?$$

Express your answer both in component and in matrix notation.

Problem 3.13 Let e_1, e_2, e_3 be a basis for a three-dimensional vector space and e'_1, e'_2, e'_3 a second basis given by

$$e'_1 = e_3,$$
$$e'_2 = e_2 + 2e_3,$$
$$e'_3 = e_1 + 2e_2 + 3e_3.$$

(a) Express the e_i in terms of the e'_j, and write out the transformation matrices $\mathsf{A} = [A^i{}_j]$ and $\mathsf{A}' = \mathsf{A}^{-1} = [A'^i{}_j]$.
(b) If $u = e_1 + e_2 + e_3$, compute its components in the e'_i basis.
(c) Let T be the linear transformation defined by

$$Te_1 = e_2 + e_3, \qquad Te_2 = e_3 + e_1, \qquad Te_3 = e_1 + e_2.$$

What is the matrix of components of T with respect to the basis e_i?
(d) By evaluating Te'_1, etc. in terms of the e'_j, write out the matrix of components of T with respect to the e'_j basis and verify the similarity transformation $\mathsf{T}' = \mathsf{A}\mathsf{T}\mathsf{A}^{-1}$.

3.7 Dual spaces

Linear functionals

A **linear functional** φ on a vector space V over a field $\mathbb{K}$ is a linear map $\varphi : V \to \mathbb{K}$, where the field of scalars $\mathbb{K}$ is regarded as a one-dimensional vector space, spanned by the element 1,

$$\varphi(au + bv) = a\varphi(u) + b\varphi(v). \tag{3.24}$$

The use of the phrase *linear functional* in place of 'linear function' is largely adopted with infinite dimensional spaces in mind. For example, if V is the space of continuous functions on the interval $[0, 1]$ and $K(x)$ is an integrable function on $[0, 1]$, let $\varphi_K : V \to \mathbb{K}$ be defined by

$$\varphi_K(f) = \int_0^1 K(y) f(y)\, \mathrm{d}y.$$

As the linear map φ_K is a function whose argument is another function, the terminology 'linear functional' seems more appropriate. In this case it is common to write the action of φ_K on the function f as $\varphi_K[f]$ in place of $\varphi_K(f)$.

Theorem 3.8 *If $\varphi : V \to \mathbb{K}$ is any linear functional then its kernel* $\ker\varphi$ *has codimension 1. Conversely, any subspace $W \subset V$ of codimension 1 defines a linear functional φ on V uniquely up to a scalar factor, such that $W = \ker\varphi$.*

Proof: The first part of the theorem follows from Example 3.22 and Eq. (3.3),

$$\operatorname{codim}(\ker\varphi) = \dim(V/(\ker\varphi)) = \dim(\operatorname{im}\varphi) = \dim\mathbb{K} = 1.$$

To prove the converse, let u be any vector not belonging to W – if no such vector exists then $W = V$ and the codimension is 0. The set of cosets $au + W = a(u + W)$ where $a \in \mathbb{K}$ form a one-dimensional vector space that must be identical with all of V/W. Every vector $v \in V$ therefore has a unique decomposition $v = au + w$ where $w \in W$, since

$$v = a'u + w' \implies (a - a')u = w - w' \implies a = a' \text{ and } w = w'.$$

A linear functional φ having kernel W has $\varphi(W) = 0$ and $\varphi(u) = c \neq 0$, for if $c = 0$ then the kernel of φ is all of V. Furthermore, given any non-zero scalar $c \in \mathbb{K}$, these two requirements define a linear functional on V as its value on any vector $v \in V$ is uniquely determined by

$$\varphi(v) = \varphi(au + w) = ac.$$

If φ' is any other such linear functional, having $c' = \varphi'(u) \neq 0$, then $\varphi' = (c'/c)\varphi$ since $\varphi'(v) = ac' = (c'/c)\varphi(v)$. ■

This proof even applies, as it stands, to infinite dimensional spaces, although in that case it is usual to impose the added stipulation that linear functionals be continuous (see Section 10.9 and Chapter 13).

Exercise: If ω and ρ are linear functionals on V, show that

$$\ker\omega = \ker\rho \iff \omega = a\rho \quad \text{for some } a \in \mathbb{K}.$$

Example 3.26 Let $V = \mathbb{K}^n$, where n is any integer ≥ 2 or possibly $n = \infty$. For convenience, we will take V to be the space of row vectors of length n here. Let W be the subspace of vectors

$$W = \{(x_1, x_2, \dots) \in \mathbb{K}^n \mid x_1 + x_2 = 0\}.$$

This is a subspace of codimension 1, for if we set $u = (1, 0, 0, \dots)$, then any vector v can be written

$$v = (v_1, v_2, v_3, \dots) = au + w$$

where $a = v_1 + v_2$ and $w = (-v_2, v_2, v_3, \dots)$. This decomposition is unique, for if $au + w = a'u + w'$ then $(a - a')u = w' - w \in W$. Hence $a = a'$ and $w = w'$, since $u \notin W$. Every coset $v + W$ can therefore be uniquely expressed as $a(u + W)$, and W has codimension 1.

Let $\varphi : V \to \mathbb{K}$ be the linear functional such that $\varphi(W) = 0$ and $\varphi(u) = 1$. Then $\varphi(v) = \varphi(au + w) = a\varphi(u) + \varphi(w) = a$, so that

$$\varphi\big((x_1, x_2, x_3, \dots)\big) = x_1 + x_2.$$

The kernel of φ is evidently W, and every other linear functional φ' having kernel W is of the form $\varphi\big((x_1, x_2, x_3, \dots)\big) = c(x_1 + x_2)$.

The dual space of a vector space

As for general linear maps, it is possible to add linear functionals and multiply them by scalars,

$$(\varphi + \omega)(u) = \varphi(u) + \omega(u), \qquad (a\omega)(u) = a\omega(u).$$

With respect to these operations the set of linear functionals on V forms a vector space over $\mathbb{K}$ called the **dual space** of V, usually denoted V^*. In keeping with earlier conventions, other possible notations for this space are $L(V, \mathbb{K})$ or $\mathrm{Hom}(V, \mathbb{K})$. Frequently, linear functionals on V will be called **covectors**, and in later chapters we will have reason to refer to them as **1-forms**.

Let V be a finite dimensional vector space, $\dim V = n$, and $\{e_1, e_2, \dots, e_n\}$ any basis for V. A linear functional ω on V is uniquely defined by assigning its values on the basis vectors,

$$w_1 = \omega(e_1), \quad w_2 = \omega(e_2), \quad \dots, \quad w_n = \omega(e_n),$$

since the value on any vector $u = u^i e_i \equiv \sum_{i=1}^n u^i e_i$ can be determined using linearity (3.24),

$$\omega(u) = \omega(u^i e_i) = u^i \omega(e_i) = w_i u^i \equiv \sum_{i=1}^n w_i u^i. \tag{3.25}$$

Define n linear functionals $\varepsilon^1, \varepsilon^2, \dots, \varepsilon^n$ by

$$\varepsilon^i(e_j) = \delta^i_{\ j} \tag{3.26}$$

where $\delta^i_{\ j}$ is the Kronecker delta defined in Eq. (3.14). Note that these equations uniquely define each linear functional ε^i, since their values are assigned on each basis vector e_j in turn.

Theorem 3.9 *The n-linear functionals* $\{\varepsilon^1, \varepsilon^2, \dots, \varepsilon^n\}$ *form a basis of* V^**, called the* **dual basis** *to* $\{e_1, \dots, e_n\}$*. Hence* $\dim V^* = n = \dim V$.

Proof: Firstly, suppose there are scalars $a_i \in \mathbb{K}$ such that

$$a_i \varepsilon^i = 0.$$

Applying the linear functional on the left-hand side of this equation to an arbitrary basis vector e_j,

$$0 = a_i \varepsilon^i(e_j) = a_i \delta^i{}_j = a_j,$$

shows that the linear functionals $\{\varepsilon^1, \dots, \varepsilon^n\}$ are linearly independent. Furthermore these linear functions span V^* since every linear functional ω on V can be written

$$\omega = w_i \varepsilon^i \quad \text{where} \quad w_i = \omega(e_i). \tag{3.27}$$

This follows from

$$\begin{aligned} w_i \varepsilon^i(u) &= w_i \varepsilon^i(u^j e_j) \\ &= w_i u^j \varepsilon^i(e_j) \\ &= w_i u^j \delta^i_j \\ &= w_i u^i \\ &= \omega(u) \quad \text{by Eq. (3.25).} \end{aligned}$$

Thus ω and $\omega' = w_i \varepsilon^i$ have the same effect on every vector $u = u^i e_i$; they are therefore identical linear functionals. The proposition that $\dim V^* = n$ follows from Corollary 3.4. ∎

Exercise: Show that the expansion (3.27) is unique; i.e., if $\omega = w'_i \varepsilon^i$ then $w'_i = w_i$ for each $i = 1, \dots, n$.

Given a basis $E = \{e_i\}$, we will frequently refer to the n numbers $w_i = \omega(e_i)$ as the **components of the linear functional** ω in this basis. Alternatively, we can think of them as the components of ω with respect to the dual basis in V^*. The formula (3.25) has a somewhat deceptive 'dot product' feel about it. In Chapter 5 a dot product will be correctly defined as a product between vectors from the *same* vector space, while (3.25) is a product between vectors from different spaces V and V^*. It is, in fact, better to think of the components u^i of a vector u from V as forming a column vector, while the components w_j of a linear functional ω form a row vector. The above product then makes sense as a matrix product between a $1 \times n$ row matrix and an $n \times 1$ column matrix. While a vector is often thought of geometrically as a directed line segment, often represented by an arrow, this is not a good way to think of a covector. Perhaps the best way to visualize a linear functional is as a set of parallel planes of vectors determined by $\omega(v) = \text{const.}$ (see Fig. 3.3).

Dual of the dual

We may enquire whether this dualizing process can be continued to generate further vector spaces such as the dual of the dual space V^{**}, etc. For finite dimensional spaces the process essentially stops at the first dual, for there is a completely natural way in which V can be

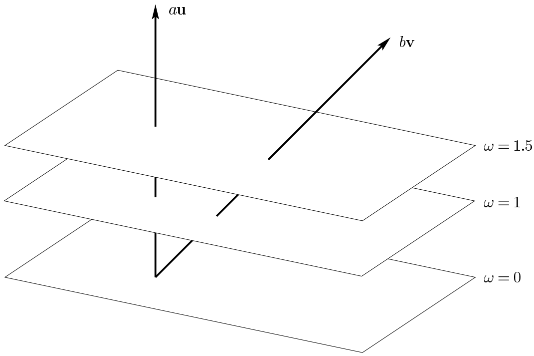

Figure 3.3 Geometrical picture of a linear functional

identified with V^{**}. To understand how V itself can be regarded as the dual space of V^*, define a linear map $\bar{v} : V^* \to \mathbb{K}$ corresponding to any vector $v \in V$ by

$$\bar{v}(\omega) = \omega(v) \quad \text{for all } \omega \in V^*. \tag{3.28}$$

The map $\bar{v}$ is a linear functional on V^*, since

$$\begin{aligned} \bar{v}(a\omega + b\rho) &= (a\omega + b\rho)(v) \\ &= a\omega(v) + b\rho(v) \\ &= a\bar{v}(\omega) + b\bar{v}(\rho). \end{aligned}$$

The map $\beta : V \to V^{**}$ defined by $\beta(v) = \bar{v}$ is linear, since

$$\begin{aligned} \beta(au + bv)(\varphi) &= \overline{au + bv}(\varphi) \quad (\varphi \in V^*) \\ &= \varphi(au + bv) \\ &= a\varphi(u) + b\varphi(v) \\ &= a\bar{u}(\varphi) + b\bar{v}(\varphi) \\ &= a\beta(u)(\varphi) + b\beta(v)(\varphi). \end{aligned}$$

As this holds for arbitrary covectors φ we have $\beta(au + bv) = a\beta(u) + b\beta(v)$. Furthermore if $e_1, e_2, \dots, e_n$ is a basis of V with dual basis $\varepsilon^1, \varepsilon^2, \dots, \varepsilon^n$, then

$$\bar{e}_i(\varepsilon^j) = \varepsilon^j(e_i) = \delta^j_i,$$

and it follows from Theorem 3.9 that $\{\beta(e_i) = \bar{e}_i\}$ is the basis of V^{**} dual to the basis $\{\varepsilon^j\}$ of V^*. The map $\beta : V \to V^{**}$ is therefore onto since every $f \in V^{**}$ can be written in the form

$$f = u^i \beta(e_i) = \beta(u) = \bar{u} \quad \text{where} \quad u = u^i e_i.$$

Since $\{\beta(e_i)\}$ is a basis, it follows from Theorem 3.5 that the components u^i, and therefore the vector u, are uniquely determined by f. The map β is thus a vector space isomorphism, as it is both onto and one-to-one.

We have shown that $V \cong V^{**}$. In itself this is to be expected since these spaces have the same dimension, but the significant thing to note is that since the defining Eq. (3.28) makes no mention of any particular choice of basis, the correspondence between V and V^{**} is totally *natural*. There is therefore no ambiguity in *identifying* $\bar{v}$ with v, and rewriting (3.28) as

$$v(\omega) = \omega(v).$$

This reciprocity between the two spaces V and V^* lends itself to the following alternative notations, which will be used interchangeably throughout this book:

$$\langle \omega, v \rangle \equiv \langle v, \omega \rangle \equiv v(\omega) \equiv \omega(v). \tag{3.29}$$

However, it should be pointed out that the identification of V and V^{**} will only work for *finite* dimensional vector spaces. In infinite dimensional spaces, every vector may be regarded as a linear functional on V^* in a natural way, but the converse is not true – there exist linear functionals on V^* that do not correspond to vectors from V.

Transformation law of covector components

By Theorem 3.9 the spaces V and V^* are in one-to-one correspondence since they have the same dimension, but unlike that described above between V and V^{**} this correspondence is not *natural*. For example, if $v = v^i e_i$ is a vector in V let υ be the linear functional whose components in the dual basis are exactly the same as the components of the original vector, $\upsilon = v^i \varepsilon^i$. While the map $v \mapsto \upsilon$ is a vector space isomorphism, the same rule applied with respect to a different basis $\{e'_i\}$ will generally lead to a different correspondence between vectors and covectors. Thus, given an arbitrary vector $v \in V$ there is no *basis-independent* way of pointing to a covector partner in V^*. Essentially this arises from the fact that the law of transformation of components of a linear functional is different from the transformation law of components for a vector.

We have seen in Section 3.6 that the transformation of a basis can be written by Eqs. (3.16) and (3.20),

$$e_i = A^j{}_i e'_j, \qquad e'_j = A'^k{}_j e_k \tag{3.30}$$

where $[A^i{}_k]$ and $[A'^i{}_k]$ are related through the inverse matrix equations (3.19). Let $\{\varepsilon^i\}$ and $\{\varepsilon'^i\}$ be the dual bases corresponding to the bases $\{e_j\}$ and $\{e'_j\}$ respectively of V,

$$\varepsilon^i(e_j) = \delta^i_j, \qquad \varepsilon'^i(e'_j) = \delta^i_j. \tag{3.31}$$

Set

$$\varepsilon^i = B^i{}_j \varepsilon'^j,$$

and substituting this and Eq. (3.30) into the first identity of (3.31) gives, after replacing the index j by k,

$$\begin{aligned}\delta^i_k &= B^i{}_j \varepsilon'^j(e_k) \\ &= B^i{}_j \varepsilon'^j(A^l{}_k e'_l) \\ &= B^i{}_j A^l{}_k \delta^j_l \\ &= B^i{}_j A^j{}_k.\end{aligned}$$

Hence $B^i{}_j = A'^i{}_j$ and the transformation of the dual basis is

$$\varepsilon^i = A'^i{}_j \varepsilon'^j. \tag{3.32}$$

Exercise: Show the transform inverse to (3.32)

$$\varepsilon'^j = A^j{}_k \varepsilon^k. \tag{3.33}$$

If $\omega = w_i \varepsilon^i$ is a linear functional having components w_i with respect to the first basis, then

$$\omega = w_i \varepsilon^i = w_i A'^i{}_j \varepsilon'^j = w'_i \varepsilon'^i$$

where

$$w'_i = A'^j{}_i w_j. \tag{3.34}$$

This is known as the **covariant vector transformation law of components**. Its inverse is

$$w_j = A^k{}_j w'_k. \tag{3.35}$$

These equations are to be compared with the contravariant vector transformation law of components of a vector $v = v^i e_i$, given by Eqs. (3.17) and (3.21),

$$v'^j = A^j{}_i v^i, \qquad v^i = A'^i{}_j v^j. \tag{3.36}$$

Exercise: Verify directly from (3.34) and (3.36) that Eq. (3.25) is basis-independent,

$$\omega(u) = w_i u^i = w'_j u'^j.$$

Exercise: Show that if the components of ω are displayed as a $1 \times n$ row matrix $\mathbf{w}^T = (w_1, w_2, \dots, w_n)$ then the transformation law (3.34) can be written as a matrix equation

$$\mathbf{w}'^T = \mathbf{w}^T \mathsf{A}.$$

Problems

Problem 3.14 Find the dual basis to the basis of $\mathbb{R}^3$ having column vector representation

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \mathbf{e}_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Problem 3.15 Let $\mathcal{P}(x)$ be the vector space of real polynomials $f(x) = a_0 + a_1 x + \dots + a_n x^n$. If $(b^0, b^1, b^2, \dots)$ is any sequence of real numbers, show that the map $\beta : \mathcal{P}(x) \to \mathbb{R}$ given by

$$\beta(f(x)) = \sum_{i=0}^{n} b^i a_i$$

is a linear functional on $\mathcal{P}(x)$.

Show that *every* linear functional β on $\mathcal{P}(x)$ can be obtained in this way from such a sequence and hence that $(\hat{\mathbb{R}}^\infty)^* = \mathbb{R}^\infty$.

Problem 3.16 Define the **annihilator** $S^\perp$ of a subset $S \subseteq V$ as the set of all linear functionals that vanish on S,

$$S^\perp = \{\omega \in V^* \mid \omega(u) = 0 \quad \forall u \in S\}.$$

(a) Show that for any subset S, $S^\perp$ is a vector subspace of V^*.
(b) If $T \subseteq S$, show that $S^\perp \subseteq T^\perp$.
(c) If U is a vector subspace of V, show that $(V/U)^* \cong U^\perp$. [*Hint*: For each ω in $U^\perp$ define the element $\bar{\omega} \in (V/U)^*$ by $\bar{\omega}(v + U) = \omega(v)$.]
(d) Show that $U^* \cong V^*/U^\perp$.
(e) If V is finite dimensional with $\dim V = n$ and W is any subspace of V with $\dim W = m$, show that $\dim W^\perp = n - m$. [*Hint*: Use a basis adapted to the subspace W by Theorem 3.7 and consider its dual basis in V^*.]
(f) Adopting the natural identification of V and V^{**}, show that $(W^\perp)^\perp = W$.

Problem 3.17 Let u be a vector in the vector space V of dimension n.

(a) If ω is a linear functional on V such that $a = \omega(u) \neq 0$, show that a basis $e_1, \dots, e_n$ can be chosen such that

$$u = e_1 \quad \text{and} \quad \omega = a\varepsilon^1$$

where $\{\varepsilon^1, \dots, \varepsilon^n\}$ is the dual basis. [*Hint*: Apply Theorem 3.7 to the vector u and try a further basis transformation of the form $e'_1 = e_1,\ e'_2 = e_2 + a_2 e_1, \dots,\ e'_n = e_n + a_n e_1$.]
(b) If $a = 0$, show that the basis may be chosen such that

$$u = e_1 \quad \text{and} \quad \omega = \varepsilon^2.$$

Problem 3.18 For the three-dimensional basis transformation of Problem 3.13 evaluate the ε'^j dual to e'_i in terms of the dual basis ε^j. What are the components of the linear functional $\omega = \varepsilon^1 + \varepsilon^2 + \varepsilon^3$ with respect to the new dual basis?

Problem 3.19 If $A : V \to V$ is a linear operator, define its **transpose** to be the linear map $A' : V^* \to V^*$ such that

$$A'\omega(u) = \omega(Au), \quad \forall\, u \in V,\ \omega \in V^*.$$

Show that this relation uniquely defines the linear operator A' and that

$$O' = O, \qquad (\mathrm{id}_V)' = \mathrm{id}_{V^*}, \qquad (aB + bA)' = aB' + bA', \quad \forall a, b \in \mathbb{K}.$$

(a) Show that $(BA)' = A'B'$.
(b) If A is an invertible operator then show that $(A')^{-1} = (A^{-1})'$.
(c) If V is finite dimensional show that $A'' = A$, if we make the natural identification of V^{**} and V.
(d) Show that the matrix of components of the transpose map A' with respect to the dual basis is the transpose of the matrix of A, $\mathsf{A}' = \mathsf{A}^T$.
(e) Using Problem 3.16 show that $\ker A' = (\mathrm{im}\, A)^\perp$.
(f) Use (3.10) to show that the rank of A' equals the rank of A.

Problem 3.20 The *row rank* of a matrix is defined as the maximum number of linearly independent rows, while its *column rank* is the maximum number of linearly independent columns.

(a) Show that the rank of a linear operator A on a finite dimensional vector space V is equal to the column rank of its matrix A with respect to any basis of V.
(b) Use parts (d) and (f) of Problem 3.19 to show that the row rank of a square matrix is equal to its column rank.

Problem 3.21 Let S be a linear operator on a vector space V.

(a) Show that the rank of S is one, $\rho(S) = 1$, if and only if there exists a non-zero vector u and a non-zero linear functional α such that

$$S(v) = u\alpha(v).$$

(b) With respect to any basis $\{e_i\}$ of V and its dual basis $\{\varepsilon^j\}$, show that

$$S^i{}_j = u^i a_j \quad \text{where} \quad u = u^i e_i, \quad \alpha = a_j \varepsilon^j.$$

(c) Show that every linear operator A of rank r can be written as a sum of r linear operators of rank one.
(d) Show that the last statement is equivalent to the assertion that for every matrix A of rank r there exist column vectors $\mathbf{u}_i$ and $\mathbf{a}_i$ $(i = 1, \dots, r)$ such that

$$\mathsf{S} = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{a}_i^T.$$

References

[1] G. Birkhoff and S. MacLane. *A Survey of Modern Algebra*. New York, MacMillan, 1953.
[2] N. B. Haaser and J. A. Sullivan. *Real Analysis*. New York, Van Nostrand Reinhold Company, 1971.
[3] S. Hassani. *Foundations of Mathematical Physics*. Boston, Allyn and Bacon, 1991.

[4] R. Geroch. *Mathematical Physics*. Chicago, The University of Chicago Press, 1985.
[5] S. Lang. *Algebra*. Reading, Mass., Addison-Wesley, 1965.
[6] L. H. Loomis and S. Sternberg. *Advanced Calculus*. Reading, Mass., Addison-Wesley, 1968.
[7] P. R. Halmos. *Finite-dimensional Vector Spaces*. New York, D. Van Nostrand Company, 1958.