Mathematics for Physics

A guided tour for graduate students

Michael Stone

and

Paul Goldbart

PIMANDER-CASAUBON
Alexandria • Florence • London

Appendix A

Linear Algebra Review

In physics we often have to work with infinite dimensional vector spaces. Navigating these vasty deeps is much easier if you have a sound grasp of the theory of finite dimensional spaces. Most physics students have studied this as undergraduates, but not always in a systematic way. In this appendix we gather together and review those parts of linear algebra that we will find useful in the main text.

A.1 Vector space

A.1.1 Axioms

A *vector space* V over a field $\mathbb{F}$ is a set equipped with two operations: a binary operation called *vector addition* which assigns to each pair of elements $\mathbf{x}$, $\mathbf{y} \in V$ a third element denoted by $\mathbf{x}+\mathbf{y}$, and *scalar multiplication* which assigns to an element $\mathbf{x} \in V$ and $\lambda \in \mathbb{F}$ a new element $\lambda\mathbf{x} \in V$. There is also a distinguished element $\mathbf{0} \in V$ such that the following axioms are obeyed:[1]

1) Vector addition is commutative: $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$.
2) Vector addition is associative: $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$.
3) Additive identity: $\mathbf{0}+\mathbf{x} = \mathbf{x}$.
4) Existence of an additive inverse: for any $\mathbf{x} \in V$, there is an element $(-\mathbf{x}) \in V$, such that $\mathbf{x}+(-\mathbf{x}) = \mathbf{0}$.
5) Scalar distributive law i) $\lambda(\mathbf{x}+\mathbf{y}) = \lambda\mathbf{x}+\lambda\mathbf{y}$.
6) Scalar distributive law ii) $(\lambda+\mu)\mathbf{x} = \lambda\mathbf{x}+\mu\mathbf{x}$.

[1] In this list 1, λ, $\mu, \in \mathbb{F}$ and $\mathbf{x}$, $\mathbf{y}, \mathbf{0} \in V$.

7) Scalar multiplicatiion is associative: $(\lambda\mu)\mathbf{x} = \lambda(\mu\mathbf{x})$.
8) Multiplicative identity: $1\mathbf{x} = \mathbf{x}$.

The elements of V are called *vectors*. We will only consider vector spaces over the field of the real numbers, $\mathbb{F} = \mathbb{R}$, or the complex numbers, $\mathbb{F} = \mathbb{C}$.

You have no doubt been working with vectors for years, and are saying to yourself "I know this stuff." Perhaps so, but to see if you really understand these axioms try the following exercise. Its value lies not so much in the solution of its parts, which are easy, as in appreciating that these commonly used properties both can and need to be proved from the axioms. (Hint: work the problems in the order given; the later parts depend on the earlier.)

Exercise A.1: Use the axioms to show that:

i) If $\mathbf{x} + \tilde{\mathbf{0}} = \mathbf{x}$, then $\tilde{\mathbf{0}} = \mathbf{0}$.
ii) We have $0\mathbf{x} = \mathbf{0}$ for any $\mathbf{x} \in V$. Here 0 is the additive identity in $\mathbb{F}$.
iii) If $\mathbf{x} + \mathbf{y} = \mathbf{0}$, then $\mathbf{y} = -\mathbf{x}$. Thus the additive inverse is unique.
iv) Given $\mathbf{x}$, $\mathbf{y}$ in V, there is a unique $\mathbf{z}$ such that $\mathbf{x}+\mathbf{z} = \mathbf{y}$, to whit $\mathbf{z} = \mathbf{x}-\mathbf{y}$.
v) $\lambda\mathbf{0} = \mathbf{0}$ for any $\lambda \in \mathbb{F}$.
vi) If $\lambda\mathbf{x} = \mathbf{0}$, then either $\mathbf{x} = \mathbf{0}$ or $\lambda = 0$.
vii) $(-1)\mathbf{x} = -\mathbf{x}$.

A.1.2 Bases and components

Let V be a vector space over $\mathbb{F}$. For the moment, this space has no additional structure beyond that of the previous section — no inner product and so no notion of what it means for two vectors to be orthogonal. There is still much that can be done, though. Here are the most basic concepts and properties that need to be understand:

i) A set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is *linearly dependent* if there exist $\lambda^\mu \in \mathbb{F}$, not all zero, such that

$$\lambda^1\mathbf{e}_1 + \lambda^2\mathbf{e}_2 + \cdots + \lambda^n\mathbf{e}_n = \mathbf{0}. \tag{A.1}$$

ii) If it is not linearly dependent, a set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is *linearly independent*. For a linearly independent set, a relation

$$\lambda^1\mathbf{e}_1 + \lambda^2\mathbf{e}_2 + \cdots + \lambda^n\mathbf{e}_n = \mathbf{0} \tag{A.2}$$

can hold only if all the λ^μ are zero.

iii) A set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is said to *span* V if for any $\mathbf{x} \in V$ there are numbers x^μ such that $\mathbf{x}$ can be written (not necessarily uniquely) as

$$\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + \cdots + x^n\mathbf{e}_n. \tag{A.3}$$

A vector space is *finite dimensional* if a finite spanning set exists.

iv) A set of vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a *basis* if it is a *maximal linearly independent set* (*i.e.* introducing any additional vector makes the set linearly dependent). An alternative definition declares a basis to be a *minimal spanning set* (*i.e.* deleting any of the $\mathbf{e}_i$ destroys the spanning property). *Exercise*: Show that these two definitions are equivalent.

v) If $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis then any $\mathbf{x} \in V$ can be written

$$\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + \ldots x^n\mathbf{e}_n, \tag{A.4}$$

where the x^μ, the *components* of the vector with respect to this basis, are unique in that two vectors coincide if and only if they have the same components.

vi) **Fundamental Theorem**: If the sets $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ are both bases for the space V then $m = n$. This invariant integer is the *dimension*, $\dim(V)$, of the space. For a proof (not difficult) see a mathematics text such as Birkhoff and McLane's *Survey of Modern Algebra*, or Halmos' *Finite Dimensional Vector Spaces*.

Suppose that $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_n\}$ are both bases, and that

$$\mathbf{e}_\nu = a^\mu_\nu \mathbf{e}'_\mu. \tag{A.5}$$

Since $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis, the $\mathbf{e}'_\nu$ can also be uniquely expressed in terms of the $\mathbf{e}_\mu$, and so the numbers a^μ_ν constitute an invertible matrix. (Note that we are, as usual, using the Einstein summation convention that repeated indices are to be summed over.) The components x'^μ of $\mathbf{x}$ in the new basis are then found by comparing the coefficients of $\mathbf{e}'_\mu$ in

$$x'^\mu \mathbf{e}'_\mu = \mathbf{x} = x^\nu \mathbf{e}_\nu = x^\nu \left(a^\mu_\nu \mathbf{e}'_\mu\right) = \left(x^\nu a^\mu_\nu\right) \mathbf{e}'_\mu \tag{A.6}$$

to be $x'^\mu = a^\mu_\nu x^\nu$, or equivalently, $x^\nu = (a^{-1})^\nu_\mu\, x'^\mu$. Note how the $\mathbf{e}_\mu$ and the x^μ transform in "opposite" directions. The components x^μ are therefore said to transform *contravariantly*.

A.2 Linear maps

Let V and W be vector spaces having dimensions n and m respectively. A *linear map*, or *linear operator*, A is a function $A : V \to W$ with the property that

$$A(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda A(\mathbf{x}) + \mu A(\mathbf{y}). \tag{A.7}$$

A.2.1 Matrices

The linear map A is an object that exists independently of any basis. Given bases $\{\mathbf{e}_\mu\}$ for V and $\{\mathbf{f}_\nu\}$ for W, however, the map may be represented by an m-by-n *matrix*. We obtain this matrix

$$\mathbf{A} = \begin{pmatrix} a^1{}_1 & a^1{}_2 & \dots & a^1{}_n \\ a^2{}_1 & a^2{}_2 & \dots & a^2{}_n \\ \vdots & \vdots & \ddots & \vdots \\ a^m{}_1 & a^m{}_2 & \dots & a^m{}_n \end{pmatrix}, \tag{A.8}$$

having entries $a^\nu{}_\mu$, by looking at the action of A on the basis elements:

$$A(\mathbf{e}_\mu) = \mathbf{f}_\nu a^\nu{}_\mu \,. \tag{A.9}$$

To make the right-hand-side of (A.9) look like a matrix product, where we sum over adjacent indices, the array $a^\nu{}_\mu$ has been written to the *right* of the basis vector.[2] The map $\mathbf{y} = A(\mathbf{x})$ is therefore

$$\mathbf{y} \equiv y^\nu \mathbf{f}_\nu = A(\mathbf{x}) = A(x^\mu \mathbf{e}_\mu) = x^\mu A(\mathbf{e}_\mu) = x^\mu(\mathbf{f}_\nu a^\nu{}_\mu) = (a^\nu{}_\mu x^\mu)\mathbf{f}_\nu, \tag{A.10}$$

whence, comparing coefficients of $\mathbf{f}_\nu$, we have

$$y^\nu = a^\nu{}_\mu x^\mu. \tag{A.11}$$

The action of the linear map on *components* is therefore given by the usual matrix multiplication from the *left*: $\mathbf{y} = \mathbf{A}\mathbf{x}$, or more explicitly

$$\begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix} = \begin{pmatrix} a^1{}_1 & a^1{}_2 & \dots & a^1{}_n \\ a^2{}_1 & a^2{}_2 & \dots & a^2{}_n \\ \vdots & \vdots & \ddots & \vdots \\ a^m{}_1 & a^m{}_2 & \dots & a^m{}_n \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}. \tag{A.12}$$

[2] You have probably seen this "backward" action before in quantum mechanics. If we use Dirac notation $|n\rangle$ for an orthonormal basis, and insert a complete set of states, $|m\rangle\langle m|$, then $A|n\rangle = |m\rangle\langle m|A|n\rangle$. The matrix $\langle m|A|n\rangle$ representing the operator A operating on a vector from the *left* thus automatically appears to the *right* of the basis vectors used to expand the result.

The *identity map* $I : V \to V$ is represented by the n-by-n matrix

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \tag{A.13}$$

which has the same entries in any basis.

Exercise A.2: Let U, V, W be vector spaces, and $A : V \to W$, $B : U \to V$ linear maps which are represented by the matrices $\mathbf{A}$ with entries $a^\mu{}_\nu$ and $\mathbf{B}$ with entries $b^\mu{}_\nu$, respectively. Use the action of the maps on basis elements to show that the map $AB : U \to W$ is represented by the matrix product $\mathbf{AB}$ whose entries are $a^\mu{}_\lambda b^\lambda{}_\nu$.

A.2.2 Range-nullspace theorem

Given a linear map $A : V \to W$, we can define two important subspaces:

i) The *kernel* or *nullspace* is defined by

$$\operatorname{Ker} A = \{\mathbf{x} \in V : A(\mathbf{x}) = \mathbf{0}\}. \tag{A.14}$$

It is a subspace of V.

ii) The *range* or *image* space is defined by

$$\operatorname{Im} A = \{\mathbf{y} \in W : \mathbf{y} = A(\mathbf{x}), \mathbf{x} \in V\}. \tag{A.15}$$

It is a subspace of the *target space* W.

The key result linking these spaces is the *range-nullspace theorem* which states that

$$\boxed{\dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A) = \dim V}$$

It is proved by taking a basis $\mathbf{n}_\mu$ for $\operatorname{Ker} A$ and extending it to a basis for the whole of V by appending $(\dim V - \dim(\operatorname{Ker} A))$ extra vectors $\mathbf{e}_\nu$. It is easy to see that the vectors $A(\mathbf{e}_\nu)$ are linearly independent and span $\operatorname{Im} A \subseteq W$. Note that this result is meaningless unless V is finite dimensional.

The number $\dim(\operatorname{Im} A)$ is the number of linearly independent columns in the matrix, and is often called the (column) rank of the matrix.

A.2.3 The dual space

Associated with the vector space V is its *dual space*, V^*, which is the set of linear maps $f : V \to \mathbb{F}$. In other words the set of linear functions $f(\ )$ that take in a vector and return a number. These functions are often also called *covectors*. (Mathematicians place the prefix *co-* in front of the name of a mathematical object to indicate a dual class of objects, consisting of the set of structure-preserving maps of the original objects into the field over which they are defined.)

Using linearity we have

$$f(\mathbf{x}) = f(x^\mu \mathbf{e}_\mu) = x^\mu f(\mathbf{e}_\mu) = x^\mu\, f_\mu. \tag{A.16}$$

The set of numbers $f_\mu = f(\mathbf{e}_\mu)$ are the components of the covector $f \in V^*$. If we change basis $\mathbf{e}_\nu = a^\mu_\nu \mathbf{e}'_\mu$ then

$$f_\nu = f(\mathbf{e}_\nu) = f(a^\mu_\nu \mathbf{e}'_\mu) = a^\mu_\nu f(\mathbf{e}'_\mu) = a^\mu_\nu f'_\mu. \tag{A.17}$$

Thus $f_\nu = a^\mu_\nu f'_\mu$ and the f_μ components transform in the same manner as the basis. They are therefore said to transform *covariantly*.

Given a basis $\mathbf{e}_\mu$ of V, we can define a *dual basis* for V^* as the set of covectors $\mathbf{e}^{*\mu} \in V^*$ such that

$$\mathbf{e}^{*\mu}(\mathbf{e}_\nu) = \delta^\mu_\nu. \tag{A.18}$$

It should be clear that this is a basis for V^*, and that f can be expanded

$$f = f_\mu \mathbf{e}^{*\mu}. \tag{A.19}$$

Although the spaces V and V^* have the same dimension, and are therefore isomorphic, there is no natural map between them. The assignment $\mathbf{e}_\mu \mapsto \mathbf{e}^{*\mu}$ is *unnatural* because it depends on the choice of basis.

One way of driving home the distinction between V and V^* is to consider the space V of fruit orders at a grocers. Assume that the grocer stocks only apples, oranges and pears. The elements of V are then vectors such as

$$\mathbf{x} = 3\text{kg}\,\mathbf{apples} + 4.5\text{kg}\,\mathbf{oranges} + 2\text{kg}\,\mathbf{pears}. \tag{A.20}$$

Take V^* to be the space of possible price lists, an example element being

$$f = (£3.00/\text{kg})\,\mathbf{apples}^* + (£2.00/\text{kg})\,\mathbf{oranges}^* + (£1.50/\text{kg})\,\mathbf{pears}^*. \tag{A.21}$$

The evaluation of f on $\mathbf{x}$

$$f(\mathbf{x}) = 3 \times £3.00 + 4.5 \times £2.00 + 2 \times £1.50 = £21.00, \tag{A.22}$$

then returns the total cost of the order. You should have no difficulty in distinguishing between a price list and box of fruit!

We may consider the original vector space V to be the dual space of V^* since, given vectors in $\mathbf{x} \in V$ and $f \in V^*$, we naturally define $\mathbf{x}(f)$ to be $f(\mathbf{x})$. Thus $(V^*)^* = V$. Instead of giving one space priority as being the set of linear functions on the other, we can treat V and V^* on an equal footing. We then speak of the *pairing* of $\mathbf{x} \in V$ with $f \in V^*$ to get a number in the field. It is then common to use the notation $(f, \mathbf{x})$ to mean either of $f(\mathbf{x})$ or $\mathbf{x}(f)$. **Warning:** despite the similarity of the notation, do not fall into the trap of thinking of the pairing $(f, \mathbf{x})$ as an inner product (see next section) of f with $\mathbf{x}$. The two objects being paired live in different spaces. In an inner product, the vectors being multiplied live in the same space.

A.3 Inner-product spaces

Some vector spaces V come equipped with an inner (or scalar) product. This additional structure allows us to relate V and V^*.

A.3.1 Inner products

We will use the symbol $\langle \mathbf{x}, \mathbf{y} \rangle$ to denote an *inner product*. An inner (or *scalar*) product is a conjugate-symmetric, sesquilinear, non-degenerate map $V \times V \to \mathbb{F}$. In this string of jargon, the phrase *conjugate symmetric* means that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*, \tag{A.23}$$

where the "$*$" denotes complex conjugation, and *sesquilinear*[3] means

$$\begin{aligned} \langle \mathbf{x}, \lambda\mathbf{y} + \mu\mathbf{z} \rangle &= \lambda\langle \mathbf{x}, \mathbf{y} \rangle + \mu\langle \mathbf{x}, \mathbf{z} \rangle, & \text{(A.24)} \\ \langle \lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z} \rangle &= \lambda^*\langle \mathbf{x}, \mathbf{z} \rangle + \mu^*\langle \mathbf{y}, \mathbf{z} \rangle. & \text{(A.25)} \end{aligned}$$

The product is therefore linear in the second slot, but *anti-linear* in the first. When our field is the real numbers $\mathbb{R}$ then the complex conjugation is

[3] *Sesqui* is a Latin prefix meaning "one-and-a-half".

redundant and the product will be symmetric

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle, \tag{A.26}$$

and bilinear

$$\begin{aligned} \langle \mathbf{x}, \lambda\mathbf{y} + \mu\mathbf{z} \rangle &= \lambda\langle \mathbf{x}, \mathbf{y})\rangle + \mu\langle \mathbf{x}, \mathbf{z} \rangle, & \text{(A.27)} \\ \langle \lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z} \rangle &= \lambda\langle \mathbf{x}, \mathbf{z} \rangle + \mu\langle \mathbf{y}, \mathbf{z} \rangle. & \text{(A.28)} \end{aligned}$$

The term *non-degenerate* means that if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for all $\mathbf{y}$, then $\mathbf{x} = \mathbf{0}$. Many inner products satisfy the stronger condition of being *positive definite*. This means that $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ unless $\mathbf{x} = \mathbf{0}$, in which case $\langle \mathbf{x}, \mathbf{x} \rangle = 0$. Positive definiteness implies non-degeneracy, but not *vice-versa*.

Given a basis $\mathbf{e}_\mu$, we can form the pairwise products

$$\langle \mathbf{e}_\mu, \mathbf{e}_\nu \rangle = g_{\mu\nu}. \tag{A.29}$$

If the array of numbers $g_{\mu\nu}$ constituting the components of the *metric tensor* turns out to be $g_{\mu\nu} = \delta_{\mu\nu}$, then we say that the basis is *orthonormal* with respect to the inner product. We will not assume orthonormality without specifically saying so. The non-degeneracy of the inner product guarantees the existence of a matrix $g^{\mu\nu}$ which is the inverse of $g_{\mu\nu}$, *i.e.* $g_{\mu\nu}g^{\nu\lambda} = \delta_\mu^\lambda$.

If we take our field to be the real numbers $\mathbb{R}$ then the additional structure provided by a non-degenerate inner product allows us to identify V with V^*. For any $f \in V^*$ we can find a vector $\mathbf{f} \in V$ such that

$$f(\mathbf{x}) = \langle \mathbf{f}, \mathbf{x} \rangle. \tag{A.30}$$

In components, we solve the equation

$$f_\mu = g_{\mu\nu} f^\nu \tag{A.31}$$

for f^ν. We find $f^\nu = g^{\nu\mu} f_\mu$. Usually, we simply identify f with $\mathbf{f}$, and hence V with V^*. We say that the *covariant components* f_μ are related to the *contravariant components* f^μ by *raising*

$$f^\mu = g^{\mu\nu} f_\nu, \tag{A.32}$$

or *lowering*

$$f_\mu = g_{\mu\nu} f^\nu, \tag{A.33}$$

the indices using the metric tensor. Obviously, this identification depends crucially on the inner product; a different inner product would, in general, identify an $f \in V^*$ with a completely different $\mathbf{f} \in V$.

A.3.2 Euclidean vectors

Consider $\mathbb{R}^n$ equipped with its Euclidean metric and associated "dot" inner product. Given a vector $\mathbf{x}$ and a basis $\mathbf{e}_\mu$ with $g_{\mu\nu} = \mathbf{e}_\mu \cdot \mathbf{e}_\nu$, we can define two sets of components for the same vector. Firstly the coefficients x^μ appearing in the basis expansion

$$\mathbf{x} = x^\mu \mathbf{e}_\mu,$$

and secondly the "components"

$$x_\mu = \mathbf{x} \cdot \mathbf{e}_\mu = g_{\mu\nu} x^\nu,$$

of $\mathbf{x}$ along the basis vectors. The x_μ are obtained from the x^μ by the same "lowering" operation as before, and so x^μ and x_μ are naturally referred to as the contravariant and covariant components, respectively, of the vector $\mathbf{x}$. When the $\mathbf{e}_\mu$ constitute an orthonormal basis, then $g_{\mu\nu} = \delta_{\mu\nu}$ and the two sets of components are numerically coincident.

A.3.3 Bra and ket vectors

When our vector space is over the field of complex numbers, the anti-linearity of the first slot of the inner product means we can no longer make a simple identification of V with V^*. Instead there is an *anti-linear* corresponence between the two spaces. The vector $\mathbf{x} \in V$ is mapped to $\langle \mathbf{x}, \ \rangle$ which, since it returns a number when a vector is inserted into its vacant slot, is an element of V^*. This mapping is anti-linear because

$$\lambda \mathbf{x} + \mu \mathbf{y} \mapsto \langle \lambda \mathbf{x} + \mu \mathbf{y}, \ \rangle = \lambda^* \langle \mathbf{x}, \ \rangle + \mu^* \langle \mathbf{y}, \ \rangle. \tag{A.34}$$

This antilinear map is probably familiar to you from quantum mechanics, where V is the space of Dirac's "ket" vectors $|\psi\rangle$ and V^* the space of "bra" vectors $\langle\psi|$. The symbol, here ψ, in each of these objects is a label distinguishing one state-vector from another. We often use the eigenvalues of some complete set set of commuting operators. To each vector $|\psi\rangle$ we use the $(\ldots)^\dagger$ map to assign it a dual vector

$$|\psi\rangle \mapsto |\psi\rangle^\dagger \ \equiv \ \langle\psi|$$

having the same labels. The dagger map is defined to be antilinear

$$(\lambda|\psi\rangle + \mu|\chi\rangle)^\dagger = \lambda^* \langle\psi| + \mu^* \langle\chi|, \tag{A.35}$$

and Dirac denoted the number resulting from the pairing of the covector $\langle\psi|$ with the vector $|\chi\rangle$ by the "bra-c-ket" symbol $\langle\psi|\chi\rangle$:

$$\langle\psi|\chi\rangle \stackrel{\text{def}}{=} (\langle\psi|, |\chi\rangle). \tag{A.36}$$

We can regard the dagger map as either determining the inner-product on V *via*

$$\langle|\psi\rangle, |\chi\rangle\rangle \stackrel{\text{def}}{=} (|\psi\rangle^\dagger, |\chi\rangle) = (\langle\psi|, |\chi\rangle) \equiv \langle\psi|\chi\rangle, \tag{A.37}$$

or being determined by it as

$$|\psi\rangle^\dagger \stackrel{\text{def}}{=} \langle|\psi\rangle, \ \ \rangle \equiv \langle\psi|. \tag{A.38}$$

When we represent our vectors by their components with respect to an orthonormal basis, the dagger map is the familiar operation of taking the conjugate transpose,

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^\dagger = (x_1^*, x_2^*, \ldots, x_n^*) \tag{A.39}$$

but this is not true in general. In a non-orthogonal basis the column vector with components x^μ is mapped to the row vector with components $(x^\dagger)_\mu = (x^\nu)^* g_{\nu\mu}$.

Much of Dirac notation tacitly assumes an orthonormal basis. For example, in the expansion

$$|\psi\rangle = \sum_n |n\rangle\langle n|\psi\rangle \tag{A.40}$$

the expansion coefficients $\langle n|\psi\rangle$ should be the *contravariant* components of $|\psi\rangle$, but the $\langle n|\psi\rangle$ have been obtained from the inner product, and so are in fact its *covariant* components. The expansion (A.40) is therefore valid only when the $|n\rangle$ constitute an orthonormal basis. This will always be the case when the labels on the states show them to be the eigenvectors of a complete commuting set of observables, but sometimes, for example, we may use the integer "n" to refer to an orbital centered on a particular atom in a crystal, and then $\langle n|m\rangle \neq \delta_{mn}$. When using such a non-orthonormal basis it is safer not to use Dirac notation.

Conjugate operator

A linear map $A : V \to W$ automatically induces a map $A^* : W^* \to V^*$. Given $f \in W^*$ we can evaluate $f(A(\mathbf{x}))$ for any $\mathbf{x}$ in V, and so $f(A(\ \ ))$ is an element of V^* that we may denote by $A^*(f)$. Thus,

$$A^*(f)(\mathbf{x}) = f(A(\mathbf{x})). \tag{A.41}$$

Functional analysts (people who spend their working day in Banach space) call A^* the *conjugate* of A. The word "conjugate" and the symbol A^* is rather unfortunate as it has the potential for generating confusion[4] — not least because the $(\ldots)^*$ map is *linear*. No complex conjugation is involved. Thus

$$(\lambda A + \mu B)^* = \lambda A^* + \mu B^*. \tag{A.42}$$

Dirac deftly sidesteps this notational problem by writing $\langle\psi|A$ for the action of the conjugate of the operator $A : V \to V$ on the bra vector $\langle\psi| \in V^*$. After setting $f \to \langle\psi|$ and $\mathbf{x} \to |\chi\rangle$, equation (A.41) therefore reads

$$(\langle\psi|A)\,|\chi\rangle = \langle\psi|\,(A|\chi\rangle)\,. \tag{A.43}$$

This shows that it does not matter where we place the parentheses, so Dirac simply drops them and uses one symbol $\langle\psi|A|\chi\rangle$ to represent both sides of (A.43). Dirac notation thus avoids the non-complex-conjugating "$*$" by suppressing the distinction between an operator and its conjugate. If, therefore, for some reason we need to make the distinction, we cannnot use Dirac notation.

Exercise A.3: If $A : V \to V$ and $B : V \to V$ show that $(AB)^* = B^*A^*$.

Exercise A.4: How does the reversal of the operator order in the previous exercise manifest itself in Dirac notation?

Exercise A.5: Show that if the linear operator A is, in a basis $\mathbf{e}_\mu$, represented by the matrix $\mathbf{A}$, then the conjugate operator A^* is represented in the dual basis $\mathbf{e}^{*\mu}$ by the transposed matrix $\mathbf{A}^T$.

[4]The terms *dual*, *transpose*, or *adjoint* are sometimes used in place of "conjugate." Each of these words brings its own capacity for confusion.

A.3.4 Adjoint operator

The "conjugate" operator of the previous section does not require an inner product for its definition, and is a map from V^* to V^*. When we do have an inner product, however, we can use it to define a different operator "conjugate" to A that, like A itself, is a map from V to V. This new conjugate is called the *adjoint* or the *Hermitian conjugate* of A. To construct it, we first remind ourselves that for any linear map $f : V \to \mathbb{C}$, there is a vector $\mathbf{f} \in V$ such that $f(\mathbf{x}) = \langle \mathbf{f}, \mathbf{x} \rangle$. (To find it we simply solve $f_\nu = (f^\mu)^* g_{\mu\nu}$ for f^μ.) We next observe that $\mathbf{x} \mapsto \langle \mathbf{y}, A\mathbf{x} \rangle$ is such a linear map, and so there is a $\mathbf{z}$ such that $\langle \mathbf{y}, A\mathbf{x} \rangle = \langle \mathbf{z}, \mathbf{x} \rangle$. It should be clear that $\mathbf{z}$ depends linearly on $\mathbf{y}$, so we may define the adjoint linear map, $A^\dagger$, by setting $A^\dagger \mathbf{y} = \mathbf{z}$. This gives us the identity

$$\boxed{\langle \mathbf{y}, A\mathbf{x} \rangle = \langle A^\dagger \mathbf{y}, \mathbf{x} \rangle}$$

The correspondence $A \mapsto A^\dagger$ is anti-linear

$$(\lambda A + \mu B)^\dagger = \lambda^* A^\dagger + \mu^* B^\dagger. \tag{A.44}$$

The adjoint of A depends on the inner product being used to define it. Different inner products give different $A^\dagger$'s.

In the particular case that our chosen basis $\mathbf{e}_\mu$ is orthonormal with respect to the inner product, *i.e.*

$$\langle \mathbf{e}_\mu, \mathbf{e}_\nu \rangle = \delta_{mu\nu}, \tag{A.45}$$

then the Hermitian conjugate $A^\dagger$ of the operator A is represented by the Hermitian conjugate matrix $\mathbf{A}^\dagger$ which is obtained from the matrix $\mathbf{A}$ by interchanging rows and columns and complex conjugating the entries.

Exercise A.6: Show that $(AB)^\dagger = B^\dagger A^\dagger$.

Exercise A.7: When the basis is not orthonormal, show that

$$(A^\dagger)^\rho{}_\sigma = (g_{\sigma\mu} A^\mu{}_\nu g^{\nu\rho})^* \,. \tag{A.46}$$

A.4 Sums and differences of vector spaces

A.4.1 Direct sums

Suppose that U and V are vector spaces. We define their *direct sum* $U \oplus V$ to be the vector space of ordered pairs $(\mathbf{u}, \mathbf{v})$ with

$$\lambda(\mathbf{u}_1, \mathbf{v}_1) + \mu(\mathbf{u}_2, \mathbf{v}_2) = (\lambda \mathbf{u}_1 + \mu \mathbf{u}_2, \lambda \mathbf{v}_1 + \mu \mathbf{v}_2). \tag{A.47}$$

The set of vectors $\{(\mathbf{u}, 0)\} \subset U \oplus V$ forms a copy of U, and $\{(0, \mathbf{v})\} \subset U \oplus V$ a copy of V. Thus U and V may be regarded as subspaces of $U \oplus V$.

If U and V are any pair of subspaces of W, we can form the space $U + V$ consisting of all elements of W that can be written as $\mathbf{u} + \mathbf{v}$ with $\mathbf{u} \in U$ and $\mathbf{v} \in V$. The decomposition $\mathbf{x} = \mathbf{u} + \mathbf{v}$ of an element $\mathbf{x} \in U + V$ into parts in U and V will be unique (in that $\mathbf{u}_1 + \mathbf{v}_1 = \mathbf{u}_2 + \mathbf{v}_2$ implies that $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{v}_1 = \mathbf{v}_2$) if and only if $U \cap V = \{\mathbf{0}\}$ where $\{\mathbf{0}\}$ is the subspace containing only the zero vector. In this case $U + V$ can be identified with $U \oplus V$.

If U is a subspace of W then we can seek a *complementary space* V such that $W = U \oplus V$, or, equivalently, $W = U + V$ with $U \cap V = \{\mathbf{0}\}$. Such complementary spaces are *not* unique. Consider $\mathbb{R}^3$, for example, with U being the vectors in the x, y plane. If $\mathbf{e}$ is any vector that does not lie in this plane then the one-dimensional space spanned by $\mathbf{e}$ is a complementary space for U.

A.4.2 Quotient spaces

We have seen that if U is a subspace of W there are many complementary subspaces V such that $W = U \oplus V$. We can however define a *unique* space that we might denote by $W - U$ and refer to as the difference of the two spaces. It is more common, however, to see this space written as W/U and referred to as the *quotient* of W *modulo* U. This quotient space is the vector space of *equivalence classes* of vectors, where we do not distinguish between two vectors in W if their difference lies in U. In other words

$$\mathbf{x} = \mathbf{y} \pmod U \quad \Leftrightarrow \quad \mathbf{x} - \mathbf{y} \in U. \tag{A.48}$$

The collection of elements in W that are equivalent to $\mathbf{x} \, (\mathrm{mod}\, U)$ composes a *coset*, written $\mathbf{x} + U$, a set whose elements are $\mathbf{x} + \mathbf{u}$ where $\mathbf{u}$ is any vector in U. These cosets are the elements of W/U.

When we have a linear map $A : U \to V$, the quotient space $V/\mathrm{Im}\, A$ is often called the *co-kernel* of A.

Given a positive-definite inner product, we can define a unique *orthogonal complement* of $U \subset W$. We define $U^\perp$ to be the set

$$U^\perp = \{\mathbf{x} \in W : \langle \mathbf{x}, \mathbf{y} \rangle = 0, \ \forall \mathbf{y} \in U\}. \tag{A.49}$$

It is easy to see that this is a linear subspace and that $U \oplus U^\perp = W$. For finite dimensional spaces

$$\boxed{\dim W/U = \dim U^\perp = \dim W - \dim U}$$

and $(U^\perp)^\perp = U$. For infinite dimensional spaces we only have $(U^\perp)^\perp \supseteq U$. (Be careful, however. If the inner product is *not* positive definite, U and $U^\perp$ may have non-zero vectors in common.)

Although they have the same dimensions, do not confuse W/U with $U^\perp$, and in particular do not use the phrase *orthogonal complement* without specifying an inner product.

A practical example of a quotient space occurs in digital imaging. A colour camera reduces the infinite-dimensional space $\mathcal{L}$ of coloured light incident on each pixel to three numbers, R, G and B, these obtained by pairing the spectral intensity with the frequency response (an element of $\mathcal{L}^*$) of the red, green and blue detectors at that point. The space of distingushable colours is therefore only three dimensional. Many different incident spectra will give the same output RGB signal, and are therefore equivalent as far as the camera is concerned. In the colour industry these equivalent colours are called *metamers*. Equivalent colours differ by spectral intensities that lie in the space $\mathcal{B}$ of *metameric black*. There is no inner product here, so it is meaningless to think of the space of distinguishable colours as being $\mathcal{B}^\perp$. It is, however, precisely what we mean by $\mathcal{L}/\mathcal{B}$.

A.4.3 Projection-operator decompositions

An operator $P : V \to V$ that obeys $P^2 = P$ is called a *projection operator*. It projects a vector $\mathbf{x} \in V$ to $P\mathbf{x} \in \operatorname{Im} P$ *along* $\operatorname{Ker} P$ — in the sense of casting a shadow onto $\operatorname{Im} P$ with the light coming from the direction $\operatorname{Ker} P$. In other words all vectors lying in $\operatorname{Ker} P$ are killed, whilst any vector already in $\operatorname{Im} P$ is left alone by P. (If $\mathbf{x} \in \operatorname{Im} P$ then $\mathbf{x} = P\mathbf{y}$ for some $\mathbf{y} \in V$, and $P\mathbf{x} = P^2\mathbf{y} = P\mathbf{y} = \mathbf{x}$.) The only vector common to both $\operatorname{Ker} P$ and $\operatorname{Im} P$ is $\mathbf{0}$, and so

$$V = \operatorname{Ker} P \oplus \operatorname{Im} P. \tag{A.50}$$

A set of projection operators P_i that are "orthogonal"

$$P_i P_j = \delta_{ij} P_i, \tag{A.51}$$

and sum to the identity operator

$$\sum_i P_i = I, \tag{A.52}$$

is called a *resolution of the identity*. The resulting equation

$$\mathbf{x} = \sum_i P_i \mathbf{x} \tag{A.53}$$

decomposes $\mathbf{x}$ uniquely into a sum of terms $P_i\mathbf{x} \in \operatorname{Im} P_i$ and so decomposes V into a direct sum of subspaces $V_i \equiv \operatorname{Im} P_i$:

$$V = \bigoplus_i V_i. \tag{A.54}$$

Exercise A.8: Let P_1 be a projection operator. Show that $P_2 = I - P_1$ is also a projection operator and $P_1P_2 = 0$. Show also that $\operatorname{Im} P_2 = \operatorname{Ker} P_1$ and $\operatorname{Ker} P_2 = \operatorname{Im} P_1$.

A.5 Inhomogeneous linear equations

Suppose we wish to solve the system of linear equations

$$\begin{aligned} a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n &= b_1 \\ a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n &= b_2 \\ &\vdots \\ a_{m1}y_1 + a_{m2}y_2 + \cdots + a_{mn}y_n &= b_m \end{aligned}$$

or, in matrix notation,

$$\mathbf{A}\mathbf{y} = \mathbf{b}, \tag{A.55}$$

where $\mathbf{A}$ is the m-by-n matrix with entries a_{ij}. Faced with such a problem, we should start by asking ourselves the questions:

i) Does a solution exist?
ii) If a solution does exist, is it unique?

These issues are best addressed by considering the matrix $\mathbf{A}$ as a linear operator $A: V \to W$, where V is n dimensional and W is m dimensional. The natural language is then that of the range and nullspaces of A. There is no solution to the equation $\mathbf{A}\mathbf{y} = \mathbf{b}$ when $\operatorname{Im} A$ is not the whole of W and $\mathbf{b}$ does not lie in $\operatorname{Im} A$. Similarly, the solution will not be unique if there are distinct vectors $\mathbf{x}_1$, $\mathbf{x}_2$ such that $A\mathbf{x}_1 = A\mathbf{x}_2$. This means that $A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, or $(\mathbf{x}_1 - \mathbf{x}_2) \in \operatorname{Ker} A$. These situations are linked, as we have seen, by the range null-space theorem:

$$\dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A) = \dim V. \tag{A.56}$$

Thus, if $m > n$ there are bound to be some vectors $\mathbf{b}$ for which no solution exists. When $m < n$ the solution cannot be unique.

A.5.1 Rank and index

Suppose $V \equiv W$ (so $m = n$ and the matrix is square) and we chose an inner product, $\langle \mathbf{x}, \mathbf{y} \rangle$, on V. Then $\mathbf{x} \in \operatorname{Ker} A$ implies that, for all $\mathbf{y}$

$$0 = \langle \mathbf{y}, A\mathbf{x} \rangle = \langle A^\dagger \mathbf{y}, \mathbf{x} \rangle, \tag{A.57}$$

or that $\mathbf{x}$ is perpendicular to the range of $A^\dagger$. Conversely, let $\mathbf{x}$ be perpendicular to the range of $A^\dagger$; then

$$\langle \mathbf{x}, A^\dagger \mathbf{y} \rangle = 0, \quad \forall \mathbf{y} \in V, \tag{A.58}$$

which means that

$$\langle A\mathbf{x}, \mathbf{y} \rangle = 0, \quad \forall \mathbf{y} \in V, \tag{A.59}$$

and, by the non-degeneracy of the inner product, this means that $A\mathbf{x} = \mathbf{0}$. The net result is that

$$\operatorname{Ker} A = (\operatorname{Im} A^\dagger)^\perp. \tag{A.60}$$

Similarly

$$\operatorname{Ker} A^\dagger = (\operatorname{Im} A)^\perp. \tag{A.61}$$

Now

$$\begin{aligned} \dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A) &= \dim V, \\ \dim(\operatorname{Ker} A^\dagger) + \dim(\operatorname{Im} A^\dagger) &= \dim V, \end{aligned} \tag{A.62}$$

but

$$\begin{aligned} \dim(\operatorname{Ker} A) &= \dim(\operatorname{Im} A^\dagger)^\perp \\ &= \dim V - \dim(\operatorname{Im} A^\dagger) \\ &= \dim(\operatorname{Ker} A^\dagger). \end{aligned}$$

Thus, for finite-dimensional square matrices, we have

$$\boxed{\dim(\operatorname{Ker} A) = \dim(\operatorname{Ker} A^\dagger)}$$

In particular, the row and column rank of a square matrix coincide.

Example: Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 3 & 4 \end{pmatrix}$$

Clearly, the number of linearly independent rows is two, since the third row is the sum of the other two. The number of linearly independent columns is also two — although less obviously so — because

$$-\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}.$$

Warning: The equality $\dim(\mathrm{Ker}\, A) = \dim(\mathrm{Ker}\, A^\dagger)$, need not hold in infinite dimensional spaces. Consider the space with basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \ldots$ indexed by the positive integers. Define $A\mathbf{e}_1 = \mathbf{e}_2$, $A\mathbf{e}_2 = \mathbf{e}_3$, and so on. This operator has $\dim(\mathrm{Ker}\, A) = 0$. The adjoint with respect to the natural inner product has $A^\dagger\mathbf{e}_1 = \mathbf{0}$, $A^\dagger\mathbf{e}_2 = \mathbf{e}_1$, $A^\dagger\mathbf{e}_3 = \mathbf{e}_2$. Thus $\mathrm{Ker}\, A^\dagger = \{\mathbf{e}_1\}$, and $\dim(\mathrm{Ker}\, A^\dagger) = 1$. The difference $\dim(\mathrm{Ker}\, A) - \dim(\mathrm{Ker}\, A^\dagger)$ is called the *index* of the operator. The index of an operator is often related to topological properties of the space on which it acts, and in this way appears in physics as the origin of *anomalies* in quantum field theory.

A.5.2 Fredholm alternative

The results of the previous section can be summarized as saying that the *Fredholm Alternative* holds for finite square matrices. The Fredholm Alternative is the set of statements

I. **Either**

  i) $A\mathbf{x} = \mathbf{b}$ has a *unique* solution,

  or

  ii) $A\mathbf{x} = \mathbf{0}$ has a solution.

II. If $A\mathbf{x} = \mathbf{0}$ has n linearly independent solutions, then so does $A^\dagger\mathbf{x} = \mathbf{0}$.

III. If alternative ii) holds, then $A\mathbf{x} = \mathbf{b}$ has *no* solution unless b is orthogonal to all solutions of $A^\dagger\mathbf{x} = \mathbf{0}$.

It should be obvious that this is a recasting of the statements that

$$\dim(\mathrm{Ker}\, A) = \dim(\mathrm{Ker}\, A^\dagger),$$

and

$$(\mathrm{Ker}\, A^{\dagger})^{\perp} = \mathrm{Im}\, A. \tag{A.63}$$

Notice that finite-dimensionality is essential here. Neither of these statement is guaranteed to be true in infinite dimensional spaces.

A.6 Determinants

A.6.1 Skew-symmetric n-linear forms

You will be familiar with the elementary definition of the determinant of an n-by-n matrix $\mathbf{A}$ having entries a_{ij}:

$$\det \mathbf{A} \equiv \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \stackrel{\text{def}}{=} \epsilon_{i_1 i_2 \dots i_n} a_{1 i_1} a_{2 i_2} \dots a_{n i_n}. \tag{A.64}$$

Here, $\epsilon_{i_1 i_2 \dots i_n}$ is the *Levi-Civita* symbol, which is skew-symmetric in all its indices and $\epsilon_{12\dots n} = 1$. From this definition we see that the determinant changes sign if any pair of its rows are interchanged, and that it is linear in each row. In other words

$$\begin{vmatrix} \lambda a_{11} + \mu b_{11} & \lambda a_{12} + \mu b_{12} & \dots & \lambda a_{1n} + \mu b_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{vmatrix}$$
$$= \lambda \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{vmatrix} + \mu \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{vmatrix}.$$

If we consider each row as being the components of a vector in an n-dimensional vector space V, we may regard the determinant as being a skew-symmetric n-linear form, *i.e.* a map

$$\omega : \overbrace{V \times V \times \dots V}^{n \text{ factors}} \to \mathbb{F} \tag{A.65}$$

which is linear in each slot,

$$\omega(\lambda\mathbf{a}+\mu\mathbf{b},\mathbf{c}_2,\ldots,\mathbf{c}_n)=\lambda\,\omega(\mathbf{a},\mathbf{c}_2,\ldots,\mathbf{c}_n)+\mu\,\omega(\mathbf{b},\mathbf{c}_2,\ldots,\mathbf{c}_n), \tag{A.66}$$

and changes sign when any two arguments are interchanged,

$$\omega(\ldots,\mathbf{a}_i,\ldots,\mathbf{a}_j,\ldots)=-\omega(\ldots,\mathbf{a}_j,\ldots,\mathbf{a}_i,\ldots). \tag{A.67}$$

We will denote the space of skew-symmetric n-linear forms on V by the symbol $\bigwedge^n(V^*)$. Let ω be an arbitrary skew-symmetric n-linear form in $\bigwedge^n(V^*)$, and let $\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$ be a basis for V. If $\mathbf{a}_i=a_{ij}\mathbf{e}_j$ $(i=1,\ldots,n)$ is a collection of n vectors[5], we compute

$$\begin{aligned}\omega(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_n) &= a_{1i_1}a_{2i_2}\ldots a_{ni_n}\omega(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\ldots,\mathbf{e}_{i_n})\\ &= a_{1i_1}a_{2i_2}\ldots a_{ni_n}\epsilon_{i_1i_2\ldots,i_n}\omega(\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n). \end{aligned}\tag{A.68}$$

In the first line we have exploited the linearity of ω in each slot, and in going from the first to the second line we have used skew-symmetry to rearrange the basis vectors in their canonical order. We deduce that all skew-symmetric n-forms are proportional to the determinant

$$\omega(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_n)\propto\begin{vmatrix}a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \ldots & a_{nn}\end{vmatrix},$$

and that the proportionality factor is the number $\omega(\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n)$. When the number of its slots is equal to the dimension of the vector space, there is therefore essentially only *one* skew-symmetric multilinear form and $\bigwedge^n(V^*)$ is a one-dimensional vector space.

Now we use the notion of skew-symmetric n-linear forms to give a powerful definition of the determinant of an *endomorphism*, *i.e.* a linear map $A:V\to V$. Let ω be a non-zero skew-symmetric n-linear form. The object

$$\omega_A(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)\stackrel{\text{def}}{=}\omega(A\mathbf{x}_1,A\mathbf{x}_2,\ldots,A\mathbf{x}_n). \tag{A.69}$$

[5] The index j on a_{ij} should really be a superscript since a_{ij} is the j-th contravariant component of the vector $\mathbf{a}_i$. We are writing it as a subscript only for compatibility with other equations in this section.

is also a skew-symmetric n-linear form. Since there is only one such object up to multiplicative constants, we must have

$$\omega(A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n) \propto \omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n). \tag{A.70}$$

We define "$\det A$" to be the constant of proportionality. Thus

$$\omega(A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_n) = \det(A)\omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n). \tag{A.71}$$

By writing this out in a basis where the linear map A is represented by the matrix $\mathbf{A}$, we easily see that

$$\det \mathbf{A} = \det A. \tag{A.72}$$

The new definition is therefore compatible with the old one. The advantage of this more sophisticated definition is that it makes no appeal to a basis, and so shows that the determinant of an endomorphism is a basis-independent concept. A byproduct is an easy proof that $\det(AB) = \det(A)\det(B)$, a result that is not so easy to establish with the elementary definition. We write

$$\begin{aligned}\det(AB)\omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) &= \omega(AB\mathbf{x}_1, AB\mathbf{x}_2, \ldots, AB\mathbf{x}_n)\\ &= \omega(A(B\mathbf{x}_1), A(B\mathbf{x}_2), \ldots, A(B\mathbf{x}_n))\\ &= \det(A)\omega(B\mathbf{x}_1, B\mathbf{x}_2, \ldots, B\mathbf{x}_n)\\ &= \det(A)\det(B)\omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n).\end{aligned} \tag{A.73}$$

Cancelling the common factor of $\omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ completes the proof.

Exercise A.9: Let ω be a skew-symmetric n-linear form on an n-dimensional vector space. Assuming that ω does not vanish identically, show that a set of n vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is linearly independent, and hence forms a basis, if, and only if, $\omega(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \neq 0$.

Exercise A.10: Extend the paring between V and its dual space V^* to a pairing between the one-dimensional $\bigwedge^n(V^*)$ and *its* dual space. Use this pairing, together with the result of exercise A.5, to show that

$$\det \mathbf{A}^T = \det A^* = [\det A]^* = [\det A]^T = \det A = \det \mathbf{A},$$

where the "$*$" denotes the conjugate operator (and not complex conjugation) and the penultimate equality holds because transposition has no effect on a one-by-one matrix. Conclude that $\det \mathbf{A} = \det \mathbf{A}^T$. A determinant is therefore unaffected by the interchange of its rows with its columns.

Exercise A.11: *Cauchy-Binet formula.* Let $\mathbf{A}$ be a m-by-n matrix and $\mathbf{B}$ be an n-by-m matrix. The matrix product $\mathbf{AB}$ is therefore defined, and is an m-by-m matrix. Let S be a subset of $\{1,\ldots,n\}$ with m elements, and let $\mathbf{A}_S$ be the m-by-m matrix whose columns are the columns of $\mathbf{A}$ corresponding to indices in S. Similarly let $\mathbf{B}_S$ be the m-by-m matrix whose *rows* are the rows of $\mathbf{B}$ with indices in S. Show that

$$\det \mathbf{AB} = \sum_S \det \mathbf{A}_S \det \mathbf{B}_S$$

where the sum is over all $n!/m!(n-m)!$ subsets S. If $m > n$ there there are no such subsets. Show that in this case $\det \mathbf{AB} = 0$.

Exercise A.12: Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix}$$

be a partitioned matrix where $\mathbf{a}$ is m-by-m, $\mathbf{b}$ is m-by-n, $\mathbf{c}$ is n-by-m, and $\mathbf{d}$ is n-by-n. By making a Gaussian decomposition

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_m & \mathbf{x} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix} \begin{pmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{y} & \mathbf{I}_n \end{pmatrix},$$

show that, for invertible $\mathbf{d}$, we have *Schur's determinant formula*[6]

$$\det \mathbf{A} = \det(\mathbf{d}) \det(\mathbf{a} - \mathbf{b}\mathbf{d}^{-1}\mathbf{c}).$$

A.6.2 The adjugate matrix

Given a square matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \tag{A.74}$$

and an element a_{ij}, we define the corresponding *minor* M_{ij} to be the determinant of the $(n-1)$-by-$(n-1)$ matrix constructed by deleting from $\mathbf{A}$ the row and column containing a_{ij}. The number

$$A_{ij} = (-1)^{i+j} M_{ij} \tag{A.75}$$

[6] I. Schur, *J. für reine und angewandte Math.*, **147** (1917) 205-232.

is then called the *co-factor* of the element a_{ij}. (It is traditional to use uppercase letters to denote co-factors.) The basic result involving co-factors is that

$$\sum_j a_{ij}A_{i'j} = \delta_{ii'}\det \mathbf{A}. \tag{A.76}$$

When $i = i'$, this is known as the *Laplace development* of the determinant about row i. We get zero when $i \neq i'$ because we are effectively developing a determinant with two equal rows. We now define the *adjugate matrix*,[7] $\mathrm{Adj}\,\mathbf{A}$, to be the transposed matrix of the co-factors:

$$(\mathrm{Adj}\,\mathbf{A})_{ij} = A_{ji}. \tag{A.77}$$

In terms of this we have

$$\mathbf{A}(\mathrm{Adj}\,\mathbf{A}) = (\det \mathbf{A})\mathbf{I}. \tag{A.78}$$

In other words

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}\mathrm{Adj}\,\mathbf{A}. \tag{A.79}$$

Each entry in the adjugate matrix is a polynomial of degree $n-1$ in the entries of the original matrix. Thus, no division is required to form it, and the adjugate matrix exists even if the inverse matrix does not.

Exercise A.13: It is possible to Laplace-develop a determinant about a *set* of rows. For example, the development of a 4-by-4 determinant about its first *two* rows is given by:

$$\begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}\begin{vmatrix} c_3 & d_3 \\ c_4 & d_4 \end{vmatrix} - \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}\begin{vmatrix} b_3 & d_3 \\ b_4 & d_4 \end{vmatrix} + \begin{vmatrix} a_1 & d_1 \\ a_2 & d_2 \end{vmatrix}\begin{vmatrix} b_3 & c_3 \\ b_4 & c_4 \end{vmatrix}$$
$$+ \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}\begin{vmatrix} a_3 & d_3 \\ a_4 & d_4 \end{vmatrix} - \begin{vmatrix} b_1 & d_1 \\ b_2 & d_2 \end{vmatrix}\begin{vmatrix} a_3 & c_3 \\ a_4 & c_4 \end{vmatrix} + \begin{vmatrix} c_1 & d_1 \\ c_2 & d_2 \end{vmatrix}\begin{vmatrix} a_3 & b_3 \\ a_4 & b_4 \end{vmatrix}$$

Understand why this formula is correct, and, using that insight, describe the general rule.

[7]Some authors rather confusingly call this the *adjoint matrix*.

Exercise A.14: *Sylvester's Lemma.*[8] Let **A** and **B** be two n-by-n matrices. Show that

$$\det \mathbf{A} \det \mathbf{B} = \sum \det \mathbf{A}' \det \mathbf{B}'$$

where $\mathbf{A}'$ and $\mathbf{B}'$ are constructed by selecting a fixed set of $k < n$ columns of **B** (which we can, without loss of generality, take to to be the first k columns) and interchanging them with k columns of **A**, preserving the order of the columns. The sum is over all $n!/k!(n-k)!$ ways of choosing columns of **A**. (Hint: Show that, without loss of generality, we can take the columns of **A** to be a set of basis vectors, and that, in this case, the lemma becomes a re-statement of your "general rule" from the previous problem.)

Cayley's theorem

You will know that the possible eigenvalues of the n-by-n matrix **A** are given by the roots of its *characteristic equation*

$$0 = \det(\mathbf{A} - \lambda\mathbf{I}) = (-1)^n \left(\lambda^n - \operatorname{tr}(\mathbf{A})\lambda^{n-1} + \cdots + (-1)^n \det(\mathbf{A})\right), \quad \text{(A.80)}$$

and have probably met with *Cayley's theorem* that asserts that every matrix obeys its own characteristic equation.

$$\mathbf{A}^n - \operatorname{tr}(\mathbf{A})\mathbf{A}^{n-1} + \cdots + (-1)^n \det(\mathbf{A})\mathbf{I} = \mathbf{0}. \quad \text{(A.81)}$$

The proof of Cayley's theorem involves the adjugate matrix. We write

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (-1)^n \left(\lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n\right) \quad \text{(A.82)}$$

and observe that

$$\det(\mathbf{A} - \lambda\mathbf{I})\mathbf{I} = (\mathbf{A} - \lambda\mathbf{I})\operatorname{Adj}(\mathbf{A} - \lambda\mathbf{I}). \quad \text{(A.83)}$$

Now $\operatorname{Adj}(\mathbf{A} - \lambda\mathbf{I})$ is a matrix-valued polynomial in λ of degree $n-1$, and it can be written

$$\operatorname{Adj}(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{C}_0\lambda^{n-1} + \mathbf{C}_1\lambda^{n-2} + \cdots + \mathbf{C}_{n-1}, \quad \text{(A.84)}$$

for some matrix coefficients $\mathbf{C}_i$. On multiplying out the equation

$$(-1)^n \left(\lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n\right)\mathbf{I} = (\mathbf{A} - \lambda\mathbf{I})(\mathbf{C}_0\lambda^{n-1} + \mathbf{C}_1\lambda^{n-2} + \cdots + \mathbf{C}_{n-1}) \quad \text{(A.85)}$$

[8] J. J. Sylvester, *Phil. Mag.* **1** (1851) 295–305.

and comparing like powers of λ, we find the relations

$$\begin{aligned}
(-1)^n\mathbf{I} &= -\mathbf{C}_0, \\
(-1)^n\alpha_1\mathbf{I} &= -\mathbf{C}_1 + \mathbf{A}\mathbf{C}_0, \\
(-1)^n\alpha_2\mathbf{I} &= -\mathbf{C}_2 + \mathbf{A}\mathbf{C}_1, \\
&\vdots \\
(-1)^n\alpha_{n-1}\mathbf{I} &= -\mathbf{C}_{n-1} + \mathbf{A}\mathbf{C}_{n-2}, \\
(-1)^n\alpha_n\mathbf{I} &= \quad \mathbf{A}\mathbf{C}_{n-1}.
\end{aligned}$$

Multiply the first equation on the left by $\mathbf{A}^n$, the second by $\mathbf{A}^{n-1}$, and so on down the last equation which we multiply by $\mathbf{A}^0 \equiv \mathbf{I}$. Now add. We find that the sum telescopes to give Cayley's theorem,

$$\mathbf{A}^n + \alpha_1\mathbf{A}^{n-1} + \cdots + \alpha_n\mathbf{I} = \mathbf{0},$$

as advertised.

A.6.3 Differentiating determinants

Suppose that the elements of $\mathbf{A}$ depend on some parameter x. From the elementary definition

$$\det \mathbf{A} = \epsilon_{i_1 i_2 \ldots i_n} a_{1i_1} a_{2i_2} \ldots a_{ni_n},$$

we find

$$\frac{d}{dx}\det \mathbf{A} = \epsilon_{i_1 i_2 \ldots i_n}\left(a'_{1i_1} a_{2i_2} \ldots a_{ni_n} + a_{1i_1} a'_{2i_2} \ldots a_{ni_n} + \cdots + a_{1i_1} a_{2i_2} \ldots a'_{ni_n}\right). \tag{A.86}$$

In other words,

$$\frac{d}{dx}\det \mathbf{A} = \begin{vmatrix} a'_{11} & a'_{12} & \ldots & a'_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a'_{21} & a'_{22} & \ldots & a'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} + \cdots + \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{n1} & a'_{n2} & \ldots & a'_{nn} \end{vmatrix}.$$

The same result can also be written more compactly as

$$\frac{d}{dx}\det \mathbf{A} = \sum_{ij} \frac{da_{ij}}{dx} A_{ij}, \tag{A.87}$$

where A_{ij} is cofactor of a_{ij}. Using the connection between the adjugate matrix and the inverse, this is equivalent to

$$\frac{1}{\det \mathbf{A}} \frac{d}{dx} \det \mathbf{A} = \operatorname{tr} \left\{ \frac{d\mathbf{A}}{dx} \mathbf{A}^{-1} \right\}, \tag{A.88}$$

or

$$\frac{d}{dx} \ln (\det \mathbf{A}) = \operatorname{tr} \left\{ \frac{d\mathbf{A}}{dx} \mathbf{A}^{-1} \right\}. \tag{A.89}$$

A special case of this formula is the result

$$\frac{\partial}{\partial a_{ij}} \ln (\det \mathbf{A}) = \left(\mathbf{A}^{-1} \right)_{ji}. \tag{A.90}$$

A.7 Diagonalization and canonical forms

An essential part of the linear algebra tool-kit is the set of techniques for the reduction of a matrix to its simplest, *canonical form*. This is often a diagonal matrix.

A.7.1 Diagonalizing linear maps

A common task is the diagonalization of a matrix $\mathbf{A}$ representing a linear map A. Let us recall some standard material relating to this:

i) If $A\mathbf{x} = \lambda \mathbf{x}$ for a non-zero vector $\mathbf{x}$, then $\mathbf{x}$ is said to be an *eigenvector* of A with *eigenvalue* λ.

ii) A linear operator A on a finite-dimensional vector space is said to be *self-adjoint*, or *Hermitian*, with respect to the inner product $\langle\ ,\ \rangle$ if $A = A^\dagger$, or equivalently if $\langle \mathbf{x}, A\mathbf{y} \rangle = \langle A\mathbf{x}, \mathbf{y} \rangle$ for all $\mathbf{x}$ and $\mathbf{y}$.

iii) If A is Hermitian with respect to a positive definite inner product $\langle\ ,\ \rangle$ then all the eigenvalues λ are real. To see that this is so, we write

$$\lambda \langle \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{x}, \lambda \mathbf{x} \rangle = \langle \mathbf{x}, A\mathbf{x} \rangle = \langle A\mathbf{x}, \mathbf{x} \rangle = \langle \lambda \mathbf{x}, \mathbf{x} \rangle = \lambda^* \langle \mathbf{x}, \mathbf{x} \rangle. \tag{A.91}$$

Because the inner product is positive definite and $\mathbf{x}$ is not zero, the factor $\langle \mathbf{x}, \mathbf{x} \rangle$ cannot be zero. We conclude that $\lambda = \lambda^*$.

iii) If A is Hermitian and λ_i and λ_j are two distinct eigenvalues with eigenvectors $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively, then $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$. To prove this, we write

$$\lambda_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle \mathbf{x}_i, A\mathbf{x}_j \rangle = \langle A\mathbf{x}_i, \mathbf{x}_j \rangle = \langle \lambda_i \mathbf{x}_i, \mathbf{x}_j \rangle = \lambda_i^* \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \tag{A.92}$$

But $\lambda_i^* = \lambda_i$, and so $(\lambda_i - \lambda_j)\langle \mathbf{x}_i, \mathbf{x}_j\rangle = 0$. Since, by assumption, $(\lambda_i - \lambda_j) \neq 0$ we must have $\langle \mathbf{x}_i, \mathbf{x}_j\rangle = 0$.

iv) An operator A is said to be *diagonalizable* if we can find a basis for V that consists of eigenvectors of A. In this basis, A is represented by the matrix $\mathbf{A} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where the λ_i are the eigenvalues.

Not all linear operators can be diagonalized. The key element determining the diagonalizability of a matrix is the *minimal polynomial equation* obeyed by the matrix representing the operator. As mentioned in the previous section, the possible eigenvalues an N-by-N matrix $\mathbf{A}$ are given by the roots of the *characteristic equation*

$$0 = \det\,(\mathbf{A} - \lambda\mathbf{I}) = (-1)^N \left(\lambda^N - \mathrm{tr}\,(\mathbf{A})\lambda^{N-1} + \cdots + (-1)^N \det\,(\mathbf{A})\right).$$

This is because a non-trivial solution to the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{A.93}$$

requires the matrix $\mathbf{A}-\lambda\mathbf{I}$ to have a non-trivial nullspace, and so $\det\,(\mathbf{A} - \lambda\mathbf{I})$ must vanish. Cayley's Theorem, which we proved in the previous section, asserts that every matrix obeys its own characteristic equation:

$$\mathbf{A}^N - \mathrm{tr}\,(\mathbf{A})\mathbf{A}^{N-1} + \cdots + (-1)^N \det\,(\mathbf{A})\mathbf{I} = \mathbf{0}.$$

The matrix $\mathbf{A}$ may, however, satisfy an equation of lower degree. For example, the characteristic equation of the matrix

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix} \tag{A.94}$$

is $(\lambda - \lambda_1)^2$. Cayley therefore asserts that $(\mathbf{A} - \lambda_1\mathbf{I})^2 = \mathbf{0}$. This is clearly true, but $\mathbf{A}$ also satisfies the equation of first degree $(\mathbf{A} - \lambda_1\mathbf{I}) = \mathbf{0}$.

The equation of lowest degree satisfied by $\mathbf{A}$ is said to be the minimal polynomial equation. It is unique up to an overall numerical factor: if two distinct minimal equations of degree n were to exist, and if we normalize them so that the coefficients of $\mathbf{A}^n$ coincide, then their difference, if non-zero, would be an equation of degree $\leq (n-1)$ obeyed by $\mathbf{A}$ — and a contradiction to the minimal equation having degree n.

If

$$P(\mathbf{A}) \equiv (\mathbf{A} - \lambda_1\mathbf{I})^{\alpha_1}(\mathbf{A} - \lambda_2\mathbf{I})^{\alpha_2} \cdots (\mathbf{A} - \lambda_n\mathbf{I})^{\alpha_n} = \mathbf{0} \tag{A.95}$$

is the minimal equation then each root λ_i is an eigenvalue of $\mathbf{A}$. To prove this, we select one factor of $(A - \lambda_i \mathbf{I})$ and write

$$P(\mathbf{A}) = (A - \lambda_i \mathbf{I})Q(\mathbf{A}), \tag{A.96}$$

where $Q(\mathbf{A})$ contains all the remaining factors in $P(\mathbf{A})$. We now observe that there must be some vector $\mathbf{y}$ such that $\mathbf{x} = Q(\mathbf{A})\mathbf{y}$ is not zero. If there were no such $\mathbf{y}$ then $Q(\mathbf{A}) = \mathbf{0}$ would be an equation of lower degree obeyed by $\mathbf{A}$ in contradiction to the assumed minimality of $P(\mathbf{A})$. Since

$$\mathbf{0} = P(\mathbf{A})\mathbf{y} = (\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x} \tag{A.97}$$

we see that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eignvalue λ_i.

Because all possible eigenvalues appear as roots of the characteristic equation, the minimal equation must have the same roots as the characteristic equation, but with equal or lower multiplicities α_i.

In the special case that $\mathbf{A}$ is self-adjoint, or Hermitian, with respect to a positive definite inner product $\langle\ ,\ \rangle$ the minimal equation has no repeated roots. Suppose that this were not so, and that $\mathbf{A}$ has minimal equation $(\mathbf{A} - \lambda\mathbf{I})^2 R(\mathbf{A}) = \mathbf{0}$ where $R(\mathbf{A})$ is a polynomial in $\mathbf{A}$. Then, for all vectors $\mathbf{x}$ we have

$$0 = \langle R\mathbf{x}, (\mathbf{A} - \lambda\mathbf{I})^2 R\mathbf{x}\rangle = \langle (\mathbf{A} - \lambda\mathbf{I})R\mathbf{x}, (\mathbf{A} - \lambda\mathbf{I})R\mathbf{x}\rangle. \tag{A.98}$$

Now the vanishing of the rightmost expression shows that $(\mathbf{A} - \lambda\mathbf{I})R(\mathbf{A})\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$. In other words

$$(\mathbf{A} - \lambda\mathbf{I})R(\mathbf{A}) = \mathbf{0}. \tag{A.99}$$

The equation with the repeated factor was not minimal therefore, and we have a contradiction.

If the equation of lowest degree satisfied by the matrix has no repeated roots, the matrix is diagonalizable; if there are repeated roots, it is not. The last statement should be obvious, because a diagonalized matrix satisfies an equation with no repeated roots, and this equation will hold in all bases, including the original one. The first statement, in combination with with the observation that the minimal equation for a Hermitian matrix has no repeated roots, shows that a Hermitian (with respect to a positive definite inner product) matrix can be diagonalized.

To establish the first statement, suppose that $\mathbf{A}$ obeys the equation

$$\mathbf{0} = P(\mathbf{A}) \equiv (\mathbf{A}-\lambda_1\mathbf{I})(\mathbf{A}-\lambda_2\mathbf{I})\cdots(\mathbf{A}-\lambda_n\mathbf{I}), \tag{A.100}$$

where the λ_i are all distinct. Then, setting $x \to \mathbf{A}$ in the identity[9]

$$\begin{aligned} 1 \;=\; & \frac{(x-\lambda_2)(x-\lambda_3)\cdots(x-\lambda_n)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)\cdots(\lambda_1-\lambda_n)} + \frac{(x-\lambda_1)(x-\lambda_3)\cdots(x-\lambda_n)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)\cdots(\lambda_2-\lambda_n)} + \cdots \\ & + \frac{(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_{n-1})}{(\lambda_n-\lambda_1)(\lambda_n-\lambda_2)\cdots(\lambda_n-\lambda_{n-1})}, \end{aligned} \tag{A.101}$$

where in each term one of the factors of the polynomial is omitted in both numerator and denominator, we may write

$$\mathbf{I} = \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_n, \tag{A.102}$$

where

$$\mathbf{P}_1 = \frac{(\mathbf{A}-\lambda_2\mathbf{I})(\mathbf{A}-\lambda_3\mathbf{I})\cdots(\mathbf{A}-\lambda_n\mathbf{I})}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)\cdots(\lambda_1-\lambda_n)}, \tag{A.103}$$

etc. Clearly $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ if $i \neq j$, because the product contains the minimal equation as a factor. Multiplying (A.102) by $\mathbf{P}_i$ therefore gives $\mathbf{P}_i^2 = \mathbf{P}_i$, showing that the $\mathbf{P}_i$ are projection operators. Further $(\mathbf{A}-\lambda_i\mathbf{I})(\mathbf{P}_i) = \mathbf{0}$, so

$$(\mathbf{A}-\lambda_i\mathbf{I})(\mathbf{P}_i\mathbf{x}) = \mathbf{0} \tag{A.104}$$

for any vector $\mathbf{x}$, and we see that $\mathbf{P}_i\mathbf{x}$, if not zero, is an eigenvector with eigenvalue λ_i. Thus $\mathbf{P}_i$ projects into the i-th eigenspace. Applying the resolution of the identity (A.102) to a vector $\mathbf{x}$ shows that it can be decomposed

$$\begin{aligned} \mathbf{x} &= \mathbf{P}_1\mathbf{x} + \mathbf{P}_2\mathbf{x} + \cdots + \mathbf{P}_n\mathbf{x} \\ &= \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n, \end{aligned} \tag{A.105}$$

where $\mathbf{x}_i$, if not zero, is an eigenvector with eigenvalue λ_i. Since any $\mathbf{x}$ can be written as a sum of eigenvectors, the eigenvectors span the space.

[9]The identity may be verified by observing that the difference of the left and right hand sides is a polynomial of degree $n-1$, which, by inspection, vanishes at the n points $x = \lambda_i$. But a polynomial that has more zeros than its degree must be identically zero.

Jordan decomposition

If the minimal polynomial has repeated roots, the matrix can still be reduced to the *Jordan canonical form*, which is diagonal except for some 1's immediately above the diagonal.

For example, suppose the characteristic equation for a 6-by-6 matrix $\mathbf{A}$ is

$$0 = \det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)^3(\lambda_2 - \lambda)^3, \tag{A.106}$$

but the minimal equation is

$$0 = (\lambda_1 - \lambda)^3(\lambda_2 - \lambda)^2. \tag{A.107}$$

Then the Jordan form of $\mathbf{A}$ might be

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix}. \tag{A.108}$$

One may easily see that (A.107) is the minimal equation for this matrix. The minimal equation alone does not uniquely specify the pattern of λ_i's and 1's in the Jordan form, though.

It is rather tedious, but quite straightforward, to show that any linear map can be reduced to a Jordan form. The proof is sketched in the following exercises:

Exercise A.15: Suppose that the linear operator T is represented by an $N \times N$ matrix, where $N > 1$. T obeys the equation

$$(T - \lambda I)^p = 0,$$

with $p = N$, but does not obey this equation for any $p < N$. Here λ is a number and I is the identity operator.

i) Show that if T has an eigenvector, the corresponding eigenvalue must be λ. Deduce that T *cannot* be diagonalized.
ii) Show that there exists a vector $\mathbf{e}_1$ such that $(T - \lambda I)^N \mathbf{e}_1 = 0$, but no lesser power of $(T - \lambda I)$ kills $\mathbf{e}_1$.

iii) Define $\mathbf{e}_2 = (T - \lambda I)\mathbf{e}_1$, $\mathbf{e}_3 = (T - \lambda I)^2\mathbf{e}_1$, *etc.* up to $\mathbf{e}_N$. Show that the vectors $\mathbf{e}_1, \dots, \mathbf{e}_N$ are linearly independent.

iv) Use $\mathbf{e}_1, \dots, \mathbf{e}_N$ as a basis for your vector space. Taking

$$\mathbf{e}_1 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_N = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

write out the matrix representing T in the $\mathbf{e}_i$ basis.

Exercise A.16: Let $T : V \to V$ be a linear map, and suppose that the minimal polynomial equation satisfied by T is

$$Q(T) = (T - \lambda_1 I)^{r_1}(T - \lambda_2 I)^{r_2} \dots (T - \lambda_n I)^{r_n} = 0.$$

Let V_{λ_i} denote the space of *generalized eigenvectors* for the eigenvalue λ_i. This is the set of $\mathbf{x}$ such that $(T - \lambda_i I)^{r_i}\mathbf{x} = 0$. You will show that

$$V = \bigoplus_i V_{\lambda_i}.$$

i) Consider the set of polynomials $Q_{\lambda_i,j}(t) = (t - \lambda_i)^{-(r_i-j+1)}Q(t)$ where $j = 1, \dots, r_i$. Show that this set of $N \equiv \sum_i r_i$ polynomials forms a basis for the vector space $\mathcal{F}_{N-1}(t)$ of polynomials in t of degree no more than $N - 1$. (Since the number of $Q_{\lambda_i,j}$ is N, and this is equal to the dimension of $\mathcal{F}_{N-1}(t)$, the claim will be established if you can show that the polynomials are linearly independent. This is easy to do: suppose that

$$\sum_{\lambda_i,j} \alpha_{\lambda_i,j} Q_{\lambda_i,j}(t) = 0.$$

Set $t = \lambda_i$ and deduce that $\alpha_{\lambda_i,1} = 0$. Knowing this, differentiate with respect to t and again set $t = \lambda_i$ and deduce that $\alpha_{\lambda_i,2} = 0$, and so on.)

ii) Since the $Q_{\lambda_i,j}$ form a basis, and since $1 \in \mathcal{F}_{N-1}$, argue that we can find $\beta_{\lambda_i,j}$ such that

$$1 = \sum_{\lambda_i,j} \beta_{\lambda_i,j} Q_{\lambda_i,j}(t).$$

Now define

$$P_i = \sum_{j=1}^{r_i} \beta_{\lambda_i,j} Q_{\lambda_i,j}(T),$$

and so

$$I = \sum_{\lambda_i} P_i, \quad (\star)$$

Use the minimal polynomial equation to deduce that $P_i P_j = 0$ if $i \neq j$. Multiplication of $\star$ by P_i then shows that $P_i P_j = \delta_{ij} P_j$. Deduce from this that $\star$ is a resolution of the identity into a sum of mutually orthogonal projection operators P_i that project onto the spaces V_{λ_i}. Conclude that any $\mathbf{x}$ can be expanded as $\mathbf{x} = \sum_i \mathbf{x}_i$ with $\mathbf{x}_i \equiv P_i \mathbf{x} \in V_{\lambda_i}$.

iii) Show that the decomposition also implies that $V_{\lambda_i} \cap V_{\lambda_j} = \{\mathbf{0}\}$ if $i \neq j$. (Hint: a vector in V_{λ_i} is called by all projectors with the possible exception of P_i and a vector in V_{λ_j} will be killed by all the projectors with the possible exception of P_j.)

iv) Put these results together to deduce that V is a direct sum of the V_{λ_i}.

v) Combine the result of part iv) with the ideas behind exercise A.15 to complete the proof of the Jordan decomposition theorem.

A.7.2 Diagonalizing quadratic forms

Do not confuse the notion of diagonalizing the matrix representing a *linear map* $A : V \to V$ with that of diagonalizing the matrix representing a *quadratic form*. A (real) quadratic form is a map $Q : V \to \mathbb{R}$, which is obtained from a symmetric bilinear form $B : V \times V \to \mathbb{R}$ by setting the two arguments, $\mathbf{x}$ and $\mathbf{y}$, in $B(\mathbf{x}, \mathbf{y})$ equal:

$$Q(\mathbf{x}) = B(\mathbf{x}, \mathbf{x}). \tag{A.109}$$

No information is lost by this specialization. We can recover the non-diagonal ($\mathbf{x} \neq \mathbf{y}$) values of B from the diagonal values, $Q(\mathbf{x})$, by using the *polarization trick*

$$B(\mathbf{x}, \mathbf{y}) = \frac{1}{2}[Q(\mathbf{x} + \mathbf{y}) - Q(\mathbf{x}) - Q(\mathbf{y})]. \tag{A.110}$$

An example of a real quadratic form is the kinetic energy term

$$T(\dot{\mathbf{x}}) = \frac{1}{2} m_{ij} \dot{x}^i \dot{x}^j = \frac{1}{2} \dot{\mathbf{x}}^T \mathbf{M} \dot{\mathbf{x}} \tag{A.111}$$

in a "small vibrations" Lagrangian. Here, $\mathbf{M}$, with entries m_{ij}, is the mass matrix.

Whilst one can diagonalize such forms by the tedious procedure of finding the eigenvalues and eigenvectors of the associated matrix, it is simpler to use Lagrange's method, which is based on repeatedly completing squares.

Consider, for example, the quadratic form

$$Q = x^2 - y^2 - z^2 + 2xy - 4xz + 6yz = (\, x,\, y,\, z\,) \begin{pmatrix} 1 & 1 & -2 \\ 1 & -1 & 3 \\ -2 & 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{A.112}$$

We complete the square involving x:

$$Q = (x + y - 2z)^2 - 2y^2 + 10yz - 5z^2, \tag{A.113}$$

where the terms outside the squared group no longer involve x. We now complete the square in y:

$$Q = (x + y - 2z)^2 - (\sqrt{2}y - \frac{5}{\sqrt{2}}z)^2 + \frac{15}{2}z^2, \tag{A.114}$$

so that the remaining term no longer contains y. Thus, on setting

$$\begin{aligned} \xi &= x + y - 2z, \\ \eta &= \sqrt{2}y - \frac{5}{\sqrt{2}}z, \\ \zeta &= \sqrt{\frac{15}{2}}z, \end{aligned}$$

we have

$$Q = \xi^2 - \eta^2 + \zeta^2 = (\, \xi,\, \eta,\, \zeta\,) \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix}. \tag{A.115}$$

If there are no x^2, y^2, or z^2 terms to get us started, then we can proceed by using $(x + y)^2$ and $(x - y)^2$. For example, consider

$$\begin{aligned} Q &= 2xy + 2yz + 2zy, \\ &= \frac{1}{2}(x + y)^2 - \frac{1}{2}(x - y)^2 + 2xz + 2yz \\ &= \frac{1}{2}(x + y)^2 + 2(x + y)z - \frac{1}{2}(x - y)^2 \\ &= \frac{1}{2}(x + y + 2z)^2 - \frac{1}{2}(x - y)^2 - 4z^2 \\ &= \xi^2 - \eta^2 - \zeta^2, \end{aligned}$$

where

$$\begin{aligned}\xi &= \frac{1}{\sqrt{2}}(x+y+2z),\\ \eta &= \frac{1}{\sqrt{2}}(x-y),\\ \zeta &= \sqrt{2}z.\end{aligned}$$

A judicious combination of these two tactics will reduce the matrix representing any real quadratic form to a matrix with ± 1's and 0's on the diagonal, and zeros elsewhere. As the egregiously asymmetric treatment of x, y, z in the last example indicates, this can be done in many ways, but *Cayley's Law of Inertia* asserts that the *signature* — the number of $+1$'s, -1's and 0's — will always be the same. Naturally, if we allow complex numbers in the redefinitions of the variables, we can always reduce the form to one with only $+1$'s and 0's.

The essential difference between diagonalizing linear maps and diagonalizing quadratic forms is that in the former case we seek matrices $\mathbf{A}$ such that $\mathbf{A}^{-1}\mathbf{M}\mathbf{A}$ is diagonal, whereas in the latter case we seek matrices $\mathbf{A}$ such that $\mathbf{A}^T\mathbf{M}\mathbf{A}$ is diagonal. Here, the superscript T denotes transposition.

Exercise A.17: Show that the matrix

$$\mathbf{Q} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

representing the quadratic form

$$Q(x,y) = ax^2 + 2bxy + cy^2$$

may be reduced to

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{or} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

depending on whether the *discriminant*, $ac - b^2$, is respectively greater than zero, less than zero, or equal to zero.

Warning: You might be tempted to refer to the discriminant $ac - b^2$ as being the determinant of Q. It is indeed the determinant of the matrix $\mathbf{Q}$, but there is no such thing as the "determinant" of the quadratic form itself. You may compute the determinant of the matrix representing Q in some basis, but if you change basis and repeat the calculation you will get a different answer. For *real* quadratic forms, however, the *sign* of the determinant stays the same, and this is all that the discriminant cares about.

A.7.3 Block-diagonalizing symplectic forms

A skew-symmetric bilinear form $\omega : V \times V \to \mathbb{R}$ is often called a *symplectic form*. Such forms play an important role in Hamiltonian dynamics and in optics. Let $\{\mathbf{e}_i\}$ be a basis for V, and set

$$\omega(\mathbf{e}_i, \mathbf{e}_j) = \omega_{ij}. \tag{A.116}$$

If $\mathbf{x} = x^i \mathbf{e}_i$ and $\mathbf{y} = y^i \mathbf{e}_i$, we therefore have

$$\omega(\mathbf{x}, \mathbf{y}) = \omega(\mathbf{e}_i, \mathbf{e}_j) x^i y^j = \omega_{ij} x^i y^j. \tag{A.117}$$

The numbers ω_{ij} can be thought of as the entries in a real skew-symmetric matrix $\mathbf{\Omega}$, in terms of which $\omega(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{\Omega} \mathbf{y}$. We cannot exactly "diagonalize" such a skew-symmetric matrix because a matrix with non-zero entries only on its principal diagonal is necessarily symmetric. We can do the next best thing, however, and reduce $\mathbf{\Omega}$ to *block diagonal* form with simple 2-by-2 skew matrices along the diagonal.

We begin by expanding ω as

$$\omega = \frac{1}{2} \omega_{ij} \mathbf{e}^{*i} \wedge, \mathbf{e}^{*j} \tag{A.118}$$

where the *wedge* (or *exterior*) product $\mathbf{e}^{*j} \wedge \mathbf{e}^{*j}$ of a pair of basis vectors in V^* denotes the particular skew-symmetric bilinear form

$$\mathbf{e}^{*i} \wedge \mathbf{e}^{*j}(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta^i_\alpha \delta^j_\beta - \delta^i_\beta \delta^j_\alpha. \tag{A.119}$$

Again, if $\mathbf{x} = x^i \mathbf{e}_i$ and $\mathbf{y} = y^i \mathbf{e}_i$, we have

$$\begin{aligned} \mathbf{e}^{*i} \wedge \mathbf{e}^{*j}(\mathbf{x}, \mathbf{y}) &= \mathbf{e}^{*i} \wedge \mathbf{e}^{*j}(x^\alpha \mathbf{e}_\alpha, y^\beta \mathbf{e}_\beta) \\ &= (\delta^i_\alpha \delta^j_\beta - \delta^i_\beta \delta^j_\alpha) x^\alpha y^\beta \\ &= x^i y^j - y^i x^j. \end{aligned} \tag{A.120}$$

Consequently

$$\omega(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \omega_{ij} (x^i y^j - y^i x^j) = \omega_{ij} x^i y^j, \tag{A.121}$$

as before. We extend the definition of the wedge product to other elements of V^* by requiring "$\wedge$" to be associative and distributive, taking note that

$$\mathbf{e}^{*i} \wedge \mathbf{e}^{*j} = -\mathbf{e}^{*j} \wedge \mathbf{e}^{*i}, \tag{A.122}$$

and so $0 = \mathbf{e}^{*1} \wedge \mathbf{e}^{*1} = \mathbf{e}^{*2} \wedge \mathbf{e}^{*2}$, *etc.*

We next show that there exists a basis $\{\mathbf{f}^{*i}\}$ of V^* such that

$$\omega = \mathbf{f}^{*1} \wedge \mathbf{f}^{*2} + \mathbf{f}^{*3} \wedge \mathbf{f}^{*4} + \cdots + \mathbf{f}^{*(p-1)} \wedge \mathbf{f}^{*p}. \tag{A.123}$$

Here, the integer $p \leq n$ is the *rank* of ω. It is necessarily an even number.

The new basis is constructed by a skew-analogue of Lagrange's method of completing the square. If

$$\omega = \frac{1}{2}\omega_{ij}\mathbf{e}^{*i} \wedge \mathbf{e}^{*j} \tag{A.124}$$

is not identically zero, we can, after re-ordering the basis if neceessary, assume that $\omega_{12} \neq 0$. Then

$$\omega = \left(\mathbf{e}^{*1} - \frac{1}{\omega_{12}}(\omega_{23}\mathbf{e}^{*3} + \cdots + \omega_{2n}\mathbf{e}^{*n})\right) \wedge (\omega_{12}\mathbf{e}^{*2} + \omega_{13}\mathbf{e}^{*3} + \cdots \omega_{1n}\mathbf{e}^{*n}) + \omega^{\{3\}} \tag{A.125}$$

where $\omega^{\{3\}} \in \bigwedge^2(V^*)$ does not contain $\mathbf{e}^{*1}$ or $\mathbf{e}^{*2}$. We set

$$\mathbf{f}^{*1} = \mathbf{e}^{*1} - \frac{1}{\omega_{12}}(\omega_{23}\mathbf{e}^{*3} + \cdots + \omega_{2n}\mathbf{e}^{*n}) \tag{A.126}$$

and

$$\mathbf{f}^{*2} = \omega_{12}\mathbf{e}^{*2} + \omega_{13}\mathbf{e}^{*3} + \cdots \omega_{1n}\mathbf{e}^{*n}. \tag{A.127}$$

Thus,

$$\omega = \mathbf{f}^{*1} \wedge \mathbf{f}^{*2} + \omega^{\{3\}}. \tag{A.128}$$

If the remainder $\omega^{\{3\}}$ is identically zero, we are done. Otherwise, we apply the same same process to $\omega^{\{3\}}$ so as to construct $\mathbf{f}^{*3}$, $\mathbf{f}^{*4}$ and $\omega^{\{5\}}$; we continue in this manner until we find a remainder, $\omega^{\{p+1\}}$, that vanishes.

Let $\{\mathbf{f}_i\}$ be the basis for V dual to the basis $\{\mathbf{f}^{*i}\}$. Then $\omega(\mathbf{f}_1, \mathbf{f}_2) = -\omega(\mathbf{f}_2, \mathbf{f}_1) = \omega(\mathbf{f}_3, \mathbf{f}_4) = -\omega(\mathbf{f}_4, \mathbf{f}_3) = 1$, and so on, all other values being zero. This shows that if we define the coefficients $a^i{}_j$ by expressing $\mathbf{f}^{*i} = a^i{}_j\mathbf{e}^{*j}$, and hence $\mathbf{e}_i = \mathbf{f}_j a^j{}_i$, then the matrix $\mathbf{\Omega}$ has been expressed as

$$\mathbf{\Omega} = \mathbf{A}^T\widetilde{\mathbf{\Omega}}\mathbf{A}, \tag{A.129}$$

where $\mathbf{A}$ is the matrix with entries $a^i{}_j$, and $\widetilde{\mathbf{\Omega}}$ is the matrix

$$\widetilde{\mathbf{\Omega}} = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & 0 & 1 & \\ & & -1 & 0 & \\ & & & & \ddots \end{pmatrix}, \tag{A.130}$$

which contains $p/2$ diagonal blocks of

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{A.131}$$

and all other entries are zero.
Example: Consider the skew bilinear form

$$\omega(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{\Omega} \mathbf{y} = (\, x^1, x^2, x^3, x^4 \,) \begin{pmatrix} 0 & 1 & 3 & 0 \\ -1 & 0 & 1 & 5 \\ -3 & -1 & 0 & 0 \\ 0 & -5 & 0 & 0 \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \\ y^3 \\ y^4 \end{pmatrix}. \tag{A.132}$$

This corresponds to

$$\omega = \mathbf{e}^{*1} \wedge \mathbf{e}^{*2} + 3\mathbf{e}^{*1} \wedge \mathbf{e}^{*3} + \mathbf{e}^{*2} \wedge \mathbf{e}^{*3} + 5\mathbf{e}^{*2} \wedge \mathbf{e}^{*4}. \tag{A.133}$$

Following our algorithm, we write ω as

$$\omega = (\mathbf{e}^{*1} - \mathbf{e}^{*3} - 5\mathbf{e}^{*4}) \wedge (\mathbf{e}^{*2} + 3\mathbf{e}^{*3}) - 15\mathbf{e}^{*3} \wedge \mathbf{e}^{*4}. \tag{A.134}$$

If we now set

$$\begin{aligned} \mathbf{f}^{*1} &= \mathbf{e}^{*1} - \mathbf{e}^{*3} - 5\mathbf{e}^{*4}, \\ \mathbf{f}^{*2} &= \mathbf{e}^{*2} + 3\mathbf{e}^{*3}, \\ \mathbf{f}^{*3} &= -15\mathbf{e}^{*3}, \\ \mathbf{f}^{*4} &= \mathbf{e}^{*4}, \end{aligned} \tag{A.135}$$

we have

$$\omega = \mathbf{f}^{*1} \wedge \mathbf{f}^{*2} + \mathbf{f}^{*3} \wedge \mathbf{f}^{*4}. \tag{A.136}$$

We have correspondingly expressed the matrix $\mathbf{\Omega}$ as

$$\begin{pmatrix} 0 & 1 & 3 & 0 \\ -1 & 0 & 1 & 5 \\ -3 & -1 & 0 & 0 \\ 0 & -5 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 3 & -15 & 0 \\ -5 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 & -5 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & -15 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{A.137}$$