1

Linear algebra

1.1 Numbers

The **natural** numbers are the positive integers and zero. **Rational** numbers are ratios of integers. **Irrational** numbers have decimal digits d_n

$$x = \sum_{n=m_x}^{\infty} \frac{d_n}{10^n} \tag{1.1}$$

that do not repeat. Thus the repeating decimals $1/2 = 0.50000\ldots$ and $1/3 = 0.\bar{3} \equiv 0.33333\ldots$ are rational, while $\pi = 3.141592654\ldots$ is irrational. Decimal arithmetic was invented in India over 1500 years ago but was not widely adopted in the Europe until the seventeenth century.

The **real** numbers $\mathbb{R}$ include the rational numbers and the irrational numbers; they correspond to all the points on an infinite line called the **real line**.

The **complex** numbers $\mathbb{C}$ are the real numbers with one new number i whose square is -1. A complex number z is a linear combination of a real number x and a real multiple iy of i

$$z = x + iy. \tag{1.2}$$

Here $x = \mathrm{Re}z$ is the **real part** of z, and $y = \mathrm{Im}z$ is its **imaginary part**. One adds complex numbers by adding their real and imaginary parts

$$z_1 + z_2 = x_1 + iy_1 + x_2 + iy_2 = x_1 + x_2 + i(y_1 + y_2). \tag{1.3}$$

Since $i^2 = -1$, the product of two complex numbers is

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + y_1 x_2). \tag{1.4}$$

The polar representation $z = r\exp(i\theta)$ of $z = x + iy$ is

$$z = x + iy = re^{i\theta} = r(\cos\theta + i\sin\theta) \tag{1.5}$$

in which r is the **modulus** or **absolute value** of z

$$r = |z| = \sqrt{x^2 + y^2} \tag{1.6}$$

and θ is its **phase** or **argument**

$$\theta = \arctan(y/x). \tag{1.7}$$

Since $\exp(2\pi i) = 1$, there is an inevitable ambiguity in the definition of the phase of any complex number: for any integer n, the phase $\theta + 2\pi n$ gives the same z as θ. In various computer languages, the function atan2(y, x) returns the angle θ in the interval $-\pi < \theta \leq \pi$ for which $(x, y) = r(\cos\theta, \sin\theta)$.

There are two common notations z^* and $\bar{z}$ for the **complex conjugate** of a complex number $z = x + iy$

$$z^* = \bar{z} = x - iy. \tag{1.8}$$

The square of the modulus of a complex number $z = x + iy$ is

$$|z|^2 = x^2 + y^2 = (x + iy)(x - iy) = \bar{z}z = z^*z. \tag{1.9}$$

The inverse of a complex number $z = x + iy$ is

$$z^{-1} = (x + iy)^{-1} = \frac{x - iy}{(x - iy)(x + iy)} = \frac{x - iy}{x^2 + y^2} = \frac{z^*}{z^*z} = \frac{z^*}{|z|^2}. \tag{1.10}$$

Grassmann numbers θ_i are **anticommuting** numbers, that is, the **anticommutator** of any two Grassmann numbers vanishes

$$\{\theta_i, \theta_j\} \equiv [\theta_i, \theta_j]_+ \equiv \theta_i\theta_j + \theta_j\theta_i = 0. \tag{1.11}$$

So the square of any Grassmann number is zero, $\theta_i^2 = 0$. We won't use these numbers until chapter 16, but they do have amusing properties. The highest monomial in N Grassmann numbers θ_i is the product $\theta_1\theta_2 \ldots \theta_N$. So the most complicated power series in two Grassmann numbers is just

$$f(\theta_1, \theta_2) = f_0 + f_1\theta_1 + f_2\theta_2 + f_{12}\theta_1\theta_2 \tag{1.12}$$

(Hermann Grassmann, 1809–1877).

1.2 Arrays

An **array** is an **ordered set** of numbers. Arrays play big roles in computer science, physics, and mathematics. They can be of any (integral) dimension.

A one-dimensional array $(a_1, a_2, \ldots, a_n)$ is variously called an ***n*-tuple**, a **row vector** when written horizontally, a **column vector** when written vertically, or an ***n*-vector**. The numbers a_k are its **entries** or **components**.

A two-dimensional array a_{ik} with i running from 1 to n and k from 1 to m is an $n \times m$ **matrix**. The numbers a_{ik} are its **entries**, **elements**, or **matrix elements**.

One can think of a matrix as a stack of row vectors or as a queue of column vectors. The entry a_{ik} is in the ith row and the kth column.

One can add together arrays of the same dimension and shape by adding their entries. Two n-tuples add as

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n) \tag{1.13}$$

and two $n \times m$ matrices a and b add as

$$(a + b)_{ik} = a_{ik} + b_{ik}. \tag{1.14}$$

One can multiply arrays by numbers. Thus z times the three-dimensional array a_{ijk} is the array with entries $z\,a_{ijk}$. One can multiply two arrays together no matter what their shapes and dimensions. The **outer product** of an n-tuple a and an m-tuple b is an $n \times m$ matrix with elements

$$(a\,b)_{ik} = a_i\,b_k \tag{1.15}$$

or an $m \times n$ matrix with entries $(ba)_{ki} = b_k a_i$. If a and b are complex, then one also can form the outer products $(\bar{a}\,b)_{ik} = \overline{a_i}\,b_k$, $(\bar{b}\,a)_{ki} = \overline{b_k}\,a_i$, and $(\bar{b}\,\bar{a})_{ki} = \overline{b_k}\,\overline{a_i}$. The outer product of a matrix a_{ik} and a three-dimensional array $b_{j\ell m}$ is a five-dimensional array

$$(a\,b)_{ikj\ell m} = a_{ik}\,b_{j\ell m}. \tag{1.16}$$

An **inner product** is possible when two arrays are of the same size in one of their dimensions. Thus the **inner product** $(a, b) \equiv \langle a|b\rangle$ or **dot-product** $a \cdot b$ of two real n-tuples a and b is

$$(a, b) = \langle a|b\rangle = a \cdot b = (a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1b_1 + \cdots + a_nb_n. \tag{1.17}$$

The inner product of two complex n-tuples often is defined as

$$(a, b) = \langle a|b\rangle = \bar{a} \cdot b = (\overline{a_1}, \ldots, \overline{a_n}) \cdot (b_1, \ldots, b_n) = \overline{a_1}\,b_1 + \cdots + \overline{a_n}\,b_n \tag{1.18}$$

or as its complex conjugate

$$(a, b)^* = \langle a|b\rangle^* = (\bar{a} \cdot b)^* = (b, a) = \langle b|a\rangle = \bar{b} \cdot a \tag{1.19}$$

so that the inner product of a vector with itself is nonnegative $(a, a) \geq 0$.

The product of an $m \times n$ matrix a_{ik} times an n-tuple b_k is the m-tuple b' whose ith component is

$$b'_i = a_{i1}b_1 + a_{i2}b_2 + \cdots + a_{in}b_n = \sum_{k=1}^{n} a_{ik}b_k. \tag{1.20}$$

This product is $b' = a\,b$ in matrix notation.

If the size n of the second dimension of a matrix a matches that of the first dimension of a matrix b, then their product $a\,b$ is a matrix with entries

$$(a\,b)_{i\ell} = a_{i1}\,b_{1\ell} + \cdots + a_{in}\,b_{n\ell}. \tag{1.21}$$

1.3 Matrices

Apart from n-tuples, the most important arrays in linear algebra are the two-dimensional arrays called matrices.

The **trace** of an $n \times n$ matrix a is the sum of its diagonal elements

$$\operatorname{Tr} a = \operatorname{tr} a = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}. \tag{1.22}$$

The trace of two matrices is independent of their order

$$\operatorname{Tr}(a\,b) = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik} b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} b_{ki} a_{ik} = \operatorname{Tr}(b a) \tag{1.23}$$

as long as the matrix elements are numbers that commute with each other. It follows that the trace is **cyclic**

$$\operatorname{Tr}(a\,b \ldots z) = \operatorname{Tr}(b \ldots z\,a)\,. \tag{1.24}$$

The **transpose** of an $n \times \ell$ matrix a is an $\ell \times n$ matrix a^{T} with entries

$$\left(a^{\mathsf{T}}\right)_{ij} = a_{ji}. \tag{1.25}$$

Some mathematicians use a prime to mean transpose, as in $a' = a^{\mathsf{T}}$, but physicists tend to use primes to label different objects or to indicate differentiation. One may show that

$$(a\,b)^{\mathsf{T}} = b^{\mathsf{T}}\,a^{\mathsf{T}}. \tag{1.26}$$

A matrix that is equal to its transpose

$$a = a^{\mathsf{T}} \tag{1.27}$$

is **symmetric**.

The (hermitian) **adjoint** of a matrix is the complex conjugate of its transpose (Charles Hermite, 1822–1901). That is, the (hermitian) adjoint $a^\dagger$ of an $N \times L$ complex matrix a is the $L \times N$ matrix with entries

$$(a^\dagger)_{ij} = (a_{ji})^* = a^*_{ji}. \tag{1.28}$$

One may show that

$$(a\,b)^\dagger = b^\dagger\,a^\dagger. \tag{1.29}$$

A matrix that is equal to its adjoint

$$(a^\dagger)_{ij} = (a_{ji})^* = a^*_{ji} = a_{ij} \tag{1.30}$$

(and which must be a square matrix) is **hermitian** or **self adjoint**

$$a = a^\dagger. \tag{1.31}$$

Example 1.1 (The Pauli matrices)

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.32}$$

are all hermitian (Wolfgang Pauli, 1900–1958). □

A real hermitian matrix is symmetric. If a matrix a is hermitian, then the quadratic form

$$\langle v|a|v\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} v_i^* a_{ij} v_j \in \mathbb{R} \tag{1.33}$$

is real for all complex n-tuples v.

The **Kronecker delta** δ_{ik} is defined to be unity if $i = k$ and zero if $i \neq k$ (Leopold Kronecker, 1823–1891). The **identity matrix** I has entries $I_{ik} = \delta_{ik}$.

The **inverse** a^{-1} of an $n \times n$ matrix a is a square matrix that satisfies

$$a^{-1}\,a = a\,a^{-1} = I \tag{1.34}$$

in which I is the $n \times n$ identity matrix.

So far we have been writing n-tuples and matrices and their elements with lower-case letters. It is equally common to use capital letters, and we will do so for the rest of this section.

A matrix U whose adjoint $U^\dagger$ is its inverse

$$U^\dagger U = U U^\dagger = I \tag{1.35}$$

is **unitary**. Unitary matrices are square.

A real unitary matrix O is **orthogonal** and obeys the rule

$$O^{\mathsf{T}} O = O O^{\mathsf{T}} = I. \tag{1.36}$$

Orthogonal matrices are square.

An $N \times N$ hermitian matrix A is **nonnegative**

$$A \geq 0 \tag{1.37}$$

if for all complex vectors V the quadratic form

$$\langle V|A|V\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} V_i^* A_{ij} V_j \geq 0 \tag{1.38}$$

is nonnegative. It is **positive** or **positive definite** if

$$\langle V|A|V\rangle > 0 \tag{1.39}$$

for all nonzero vectors $|V\rangle$.

Example 1.2 (Kinds of positivity) The nonsymmetric, nonhermitian 2×2 matrix

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{1.40}$$

is positive on the space of all real 2-vectors but not on the space of all complex 2-vectors. □

Example 1.3 (Representations of imaginary and Grassmann numbers) The 2×2 matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{1.41}$$

can represent the number i since

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I. \tag{1.42}$$

The 2×2 matrix

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{1.43}$$

can represent a Grassmann number since

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0. \tag{1.44}$$

To represent two Grassmann numbers, one needs 4×4 matrices, such as

$$\theta_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \theta_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.45}$$

The matrices that represent n Grassmann numbers are $2^n \times 2^n$. □

Example 1.4 (Fermions) The matrices (1.45) also can represent lowering or annihilation operators for a system of two fermionic states. For $a_1 = \theta_1$ and $a_2 = \theta_2$ and their adjoints $a_1^\dagger$ and $a_2^\dagger$, the creation operators satisfy the anticommutation relations

$$\{a_i, a_k^\dagger\} = \delta_{ik} \quad \text{and} \quad \{a_i, a_k\} = \{a_i^\dagger, a_k^\dagger\} = 0 \tag{1.46}$$

where i and k take the values 1 or 2. In particular, the relation $(a_i^\dagger)^2 = 0$ implements **Pauli's exclusion principle**, the rule that no state of a fermion can be doubly occupied. □

1.4 Vectors

Vectors are things that can be multiplied by numbers and added together to form other vectors in the same **vector space**. So if U and V are vectors in a vector space S over a set F of numbers x and y and so forth, then

$$W = x\,U + y\,V \tag{1.47}$$

also is a vector in the vector space S.

A **basis** for a vector space S is a set of vectors B_k for $k = 1, \ldots, N$ in terms of which every vector U in S can be expressed as a linear combination

$$U = u_1 B_1 + u_2 B_2 + \cdots + u_N B_N \tag{1.48}$$

with numbers u_k in F. The numbers u_k are the **components** of the vector U in the basis B_k.

Example 1.5 (Hardware store) Suppose the vector W represents a certain kind of washer and the vector N represents a certain kind of nail. Then if n and m are natural numbers, the vector

$$H = nW + mN \tag{1.49}$$

would represent a possible inventory of a very simple hardware store. The vector space of all such vectors H would include all possible inventories of the store. That space is a two-dimensional vector space over the natural numbers, and the two vectors W and N form a basis for it. □

Example 1.6 (Complex numbers) The complex numbers are a vector space. Two of its vectors are the number 1 and the number i; the vector space of complex numbers is then the set of all linear combinations

$$z = x1 + yi = x + iy. \tag{1.50}$$

So the complex numbers are a two-dimensional vector space over the real numbers, and the vectors 1 and i are a basis for it.

The complex numbers also form a one-dimensional vector space over the complex numbers. Here any nonzero real or complex number, for instance the number 1, can be a basis consisting of the single vector 1. This one-dimensional vector space is the set of all $z = z1$ for arbitrary complex z. □

Example 1.7 (2-space) Ordinary flat two-dimensional space is the set of all linear combinations

$$r = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{1.51}$$

in which x and y are real numbers and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are perpendicular vectors of unit length (unit vectors). This vector space, called $\mathbb{R}^2$, is a 2-d space over the reals.

Note that the same vector r can be described either by the basis vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ or by any other set of basis vectors, such as $-\hat{\mathbf{y}}$ and $\hat{\mathbf{x}}$

$$r = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} = -y(-\hat{\mathbf{y}}) + x\hat{\mathbf{x}}. \tag{1.52}$$

So the components of the vector r are (x, y) in the $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}\}$ basis and $(-y, x)$ in the $\{-\hat{\mathbf{y}}, \hat{\mathbf{x}}\}$ basis. **Each vector is unique, but its components depend upon the basis.** □

Example 1.8 (3-space) Ordinary flat three-dimensional space is the set of all linear combinations

$$r = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \tag{1.53}$$

in which x, y, and z are real numbers. It is a 3-d space over the reals. □

Example 1.9 (Matrices) Arrays of a given dimension and size can be added and multiplied by numbers, and so they form a vector space. For instance, all complex three-dimensional arrays a_{ijk} in which $1 \leq i \leq 3$, $1 \leq j \leq 4$, and $1 \leq k \leq 5$ form a vector space over the complex numbers. □

Example 1.10 (Partial derivatives) Derivatives are vectors, so are partial derivatives. For instance, the linear combinations of x and y partial derivatives taken at $x = y = 0$

$$a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y} \tag{1.54}$$

form a vector space. □

Example 1.11 (Functions) The space of all linear combinations of a set of functions $f_i(x)$ defined on an interval $[a, b]$

$$f(x) = \sum_i z_i f_i(x) \tag{1.55}$$

is a vector space over the natural, real, or complex numbers $\{z_i\}$. □

Example 1.12 (States) In quantum mechanics, a state is represented by a vector, often written as ψ or in Dirac's notation as $|\psi\rangle$. If c_1 and c_2 are complex numbers, and $|\psi_1\rangle$ and $|\psi_2\rangle$ are any two states, then the linear combination

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle \tag{1.56}$$

also is a possible state of the system. □

1.5 Linear operators

A **linear operator** A maps each vector U in its **domain** into a vector $U' = A(U) \equiv AU$ in its **range** in a way that is linear. So if U and V are two vectors in its domain and b and c are numbers, then

$$A(bU + cV) = bA(U) + cA(V) = bAU + cAV. \tag{1.57}$$

If the domain and the range are the same vector space S, then A maps each basis vector B_i of S into a linear combination of the basis vectors B_k

$$AB_i = a_{1i}B_1 + a_{2i}B_2 + \cdots + a_{Ni}B_N = \sum_{k=1}^{N} a_{ki}\, B_k. \tag{1.58}$$

The square matrix a_{ki} **represents** the linear operator A in the B_k basis. The effect of A on any vector $U = u_1B_1 + u_2B_2 + \cdots + u_NB_N$ in S then is

$$\begin{aligned} AU = A\left(\sum_{i=1}^{N} u_iB_i\right) &= \sum_{i=1}^{N} u_iAB_i = \sum_{i=1}^{N} u_i \sum_{k=1}^{N} a_{ki}\, B_k \\ &= \sum_{k=1}^{N}\left(\sum_{i=1}^{N} a_{ki}u_i\right) B_k. \end{aligned} \tag{1.59}$$

So the kth component u'_k of the vector $U' = AU$ is

$$u'_k = a_{k1}u_1 + a_{k2}u_2 + \cdots + a_{kN}u_N = \sum_{i=1}^{N} a_{ki}\, u_i. \tag{1.60}$$

Thus the column vector u' of the components u'_k of the vector $U' = AU$ is the product $u' = a\,u$ of the matrix with elements a_{ki} that represents the linear operator A in the B_k basis and the column vector with components u_i that represents the vector U in that basis. So in each basis, vectors and linear operators are represented by column vectors and matrices.

Each linear operator is unique, but its matrix depends upon the basis. If we change from the B_k basis to another basis B'_k

$$B_k = \sum_{\ell=1}^{N} u_{\ell k}\, B'_\ell \tag{1.61}$$

in which the $N \times N$ matrix $u_{\ell k}$ has an inverse matrix u^{-1}_{ki} so that

$$\sum_{k=1}^{N} u^{-1}_{ki} B_k = \sum_{k=1}^{N} u^{-1}_{ki} \sum_{\ell=1}^{N} u_{\ell k} B'_\ell = \sum_{\ell=1}^{N}\left(\sum_{k=1}^{N} u_{\ell k} u^{-1}_{ki}\right) B'_\ell = \sum_{\ell=1}^{N} \delta_{\ell i} B'_\ell = B'_i, \tag{1.62}$$

then the new basis vectors B'_i are given by

$$B'_i = \sum_{k=1}^{N} u^{-1}_{ki} B_k. \tag{1.63}$$

Thus (exercise 1.9) the linear operator A maps the basis vector B'_i to

$$AB'_i = \sum_{k=1}^{N} u^{-1}_{ki} AB_k = \sum_{j,k=1}^{N} u^{-1}_{ki} a_{jk} B_j = \sum_{j,k,\ell=1}^{N} u_{\ell j} a_{jk} u^{-1}_{ki} B'_\ell. \tag{1.64}$$

So the matrix a' that represents A in the B' basis is related to the matrix a that represents it in the B basis by a **similarity transformation**

$$a'_{\ell i} = \sum_{jk=1}^{N} u_{\ell j} a_{jk} u^{-1}_{ki} \quad \text{or} \quad a' = u\, a\, u^{-1} \tag{1.65}$$

in matrix notation.

Example 1.13 (Change of basis) Let the action of the linear operator A on the basis vectors $\{B_1, B_2\}$ be $AB_1 = B_2$ and $AB_2 = 0$. If the column vectors

$$b_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad b_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.66}$$

represent the basis vectors B_1 and B_2, then the matrix

$$a = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{1.67}$$

represents the linear operator A. But if we use the basis vectors

$$B'_1 = \frac{1}{\sqrt{2}}(B_1 + B_2) \quad \text{and} \quad B'_2 = \frac{1}{\sqrt{2}}(B_1 - B_2) \tag{1.68}$$

then the vectors

$$b'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad b'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{1.69}$$

would represent B_1 and B_2, and the matrix

$$a' = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \tag{1.70}$$

would represent the linear operator A (exercise 1.10). □

A linear operator A also may map a vector space S with basis B_k into a different vector space T with its own basis C_k. In this case, A maps the basis vector B_i into a linear combination of the basis vectors C_k

$$AB_i = \sum_{k=1}^{M} a_{ki} C_k \tag{1.71}$$

and an arbitrary vector $U = u_1 B_1 + \cdots + u_N B_N$ in S into the vector

$$AU = \sum_{k=1}^{M} \left(\sum_{i=1}^{N} a_{ki} u_i \right) C_k \tag{1.72}$$

in T.

1.6 Inner products

Most of the vector spaces used by physicists have an inner product. A **positive-definite inner product** associates a number (f, g) with every ordered pair of vectors f and g in the vector space V and satisfies the rules

$$(f, g) = (g, f)^* \tag{1.73}$$

$$(f, z\,g + w\,h) = z\,(f, g) + w\,(f, h) \tag{1.74}$$

$$(f, f) \geq 0 \quad \text{and} \quad (f, f) = 0 \iff f = 0 \tag{1.75}$$

in which f, g, and h are vectors, and z and w are numbers. The first rule says that the inner product is **hermitian**; the second rule says that it is **linear** in the second vector $z\,g + w\,h$ of the pair; and the third rule says that it is **positive definite**. The first two rules imply that (exercise 1.11) the inner product is **antilinear** in the first vector of the pair

$$(z\,g + w\,h, f) = z^*(g, f) + w^*(h, f). \tag{1.76}$$

A **Schwarz inner product** satisfies the first two rules (1.73, 1.74) for an inner product and the fourth (1.76) but only the first part of the third (1.75)

$$(f, f) \geq 0. \tag{1.77}$$

This condition of **nonnegativity** implies (exercise 1.15) that a vector f of zero length must be orthogonal to all vectors g in the vector space V

$$(f, f) = 0 \implies (g, f) = 0 \text{ for all } g \in V. \tag{1.78}$$

So a Schwarz inner product is *almost* positive definite.

Inner products of 4-vectors can be negative. To accommodate them we define an **indefinite** inner product without regard to positivity as one that satisfies the first two rules (1.73 & 1.74) and therefore also the fourth rule (1.76) and that instead of being positive definite is **nondegenerate**

$$(f, g) = 0 \text{ for all } f \in V \implies g = 0. \tag{1.79}$$

This rule says that only the zero vector is orthogonal to all the vectors of the space. The positive-definite condition (1.75) is stronger than and implies nondegeneracy (1.79) (exercise 1.14).

Apart from the indefinite inner products of 4-vectors in special and general relativity, most of the inner products physicists use are Schwarz inner products or positive-definite inner products. For such inner products, we can define the **norm** $|f| = \| f \|$ of a vector f as the square-root of the nonnegative inner product (f,f)

$$\| f \| = \sqrt{(f,f)}. \tag{1.80}$$

The **distance** between two vectors f and g is the norm of their difference

$$\| f - g \| . \tag{1.81}$$

Example 1.14 (Euclidean space) The space of real vectors U, V with N components U_i, V_i forms an N-dimensional vector space over the real numbers with an inner product

$$(U, V) = \sum_{i=1}^{N} U_i V_i \tag{1.82}$$

that is nonnegative when the two vectors are the same

$$(U, U) = \sum_{i=1}^{N} U_i U_i = \sum_{i=1}^{N} U_i^2 \geq 0 \tag{1.83}$$

and vanishes only if all the components U_i are zero, that is, if the vector $U = 0$. Thus the inner product (1.82) is positive definite. When (U, V) is zero, the vectors U and V are **orthogonal**. □

Example 1.15 (Complex euclidean space) The space of complex vectors with N components U_i, V_i forms an N-dimensional vector space over the complex numbers with inner product

$$(U, V) = \sum_{i=1}^{N} U_i^* V_i = (V, U)^*. \tag{1.84}$$

The inner product (U, U) is nonnegative and vanishes

$$(U, U) = \sum_{i=1}^{N} U_i^* U_i = \sum_{i=1}^{N} |U_i|^2 \geq 0 \tag{1.85}$$

only if $U = 0$. So the inner product (1.84) is positive definite. If (U, V) is zero, then U and V are orthogonal. □

Example 1.16 (Complex matrices) For the vector space of $N \times L$ complex matrices A, B, ..., the trace of the adjoint (1.28) of A multiplied by B is an inner product

$$(A, B) = \mathrm{Tr} A^\dagger B = \sum_{i=1}^{N} \sum_{j=1}^{L} (A^\dagger)_{ji} B_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{L} A_{ij}^* B_{ij} \tag{1.86}$$

that is nonnegative when the matrices are the same

$$(A, A) = \mathrm{Tr} A^\dagger A = \sum_{i=1}^{N} \sum_{j=1}^{L} A_{ij}^* A_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{L} |A_{ij}|^2 \geq 0 \tag{1.87}$$

and zero only when $A = 0$. So this inner product is positive definite. □

A vector space with a positive-definite inner product (1.73–1.77) is called an **inner-product space**, a **metric space**, or a **pre-Hilbert space**.

A sequence of vectors f_n is a **Cauchy sequence** if for every $\epsilon > 0$ there is an integer $N(\epsilon)$ such that $\| f_n - f_m \| < \epsilon$ whenever both n and m exceed $N(\epsilon)$. A sequence of vectors f_n **converges** to a vector f if for every $\epsilon > 0$ there is an integer $N(\epsilon)$ such that $\| f - f_n \| < \epsilon$ whenever n exceeds $N(\epsilon)$. An inner-product space with a norm defined as in (1.80) is **complete** if each of its Cauchy sequences converges to a vector in that space. A **Hilbert space** is a complete inner-product space. Every finite-dimensional inner-product space is complete and so is a Hilbert space. But the term *Hilbert space* more often is used to describe infinite-dimensional complete inner-product spaces, such as the space of all square-integrable functions (David Hilbert, 1862–1943).

Example 1.17 (The Hilbert space of square-integrable functions) For the vector space of functions (1.55), a natural inner product is

$$(f, g) = \int_a^b dx f^*(x) g(x). \tag{1.88}$$

The squared norm $\| f \|$ of a function $f(x)$ is

$$\| f \|^2 = \int_a^b dx |f(x)|^2. \tag{1.89}$$

A function is **square integrable** if its norm is finite. The space of all square-integrable functions is an inner-product space; it also is complete and so is a Hilbert space. □

Example 1.18 (Minkowski inner product) The Minkowski or Lorentz inner product (p, x) of two 4-vectors $p = (E/c, p_1, p_2, p_3)$ and $x = (ct, x_1, x_2, x_3)$ is

$\boldsymbol{p}\cdot\boldsymbol{x} - Et$. It is indefinite, nondegenerate, and invariant under Lorentz transformations, and often is written as $p\cdot x$ or as px. If p is the 4-momentum of a freely moving physical particle of mass m, then

$$p\cdot p = \boldsymbol{p}\cdot\boldsymbol{p} - E^2/c^2 = -c^2m^2 \leq 0. \tag{1.90}$$

The Minkowski inner product satisfies the rules (1.73, 1.75, and 1.79), but it is **not positive definite**, and it does not satisfy the Schwarz inequality (Hermann Minkowski, 1864–1909; Hendrik Lorentz, 1853–1928). □

1.7 The Cauchy–Schwarz inequality

For any two vectors f and g, the Schwarz inequality

$$(f,f)(g,g) \geq |(f,g)|^2 \tag{1.91}$$

holds for any Schwarz inner product (and so for any positive-definite inner product). The condition (1.77) of nonnegativity ensures that for any complex number λ the inner product of the vector $f - \lambda g$ with itself is nonnegative

$$(f-\lambda g, f-\lambda g) = (f,f) - \lambda^*(g,f) - \lambda(f,g) + |\lambda|^2(g,g) \geq 0. \tag{1.92}$$

Now if $(g,g) = 0$, then for $(f-\lambda g, f-\lambda g)$ to remain nonnegative for all complex values of λ it is necessary that $(f,g) = 0$ also vanish (exercise 1.15). Thus if $(g,g) = 0$, then the Schwarz inequality (1.91) is trivially true because both sides of it vanish. So we assume that $(g,g) > 0$ and set $\lambda = (g,f)/(g,g)$. The inequality (1.92) then gives us

$$(f-\lambda g, f-\lambda g) = \left(f - \frac{(g,f)}{(g,g)}g, f - \frac{(g,f)}{(g,g)}g\right) = (f,f) - \frac{(f,g)(g,f)}{(g,g)} \geq 0$$

which is the Schwarz inequality (1.91) (Hermann Schwarz, 1843–1921)

$$(f,f)(g,g) \geq |(f,g)|^2. \tag{1.93}$$

Taking the square-root of each side, we get

$$\| f \| \| g \| \geq |(f,g)|. \tag{1.94}$$

Example 1.19 (Some Schwarz inequalities) For the dot-product of two real 3-vectors $\boldsymbol{r}$ and $\boldsymbol{R}$, the Cauchy–Schwarz inequality is

$$(\boldsymbol{r}\cdot\boldsymbol{r})\,(\boldsymbol{R}\cdot\boldsymbol{R}) \geq (\boldsymbol{r}\cdot\boldsymbol{R})^2 = (\boldsymbol{r}\cdot\boldsymbol{r})\,(\boldsymbol{R}\cdot\boldsymbol{R})\cos^2\theta \tag{1.95}$$

where θ is the angle between $\boldsymbol{r}$ and $\boldsymbol{R}$.

The Schwarz inequality for two real n-vectors $\boldsymbol{x}$ is

$$(\boldsymbol{x}\cdot\boldsymbol{x})\,(\boldsymbol{y}\cdot\boldsymbol{y}) \geq (\boldsymbol{x}\cdot\boldsymbol{y})^2 = (\boldsymbol{x}\cdot\boldsymbol{x})\,(\boldsymbol{y}\cdot\boldsymbol{y})\cos^2\theta \tag{1.96}$$

and it implies (Exercise 1.16) that

$$\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \geq \|\boldsymbol{x}+\boldsymbol{y}\|. \tag{1.97}$$

For two complex n-vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, the Schwarz inequality is

$$(\boldsymbol{u}^*\cdot\boldsymbol{u})\,(\boldsymbol{v}^*\cdot\boldsymbol{v}) \geq |\boldsymbol{u}^*\cdot\boldsymbol{v}|^2 = (\boldsymbol{u}^*\cdot\boldsymbol{u})\,(\boldsymbol{v}^*\cdot\boldsymbol{v})\cos^2\theta \tag{1.98}$$

and it implies (exercise 1.17) that

$$\|\boldsymbol{u}\| + \|\boldsymbol{v}\| \geq \|\boldsymbol{u}+\boldsymbol{v}\|. \tag{1.99}$$

The inner product (1.88) of two complex functions f and g provides a somewhat different instance

$$\int_a^b dx\,|f(x)|^2 \int_a^b dx\,|g(x)|^2 \geq \left|\int_a^b dx\, f^*(x)g(x)\right|^2 \tag{1.100}$$

of the Schwarz inequality. □

1.8 Linear independence and completeness

A set of N vectors $V_1, V_2, \ldots, V_N$ is **linearly dependent** if there exist numbers c_i, *not all zero*, such that the linear combination

$$c_1V_1 + \cdots + c_NV_N = 0 \tag{1.101}$$

vanishes. A set of vectors is **linearly independent** if it is not linearly dependent.

A set $\{V_i\}$ of linearly independent vectors is **maximal** in a vector space S if the addition of any other vector U in S to the set $\{V_i\}$ makes the enlarged set $\{U, V_i\}$ linearly dependent.

A set of N linearly independent vectors $V_1, V_2, \ldots, V_N$ that is maximal in a vector space S can represent any vector U in the space S as a linear combination of its vectors, $U = u_1V_1 + \cdots + u_NV_N$. For if we enlarge the maximal set $\{V_i\}$ by including in it any vector U not already in it, then the bigger set $\{U, V_i\}$ will be linearly dependent. Thus there will be numbers $c, c_1, \ldots, c_N$, not all zero, that make the sum

$$cU + c_1V_1 + \cdots + c_NV_N = 0 \tag{1.102}$$

vanish. Now if c were 0, then the set $\{V_i\}$ would be linearly dependent. Thus $c \neq 0$, and so we may divide by c and express the arbitrary vector U as a linear combination of the vectors V_i

$$U = -\frac{1}{c}(c_1V_1 + \cdots + c_NV_N) = u_1V_1 + \cdots + u_NV_N \tag{1.103}$$

with $u_k = -c_k/c$. So a set of linearly independent vectors $\{V_i\}$ that is maximal in a space S can represent every vector U in S as a linear combination

$U = u_1 V_1 + \ldots + u_N V_N$ of its vectors. The set $\{V_i\}$ **spans** the space S; it is a **complete** set of vectors in the space S.

A set of vectors $\{V_i\}$ that spans a vector space S provides a **basis** for that space because the set lets us represent an arbitrary vector U in S as a linear combination of the basis vectors $\{V_i\}$. If the vectors of a basis are linearly dependent, then at least one of them is superfluous, and so it is convenient to have the vectors of a basis be linearly independent.

1.9 Dimension of a vector space

If $V_1, \ldots, V_N$ and $W_1, \ldots, W_M$ are two maximal sets of N and M linearly independent vectors in a vector space S, then $N = M$.

Suppose $M < N$. Since the Us are complete, they span S, and so we may express each of the N vectors V_i in terms of the M vectors W_j

$$V_i = \sum_{j=1}^{M} A_{ij} W_j. \tag{1.104}$$

Let A_j be the vector with components A_{ij}. There are $M < N$ such vectors, and each has $N > M$ components. So it is always possible to find a nonzero N-dimensional vector C with components c_i that is orthogonal to all M vectors A_j

$$\sum_{i=1}^{N} c_i A_{ij} = 0. \tag{1.105}$$

Thus the linear combination

$$\sum_{i=1}^{N} c_i V_i = \sum_{i=1}^{N} \sum_{j=1}^{M} c_i A_{ij} W_j = 0 \tag{1.106}$$

vanishes, which implies that the N vectors V_i are linearly dependent. Since these vectors are by assumption linearly independent, it follows that $N \leq M$.

Similarly, one may show that $M \leq N$. Thus $M = N$.

The number of vectors in a maximal set of linearly independent vectors in a vector space S is the **dimension** of the vector space. Any N linearly independent vectors in an N-dimensional space form a **basis** for it.

1.10 Orthonormal vectors

Suppose the vectors $V_1, V_2, \ldots, V_N$ are linearly independent. Then we can make out of them a set of N vectors U_i that are orthonormal

$$(U_i, U_j) = \delta_{ij}. \tag{1.107}$$

There are many ways to do this, because there are many such sets of orthonormal vectors. We will use the Gram–Schmidt method. We set

$$U_1 = \frac{V_1}{\sqrt{(V_1, V_1)}}, \tag{1.108}$$

so the first vector U_1 is normalized. Next we set $u_2 = V_2 + c_{12} U_1$ and require that u_2 be orthogonal to U_1

$$0 = (U_1, u_2) = (U_1, c_{12} U_1 + V_2) = c_{12} + (U_1, V_2). \tag{1.109}$$

Thus $c_{12} = -(U_1, V_2)$, and so

$$u_2 = V_2 - (U_1, V_2)\, U_1. \tag{1.110}$$

The normalized vector U_2 then is

$$U_2 = \frac{u_2}{\sqrt{(u_2, u_2)}}. \tag{1.111}$$

We next set $u_3 = V_3 + c_{13} U_1 + c_{23} U_2$ and ask that u_3 be orthogonal to U_1

$$0 = (U_1, u_3) = (U_1, c_{13} U_1 + c_{23} U_2 + V_3) = c_{13} + (U_1, V_3) \tag{1.112}$$

and also to U_2

$$0 = (U_2, u_3) = (U_2, c_{13} U_1 + c_{23} U_2 + V_3) = c_{23} + (U_2, V_3). \tag{1.113}$$

So $c_{13} = -(U_1, V_3)$ and $c_{23} = -(U_2, V_3)$, and we have

$$u_3 = V_3 - (U_1, V_3)\, U_1 - (U_2, V_3)\, U_2. \tag{1.114}$$

The normalized vector U_3 then is

$$U_3 = \frac{u_3}{\sqrt{(u_3, u_3)}}. \tag{1.115}$$

We may continue in this way until we reach the last of the N linearly independent vectors. We require the kth unnormalized vector u_k

$$u_k = V_k + \sum_{i=1}^{k-1} c_{ik}\, U_i \tag{1.116}$$

to be orthogonal to the $k-1$ vectors U_i and find that $c_{ik} = -(U_i, V_k)$ so that

$$u_k = V_k - \sum_{i=1}^{k-1} (U_i, V_k)\, U_i. \tag{1.117}$$

The normalized vector then is

$$U_k = \frac{u_k}{\sqrt{(u_k, u_k)}}. \tag{1.118}$$

A basis is more convenient if its vectors are orthonormal.

1.11 Outer products

From any two vectors f and g, we may make an operator A that takes any vector h into the vector f with coefficient (g, h)

$$A\,h = f\,(g,h). \tag{1.119}$$

Since for any vectors e, h and numbers z, w

$$A\,(z\,h + w\,e) = f\,(g, z\,h + w\,e) = z f\,(g,h) + w f\,(g,e) = z\,A\,h + w\,A\,e \tag{1.120}$$

it follows that A is linear.

If in some basis f, g, and h are vectors with components f_i, g_i, and h_i, then the linear transformation is

$$(Ah)_i = \sum_{j=1}^{N} A_{ij}\,h_j = f_i \sum_{j=1}^{N} g_j^*\,h_j \tag{1.121}$$

and in that basis A is the matrix with entries

$$A_{ij} = f_i\,g_j^*. \tag{1.122}$$

It is the **outer product** of the vectors f and g.

Example 1.20 (Outer product) If in some basis the vectors f and g are

$$f = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} i \\ 1 \\ 3i \end{pmatrix} \tag{1.123}$$

then their outer product is the matrix

$$A = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \begin{pmatrix} -i & 1 & -3i \end{pmatrix} = \begin{pmatrix} -2i & 2 & -6i \\ -3i & 3 & -9i \end{pmatrix}. \tag{1.124}$$

Dirac developed a notation that handles outer products very easily. □

Example 1.21 (Outer products) If the vectors $f = |f\rangle$ and $g = |g\rangle$ are

$$|f\rangle = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad |g\rangle = \begin{pmatrix} z \\ w \end{pmatrix} \tag{1.125}$$

then their outer products are

$$|f\rangle\langle g| = \begin{pmatrix} az^* & aw^* \\ bz^* & bw^* \\ cz^* & cw^* \end{pmatrix} \quad \text{and} \quad |g\rangle\langle f| = \begin{pmatrix} za^* & zb^* & zc^* \\ wa^* & wb^* & wc^* \end{pmatrix} \tag{1.126}$$

as well as

$$|f\rangle\langle f| = \begin{pmatrix} aa^* & ab^* & ac^* \\ ba^* & bb^* & bc^* \\ ca^* & cb^* & cc^* \end{pmatrix} \quad \text{and} \quad |g\rangle\langle g| = \begin{pmatrix} zz^* & zw^* \\ wz^* & ww^* \end{pmatrix}. \tag{1.127}$$

Students should feel free to write down their own examples. □

1.12 Dirac notation

Outer products are important in quantum mechanics, and so Dirac invented a notation for linear algebra that makes them easy to write. In his notation, a vector f is a **ket** $f = |f\rangle$. The new thing in his notation is the **bra** $\langle g|$. The inner product of two vectors (g,f) is the **bracket** $(g,f) = \langle g|f\rangle$. A matrix element (g, Af) is then $(g, Af) = \langle g|A|f\rangle$ in which the bra and ket bracket the operator. In Dirac notation, the outer product $A\,h = f\,(g,h)$ reads $A\,|h\rangle = |f\rangle\langle g|h\rangle$, so that the outer product A itself is $A = |f\rangle\langle g|$. Before Dirac, bras were implicit in the definition of the inner product, but they did not appear explicitly; there was no way to write the bra $\langle g|$ or the operator $|f\rangle\langle g|$.

If the kets $|n\rangle$ form an orthonormal basis in an N-dimensional vector space, then we can expand an arbitrary ket in the space as

$$|f\rangle = \sum_{n=1}^{N} c_n |n\rangle. \tag{1.128}$$

Since the basis vectors are orthonormal $\langle \ell|n\rangle = \delta_{\ell n}$, we can identify the coefficients c_n by forming the inner product

$$\langle \ell|f\rangle = \sum_{n=1}^{N} c_n\,\langle \ell|n\rangle = \sum_{n=1}^{N} c_n\,\delta_{\ell,n} = c_\ell. \tag{1.129}$$

The original expansion (1.128) then must be

$$|f\rangle = \sum_{n=1}^{N} c_n |n\rangle = \sum_{n=1}^{N} \langle n|f\rangle\,|n\rangle = \sum_{n=1}^{N} |n\rangle\,\langle n|f\rangle = \left(\sum_{n=1}^{N} |n\rangle\,\langle n|\right) |f\rangle. \tag{1.130}$$

Since this equation must hold for *every* vector $|f\rangle$ in the space, it follows that the sum of outer products within the parentheses is the identity operator for the space

$$I = \sum_{n=1}^{N} |n\rangle\,\langle n|. \tag{1.131}$$

Every set of kets $|\alpha_n\rangle$ that forms an orthonormal basis $\langle\alpha_n|\alpha_\ell\rangle = \delta_{n\ell}$ for the space gives us an equivalent representation of the identity operator

$$I = \sum_{n=1}^{N} |\alpha_n\rangle \langle\alpha_n| = \sum_{n=1}^{N} |n\rangle \langle n|. \tag{1.132}$$

Before Dirac, one could not write such equations. They provide for every vector $|f\rangle$ in the space the expansions

$$|f\rangle = \sum_{n=1}^{N} |\alpha_n\rangle \langle\alpha_n|f\rangle = \sum_{n=1}^{N} |n\rangle \langle n|f\rangle. \tag{1.133}$$

Example 1.22 (Inner-product rules) In Dirac's notation, the rules (1.73–1.76) of a positive-definite inner product are

$$\langle f|g\rangle = \langle g|f\rangle^* \tag{1.134}$$

$$\langle f|z_1 g_1 + z_2 g_2\rangle = z_1\langle f|g_1\rangle + z_2\langle f|g_2\rangle \tag{1.135}$$

$$\langle z_1 f_1 + z_2 f_2|g\rangle = z_1^*\langle f_1|g\rangle + z_2^*\langle f_2|g\rangle \tag{1.136}$$

$$\langle f|f\rangle \geq 0 \quad \text{and} \quad \langle f|f\rangle = 0 \iff f = 0. \tag{1.137}$$

Usually states in Dirac notation are labeled $|\psi\rangle$ or by their quantum numbers $|n, l, m\rangle$, and one rarely sees plus signs or complex numbers or operators inside bras or kets. But one should. □

Example 1.23 (Gram–Schmidt) In Dirac notation, the formula (1.117) for the kth orthogonal linear combination of the vectors $|V_\ell\rangle$ is

$$|u_k\rangle = |V_k\rangle - \sum_{i=1}^{k-1} |U_i\rangle\langle U_i|V_k\rangle = \left(I - \sum_{i=1}^{k-1} |U_i\rangle\langle U_i|\right)|V_k\rangle \tag{1.138}$$

and the formula (1.118) for the kth orthonormal linear combination of the vectors $|V_\ell\rangle$ is

$$|U_k\rangle = \frac{|u_k\rangle}{\sqrt{\langle u_k|u_k\rangle}}. \tag{1.139}$$

The vectors $|U_k\rangle$ are not unique; they vary with the order of the $|V_k\rangle$. □

Vectors and linear operators are abstract. The numbers we compute with are inner products like $\langle g|f\rangle$ and $\langle g|A|f\rangle$. In terms of N orthonormal basis vectors $|n\rangle$ with $f_n = \langle n|f\rangle$ and $g_n^* = \langle g|n\rangle$, we can use the expansion (1.131) to write these inner products as

$$\langle g|f\rangle = \langle g|I|f\rangle = \sum_{n=1}^{N} \langle g|n\rangle\langle n|f\rangle = \sum_{n=1}^{N} g_n^* f_n,$$

$$\langle g|A|f\rangle = \langle g|IAI|f\rangle = \sum_{n,\ell=1}^{N} \langle g|n\rangle\langle n|A|\ell\rangle\langle\ell|f\rangle = \sum_{n,\ell=1}^{N} g_n^* A_{n\ell} f_\ell \tag{1.140}$$

in which $A_{n\ell} = \langle n|A|\ell\rangle$. We often gather the inner products $f_\ell = \langle\ell|f\rangle$ into a column vector f with components $f_\ell = \langle\ell|f\rangle$

$$f = \begin{pmatrix} \langle 1|f\rangle \\ \langle 2|f\rangle \\ \vdots \\ \langle N|f\rangle \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_3 \end{pmatrix} \tag{1.141}$$

and the $\langle n|A|\ell\rangle$ into a matrix A with matrix elements $A_{n\ell} = \langle n|A|\ell\rangle$. If we also line up the inner products $\langle g|n\rangle = \langle g|n\rangle^*$ in a row vector that is the transpose of the complex conjugate of the column vector g

$$g^\dagger = \left(\langle 1|g\rangle^*, \langle 2|g\rangle^*, \ldots, \langle N|g\rangle^*\right) = \left(g_1^*, g_2^*, \ldots, g_N^*\right) \tag{1.142}$$

then we can write inner products in matrix notation as $\langle g|f\rangle = g^\dagger f$ and as $\langle g|A|f\rangle = g^\dagger A f$.

If we switch to a different basis, say from $|n\rangle$s to $|\alpha_n\rangle$s, then the components of the column vectors change from $f_n = \langle n|f\rangle$ to $f'_n = \langle\alpha_n|f\rangle$, and similarly those of the row vectors $g^\dagger$ and of the matrix A change, but the bras, the kets, the linear operators, and the inner products $\langle g|f\rangle$ and $\langle g|A|f\rangle$ do not change because the identity operator is basis independent (1.132)

$$\langle g|f\rangle = \sum_{n=1}^{N} \langle g|n\rangle\langle n|f\rangle = \sum_{n=1}^{N} \langle g|\alpha_n\rangle\langle\alpha_n|f\rangle,$$

$$\langle g|A|f\rangle = \sum_{n,\ell=1}^{N} \langle g|n\rangle\langle n|A|\ell\rangle\langle\ell|f\rangle = \sum_{n,\ell=1}^{N} \langle g|\alpha_n\rangle\langle\alpha_n|A|\alpha_\ell\rangle\langle\alpha_\ell|f\rangle. \tag{1.143}$$

Dirac's outer products show how to change from one basis to another. The sum of outer products

$$U = \sum_{n=1}^{N} |\alpha_n\rangle\langle n| \tag{1.144}$$

maps the ket $|\ell\rangle$ of the orthonormal basis we started with into $|\alpha_\ell\rangle$

$$U|\ell\rangle = \sum_{n=1}^{N} |\alpha_n\rangle\langle n|\ell\rangle = \sum_{n=1}^{N} |\alpha_n\rangle\,\delta_{n\ell} = |\alpha_\ell\rangle. \tag{1.145}$$

Example 1.24 (A simple change of basis) If the ket $|\alpha_n\rangle$ of the new basis is simply $|\alpha_n\rangle = |n+1\rangle$ with $|\alpha_N\rangle = |N+1\rangle \equiv |1\rangle$ then the operator that maps the N kets $|n\rangle$ into the kets $|\alpha_n\rangle$ is

$$U = \sum_{n=1}^{N} |\alpha_n\rangle\langle n| = \sum_{n=1}^{N} |n+1\rangle\langle n|. \tag{1.146}$$

The square U^2 of U also changes the basis; it sends $|n\rangle$ to $|n+2\rangle$. The set of operators U^k for $k = 1, 2, \ldots, N$ forms a group known as Z_N. □

1.13 The adjoint of an operator

In Dirac's notation, the most general linear operator on an N-dimensional vector space is a sum of dyadics like $z\,|n\rangle\langle\ell|$ in which z is a complex number and the kets $|n\rangle$ and $|\ell\rangle$ are two of the N orthonormal kets that make up a basis for the space. The **adjoint** of this basic linear operator is

$$(z\,|n\rangle\langle\ell|)^\dagger = z^*\,|\ell\rangle\langle n|. \tag{1.147}$$

Thus with $z = \langle n|A|\ell\rangle$, the most general linear operator on the space is

$$A = IAI = \sum_{n,\ell=1}^{N} |n\rangle\langle n|A|\ell\rangle\langle\ell| \tag{1.148}$$

and its adjoint $A^\dagger$ is the operator $IA^\dagger I$

$$A^\dagger = \sum_{n,\ell=1}^{N} |n\rangle\langle n|A^\dagger|\ell\rangle\langle\ell| = \sum_{n,\ell=1}^{N} |\ell\rangle\langle n|A|\ell\rangle^*\langle n| = \sum_{n,\ell=1}^{N} |n\rangle\langle\ell|A|n\rangle^*\langle\ell|.$$

It follows that $\langle n|A^\dagger|\ell\rangle = \langle\ell|A|n\rangle^*$ so that the matrix $A^\dagger_{n\ell}$ that represents $A^\dagger$ in this basis is

$$A^\dagger_{n\ell} = \langle n|A^\dagger|\ell\rangle = \langle\ell|A|n\rangle^* = A^*_{\ell n} = A^{*\mathrm{T}}_{n\ell} \tag{1.149}$$

in agreement with our definition (1.28) of the adjoint of a matrix as the transpose of its complex conjugate, $A^\dagger = A^{*\mathrm{T}}$. We also have

$$\langle g|A^\dagger f\rangle = \langle g|A^\dagger|f\rangle = \langle f|A|g\rangle^* = \langle f|Ag\rangle^* = \langle Ag|f\rangle. \tag{1.150}$$

Taking the adjoint of the adjoint is by (1.147)

$$\left[(z\,|n\rangle\langle\ell|)^\dagger\right]^\dagger = \left[z^*\,|\ell\rangle\langle n|\right]^\dagger = z\,|n\rangle\langle\ell| \tag{1.151}$$

the same as doing nothing at all. This also follows from the matrix formula (1.149) because both $(A^*)^* = A$ and $(A^\mathrm{T})^\mathrm{T} = A$, and so

$$\left(A^\dagger\right)^\dagger = \left(A^{*\mathrm{T}}\right)^{*\mathrm{T}} = A, \tag{1.152}$$

the adjoint of the adjoint of a matrix is the original matrix.

Before Dirac, the adjoint $A^\dagger$ of a linear operator A was defined by

$$(g, A^\dagger f) = (Ag, f) = (f, Ag)^*. \tag{1.153}$$

This definition also implies that $A^{\dagger\dagger} = A$ since

$$(g, A^{\dagger\dagger} f) = (A^\dagger g, f) = (f, A^\dagger g)^* = (Af, g)^* = (g, Af). \tag{1.154}$$

We also have $(g, Af) = (g, A^{\dagger\dagger} f) = (A^\dagger g, f)$.

1.14 Self-adjoint or hermitian linear operators

An operator A that is equal to its adjoint, $A^\dagger = A$, is **self adjoint** or **hermitian**. In view of (1.149), the matrix elements of a self-adjoint linear operator A satisfy $\langle n|A^\dagger|\ell\rangle = \langle\ell|A|n\rangle^* = \langle n|A|\ell\rangle$ in any orthonormal basis. So a matrix that represents a hermitian operator is equal to the transpose of its complex conjugate

$$A_{n\ell} = \langle n|A|\ell\rangle = \langle n|A^\dagger|\ell\rangle = \langle\ell|A|n\rangle^* = A^{*T}_{n\ell} = A^\dagger_{n\ell}. \tag{1.155}$$

We also have

$$\langle g|\,A\,|f\rangle = \langle Ag|f\rangle = \langle f|Ag\rangle^* = \langle f|\,A\,|g\rangle^* \tag{1.156}$$

and in pre-Dirac notation

$$(g, Af) = (Ag, f) = (f, Ag)^*. \tag{1.157}$$

A matrix A_{ij} that is **real and symmetric** or **imaginary and antisymmetric** is hermitian. But a self-adjoint linear operator A that is represented by a matrix A_{ij} that is real and symmetric (or imaginary and antisymmetric) in one orthonormal basis will not in general be represented by a matrix that is real and symmetric (or imaginary and antisymmetric) in a different orthonormal basis, but it will be represented by a hermitian matrix in every orthonormal basis.

A ket $|a'\rangle$ is an **eigenvector** of a linear operator A with **eigenvalue** a' if $A|a'\rangle = a'|a'\rangle$. As we'll see in section 1.28, hermitian matrices have real eigenvalues and complete sets of orthonormal eigenvectors. Hermitian operators and matrices represent physical variables in quantum mechanics.

1.15 Real, symmetric linear operators

In quantum mechanics, we usually consider complex vector spaces, that is, spaces in which the vectors $|f\rangle$ are complex linear combinations

$$|f\rangle = \sum_{i=1}^{N} z_i\,|i\rangle \tag{1.158}$$

of complex orthonormal basis vectors $|i\rangle$.

But real vector spaces also are of interest. A real vector space is a vector space in which the vectors $|f\rangle$ are real linear combinations

$$|f\rangle = \sum_{n=1}^{N} x_n |n\rangle \tag{1.159}$$

of real orthonormal basis vectors, $x_n^* = x_n$ and $|n\rangle^* = |n\rangle$.

A real linear operator A on a real vector space

$$A = \sum_{n,m=1}^{N} |n\rangle\langle n|A|m\rangle\langle m| = \sum_{n,m=1}^{N} |n\rangle A_{nm}\langle m| \tag{1.160}$$

is represented by a real matrix $A_{nm}^* = A_{nm}$. A real linear operator A that is self adjoint on a real vector space satisfies the condition (1.157) of hermiticity but with the understanding that complex conjugation has no effect

$$(g, Af) = (Ag, f) = (f, Ag)^* = (f, Ag). \tag{1.161}$$

Thus its matrix elements are symmetric, $\langle g|A|f\rangle = \langle f|A|g\rangle$. Since A is hermitian as well as real, the matrix A_{nm} that represents it (in a real basis) is real and hermitian, and so is symmetric $A_{nm} = A_{nm}^* = A_{mn}$.

1.16 Unitary operators

A **unitary operator** U is one whose adjoint is its inverse

$$U U^\dagger = U^\dagger U = I. \tag{1.162}$$

Any operator that changes from one orthonormal basis $|n\rangle$ to another $|\alpha_n\rangle$

$$U = \sum_{n=1}^{N} |\alpha_n\rangle\langle n| \tag{1.163}$$

is unitary since

$$\begin{aligned} UU^\dagger &= \sum_{n=1}^{N} |\alpha_n\rangle\langle n| \sum_{m=1}^{N} |m\rangle\langle\alpha_m| = \sum_{n,m=1}^{N} |\alpha_n\rangle\langle n|m\rangle\langle\alpha_m| \\ &= \sum_{n,m=1}^{N} |\alpha_n\rangle\delta_{n,m}\langle\alpha_m| = \sum_{n=1}^{N} |\alpha_n\rangle\langle\alpha_n| = I \end{aligned} \tag{1.164}$$

as well as

$$U^\dagger U = \sum_{m=1}^{N} |m\rangle\langle\alpha_m| \sum_{n=1}^{N} |\alpha_n\rangle\langle n| = \sum_{n=1}^{N} |n\rangle\langle n| = I. \tag{1.165}$$

A unitary operator maps any orthonormal basis $|n\rangle$ into another orthonormal basis $|\alpha_n\rangle$. For if $|\alpha_n\rangle = U|n\rangle$, then $\langle\alpha_n|\alpha_m\rangle = \delta_{n,m}$ (exercise 1.22). If we multiply the relation $|\alpha_n\rangle = U|n\rangle$ by the bra $\langle n|$ and then sum over the index n, we get

$$\sum_{n=1}^{N} |\alpha_n\rangle\langle n| = \sum_{n=1}^{N} U|n\rangle\langle n| = U\sum_{n=1}^{N} |n\rangle\langle n| = U. \tag{1.166}$$

Every unitary operator is a basis-changing operator, and *vice versa*.

Inner products do not change under unitary transformations because $\langle g|f\rangle = \langle g|U^\dagger U|f\rangle = \langle Ug|U|f\rangle = \langle Ug|Uf\rangle$, which in pre-Dirac notation is $(g,f) = (g, U^\dagger Uf) = (Ug, Uf)$.

Unitary matrices have unimodular determinants because the determinant of the product of two matrices is the product of their determinants (1.204) and because transposition doesn't change the value of a determinant (1.194)

$$1 = |I| = |UU^\dagger| = |U||U^\dagger| = |U||U^\mathsf{T}|^* = |U||U|^*. \tag{1.167}$$

A unitary matrix that is real is **orthogonal** and satisfies

$$OO^\mathsf{T} = O^\mathsf{T}O = I. \tag{1.168}$$

1.17 Hilbert space

We have mostly been talking about linear operators that act on finite-dimensional vector spaces and that can be represented by matrices. But infinite-dimensional vector spaces and the linear operators that act on them play central roles in electrodynamics and quantum mechanics. For instance, the Hilbert space $\mathcal{H}$ of all "wave" functions $\psi(\boldsymbol{x}, t)$ that are square integrable over three-dimensional space at all times t is of infinite dimension.

In one space dimension, the state $|x'\rangle$ represents a particle at position x' and is an eigenstate of the hermitian position operator x with eigenvalue x', that is, $x|x'\rangle = x'|x'\rangle$. These states form a basis that is orthogonal in the sense that $\langle x|x'\rangle = 0$ for $x \neq x'$ and normalized in the sense that $\langle x|x'\rangle = \delta(x - x')$ in which $\delta(x - x')$ is Dirac's delta function. The delta function $\delta(x - x')$ actually is a **functional** $\delta_{x'}$ that maps any suitably smooth function f into

$$\delta_{x'}[f] = \int \delta(x - x')f(x)\,dx = f(x'), \tag{1.169}$$

its value at x'.

Another basis for the Hilbert space of one-dimensional quantum mechanics is made of the states $|p\rangle$ of well-defined momentum. The state $|p'\rangle$ represents a particle or system with momentum p'. It is an eigenstate of the hermitian

momentum operator p with eigenvalue p', that is, $p|p'\rangle = p'|p'\rangle$. The momentum states also are orthonormal in Dirac's sense, $\langle p|p'\rangle = \delta(p - p')$.

The operator that translates a system in space by a distance a is

$$U(a) = \int |x + a\rangle\langle x|\, dx. \tag{1.170}$$

It maps the state $|x'\rangle$ to the state $|x' + a\rangle$ and is unitary (exercise 1.23). Remarkably, this translation operator is an exponential of the momentum operator $U(a) = \exp(-i\,p\,a/\hbar)$ in which $\hbar = h/2\pi = 1.054 \times 10^{-34}$ Js is Planck's constant divided by 2π.

In two dimensions, with basis states $|x, y\rangle$ that are orthonormal in Dirac's sense, $\langle x, y|x', y'\rangle = \delta(x - x')\delta(y - y')$, the unitary operator

$$U(\theta) = \int |x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta\rangle\langle x, y|\, dxdy \tag{1.171}$$

rotates a system in space by the angle θ. This rotation operator is the exponential $U(\theta) = \exp(-i\,\theta\,L_z/\hbar)$ in which the z component of the angular momentum is $L_z = x\,p_y - y\,p_x$.

We may carry most of our intuition about matrices over to these unitary transformations that change from one infinite basis to another. But we must use common sense and keep in mind that infinite sums and integrals do not always converge.

1.18 Antiunitary, antilinear operators

Certain maps on states $|\psi\rangle \to |\psi'\rangle$, such as those involving time reversal, are implemented by operators K that are **antilinear**

$$K(z\psi + w\phi) = K(z|\psi\rangle + w|\phi\rangle) = z^*K|\psi\rangle + w^*K|\phi\rangle = z^*K\psi + w^*K\phi \tag{1.172}$$

and **antiunitary**

$$(K\phi, K\psi) = \langle K\phi|K\psi\rangle = (\phi, \psi)^* = \langle\phi|\psi\rangle^* = \langle\psi|\phi\rangle = (\psi, \phi)\,. \tag{1.173}$$

In Dirac notation, these rules are $K(z|\psi\rangle) = z^*\langle\psi|$ and $K(w\langle\phi|) = w^*|\phi\rangle$.

1.19 Symmetry in quantum mechanics

In quantum mechanics, a symmetry is a map of states $|\psi\rangle \to |\psi'\rangle$ and $|\phi\rangle \to |\phi'\rangle$ that preserves inner products and probabilities

$$|\langle\phi'|\psi'\rangle|^2 = |\langle\phi|\psi\rangle|^2. \tag{1.174}$$

Eugene Wigner (1902–1995) showed that every symmetry in quantum mechanics can be represented either by an operator U that is linear and unitary or by an operator K that is antilinear and antiunitary. The antilinear, antiunitary case seems to occur only when the symmetry involves time reversal. Most symmetries are represented by operators that are linear and unitary. Unitary operators are of great importance in quantum mechanics. We use them to represent rotations, translations, Lorentz transformations, and internal-symmetry transformations.

1.20 Determinants

The **determinant** of a 2×2 matrix A is

$$\det A = |A| = A_{11}A_{22} - A_{21}A_{12}. \tag{1.175}$$

In terms of the 2×2 antisymmetric ($e_{ij} = -e_{ji}$) matrix $e_{12} = 1 = -e_{21}$ with $e_{11} = e_{22} = 0$, this determinant is

$$\det A = \sum_{i=1}^{2}\sum_{j=1}^{2} e_{ij}A_{i1}A_{j2}. \tag{1.176}$$

It's also true that

$$e_{k\ell}\det A = \sum_{i=1}^{2}\sum_{j=1}^{2} e_{ij}A_{ik}A_{j\ell}. \tag{1.177}$$

These definitions and results extend to any square matrix. If A is a 3×3 matrix, then its determinant is

$$\det A = \sum_{i,j,k=1}^{3} e_{ijk}A_{i1}A_{j2}A_{k3} \tag{1.178}$$

in which e_{ijk} is totally antisymmetric with $e_{123} = 1$, and the sums over i, j, and k run from 1 to 3. More explicitly, this determinant is

$$\begin{aligned}\det A &= \sum_{i,j,k=1}^{3} e_{ijk}A_{i1}A_{j2}A_{k3} \\ &= \sum_{i=1}^{3} A_{i1}\sum_{j,k=1}^{3} e_{ijk}A_{j2}A_{k3} \\ &= A_{11}(A_{22}A_{33} - A_{32}A_{23}) + A_{21}(A_{32}A_{13} - A_{12}A_{33}) \\ &\quad + A_{31}(A_{12}A_{23} - A_{22}A_{13})\,.\end{aligned} \tag{1.179}$$

The terms within parentheses are the 2×2 determinants (called **minors**) of the matrix A without column 1 and row i, multiplied by $(-1)^{1+i}$:

$$\begin{aligned}\det A &= A_{11}(-1)^2 (A_{22}A_{33} - A_{32}A_{23}) + A_{21}(-1)^3 (A_{12}A_{33} - A_{32}A_{13}) \\ &\quad + A_{31}(-1)^4 (A_{12}A_{23} - A_{22}A_{13}) \\ &= A_{11}C_{11} + A_{21}C_{21} + A_{31}C_{31}\end{aligned} \tag{1.180}$$

The minors multiplied by $(-1)^{1+i}$ are called **cofactors**:

$$\begin{aligned}C_{11} &= A_{22}A_{33} - A_{23}A_{32}, \\ C_{21} &= A_{32}A_{13} - A_{12}A_{33}, \\ C_{31} &= A_{12}A_{23} - A_{22}A_{13}.\end{aligned} \tag{1.181}$$

Example 1.25 (Determinant of a 3×3 matrix) The determinant of a 3×3 matrix is the dot-product of the vector of its first row with the cross-product of the vectors of its second and third rows:

$$\begin{vmatrix} U_1 & U_2 & U_3 \\ V_1 & V_2 & V_3 \\ W_1 & W_2 & W_3 \end{vmatrix} = \sum_{i,j,k=1}^{3} e_{ijk} U_i V_j W_k = \sum_{i=1}^{3} U_i (V \times W)_i = \boldsymbol{U} \cdot (\boldsymbol{V} \times \boldsymbol{W})$$

which is called the scalar triple product. □

Laplace used the totally antisymmetric symbol $e_{i_1 i_2 \ldots i_N}$ with N indices and with $e_{123\ldots N} = 1$ to define the determinant of an $N \times N$ matrix A as

$$\det A = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} e_{i_1 i_2 \ldots i_N} A_{i_1 1} A_{i_2 2} \ldots A_{i_N N} \tag{1.182}$$

in which the sums over $i_1 \ldots i_N$ run from 1 to N. In terms of cofactors, two forms of his expansion of this determinant are

$$\det A = \sum_{i=1}^{N} A_{ik} C_{ik} = \sum_{k=1}^{N} A_{ik} C_{ik} \tag{1.183}$$

in which the first sum is over the row index i but not the (arbitrary) column index k, and the second sum is over the column index k but not the (arbitrary) row index i. The cofactor C_{ik} is $(-1)^{i+k} M_{ik}$ in which the minor M_{ik} is the determinant of the $(N-1) \times (N-1)$ matrix A without its ith row and kth column. It's also true that

$$e_{k_1 k_2 \ldots k_N} \det A = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} e_{i_1 i_2 \ldots i_N} A_{i_1 k_1} A_{i_2 k_2} \ldots A_{i_N k_N}. \tag{1.184}$$

The key feature of a determinant is that it is an *antisymmetric* combination of products of the elements A_{ik} of a matrix A. One implication of this antisymmetry is that the interchange of any two rows or any two columns changes the sign of the determinant. Another is that if one adds a multiple of one column to another column, for example a multiple xA_{i2} of column 2 to column 1, then the determinant

$$\det A' = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} e_{i_1 i_2 \ldots i_N} \left(A_{i_1 1} + x A_{i_1 2}\right) A_{i_2 2} \ldots A_{i_N N} \tag{1.185}$$

is unchanged. The reason is that the extra term $\delta \det A$ vanishes

$$\delta \det A = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} x\, e_{i_1 i_2 \ldots i_N} A_{i_1 2} A_{i_2 2} \ldots A_{i_N N} = 0 \tag{1.186}$$

because it is proportional to a sum of products of a factor $e_{i_1 i_2 \ldots i_N}$ that is antisymmetric in i_1 and i_2 and a factor $A_{i_1 2} A_{i_2 2}$ that is symmetric in these indices. For instance, when i_1 and i_2 are 5 & 7 and 7 & 5, the two terms cancel

$$e_{57\ldots i_N} A_{52} A_{72} \ldots A_{i_N N} + e_{75\ldots i_N} A_{72} A_{52} \ldots A_{i_N N} = 0 \tag{1.187}$$

because $e_{57\ldots i_N} = -e_{75\ldots i_N}$.

By repeated additions of $x_2 A_{i2}$, $x_3 A_{i3}$, etc. to A_{i1}, we can change the first column of the matrix A to a linear combination of all the columns

$$A_{i1} \longrightarrow A_{i1} + \sum_{k=2}^{N} x_k A_{ik} \tag{1.188}$$

without changing $\det A$. In this linear combination, the coefficients x_k are arbitrary. The analogous operation with arbitrary y_k

$$A_{i\ell} \longrightarrow A_{i\ell} + \sum_{k=1, k \neq \ell}^{N} y_k A_{ik} \tag{1.189}$$

replaces the ℓth column by a linear combination of all the columns without changing $\det A$.

Suppose that the columns of an $N \times N$ matrix A are linearly dependent (section 1.8), so that the linear combination of columns

$$\sum_{k=1}^{N} y_k A_{ik} = 0 \quad \text{for } i = 1, \ldots N \tag{1.190}$$

vanishes for some coefficients y_k not all zero. Suppose $y_1 \neq 0$. Then by adding suitable linear combinations of columns 2 through N to column 1, we could make all the modified elements A'_{i1} of column 1 vanish without changing $\det A$.

But then $\det A$ as given by (1.182) would vanish. **Thus the determinant of any matrix whose columns are linearly dependent must vanish.**

The converse also is true: if columns of a matrix are linearly independent, then the determinant of that matrix can not vanish. The reason is that any linearly independent set of vectors is complete (section 1.8). Thus if the columns of a matrix A are linearly independent and therefore complete, some linear combination of all columns 2 through N when added to column 1 will convert column 1 into a (nonzero) multiple of the N-dimensional column vector $(1, 0, 0, \ldots 0)$, say $(c_1, 0, 0, \ldots 0)$. Similar operations will convert column 2 into a (nonzero) multiple of the column vector $(0, 1, 0, \ldots 0)$, say $(0, c_2, 0, \ldots 0)$. Continuing in this way, we may convert the matrix A to a matrix with nonzero entries along the main diagonal and zeros everywhere else. The determinant $\det A$ is then the product of the nonzero diagonal entries $c_1 c_2 \ldots c_N \neq 0$, and so $\det A$ can not vanish.

We may extend these arguments to the rows of a matrix. The addition to row k of a linear combination of the other rows

$$A_{ki} \longrightarrow A_{ki} + \sum_{\ell=1, \ell \neq k}^{N} z_\ell A_{\ell i} \tag{1.191}$$

does not change the value of the determinant. In this way, one may show that the determinant of a matrix vanishes if and only if its rows are linearly dependent. The reason why these results apply to the rows as well as to the columns is that the determinant of a matrix A may be defined either in terms of the columns as in definitions (1.182 & 1.184) or in terms of the rows:

$$\det A = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} e_{i_1 i_2 \ldots i_N} A_{1 i_1} A_{2 i_2} \ldots A_{N i_N}, \tag{1.192}$$

$$e_{k_1 k_2 \ldots k_N} \det A = \sum_{i_1, i_2, \ldots, i_N = 1}^{N} e_{i_1 i_2 \ldots i_N} A_{k_1 i_1} A_{k_2 i_2} \ldots A_{k_N i_N}. \tag{1.193}$$

These and other properties of determinants follow from a study of **permutations** (section 10.13). Detailed proofs are in Aitken (1959).

By comparing the row (1.182 & 1.184) and column (1.192 & 1.193) definitions of determinants, we see that the determinant of the transpose of a matrix is the same as the determinant of the matrix itself:

$$\det \left(A^{\mathsf{T}}\right) = \det A. \tag{1.194}$$

Let us return for a moment to Laplace's expansion (1.183) of the determinant $\det A$ of an $N \times N$ matrix A as a sum of $A_{ik} C_{ik}$ over the row index i with the column index k held fixed

$$\det A = \sum_{i=1}^{N} A_{ik} C_{ik} \tag{1.195}$$

in order to prove that

$$\delta_{k\ell} \det A = \sum_{i=1}^{N} A_{ik} C_{i\ell}. \tag{1.196}$$

For $k = \ell$, this formula just repeats Laplace's expansion (1.195). But for $k \neq \ell$, it is Laplace's expansion for the determinant of a matrix A' that is the same as A but with its ℓth column replaced by its kth one. Since the matrix A' has two identical columns, its determinant vanishes, which explains (1.196) for $k \neq \ell$.

This rule (1.196) provides a formula for the inverse of a matrix A whose determinant does not vanish. Such matrices are said to be **nonsingular**. The inverse A^{-1} of an $N \times N$ nonsingular matrix A is the transpose of the matrix of cofactors divided by $\det A$

$$\left(A^{-1}\right)_{\ell i} = \frac{C_{i\ell}}{\det A} \quad \text{or} \quad A^{-1} = \frac{C^{\mathsf{T}}}{\det A}. \tag{1.197}$$

To verify this formula, we use it for A^{-1} in the product $A^{-1}A$ and note that by (1.196) the ℓkth entry of the product $A^{-1}A$ is just $\delta_{\ell k}$

$$\left(A^{-1}A\right)_{\ell k} = \sum_{i=1}^{N} \left(A^{-1}\right)_{\ell i} A_{ik} = \sum_{i=1}^{N} \frac{C_{i\ell}}{\det A} A_{ik} = \delta_{\ell k}. \tag{1.198}$$

Example 1.26 (Inverting a 2×2 matrix) Let's apply our formula (1.197) to find the inverse of the general 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{1.199}$$

We find then

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{1.200}$$

which is the correct inverse as long as $ad \neq bc$. □

The simple example of matrix multiplication

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a & xa+b & ya+zb+c \\ d & xd+e & yd+ze+f \\ g & xg+h & yg+zh+i \end{pmatrix} \tag{1.201}$$

shows that the operations (1.189) on columns that don't change the value of the determinant can be written as matrix multiplication from the right by a matrix

that has unity on its main diagonal and zeros below. Now consider the matrix product

$$\begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix} \tag{1.202}$$

in which A and B are $N \times N$ matrices, I is the $N \times N$ identity matrix, and 0 is the $N \times N$ matrix of all zeros. The second matrix on the left-hand side has unity on its main diagonal and zeros below, and so it does not change the value of the determinant of the matrix to its left, which then must equal that of the matrix on the right-hand side:

$$\det \begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} = \det \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix}. \tag{1.203}$$

By using Laplace's expansion (1.183) along the first column to evaluate the determinant on the left-hand side and his expansion along the last row to compute the determinant on the right-hand side, one finds that **the determinant of the product of two matrices is the product of the determinants**

$$\det A \det B = \det AB. \tag{1.204}$$

Example 1.27 (Two 2×2 matrices) When the matrices A and B are both 2×2, the two sides of (1.203) are

$$\det \begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{pmatrix}$$
$$= a_{11}a_{22} \det B - a_{21}a_{12} \det B = \det A \det B \tag{1.205}$$

and

$$\det \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} & ab_{11} & ab_{12} \\ a_{21} & a_{22} & ab_{21} & ab_{22} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$
$$= (-1)C_{42} = (-1)(-1) \det AB = \det AB \tag{1.206}$$

and so they give the product rule $\det A \det B = \det AB$. □

Often one uses the notation $|A| = \det A$ to denote a determinant. In this more compact notation, the obvious generalization of the product rule is

$$|ABC \ldots Z| = |A||B| \ldots |Z|. \tag{1.207}$$

The product rule (1.204) implies that $\det\left(A^{-1}\right)$ is $1/\det A$ since

$$1 = \det I = \det\left(AA^{-1}\right) = \det A \det\left(A^{-1}\right). \tag{1.208}$$

Incidentally, Gauss, Jordan, and modern mathematicians have developed much faster ways of computing determinants and matrix inverses than those (1.183 & 1.197) due to Laplace. Octave, Matlab, Maple, and Mathematica use these modern techniques, which also are freely available as programs in C and FORTRAN from www.netlib.org/lapack.

Example 1.28 (Numerical tricks) Adding multiples of rows to other rows does not change the value of a determinant, and interchanging two rows only changes a determinant by a minus sign. So we can use these operations, which leave determinants invariant, to make a matrix **upper triangular**, a form in which its determinant is just the product of the factors on its diagonal. For instance, to make the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -6 & 3 \\ 4 & 2 & -5 \end{pmatrix} \tag{1.209}$$

upper triangular, we add twice the first row to the second row

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 5 \\ 4 & 2 & -5 \end{pmatrix} \tag{1.210}$$

and then subtract four times the first row from the third

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 5 \\ 0 & -6 & -9 \end{pmatrix}. \tag{1.211}$$

Next, we subtract three times the second row from the third

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 5 \\ 0 & 0 & -24 \end{pmatrix}. \tag{1.212}$$

We now find as the determinant of A the product of its diagonal elements:

$$|A| = 1(-2)(-24) = 48. \tag{1.213}$$

The Matlab command is $d = \det(A)$. □

1.21 Systems of linear equations

Suppose we wish to solve the system of N linear equations

$$\sum_{k=1}^{N} A_{ik} x_k = y_i \tag{1.214}$$

for N unknowns x_k. In matrix notation, with A an $N \times N$ matrix and x and y N-vectors, this system of equations is $A\,x = y$. If the matrix A is **nonsingular**, that is, if $\det(A) \neq 0$, then it has an inverse A^{-1} given by (1.197), and we may multiply both sides of $A\,x = y$ by A^{-1} and so find x as $x = A^{-1}\,y$. When A is nonsingular, this is the unique solution to (1.214).

When A is singular, $\det(A) = 0$, and so its columns are linearly dependent (section 1.20). In this case, the linear dependence of the columns of A implies that $A\,z = 0$ for some nonzero vector z. Thus if x is a solution, so that $A\,x = y$, then $A(x + cz) = A\,x + c\,A\,z = y$ implies that $x + cz$ for all c also is a solution. So if $\det(A) = 0$, then there may be solutions, but there can be no unique solution. Whether equation (1.214) has any solutions when $\det(A) = 0$ depends on whether the vector y can be expressed as a linear combination of the columns of A. Since these columns are linearly dependent, they span a subspace of fewer than N dimensions, and so (1.214) has solutions only when the N-vector y lies in that subspace.

A system of $M < N$ equations

$$\sum_{k=1}^{N} A_{ik} x_k = y_i \quad \text{for} \quad i = 1, 2, \ldots, M \tag{1.215}$$

in N unknowns is **under-determined**. As long as at least M of the N columns A_{ik} of the matrix A are linearly independent, such a system always has solutions, but they will not be unique.

1.22 Linear least squares

Suppose we have a system of $M > N$ equations in N unknowns x_k

$$\sum_{k=1}^{N} A_{ik} x_k = y_i \quad \text{for} \quad i = 1, 2, \ldots, M. \tag{1.216}$$

This problem is **over-determined** and, in general, has no solution, but it does have an approximate solution due to Carl Gauss (1777–1855).

If the matrix A and the vector y are real, then Gauss's solution is the N values x_k that minimize the sum E of the squares of the errors

$$E = \sum_{i=1}^{M} \left(y_i - \sum_{k=1}^{N} A_{ik} x_k \right)^2 . \tag{1.217}$$

The minimizing values x_k make the N derivatives of E vanish

$$\frac{\partial E}{\partial x_\ell} = 0 = \sum_{i=1}^{M} 2 \left(y_i - \sum_{k=1}^{N} A_{ik} x_k \right) (-A_{i\ell}) \tag{1.218}$$

or in matrix notation $A^{\mathsf{T}} y = A^{\mathsf{T}} A x$. Since A is real, the matrix $A^{\mathsf{T}} A$ is nonnegative (1.38); if it also is positive (1.39), then it has an inverse, and our **least-squares solution** is

$$x = \left(A^{\mathsf{T}} A \right)^{-1} A^{\mathsf{T}} y. \tag{1.219}$$

If the matrix A and the vector y are complex, and if the matrix $A^\dagger A$ is positive, then one may show (exercise 1.25) that Gauss's solution is

$$x = \left(A^\dagger A \right)^{-1} A^\dagger y. \tag{1.220}$$

1.23 Lagrange multipliers

The maxima and minima of a function $f(x)$ of several variables $x_1, x_2, \ldots, x_n$ are among the points at which its gradient vanishes

$$\nabla f(x) = 0. \tag{1.221}$$

These are the stationary points of f.

Example 1.29 (Minimum) For instance, if $f(x) = x_1^2 + 2x_2^2 + 3x_3^2$, then its minimum is at

$$\nabla f(x) = (2x_1, 4x_2, 6x_3) = 0 \tag{1.222}$$

that is, at $x_1 = x_2 = x_3 = 0$. □

But how do we find the extrema of $f(x)$ if x must satisfy a constraint? We use a Lagrange multiplier (Joseph-Louis Lagrange, 1736–1813).

In the case of one constraint $c(x) = 0$, we no longer expect the gradient $\nabla f(x)$ to vanish, but its projection $dx \cdot \nabla f(x)$ must vanish in those directions dx that preserve the constraint. So $dx \cdot \nabla f(x) = 0$ for all dx that make the dot-product $dx \cdot \nabla c(x)$ vanish. This means that $\nabla f(x)$ and $\nabla c(x)$ must be parallel. So the extrema of $f(x)$ subject to the constraint $c(x) = 0$ satisfy two equations

$$\nabla f(x) = \lambda \nabla c(x) \quad \text{and} \quad c(x) = 0. \tag{1.223}$$

These equations define the extrema of the unconstrained function

$$L(x,\lambda) = f(x) - \lambda\, c(x) \tag{1.224}$$

of the $n+1$ variables $x, \ldots, x_n, \lambda$

$$\nabla L(x,\lambda) = \nabla f(x) - \lambda\, \nabla c(x) = 0 \quad \text{and} \quad \frac{\partial L(x,\lambda)}{\partial \lambda} = -c(x) = 0. \tag{1.225}$$

The variable λ is a **Lagrange multiplier**.

In the case of k constraints $c_1(x) = 0, \ldots, c_k(x) = 0$, the projection ∇f must vanish in those directions dx that preserve all the constraints. So $dx \cdot \nabla f(x) = 0$ for all dx that make all $dx \cdot \nabla c_j(x) = 0$ for $j = 1, \ldots, k$. The gradient ∇f will satisfy this requirement if it's a linear combination

$$\nabla f = \lambda_1 \nabla c_1 + \cdots + \lambda_k \nabla c_k \tag{1.226}$$

of the k gradients because then $dx \cdot \nabla f$ will vanish if $dx \cdot \nabla c_j = 0$ for $j = 1, \ldots, k$. The extrema also must satisfy the constraints

$$c_1(x) = 0, \ldots, c_k(x) = 0. \tag{1.227}$$

Equations (1.226 & 1.227) define the extrema of the unconstrained function

$$L(x,\lambda) = f(x) - \lambda_1\, c_1(x) + \cdots \lambda_k\, c_k(x) \tag{1.228}$$

of the $n+k$ variables x and λ

$$\nabla L(x,\lambda) = \nabla f(x) - \lambda\, \nabla c_1(x) - \cdots - \lambda\, \nabla c_k(x) = 0 \tag{1.229}$$

and

$$\frac{\partial L(x,\lambda)}{\partial \lambda_j} = -c_j(x) = 0 \quad \text{for} \quad j = 1, \ldots, k. \tag{1.230}$$

Example 1.30 (Constrained extrema and eigenvectors) Suppose we want to find the extrema of a real, symmetric quadratic form $f(x) = x^{\mathsf{T}} A\, x$ subject to the constraint $c(x) = x \cdot x - 1$, which says that the vector x is of unit length. We form the function

$$L(x,\lambda) = x^{\mathsf{T}} A\, x - \lambda\,(x \cdot x - 1) \tag{1.231}$$

and since the matrix A is real and symmetric, we find its unconstrained extrema as

$$\nabla L(x,\lambda) = 2A\, x - 2\lambda\, x = 0 \quad \text{and} \quad x \cdot x = 1. \tag{1.232}$$

The extrema of $f(x) = x^{\mathsf{T}} A\, x$ subject to the constraint $c(x) = x \cdot x - 1$ are the normalized **eigenvectors**

$$A\, x = \lambda\, x \quad \text{and} \quad x \cdot x = 1 \tag{1.233}$$

of the real, symmetric matrix A. □

1.24 Eigenvectors

If a linear operator A maps a nonzero vector $|u\rangle$ into a multiple of itself

$$A|u\rangle = \lambda |u\rangle \tag{1.234}$$

then the vector $|u\rangle$ is an **eigenvector** of A with **eigenvalue** λ. (The German adjective *eigen* means special or proper.)

If the vectors $\{|k\rangle\}$ for $k = 1, \ldots, N$ form a basis for the vector space in which A acts, then we can write the identity operator for the space as $I = |1\rangle\langle 1| + \cdots + |N\rangle\langle N|$. By inserting this formula for I twice into the eigenvector equation (1.234), we can write it as

$$\sum_{\ell=1}^{N} \langle k|A|\ell\rangle\langle \ell|u\rangle = \lambda\, \langle k|u\rangle. \tag{1.235}$$

In matrix notation, with $A_{k\ell} = \langle k|A|\ell\rangle$ and $u_\ell = \langle \ell|u\rangle$, this is $A\, u = \lambda\, u$.

Example 1.31 (Eigenvalues of an orthogonal matrix) The matrix equation

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ \pm i \end{pmatrix} = e^{\pm i\theta} \begin{pmatrix} 1 \\ \pm i \end{pmatrix} \tag{1.236}$$

tells us that the eigenvectors of this 2×2 orthogonal matrix are the 2-tuples $(1, \pm i)$ with eigenvalues $e^{\pm i\theta}$. The eigenvalues λ of a unitary (and of an orthogonal) matrix are unimodular, $|\lambda| = 1$ (exercise 1.26). □

Example 1.32 (Eigenvalues of an antisymmetric matrix) Let us consider an eigenvector equation for a matrix A that is antisymmetric

$$\sum_{k=1}^{N} A_{ik}\, u_k = \lambda\, u_i. \tag{1.237}$$

The antisymmetry $A_{ik} = -A_{ki}$ of A implies that

$$\sum_{i,k=1}^{N} u_i\, A_{ik}\, u_k = 0. \tag{1.238}$$

Thus the last two relations imply that

$$0 = \sum_{i,k=1}^{N} u_i\, A_{ik}\, u_k = \lambda \sum_{i=1}^{N} u_i^2 = 0. \tag{1.239}$$

Thus either the eigenvalue λ or the dot-product of the eigenvector with itself vanishes.

A subspace $c_\ell|u_\ell\rangle + \cdots + c_r|u_r\rangle$ spanned by any set of eigenvectors of a matrix A is left invariant by its action, that is

$$A\,(c_\ell|u_\ell\rangle + \cdots + c_r|u_r\rangle) = c_\ell\lambda_\ell|u_\ell\rangle + \cdots + c_r\lambda_r|u_r\rangle. \tag{1.240}$$

□

Eigenvectors span **invariant subspaces**.

1.25 Eigenvectors of a square matrix

Let A be an $N \times N$ matrix with complex entries A_{ik}. A vector V with N entries V_k (not all zero) is an **eigenvector** of A with **eigenvalue** λ if

$$AV = \lambda V \iff \sum_{k=1}^{N} A_{ik}V_k = \lambda V_i. \tag{1.241}$$

Every $N \times N$ matrix A has N eigenvectors $V^{(\ell)}$ and eigenvalues λ_ℓ

$$AV^{(\ell)} = \lambda_\ell V^{(\ell)} \tag{1.242}$$

for $\ell = 1 \ldots N$. To see why, we write the top equation (1.241) as

$$\sum_{k=1}^{N} (A_{ik} - \lambda\delta_{ik})\, V_k = 0 \tag{1.243}$$

or in matrix notation as $(A - \lambda I)\, V = 0$ in which I is the $N \times N$ matrix with entries $I_{ik} = \delta_{ik}$. This equation and (1.243) say that the columns of the matrix $A - \lambda I$, considered as vectors, are linearly dependent, as defined in section 1.8. We saw in section 1.20 that the columns of a matrix $A - \lambda I$ are linearly dependent if and only if the determinant $|A - \lambda I|$ vanishes. Thus a nonzero solution of the eigenvalue equation (1.241) exists if and only if the determinant

$$\det(A - \lambda I) = |A - \lambda I| = 0 \tag{1.244}$$

vanishes. This requirement that the determinant of $A - \lambda I$ vanishes is called the **characteristic equation**. For an $N \times N$ matrix A, it is a polynomial equation of the Nth degree in the unknown eigenvalue λ

$$|A - \lambda I| \equiv P(\lambda, A) = |A| + \cdots + (-1)^{N-1}\lambda^{N-1}\,\mathrm{Tr}A + (-1)^N\lambda^N = \sum_{k=0}^{N} p_k\lambda^k = 0 \tag{1.245}$$

in which $p_0 = |A|$, $p_{N-1} = (-1)^{N-1}\mathrm{Tr}A$, and $p_N = (-1)^N$. (All the p_ks are basis independent.) By the fundamental theorem of algebra (section 5.9), the characteristic equation always has N roots or solutions λ_ℓ lying somewhere in the complex plane. Thus the characteristic polynomial has the factored form

$$P(\lambda, A) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_N - \lambda). \tag{1.246}$$

For every root λ_ℓ, there is a nonzero eigenvector $V^{(\ell)}$ whose components $V_k^{(\ell)}$ are the coefficients that make the N vectors $A_{ik} - \lambda_\ell\,\delta_{ik}$ that are the columns of the matrix $A - \lambda_\ell I$ sum to zero in (1.243). Thus **every $N \times N$ matrix has N eigenvalues λ_ℓ and N eigenvectors $V^{(\ell)}$**.

The $N \times N$ diagonal matrix $D_{k\ell} = \delta_{k\ell}\,\lambda_\ell$ is the **canonical form** of the matrix A; the matrix $V_{k\ell} = V_k^{(\ell)}$ whose columns are the eigenvectors $V^{(\ell)}$ of A is the **modal matrix**; and $AV = VD$.

Example 1.33 (The canonical form of a 3×3 matrix) If in Matlab we set $A =$ [0 1 2; 3 4 5; 6 7 8] and enter $[V, D] =$ eig(A), then we get

$$V = \begin{pmatrix} 0.1648 & 0.7997 & 0.4082 \\ 0.5058 & 0.1042 & -0.8165 \\ 0.8468 & -0.5913 & 0.4082 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 13.3485 & 0 & 0 \\ 0 & -1.3485 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and one may check that $AV = VD$. □

Setting $\lambda = 0$ in the factored form (1.246) of $P(\lambda, A)$ and in the characteristic equation (1.245), we see that **the determinant of every $N \times N$ matrix is the product of its N eigenvalues**

$$P(0, A) = |A| = p_0 = \lambda_1\lambda_2\ldots\lambda_N. \tag{1.247}$$

These N roots usually are all different, and when they are, the eigenvectors $V^{(\ell)}$ are linearly independent. The first eigenvector is trivially linearly independent. Let's assume that the first $K < N$ eigenvectors are linearly independent; we'll show that the first $K+1$ eigenvectors are linearly independent. If they were linearly dependent, then there would be $K + 1$ numbers c_ℓ, not all zero, such that

$$\sum_{\ell=1}^{K+1} c_\ell\, V^{(\ell)} = 0. \tag{1.248}$$

First we multiply this equation from the left by the linear operator A and use the eigenvalue equation (1.242)

$$A\sum_{\ell=1}^{K+1} c_\ell\, V^{(\ell)} = \sum_{\ell=1}^{K+1} c_\ell\, AV^{(\ell)} = \sum_{\ell=1}^{K+1} c_\ell\,\lambda_\ell\, V^{(\ell)} = 0. \tag{1.249}$$

Now we multiply the same equation (1.248) by λ_{K+1}

$$\sum_{\ell=1}^{K+1} c_\ell\,\lambda_N\, V^{(\ell)} = 0 \tag{1.250}$$

and subtract the product (1.250) from (1.249). The terms with $\ell = K+1$ cancel leaving

$$\sum_{\ell=1}^{K} c_\ell \, (\lambda_\ell - \lambda_N) \, V^{(\ell)} = 0 \tag{1.251}$$

in which all the factors $(\lambda_\ell - \lambda_{K+1})$ are different from zero since by assumption all the eigenvalues are different. But this last equation says that the first K eigenvectors are linearly dependent, which contradicts our assumption that they were linearly independent. This contradiction tells us that **if all N eigenvectors of an $N \times N$ square matrix have different eigenvalues, then they are linearly independent.**

An eigenvalue λ that is a single root of the characteristic equation (1.245) is associated with a single eigenvector; it is called a **simple eigenvalue**. An eigenvalue λ that is an nth root of the characteristic equation is associated with n eigenvectors; it is said to be an ***n*-fold degenerate eigenvalue** or to have **algebraic multiplicity** n. Its **geometric multiplicity** is the number $n' \leq n$ of linearly independent eigenvectors with eigenvalue λ. A matrix with $n' < n$ for any eigenvalue λ is **defective**. Thus an $N \times N$ matrix with fewer than N linearly independent eigenvectors is defective.

Example 1.34 (A defective 2×2 matrix) Each of the 2×2 matrices

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{1.252}$$

has only one linearly independent eigenvector and so is defective. □

Suppose A is an $N \times N$ matrix that is not defective. We may use its N linearly independent eigenvectors $V^{(\ell)} = |\ell\rangle$ to define the columns of an $N \times N$ matrix S as $S_{k\ell} = V_k^{(\ell)}$. In terms of S, the eigenvalue equation (1.242) takes the form

$$\sum_{k=1}^{N} A_{ik} S_{k\ell} = \lambda_\ell S_{i\ell}. \tag{1.253}$$

Since the columns of S are linearly independent, the determinant of S does not vanish – the matrix S is **nonsingular** – and so its inverse S^{-1} is well defined by (1.197). So we may multiply this equation by S^{-1} and get

$$\sum_{i,k=1}^{N} \left(S^{-1}\right)_{ni} A_{ik} S_{k\ell} = \sum_{i=1}^{N} \lambda_\ell \left(S^{-1}\right)_{ni} S_{i\ell} = \lambda_n \delta_{n\ell} = \lambda_\ell \tag{1.254}$$

or in matrix notation

$$S^{-1} A S = A^{(\mathrm{d})} \tag{1.255}$$

in which $A^{(\mathrm{d})}$ is the diagonal form of the matrix A in which its eigenvalues λ_ℓ are arranged along its main diagonal with zeros elsewhere. This equation (1.255) is a **similarity transformation**. Thus **every nondefective square matrix can be diagonalized by a similarity transformation** $S^{-1}AS = A^{(\mathrm{d})}$ and can be generated from its diagonal form by the inverse $A = SA^{(\mathrm{d})}S^{-1}$ of that similarity transformation. By using the product rule (1.207), we see that the determinant of any nondefective square matrix is the product of its eigenvalues

$$|A| = |SA^{(\mathrm{d})}S^{-1}| = |S|\,|A^{(\mathrm{d})}|\,|S^{-1}| = |SS^{-1}|\,|A^{(\mathrm{d})}| = |A^{(\mathrm{d})}| = \prod_{\ell=1}^{N} \lambda_\ell, \tag{1.256}$$

which is a special case of (1.247).

1.26 A matrix obeys its characteristic equation

Every square matrix obeys its characteristic equation (1.245). That is, the characteristic equation

$$P(\lambda, A) = |A - \lambda I| = \sum_{k=0}^{N} p_k \, \lambda^k = 0 \tag{1.257}$$

remains true when the matrix A replaces the variable λ

$$P(A, A) = \sum_{k=0}^{N} p_k \, A^k = 0. \tag{1.258}$$

To see why, we use the formula (1.197) for the inverse of the matrix $A - \lambda I$

$$(A - \lambda I)^{-1} = \frac{C(\lambda, A)^{\mathsf{T}}}{|A - \lambda I|} \tag{1.259}$$

in which $C(\lambda, A)^{\mathsf{T}}$ is the transpose of the matrix of cofactors of the matrix $A - \lambda I$. Since $|A - \lambda I| = P(\lambda, A)$, we have, rearranging,

$$(A - \lambda I)\, C(\lambda, A)^{\mathsf{T}} = |A - \lambda I|\, I = P(\lambda, A)\, I. \tag{1.260}$$

The transpose of the matrix of cofactors of the matrix $A - \lambda I$ is a polynomial in λ with matrix coefficients

$$C(\lambda, A)^{\mathsf{T}} = C_0 + C_1\lambda + \cdots + C_{N-1}\lambda^{N-1}. \tag{1.261}$$

The left-hand side of equation (1.260) is then

$$\begin{aligned}(A - \lambda I)C(\lambda, A)^{\mathsf{T}} = {} & AC_0 + (AC_1 - C_0)\lambda + (AC_2 - C_1)\lambda^2 + \cdots \\ & + (AC_{N-1} - C_{N-2})\lambda^{N-1} - C_{N-1}\lambda^N.\end{aligned} \tag{1.262}$$

Equating equal powers of λ on both sides of (1.260), we have, using (1.257) and (1.262),

$$\begin{aligned} AC_0 &= p_0 I, \\ AC_1 - C_0 &= p_1 I, \\ AC_2 - C_1 &= p_2 I, \\ \cdots &= \cdots \\ AC_{N-1} - C_{N-2} &= p_{N-1} I, \\ -C_{N-1} &= p_N I. \end{aligned} \tag{1.263}$$

We now multiply from the left the first of these equations by I, the second by A, the third by $A^2, \ldots$, and the last by A^N and then add the resulting equations. All the terms on the left-hand sides cancel, while the sum of those on the right gives $P(A, A)$. Thus a square matrix A obeys its characteristic equation $0 = P(A, A)$ or

$$0 = \sum_{k=0}^{N} p_k A^k = |A| I + p_1 A + \cdots + (-1)^{N-1} (\mathrm{Tr} A) A^{N-1} + (-1)^N A^N, \tag{1.264}$$

a result known as the **Cayley–Hamilton theorem** (Arthur Cayley, 1821–1895, and William Hamilton, 1805–1865). This derivation is due to Israel Gelfand (1913–2009) (Gelfand, 1961, pp. 89–90).

Because every $N \times N$ matrix A obeys its characteristic equation, its Nth power A^N can be expressed as a linear combination of its lesser powers

$$A^N = (-1)^{N-1} \left(|A| I + p_1 A + p_2 A^2 + \cdots + (-1)^{N-1} (\mathrm{Tr} A) A^{N-1} \right). \tag{1.265}$$

For instance, the square A^2 of every 2×2 matrix is given by

$$A^2 = -|A| I + (\mathrm{Tr} A) A. \tag{1.266}$$

Example 1.35 (Spin-one-half rotation matrix) If $\boldsymbol{\theta}$ is a real 3-vector and $\boldsymbol{\sigma}$ is the 3-vector of Pauli matrices (1.32), then the square of the traceless 2×2 matrix $A = \boldsymbol{\theta} \cdot \boldsymbol{\sigma}$ is

$$(\boldsymbol{\theta} \cdot \boldsymbol{\sigma})^2 - |\boldsymbol{\theta} \cdot \boldsymbol{\sigma}| = -\begin{vmatrix} \theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & -\theta_3 \end{vmatrix} I = \theta^2 I \tag{1.267}$$

in which $\theta^2 = \boldsymbol{\theta} \cdot \boldsymbol{\theta}$. One may use this identity to show (exercise (1.28)) that

$$\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{\sigma}/2) = \cos(\theta/2) - i\hat{\boldsymbol{\theta}} \cdot \boldsymbol{\sigma} \sin(\theta/2) \tag{1.268}$$

in which $\hat{\boldsymbol{\theta}}$ is a unit 3-vector. For a spin-one-half object, this matrix represents a right-handed rotation of θ radians about the axis $\hat{\boldsymbol{\theta}}$. □

1.27 Functions of matrices

What sense can we make of a function f of an $N \times N$ matrix A and how would we compute it? One way is to use the characteristic equation (1.265) to express every power of A in terms of I, A, $\ldots$, A^{N-1} and the coefficients $p_0 = |A|, p_1, p_2, \ldots, p_{N-2}$, and $p_{N-1} = (-1)^{N-1} \mathrm{Tr} A$. Then if $f(x)$ is a polynomial or a function with a convergent power series

$$f(x) = \sum_{k=0}^{\infty} c_k x^k \tag{1.269}$$

in principle we may express $f(A)$ in terms of N functions $f_k(\boldsymbol{p})$ of the coefficients $\boldsymbol{p} \equiv (p_0, \ldots, p_{N-1})$ as

$$f(A) = \sum_{k=0}^{N-1} f_k(\boldsymbol{p}) A^k. \tag{1.270}$$

The identity (1.268) for $\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{\sigma}/2)$ is an $N = 2$ example of this technique, which can become challenging when $N > 3$.

Example 1.36 (The 3×3 rotation matrix) In exercise (1.29), one finds the characteristic equation (1.264) for the 3×3 matrix $-i\boldsymbol{\theta} \cdot \boldsymbol{J}$ in which $(J_k)_{ij} = i\epsilon_{ikj}$, and ϵ_{ijk} is totally antisymmetric with $\epsilon_{123} = 1$. The generators J_k satisfy the commutation relations $[J_i, J_j] = i\epsilon_{ijk} J_k$ in which sums over repeated indices from 1 to 3 are understood. In exercise (1.31), one uses this characteristic equation for $-i\boldsymbol{\theta} \cdot \boldsymbol{J}$ to show that the 3×3 real orthogonal matrix $\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J})$, which represents a right-handed rotation by θ radians about the axis $\hat{\boldsymbol{\theta}}$, is

$$\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J}) = \cos\theta \, I - i\hat{\boldsymbol{\theta}} \cdot \boldsymbol{J} \sin\theta + (1 - \cos\theta) \hat{\boldsymbol{\theta}} (\hat{\boldsymbol{\theta}})^{\mathrm{T}} \tag{1.271}$$

or

$$\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J})_{ij} = \delta_{ij} \cos\theta - \sin\theta \, \epsilon_{ijk} \hat{\theta}_k + (1 - \cos\theta) \hat{\theta}_i \hat{\theta}_j \tag{1.272}$$

in terms of indices. □

Direct use of the characteristic equation can become unwieldy for larger values of N. Fortunately, another trick is available if A is a nondefective square matrix, and if the power series (1.269) for $f(x)$ converges. For then A is related to its diagonal form $A^{(\mathrm{d})}$ by a similarity transformation (1.255), and we may define $f(A)$ as

$$f(A) = S f(A^{(\mathrm{d})}) S^{-1} \tag{1.273}$$

in which $f(A^{(\mathrm{d})})$ is the diagonal matrix with entries $f(a_\ell)$

$$f(A^{(\mathrm{d})}) = \begin{pmatrix} f(a_1) & 0 & 0 & \cdots \\ 0 & f(a_2) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & f(a_N) \end{pmatrix}, \tag{1.274}$$

in which $a_1, a_2, \ldots, a_N$ are the eigenvalues of the matrix A. This definition makes sense if $f(A)$ is a series in powers of A because then

$$f(A) = \sum_{n=0}^{\infty} c_n A^n = \sum_{n=0}^{\infty} c_n \left(SA^{(\mathrm{d})}S^{-1}\right)^n. \tag{1.275}$$

So since $S^{-1}S = I$, we have $\left(SA^{(\mathrm{d})}S^{-1}\right)^n = S\left(A^{(\mathrm{d})}\right)^n S^{-1}$ and thus

$$f(A) = S\left[\sum_{n=0}^{\infty} c_n \left(A^{(\mathrm{d})}\right)^n\right] S^{-1} = Sf(A^{(\mathrm{d})})S^{-1}, \tag{1.276}$$

which is (1.273).

Example 1.37 (The time-evolution operator) In quantum mechanics, the time-evolution operator is the exponential $\exp(-iHt/\hbar)$ where $H = H^\dagger$ is a hermitian linear operator, the hamiltonian (William Rowan Hamilton, 1805–1865), and $\hbar = h/(2\pi) = 1.054 \times 10^{-34}$ Js where h is constant (Max Planck, 1858–1947). As we'll see in the next section, hermitian operators are never defective, so H can be diagonalized by a similarity transformation

$$H = SH^{(\mathrm{d})}S^{-1}. \tag{1.277}$$

The diagonal elements of the diagonal matrix $H^{(\mathrm{d})}$ are the **energies** E_ℓ of the states of the system described by the hamiltonian H. The time-evolution operator $U(t)$ then is

$$U(t) = S\exp(-iH^{(\mathrm{d})}t/\hbar)\,S^{-1}. \tag{1.278}$$

For a three-state system with angular frequencies $\omega_i = E_i/\hbar$, it is

$$U(t) = S\begin{pmatrix} e^{-i\omega_1 t} & 0 & 0 \\ 0 & e^{-i\omega_2 t} & \\ 0 & 0 & e^{-i\omega_3 t} \end{pmatrix} S^{-1} \tag{1.279}$$

in which the angular frequencies are $\omega_\ell = E_\ell/\hbar$. □

Example 1.38 (Entropy) The **entropy** S of a system described by a density operator ρ is the trace

$$S = -k\,\mathrm{Tr}\,(\rho \ln \rho) \tag{1.280}$$

in which $k = 1.38 \times 10^{-23}$ J/K is the constant named after Ludwig Boltzmann (1844–1906). The density operator ρ is hermitian, nonnegative, and of unit trace.

Since ρ is hermitian, the matrix that represents it is never defective (section 1.28), and so it can be diagonalized by a similarity transformation $\rho = S\,\rho^{(\mathrm{d})}\,S^{-1}$. By (1.24), $\mathrm{Tr}ABC = \mathrm{Tr}BCA$, so we can write S as

$$S = -k\mathrm{Tr}\left(S\,\rho^{(\mathrm{d})}\,S^{-1}\,S\,\ln(\rho^{(\mathrm{d})})\,S^{-1}\right) = -k\mathrm{Tr}\left(\rho^{(\mathrm{d})}\,\ln(\rho^{(\mathrm{d})})\right). \tag{1.281}$$

A vanishing eigenvalue $\rho_k^{(\mathrm{d})} = 0$ contributes nothing to this trace since $\lim_{x\to 0} x\ln x = 0$. If the system has three states, populated with probabilities ρ_i, the elements of $\rho^{(\mathrm{d})}$, then the sum

$$\begin{aligned} S &= -k\,(\rho_1 \ln \rho_1 + \rho_2 \ln \rho_2 + \rho_3 \ln \rho_3) \\ &= k\,[\rho_1 \ln(1/\rho_1) + \rho_2 \ln(1/\rho_2) + \rho_3 \ln(1/\rho_3)] \end{aligned} \tag{1.282}$$

is its entropy. □

1.28 Hermitian matrices

Hermitian matrices have very nice properties. By definition (1.30), a hermitian matrix A is square and unchanged by hermitian conjugation $A^\dagger = A$. Since it is square, the results of section 1.25 ensure that an $N \times N$ hermitian matrix A has N eigenvectors $|n\rangle$ with eigenvalues a_n

$$A|n\rangle = a_n|n\rangle. \tag{1.283}$$

In fact, all its eigenvalues are real. To see why, we take the adjoint

$$\langle n|A^\dagger = a_n^*\langle n| \tag{1.284}$$

and use the property $A^\dagger = A$ to find

$$\langle n|A^\dagger = \langle n|A = a_n^*\langle n|. \tag{1.285}$$

We now form the inner product of both sides of this equation with the ket $|n\rangle$ and use the eigenvalue equation (1.283) to get

$$\langle n|A|n\rangle = a_n\langle n|n\rangle = a_n^*\langle n|n\rangle, \tag{1.286}$$

which (since $\langle n|n\rangle > 0$) tells us that the eigenvalues are real

$$a_n^* = a_n. \tag{1.287}$$

Since $A^\dagger = A$, the matrix elements of A between two of its eigenvectors satisfy

$$a_m^*\langle m|n\rangle = (a_m\langle n|m\rangle)^* = \langle n|A|m\rangle^* = \langle m|A^\dagger|n\rangle = \langle m|A|n\rangle = a_n\langle m|n\rangle, \tag{1.288}$$

which implies that

$$\left(a_m^* - a_n\right)\langle m|n\rangle = 0. \tag{1.289}$$

But by (1.287), the eigenvalues a_m are real, and so we have

$$(a_m - a_n)\,\langle m|n\rangle = 0, \tag{1.290}$$

which tells us that when the eigenvalues are different, the eigenvectors are orthogonal. In the absence of a symmetry, all n eigenvalues usually are different, and so the eigenvectors usually are mutually orthogonal.

When two or more eigenvectors $|n_\alpha\rangle$ of a hermitian matrix have the same eigenvalue a_n, their eigenvalues are said to be **degenerate**. In this case, any linear combination of the degenerate eigenvectors also will be an eigenvector with the same eigenvalue a_n

$$A\left(\sum_{\alpha\in D} c_\alpha |n_\alpha\rangle\right) = a_n\left(\sum_{\alpha\in D} c_\alpha |n_\alpha\rangle\right) \tag{1.291}$$

where D is the set of labels α of the eigenvectors with the same eigenvalue. If the degenerate eigenvectors $|n_\alpha\rangle$ are linearly independent, then we may use the Gramm–Schmidt procedure (1.108–1.118) to choose the coefficients c_α so as to construct degenerate eigenvectors that are orthogonal to each other and to the nondegenerate eigenvectors. We then may normalize these mutually orthogonal eigenvectors.

But two related questions arise. Are the degenerate eigenvectors $|n_\alpha\rangle$ linearly independent? And if so, what orthonormal linear combinations of them should we choose for a given physical problem? Let's consider the second question first.

We know (section 1.16) that unitary transformations preserve the orthonormality of a basis. Any unitary transformation that commutes with the matrix A

$$[A, U] = 0 \tag{1.292}$$

maps each set of orthonormal degenerate eigenvectors of A into another set of orthonormal degenerate eigenvectors of A with the same eigenvalue because

$$AU|n_\alpha\rangle = UA|n_\alpha\rangle = a_n\, U|n_\alpha\rangle. \tag{1.293}$$

So there's a huge spectrum of choices for the orthonormal degenerate eigenvectors of A with the same eigenvalue. What is the right set for a given physical problem?

A sensible way to proceed is to add to the matrix A a second hermitian matrix B multiplied by a tiny, real scale factor ϵ

$$A(\epsilon) = A + \epsilon B. \tag{1.294}$$

The matrix B must completely break whatever symmetry led to the degeneracy in the eigenvalues of A. Ideally, the matrix B should be one that represents a modification of A that is physically plausible and relevant to the problem at

hand. The hermitian matrix $A(\epsilon)$ then will have N different eigenvalues $a_n(\epsilon)$ and N orthonormal nondegenerate eigenvectors

$$A(\epsilon)|n_\beta, \epsilon\rangle = a_{n_\beta}(\epsilon)|n_\beta, \epsilon\rangle. \tag{1.295}$$

These eigenvectors $|n_\beta, \epsilon\rangle$ of $A(\epsilon)$ are orthogonal to each other

$$\langle n_\beta, \epsilon|n_{\beta'}, \epsilon\rangle = \delta_{\beta,\beta'} \tag{1.296}$$

and to the eigenvectors of $A(\epsilon)$ with other eigenvalues, and they remain so as we take the limit

$$|n_\beta\rangle = \lim_{\epsilon\to 0} |n_\beta, \epsilon\rangle. \tag{1.297}$$

We may choose them as the orthogonal degenerate eigenvectors of A. Since one always may find a crooked hermitian matrix B that breaks any particular symmetry, it follows that every $N \times N$ hermitian matrix A possesses N orthonormal eigenvectors, which are complete in the vector space in which A acts. (Any N linearly independent vectors span their N-dimensional vector space, as explained in section 1.9.)

Now let's return to the first question and again show that an $N \times N$ hermitian matrix has N orthogonal eigenvectors. To do this, we'll first show that the space of vectors orthogonal to an eigenvector $|n\rangle$ of a hermitian operator A

$$A|n\rangle = \lambda|n\rangle \tag{1.298}$$

is **invariant** under the action of A – that is, $\langle n|y\rangle = 0$ implies $\langle n|A|y\rangle = 0$. We use successively the definition of $A^\dagger$, the hermiticity of A, the eigenvector equation (1.298), the definition of the inner product, and the reality of the eigenvalues of a hermitian matrix:

$$\langle n|A|y\rangle = \langle A^\dagger n|y\rangle = \langle An|y\rangle = \langle \lambda n|y\rangle = \bar{\lambda}\langle n|y\rangle = \lambda\langle n|y\rangle = 0. \tag{1.299}$$

Thus the space of vectors orthogonal to an eigenvector of a hermitian operator A is invariant under the action of that operator.

Now a hermitian operator A acting on an N-dimensional vector space S is represented by an $N \times N$ hermitian matrix, and so it has at least one eigenvector $|1\rangle$. The subspace of S consisting of all vectors orthogonal to $|1\rangle$ is an $(N-1)$-dimensional vector space S_{N-1} that is invariant under the action of A. On this space S_{N-1}, the operator A is represented by an $(N-1) \times (N-1)$ hermitian matrix A_{N-1}. This matrix has at least one eigenvector $|2\rangle$. The subspace of S_{N-1} consisting of all vectors orthogonal to $|2\rangle$ is an $(N-2)$-dimensional vector space S_{N-2} that is invariant under the action of A. On S_{N-2}, the operator A is represented by an $(N-2) \times (N-2)$ hermitian matrix A_{N-2}, which has at least one eigenvector $|3\rangle$. By construction, the vectors $|1\rangle$, $|2\rangle$, and $|3\rangle$ are mutually orthogonal. Continuing in this way, we see that A **has** N **orthogonal eigenvectors** $|k\rangle$ for $k = 1, 2, \ldots, N$. Thus no hermitian matrix is defective.

The N orthogonal eigenvectors $|k\rangle$ of an $N \times N$ matrix A can be normalized and used to write the $N \times N$ identity operator I as

$$I = \sum_{k=1}^{N} |k\rangle\langle k|. \tag{1.300}$$

On multiplying from the left by the matrix A, we find

$$A = AI = A\sum_{k=1}^{N} |k\rangle\langle k| = \sum_{k=1}^{N} a_k |k\rangle\langle k|, \tag{1.301}$$

which is the diagonal form of the hermitian matrix A. This expansion of A as a sum over outer products of its eigenstates multiplied by their eigenvalues exhibits the possible values a_k of the physical quantity represented by the matrix A when selective, nondestructive measurements $|k\rangle\langle k|$ of the quantity A are done.

The hermitian matrix A is diagonal in the basis of its eigenstates $|k\rangle$

$$A_{kj} = \langle k|A|j\rangle = a_k \delta_{kj}. \tag{1.302}$$

But in any other basis $|\alpha_k\rangle$, the matrix A appears as

$$A_{k\ell} = \langle \alpha_k|A|\alpha_\ell\rangle = \sum_{n=1}^{N} \langle \alpha_k|n\rangle a_n \langle n|\alpha_\ell\rangle. \tag{1.303}$$

The unitary matrix $U_{kn} = \langle \alpha_k|n\rangle$ relates the matrix $A_{k\ell}$ in an arbitrary basis to its diagonal form $A = UA^{(\mathrm{d})}U^\dagger$ in which $A^{(\mathrm{d})}$ is the diagonal matrix $A^{(\mathrm{d})}_{nm} = a_n \delta_{nm}$. An arbitrary $N \times N$ hermitian matrix A can be diagonalized by a unitary transformation.

A matrix that is **real and symmetric** is hermitian; so is one that is **imaginary and antisymmetric**. A real, symmetric matrix R can be diagonalized by an **orthogonal transformation**

$$R = O\,R^{(\mathrm{d})}O^{\mathsf{T}} \tag{1.304}$$

in which the matrix O is a real unitary matrix, that is, an orthogonal matrix (1.168).

Example 1.39 (The seesaw mechanism) Suppose we wish to find the eigenvalues of the real, symmetric mass matrix

$$\mathcal{M} = \begin{pmatrix} 0 & m \\ m & M \end{pmatrix} \tag{1.305}$$

in which m is an ordinary mass and M is a huge mass. The eigenvalues μ of this hermitian mass matrix satisfy $\det(\mathcal{M} - \mu I) = \mu(\mu - M) - m^2 = 0$ with

solutions $\mu_\pm = \left(M \pm \sqrt{M^2 + 4m^2}\right)/2$. The larger mass $\mu_+ \approx M + m^2/M$ is approximately the huge mass M and the smaller mass $\mu_- \approx -m^2/M$ is very tiny. The physical mass of a fermion is the absolute value of its mass parameter, here m^2/M.

The product of the two eigenvalues is the constant $\mu_+\mu_- = \det \mathcal{M} = -m^2$ so as μ_- goes down, μ_+ must go up. In 1975, Gell-Mann, Ramond, Slansky, and Jerry Stephenson invented this "**seesaw**" mechanism as an explanation of why neutrinos have such small masses, less than 1 eV/c^2. If $mc^2 = 10$ MeV, and $\mu_- c^2 \approx 0.01$ eV, which is a plausible light-neutrino mass, then the rest energy of the huge mass would be $Mc^2 = 10^7$ GeV. This huge mass would point at new physics, beyond the standard model. Yet the small masses of the neutrinos may be related to the weakness of their interactions. □

If we return to the orthogonal transformation (1.304) and multiply column ℓ of the matrix O and row ℓ of the matrix O^{T} by $\sqrt{|R^{(\mathrm{d})}_\ell|}$, then we arrive at the **congruency transformation** of Sylvester's theorem

$$R = C\,\hat{R}^{(\mathrm{d})}C^{\mathsf{T}} \tag{1.306}$$

in which the diagonal entries $\hat{R}^{(\mathrm{d})}_\ell$ are either ± 1 or 0 because the matrices C and C^{T} have absorbed the moduli $|R^{(\mathrm{d})}_\ell|$.

Example 1.40 (Equivalence principle) If G is a real, symmetric 4×4 matrix then there's a real 4×4 matrix $D = C^{\mathsf{T}-1}$ such that

$$G_{\mathrm{d}} = D^{\mathsf{T}} G D = \begin{pmatrix} g_1 & 0 & 0 & 0 \\ 0 & g_2 & 0 & 0 \\ 0 & 0 & g_3 & 0 \\ 0 & 0 & 0 & g_4 \end{pmatrix} \tag{1.307}$$

in which the diagonal entries g_I are ± 1 or 0. Thus there's a real 4×4 matrix D that casts the real nonsingular symmetric metric g_{jk} of space-time at any given point into the diagonal metric $\eta_{j\ell}$ of flat space-time by the congruence

$$g_{\mathrm{d}} = D^{\mathsf{T}} g D = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \eta. \tag{1.308}$$

Usually one needs different Ds at different points. Since one can implement the congruence by changing coordinates, it follows that in any gravitational field, one may choose free-fall coordinates in which all physical laws take the same form as in special relativity without acceleration or gravitation at least over suitably small volumes of space-time (section 11.39). □

1.29 Normal matrices

The largest set of matrices that can be diagonalized by a unitary transformation is the set of **normal** matrices. These are square matrices that commute with their adjoints

$$[A, A^\dagger] = AA^\dagger - A^\dagger A = 0. \tag{1.309}$$

This broad class of matrices includes not only hermitian matrices but also unitary matrices since

$$[U, U^\dagger] = UU^\dagger - U^\dagger U = I - I = 0. \tag{1.310}$$

To see why a normal matrix can be diagonalized by a unitary transformation, let us consider an $N \times N$ normal matrix V which (since it is square (section 1.25)) has N eigenvectors $|n\rangle$ with eigenvalues v_n

$$(V - v_n I)\,|n\rangle = 0. \tag{1.311}$$

The square of the norm (1.80) of this vector must vanish

$$\| (V - v_n I)\,|n\rangle \|^2 = \langle n|\,(V - v_n I)^\dagger\,(V - v_n I)\,|n\rangle = 0. \tag{1.312}$$

But since V is normal, we also have

$$\langle n|\,(V - v_n I)^\dagger\,(V - v_n I)\,|n\rangle = \langle n|\,(V - v_n I)\,(V - v_n I)^\dagger\,|n\rangle. \tag{1.313}$$

So the square of the norm of the vector $\left(V^\dagger - v_n^* I\right)|n\rangle = (V - v_n I)^\dagger\,|n\rangle$ also vanishes $\| \left(V^\dagger - v_n^* I\right)|n\rangle \|^2 = 0$, which tells us that $|n\rangle$ also is an eigenvector of $V^\dagger$ with eigenvalue v_n^*

$$V^\dagger|n\rangle = v_n^*|n\rangle \quad \text{and so} \quad \langle n|V = v_n\langle n|. \tag{1.314}$$

If now $|m\rangle$ is an eigenvector of V with eigenvalue v_m

$$V|m\rangle = v_m|m\rangle = 0 \tag{1.315}$$

then we have

$$\langle n|V|m\rangle = v_m\langle n|m\rangle \tag{1.316}$$

and from (1.314)

$$\langle n|V|m\rangle = v_n\langle n|m\rangle. \tag{1.317}$$

Subtracting (1.316) from (1.317), we get

$$(v_n - v_m)\,\langle m|n\rangle = 0, \tag{1.318}$$

which shows that **any two eigenvectors of a normal matrix V with different eigenvalues are orthogonal.**

Usually, all N eigenvalues of an $N \times N$ normal matrix are different. In this case, all the eigenvectors are orthogonal and may be individually normalized.

But even when a set D of eigenvectors has the same (degenerate) eigenvalue, one may use the argument (1.291–1.297) to find a suitable set of orthonormal eigenvectors with that eigenvalue. Thus **every $N \times N$ normal matrix has N orthonormal eigenvectors.** It follows then from the argument of equations (1.300–1.303) that every $N \times N$ normal matrix V can be diagonalized by an $N \times N$ unitary matrix U

$$V = UV^{(\mathrm{d})}U^\dagger \tag{1.319}$$

whose nth column $U_{kn} = \langle\alpha_k|n\rangle$ is the eigenvector $|n\rangle$ in the arbitrary basis $|\alpha_k\rangle$ of the matrix $V_{k\ell} = \langle\alpha_k|V|\alpha_\ell\rangle$ as in (1.303).

Since the eigenstates $|n\rangle$ of a normal matrix A

$$A|n\rangle = a_n|n\rangle \tag{1.320}$$

are complete and orthonormal, we can write the identity operator I as

$$I = \sum_{n=1}^{N} |n\rangle\langle n|. \tag{1.321}$$

The product AI is A itself, so

$$A = AI = A\sum_{n=1}^{N} |n\rangle\langle n| = \sum_{n=1}^{N} a_n\,|n\rangle\langle n|. \tag{1.322}$$

It follows therefore that if f is a function, then $f(A)$ is

$$f(A) = \sum_{n=1}^{N} f(a_n)\,|n\rangle\langle n|, \tag{1.323}$$

which is simpler than the expression (1.273) for an arbitrary nondefective matrix. This is a good way to think about functions of normal matrices.

Example 1.41 How do we handle the operator $\exp(-iHt/\hbar)$ that translates states in time by t? The hamiltonian H is hermitian and so is normal. Its orthonormal eigenstates $|n\rangle$ are the energy levels E_n

$$H|n\rangle = E_n|n\rangle. \tag{1.324}$$

So we apply (1.323) with $A \to H$ and get

$$e^{-iHt/\hbar} = \sum_{n=1}^{N} e^{-iE_n t/\hbar}\,|n\rangle\langle n|, \tag{1.325}$$

which lets us compute the time evolution of any state $|\psi\rangle$ as

$$e^{-iHt/\hbar}|\psi\rangle = \sum_{n=1}^{N} e^{-iE_n t/\hbar}\,|n\rangle\langle n|\psi\rangle \tag{1.326}$$

if we know the eigenstates $|n\rangle$ and eigenvalues E_n of the hamiltonian H. □

The determinant $|V|$ of a normal matrix V satisfies the identities

$$|V| = \exp[\mathrm{Tr}(\ln V)], \quad \ln|V| = \mathrm{Tr}(\ln V), \quad \text{and} \quad \delta \ln|V| = \mathrm{Tr}\left(V^{-1}\delta V\right). \tag{1.327}$$

1.30 Compatible normal matrices

Two normal matrices A and B that **commute**

$$[A, B] \equiv AB - BA = 0 \tag{1.328}$$

are said to be **compatible**. Since these operators are normal, they have complete sets of orthonormal eigenvectors. If $|u\rangle$ is an eigenvector of A with eigenvalue z, then so is $B|u\rangle$ since

$$AB|u\rangle = BA|u\rangle = Bz|u\rangle = z\,B|u\rangle. \tag{1.329}$$

We have seen that any normal matrix A can be written as a sum (1.322) of outer products

$$A = \sum_{n=1}^{N} |a_n\rangle a_n \langle a_n| \tag{1.330}$$

of its orthonormal eigenvectors $|a_n\rangle$, which are complete in the N-dimensional vector space $\mathcal{S}$ on which A acts. Suppose now that the eigenvalues a_n of A are nondegenerate, and that B is another normal matrix acting on $\mathcal{S}$ and that the matrices A and B are compatible. Then in the basis provided by the eigenvectors (or eigenstates) $|a_n\rangle$ of the matrix A, the matrix B must satisfy

$$0 = \langle a_n|AB - BA|a_k\rangle = (a_n - a_k)\,\langle a_n|B|a_k\rangle, \tag{1.331}$$

which says that $\langle a_n|B|a_k\rangle$ is zero unless $a_n = a_k$. Thus if the eigenvalues a_n of the operator A are nondegenerate, then the operator B is diagonal

$$B = IBI = \sum_{n=1}^{N} |a_n\rangle\langle a_n|B\sum_{k=1}^{N} |a_k\rangle\langle a_k| = \sum_{n=1}^{N} |a_n\rangle\langle a_n|B|a_n\rangle\langle a_n| \tag{1.332}$$

in the $|a_n\rangle$ basis. Moreover B maps each eigenket $|a_k\rangle$ of A into

$$B|a_k\rangle = \sum_{n=1}^{N} |a_n\rangle\langle a_n|B|a_n\rangle\langle a_n|a_k\rangle = \sum_{n=1}^{N} |a_n\rangle\langle a_n|B|a_n\rangle\delta_{nk} = \langle a_k|B|a_k\rangle|a_k\rangle, \tag{1.333}$$

which says that each eigenvector $|a_k\rangle$ of the matrix A also is an eigenvector of the matrix B with eigenvalue $\langle a_k|B|a_k\rangle$. Thus **two compatible normal matrices can be simultaneously diagonalized** if one of them has nondegenerate eigenvalues.

If A's eigenvalues a_n are degenerate, each eigenvalue a_n may have d_n orthonormal eigenvectors $|a_n, k\rangle$ for $k = 1, \ldots, d_n$. In this case, the matrix elements $\langle a_n, k|B|a_m, k'\rangle$ of B are zero unless the eigenvalues are the same, $a_n = a_m$. The matrix representing the operator B in this basis consists of square, $d_n \times d_n$, normal submatrices $\langle a_n, k|B|a_n, k'\rangle$ arranged along its main diagonal; it is said to be in **block-diagonal form**. Since each submatrix is a $d_n \times d_n$ normal matrix, we may find linear combinations $|a_n, b_k\rangle$ of the degenerate eigenvectors $|a_n, k\rangle$ that are orthonormal eigenvectors of both compatible operators

$$A|a_n, b_k\rangle = a_n|a_n, b_k\rangle \quad \text{and} \quad B|a_n, b_k\rangle = b_k|a_n, b_k\rangle. \tag{1.334}$$

Thus one can simultaneously diagonalize any two compatible operators.

The converse also is true: if the operators A and B can be simultaneously diagonalized as in (1.334), then they commute

$$AB|a_n, b_k\rangle = Ab_k|a_n, b_k\rangle = a_n b_k|a_n, b_k\rangle = a_n B|a_n, b_k\rangle = BA|a_n, b_k\rangle$$

and so are compatible. Normal matrices can be simultaneously diagonalized if and only if they are compatible, that is, if and only if they commute.

In quantum mechanics, compatible hermitian operators represent physical observables that can be measured simultaneously to arbitrary precision (in principle). A set of compatible hermitian operators $\{A, B, C, \ldots\}$ is said to be **complete** if to every set of eigenvalues $\{a_n, b_k, c_\ell, \ldots\}$ there is only a single eigenvector $|a_n, b_k, c_\ell, \ldots\rangle$.

Example 1.42 (Compatible photon observables) The state of a photon is completely characterized by its momentum and its angular momentum about its direction of motion. For a photon, the momentum operator $\boldsymbol{P}$ and the dot-product $\boldsymbol{J}\cdot\boldsymbol{P}$ of the angular momentum $\boldsymbol{J}$ with the momentum form a complete set of compatible hermitian observables. Incidentally, because its mass is zero, the angular momentum $\boldsymbol{J}$ of a photon about its direction of motion can have only two values $\pm\hbar$, which correspond to its two possible states of circular polarization. □

Example 1.43 (Thermal density operator) A **density operator** ρ is the most general description of a quantum-mechanical system. It is hermitian, positive, and of unit trace. Since it is hermitian, it can be diagonalized (section 1.28)

$$\rho = \sum_n |n\rangle\langle n|\rho|n\rangle\langle n| \tag{1.335}$$

and its eigenvalues $\rho_n = \langle n|\rho|n\rangle$ are real. Each ρ_n is the probability that the system is in the state $|n\rangle$ and so is nonnegative. The unit-trace rule

$$\sum_n \rho_n = 1 \tag{1.336}$$

ensures that these probabilities add up to one – the system is in some state.

The mean value of an operator F is the trace, $\langle F\rangle = \mathrm{Tr}(\rho F)$. So the average energy E is the trace, $E = \langle H\rangle = \mathrm{Tr}(\rho H)$. The **entropy operator** S is the negative logarithm of the density operator multiplied by Boltzmann's constant $S = -k\ln\rho$, and the mean entropy S is $S = \langle S\rangle = -k\mathrm{Tr}(\rho\ln\rho)$.

A density operator that describes a system in thermal equilibrium at a constant temperature T is time independent and so commutes with the hamiltonian, $[\rho, H] = 0$. Since ρ and H commute, they are compatible operators (1.328), and so they can be simultaneously diagonalized. Each eigenstate $|n\rangle$ of ρ is an eigenstate of H; its energy E_n is its eigenvalue, $H|n\rangle = E_n|n\rangle$.

If we have no information about the state of the system other than its mean energy E, then we take ρ to be the density operator that maximizes the mean entropy S while respecting the constraints

$$c_1 = \sum_n \rho_n - 1 = 0 \quad \text{and} \quad c_2 = \mathrm{Tr}(\rho H) - E = 0. \tag{1.337}$$

We introduce two Lagrange multipliers (section 1.23) and maximize the unconstrained function

$$\begin{aligned} L(\rho, \lambda_1, \lambda_2) &= S - \lambda_1 c_1 - \lambda_2 c_2 \\ &= -k\sum_n \rho_n \ln\rho_n - \lambda_1\left[\sum_n \rho_n - 1\right] - \lambda_2\left[\sum_n \rho_n E_n - E\right] \end{aligned} \tag{1.338}$$

by setting its derivatives with respect to ρ_n, λ_1, and λ_2 equal to zero

$$\frac{\partial L}{\partial \rho_n} = -k(\ln\rho_n + 1) - \lambda_1 - \lambda_2 E_n = 0, \tag{1.339}$$

$$\frac{\partial L}{\partial \lambda_1} = \sum_n \rho_n - 1 = 0, \tag{1.340}$$

$$\frac{\partial L}{\partial \lambda_2} = \sum_n \rho_n E_n - E = 0. \tag{1.341}$$

The first (1.339) of these conditions implies that

$$\rho_n = \exp[-(\lambda_1 + \lambda_2 E_n + k)/k]. \tag{1.342}$$

We satisfy the second condition (1.340) by choosing λ_1 so that

$$\rho_n = \frac{\exp(-\lambda_2 E_n/k)}{\sum_n \exp(-\lambda_2 E_n/k)}. \tag{1.343}$$

Setting $\lambda_2 = 1/T$, we define the temperature T so that ρ satisfies the third condition (1.341). Its eigenvalue ρ_n then is

$$\rho_n = \frac{\exp(-E_n/kT)}{\sum_n \exp(-E_n/kT)}. \tag{1.344}$$

In terms of the inverse temperature $\beta \equiv 1/(kT)$, the density operator is

$$\rho = \frac{e^{-\beta H}}{\mathrm{Tr}\left(e^{-\beta H}\right)}, \tag{1.345}$$

which is the **Boltzmann distribution**. □

1.31 The singular-value decomposition

Every complex $M \times N$ rectangular matrix A is the product of an $M \times M$ unitary matrix U, an $M \times N$ rectangular matrix Σ that is zero except on its main diagonal, which consists of its nonnegative singular values S_k, and an $N \times N$ unitary matrix $V^\dagger$

$$A = U\,\Sigma\,V^\dagger. \tag{1.346}$$

This singular-value decomposition (SVD) is a key theorem of matrix algebra.

Suppose A is a linear operator that maps vectors in an N-dimensional vector space V_N into vectors in an M-dimensional vector space V_M. The spaces V_N and V_M will have infinitely many orthonormal bases $\{|n, a\rangle \in V_N\}$ and $\{|m, b\rangle \in V_M\}$ labeled by continuous parameters a and b. Each pair of bases provides a resolution of the identity operator I_N for V_N and I_M for V_M

$$I_N = \sum_{n=1}^{N} |n, a\rangle\langle n, a| \quad \text{and} \quad I_M = \sum_{m=1}^{M} |m, b\rangle\langle m, b|. \tag{1.347}$$

These identity operators give us many ways of writing the linear operator A

$$A = I_M A I_N = \sum_{m=1}^{M}\sum_{n=1}^{N} |m, b\rangle\langle m, b|A|n, a\rangle\langle n, a|, \tag{1.348}$$

in which the $\langle m, b|A|n, a\rangle$ are the elements of a complex $M \times N$ matrix. The singular-value decomposition of the linear operator A is a choice among all these expressions for I_N and I_M that expresses A as

$$A = \sum_{k=1}^{\min(M,N)} |U_k\rangle S_k \langle V_k| \tag{1.349}$$

in which the $\min(M, N)$ singular values S_k are nonnegative

$$S_k \geq 0. \tag{1.350}$$

Let's use the notation $|An\rangle \equiv A|n\rangle$ for the image of a vector $|n\rangle$ in an orthonormal basis $\{|n\rangle\}$ of V_N under the map A. We seek a special orthonormal basis $\{|n\rangle\}$ of V_N that has the property that the vectors $|An\rangle$ are orthogonal. This special basis $\{|n\rangle\}$ of V_N is the set of N orthonormal eigenstates of the $N \times N$ (nonnegative) hermitian operator $A^\dagger A$

$$A^\dagger A|n\rangle = e_n|n\rangle. \tag{1.351}$$

For since $A|n'\rangle = |An'\rangle$ and $A^\dagger A|n\rangle = e_n|n\rangle$, it follows that

$$\langle An'|An\rangle = \langle n'|A^\dagger A|n\rangle = e_n\langle n'|n\rangle = e_n\delta_{n'n}, \tag{1.352}$$

which shows that the vectors $|An\rangle$ are orthogonal and that their eigenvalues $e_n = \langle An|An\rangle$ are nonnegative. This is the essence of the singular-value decomposition.

If $N = M$, so that matrices $\langle m, b|A|n, a\rangle$ representing the linear operator A are square, then the $N = M$ singular values S_n are the nonnegative square-roots of the eigenvalues e_n

$$S_n = \sqrt{e_n} = \sqrt{\langle An|An\rangle} \geq 0. \tag{1.353}$$

We therefore may normalize each vector $|An\rangle$ whose singular value S_n is positive as

$$|m_n\rangle = \frac{1}{S_n}|An\rangle \quad \text{for} \quad S_n > 0 \tag{1.354}$$

so that the vectors $\{|m_n\rangle\}$ with positive singular values are orthonormal

$$\langle m_{n'}|m_n\rangle = \delta_{n',n}. \tag{1.355}$$

If only $P < N$ of the singular values are positive, then we may augment this set of P vectors $\{|m_n\rangle\}$ with $N - P = M - P$ new normalized vectors $|m_{n'}\rangle$ that are orthogonal to each other and to the P vectors defined by (1.354) (with positive singular values $S_n > 0$) so that the set of $N = M$ vectors $\{|m_n\rangle, |m_{n'}\rangle\}$ are complete and orthonormal in the space $V_{M=N}$.

If $N > M$, then A maps the N-dimensional space V_N into the smaller M-dimensional space V_M, and so A must annihilate $N - M$ basis vectors

$$A|n'\rangle = 0 \quad \text{for} \quad M < n' \leq N. \tag{1.356}$$

In this case, there are only M singular values S_n of which Z may be zero. The Z vectors $|An\rangle = A|n\rangle$ with vanishing S_ns are vectors of length zero; for these values of n, the matrix A maps the vector $|n\rangle$ to the zero vector. If there are more than $N - M$ zero-length vectors $|An\rangle = A|n\rangle$, then we must

replace the extra ones by new normalized vectors $|m_{n'}\rangle$ that are orthogonal to each other and to the vectors defined by (1.354) so that we have M orthonormal vectors in the augmented set $\{|m_n\rangle, |m_{n'}\rangle\}$. These vectors then form a basis for V_M.

When $N \leq M$, there are only N singular values S_n of which Z may be zero. If Z of the S_ns vanish, then one must add $Q = Z + M - N$ new normalized vectors $|m_{n'}\rangle$ that are orthogonal to each other and to the vectors defined by (1.354)

$$\langle m_{n'}|m_n\rangle = \frac{1}{S_n}\langle m_{n'}|A|n\rangle = 0 \quad \text{for} \quad n' > N - Z \quad \text{and} \quad S_n > 0 \tag{1.357}$$

so that we have M orthonormal vectors in the augmented set $\{|m_n\rangle, |m_{n'}\rangle\}$. These vectors then form a basis for V_M.

In both cases, $N > M$ and $M \geq N$, there are $\min(M, N)$ singular values, Z of which may be zero. We may choose the new vectors $\{|m_{n'}\rangle\}$ arbitrarily – as long as the augmented set $\{|m_n\rangle, |m_{n'}\rangle\}$ includes all the vectors defined by (1.354) and forms an orthonormal basis for V_M.

We now have two special orthonormal bases: the N N-dimensional eigenvectors $|n\rangle \in V_N$ that satisfy (1.351) and the M M-dimensional vectors $|m_n\rangle \in V_M$. To make the singular-value decomposition of the linear operator A, we choose as the identity operators I_N for the N-dimensional space V_N and I_M for the M-dimensional space V_M the sums

$$I_N = \sum_{n=1}^{N} |n\rangle\langle n| \quad \text{and} \quad I_M = \sum_{n'=1}^{M} |m_{n'}\rangle\langle m_{n'}|. \tag{1.358}$$

The singular-value decomposition of A then is

$$A = I_M A I_N = \sum_{n'=1}^{M} |m_{n'}\rangle\langle m_{n'}|A \sum_{n=1}^{N} |n\rangle\langle n|. \tag{1.359}$$

There are $\min(M, N)$ singular values S_n, all nonnegative. For the positive singular values, equations (1.352 & 1.354) show that the matrix element $\langle m_{n'}|A|n\rangle$ vanishes unless $n' = n$

$$\langle m_{n'}|A|n\rangle = \frac{1}{S_{n'}}\langle An'|An\rangle = S_{n'}\,\delta_{n'n}. \tag{1.360}$$

For the Z vanishing singular values, equation (1.353) shows that $A|n\rangle = 0$ and so

$$\langle m_{n'}|A|n\rangle = 0. \tag{1.361}$$

Thus only the $\min(M, N) - Z$ singular values that are positive contribute to the singular-value decomposition (1.359). If $N > M$, then there can be at most M nonzero eigenvalues e_n. If $N \leq M$, there can be at most N nonzero e_ns. The final

form of the singular-value decomposition then is a sum of dyadics weighted by the positive singular values

$$A = \sum_{n=1}^{\min(M,N)} |m_n\rangle S_n \langle n| = \sum_{n=1}^{\min(M,N)-Z} |m_n\rangle S_n \langle n|. \tag{1.362}$$

The vectors $|m_n\rangle$ and $|n\rangle$ respectively are the left and right singular vectors. The nonnegative numbers S_n are the singular values.

The linear operator A maps the $\min(M, N)$ right singular vectors $|n\rangle$ into the $\min(M, N)$ left singular vectors $S_n|m_n\rangle$ scaled by their singular values

$$A|n\rangle = S_n|m_n\rangle \tag{1.363}$$

and its adjoint $A^\dagger$ maps the $\min(M, N)$ left singular vectors $|m_n\rangle$ into the $\min(M, N)$ right singular vectors $|n\rangle$ scaled by their singular values

$$A^\dagger|m_n\rangle = S_n|n\rangle. \tag{1.364}$$

The N-dimensional vector space V_N is the **domain** of the linear operator A. If $N > M$, then A annihilates (at least) $N - M$ of the basis vectors $|n\rangle$. The **null space** or **kernel** of A is the space spanned by the basis vectors $|n\rangle$ that A annihilates. The vector space spanned by the left singular vectors $|m_n\rangle$ with nonzero singular values $S_n > 0$ is the **range** or **image** of A. It follows from the singular-value decomposition (1.362) that the dimension N of the domain is equal to the dimension of the kernel $N - M$ plus that of the range M, a result called the **rank-nullity theorem**.

Incidentally, the vectors $|m_n\rangle$ are the eigenstates of the hermitian matrix $A\,A^\dagger$ as one may see from the explicit product of the expansion (1.362) with its adjoint

$$\begin{aligned} A\,A^\dagger &= \sum_{n=1}^{\min(M,N)} |m_n\rangle S_n \langle n| \sum_{n'=1}^{\min(M,N)} |n'\rangle S_{n'} \langle m_{n'}| \\ &= \sum_{n=1}^{\min(M,N)} \sum_{n'=1}^{\min(M,N)} |m_n\rangle S_n \delta_{nn'} S_{n'} \langle m_{n'}| \\ &= \sum_{n=1}^{\min(M,N)} |m_n\rangle S_n^2 \langle m_n|, \end{aligned} \tag{1.365}$$

which shows that $|m_n\rangle$ is an eigenvector of $A\,A^\dagger$ with eigenvalue $e_n = S_n^2$,

$$A\,A^\dagger|m_n\rangle = S_n^2|m_n\rangle. \tag{1.366}$$

The SVD expansion (1.362) usually is written as a product of three explicit matrices, $A = U\Sigma V^\dagger$. The middle matrix Σ is an $M \times N$ matrix with the $\min(M, N)$ singular values $S_n = \sqrt{e_n}$ on its main diagonal and zeros elsewhere.

By convention, one writes the S_n in decreasing order with the biggest S_n as entry Σ_{11}. The first matrix U and the third matrix $V^\dagger$ depend upon the bases one uses to represent the linear operator A. If these basis vectors are $|\alpha_k\rangle$ and $|\beta_\ell\rangle$, then

$$A_{k\ell} = \langle\alpha_k|A|\beta_\ell\rangle = \sum_{n=1}^{\min(M,N)} \langle\alpha_k|m_n\rangle S_n \langle n|\beta_\ell\rangle \tag{1.367}$$

so that the k, nth entry in the matrix U is $U_{kn} = \langle\alpha_k|m_n\rangle$. The columns of the matrix U are the left singular vectors of the matrix A:

$$\begin{pmatrix} U_{1n} \\ U_{2n} \\ \vdots \\ U_{Mn} \end{pmatrix} = \begin{pmatrix} \langle\alpha_1|m_n\rangle \\ \langle\alpha_2|m_n\rangle \\ \vdots \\ \langle\alpha_M|m_n\rangle \end{pmatrix}. \tag{1.368}$$

Similarly, the n, ℓth entry of the matrix $V^\dagger$ is $\left(V^\dagger\right)_{n,\ell} = \langle n|\beta_\ell\rangle$. Thus $V_{\ell,n} = \left(V^\intercal\right)_{n,\ell} = \langle n|\beta_\ell\rangle^* = \langle\beta_\ell|n\rangle$. The columns of the matrix V are the right singular vectors of the matrix A

$$\begin{pmatrix} V_{1n} \\ V_{2n} \\ \vdots \\ V_{Nn} \end{pmatrix} = \begin{pmatrix} \langle\beta_1|n\rangle \\ \langle\beta_2|n\rangle \\ \vdots \\ \langle\beta_N|n\rangle \end{pmatrix}. \tag{1.369}$$

Since the columns of U and of V respectively are M and N orthonormal vectors, both of these matrices are unitary, that is $U^\dagger U = I_M$ and $V^\dagger V = I_N$ are the $M \times M$ and $N \times N$ identity matrices. The matrix form of the singular-value decomposition of A then is

$$A_{k\ell} = \sum_{m=1}^{M} \sum_{n=1}^{N} U_{km} \Sigma_{mn} V^\dagger_{n\ell} = \sum_{n=1}^{\min(M,N)} U_{kn} S_n V^\dagger_{n\ell} \tag{1.370}$$

or in matrix notation

$$A = U\Sigma V^\dagger. \tag{1.371}$$

The usual statement of the SVD theorem is: Every $M \times N$ complex matrix A can be written as the matrix product of an $M \times M$ unitary matrix U, an $M \times N$ matrix Σ that is zero except for its $\min(M, N)$ nonnegative diagonal elements, and an $N \times N$ unitary matrix $V^\dagger$

$$A = U\,\Sigma\,V^\dagger. \tag{1.372}$$

The first $\min(M, N)$ diagonal elements of S are the singular values S_k. They are real and nonnegative. The first $\min(M, N)$ columns of U and V are the left and right singular vectors of A. The last $\max(N - M, 0) + Z$ columns (1.369)

of the matrix V span the null space or kernel of A, and the first $\min(M, N) - Z$ columns (1.368) of the matrix U span the range of A.

Example 1.44 (Singular-value decomposition of a 2×3 matrix) If A is

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.373}$$

then the positive hermitian matrix $A^\dagger A$ is

$$A^\dagger A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \tag{1.374}$$

The normalized eigenvectors and eigenvalues of $A^\dagger A$ are

$$|1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \; e_1 = 2; \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; e_2 = 1; \quad |3\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \; e_3 = 0. \tag{1.375}$$

The third eigenvalue e_3 had to vanish because A is a 3×2 matrix.

The vector $A|1\rangle$ is (as a row vector) $|A1\rangle = A|1\rangle = (0, \sqrt{2})$, and its norm is $\sqrt{\langle 1|A^\dagger A|1\rangle} = \sqrt{2}$, so the normalized vector $|m_1\rangle$ is $|m_1\rangle = |A1\rangle/\sqrt{2} = (0; 1)$. Similarly, the vector $|m_2\rangle$ is $|m_2\rangle = A|2\rangle/\sqrt{\langle 2|A^\dagger A|2\rangle} = (1, 0)$. The SVD of A then is

$$A = \sum_{n=1}^{2} |m_n\rangle S_n \langle n| = U \Sigma V^\dagger \tag{1.376}$$

where $S_n = \sqrt{e_n}$. The unitary matrices $U_{k,n} = \langle \alpha_k | m_n \rangle$ and $V_{k,n} = \langle \beta_k | n \rangle$ are

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.377}$$

and the diagonal matrix Σ is

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{1.378}$$

So finally the SVD of $A = U \Sigma V^\dagger$ is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}. \tag{1.379}$$

The null space or kernel of A is the set of vectors that are real multiples c

$$N_A = \frac{c}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \tag{1.380}$$

of the third column of the matrix V displayed in (1.377). □

Example 1.45 (Matlab's SVD) Matlab's command [U,S,V] = svd(X) performs the singular-value decomposition of the matrix X. For instance

```
>> X = rand(3,3) + i*rand(3,3)

    0.6551 + 0.2551i  0.4984 + 0.8909i   0.5853 + 0.1386i
X = 0.1626 + 0.5060i  0.9597 + 0.9593i   0.2238 + 0.1493i
    0.1190 + 0.6991i  0.3404 + 0.5472i   0.7513 + 0.2575i
>> [U,S,V] = svd(X)

    -0.3689 - 0.4587i  0.4056 - 0.2075i  0.4362 - 0.5055i
U = -0.3766 - 0.5002i -0.5792 - 0.2810i  0.0646 + 0.4351i
    -0.2178 - 0.4626i  0.1142 + 0.6041i -0.5938 - 0.0901i

    2.2335         0         0
S =      0    0.7172         0
         0         0    0.3742

    -0.4577                0.5749                0.6783
V = -0.7885 - 0.0255i -0.6118 - 0.0497i -0.0135 + 0.0249i
    -0.3229 - 0.2527i  0.3881 + 0.3769i -0.5469 - 0.4900i.
```

The singular values are 2.2335, 0.7172, and 0.3742. □

We may use the SVD to solve, when possible, the matrix equation

$$A\,|x\rangle = |y\rangle \tag{1.381}$$

for the N-dimensional vector $|x\rangle$ in terms of the M-dimensional vector $|y\rangle$ and the $M \times N$ matrix A. Using the SVD expansion (1.362), we have

$$\sum_{n=1}^{\min(M,N)} |m_n\rangle S_n \langle n|x\rangle = |y\rangle. \tag{1.382}$$

The orthonormality (1.355) of the vectors $|m_n\rangle$ then tells us that

$$S_n \langle n|x\rangle = \langle m_n|y\rangle. \tag{1.383}$$

If the singular value is positive $S_n > 0$ whenever $\langle m_n|y\rangle \neq 0$, then we may divide by the singular value to get $\langle n|x\rangle = \langle m_n|y\rangle / S_n$ and so find the solution

$$|x\rangle = \sum_{n=1}^{\min(M,N)} \frac{\langle m_n|y\rangle}{S_n} |n\rangle. \tag{1.384}$$

But this solution is not always available or unique.

For instance, if for some n' the inner product $\langle m_{n'}|y\rangle \neq 0$ while the singular value $S_{n'} = 0$, then there is no solution to equation (1.381). This problem often occurs when $M > N$.

Example 1.46 Suppose A is the 3×2 matrix

$$A = \begin{pmatrix} r_1 & p_1 \\ r_2 & p_2 \\ r_3 & p_3 \end{pmatrix} \tag{1.385}$$

and the vector $|y\rangle$ is the cross-product $|y\rangle = \boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$. Then no solution $|x\rangle$ exists to the equation $A|x\rangle = |y\rangle$ (unless $\boldsymbol{r}$ and $\boldsymbol{p}$ are parallel) because $A|x\rangle$ is a linear combination of the vectors $\boldsymbol{r}$ and $\boldsymbol{p}$ while $|y\rangle = \boldsymbol{L}$ is perpendicular to both $\boldsymbol{r}$ and $\boldsymbol{p}$. □

Even when the matrix A is square, the equation (1.381) sometimes has no solutions. For instance, if A is a square matrix that vanishes, $A = 0$, then (1.381) has no solutions whenever $|y\rangle \neq 0$. And when $N > M$, as in for instance

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{1.386}$$

the solution (1.384) is never unique, for we may add to it any linear combination of the vectors $|n\rangle$ that A annihilates for $M < n \leq N$

$$|x\rangle = \sum_{n=1}^{\min(M,N)} \frac{\langle m_n|y\rangle}{S_n} |n\rangle + \sum_{n=M+1}^{N} x_n |n\rangle. \tag{1.387}$$

These are the vectors $|n\rangle$ for $M < n \leq N$ which A maps to zero since they do not occur in the sum (1.362), which stops at $n = \min(M, N) < N$.

Example 1.47 (The CKM matrix) In the standard model, the mass matrix of the d, s, and b quarks is a 3×3 complex, symmetric matrix M. Since M is symmetric ($M = M^{\mathsf{T}}$), its adjoint is its complex conjugate, $M^\dagger = M^*$. So the right singular vectors $|n\rangle$ are the eigenstates of M^*M as in (1.351)

$$M^*M|n\rangle = S_n^2|n\rangle \tag{1.388}$$

and the left singular vectors $|m_n\rangle$ are the eigenstates of MM^* as in (1.366)

$$MM^*|m_n\rangle = (M^*M)^*\,|m_n\rangle = S_n^2|m_n\rangle. \tag{1.389}$$

Thus the left singular vectors are just the complex conjugates of the right singular vectors, $|m_n\rangle = |n\rangle^*$. But this means that the unitary matrix V is the complex conjugate of the unitary matrix U, so the SVD of M is (Autonne, 1915)

$$M = U\Sigma U^{\mathsf{T}}. \tag{1.390}$$

The masses of the quarks then are the nonnegative singular values S_n along the diagonal of the matrix Σ. By redefining the quark fields, one may make the (CKM) matrix U real – except for a single complex phase, which causes a violation of charge-conjugation-parity (CP) symmetry. A similar matrix determines the neutrino masses. □

1.32 The Moore–Penrose pseudoinverse

Although a matrix A has an inverse A^{-1} if and only if it is square and has a nonzero determinant, one may use the singular-value decomposition to make a pseudoinverse A^+ for an arbitrary $M \times N$ matrix A. If the singular-value decomposition of the matrix A is

$$A = U\,\Sigma\,V^\dagger \tag{1.391}$$

then the Moore–Penrose pseudoinverse (Eliakim H. Moore, 1862–1932, Roger Penrose, 1931–) is

$$A^+ = V\,\Sigma^+\,U^\dagger \tag{1.392}$$

in which Σ^+ is the transpose of the matrix Σ with every nonzero entry replaced by its inverse (and the zeros left as they are). One may show that the pseudoinverse A^+ satisfies the four relations

$$\begin{aligned} A\,A^+\,A &= A \quad \text{and} \quad A^+\,A\,A^+ = A^+, \\ (A\,A^+)^\dagger &= A\,A^+ \quad \text{and} \quad (A^+\,A)^\dagger = A^+\,A \end{aligned} \tag{1.393}$$

and that it is the only matrix that does so.

Suppose that all the singular values of the $M \times N$ matrix A are positive. In this case, if A has more rows than columns, so that $M > N$, then the product AA^+ is the $N \times N$ identity matrix I_N

$$A^+A = V^\dagger\Sigma^+\Sigma V = V^\dagger I_N V = I_N \tag{1.394}$$

and AA^+ is an $M \times M$ matrix that is not the identity matrix I_M. If instead A has more columns than rows, so that $N > M$, then AA^+ is the $M \times M$ identity matrix I_M

$$AA^+ = U\Sigma\Sigma^+U^\dagger = UI_M U^\dagger = I_M \tag{1.395}$$

but A^+A is an $N \times N$ matrix that is not the identity matrix I_N. If the matrix A is square with positive singular values, then it has a true inverse A^{-1} which is equal to its pseudoinverse

$$A^{-1} = A^+. \tag{1.396}$$

If the columns of A are linearly independent, then the matrix $A^\dagger A$ has an inverse, and the pseudoinverse is

$$A^+ = \left(A^\dagger A\right)^{-1} A^\dagger. \tag{1.397}$$

The solution (1.220) to the complex least-squares method used this pseudoinverse.

If the rows of A are linearly independent, then the matrix $AA^\dagger$ has an inverse, and the pseudoinverse is

$$A^+ = A^\dagger \left(AA^\dagger\right)^{-1}. \tag{1.398}$$

If both the rows and the columns of A are linearly independent, then the matrix A has an inverse A^{-1} which is its pseudoinverse

$$A^{-1} = A^+. \tag{1.399}$$

Example 1.48 (The pseudoinverse of a 2×3 matrix) The pseudoinverse A^+ of the matrix A

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.400}$$

with singular-value decomposition (1.379) is

$$\begin{aligned} A^+ &= V \Sigma^+ U^\dagger \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \end{aligned} \tag{1.401}$$

which satisfies the four conditions (1.393). The product $A\,A^+$ gives the 2×2 identity matrix

$$A\,A^+ = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.402}$$

which is an instance of (1.395). Moreover, the rows of A are linearly independent, and so the simple rule (1.398) works:

$$\begin{aligned} A^+ &= A^\dagger \left(AA^\dagger\right)^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \left(\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \right)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \end{aligned} \tag{1.403}$$

which is (1.401).

The columns of the matrix A are not linearly independent, however, and so the simple rule (1.397) fails. Thus the product A^+A

$$A^+A = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.404}$$

is not the 3×3 identity matrix which it would be if (1.397) held. □

1.33 The rank of a matrix

Four equivalent definitions of the **rank** $R(A)$ of an $M \times N$ matrix A are:

1 the number of its linearly independent rows,
2 the number of its linearly independent columns,
3 the number of its nonzero singular values, and
4 the number of rows in its biggest square nonsingular submatrix.

A matrix of rank zero has no nonzero singular values and so is zero.

Example 1.49 (Rank) The 3×4 matrix

$$A = \begin{pmatrix} 1 & 0 & 1 & -2 \\ 2 & 2 & 0 & 2 \\ 4 & 3 & 1 & 1 \end{pmatrix} \tag{1.405}$$

has three rows, so its rank can be at most 3. But twice the first row added to thrice the second row equals twice the third row or

$$2r_1 + 3r_2 - 2r_3 = 0 \tag{1.406}$$

so $R(A) \leq 2$. The first two rows obviously are not parallel, so they are linearly independent. Thus the number of linearly independent rows of A is 2, and so A has rank 2. □

1.34 Software

Free, high-quality software for virtually all numerical problems in linear algebra are available in LAPACK – the Linear Algebra PACKage. The FORTRAN version is available at the web-site www.netlib.org/lapack/ and the C++ version at math.nist.gov/tnt/.

Matlab is a superb commercial program for numerical problems. A free GNU version of it is available at www.gnu.org/software/octave/. Maple and Mathematica are good commercial programs for symbolic problems.

1.35 The tensor/direct product

The **tensor product** (also called the **direct product**) is simple, but it can confuse students if they see it for the first time in a course on quantum mechanics. The tensor product is used to describe composite systems, such as an angular momentum composed of orbital and spin angular momenta.

If A is an $M \times N$ matrix with elements A_{ij} and Λ is a $K \times L$ matrix with elements $\Lambda_{\alpha\beta}$, then their **direct product** $C = A \otimes \Lambda$ is an $MK \times NL$ matrix with elements $C_{i\alpha j\beta} = A_{ij}\,\Lambda_{\alpha\beta}$. This direct-product matrix $A \otimes \Lambda$ maps the vector $V_{j\beta}$ into the vector

$$W_{i\alpha} = \sum_{j=1}^{N}\sum_{\beta=1}^{L} C_{i\alpha j\beta}\, V_{j\beta} = \sum_{j=1}^{N}\sum_{\beta=1}^{L} A_{ij}\,\Lambda_{\alpha\beta}\, V_{j\beta}. \tag{1.407}$$

In this sum, the second indices of A and Λ match those of the vector V. The most important case is when both A and Λ are square matrices, as will be their product $C = A \otimes \Lambda$. We'll focus on this case in the rest of this section.

The key idea here is that the direct product is a product of two operators that act on two different spaces. The operator A acts on the space S spanned by the N kets $|i\rangle$, and the operator Λ acts on the space Σ spanned by the K kets $|\alpha\rangle$. Let us assume that both operators map into these spaces, so that we may write them as

$$A = I_S A I_S = \sum_{i,j=1}^{N} |i\rangle\langle i|A|j\rangle\langle j| \tag{1.408}$$

and as

$$\Lambda = I_\Sigma \Lambda I_\Sigma = \sum_{\alpha,\beta=1}^{K} |\alpha\rangle\langle\alpha|\Lambda|\beta\rangle\langle\beta|. \tag{1.409}$$

Then the direct product $C = A \otimes \Lambda$

$$C = A \otimes \Lambda = \sum_{i,j=1}^{N}\sum_{\alpha,\beta=1}^{K} |i\rangle \otimes |\alpha\rangle\, \langle i|A|j\rangle\langle\alpha|\Lambda|\beta\rangle\, \langle j| \otimes \langle\beta| \tag{1.410}$$

acts on the direct product of the two vector spaces $S \otimes \Sigma$, which is spanned by the direct-product kets $|i, \alpha\rangle = |i\rangle\,|\alpha\rangle = |i\rangle \otimes |\alpha\rangle$.

In general, the direct-product space $S \otimes \Sigma$ is much bigger than the spaces S and Σ. For although $S \otimes \Sigma$ is spanned by the direct-product kets $|i\rangle \otimes |\alpha\rangle$, most vectors in the space $S \otimes \Sigma$ are of the form

$$|\psi\rangle = \sum_{i=1}^{N}\sum_{\alpha=1}^{K} \psi(i,\alpha)|i\rangle \otimes |\alpha\rangle \tag{1.411}$$

and not the direct product $|s\rangle \otimes |\sigma\rangle$ of a pair of vectors $|s\rangle \in S$ and $|\sigma\rangle \in \Sigma$

$$\begin{aligned} |s\rangle \otimes |\sigma\rangle &= \left(\sum_{i=1}^{N} s_i|i\rangle\right) \otimes \left(\sum_{\alpha=1}^{K} \sigma_\alpha|\alpha\rangle\right) \\ &= \sum_{i=1}^{N}\sum_{\alpha=1}^{K} s_i\sigma_\alpha|i\rangle \otimes |\alpha\rangle. \end{aligned} \tag{1.412}$$

Using the simpler notation $|i, \alpha\rangle$ for $|i\rangle \otimes |\alpha\rangle$, we may write the action of the direct-product operator $A \otimes \Lambda$ on the state

$$|\psi\rangle = \sum_{i=1}^{N}\sum_{\alpha=1}^{K} |i,\alpha\rangle\langle i,\alpha|\psi\rangle \tag{1.413}$$

as

$$(A \otimes \Lambda)|\psi\rangle = \sum_{i,j=1}^{N}\sum_{\alpha,\beta=1}^{K} |i,\alpha\rangle\, \langle i|A|j\rangle\langle\alpha|\Lambda|\beta\rangle\, \langle j,\beta|\psi\rangle. \tag{1.414}$$

Example 1.50 (States of the hydrogen atom) Suppose the states $|n, \ell, m\rangle$ are the eigenvectors of the hamiltonian H, the square $\boldsymbol{L}^2$ of the orbital angular momentum $\boldsymbol{L}$, and the third component of the orbital angular momentum L_3 for a hydrogen atom without spin:

$$\begin{aligned} H|n,\ell,m\rangle &= E_n|n,\ell,m\rangle, \\ \boldsymbol{L}^2|n,\ell,m\rangle &= \hbar^2\ell(\ell+1)|n,\ell,m\rangle, \\ L_3|n,\ell,m\rangle &= \hbar m|n,\ell,m\rangle. \end{aligned} \tag{1.415}$$

Suppose the states $|\sigma\rangle$ for $\sigma = \pm$ are the eigenstates of the third component S_3 of the operator $\boldsymbol{S}$ that represents the spin of the electron

$$S_3|\sigma\rangle = \sigma\frac{\hbar}{2}|\sigma\rangle. \tag{1.416}$$

Then the direct- or tensor-product states

$$|n,\ell,m,\sigma\rangle \equiv |n,\ell,m\rangle \otimes |\sigma\rangle \equiv |n,\ell,m\rangle|\sigma\rangle \tag{1.417}$$

represent a hydrogen atom including the spin of its electron. They are eigenvectors of all four operators H, $\boldsymbol{L}^2$, L_3, and S_3:

$$\begin{aligned} H|n,\ell,m,\sigma\rangle &= E_n|n,\ell,m,\sigma\rangle, & \boldsymbol{L}^2|n,\ell,m,\sigma\rangle &= \hbar^2\ell(\ell+1)|n,\ell,m,\sigma\rangle, \\ L_3|n,\ell,m,\sigma\rangle &= \hbar m|n,\ell,m,\sigma\rangle, & S_3|n,\ell,m,\sigma\rangle &= \sigma\hbar|n,\ell,m,\sigma\rangle. \end{aligned} \tag{1.418}$$

Suitable linear combinations of these states are eigenvectors of the square $\boldsymbol{J}^2$ of the composite angular momentum $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$ as well as of J_3, L_3, and S_3. □

Example 1.51 (Adding two spins) The smallest positive value of angular momentum is $\hbar/2$. The spin-one-half angular momentum operators $\boldsymbol{S}$ are represented by three 2×2 matrices

$$S_a = \frac{\hbar}{2}\sigma_a \tag{1.419}$$

in which the σ_a are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.420}$$

Consider two spin operators $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ acting on two spin-one-half systems. The states $|\pm\rangle_1$ are eigenstates of $S_3^{(1)}$, and the states $|\pm\rangle_2$ are eigenstates of $S_3^{(2)}$

$$S_3^{(1)}|\pm\rangle_1 = \pm\frac{\hbar}{2}|\pm\rangle_1 \quad \text{and} \quad S_3^{(2)}|\pm\rangle_2 = \pm\frac{\hbar}{2}|\pm\rangle_2. \tag{1.421}$$

Then the direct-product states $|\pm,\pm\rangle = |\pm\rangle_1|\pm\rangle_2 = |\pm\rangle_1 \otimes |\pm\rangle_2$ are eigenstates of both $S_3^{(1)}$ and $S_3^{(2)}$

$$S_3^{(1)}|\pm,s_2\rangle = \pm\frac{\hbar}{2}|+,s_2\rangle \quad \text{and} \quad S_3^{(2)}|s_1,\pm\rangle = \pm\frac{\hbar}{2}|s_1,\pm\rangle. \tag{1.422}$$

These states also are eigenstates of the third component of the spin operator of the combined system

$$S_3 = S_3^{(1)} + S_3^{(2)}, \quad \text{that is} \quad S_3|s_1,s_2\rangle = \frac{\hbar}{2}(s_1+s_2)|s_1,s_2\rangle. \tag{1.423}$$

Thus $S_3|+,+\rangle = \hbar|+,+\rangle$, and $S_3|-,-\rangle = -\hbar|-,-\rangle$, while $S_3|+,-\rangle = 0$ and $S_3|-,+\rangle = 0$.

Now let's consider the effect of the operator S_1^2 on the state $|++\rangle$

$$\begin{aligned} S_1^2|++\rangle &= \left(S_1^{(1)} + S_1^{(2)}\right)^2|++\rangle = \frac{\hbar^2}{4}\left(\sigma_1^{(1)} + \sigma_1^{(2)}\right)^2|++\rangle \\ &= \frac{\hbar^2}{2}\left(1 + \sigma_1^{(1)}\sigma_1^{(2)}\right)|++\rangle = \frac{\hbar^2}{2}\left(|++\rangle + \sigma_1^{(1)}|+\rangle\sigma_1^{(2)}|+\rangle\right) \\ &= \frac{\hbar^2}{2}\left(|++\rangle + |--\rangle\right). \end{aligned} \tag{1.424}$$

The rest of this example will be left to exercise 1.36. □

1.36 Density operators

A general quantum-mechanical system is represented by a **density operator** ρ that is hermitian $\rho^\dagger = \rho$, of unit trace $\mathrm{Tr}\rho = 1$, and positive $\langle\psi|\rho|\psi\rangle \geq 0$ for all kets $|\psi\rangle$.

If the state $|\psi\rangle$ is normalized, then $\langle\psi|\rho|\psi\rangle$ is the nonnegative probability that the system is in that state. This probability is real because the density matrix is hermitian. If $\{|n\rangle\}$ is any complete set of orthonormal states,

$$I = \sum_n |n\rangle\langle n|, \tag{1.425}$$

then the probability that the system is in the state $|n\rangle$ is

$$p_n = \langle n|\rho|n\rangle = \mathrm{Tr}\,(\rho|n\rangle\langle n|)\,. \tag{1.426}$$

Since $\mathrm{Tr}\rho = 1$, the sum of these probabilities is unity

$$\sum_n p_n = \sum_n \langle n|\rho|n\rangle = \mathrm{Tr}\left(\rho\sum_n |n\rangle\langle n|\right) = \mathrm{Tr}\,(\rho I) = \mathrm{Tr}\rho = 1. \tag{1.427}$$

A system that is measured to be in a state $|n\rangle$ cannot simultaneously be measured to be in an orthogonal state $|m\rangle$. The probabilities sum to unity because the system must be in some state.

Since the density operator ρ, is hermitian, it has a complete, orthonormal set of eigenvectors $|k\rangle$, all of which have nonnegative eigenvalues ρ_k

$$\rho|k\rangle = \rho_k|k\rangle. \tag{1.428}$$

They afford for it the expansion

$$\rho = \sum_{k=1}^{N} \rho_k|k\rangle\langle k| \tag{1.429}$$

in which the eigenvalue ρ_k is the probability that the system is in the state $|k\rangle$.

1.37 Correlation functions

We can define two Schwarz inner products for a density matrix ρ. If $|f\rangle$ and $|g\rangle$ are two states, then the inner product

$$(f,g) \equiv \langle f|\rho|g\rangle \tag{1.430}$$

for $g = f$ is nonnegative, $(f,f) = \langle f|\rho|f\rangle \geq 0$, and satisfies the other conditions (1.73, 1.74, & 1.76) for a Schwarz inner product.

The second Schwarz inner product applies to operators A and B and is defined (Titulaer and Glauber, 1965) as

$$(A, B) = \mathrm{Tr}\left(\rho A^\dagger B\right) = \mathrm{Tr}\left(B\rho A^\dagger\right) = \mathrm{Tr}\left(A^\dagger B\rho\right). \tag{1.431}$$

This inner product is nonnegative when $A = B$ and obeys the other rules (1.73, 1.74, & 1.76) for a Schwarz inner product.

These two degenerate inner products are not inner products in the strict sense of (1.73–1.79), but they are Schwarz inner products, and so (1.92–1.93) they satisfy the Schwarz inequality (1.93)

$$(f,f)(g,g) \geq |(f,g)|^2. \tag{1.432}$$

Applied to the first, vector, Schwarz inner product (1.430), the Schwarz inequality gives

$$\langle f|\rho|f\rangle\langle g|\rho|g\rangle \geq |\langle f|\rho|g\rangle|^2, \tag{1.433}$$

which is a useful property of density matrices. Application of the Schwarz inequality to the second, operator, Schwarz inner product (1.431) gives (Titulaer and Glauber, 1965)

$$\mathrm{Tr}\left(\rho A^\dagger A\right)\mathrm{Tr}\left(\rho B^\dagger B\right) \geq \left|\mathrm{Tr}\left(\rho A^\dagger B\right)\right|^2. \tag{1.434}$$

The operator $E_i(x)$ that represents the ith component of the electric field at the point x is the hermitian sum of the "positive-frequency" part $E_i^{(+)}(x)$ and its adjoint $E_i^{(-)}(x) = (E_i^{(+)}(x))^\dagger$

$$E_i(x) = E_i^{(+)}(x) + E_i^{(-)}(x). \tag{1.435}$$

Glauber has defined the first-order correlation function $G_{ij}^{(1)}(x,y)$ as (Glauber, 1963b)

$$G_{ij}^{(1)}(x,y) = \mathrm{Tr}\left(\rho E_i^{(-)}(x)E_j^{(+)}(y)\right) \tag{1.436}$$

or in terms of the operator inner product (1.431) as

$$G_{ij}^{(1)}(x,y) = \left(E_i^{(+)}(x), E_j^{(+)}(y)\right). \tag{1.437}$$

By setting $A = E_i^{(+)}(x)$, etc., it follows then from the Schwarz inequality (1.434) that the correlation function $G_{ij}^{(1)}(x,y)$ is bounded by (Titulaer and Glauber, 1965)

$$|G_{ij}^{(1)}(x,y)|^2 \leq G_{ii}^{(1)}(x,x)G_{jj}^{(1)}(y,y). \tag{1.438}$$

Interference fringes are sharpest when this inequality is saturated:

$$|G_{ij}^{(1)}(x,y)|^2 = G_{ii}^{(1)}(x,x)G_{jj}^{(1)}(y,y), \tag{1.439}$$

which can occur only if the correlation function $G_{ij}^{(1)}(x,y)$ factorizes (Titulaer and Glauber, 1965)

$$G_{ij}^{(1)}(x,y) = \mathcal{E}_i^*(x)\mathcal{E}_j(y) \tag{1.440}$$

as it does when the density operator is an outer product of coherent states

$$\rho = |\{\alpha_k\}\rangle\langle\{\alpha_k\}|, \tag{1.441}$$

which are eigenstates of $E_i^{(+)}(x)$ with eigenvalue $\mathcal{E}_i(x)$ (Glauber, 1963b, a)

$$E_i^{(+)}(x)|\{\alpha_k\}\rangle = \mathcal{E}_i(x)|\{\alpha_k\}\rangle. \tag{1.442}$$

The higher-order correlation functions

$$G_{i_1\ldots i_{2n}}^{(n)}(x_1\ldots x_{2n}) = \mathrm{Tr}\left(\rho E_{i_1}^{(-)}(x_1)\ldots E_{i_n}^{(-)}(x_n)E_{i_{n+1}}^{(+)}(x_{n+1})\ldots E_{i_{2n}}^{(+)}(x_n)\right) \tag{1.443}$$

satisfy similar inequalities (Glauber, 1963b), which also follow from the Schwarz inequality (1.434).

Exercises

1.1 Why is the most complicated function of two Grassmann numbers a polynomial with at most four terms as in (1.12)?

1.2 Derive the cyclicity (1.24) of the trace from (1.23).

1.3 Show that $(AB)^\intercal = B^\intercal A^\intercal$, which is (1.26).

1.4 Show that a real hermitian matrix is symmetric.

1.5 Show that $(AB)^\dagger = B^\dagger A^\dagger$, which is (1.29).

1.6 Show that the matrix (1.40) is positive on the space of all real 2-vectors but not on the space of all complex 2-vectors.

1.7 Show that the two 4×4 matrices (1.45) satisfy Grassmann's algebra (1.11) for $N = 2$.

1.8 Show that the operators $a_i = \theta_i$ defined in terms of the Grassmann matrices (1.45) and their adjoints $a_i^\dagger = \theta_i^\dagger$ satisfy the anticommutation relations (1.46) of the creation and annihilation operators for a system with two fermionic states.

1.9 Derive (1.64) from (1.61–1.63).

1.10 Fill in the steps leading to the formulas (1.69) for the vectors b_1' and b_2' and the formula (1.70) for the matrix a'.

1.11 Show that the antilinearity (1.76) of the inner product follows from its first two properties (1.73 & 1.74).

1.12 Show that the Minkowski product $(x,y) = x^0y^0 - \boldsymbol{x}\cdot\boldsymbol{y}$ of two 4-vectors x and y is an inner product that obeys the rules (1.73, 1.74, and 1.79).

1.13 Show that if $f = 0$, then the linearity (1.74) of the inner product implies that (f,f) and (g,f) vanish.

1.14 Show that the condition (1.75) of being positive definite implies nondegeneracy (1.79).

1.15 Show that the nonnegativity (1.77) of the Schwarz inner product implies the condition (1.78). Hint: the inequality $(f - \lambda g, f - \lambda g) \geq 0$ must hold for every complex λ and for all vectors f and g.

1.16 Show that the inequality (1.97) follows from the Schwarz inequality (1.96).

1.17 Show that the inequality (1.99) follows from the Schwarz inequality (1.98).

1.18 Use the Gram–Schmidt method to find orthonormal linear combinations of the three vectors

$$s_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad s_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad s_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{1.444}$$

1.19 Now use the Gram–Schmidt method to find orthonormal linear combinations of the same three vectors but in a different order

$$s'_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad s'_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad s'_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{1.445}$$

Did you get the same orthonormal vectors as in the previous exercise?

1.20 Derive the linearity (1.120) of the outer product from its definition (1.119).

1.21 Show that a linear operator A that is represented by a hermitian matrix (1.155) in an orthonormal basis satisfies $(g, Af) = (Ag, f)$.

1.22 Show that a unitary operator maps one orthonormal basis into another.

1.23 Show that the integral (1.170) defines a unitary operator that maps the state $|x'\rangle$ to the state $|x' + a\rangle$.

1.24 For the 2×2 matrices

$$A = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & -1 \\ 4 & -3 \end{pmatrix} \tag{1.446}$$

verify equations (1.202–1.204).

1.25 Derive the least-squares solution (1.220) for complex A, x, and y when the matrix $A^\dagger A$ is positive.

1.26 Show that the eigenvalues λ of a unitary matrix are unimodular, that is, $|\lambda| = 1$.

1.27 What are the eigenvalues and eigenvectors of the two defective matrices (1.252)?

1.28 Use (1.267) to derive expression (1.268) for the 2×2 rotation matrix $\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{\sigma}/2)$.

1.29 Compute the characteristic equation for the matrix $-i\boldsymbol{\theta} \cdot \boldsymbol{J}$ in which the generators are $(J_k)_{ij} = i\epsilon_{ikj}$ and ϵ_{ijk} is totally antisymmetric with $\epsilon_{123} = 1$.

1.30 Show that the sum of the eigenvalues of a normal antisymmetric matrix vanishes.

1.31 Use the characteristic equation of exercise 1.29 to derive identities (1.271) and (1.272) for the 3×3 real orthogonal matrix $\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J})$.

1.32 Consider the 2×3 matrix A

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -3 & 0 & 1 \end{pmatrix}. \tag{1.447}$$

Perform the singular value decomposition $A = USV^{\mathrm{T}}$, where V^{T} is the transpose of V. Find the singular values and the real orthogonal matrices U and V. Students may use Lapack, Octave, Matlab, Maple or any other program to do this exercise.

1.33 Consider the 6×9 matrix A with elements

$$A_{j,k} = x + x^j + i(y - y^k) \tag{1.448}$$

in which $x = 1.1$ and $y = 1.02$. Find the singular values, and the first left and right singular vectors. Students may use Lapack, Octave, Matlab, Maple or any other program to do this exercise.

1.34 Show that the totally antisymmetric Levi-Civita symbol ϵ_{ijk} satisfies the useful relation

$$\sum_{i=1}^{3} \epsilon_{ijk}\, \epsilon_{imn} = \delta_{jn}\, \delta_{km} - \delta_{jm}\, \delta_{kn}. \tag{1.449}$$

1.35 Consider the hamiltonian

$$H = \tfrac{1}{2}\hbar\omega\sigma_3 \tag{1.450}$$

where σ_3 is defined in (1.420). The entropy S of this system at temperature T is

$$S = -k\mathrm{Tr}\,[\rho \ln(\rho)] \tag{1.451}$$

in which the density operator ρ is

$$\rho = \frac{e^{-H/(kT)}}{\mathrm{Tr}\left[e^{-H/(kT)}\right]}. \tag{1.452}$$

Find expressions for the density operator ρ and its entropy S.

1.36 Find the action of the operator $\boldsymbol{S}^2 = \left(\boldsymbol{S}^{(1)} + \boldsymbol{S}^{(2)}\right)^2$ defined by (1.419) on the four states $|\pm\pm\rangle$ and then find the eigenstates and eigenvalues of $\boldsymbol{S}^2$ in the space spanned by these four states.

1.37 A system that has three fermionic states has three creation operators $a_i^\dagger$ and three annihilation operators a_k which satisfy the anticommutation relations $\{a_i, a_k^\dagger\} = \delta_{ik}$ and $\{a_i, a_k\} = \{a_i^\dagger, a_k^\dagger\} = 0$ for $i, k = 1, 2, 3$. The eight states of the system are $|v, u, t\rangle \equiv (a_3^\dagger)^t (a_2^\dagger)^u (a_1^\dagger)^v |0, 0, 0\rangle$. We can represent them by eight 8-vectors, each of which has seven 0s with a 1 in position $5v+3u+t$. How big should the matrices that represent the creation and annihilation operators be? Write down the three matrices that represent the three creation operators.

1.38 Show that the Schwarz inner product (1.430) is degenerate because it can violate (1.79) for certain density operators and certain pairs of states.

1.39 Show that the Schwarz inner product (1.431) is degenerate because it can violate (1.79) for certain density operators and certain pairs of operators.

1.40 The coherent state $|\{\alpha_k\}\rangle$ is an eigenstate of the annihilation operator a_k with eigenvalue α_k for each mode k of the electromagnetic field, $a_k|\{\alpha_k\}\rangle = \alpha_k|\{\alpha_k\}\rangle$. The positive-frequency part $E_i^{(+)}(x)$ of the electric field is a linear combination of the annihilation operators

$$E_i^{(+)}(x) = \sum_k a_k \, \mathcal{E}_i^{(+)}(k) \, e^{i(kx-\omega t)}. \tag{1.453}$$

Show that $|\{\alpha_k\}\rangle$ is an eigenstate of $E_i^{(+)}(x)$ as in (1.442) and find its eigenvalue $\mathcal{E}_i(x)$.

2

Fourier series

2.1 Complex Fourier series

The phases $\exp(inx)/\sqrt{2\pi}$, one for each integer n, are orthonormal on an interval of length 2π

$$\int_0^{2\pi} \left(\frac{e^{imx}}{\sqrt{2\pi}}\right)^* \frac{e^{inx}}{\sqrt{2\pi}} \, dx = \int_0^{2\pi} \frac{e^{i(n-m)x}}{2\pi} \, dx = \delta_{m,n} \tag{2.1}$$

where $\delta_{n,m} = 1$ if $n = m$, and $\delta_{n,m} = 0$ if $n \neq m$. So if a function $f(x)$ is a sum of these phases

$$f(x) = \sum_{n=-\infty}^{\infty} f_n \frac{e^{inx}}{\sqrt{2\pi}} \tag{2.2}$$

then their orthonormality (2.1) gives the nth coefficient f_n as the integral

$$\int_0^{2\pi} \frac{e^{-inx}}{\sqrt{2\pi}} f(x) \, dx = \int_0^{2\pi} \frac{e^{-inx}}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} f_m \frac{e^{imx}}{\sqrt{2\pi}} \, dx = \sum_{m=-\infty}^{\infty} \delta_{n,m} f_m = f_n \tag{2.3}$$

(Joseph Fourier, 1768–1830).

The **Fourier series** (2.2) is **periodic** with period 2π because the phases are periodic with period 2π, $\exp(in(x + 2\pi)) = \exp(inx)$. Thus even if the function $f(x)$ which we use in (2.3) to make the Fourier coefficients f_n is not periodic, its Fourier series (2.2) will nevertheless be strictly periodic, as illustrated by Figs. 2.2 & 2.4.