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 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 of $z = x + iy$ is

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

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 $z = re^{i\theta}$: 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, $\theta_i^2 = 0$. These numbers have amusing properties (used in chapter 20). For example, because $\theta_1\theta_2 = -\theta_2\theta_1$ and $\theta_1^2 = \theta_2^2 = 0$, the most general function of two Grassmann numbers is

$$f(\theta_1, \theta_2) = a + b\,\theta_1 + c\,\theta_2 + d\,\theta_1\theta_2 \tag{1.12}$$

and $1/(1 + a\,\theta_i) = 1 - a\,\theta_i$ in which a, b, c, d are complex numbers (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 $(\overline{a}\, b)_{ik} = \overline{a_i}\, b_k$, $(\overline{b}\, a)_{ki} = \overline{b_k}\, a_i$, and $(\overline{b}\, \overline{a})_{ki} = \overline{b_k}\, \overline{a_i}$. 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

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

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

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

or as its complex conjugate

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

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

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

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

$$(a\,b)_{i\ell} = a_{i1}\,b_{1\ell} + \ldots + a_{in}\,b_{n\ell} = \sum_{k=1}^{n} a_{ik}\,b_{k\ell}. \tag{1.21}$$

1.3 Matrices

Matrices are two-dimensional arrays.

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

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

The trace of the product of two matrices is independent of their order

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

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

$$\mathrm{Tr}\,(a\,b\,c\ldots z) = \mathrm{Tr}\,(b\,c\ldots z\,a) = \mathrm{Tr}\,(c\ldots z\,a\,b) = \ldots \tag{1.24}$$

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

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

Mathematicians often 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 transposition inverts the order of multiplication

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

A matrix that is equal to its transpose

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

is **symmetric**, $a_{ij} = a_{ji}$.

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

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

One may show that

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

A matrix that is equal to its adjoint

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

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

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

(Charles Hermite 1822–1901).

Example 1.1 (The Pauli Matrices) All three of Pauli's 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 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** $\boldsymbol{\delta_{ik}}$ is defined to be unity if $i = k$ and zero if $i \neq k$

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \tag{1.34}$$

(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

$$a^{-1}\, a = a\, a^{-1} = I \tag{1.35}$$

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

$$U^\dagger U = U U^\dagger = I \tag{1.36}$$

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

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

$$O^\mathsf{T} O = O O^\mathsf{T} = I. \tag{1.37}$$

Orthogonal matrices are square.

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

$$A \geq 0 \tag{1.38}$$

if for all complex vectors V the quadratic form

$$\langle V|A|V\rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} V_i^* A_{ij} V_j \geq 0 \tag{1.39}$$

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

$$\langle V|A|V\rangle > 0 \tag{1.40}$$

for all nonzero vectors $|V\rangle$.

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

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

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

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

can represent the number i since

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I. \tag{1.43}$$

The 2×2 matrix

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

can represent a Grassmann number since

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0. \tag{1.45}$$

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

$$\theta_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \theta_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.46}$$

The matrices that represent n Grassmann numbers are $2^n \times 2^n$ and have 2^n rows and 2^n columns. □

Example 1.4 (Fermions) The matrices (1.46) 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 operaors, satisfy the anticommutation relations

$$\{a_i, a_k^\dagger\} = \delta_{ik} \quad \text{and} \quad \{a_i, a_k\} = \{a_i^\dagger, a_k^\dagger\} = 0 \tag{1.47}$$

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

$$W = x\,U + y\,V \tag{1.48}$$

also is a vector in the vector space S.

A **basis** for a vector space S is a set B of vectors B_k for $k = 1 \dots n$ in

terms of which every vector U in S can be expressed as a linear combination

$$U = u_1 B_1 + u_2 B_2 + \ldots + u_n B_n \tag{1.49}$$

with numbers u_k in F. The numbers u_k are the **components** of the vector U in the basis B. If the **basis vectors** B_k are **orthonormal**, that is, if their inner products are $(B_k, B_\ell) = \langle B_k | B_\ell \rangle = \bar{B}_k \cdot B_\ell = \delta_{k\ell}$, then we might represent the vector U as the n-tuple $(u_1, u_2, \ldots, u_n)$ with $u_k = \langle B_k | U \rangle$ or as the corresponding column vector.

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

$$H = nW + mN \tag{1.50}$$

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

$$z = x1 + yi = x + iy. \tag{1.51}$$

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

$$\boldsymbol{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{1.52}$$

in which x and y are real numbers and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are perpendicular vectors of unit length (unit vectors with $\hat{x} \cdot \hat{x} = 1 = \hat{y} \cdot \hat{y}$ and $\hat{x} \cdot \hat{y} = 0$). This vector space, called $\mathbb{R}^2$, is a 2-d space over the reals.

The vector $\boldsymbol{r}$ can be described by the basis vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ and also by

any other set of basis vectors, such as $-\hat{\mathbf{y}}$ and $\hat{\mathbf{x}}$

$$\boldsymbol{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} = -y(-\hat{\mathbf{y}}) + x\hat{\mathbf{x}}. \tag{1.53}$$

The components of the vector $\boldsymbol{r}$ 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

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \tag{1.54}$$

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 $x = y = 0$

$$a\,\frac{\partial}{\partial x} + b\,\frac{\partial}{\partial y} \tag{1.55}$$

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 $[a, b]$

$$f(x) = \sum_i z_i\, f_i(x) \tag{1.56}$$

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

Example 1.12 (States in quantum mechanics) In quantum mechanics, if the properties of a system have been measured as completely as possible, then the system (or our knowledge of it) is said to be in a **state**, often called a **pure state**, and is represented by a vector ψ or $|\psi\rangle$ in Dirac's notation. If the properties of a system have not been measured as completely as possible, then the system (or our knowledge of it) is said to be in a **mixture** or a **mixed state**, and is represented by a density operator (section 1.35).

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

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle \tag{1.57}$$

also is a possible state of the system.

A harmonic oscillator in its kth excited state is in a state described by a vector $|k\rangle$. A particle exactly at position q is in a state described by a vector $|q\rangle$. An electron moving with momentum $\boldsymbol{p}$ and spin σ is in a state represented by a vector $|\boldsymbol{p}, \sigma\rangle$. A hydrogen atom at rest in its ground state is in a state $|E_0\rangle$. □

Example 1.13 (Polarization of photons and gravitons) The general state of a photon of momentum $\vec{k}$ is one of elliptical polarization

$$|\vec{k}, \theta, \phi\rangle = \cos\theta\, e^{i\phi}|\vec{k}, +\rangle + \sin\theta\, e^{-i\phi}|\vec{k}, -\rangle \tag{1.58}$$

in which the states of positive and negative helicity $|\vec{k}, \pm\rangle$ represent a photon whose angular momentum $\pm\hbar$ is parallel or antiparallel to its momentum $\vec{k}$. If $\theta = \pi/4 + n\pi$, the polarization is linear, and the electric field is parallel to an axis that depends upon ϕ and is perpendicular to $\vec{k}$.

The general state of a graviton of momentum $\vec{k}$ also is one of elliptical polarization (1.58), but now the states of positive and negative helicity $|\vec{k}, \pm\rangle$ have angular momentum $\pm 2\hbar$ parallel or antiparallel to the momentum $\vec{k}$. Linear polarization again is $\theta = \pi/4 + n\pi$. The state $|\vec{k}, +\rangle$ represents space being stretched and squeezed along one axis while being squeezed and stretched along another axis, both axes perpendicular to each other and to $\vec{k}$. In the state $|\vec{k}, \times\rangle$, the stretching and squeezing axes are rotated by 45° about $\vec{k}$ relative to those of $|\vec{k}, +\rangle$. □

1.5 Linear operators

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

$$A(bV + cW) = bA(V) + cA(W) = bA\,V + cA\,W. \tag{1.59}$$

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

$$A\,B_i = a_{1i}B_1 + a_{2i}B_2 + \ldots + a_{ni}B_n = \sum_{k=1}^{n} a_{ki}\,B_k \tag{1.60}$$

a formula that is clearer in Dirac's notation (section 1.12). The square matrix a_{ki} **represents** the linear operator A in the B_k basis. The effect of A on any vector $V = u_1B_1 + u_2B_2 + \ldots + u_nB_n$ in S then is

$$A\,V = A\sum_{i=1}^{n} u_iB_i = \sum_{i=1}^{n} u_iAB_i = \sum_{i,k=1}^{n} u_ia_{ki}B_k = \sum_{i,k=1}^{n} a_{ki}u_iB_k. \tag{1.61}$$

So the kth component u'_k of the vector $V' = A\,V$ is

$$u'_k = a_{k1}u_1 + a_{k2}u_2 + \ldots + a_{kn}u_n = \sum_{i=1}^{n} a_{ki}\,u_i. \tag{1.62}$$

Thus the column vector u' of the components u'_k of the vector $V' = A\,V$ 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 V in that basis. 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'_i

$$B'_i = \sum_{\ell=1}^{n} u_{ki}\,B_k \tag{1.63}$$

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

$$\sum_{k=1}^{n} u^{-1}_{ki}B'_k = \sum_{k=1}^{n} u^{-1}_{ki}\sum_{\ell=1}^{n} u_{\ell k}B_\ell = \sum_{\ell=1}^{n}\left(\sum_{k=1}^{n} u_{\ell k}u^{-1}_{ki}\right)B_\ell = \sum_{\ell=1}^{n}\delta_{\ell i}B_\ell = B_i \tag{1.64}$$

then the old basis vectors B_i are given by

$$B_i = \sum_{k=1}^{n} u^{-1}_{ki}B'_k. \tag{1.65}$$

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

$$A\,B'_i = \sum_{k=1}^{n} u_{ki}\,A\,B_k = \sum_{j,k=1}^{n} u_{ki}a_{jk}B_j = \sum_{j,k,\ell=1}^{n} u_{ki}a_{jk}u^{-1}_{\ell j}\,B'_\ell. \tag{1.66}$$

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**

$$a'_{\ell i} = \sum_{jk=1}^{n} u^{-1}_{\ell j} a_{jk} u_{ki} \quad \text{or} \quad a' = u^{-1}\, a\, u \tag{1.67}$$

in matrix notation. If the matrix u is **unitary**, then its inverse is its hermitian adjoint

$$u^{-1} = u^\dagger \tag{1.68}$$

and the similarity transformation (1.67) is

$$a'_{\ell i} = \sum_{jk=1}^{n} u^\dagger_{\ell j} a_{jk} u_{ki} = \sum_{jk=1}^{n} u^*_{j\ell} a_{jk} u_{ik} \quad \text{or} \quad a' = u^\dagger\, a\, u. \tag{1.69}$$

Because traces are cyclic, they are invariant under similarity transformations

$$\mathrm{Tr}(a') = \mathrm{Tr}(u\, a\, u^{-1}) = \mathrm{Tr}(a\, u^{-1}\, u) = \mathrm{Tr}(a). \tag{1.70}$$

Example 1.14 (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

$$b_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad b_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.71}$$

represent the basis vectors B_1 and B_2, then the matrix

$$a = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{1.72}$$

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

$$B'_1 = \frac{1}{\sqrt{2}}\,(B_1 + B_2) \quad \text{and} \quad B'_2 = \frac{1}{\sqrt{2}}\,(B_1 - B_2) \tag{1.73}$$

then the vectors

$$b'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad b'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{1.74}$$

would represent B_1 and B_2, and the matrix

$$a' = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \tag{1.75}$$

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.

$$A\,B_i = \sum_{k=1}^{M} a_{ki}\,C_k. \tag{1.76}$$

It then maps an arbitrary vector $V = u_1 B_1 + \ldots + u_n B_n$ in S into the vector

$$A\,V = \sum_{k=1}^{M} \left(\sum_{i=1}^{n} a_{ki}\,u_i \right) C_k \tag{1.77}$$

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 obeys the rules

$$(f, g) = (g, f)^* \tag{1.78}$$

$$(f, z\,g + w\,h) = z\,(f, g) + w\,(f, h) \tag{1.79}$$

$$(f, f) \geq 0 \quad \text{and} \quad (f, f) = 0 \iff f = 0 \tag{1.80}$$

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 **anti-linear** in the first vector of the pair

$$(z\,g + w\,h, f) = z^*(g, f) + w^*(h, f). \tag{1.81}$$

A **Schwarz inner product** obeys the first two rules (1.78, 1.79) for an inner product and the fourth (1.81) but only the first part of the third (1.80)

$$(f, f) \geq 0. \tag{1.82}$$

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

$$(f, f) = 0 \implies (g, f) = 0 \text{ for all } g \in V. \tag{1.83}$$

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

Inner products of 4-vectors can be negative. To accomodate them we define an **indefinite** inner product without regard to positivity as one that satisfies the first two rules (1.78 & 1.79) and therefore also the fourth rule (1.81) and that instead of being positive definite is **nondegenerate**

$$(f, g) = 0 \text{ for all } f \in V \implies g = 0. \tag{1.84}$$

This rule says that only the zero vector is orthogonal to all the vectors of the space. The positive-definite condition (1.80) is stronger than and implies nondegeneracy (1.84) (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)

$$\| f \| = \sqrt{(f, f)}. \tag{1.85}$$

A vector $\hat{f} = f / \| f \|$ has unit norm and is said to be **normalized**. Two measures of the distance between two normalized vectors f and g are the **norm of their difference** and the **Bures distance**

$$D(f, g) = \| f - g \| \quad \text{and} \quad D_B(f, g) = \arccos(|(f, g)|). \tag{1.86}$$

Example 1.15 (Euclidian 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

$$(U, V) = \sum_{i=1}^{n} U_i V_i \tag{1.87}$$

that is nonnegative when the two vectors are the same

$$(U, U) = \sum_{i=1}^{n} U_i U_i = \sum_{i=1}^{n} U_i^2 \geq 0 \tag{1.88}$$

and vanishes only if all the components U_i are zero, that is, if the vector $U = 0$. Thus the inner product (1.87) is positive definite. When (U, V) is zero, the vectors U and V are **orthogonal**. □

Example 1.16 (Complex Euclidian 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

$$(U, V) = \sum_{i=1}^{n} U_i^* V_i = (V, U)^*. \tag{1.89}$$

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

$$(U, U) = \sum_{i=1}^{n} U_i^* U_i = \sum_{i=1}^{n} |U_i|^2 \geq 0 \tag{1.90}$$

only if $U = 0$. So the inner product (1.89) is positive definite. If (U, V) is zero, then U and V are orthogonal. □

Example 1.17 (Complex Matrices) For the vector space of $n \times m$ complex matrices A, B, ..., the trace of the adjoint (1.28) of A multiplied by B is an inner product

$$(A, B) = \mathrm{Tr} A^\dagger B = \sum_{i=1}^{n} \sum_{j=1}^{m} (A^\dagger)_{ji} B_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}^* B_{ij} \tag{1.91}$$

that is nonnegative when the matrices are the same

$$(A, A) = \mathrm{Tr} A^\dagger A = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}^* A_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} |A_{ij}|^2 \geq 0 \tag{1.92}$$

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

A vector space with a positive-definite inner product (1.78–1.81) 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 (1.85) 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. An infinite-dimensional complete inner-product space, such as the space of all square-integrable functions, also is a Hilbert space (David Hilbert, 1862–1943).

Example 1.18 (Hilbert Space of Square-Integrable Functions) For the vector space of functions (1.56), a natural inner product is

$$(f, g) = \int_a^b dx\, f^*(x)\, g(x). \tag{1.93}$$

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

$$\| f \|^2 = \int_a^b dx \, |f(x)|^2. \tag{1.94}$$

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 1.19 (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 (1.84), and invariant under Lorentz transformations, and often is written as $p \cdot x$ or as $p\,x$. If p is the 4-momentum of a freely moving physical particle of mass m, then

$$p \cdot p = \boldsymbol{p} \cdot \boldsymbol{p} - E^2/c^2 = -c^2 m^2 \leq 0. \tag{1.95}$$

The Minkowski inner product satisfies the rules (1.78, 1.79, and 1.84), but it is **not positive definite**, and it does not satisfy the Schwarz inequality (Hermann Minkowski 1864–1909, Hendrik Lorentz 1853–1928). □

Example 1.20 (Inner products in quantum mechanics) The probability $P(\phi|\psi)$ that a system in the state $|\psi\rangle$ will be measured to be in the state $|\phi\rangle$ is the absolute value squared of the inner product $\langle\phi|\psi\rangle$ divided by the squared norms of the two states

$$P(\phi|\psi) = \frac{|\langle\phi|\psi\rangle|^2}{\langle\phi|\phi\rangle\langle\psi|\psi\rangle}. \tag{1.96}$$

If the two states are normalized, then the probability is just the absolute value squared of their inner product, $P(\phi|\psi) = |\langle\phi|\psi\rangle|^2$. □

1.7 Cauchy–Schwarz inequalities

For any two vectors f and g, the Schwarz inequality

$$(f, f)\,(g, g) \geq |(f, g)|^2 \tag{1.97}$$

holds for any Schwarz inner product (and so for any positive-definite inner product). The condition (1.82) of **nonnegativity** ensures that for any complex number λ the inner product of the vector $f - \lambda g$ with itself is nonnegative

$$(f - \lambda g, f - \lambda g) = (f, f) - \lambda^*(g, f) - \lambda(f, g) + |\lambda|^2 (g, g) \geq 0. \tag{1.98}$$

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 (1.97) 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 (1.98) 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 (1.97)

$$(f, f)(g, g) \geq |(f, g)|^2. \tag{1.99}$$

Taking the square root of each side, we have

$$\| f \| \| g \| \geq |(f, g)| \tag{1.100}$$

(Hermann Schwarz 1843–1921).

Example 1.21 (Some Schwarz Inequalities) For the dot product of two real 3-vectors $\boldsymbol{r}$ & $\boldsymbol{R}$, the Cauchy-Schwarz inequality is

$$(\boldsymbol{r} \cdot \boldsymbol{r})\ (\boldsymbol{R} \cdot \boldsymbol{R}) \geq (\boldsymbol{r} \cdot \boldsymbol{R})^2 = (\boldsymbol{r} \cdot \boldsymbol{r})\ (\boldsymbol{R} \cdot \boldsymbol{R}) \cos^2\theta \tag{1.101}$$

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

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

$$(\boldsymbol{x} \cdot \boldsymbol{x})\ (\boldsymbol{y} \cdot \boldsymbol{y}) \geq (\boldsymbol{x} \cdot \boldsymbol{y})^2 = (\boldsymbol{x} \cdot \boldsymbol{x})\ (\boldsymbol{y} \cdot \boldsymbol{y}) \cos^2\theta \tag{1.102}$$

and it implies (exercise 1.16) that

$$\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \geq \|\boldsymbol{x} + \boldsymbol{y}\|. \tag{1.103}$$

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

$$(\boldsymbol{u}^* \cdot \boldsymbol{u})\ (\boldsymbol{v}^* \cdot \boldsymbol{v}) \geq |\boldsymbol{u}^* \cdot \boldsymbol{v}|^2 = (\boldsymbol{u}^* \cdot \boldsymbol{u})\ (\boldsymbol{v}^* \cdot \boldsymbol{v}) \cos^2\theta \tag{1.104}$$

and it implies (exercise 1.17) that

$$\|\boldsymbol{u}\| + \|\boldsymbol{v}\| \geq \|\boldsymbol{u} + \boldsymbol{v}\|. \tag{1.105}$$

The inner product (1.93) of two complex functions f and g provides another example

$$\int_a^b dx\, |f(x)|^2 \int_a^b dx\, |g(x)|^2 \geq \left| \int_a^b dx\, f^*(x)\, g(x) \right|^2 \tag{1.106}$$

of the Schwarz inequality. □

1.8 Linear independence and completeness

A set of n vectors V_1, V_2, ..., V_n is **linearly dependent** if there exist numbers c_i, *not all zero*, such that the linear combination

$$c_1 V_1 + \ldots + c_n V_n = 0 \tag{1.107}$$

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, ..., 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, $U = u_1 V_1 + \ldots + u_n V_n$. 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_0, c_1, ..., c_n, not all zero, that make the sum

$$c_0 U + c_1 V_1 + \ldots + c_n V_n = 0 \tag{1.108}$$

vanish. Now if c_0 were 0, then the set $\{V_i\}$ would be linearly dependent. Thus $c_0 \neq 0$, and so we may divide by c_0 and express the arbitrary vector U as a linear combination of the vectors V_i

$$U = -\frac{1}{c_0}\left(c_1 V_1 + \ldots + c_n V_n\right) = u_1 V_1 + \ldots + u_n V_n \tag{1.109}$$

with $u_k = -c_k/c_0$. Thus 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. Such a set $\{V_i\}$ of linearly independent vectors that is maximal in a space S is called a **basis** for S; it **spans** S; it is a **complete** set of vectors in S.

1.9 Dimension of a vector space

If $V_1, \ldots, V_n$ and $W_1, \ldots, W_m$ are any two bases for a vector space S, then $n = m$.

To see why, suppose that the n vectors $C_1, C_2, \ldots, C_n$ are complete in a vector space S, and that the m vectors $L_1, L_2, \ldots, L_m$ in S are linearly independent (Halmos, 1958, sec. 1.8). Since the C's are complete, the set of vectors $L_m, C_1, \ldots, C_n$ is linearly dependent. So we can omit one of the C's

and the remaining set $L_m, C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n$ still spans S. Repeating this argument, we find that the vectors

$$L_{m-1}, L_m, C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n \tag{1.110}$$

are linearly dependent, and that the vectors

$$L_{m-1}, L_m, C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_{j-1}, C_{j+1}, \ldots, C_n \tag{1.111}$$

still span S. We continue to repeat these steps until we run out of L's or C's. If n were less than m, then we'd end up with a set of vectors $L_k, \ldots, L_m$ that would be complete and therefore each of the vectors $L_1, \ldots, L_{k-1}$ would have to be linear combinations of the vectors $L_k, \ldots, L_m$. But the L's by assumption are linearly independent. So $n \geq m$. Thus if both the C's and the L's are bases for the same space S, and so are both complete and linearly independent in it, then both $n \geq m$ and $m \geq n$. So all the bases of a vector space consist of the same number of vectors. This number is the **dimension** of the space.

The steps of the above demonstration stop for $n = m$ when the m linearly independent L's have replaced the n complete C's leaving us with $n = m$ linearly independent L's that are complete. Thus in a vector space of n dimensions, every set of n linearly independent vectors is complete and so forms a basis for the space.

1.10 Orthonormal vectors

Suppose the vectors V_1, V_2, ..., V_n are linearly independent. Then we can make out of them a set of n vectors U_i that are orthonormal

$$(U_i, U_j) = \delta_{ij}. \tag{1.112}$$

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

$$U_1 = \frac{V_1}{\sqrt{(V_1, V_1)}} \tag{1.113}$$

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

$$0 = (U_1, u_2) = (U_1,\, c_{12}U_1 + V_2) = c_{12} + (U_1, V_2). \tag{1.114}$$

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

$$u_2 = V_2 - (U_1, V_2)\, U_1. \tag{1.115}$$

The normalized vector U_2 then is

$$U_2 = \frac{u_2}{\sqrt{(u_2, u_2)}}. \tag{1.116}$$

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

$$0 = (U_1, u_3) = (U_1,\, c_{13}U_1 + c_{23}U_2 + V_3) = c_{13} + (U_1, V_3) \tag{1.117}$$

and also to U_2

$$0 = (U_2, u_3) = (U_2,\, c_{13}U_1 + c_{23}U_2 + V_3) = c_{23} + (U_2, V_3). \tag{1.118}$$

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

$$u_3 = V_3 - (U_1, V_3)\, U_1 - (U_2, V_3)\, U_2. \tag{1.119}$$

The normalized vector U_3 then is

$$U_3 = \frac{u_3}{\sqrt{(u_3, u_3)}}. \tag{1.120}$$

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

$$u_k = V_k + \sum_{i=1}^{k-1} c_{ik}\, U_i \tag{1.121}$$

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

$$u_k = V_k - \sum_{i=1}^{k-1} (U_i, V_k)\, U_i. \tag{1.122}$$

The normalized vector then is

$$U_k = \frac{u_k}{\sqrt{(u_k, u_k)}}. \tag{1.123}$$

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 **outer-product** operator A that maps any vector h into the vector f multiplied by the inner product (g, h)

$$A\, h = f\, (g, h) = (g, h)\, f. \tag{1.124}$$

The operator A is linear because for any vectors e, h and numbers z, w

$$A\,(z\,h + w\,e) = (g, z\,h + w\,e)\,f = z\,(g,h)\,f + w\,(g,e)\,f = z\,A\,h + w\,A\,e. \tag{1.125}$$

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

$$(Ah)_i = \sum_{j=1}^{n} A_{ij}\,h_j = f_i \sum_{j=1}^{n} g_j^*\,h_j \tag{1.126}$$

and in that basis A is the matrix with entries

$$A_{ij} = f_i\,g_j^*. \tag{1.127}$$

It is the **outer product** of the vectors f and g^*. The outer product of g and f^* is different, $B_{ij} = g_i\,f_j^*$.

Example 1.22 (Outer Product) If in some basis the vectors f and g are

$$f = \begin{pmatrix} 2 \\ 3i \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} i \\ 1 \\ 3i \end{pmatrix} \tag{1.128}$$

then their outer products are the matrices

$$A = \begin{pmatrix} 2 \\ 3i \end{pmatrix} \begin{pmatrix} -i & 1 & -3i \end{pmatrix} = \begin{pmatrix} -2i & 2 & -6i \\ 3 & 3i & 9 \end{pmatrix} \tag{1.129}$$

and

$$B = \begin{pmatrix} i \\ 1 \\ 3i \end{pmatrix} \begin{pmatrix} 2 & -3i \end{pmatrix} = \begin{pmatrix} 2i & 3 \\ 2 & -3i \\ 6i & 9 \end{pmatrix}. \tag{1.130}$$

□

Example 1.23 (Dirac's outer products) Dirac's notation for outer products is neat. If the vectors $f = |f\rangle$ and $g = |g\rangle$ are

$$|f\rangle = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad |g\rangle = \begin{pmatrix} z \\ w \end{pmatrix} \tag{1.131}$$

then their outer products are

$$|f\rangle\langle g| = \begin{pmatrix} az^* & aw^* \\ bz^* & bw^* \\ cz^* & cw^* \end{pmatrix} \quad \text{and} \quad |g\rangle\langle f| = \begin{pmatrix} za^* & zb^* & zc^* \\ wa^* & wb^* & wc^* \end{pmatrix} \tag{1.132}$$

as well as

$$|f\rangle\langle f| = \begin{pmatrix} aa^* & ab^* & ac^* \\ ba^* & bb^* & bc^* \\ ca^* & cb^* & cc^* \end{pmatrix} \quad \text{and} \quad |g\rangle\langle g| = \begin{pmatrix} zz^* & zw^* \\ wz^* & ww^* \end{pmatrix}. \tag{1.133}$$

□

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 **bra**c**ket** $(g, f) = \langle g|f\rangle$. A matrix element (g, cf) of an operator c then is $(g, cf) = \langle g|c|f\rangle$ in which the bra and ket bracket the operator c.

In Dirac notation, an outer product like (1.124) $A\,h = (g, h)\,f = f\,(g, h)$ reads $A\,|h\rangle = |f\rangle\langle g|h\rangle$, and the outer product A itself is $A = |f\rangle\langle g|$.

The bra $\langle g|$ is the **adjoint** of the ket $|g\rangle$, and the ket $|f\rangle$ is the adjoint of the bra $\langle f|$

$$\langle g| = (|g\rangle)^\dagger \text{ and } |f\rangle = (\langle f|)^\dagger, \quad \text{so } \langle g|^{\dagger\dagger} = \langle g| \text{ and } |f\rangle^{\dagger\dagger} = |f\rangle. \tag{1.134}$$

The adjoint of an outer product is

$$(\,z\,|f\rangle\langle g|\,)^\dagger = z^*\,|g\rangle\langle f|. \tag{1.135}$$

In Dirac's notation, the most general linear operator is an arbitrary linear combination of outer products

$$A = \sum_{k\ell} z_{k\ell}\,|k\rangle\langle\ell|. \tag{1.136}$$

Its adjoint is

$$A^\dagger = \sum_{k\ell} z^*_{k\ell}\,|\ell\rangle\langle k|. \tag{1.137}$$

The adjoint of a ket $|h\rangle = A|f\rangle$ is

$$(|h\rangle)^\dagger = (A|f\rangle)^\dagger = \left(\sum_{k\ell} z_{k\ell}\,|k\rangle\langle\ell|f\rangle\right)^\dagger = \sum_{k\ell} z^*_{k\ell}\,\langle f|\ell\rangle\langle k| = \langle f|A^\dagger. \tag{1.138}$$

Before Dirac, bras were implicit in the definition of the inner product, but they did not appear explicitly; there was no simple way to write the bra $\langle g|$ or the outer product $|f\rangle\langle g|$.

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

$$|f\rangle = \sum_{k=1}^{n} c_k |k\rangle. \tag{1.139}$$

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

$$\langle \ell | f \rangle = \sum_{k=1}^{n} c_k \langle \ell | k \rangle = \sum_{k=1}^{n} c_k \, \delta_{\ell,k} = c_\ell. \tag{1.140}$$

The original expasion (1.139) then must be

$$|f\rangle = \sum_{k=1}^{n} c_k |k\rangle = \sum_{k=1}^{n} \langle k | f \rangle \, |k\rangle = \sum_{k=1}^{n} |k\rangle \, \langle k | f \rangle = \left(\sum_{k=1}^{n} |k\rangle \, \langle k| \right) |f\rangle. \tag{1.141}$$

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

$$I = \sum_{k=1}^{n} |k\rangle \, \langle k|. \tag{1.142}$$

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

$$I = \sum_{j=1}^{n} |\alpha_j\rangle \, \langle \alpha_j| = \sum_{k=1}^{n} |k\rangle \, \langle k|. \tag{1.143}$$

These resolutions of the identity operator give every vector $|f\rangle$ in the space the expansions

$$|f\rangle = \sum_{j=1}^{n} |\alpha_j\rangle \, \langle \alpha_j | f \rangle = \sum_{k=1}^{n} |k\rangle \, \langle k | f \rangle. \tag{1.144}$$

Example 1.24 (Linear operators represented as matrices) The equations (1.60–1.67) that relate linear operators to the matrices that represent them are much clearer in Dirac's notation. If the kets $|B_k\rangle$ are n orthonormal basis vectors, that is, if $\langle B_k | B_\ell \rangle = \delta_{k\ell}$, for a vector space S, then a linear operator A acting on S maps the basis vector $|B_i\rangle$ into (1.60)

$$A|B_i\rangle = \sum_{k=1}^{n} |B_k\rangle \langle B_k | A | B_i \rangle = \sum_{k=1}^{n} a_{ki} \, |B_k\rangle, \tag{1.145}$$

and the matrix that represents the linear operator A in the $|B_k\rangle$ basis is $a_{ki} = \langle B_k|A|B_i\rangle$. If a unitary operator U maps these basis vectors into $|B'_k\rangle = U|B_k\rangle$, then in this new basis the matrix that represents A as in (1.138) is

$$\begin{aligned} a'_{\ell i} &= \langle B'_\ell|A|B'_i\rangle = \langle B_\ell|U^\dagger\, A\, U|B_i\rangle \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n}\langle B_\ell|U^\dagger|B_j\rangle\langle B_j|A|B_k\rangle\langle B_k|U|B_i\rangle = \sum_{j=1}^{n}\sum_{k=1}^{n} u^\dagger_{\ell j} a_{jk} u_{ki} \end{aligned} \tag{1.146}$$

or $a' = u^\dagger\, a\, u$ in matrix notation. □

Example 1.25 (Inner-product rules) In Dirac's notation, the rules (1.78—1.81), of a positive-definite inner product are

$$\begin{aligned} \langle f|g\rangle &= \langle g|f\rangle^* \\ \langle f|z_1 g_1 + z_2 g_2\rangle &= z_1\langle f|g_1\rangle + z_2\langle f|g_2\rangle \\ \langle z_1 f_1 + z_2 f_2|g\rangle &= z_1^*\langle f_1|g\rangle + z_2^*\langle f_2|g\rangle \\ \langle f|f\rangle \geq 0 \text{ and } \langle f|f\rangle = 0 &\iff f = 0. \end{aligned} \tag{1.147}$$

States in Dirac notation often 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 1.26 (Gram Schmidt) In Dirac notation, the formula (1.122) for the kth orthogonal linear combination of the vectors $|V_\ell\rangle$ is

$$|u_k\rangle = |V_k\rangle - \sum_{i=1}^{k-1}|U_i\rangle\langle U_i|V_k\rangle = \left(I - \sum_{i=1}^{k-1}|U_i\rangle\langle U_i|\right)|V_k\rangle \tag{1.148}$$

and the formula (1.123) for the kth orthonormal linear combination of the vectors $|V_\ell\rangle$ is

$$|U_k\rangle = \frac{|u_k\rangle}{\sqrt{\langle u_k|u_k\rangle}}. \tag{1.149}$$

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 $|j\rangle$ with $f_j = \langle j|f\rangle$ and $g^*_j = \langle g|j\rangle$, we can use the expansion (1.142)

of the identity operator to write these inner products as

$$\begin{aligned} \langle g|f\rangle = \langle g|I|f\rangle &= \sum_{j=1}^{n} \langle g|j\rangle\langle j|f\rangle = \sum_{j=1}^{n} g_j^* f_j \\ \langle g|A|f\rangle = \langle g|IAI|f\rangle &= \sum_{j,\ell=1}^{n} \langle g|j\rangle\langle j|A|\ell\rangle\langle \ell|f\rangle = \sum_{j,\ell=1}^{n} g_j^*\, A_{j\ell}\, f_\ell \end{aligned} \tag{1.150}$$

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

$$f = \begin{pmatrix} \langle 1|f\rangle \\ \langle 2|f\rangle \\ \vdots \\ \langle n|f\rangle \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} \tag{1.151}$$

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

$$g^\dagger = (\langle 1|g\rangle^*, \langle 2|g\rangle^*, \ldots, \langle n|g\rangle^*) = (g_1^*, g_2^*, \ldots, g_n^*) \tag{1.152}$$

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

One can compute the inner product $\langle g, f\rangle$ of two vectors f and g by doing the sum (1.150) of $g_j^* f_j$ over the index j only if one knows their components f_j and g_j which are their inner products $f_j = \langle j|f\rangle$ and $g_j = \langle j|g\rangle$ with the orthonormal states $|j\rangle$ of some basis. Thus an inner product implies the existence of an orthonormal basis and a representation of the identity operator

$$I = \sum_{j=1}^{n} |j\rangle\langle j|. \tag{1.153}$$

If we switch to a different basis, say from $|k\rangle$'s to $|\alpha_k\rangle$'s, then the components of the column vectors change from $f_k = \langle k|f\rangle$ to $f'_k = \langle \alpha_k|f\rangle$, 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 (1.143)

$$\begin{aligned}\langle g|f\rangle &= \sum_{k=1}^{n}\langle g|k\rangle\langle k|f\rangle = \sum_{k=1}^{n}\langle g|\alpha_k\rangle\langle\alpha_k|f\rangle \\ \langle g|A|f\rangle &= \sum_{k,\ell=1}^{n}\langle g|k\rangle\langle k|A|\ell\rangle\langle\ell|f\rangle = \sum_{k,\ell=1}^{n}\langle g|\alpha_k\rangle\langle\alpha_k|A|\alpha_\ell\rangle\langle\alpha_\ell|f\rangle.\end{aligned} \tag{1.154}$$

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

$$U = \sum_{k=1}^{n}|\alpha_k\rangle\langle k| \tag{1.155}$$

maps the ket $|\ell\rangle$ of one orthonormal basis into that $|\alpha_\ell\rangle$ of another

$$U|\ell\rangle = \sum_{k=1}^{n}|\alpha_k\rangle\langle k|\ell\rangle = \sum_{k=1}^{n}|\alpha_k\rangle\,\delta_{k\ell} = |\alpha_\ell\rangle. \tag{1.156}$$

Example 1.27 (Simple change of basis) If the ket $|\alpha_k\rangle$ of the new basis is simply $|\alpha_k\rangle = |k+1\rangle$ with $|\alpha_n\rangle = |n+1\rangle \equiv |1\rangle$, then the operator that maps the n kets $|k\rangle$ into the kets $|\alpha_k\rangle$ is

$$U = \sum_{k=1}^{n}|\alpha_k\rangle\langle k| = \sum_{k=1}^{n}|k+1\rangle\langle k|. \tag{1.157}$$

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

To compute the inner product (U, V) of two vectors U and V, one needs the components U_i and V_i of these vectors in order to do the sum (1.89) of $U_i^* V_i$ over the index i.

1.13 Adjoints of operators

In Dirac's notation, the most general linear operator (1.136) on an n-dimensional vector space is a sum of outer products $z\,|k\rangle\langle\ell|$ in which z is a complex number and the kets $|k\rangle$ and $|\ell\rangle$ are two of the n orthonormal kets that make up a basis for the space. The **adjoint** (1.135) of this basic linear operator is

$$(z\,|k\rangle\langle\ell|)^\dagger = z^*\,|\ell\rangle\langle k|. \tag{1.158}$$

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

$$A = IAI = \sum_{k,\ell=1}^{n} |k\rangle\langle k|A|\ell\rangle\langle\ell| \tag{1.159}$$

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

$$A^\dagger = \sum_{k,\ell=1}^{n} |\ell\rangle\langle\ell|A^\dagger|k\rangle\langle k| = \sum_{k,\ell=1}^{n} |\ell\rangle\langle k|A|\ell\rangle^*\langle k|. \tag{1.160}$$

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

$$A^\dagger_{\ell k} = \langle\ell|A^\dagger|k\rangle = \langle k|A|\ell\rangle^* = A^*_{\ell k} = A^{*\mathsf{T}}_{k\ell} \tag{1.161}$$

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

$$\langle g|A^\dagger f\rangle = \langle g|A^\dagger|f\rangle = \langle f|A|g\rangle^* = \langle f|Ag\rangle^* = \langle Ag|f\rangle. \tag{1.162}$$

Taking the adjoint of the adjoint is by (1.158)

$$\left[(z\,|k\rangle\langle\ell|)^\dagger\right]^\dagger = [z^*\,|\ell\rangle\langle k|]^\dagger = z\,|k\rangle\langle\ell| \tag{1.163}$$

the same as doing nothing at all. This also follows from the matrix formula (1.161) because both $(A^*)^* = A$ and $(A^\mathsf{T})^\mathsf{T} = A$, and so

$$\left(A^\dagger\right)^\dagger = \left(A^{*\mathsf{T}}\right)^{*\,\mathsf{T}} = A \tag{1.164}$$

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

$$(g, A^\dagger f) = (A\,g, f) = (f, A\,g)^*. \tag{1.165}$$

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

$$(g, A^{\dagger\dagger} f) = (A^\dagger g, f) = (f, A^\dagger g)^* = (Af, g)^* = (g, Af). \tag{1.166}$$

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 (1.161), the matrix elements of a self-adjoint linear operator

A satisfy $\langle k|A^\dagger|\ell\rangle = \langle \ell|A|k\rangle^* = \langle k|A|\ell\rangle$ in any orthonormal basis. So a matrix that represents a hermitian operator is equal to the transpose of its complex conjugate

$$A_{k\ell} = \langle k|A|\ell\rangle = \langle k|A^\dagger|\ell\rangle = \langle \ell|A|k\rangle^* = A^{*\mathsf{T}}_{k\ell} = A^\dagger_{k\ell}. \tag{1.167}$$

We also have

$$\langle g|\,A\,|f\rangle = \langle A\,g|f\rangle = \langle f|A\,g\rangle^* = \langle f|\,A\,|g\rangle^* \tag{1.168}$$

and in pre-Dirac notation

$$(g, A\,f) = (A\,g, f) = (f, A\,g)^*. \tag{1.169}$$

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 1.29, hermitian matrices have real eigenvalues and complete sets of orthonormal eigenvectors. Hermitian operators and matrices represent physical variables in quantum mechanics.

Example 1.28 (Fierz identities for $n \times n$ hermitian matrices) The n^2 $n \times n$ hermitian matrices t^a form a vector space with an inner product $\langle a|b\rangle$ (section 1.6) defined by the trace (1.22) $\langle a|b\rangle = \mathrm{Tr}(t^a\,t^b)$. One can use the Gram-Schmidt method (section 1.10) to make them orthonormal, so that

$$\langle a|b\rangle = \mathrm{Tr}(t^a\,t^b) = \sum_{i,k=1}^{n} t^a_{ik}\,t^b_{ki} = \delta_{ab}. \tag{1.170}$$

Then the sum of their n^2 outer products (1.22) is the identity matrix of the n^2-dimensional vector space

$$\left(\sum_{a=1}^{n^2} |a\rangle\langle a|\right)_{ij,k\ell} = \sum_{a=1}^{n^2} t^a_{ij}\,t^a_{k\ell} = I_{ik,\ell j} = \delta_{i\ell}\,\delta_{kj} \tag{1.171}$$

because

$$t^b_{ij} = (|b\rangle)_{ij} = \sum_{a=1}^{n^2} (|a\rangle)_{ij}\langle a|b\rangle = \sum_{a=1}^{n^2} t^a_{ij}\mathrm{Tr}(t^a\,t^b) = \sum_{a=1}^{n^2}\sum_{k,\ell=1}^{n} t^a_{ij}t^a_{k\ell}\,t^b_{\ell k}. \tag{1.172}$$

(Markus Fierz, 1912–2006) □

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

$$|f\rangle = \sum_{k=1}^{n} z_k \, |k\rangle \tag{1.173}$$

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

$$|f\rangle = \sum_{k=1}^{n} x_k \, |k\rangle \tag{1.174}$$

of real orthonormal basis vectors, $x_k^* = x_k$ and $|k\rangle^* = |k\rangle$.

A real linear operator A on a real vector space

$$A = \sum_{k,\ell=1}^{n} |k\rangle\langle k|A|\ell\rangle\langle\ell| = \sum_{k,\ell=1}^{n} |k\rangle A_{k\ell}\langle\ell| \tag{1.175}$$

is represented by a real matrix $A_{k\ell}^* = A_{k\ell}$. A real linear operator A that is self adjoint on a real vector space satisfies the condition (1.169) of hermiticity but with the understanding that complex conjugation has no effect

$$(g, A\, f) = (A\, g, f) = (f, A\, g)^* = (f, A\, g). \tag{1.176}$$

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 $A_{k\ell}$ that represents it (in a real basis) is real and hermitian, and so is symmetric $A_{k\ell} = A_{\ell k}^* = A_{\ell k}$.

1.16 Unitary operators

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

$$U\, U^\dagger = U^\dagger\, U = I. \tag{1.177}$$

Any operator that changes from one orthonormal basis $|k\rangle$ to another $|\alpha_k\rangle$

$$U = \sum_{k=1}^{n} |\alpha_k\rangle\langle k| \tag{1.178}$$

is unitary since

$$\begin{aligned} UU^\dagger &= \sum_{k=1}^{n} |\alpha_k\rangle\langle k| \sum_{\ell=1}^{n} |\ell\rangle\langle\alpha_\ell| = \sum_{k,\ell=1}^{n} |\alpha_k\rangle\langle k|\ell\rangle\langle\alpha_\ell| \\ &= \sum_{k,\ell=1}^{n} |\alpha_k\rangle\delta_{k,\ell}\langle\alpha_\ell| = \sum_{k=1}^{n} |\alpha_k\rangle\langle\alpha_k| = I \end{aligned} \tag{1.179}$$

as well as

$$U^\dagger U = \sum_{\ell=1}^{n} |\ell\rangle\langle\alpha_\ell| \sum_{k=1}^{n} |\alpha_k\rangle\langle k| = \sum_{k=1}^{n} |k\rangle\langle k| = I. \tag{1.180}$$

A unitary operator maps every orthonormal basis $|k\rangle$ into another orthonormal basis $|\alpha_k\rangle$. For if $|\alpha_k\rangle = U|k\rangle$, then the vectors $|\alpha_k\rangle$ are orthonormal $\langle\alpha_k|\alpha_\ell\rangle = \delta_{k,\ell}$ (exercise 1.22). They also are complete because they provide a resolution of the identity operator

$$\sum_{k=1}^{n} |\alpha_k\rangle\langle\alpha_k| = \sum_{k=1}^{n} U|k\rangle\langle k|U^\dagger = U\,I\,U^\dagger = U\,U^\dagger = I. \tag{1.181}$$

If we multiply the relation $|\alpha_k\rangle = U|k\rangle$ by the bra $\langle k|$ and then sum over the index k, we get

$$\sum_{k=1}^{n} |\alpha_k\rangle\langle k| = \sum_{k=1}^{n} U|k\rangle\langle k| = U \sum_{k=1}^{n} |k\rangle\langle k| = U. \tag{1.182}$$

Every unitary operator maps every orthonormal basis into another orthonormal basis or into itself.

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 $(g, f) = (g, U^\dagger U f) = (Ug, Uf)$.

Unitary matrices have unimodular determinants, $|\det U| = 1$, because the determinant of the product of two matrices is the product of their determinants (1.222) and because transposition doesn't change the value of a determinant (1.205)

$$1 = \det I = \det(UU^\dagger) = \det U \det U^\dagger = \det U (\det(U^\mathsf{T}))^* = \det U (\det U)^*. \tag{1.183}$$

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

$$OO^\mathsf{T} = O^\mathsf{T} O = I. \tag{1.184}$$

1.17 Hilbert spaces

We have mainly 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 its value at x'

$$\delta_{x'}[f] = \int \delta(x - x')\, f(x)\, dx = f(x'). \tag{1.185}$$

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

$$U(a) = \int |x + a\rangle\langle x|\, dx. \tag{1.186}$$

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

$$U(\theta) = \int |x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta\rangle\langle x, y|\, dxdy \tag{1.187}$$

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**

$$K\left(z\psi + w\phi\right) = K\left(z|\psi\rangle + w|\phi\rangle\right) = z^* K|\psi\rangle + w^* K|\phi\rangle = z^* K\psi + w^* K\phi \tag{1.188}$$

and **antiunitary**

$$(K\phi, K\psi) = \langle K\phi|K\psi\rangle = (\phi,\psi)^* = \langle\phi|\psi\rangle^* = \langle\psi|\phi\rangle = (\psi,\phi)\,. \tag{1.189}$$

The adjoint $K^\dagger$ of an antiunitary operator K is defined by $\langle K^\dagger\phi|\psi\rangle = \langle\phi|K|\psi\rangle^*$ so that $\langle K^\dagger K\phi|\psi\rangle = \langle K\phi|K\psi\rangle^* = \langle\phi|\psi\rangle^{**} = \langle\phi|\psi\rangle$.

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 probabilities

$$|\langle\phi'|\psi'\rangle|^2 = |\langle\phi|\psi\rangle|^2. \tag{1.190}$$

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 occurs when a 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

$$\det A = |A| = A_{11}A_{22} - A_{21}A_{12}. \tag{1.191}$$

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

$$\det A = \sum_{i=1}^{2}\sum_{j=1}^{2} e_{ij}A_{i1}A_{j2} = \sum_{i=1}^{2}\sum_{j=1}^{2} e_{ij}A_{1i}A_{2j}. \tag{1.192}$$

It's also true that

$$e_{k\ell} \det A = \sum_{i=1}^{2} \sum_{j=1}^{2} e_{ij} A_{ik} A_{j\ell}. \tag{1.193}$$

Example 1.29 (Area of a parallelogram) Two 2-vectors $V = (V_1, V_2)$ and $W = (W_1, W_2)$ define a parallelogram whose area is the absolute value of a 2×2 determinant

$$\text{area}(V, W) = \left| \det \begin{pmatrix} V_1 & V_2 \\ W_1 & W_2 \end{pmatrix} \right| = | \, V_1 \, W_2 - V_2 \, W_1 \, | \, . \tag{1.194}$$

To check this formula, rotate the coordinates so that the 2-vector V runs from the origin along the x-axis. Then $V_2 = 0$, and the determinant is $V_1 W_2$ which is the base V_1 of the parallelogram times its height W_2. □

These definitions (1.191–1.193) extend to any square matrix. If A is a 3×3 matrix, then its determinant is

$$\det A = \sum_{i,j,k=1}^{3} e_{ijk} \, A_{i1} A_{j2} A_{k3} = \sum_{i,j,k=1}^{3} e_{ijk} \, A_{1i} A_{2j} A_{3k} \tag{1.195}$$

in which e_{ijk} is the totally antisymmetric Levi-Civita symbol whose nonzero values are

$$e_{123} = e_{231} = e_{312} = 1, \quad \text{and} \quad e_{213} = e_{132} = e_{321} = -1. \tag{1.196}$$

The symbol vanishes whenever an index appears twice, thus

$$e_{111} = e_{112} = e_{113} = e_{222} = e_{221} = e_{223} = e_{333} = e_{331} = e_{332} = 0 \tag{1.197}$$

and so forth. The sums over i, j, and k run from 1 to 3

$$\begin{aligned} \det A = & \sum_{i=1}^{3} A_{i1} \sum_{j,k=1}^{3} e_{ijk} \, A_{j2} A_{k3} \\ = & A_{11} \, (A_{22} A_{33} - A_{32} A_{23}) + A_{21} \, (A_{32} A_{13} - A_{12} A_{33}) \\ & + A_{31} \, (A_{12} A_{23} - A_{22} A_{13}) \, . \end{aligned} \tag{1.198}$$

The **minor** $M_{i\ell}$ of the matrix A is the 2×2 determinant of the matrix A without row i and column ℓ, and the **cofactor** $C_{i\ell}$ is the minor $M_{i\ell}$

multiplied by $(-1)^{i+\ell}$. Thus $\det A$ is the sum

$$\begin{aligned}\det A &= A_{11}(-1)^2\,(A_{22}A_{33} - A_{32}A_{23}) + A_{21}(-1)^3\,(A_{12}A_{33} - A_{32}A_{13}) \\ &\quad + A_{31}(-1)^4\,(A_{12}A_{23} - A_{22}A_{13}) \\ &= A_{11}C_{11} + A_{21}C_{21} + A_{31}C_{31}\end{aligned} \tag{1.199}$$

of the products $A_{i1}C_{i1} = A_{i1}(-1)^{i+1}M_{i1}$ where

$$\begin{aligned}C_{11} &= (-1)^2 M_{11} = A_{22}A_{33} - A_{23}A_{32} \\ C_{21} &= (-1)^3 M_{21} = A_{32}A_{13} - A_{12}A_{33} \\ C_{31} &= (-1)^4 M_{31} = A_{12}A_{23} - A_{22}A_{13}.\end{aligned} \tag{1.200}$$

Example 1.30 (Volume of a parallelepiped) 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

$$\begin{vmatrix} U_1 & U_2 & U_3 \\ V_1 & V_2 & V_3 \\ W_1 & W_2 & W_3 \end{vmatrix} = \sum_{ijk=1}^{3} e_{ijk}\,U_iV_jW_k = \sum_{i=1}^{3} U_i\,(\boldsymbol{V} \times \boldsymbol{W})_i = \boldsymbol{U} \cdot (\boldsymbol{V} \times \boldsymbol{W}). \tag{1.201}$$

The absolute value of this **scalar triple product** is the volume of the parallelepiped defined by U, V, and W as one can see by placing the parallelepiped so the vector U runs from the origin along the x axis. The 3×3 determinant (1.201) then is $U_1(V_2W_3 - V_3W_2)$ which is the height of the parallelepiped times the area (1.194) of its base. □

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

$$\det A = \sum_{i_1 i_2 \ldots i_n = 1}^{n} e_{i_1 i_2 \ldots i_n} A_{i_1 1} A_{i_2 2} \ldots A_{i_n n} \tag{1.202}$$

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

$$\det A = \sum_{i=1}^{n} A_{ik}C_{ik} = \sum_{k=1}^{n} A_{ik}C_{ik} \tag{1.203}$$

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

$$\begin{aligned} e_{k_1k_2\ldots k_n}\det A &= \sum_{i_1i_2\ldots i_n=1}^{n} e_{i_1i_2\ldots i_n}A_{i_1k_1}A_{i_2k_2}\ldots A_{i_nk_n} \\ &= \sum_{i_1i_2\ldots i_n=1}^{n} e_{i_1i_2\ldots i_n}A_{k_1i_1}A_{k_2i_2}\ldots A_{k_ni_n}. \end{aligned} \tag{1.204}$$

In particular, since $e_{12\ldots n}=1$, the determinant of the transpose of a matrix is equal to the determinant (1.202) of the matrix

$$\det A^{\mathsf{T}} = \sum_{i_1i_2\ldots i_n=1}^{n} e_{i_1i_2\ldots i_n}A_{1i_1}A_{2i_2}\ldots A_{ni_n} = \det A. \tag{1.205}$$

The interchange $A\to A^{\mathsf{T}}$ of the rows and columns of a matrix has no effect on its determinant.

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

$$\det A' = \sum_{i_1i_2\ldots i_n=1}^{n} e_{i_1i_2\ldots i_n}\left(A_{i_11}+xA_{i_12}\right)A_{i_22}\ldots A_{i_nn} \tag{1.206}$$

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

$$\delta\det A = \sum_{i_1i_2\ldots i_n=1}^{n} x\, e_{i_1i_2\ldots i_n}\, A_{i_12}A_{i_22}\ldots A_{i_nn} = 0 \tag{1.207}$$

because it is proportional to a sum of products of a factor $e_{i_1i_2\ldots i_n}$ that is antisymmetric in i_1 and i_2 and a factor $A_{i_12}A_{i_22}$ that is symmetric in these indices. For instance, when i_1 and i_2 are 5 & 7 and 7 & 5, the two terms cancel

$$e_{57\ldots i_n}A_{52}A_{72}\ldots A_{i_nn} + e_{75\ldots i_n}A_{72}A_{52}\ldots A_{i_nn} = 0 \tag{1.208}$$

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

By repeated additions of x_2A_{i2}, x_3A_{i3}, and so forth to A_{i1}, we can change the first column of the matrix A to a linear combination of all the columns

$$A_{i1}\longrightarrow A_{i1}+\sum_{k=2}^{n}x_kA_{ik} \tag{1.209}$$

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

$$A_{i\ell} \longrightarrow A_{i\ell} + \sum_{k=1, k\neq\ell}^{n} y_k A_{ik} \tag{1.210}$$

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

$$\sum_{k=1}^{n} y_k A_{ik} = 0 \quad \text{for } i = 1, \dots n \tag{1.211}$$

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 (1.202) would vanish. **Thus the determinant of any matrix whose columns are linearly dependent must vanish**.

Now suppose that the columns of an $n \times n$ matrix are linearly independent. Then the determinant of the matrix cannot vanish because any linearly independent set of n vectors in a vector space of n dimensions 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 $(1, 0, 0, \dots 0)$, say $(c_1, 0, 0, \dots 0)$. Similar operations will convert column 2 into a nonzero multiple of the column vector $(0, 1, 0, \dots 0)$, say $(0, c_2, 0, \dots 0)$. Continuing in this way, we may convert the matrix A to a matrix with nonzero entries c_i along the main diagonal and zeros everywhere else. The determinant $\det A$ then is the product $c_1 c_2 \dots c_n$ of the nonzero diagonal entries c_i's, and so $\det A$ cannot 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

$$A_{ki} \longrightarrow A_{ki} + \sum_{\ell=1, \ell\neq k}^{n} z_\ell A_{\ell i} \tag{1.212}$$

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 or in terms of the rows as in the definitions (1.202) & 1.204). These and other properties of determinants follow from a study of **permutations** (section 11.13). Detailed proofs are in (Aitken, 1959).

Let us return for a moment to Laplace's expansion (1.203) 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

$$\det A = \sum_{i=1}^{n} A_{ik} C_{ik} = \sum_{i=1}^{n} A_{ki} C_{ki} \tag{1.213}$$

in order to prove that

$$\delta_{k\ell} \det A = \sum_{i=1}^{n} A_{ik} C_{i\ell} = \sum_{i=1}^{n} A_{ki} C_{\ell i}. \tag{1.214}$$

For $k = \ell$, this formula just repeats Laplace's expansion (1.213). But for $k \neq \ell$, it is Laplace's expansion for the determinant of a matrix that has two copies of its kth column. Since the determinant of a matrix with two identical columns vanishes, the rule (1.214) also is true for $k \neq \ell$.

The rule (1.214) 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 the determinant of the matrix

$$\left(A^{-1}\right)_{\ell i} = \frac{C_{i\ell}}{\det A} \quad \text{or} \quad A^{-1} = \frac{C^{\mathsf{T}}}{\det A}. \tag{1.215}$$

To verify this formula, we use it for A^{-1} in the product $A^{-1}A$ and note that by (1.214) the ℓkth entry of the product $A^{-1}A$ is just $\delta_{\ell k}$

$$\left(A^{-1}A\right)_{\ell k} = \sum_{i=1}^{n} \left(A^{-1}\right)_{\ell i} A_{ik} = \sum_{i=1}^{n} \frac{C_{i\ell}}{\det A} A_{ik} = \delta_{\ell k}. \tag{1.216}$$

Example 1.31 (Inverting a 2×2 Matrix) Our formula (1.215) for the inverse of the general 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{1.217}$$

gives

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{1.218}$$

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

The simple example of matrix multiplication

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a & xa+b & ya+zb+c \\ d & xd+e & yd+ze+f \\ g & xg+h & yg+zh+i \end{pmatrix} \tag{1.219}$$

shows that the operations (1.210) 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

$$\begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix} \tag{1.220}$$

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:

$$\det \begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} = \det \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix}. \tag{1.221}$$

By using Laplace's expansion (1.203) 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**

$$\det A \det B = \det AB. \tag{1.222}$$

Example 1.32 (Two 2×2 Matrices) When the matrices A and B are both 2×2, the two sides of (1.221) are

$$\begin{aligned} \det \begin{pmatrix} A & 0 \\ -I & B \end{pmatrix} &= \det \begin{pmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{pmatrix} \\ &= a_{11}a_{22} \det B - a_{21}a_{12} \det B = \det A \det B \end{aligned} \tag{1.223}$$

and

$$\det \begin{pmatrix} A & AB \\ -I & 0 \end{pmatrix} = \det \begin{pmatrix} a_{11} & a_{12} & (ab)_{11} & (ab)_{12} \\ a_{21} & a_{22} & (ab)_{21} & (ab)_{22} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{1.224}$$
$$= (-1)C_{42} = (-1)(-1)\det AB = \det AB$$

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

$$|ABC \dots Z| = |A||B| \dots |Z|. \tag{1.225}$$

The product rule (1.222) implies that $\det\left(A^{-1}\right)$ is $1/\det A$ since

$$1 = \det I = \det\left(AA^{-1}\right) = \det A \det\left(A^{-1}\right). \tag{1.226}$$

Example 1.33 (Derivative of the logarithm of a determinant) We see from our formula (1.213) for $\det A$ that its derivative with respect to any given element A_{ik} is the corresponding cofactor C_{ik}

$$\frac{\partial \det A}{\partial A_{ik}} = C_{ik} \tag{1.227}$$

because the cofactors C_{ij} and C_{jk} for all j are independent of A_{ik}. Thus the derivative of the logarithm of this determinant with respect to any parameter β is

$$\frac{\partial \ln \det A}{\partial \beta} = \frac{1}{\det A} \sum_{ik} \frac{\partial \det A}{\partial A_{ik}} \frac{\partial A_{ik}}{\partial \beta} = \sum_{ik} \frac{C_{ik}}{\det A} \frac{\partial A_{ik}}{\partial \beta}$$
$$= \sum_{ik} A^{-1}_{ki} \frac{\partial A_{ik}}{\partial \beta} = \mathrm{Tr}\left(A^{-1} \frac{\partial A}{\partial \beta}\right). \tag{1.228}$$

□

Example 1.34 (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 the absolute values of determinants invariant, to make a matrix **upper triangular**, a form in which its determinant is just the product of the factors on its diagonal. Thus to make the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -6 & 3 \\ 4 & 2 & -5 \end{pmatrix} \tag{1.229}$$

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

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

and then subtract four times the first row from the third

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 5 \\ 0 & -6 & -9 \end{pmatrix}. \tag{1.230}$$

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

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

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

$$|A| = 1(-2)(-24) = 48. \tag{1.231}$$

□

Incidentally, Gauss, Jordan, and modern mathematicians have developed much faster ways of computing determinants and matrix inverses than those (1.203 & 1.215) due to Laplace. Sage, Octave, Matlab, Maple, Mathematica, and Python use these modern techniques, which are freely available as programs in C and FORTRAN from www.netlib.org/lapack.

Example 1.35 (Using Matlab) The Matlab command to make the matrix (1.229) is A = [1 2 1; -2 -6 3; 4 2 -5] . The command d = det(A) gives its determinant, d = 48, and Ainv = $\mathrm{A}^{(-1)}$ gives its inverse

```
Ainv =    0.5000    0.2500    0.2500
          0.0417   -0.1875   -0.1042
          0.4167    0.1250   -0.0417 .
```

□

The **permanent** of a square $n \times n$ matrix A_{ik} is the sum over all permutations $1, 2, \ldots, n \to s_1, s_2, \ldots, s_n$ of the products $A_{s_1 1} A_{s_2 2} \cdots A_{s_n n}$

$$\mathrm{perm}(A) = \sum_s A_{s_1 1} A_{s_2 2} \cdots A_{s_n n}. \tag{1.232}$$

1.21 Jacobians

When one changes variables in a multiple integral from coordinates x_1, x_2 and area element $dx_1 dx_2$, one must find the new element of area in terms of the new variables y_1, y_2. If $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ are unit vectors in the x_1 and x_2 directions, then as the new coordinates (y_1, y_2) change by dy_1 and dy_2, the point they represent moves by

$$\boldsymbol{dy}^1 = \left(\frac{\partial x_1}{\partial y_1} \hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial y_1} \hat{\boldsymbol{x}}_2 \right) dy_1 \quad \text{and by} \quad \boldsymbol{dy}^2 = \left(\frac{\partial x_1}{\partial y_2} \hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial y_2} \hat{\boldsymbol{x}}_2 \right) dy_2. \tag{1.233}$$

These vectors, $\boldsymbol{dy}^1$ and $\boldsymbol{dy}^2$ define a parallelogram whose area (1.194) is the absolute value of a determinant

$$\text{area}(\boldsymbol{dy}^1, \boldsymbol{dy}^2) = \left| \det \begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_2}{\partial y_1} \\ \dfrac{\partial x_1}{\partial y_2} & \dfrac{\partial x_2}{\partial y_2} \end{pmatrix} \right| dy_1\, dy_2. \tag{1.234}$$

The determinant itself is a **jacobian**

$$J = J(x/y) = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \det \begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_2}{\partial y_1} \\ \dfrac{\partial x_1}{\partial y_2} & \dfrac{\partial x_2}{\partial y_2} \end{pmatrix}. \tag{1.235}$$

The two equal integrals are

$$\iint_{R_x} f(x_1, x_2)\, dx_1 dx_2 = \iint_{R_y} f \circ x(y_1, y_2)) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1 dy_2 \tag{1.236}$$

in which $f \circ x(y_1, y_2) = f(x_1(y_1, y_2), x_2(y_1, y_2))$ and R_x and R_y are the same region in the two coordinate systems.

In 3 dimensions, with $j = 1, 2$, and 3, the 3 vectors

$$\boldsymbol{dy}^j = \left(\frac{\partial x_1}{\partial y_j} \hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial y_j} \hat{\boldsymbol{x}}_2 + \frac{\partial x_3}{\partial y_j} \hat{\boldsymbol{x}}_3 \right) dy_j \tag{1.237}$$

define a parallelepiped whose volume (1.201) is the absolute value of the determinant

$$\text{volume}(\boldsymbol{dy}^1, \boldsymbol{dy}^2, \boldsymbol{dy}^3) = \left| \det \begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_3}{\partial y_1} \\ \dfrac{\partial x_1}{\partial y_2} & \dfrac{\partial x_2}{\partial y_2} & \dfrac{\partial x_3}{\partial y_2} \\ \dfrac{\partial x_1}{\partial y_3} & \dfrac{\partial x_2}{\partial y_3} & \dfrac{\partial x_3}{\partial y_2} \end{pmatrix} \right| dy_1\, dy_2\, dy_3. \tag{1.238}$$

The equal integrals are

$$\iiint_{R_x} f(\vec{x})\, d^3x = \iiint_{R_y} f \circ x(\vec{y}))\, \left|\frac{\partial(x_1, x_2, x_3)}{\partial(y_1, y_2, y_3)}\right| d^3y \tag{1.239}$$

in which $d^3x = dx_1 dx_2 dx_3$, $d^3y = dy_1 dy_2 dy_3$, $f \circ x(\vec{y}) = f(x_1(\vec{y}), x_2(\vec{y}), x_3(\vec{y}))$, and R_x and R_y are the same region in the two coordinate systems.

For n-dimensional integrals over $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, the rule is similar

$$\int_{R_x} f(x)\, d^nx = \int_{R_y} f \circ x(y)\, \left|\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}\right| d^ny \tag{1.240}$$

and uses the absolute value of the n-dimensional jacobian

$$J = J(x/y) = \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}. \tag{1.241}$$

Since the determinant of the transpose of a matrix is the same (1.205) as the determinant of the matrix, some people write jacobians with their rows and columns interchanged.

1.22 Systems of linear equations

Suppose we wish to solve the system of n linear equations

$$\sum_{k=1}^{n} A_{ik} x_k = y_i \tag{1.242}$$

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 (1.215), and we may multiply both sides of $A\,x = y$ by A^{-1} and so find $x = A^{-1}\,y$. When A is nonsingular, this is the unique solution to (1.242).

When A is singular, its determinant vanishes $\det(A) = 0$, 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 non-zero vector z. Thus if x satisfies $A\,x = y$, then so does $x + cz$ for any constant c because $A(x + cz) = A\,x + c\,A\,z = y$. So if $\det(A) = 0$, then the equation $A\,x = y$ may have solutions, but they will not be unique. Whether equation (1.242) 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 (1.242) has solutions only when the n-vector y lies in that subspace.

A system of $m < n$ equations

$$\sum_{k=1}^{n} A_{ik} x_k = y_i \quad \text{for} \quad i = 1, 2, \ldots, m \tag{1.243}$$

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 may not be unique.

1.23 Linear least squares

Suppose we have a system of $m > n$ equations in n unknowns x_k

$$\sum_{k=1}^{n} A_{ik} x_k = y_i \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{1.244}$$

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

$$E = \sum_{i=1}^{m} \left(y_i - \sum_{k=1}^{n} A_{ik} x_k \right)^2 . \tag{1.245}$$

The minimizing values x_k make the n derivatives of E vanish

$$\frac{\partial E}{\partial x_\ell} = 0 = \sum_{i=1}^{m} 2 \left(y_i - \sum_{k=1}^{n} A_{ik} x_k \right) (-A_{i\ell}) \tag{1.246}$$

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 (1.39); if it also is positive (1.40), then it has an inverse, and our **least-squares solution** is

$$x = \left(A^{\mathsf{T}} A \right)^{-1} A^{\mathsf{T}} y. \tag{1.247}$$

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

$$x = \left(A^\dagger A \right)^{-1} A^\dagger \, y. \tag{1.248}$$

The operators $\left(A^{\mathsf{T}} A \right)^{-1} A^{\mathsf{T}}$ and $\left(A^\dagger A \right)^{-1} A^\dagger$ are **pseudoinverses** (section 1.33).

1.24 Lagrange multipliers

The maxima and minima of a function $f(x)$ of $x = (x_1, x_2, \ldots, x_n)$ are among the points at which its gradient vanishes $\nabla f(x) = 0$, that is,

$$\frac{\partial f(x)}{\partial x_j} = 0 \tag{1.249}$$

for $j = 1, \ldots, n$. These are **stationary** points of f.

Example 1.36 (Minimum) For instance, if $f(x) = x_1^2 + 2x_2^2 + 3x_3^2$, then its minimum is at

$$\nabla f(x) = (2x_1, 4x_2, 6x_3) = 0 \tag{1.250}$$

that is, at $x_1 = x_2 = x_3 = 0$. □

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

In the case of one constraint $c(x) = 0$, we expect the gradient $\nabla f(x)$ to 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. That is, $\nabla f(x)$ and $\nabla c(x)$ must be parallel. So the extrema of $f(x)$ subject to the constraint $c(x) = 0$ satisfy the equations

$$\nabla f(x) = \lambda \nabla c(x) \quad \text{and} \quad c(x) = 0. \tag{1.251}$$

These $n + 1$ equations define the extrema of the unconstrained function

$$L(x, \lambda) = f(x) - \lambda\, c(x) \tag{1.252}$$

of the $n + 1$ variables $x_1, \ldots, x_n, \lambda$

$$\frac{\partial L(x,\lambda)}{\partial x_j} = \frac{\partial \left(f(x) - \lambda\, c(x)\right)}{\partial x_j} = 0 \quad \text{and} \quad \frac{\partial L(x,\lambda)}{\partial \lambda} = -c(x) = 0. \tag{1.253}$$

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

In the case of k constraints $c_1(x) = 0$, ..., $c_k(x) = 0$, the projection of ∇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

$$\nabla f = \lambda_1 \nabla c_1 + \ldots + \lambda_k \nabla c_k \tag{1.254}$$

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

$$c_1(x) = 0, \ldots, c_k(x) = 0. \tag{1.255}$$

The $n+k$ equations (1.254 & 1.255) define the extrema of the unconstrained function

$$L(x, \lambda) = f(x) - \lambda_1 \, c_1(x) \ldots - \lambda_k \, c_k(x) \tag{1.256}$$

of the $n + k$ variables x and λ

$$\nabla L(x, \lambda) = \nabla f(x) - \lambda_1 \, \nabla c_1(x) \ldots - \lambda_k \, \nabla c_k(x) = 0 \tag{1.257}$$

and

$$\frac{\partial L(x, \lambda)}{\partial \lambda_j} = \, - c_j(x) = 0 \quad \text{for } \, j = 1, \ldots, k. \tag{1.258}$$

Example 1.37 (Constrained Extrema and Eigenvectors) Suppose we want to find the extrema of a real, symmetric quadratic form

$$f(x) = x^{\mathsf{T}} A \, x = \sum_{i,j=1}^{n} x_i \, A_{ij} \, x_j \tag{1.259}$$

subject to the constraint $c(x) = x \cdot x - 1$ which says that the n-vector x is of unit length. We form the function

$$L(x, \lambda) = x^{\mathsf{T}} A \, x - \lambda \, (x \cdot x - 1) \tag{1.260}$$

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

$$\nabla L(x, \lambda) = 2A \, x - 2\lambda \, x = 0 \quad \text{and} \quad x \cdot x = 1. \tag{1.261}$$

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

$$A \, x = \lambda \, x \quad \text{and} \quad x \cdot x = 1 \tag{1.262}$$

of the real, symmetric matrix A. □

1.25 Eigenvectors and eigenvalues

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

$$A|u\rangle = \lambda |u\rangle \tag{1.263}$$

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

If the vectors $|k\rangle$ for $k = 1, \ldots, n$ form an orthonormal basis for the vector space in which A acts, then we can write the identity operator for

the space as $I = |1\rangle\langle 1| + \ldots + |n\rangle\langle n|$. By inserting this formula for I into the eigenvector equation (1.263), we get

$$\sum_{\ell=1}^{n} \langle k|A|\ell\rangle\langle \ell|u\rangle = \lambda \langle k|u\rangle. \tag{1.264}$$

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

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

$$A\left(\sum_{k\in S} c_k|u_k\rangle\right) = \sum_{k\in S} c_k\, A|u_k\rangle = \sum_{k\in S} c_k\,\lambda_k|u_k\rangle = \sum_{k\in S} c'_k|u_k\rangle \tag{1.265}$$

with $c'_k = c_k\lambda_k$. Eigenvectors span **invariant subspaces**.

Example 1.38 (Eigenvalues of an Orthogonal Matrix) The matrix equation

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ \pm i \end{pmatrix} = e^{\pm i\theta} \begin{pmatrix} 1 \\ \pm i \end{pmatrix} \tag{1.266}$$

tells us that the eigenvectors of this 2×2 orthogonal matrix are $(1, \pm i)$ 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 1.39 (Eigenvalues of an Antisymmetric Matrix) Let us consider an eigenvector equation for a matrix A that is antisymmetric

$$\sum_{k=1}^{n} A_{ik}\, u_k = \lambda\, u_i. \tag{1.267}$$

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

$$\sum_{i,k=1}^{n} u_i\, A_{ik}\, u_k = 0. \tag{1.268}$$

Thus the last two relations imply that

$$0 = \sum_{i,k=1}^{n} u_i\, A_{ik}\, u_k = \lambda \sum_{i=1}^{n} u_i^2 = 0. \tag{1.269}$$

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

1.26 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

$$\sum_{k=1}^{n} A_{ik} V_k = \lambda V_i \quad \text{or} \quad AV = \lambda V \tag{1.270}$$

in matrix notation. Every $n \times n$ matrix A has n eigenvectors $V^{(\ell)}$ and eigenvalues λ_ℓ

$$AV^{(\ell)} = \lambda_\ell V^{(\ell)} \tag{1.271}$$

for $\ell = 1 \dots n$. To see why, we write the top equation (1.270) as

$$\sum_{k=1}^{n} (A_{ik} - \lambda \, \delta_{ik}) \, V_k = 0 \tag{1.272}$$

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 (1.272) say that the columns of the matrix $A - \lambda I$, considered as vectors, are linearly dependent (section 1.8). The columns of a matrix $A - \lambda I$ are linearly dependent if and only if the determinant $|A - \lambda I|$ vanishes (section 1.20). Thus a solution of the eigenvalue equation (1.270) exists if and only if the determinant of $A - \lambda I$ vanishes

$$\det (A - \lambda I) = |A - \lambda I| = 0. \tag{1.273}$$

This vanishing of the determinant of $A - \lambda I$ is the **characteristic equation** of the matrix A. For an $n \times n$ matrix A, it is a polynomial equation of the nth degree in the unknown eigenvalue λ

$$\begin{aligned} 0 = |A - \lambda I| &= |A| + \cdots + (-1)^{n-1} \lambda^{n-1} \, \mathrm{Tr} A + (-1)^n \lambda^n \\ &= P(\lambda, A) = \sum_{k=0}^{n} p_k \, \lambda^k \end{aligned} \tag{1.274}$$

in which $p_0 = |A|$, $p_{n-1} = (-1)^{n-1} \mathrm{Tr} A$, and $p_n = (-1)^n$.

All the p_k's are basis independent. For if S is any nonsingular matrix, then multiplication rules (1.222 & 1.226) for determinants imply that the determinant $|A - \lambda I|$ is invariant when A undergoes a similarity transformation (1.67 & (1.278) 1.284) $A \to A' = S^{-1} A S$

$$\begin{aligned} P(\lambda, A') = P(\lambda, S^{-1} A S) &= |S^{-1} A S - \lambda I| = |S^{-1} (A - \lambda I) S| \\ &= |S^{-1}| |A - \lambda I| |S| = |A - \lambda I| = P(\lambda, A). \end{aligned} \tag{1.275}$$

By the fundamental theorem of algebra (section 6.9), the characteristic

equation (1.274) always has n roots or solutions λ_ℓ lying somewhere in the complex plane. Thus the **characteristic polynomial** $P(\lambda, A)$ has the factored form

$$P(\lambda, A) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda). \tag{1.276}$$

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 (1.272). Thus **every $n \times n$ matrix has n eigenvalues λ_ℓ and n eigenvectors $V^{(\ell)}$**.

The $n \times n$ diagonal matrix $A_{k\ell}^{(d)} = \delta_{k\ell} \, \lambda_\ell$ 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 $A\,V = V\,A_d$ or more explicitly

$$\sum_{k=1}^{n} A_{ik} \, V_{k\ell} = \sum_{k=1}^{n} A_{ik} \, V_k^{(\ell)} = \lambda_\ell \, V_i^{(\ell)} = \sum_{k=1}^{n} V_{ik} \, \delta_{k\ell} \lambda_\ell = \sum_{k=1}^{n} V_{ik} \, A_{k\ell}^{(d)}. \tag{1.277}$$

If the eigenvectors $V_{k\ell}$ are linearly independent, then the matrix V, of which they are the columns, is nonsingular and has an inverse V^{-1}. The similarity transformation

$$V^{-1} \, A \, V = A^{(d)} \tag{1.278}$$

diagonalizes the matrix A.

Example 1.40 (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 $[V, D] = \text{eig}(A)$, then we get

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

and one may check that $A\,V = V\,A_d$ and that $V^{-1}\,A\,V = A_d$. □

Setting $\lambda = 0$ in the factored form (1.276) of $P(\lambda, A)$ and in the characteristic equation (1.274), we see that **the determinant of every $n \times n$ matrix is the product of its n eigenvalues**

$$P(0, A) = |A| = p_0 = \lambda_1 \lambda_2 \dots \lambda_n. \tag{1.279}$$

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

$$\sum_{\ell=1}^{k+1} c_\ell V^{(\ell)} = 0. \tag{1.280}$$

First we multiply this equation from the left by the linear operator A and use the eigenvalue equation (1.271)

$$A \sum_{\ell=1}^{k+1} c_\ell \, V^{(\ell)} = \sum_{\ell=1}^{k+1} c_\ell \, A \, V^{(\ell)} = \sum_{\ell=1}^{k+1} c_\ell \, \lambda_\ell \, V^{(\ell)} = 0. \tag{1.281}$$

Now we multiply the same equation (1.280) by λ_{k+1}

$$\sum_{\ell=1}^{k+1} c_\ell \, \lambda_{k+1} \, V^{(\ell)} = 0 \tag{1.282}$$

and subtract the product (1.282) from (1.281) . The terms with $\ell = k+1$ cancel leaving

$$\sum_{\ell=1}^{k} c_\ell \, (\lambda_\ell - \lambda_{k+1}) \, V^{(\ell)} = 0 \tag{1.283}$$

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**. Similarly, if any $k < n$ 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 (1.274) is associated with a single eigenvector; it is called a **simple eigenvalue**. An eigenvalue λ that is a root of multiplicity n 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. **Thus every nondefective square matrix A can be diagonalized by a similarity transformation**

$$\boldsymbol{V^{-1} A V = A^{(d)}} \tag{1.284}$$

(1.278). The elements of the main diagonal of the matrix $A^{(d)}$ are the eigenvalues of the matrix A. Thus the trace of every nondefective matrix A is the sum of its eigenvalues, $\mathrm{Tr}A = \mathrm{Tr}A^{(d)} = \lambda_a + \cdots + \lambda_n$. The columns of the matrix V are the eigenvectors of the matrix A.

Since the determinant of every matrix A is the product (1.279) of its eigenvalues, $\det A = |A| = \lambda_1\lambda_2 \ldots \lambda_n$, the determinant of every nondefective matrix $A = e^L$ is the exponential of the trace of its logarithm

$$\det A = \exp\left[\mathrm{Tr}\left(\log A\right)\right] \quad \text{and} \quad \det A = \det(e^L) = \exp[\mathrm{Tr}(L)]. \tag{1.285}$$

Example 1.41 (A Defective 2×2 Matrix) Each of the 2×2 matrices

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{1.286}$$

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

1.27 A matrix obeys its characteristic equation

Every square matrix obeys its characteristic equation (1.274). That is, the characteristic equation

$$P(\lambda, A) = |A - \lambda I| = \sum_{k=0}^{n} p_k \, \lambda^k = 0 \tag{1.287}$$

remains true when the matrix A replaces the variable λ

$$P(A, A) = \sum_{k=0}^{n} p_k \, A^k = 0. \tag{1.288}$$

To see why, we use the formula (1.215) for the inverse of the matrix $A - \lambda I$

$$(A - \lambda I)^{-1} = \frac{C(\lambda, A)^{\mathsf{T}}}{|A - \lambda I|} \tag{1.289}$$

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,

$$(A - \lambda I)\, C(\lambda, A)^{\mathsf{T}} = |A - \lambda I|\, I = P(\lambda, A)\, I. \tag{1.290}$$

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

$$C(\lambda, A)^{\mathsf{T}} = C_0 + C_1\lambda + \ldots + C_{n-1}\lambda^{n-1}. \tag{1.291}$$

Combining these last two equations (1.290 & 1.291) with the characteristic equation (1.287), we have

$$\begin{aligned}(A-\lambda I)C(\lambda,A)^{\mathsf{T}} &= AC_0 + (AC_1 - C_0)\lambda + (AC_2 - C_1)\lambda^2 + \dots \\ &\quad + (AC_{n-1} - C_{n-2})\lambda^{n-1} - C_{n-1}\lambda^n \\ &= \sum_{k=0}^{n} p_k\,\lambda^k.\end{aligned} \tag{1.292}$$

Equating equal powers of λ on both sides of this equation, we find

$$\begin{aligned} AC_0 &= p_0 I \\ AC_1 - C_0 &= p_1 I \\ AC_2 - C_1 &= p_2 I \\ \dots &= \dots \\ AC_{n-1} - C_{n-2} &= p_{n-1} I \\ -C_{n-1} &= p_n I. \end{aligned} \tag{1.293}$$

We now multiply from the left the first of these equations by I, the second by A, the third by A^2, ..., 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

$$0 = \sum_{k=0}^{n} p_k\,A^k = |A|\,I + p_1 A + \ldots + (-1)^{n-1}(\mathrm{Tr}A)\,A^{n-1} + (-1)^n\,A^n \tag{1.294}$$

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

$$A^n = (-1)^{n-1}\big(|A|\,I + p_1 A + p_2 A^2 + \ldots + (-1)^{n-1}(\mathrm{Tr}A)\,A^{n-1}\big). \tag{1.295}$$

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

$$A^2 = -|A|I + (\mathrm{Tr}A)A. \tag{1.296}$$

Example 1.42 (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

$$(\boldsymbol{\theta}\cdot\boldsymbol{\sigma})^2 = -\,|\boldsymbol{\theta}\cdot\boldsymbol{\sigma}|\,I = -\begin{vmatrix} \theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & -\theta_3 \end{vmatrix} I = \theta^2\,I \tag{1.297}$$

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

$$\exp\left(-i\boldsymbol{\theta} \cdot \boldsymbol{\sigma}/2\right) = \cos(\theta/2)\, I - i\hat{\boldsymbol{\theta}} \cdot \boldsymbol{\sigma}\, \sin(\theta/2) \tag{1.298}$$

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

1.28 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 to express (1.295) every power of A in terms of I, A, ..., 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

$$f(x) = \sum_{k=0}^{\infty} c_k\, x^k \tag{1.299}$$

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

$$f(A) = \sum_{k=0}^{n-1} f_k(\boldsymbol{p})\, A^k. \tag{1.300}$$

The identity (1.298) for $\exp\left(-i\boldsymbol{\theta} \cdot \boldsymbol{\sigma}/2\right)$ is an $n = 2$ example of this technique which can become challenging when $n > 3$.

Example 1.43 (The 3×3 Rotation Matrix) In exercise (1.29), one finds the characteristic equation (1.294) 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 (1.30), one uses the 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

$$\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J}) = \cos\theta\, I - i\hat{\boldsymbol{\theta}} \cdot \boldsymbol{J}\, \sin\theta + (1 - \cos\theta)\, \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})^{\mathsf{T}} \tag{1.301}$$

or

$$\exp(-i\boldsymbol{\theta} \cdot \boldsymbol{J})_{ij} = \delta_{ij} \cos\theta - \sin\theta\, \epsilon_{ijk}\hat{\theta}_k + (1 - \cos\theta)\, \hat{\theta}_i\hat{\theta}_j \tag{1.302}$$

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 (1.299) for $f(x)$ converges. For then A is related to its diagonal form $A^{(d)}$ by a similarity transformation (1.278), and we may define $f(A)$ as

$$f(A) = S f(A^{(d)}) S^{-1} \tag{1.303}$$

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

$$f(A^{(d)}) = \begin{pmatrix} f(a_1) & 0 & 0 & \dots \\ 0 & f(a_2) & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & f(a_n) \end{pmatrix} \tag{1.304}$$

and a_1, a_1, $\dots a_n$ are the eigenvalues of the matrix A. This definition makes sense if $f(A)$ is a series in powers of A because then

$$f(A) = \sum_{k=0}^{\infty} c_k A^k = \sum_{k=0}^{\infty} c_k \left(S A^{(d)} S^{-1} \right)^k. \tag{1.305}$$

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

$$f(A) = S \left[\sum_{k=0}^{\infty} c_k \left(A^{(d)} \right)^k \right] S^{-1} = S f(A^{(d)}) S^{-1} \tag{1.306}$$

which is (1.303).

Example 1.44 (Momentum operators generate spatial translations) The position operator x and the momentum operator p obey the commutation relation $[x, p] = xp - px = i\hbar$. Thus the a-derivative $\dot{x}(a)$ of the operator $x(a) = e^{iap/\hbar}\, x\, e^{-iap/\hbar}$ is unity

$$\dot{x}(a) = e^{iap/\hbar}\, (i/\hbar)[p, x]\, e^{-iap/\hbar} = e^{iap/\hbar}\, e^{-iap/\hbar} = 1. \tag{1.307}$$

Since $x(0) = x$, we see that the unitary transformation $U(a) = e^{iap/\hbar}$ moves x to $x + a$

$$e^{iap/\hbar}\, x\, e^{-iap/\hbar} = x(a) = x(0) + \int_0^a \dot{x}(a')\, da' = x + a. \tag{1.308}$$

□

Example 1.45 (Glauber's identity) The commutator of the annihilation operator a and the creation operator $a^\dagger$ for a given mode is the number 1

$$[a, a^\dagger] = a\, a^\dagger - a^\dagger\, a = 1. \tag{1.309}$$

Thus a and $a^\dagger$ commute with their commutator $[a, a^\dagger] = 1$ just as x and p commute with their commutator $[x, p] = i\hbar$.

Suppose that A and B are any two operators that commute with their commutator $[A, B] = AB - BA$

$$[A, [A, B]] = [B, [A, B]] = 0. \tag{1.310}$$

As in the $[x, p]$ example (1.44), we define $A_B(t) = e^{-tB}\, A\, e^{tB}$ and note that because $[B, [A, B]] = 0$, its t-derivative is simply

$$\dot{A}_B(t) = e^{-tB}\, [A, B]\, e^{tB} = [A, B]. \tag{1.311}$$

Since $A_B(0) = A$, an integration gives

$$A_B(t) = A + \int_0^t \dot{A}(t)\, dt = A + \int_0^t [A, B]\, dt = A + t\, [A, B]. \tag{1.312}$$

Multiplication from the left by e^{tB} now gives $e^{tB}\, A_B(t)$ as

$$e^{tB}\, A_B(t) = A\, e^{tB} = e^{tB}\, (A + t\, [A, B])\,. \tag{1.313}$$

Now we define

$$G(t) = e^{tA}\, e^{tB}\, e^{-t(A+B)} \tag{1.314}$$

and use our formula (1.313) to compute its t-derivative as

$$\begin{aligned} \dot{G}(t) &= e^{tA} \left(A\, e^{tB} + e^{tB}\, B - \, e^{tB}(A + B) \right) e^{-t(A+B)} \\ &= e^{tA} \left(e^{tB}\, (A + t\, [A, B]) + e^{tB}\, B - \, e^{tB}(A + B) \right) e^{-t(A+B)} \\ &= e^{tA}\, e^{tB}\, t\, [A, B]\, e^{t(A+B)} = t\, [A, B]\, G(t) = t\, G(t)\, [A, B]. \end{aligned} \tag{1.315}$$

Since $\dot{G}(t)$, $G(t)$, and $[A, B]$ all commute with each other, we can integrate this operator equation

$$\frac{d}{dt} \log G(t) = \frac{\dot{G}(t)}{G(t)} = t\, [A, B] \tag{1.316}$$

from 0 to 1 and get since $G(0) = 1$

$$\log G(1) - \log G(0) = \log G(1) = \frac{1}{2}\, [A, B]. \tag{1.317}$$

Thus $G(1) = e^{[A,B]/2}$ and so

$$e^A\, e^B\, e^{-(A+B)} = e^{\frac{1}{2}[A,B]} \quad \text{or} \quad e^A\, e^B = e^{A+B+\frac{1}{2}[A,B]} \tag{1.318}$$

which is Glauber's identity. □

Example 1.46 (Chemical reactions) The chemical reactions [A] $\xrightarrow{a}$ [B], [B] $\xrightarrow{b}$ [A], and [B] $\xrightarrow{c}$ [C] make the concentrations $[A] \equiv A$, $[B] \equiv B$, and $[C] \equiv C$ of three kinds of molecules vary with time as

$$\dot{A} = -aA + bB, \quad \dot{B} = aA - (b+c)B \quad \text{and} \quad \dot{C} = cB. \tag{1.319}$$

We can group these concentrations into a 3-vector $V = (A, B, C)$ and write the three equations (1.319) as $\dot{V} = K\,V$ in which K is the matrix

$$K = \begin{pmatrix} -a & b & 0 \\ a & -b-c & 0 \\ 0 & c & 0 \end{pmatrix}. \tag{1.320}$$

The solution to the differential equation $\dot{V} = K\,V$ is $V(t) = e^{Kt}\,V(0)$.

The eigenvalues of the matrix K are the roots of the cubic equation $\det(K - \lambda I) = 0$. One root vanishes, and the other two are the roots of the quadratic equation $\lambda^2 + (a+b+c)\lambda + ac = 0$. Their sum is the trace $\mathrm{Tr}K = -(a+b+c)$. They are real when a, b, and c are positive but are complex when $4ac > (a+b+c)^2$. The eigenvectors are complete unless $4ac = (a+b+c)^2$, but are not orthogonal unless $c = 0$.

The time evolution of the concentrations $[A]$ (dashdot), $[B]$ (solid), and $[C]$ (dashes) are plotted in Fig. 1.1 for the initial conditions $[A] = 1$ and $[B] = [C] = 0$ and rates $a = 0.15$, $b = 0.1$, and $c = 0.1$. The Matlab code is in the repository Linear_algebra at `github.com/kevinecahill`.

□

Example 1.47 (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 $\hbar = h/(2\pi) = 1.054\times10^{-34}$ Js where h is Planck's 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

$$H = SH^{(d)}S^{-1}. \tag{1.321}$$

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

$$U(t) = S\,\exp(-iH^{(d)}t/\hbar)\,S^{-1}. \tag{1.322}$$

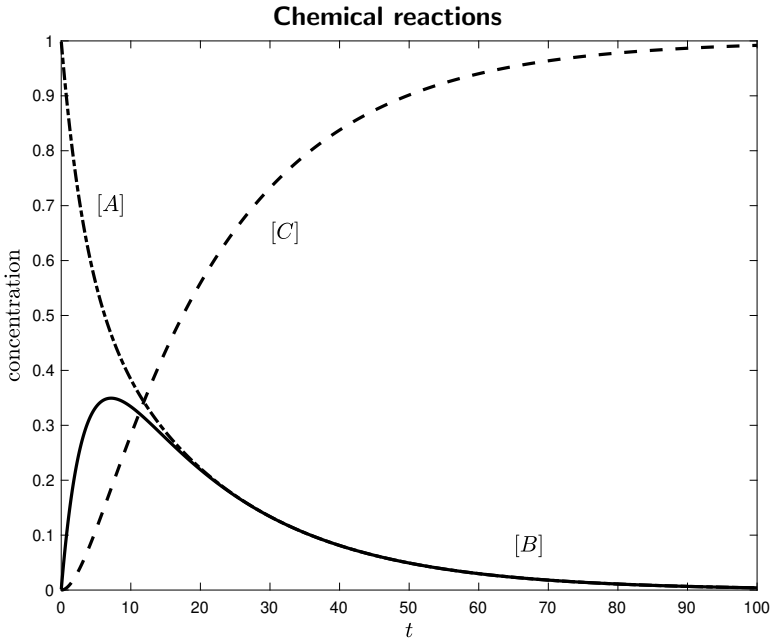

Figure 1.1 The concentrations $[A]$ (dashdot), $[B]$ (solid), and $[C]$ (dashes) as given by the matrix equation $V(t) = e^{Kt}V(0)$ for the initial conditions $[A] = 1$ and $[B] = [C] = 0$ and rates $a = 0.15$, $b = 0.1$, and $c = 0.1$.

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

$$U(t) = S \begin{pmatrix} e^{-i\omega_1 t} & 0 & 0 \\ 0 & e^{-i\omega_2 t} & \\ 0 & 0 & e^{-i\omega_3 t} \end{pmatrix} S^{-1}. \tag{1.323}$$

□

Example 1.48 (Entropy) The **entropy** S of a system described by a density operator ρ is the trace $S = -k\,\mathrm{Tr}\,(\rho \ln \rho)$ 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 1.29), and so it can be diagonalized by a similarity transformation $\rho = S\,\rho^{(d)}\,S^{-1}$. By (1.24), $\mathrm{Tr}ABC = \mathrm{Tr}BCA$, so we can write S as

$$S = -k\mathrm{Tr}\left(S\,\rho^{(d)}\,S^{-1}\,S\,\ln(\rho^{(d)})\,S^{-1}\right) = -k\mathrm{Tr}\left(\rho^{(d)}\,\ln(\rho^{(d)})\right). \tag{1.324}$$

A vanishing eigenvalue $\rho_k^{(d)} = 0$ 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 $\rho^{(d)}$, then the sum

$$\begin{aligned} S &= -k\left(\rho_1 \ln \rho_1 + \rho_2 \ln \rho_2 + \rho_3 \ln \rho_3\right) \\ &= k\left[\rho_1 \ln\left(1/\rho_1\right) + \rho_2 \ln\left(1/\rho_2\right) + \rho_3 \ln\left(1/\rho_3\right)\right] \end{aligned} \tag{1.325}$$

is its entropy. □

Example 1.49 (Logarithm of a determinant) Since every nondefective $n \times n$ matrix A may be diagonalized by a similarity transformation, its determinant is the product of its eigenvalues and its trace is the sum of them, and so the logarithm of its determinant is the trace of its logarithm

$$\ln \det A = \ln(\lambda_1 \dots \lambda_n) = \ln(\lambda_1) + \cdots + \ln(\lambda_n) = \mathrm{Tr}(\ln A). \tag{1.326}$$

When none of A's eigenvalues vanishes, this relation implies the earlier result (1.228) that the variation of A's determinant is

$$\delta \det A = \det A \ \mathrm{Tr}(A^{-1}\delta A). \tag{1.327}$$

□

1.29 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 1.26 ensure that an $n \times n$ hermitian matrix A has n eigenvectors $|k\rangle$ with eigenvalues a_k

$$A|k\rangle = a_k|k\rangle. \tag{1.328}$$

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

$$\langle k|A^\dagger = a_k^* \langle k| \tag{1.329}$$

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

$$\langle k|A^\dagger = \langle k|A = a_k^* \langle k|. \tag{1.330}$$

We now form the inner product of both sides of this equation with the ket $|k\rangle$ and use the eigenvalue equation (1.328) to get

$$\langle k|A|k\rangle = a_k \langle k|k\rangle = a_k^* \langle k|k\rangle \tag{1.331}$$

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

$$a_k^* = a_k. \tag{1.332}$$

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

$$a_m^* \langle m|k\rangle = (a_m \langle k|m\rangle)^* = \langle k|A|m\rangle^* = \langle m|A^\dagger|k\rangle = \langle m|A|k\rangle = a_k \langle m|k\rangle \tag{1.333}$$

which implies that

$$(a_m^* - a_k)\, \langle m|k\rangle = 0. \tag{1.334}$$

But by (1.332), the eigenvalues a_m are real, and so we have

$$(a_m - a_k)\, \langle m|k\rangle = 0 \tag{1.335}$$

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 $|k_\alpha\rangle$ of a hermitian matrix have the same eigenvalue a_k, 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_k

$$A \left(\sum_{\alpha \in D} c_\alpha |k_\alpha\rangle \right) = a_k \left(\sum_{\alpha \in D} c_\alpha |k_\alpha\rangle \right) \tag{1.336}$$

where D is the set of labels α of the eigenvectors with the same eigenvalue. If the degenerate eigenvectors $|k_\alpha\rangle$ are linearly independent, then we may use the Gram-Schmidt procedure (1.113–1.123) 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 $|k_\alpha\rangle$ 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 that unitary transformations preserve the orthonormality of a basis (section 1.16). Any unitary transformation that commutes with the matrix A

$$[A, U] = 0 \tag{1.337}$$

represents a symmetry of A and maps each set of orthonormal degenerate eigenvectors of A into another set of orthonormal degenerate eigenvectors of A with the same eigenvalue because

$$AU|k_\alpha\rangle = UA|k_\alpha\rangle = a_k\, U|k_\alpha\rangle. \tag{1.338}$$

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

$$A(\epsilon) = A + \epsilon B. \tag{1.339}$$

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 $a_k(\epsilon)$ and n orthonormal nondegenerate eigenvectors

$$A(\epsilon)|k_\beta, \epsilon\rangle = a_{k_\beta}(\epsilon)|k_\beta, \epsilon\rangle. \tag{1.340}$$

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

$$\langle k_\beta, \epsilon|k_{\beta'}, \epsilon\rangle = \delta_{\beta,\beta'} \tag{1.341}$$

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

$$|k_\beta\rangle = \lim_{\epsilon\to 0} |k_\beta, \epsilon\rangle. \tag{1.342}$$

We may choose them as the orthogonal degenerate eigenvectors of A. Since one can always 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 show by a different argument that an $n \times n$ hermitian matrix has n orthogonal eigenvectors. To do this, we first note that the space $S_{\perp,k}$ of vectors $|y\rangle$ orthogonal to an eigenvector $|k\rangle$ of a hermitian operator A

$$A|k\rangle = a_k|k\rangle \tag{1.343}$$

is **invariant** under the action of A, that is, $\langle k|y\rangle = 0$ implies

$$\langle k|A|y\rangle = a_k\langle k|y\rangle = 0. \tag{1.344}$$

Thus if the vector $|y\rangle$ is in the space $S_{\perp,k}$ of vectors orthogonal to an eigenvector $|k\rangle$ of a hermitian operator A, then the vector $A|y\rangle$ also is in the space $S_{\perp,k}$. This space is invariant under the action of A.

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 $S_{\perp,1}$ 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 $S_{\perp,2}$ 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 **hermitian matrices are nondefective**.

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

$$I = \sum_{k=1}^{n} |k\rangle\langle k|. \tag{1.345}$$

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

$$A = AI = A\sum_{k=1}^{n} |k\rangle\langle k| = \sum_{k=1}^{n} a_k |k\rangle\langle k| \tag{1.346}$$

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 made.

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

$$A_{kj} = \langle k|A|j\rangle = a_k \delta_{kj}. \tag{1.347}$$

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

$$A_{k\ell} = \langle \alpha_k|A|\alpha_\ell\rangle = \sum_{n=1}^{n} \langle \alpha_k|n\rangle a_n \langle n|\alpha_\ell\rangle. \tag{1.348}$$

The unitary matrix $U_{kn} = \langle \alpha_k|n\rangle$ relates the matrix $A_{k\ell}$ in an arbitrary basis to its diagonal form $A = UA^{(d)}U^\dagger$ in which $A^{(d)}$ is the diagonal matrix $A^{(d)}_{nm} = a_n\, \delta_{nm}$. 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**

$$R = O\, R^{(d)} O^{\mathsf{T}} \tag{1.349}$$

in which the matrix O is a real unitary matrix, that is, an orthogonal matrix (1.184).

Example 1.50 (The Seesaw Mechanism) Suppose we wish to find the eigenvalues of the real, symmetric mass matrix

$$\mathcal{M} = \begin{pmatrix} 0 & m \\ m & M \end{pmatrix} \tag{1.350}$$

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 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. Minkowski, Yanagida, and Gell-Mann, Ramond, and Slansky 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 suggesting new physics at that scale. But if we set $mc^2 = 0.28$ MeV and use $m_\nu = 0.45$ eV as an average neutrino mass, then the big mass is only $Mc^2 = 173$ GeV, the mass of the top. Also, the small masses of the neutrinos may be related to the weakness of their interactions. □

If we return to the orthogonal transformation (1.349) and multiply column ℓ of the matrix O and row ℓ of the matrix O^{T} by $\sqrt{|R^{(d)}_\ell|}$, then we arrive at the **congruency transformation** of Sylvester's theorem

$$R = C\, \hat{R}^{(d)} C^{\mathsf{T}} \tag{1.351}$$

in which the diagonal entries $\hat{R}^{(d)}_\ell$ are either ± 1 or 0 because the matrices $C_{k\ell} = \sqrt{|R^{(d)}_\ell|}\, O_{k\ell}$ and C^{T} have absorbed the factors $|R^{(d)}_\ell|$.

Example 1.51 (Principle of equivalence) If G is a real, symmetric 4×4

matrix then there's a real 4×4 matrix $D = C^{\mathsf{T}-1}$ such that

$$G_d = D^{\mathsf{T}} G D = \begin{pmatrix} g_1 & 0 & 0 & 0 \\ 0 & g_2 & 0 & 0 \\ 0 & 0 & g_3 & 0 \\ 0 & 0 & 0 & g_4 \end{pmatrix} \tag{1.352}$$

in which the diagonal entries g_i are ± 1 or 0. Thus there's a real 4×4 matrix D that casts any real symmetric metric g_{ik} of spacetime with three positive and one negative eigenvalues into the diagonal metric $\eta_{j\ell}$ of flat spacetime by the congruence

$$g_d = D^{\mathsf{T}} g D = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \eta \tag{1.353}$$

at any given point x of spacetime. Usually one needs different D's at different points. The principle of equivalence (section 13.25) says that in the new free-fall coordinates, all physical laws take the same form as in special relativity without acceleration or gravitation in a suitably small region of spacetime about the point x. □

1.30 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

$$[V, V^\dagger] = V V^\dagger - V^\dagger V = 0. \tag{1.354}$$

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

$$[U, U^\dagger] = U U^\dagger - U^\dagger U = I - I = 0. \tag{1.355}$$

A matrix $V = U^\dagger V^{(d)} U$ that can be diagonalized by a unitary transformation U commutes with its adjoint $V^\dagger = U^\dagger V^{(d)*} U$ and so is normal because the commutator of any two diagonal matrices vanishes

$$[V, V^\dagger] = [U^\dagger V^{(d)} U, U^\dagger V^{(d)*} U] = U^\dagger [V, V^{(d)*}] U = 0. \tag{1.356}$$

To see why a normal matrix can be diagonalized by a unitary transformation, we consider an $n \times n$ normal matrix V which since it is square has

n eigenvectors $|k\rangle$ with eigenvalues v_k

$$(V - v_k I)\,|k\rangle = 0 \tag{1.357}$$

(section 1.26). The square of the norm (1.85) of this vector must vanish

$$\|\,(V - v_k I)\,|k\rangle\,\|^2 = \langle k|\,(V - v_k I)^\dagger\,(V - v_k I)\,|k\rangle = 0. \tag{1.358}$$

But since V is normal, we also have

$$\langle k|\,(V - v_k I)^\dagger\,(V - v_k I)\,|k\rangle = \langle k|\,(V - v_k I)\,(V - v_k I)^\dagger\,|k\rangle. \tag{1.359}$$

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

$$V^\dagger|k\rangle = v_k^*|k\rangle \quad \text{and so} \quad \langle k|V = v_k\langle k|. \tag{1.360}$$

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

$$V|m\rangle = v_m|m\rangle \tag{1.361}$$

then

$$\langle k|V|m\rangle = v_m\langle k|m\rangle \tag{1.362}$$

and also by (1.360)

$$\langle k|V|m\rangle = v_k\langle k|m\rangle. \tag{1.363}$$

Subtracting (1.362) from (1.363), we get

$$(v_k - v_m)\,\langle k|m\rangle = 0 \tag{1.364}$$

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

To see that a normal $n \times n$ matrix V has n orthogonal eigenvectors, we first note that if $|y\rangle$ is any vector that is orthogonal to any eigenvector $|k\rangle$ of the matrix V, that is both $\langle k|y\rangle = 0$ and $V|k\rangle = v_k|k\rangle$, then the property (1.360) implies that

$$\langle k|V|y\rangle = v_k\langle k|y\rangle = 0. \tag{1.365}$$

Thus the space of vectors orthogonal to an eigenvector of a normal matrix V is invariant under the action of V. The argument following the analogous equation (1.344) applies also to normal matrices and shows that **every $n \times n$ normal matrix has n orthonormal eigenvectors**. It follows then

from the argument of equations (1.345–1.348) that every $n \times n$ normal matrix V can be diagonalized by an $n \times n$ unitary matrix U

$$V = U V^{(d)} U^\dagger \tag{1.366}$$

whose kth column $U_{\ell k} = \langle \alpha_\ell | k \rangle$ is the eigenvector $|k\rangle$ in the arbitrary basis $|\alpha_\ell\rangle$ of the matrix $V_{m\ell} = \langle \alpha_m | V | \alpha_\ell \rangle$ as in (1.348).

Since the eigenstates $|k\rangle$ of a normal matrix V

$$V|k\rangle = v_k |k\rangle \tag{1.367}$$

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

$$I = \sum_{k=1}^{n} |k\rangle\langle k|. \tag{1.368}$$

The product VI is V itself, so

$$V = V\, I = V \sum_{k=1}^{n} |k\rangle\langle k| = \sum_{k=1}^{n} v_k\, |k\rangle\langle k|. \tag{1.369}$$

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

$$f(V) = \sum_{k=1}^{n} f(v_k)\, |k\rangle\langle k| \tag{1.370}$$

which is simpler than the corresponding formula (1.303) for an arbitrary nondefective matrix. This is a good way to think about functions of normal matrices.

Example 1.52 (Time-evolution operator) 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 $|k\rangle$ have energy E_k

$$H|k\rangle = E_k |k\rangle. \tag{1.371}$$

So we apply (1.370) with $V \to H$ and get

$$e^{-iHt/\hbar} = \sum_{k=1}^{n} e^{-iE_k t/\hbar}\, |k\rangle\langle k| \tag{1.372}$$

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

$$e^{-iHt/\hbar} |\psi\rangle = \sum_{k=1}^{n} e^{-iE_k t/\hbar}\, |k\rangle\langle k|\psi\rangle \tag{1.373}$$

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

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

$$|V| = \exp\left[\mathrm{Tr}(\ln V)\right], \quad \ln|V| = \mathrm{Tr}(\ln V), \quad \text{and} \quad \delta \ln|V| = \mathrm{Tr}\left(V^{-1}\delta V\right). \tag{1.374}$$

1.31 Compatible normal matrices

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

$$[A, B] \equiv AB - BA = 0 \tag{1.375}$$

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

$$AB|u\rangle = BA|u\rangle = Bz|u\rangle = z\,B|u\rangle. \tag{1.376}$$

We have seen that any normal matrix A can be written as a sum (1.30) of outer products

$$A = \sum_{k=1}^{n} |a_k\rangle a_k \langle a_k| \tag{1.377}$$

of its orthonormal eigenvectors $|a_k\rangle$ which are complete in the n-dimensional vector space $\mathcal{S}$ on which A acts. Suppose now that the eigenvalues a_k 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) $|a_k\rangle$ of the matrix A, the matrix B must satisfy

$$0 = \langle a_\ell|AB - BA|a_k\rangle = (a_\ell - a_k)\,\langle a_\ell|B|a_k\rangle \tag{1.378}$$

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

$$B = IBI = \sum_{\ell=1}^{n} |a_\ell\rangle\langle a_\ell|B\sum_{k=1}^{n} |a_k\rangle\langle a_k| = \sum_{\ell=1}^{n} |a_\ell\rangle\langle a_\ell|B|a_\ell\rangle\langle a_\ell| \tag{1.379}$$

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

$$B|a_k\rangle = \sum_{\ell=1}^{n} |a_\ell\rangle\langle a_\ell|B|a_\ell\rangle\langle a_\ell|a_k\rangle = \sum_{\ell=1}^{n} |a_\ell\rangle\langle a_\ell|B|a_\ell\rangle\delta_{\ell k} = \langle a_k|B|a_k\rangle|a_k\rangle \tag{1.380}$$

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

$$A|a_\ell, b_k\rangle = a_\ell|a_\ell, b_k\rangle \quad \text{and} \quad B|a_\ell, b_k\rangle = b_k|a_\ell, b_k\rangle. \tag{1.381}$$

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 (1.381), then they commute

$$AB|a_\ell, b_k\rangle = Ab_k|a_\ell, b_k\rangle = a_\ell b_k|a_\ell, b_k\rangle = a_\ell B|a_\ell, b_k\rangle = BA|a_\ell, b_k\rangle \tag{1.382}$$

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 $\{a_j, b_k, c_\ell, \ldots\}$ there is only a single eigenvector $|a_j, b_k, c_\ell, \ldots\rangle$.

Example 1.53 (Compatible Photon Observables) For example, 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 1.54 (Thermal density operator) A **density operator** ρ is the most general description of a quantum-mechanical system. It is hermitian,

positive definite, and of unit trace. Since it is hermitian, it can be diagonalized (section 1.29)

$$\rho = \sum_n |n\rangle\langle n|\rho|n\rangle\langle n| \tag{1.383}$$

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

$$\sum_n \rho_n = 1. \tag{1.384}$$

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 (1.375), 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 $c_1 = \sum_n \rho_n - 1 = 0$ and $c_2 = \mathrm{Tr}(\rho H) - E = 0$. We introduce two Lagrange multipliers (section 1.24) and maximize the unconstrained function

$$\begin{aligned} L(\rho, \lambda_1, \lambda_2) &= S - \lambda_1 c_1 - \lambda_2 c_2 \\ &= -k \sum_n \rho_n \ln \rho_n - \lambda_1 \left(\sum_n \rho_n - 1 \right) - \lambda_2 \left(\sum_n \rho_n E_n - E \right) \end{aligned} \tag{1.385}$$

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

$$\frac{\partial L}{\partial \rho_n} = -k\,(\ln \rho_n + 1) - \lambda_1 - \lambda_2 E_n = 0 \tag{1.386}$$

$$\frac{\partial L}{\partial \lambda_1} = \sum_n \rho_n - 1 = 0 \tag{1.387}$$

$$\frac{\partial L}{\partial \lambda_2} = \sum_n \rho_n E_n - E = 0. \tag{1.388}$$

The first (1.386) of these conditions implies that

$$\rho_n = \exp\left[-(\lambda_1 + \lambda_2 E_n + k)/k\right]. \tag{1.389}$$

We satisfy the second condition (1.387) by choosing λ_1 so that

$$\rho_n = \frac{\exp(-\lambda_2 E_n/k)}{\sum_n \exp(-\lambda_2 E_n/k)}. \tag{1.390}$$

Setting $\lambda_2 = 1/T$, we define the temperature T so that ρ satisfies the third condition (1.388). Its eigenvalue ρ_n then is

$$\rho_n = \frac{\exp(-E_n/kT)}{\sum_n \exp(-E_n/kT)}. \tag{1.391}$$

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

$$\rho = \frac{e^{-\beta H}}{\mathrm{Tr}\,(e^{-\beta H})} \tag{1.392}$$

which is the **Boltzmann distribution**, also called the **canonical ensemble**. □

Example 1.55 (Grand canonical ensemble) Lagrange's function for the density operator of a system of maximum entropy $S = -k\mathrm{Tr}(\rho \ln \rho)$ given a fixed mean energy $E = \mathrm{Tr}(\rho H)$ and a fixed mean number of particles $\langle N\rangle = \mathrm{Tr}(\rho N)$, in which N is the number operator $N|n\rangle = N_n|n\rangle$, is

$$\begin{aligned} L(\rho, \lambda_1, \lambda_2, \lambda_3) = & -k\sum_n \rho_n \ln \rho_n - \lambda_1\left(\sum_n \rho_n - 1\right) \\ & - \lambda_2\left(\sum_n \rho_n E_n - E\right) - \lambda_3\left(\sum_n \rho_n N_n - \langle N\rangle\right). \end{aligned} \tag{1.393}$$

Setting the partial derivative of L with respect to ρ_n

$$\frac{\partial L}{\partial \rho_n} = -k\,(\ln \rho_n + 1) - \lambda_1 - \lambda_2 E_n - \lambda_3 N_n = 0 \tag{1.394}$$

as well as the partial derivatives of L with respect to the three Lagrange multipliers λ_i equal to zero, we get

$$\rho = \frac{e^{-\beta(H-\mu N)}}{\mathrm{Tr}(e^{-\beta(H-\mu N)})} \tag{1.395}$$

in which μ is the **chemical potential**. □

1.32 Singular-value decompositions

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 A's nonnegative singular values S_k, and an $n \times n$ unitary matrix $V^\dagger$

$$A = U\,\Sigma\,V^\dagger \quad \text{or} \quad A_{ik} = \sum_{\ell=1}^{\min(m,n)} U_{i\ell}\,S_\ell\,V^\dagger_{\ell k}. \tag{1.396}$$

This singular-value decomposition 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 $\{|a_j\rangle \in V_n\}$ and $\{|b_k\rangle \in V_m\}$ labeled by 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

$$I_n = \sum_{j=1}^{n} |a_j\rangle\langle a_j| \quad \text{and} \quad I_m = \sum_{k=1}^{m} |b_k\rangle\langle b_k| \tag{1.397}$$

and lets us write linear operator A as

$$A = I_m A I_n = \sum_{k=1}^{m}\sum_{j=1}^{n} |b_k\rangle\langle b_k|A|a_j\rangle\langle a_j| \tag{1.398}$$

in which the $\langle b_k|A|a_j\rangle$ are the elements of a complex $m \times n$ matrix.

The singular-value decomposition of the linear operator A is a choice of two special bases $\{|a_j\rangle\}$ and $\{|b_j\rangle\}$ that make $\langle b_k|A|a_j\rangle = S_j\delta_{kj}$ and so express A as

$$A = \sum_j |b_j\rangle S_j \langle a_j| \tag{1.399}$$

in which the sum is over the nonzero singular values S_j, which will turn out to be positive.

The kets of the special basis $\{|a_j\rangle\}$ are the eigenstates of the hermitian operator $A^\dagger A$

$$A^\dagger A|a_j\rangle = e_j|a_j\rangle. \tag{1.400}$$

These states $\{|a_j\rangle\}$ are orthogonal because $A^\dagger A$ is hermitian, and we may choose them to be normalized. The eigenvalue e_j is the squared length of the ket $A|a_j\rangle$ and so is positive or zero

$$\langle a_j|A^\dagger A|a_j\rangle = e_j\,\langle a_j|a_j\rangle = e_j \geq 0. \tag{1.401}$$

The singular values are the square roots of these eigenvalues

$$S_j = \sqrt{e_j} = \sqrt{\langle a_j | A^\dagger A | a_j \rangle}. \tag{1.402}$$

For $S_j > 0$, the special ket $|b_j\rangle$ is the suitably normalized image $A|a_j\rangle$ of the special ket $|a_j\rangle$

$$|b_j\rangle = \frac{A|a_j\rangle}{S_j}; \tag{1.403}$$

for $S_j = 0$, the ket $|b_j\rangle$ vanishes. The nonzero special kets $|b_j\rangle$ are orthonormal

$$\langle b_k | b_j \rangle = \frac{1}{S_k S_j} \langle a_k | A^\dagger A | a_j \rangle = \frac{e_j}{S_k S_j} \langle a_k | a_j \rangle = \frac{e_j}{S_k S_j} \delta_{kj} = \delta_{kj}. \tag{1.404}$$

The number of positive singular values, $S_j > 0$, is at most n. It also is at most m because each nonzero ket $|b_j\rangle$ is an orthonormal vector in the space V_m which has only m dimensions. So the number of positive singular values, $S_j > 0$, is at most $\min(m, n)$, the smaller of m and n.

The singular-value decomposition of the linear operator A then is the sum

$$A = A I_n = A \sum_{j=1}^{n} |a_j\rangle\langle a_j| = \sum_{j=1}^{n} A\,|a_j\rangle\langle a_j| = \sum_{j=1}^{n} |b_j\rangle\, S_j\, \langle a_j| \tag{1.405}$$

in which at most $\min(m, n)$ of the singular values are positive.

In terms of any two bases, $|k\rangle$ for $k = 1, \ldots, m$ for the space V_m and $|\ell\rangle$ for $\ell = 1, \ldots, n$ for the space V_n, and their identity operators

$$I_m = \sum_{k=1}^{m} |k\rangle\langle k| \quad \text{and} \quad I_n = \sum_{\ell=1}^{n} |\ell\rangle\langle \ell| \tag{1.406}$$

the singular-value decomposition of the linear operator A is

$$\begin{aligned} A &= \sum_{k=1}^{m} |k\rangle\langle k| A \sum_{\ell=1}^{n} |\ell\rangle\langle\ell| = \sum_{k=1}^{m}\sum_{j=1}^{n}\sum_{\ell=1}^{n} |k\rangle\langle k|b_j\rangle\, S_j\, \langle a_j|\ell\rangle\langle\ell| \\ &= \sum_{k=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{\ell=1}^{n} |k\rangle\langle k|b_i\rangle\, S_i\, \delta_{ij} \langle a_j|\ell\rangle\langle\ell| = U\,\Sigma\,V^\dagger. \end{aligned} \tag{1.407}$$

In this expansion, the k, i matrix element of the $m \times m$ unitary matrix U is $U_{ki} = \langle k|b_i\rangle$, the i, j element of the $m \times n$ matrix Σ is $\Sigma_{ij} = S_j\,\delta_{ij}$, and the j, ℓ matrix element of the $n \times n$ unitary matrix $V^\dagger$ is $V^\dagger_{j\ell} = \langle a_j|\ell\rangle$. Thus $V^*_{\ell j} = \langle a_j|\ell\rangle$, and so $V_{\ell j} = \langle a_j|\ell\rangle^* = \langle \ell|a_j\rangle$.

The vectors $|b_j\rangle$ and $|a_j\rangle$ respectively are the left and right singular vectors. Incidentally, the singular-value decomposition (1.405) shows that the left singular vectors $|b_j\rangle$ are the eigenvectors of $A\,A^\dagger$

$$\begin{aligned} A\,A^\dagger &= \sum_{j=1}^{n} |b_j\rangle\, S_j\, \langle a_j| \sum_{k=1}^{n} |a_k\rangle\, S_k\, \langle b_k| = \sum_{j,k=1}^{n} |b_j\rangle\, S_j\, \langle a_j|a_k\rangle\, S_k\, \langle b_k| \\ &= \sum_{j,k=1}^{n} |b_j\rangle\, S_j\, \delta_{jk}\, S_k\, \langle b_k| = \sum_{j=1}^{n} |b_j\rangle\, S_j^2\, \langle b_j| \end{aligned} \tag{1.408}$$

just as (1.400) the right singular vectors $|a_j\rangle$ are the eigenvectors of $A^\dagger A$.

The kets $|a_j\rangle$ whose singular values vanish, $S_j = 0$, span the **null space** or **kernel** of the linear operator A.

Example 1.56 (Singular-Value Decomposition of a 2×3 Matrix) If A is

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.409}$$

then the positive hermitian matrix $A^\dagger A$ is

$$A^\dagger A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \tag{1.410}$$

The normalized eigenvectors and eigenvalues of $A^\dagger A$ are

$$|a_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad |a_2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |a_3\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \tag{1.411}$$

and their eigenvalues are $e_1 = 2, e_2 = 1$, and $e_3 = 0$. The third eigenvalue e_3 had to vanish because A is a 3×2 matrix.

The vector $A|a_1\rangle$ (as a row vector) is $(0, \sqrt{2})$, and its norm is $\sqrt{2}$, so the normalized vector is $|b_1\rangle = (0, 1)$. Similarly, the vector $|b_2\rangle$ is $A|a_2\rangle = (1, 0)$. The SVD of A then is

$$A = \sum_{n=1}^{2} |b_j\rangle S_j \langle a_j| = U \Sigma V^\dagger \tag{1.412}$$

where $S_n = \sqrt{e_n}$. The unitary matrices are $U_{k,n} = \langle k|b_n\rangle$ and $V_{\ell,j} = \langle \ell|a_j\rangle$ are

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.413}$$

and the diagonal matrix Σ is

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{1.414}$$

So finally the SVD of $A = U\Sigma V^\dagger$ is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}. \tag{1.415}$$

The null space or kernel of A is the set of vectors that are real multiples $c|a_3\rangle$ of the eigenvector $|a_3\rangle$ which has a zero eigenvalue, $e_3 = 0$. It is the third column of the matrix V displayed in (1.413). □

Example 1.57 (Matlab's singular-value decomposition) Matlab's command [U,S,V] = svd(X) performs the singular-value decomposition (SVD) 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

$$A\,|x\rangle = |y\rangle \tag{1.416}$$

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 (1.405), we have

$$\sum_{j=1}^{\min(m,n)} |b_j\rangle S_j \langle a_j|x\rangle = |y\rangle. \tag{1.417}$$

The orthonormality (1.404) of the vectors $|b_j\rangle$ then tells us that

$$S_j \langle a_j|x\rangle = \langle b_j|y\rangle. \tag{1.418}$$

If the singular value is positive, $S_j > 0$, then we may divide by it to get $\langle a_j|x\rangle = \langle b_j|y\rangle / S_j$ and so find the solution

$$|x\rangle = \sum_{j=1}^{\min(m,n)} \frac{\langle b_j|y\rangle}{S_j} |a_j\rangle. \tag{1.419}$$

But this solution is not always available or unique.

For instance, if for some ℓ the inner product $\langle b_\ell|y\rangle \neq 0$ while the singular value $S_\ell = 0$, then there is no solution to equation (1.416). This problem occurs when $m > n$ because there are at most $n < m$ nonzero singular values.

Example 1.58 Suppose A is the 3×2 matrix

$$A = \begin{pmatrix} r_1 & p_1 \\ r_2 & p_2 \\ r_3 & p_3 \end{pmatrix} \tag{1.420}$$

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 (1.416) sometimes has no solutions. For instance, if A is a square defective matrix (section 1.26), then $A|x\rangle = |y\rangle$ will fail to have a solution when the vector $|y\rangle$ lies outside the space spanned by the linearly dependent eigenvectors of the matrix A.

And when $n > m$, as in for instance

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{1.421}$$

the solution (1.419) is never unique, for we may add to it any linear combi-

nation of the vectors $|a_j\rangle$ that have zero as their singular values

$$|x\rangle = \sum_{j=1}^{\min(m,n)} \frac{\langle b_j|y\rangle}{S_n} |a_j\rangle + \sum_{j,S_j=0} x_j|a_j\rangle \tag{1.422}$$

of which there are at least $n - m$.

Example 1.59 (CKM matrix) In the standard model, the mass matrices of the u, c, t and d, s, b quarks are 3×3 complex matrices M_u and M_d with singular-value decompositions $M_u = U_u \Sigma_u V_u^\dagger$ and $M_d = U_d \Sigma_d V_d^\dagger$ whose singular-values are the quark masses. The unitary CKM matrix $U_u^\dagger U_d$ (Cabibbo, Kobayashi, Maskawa) describes transitions among the quarks mediated by the $W^\pm$ gauge bosons. By redefining the quark fields, one may make the CKM matrix real, apart from a phase that violates charge-conjugation-parity (CP) symmetry. □

The adjoint of a complex symmetric matrix M is its complex conjugate, $M^\dagger = M^*$. So by (1.400), its right singular vectors $|n\rangle$ are the eigenstates of M^*M

$$M^*M|n\rangle = S_n^2|n\rangle \tag{1.423}$$

and by (1.408) its left singular vectors $|m_n\rangle$ are the eigenstates of MM^*

$$MM^*|m_n\rangle = (M^*M)^* |m_n\rangle = S_n^2|m_n\rangle. \tag{1.424}$$

Thus its left singular vectors are the complex conjugates of its right singular vectors, $|m_n\rangle = |n\rangle^*$. So the unitary matrix V is the complex conjugate of the unitary matrix U, and the SVD of M is (Autonne, 1915)

$$M = U \Sigma U^\mathsf{T}. \tag{1.425}$$

1.33 Moore-Penrose pseudoinverses

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

$$A = U \, \Sigma \, V^\dagger \tag{1.426}$$

then the Moore-Penrose pseudoinverse (Eliakim H. Moore 1862–1932, Roger Penrose 1931–) is

$$A^+ = V \, \Sigma^+ \, U^\dagger \tag{1.427}$$

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

$$\begin{aligned} A\,A^+\,A &= A, \quad A^+A\,A^+ = A^+, \\ \left(A\,A^+\right)^\dagger &= A\,A^+, \quad \text{and} \quad \left(A^+A\right)^\dagger = A^+\,A \end{aligned} \tag{1.428}$$

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 A^+A is the $n \times n$ identity matrix I_n

$$A^+A = V^\dagger \Sigma^+ \Sigma V = V^\dagger I_n V = I_n \tag{1.429}$$

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

$$AA^+ = U \Sigma \Sigma^+ U^\dagger = U I_m U^\dagger = I_m \tag{1.430}$$

and 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

$$A^{-1} = A^+. \tag{1.431}$$

If the columns of A are linearly independent, then the matrix $A^\dagger A$ has an inverse, and the pseudoinverse is

$$A^+ = \left(A^\dagger A\right)^{-1} A^\dagger. \tag{1.432}$$

The solution (1.248) 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

$$A^+ = A^\dagger \left(AA^\dagger\right)^{-1}. \tag{1.433}$$

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

$$A^{-1} = A^+. \tag{1.434}$$

Example 1.60 (The pseudoinverse of a 2×3 matrix) The pseudoinverse A^+ of the matrix A

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.435}$$

with singular-value decomposition (1.415) is

$$\begin{aligned} A^+ &= V\,\Sigma^+\,U^\dagger \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} \end{aligned} \tag{1.436}$$

which satisfies the four conditions (1.428). The product $A\,A^+$ gives the 2×2 identity matrix

$$A\,A^+ = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.437}$$

which is an instance of (1.430). Moreover, the rows of A are linearly independent, and so the simple rule (1.433) works:

$$\begin{aligned} A^+ &= A^\dagger \left(AA^\dagger\right)^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \left(\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \right)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} \end{aligned} \tag{1.438}$$

which is (1.436).

The columns of the matrix A are not linearly independent, however, and so the simple rule (1.432) fails. Thus the product A^+A

$$A^+A = \begin{pmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{1.439}$$

is not the 3×3 identity matrix which it would be if (1.432) held. □

1.34 Tensor products and entanglement

Tensor products are used to describe composite systems, such as the spins of two electrons. The terms **direct product** and tensor product sometimes are used interchangeably.

If A is an $n \times n$ matrix with elements A_{ij} and B is an $m \times m$ matrix with elements $B_{k\ell}$, then their **tensor product** $C = A \otimes B$ is an $nm \times nm$ matrix with elements $C_{ik,j\ell} = A_{ij}\, B_{k\ell}$. This tensor-product matrix $A \otimes B$ maps a vector $V_{j\ell}$ into the vector

$$W_{ik} = \sum_{j=1}^{n} \sum_{\ell=1}^{m} C_{ik,j\ell}\, V_{j\ell} = \sum_{j=1}^{n} \sum_{\ell=1}^{m} A_{ij}\, B_{k\ell}\, V_{j\ell} \tag{1.440}$$

in which the second double index $j\ell$ of C and the second indices of A and B match the double index $j\ell$ of the vector V.

A tensor-product operator is a product of two operators that act on two different vector spaces. Suppose that an operator A acts on a space S spanned by n kets $|i\rangle$, and that an operator B acts on a space T spanned by m kets $|k\rangle$, and that both operators map vectors into their spaces S and T. Then we may write A as

$$A = I_S A I_S = \sum_{i,j=1}^{n} |i\rangle\langle i|A|j\rangle\langle j| \tag{1.441}$$

and B as

$$B = I_T B I_T = \sum_{k,s=1}^{m} |k\rangle\langle k|B|\ell\rangle\langle \ell|. \tag{1.442}$$

Their tensor product $C = A \otimes B$ is

$$C = A \otimes B = \sum_{i,j=1}^{n} \sum_{k,\ell=1}^{m} |i\rangle \otimes |k\rangle\, \langle i|A|j\rangle\langle k|B|\ell\rangle\, \langle j| \otimes \langle \ell| \tag{1.443}$$

and it acts on the tensor-product vector space $S \otimes T$ which is spanned by the tensor-product kets $|i, k\rangle = |i\rangle\, |k\rangle = |i\rangle \otimes |k\rangle$ and has dimension nm.

An arbitrary vector in the space $S \otimes T$ is of the form

$$|\psi\rangle = \sum_{i=1}^{n} \sum_{k=1}^{m} \psi(i,k)\, |i\rangle \otimes |k\rangle = \sum_{i=1}^{n} \sum_{k=1}^{m} |i,k\rangle\langle i,k|\psi\rangle. \tag{1.444}$$

Vectors $|\phi_S, \chi_T\rangle$ that are tensor products $|\phi_S\rangle \otimes |\chi_T\rangle$ of two vectors $|\phi_S\rangle \in S$

and $|\chi_T\rangle \in T$

$$|\phi_S\rangle \otimes |\chi_T\rangle = \left(\sum_{i=1}^{n} \phi_i |i\rangle\right) \otimes \left(\sum_{k=1}^{m} \chi_k |k\rangle\right) = \sum_{i=1}^{n}\sum_{k=1}^{m} \phi_i \chi_k |i,k\rangle \tag{1.445}$$

are **separable**. States represented by vectors that are not separable are said to be **entangled**. Most states in a tensor-product space are entangled.

In the simpler notation $|i,k\rangle$ for $|i\rangle \otimes |k\rangle$, a tensor-product operator $A \otimes B$ maps an arbitrary vector (1.444) to

$$(A \otimes B)\,|\psi\rangle = \sum_{i,j=1}^{n}\sum_{k,\ell=1}^{m} |i,k\rangle\,\langle i|A|j\rangle\langle k|B|\ell\rangle\,\langle j,\ell|\psi\rangle. \tag{1.446}$$

Direct-product operators are special. An arbitrary linear operator on the space $S \otimes T$

$$D = \sum_{i,j=1}^{n}\sum_{k,\ell=1}^{m} |i,k\rangle\langle i,k|D|j,\ell\rangle\langle j,\ell| \tag{1.447}$$

maps an arbitrary vector (1.444) into the vector

$$D\,|\psi\rangle = \sum_{i,j=1}^{n}\sum_{k,\ell=1}^{m} |i,k\rangle\langle i,k|D|j,\ell\rangle\langle j,\ell|\psi\rangle. \tag{1.448}$$

Example 1.61 (States of the hydrogen atom) Suppose the state $|n,\ell,m\rangle$ is an eigenvector 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 of a hydrogen atom without spin:

$$\begin{aligned} H|n,\ell,m\rangle &= E_n|n,\ell,m\rangle \\ \boldsymbol{L}^2|n,\ell,m\rangle &= \hbar^2\ell(\ell+1)|n,\ell,m\rangle \\ L_3|n,\ell,m\rangle &= \hbar m|n,\ell,m\rangle. \end{aligned} \tag{1.449}$$

The state $|n,\ell,m\rangle = |n\rangle \otimes |\ell,m\rangle$ is separable. Suppose the states $|\sigma\rangle$ for $\sigma = \pm$ are eigenstates of the third component S_3 of the operator $\boldsymbol{S}$ that represents the spin of the electron

$$S_3|\sigma\rangle = \sigma\frac{\hbar}{2}|\sigma\rangle. \tag{1.450}$$

The separable, tensor-product states

$$|n,\ell,m,\sigma\rangle \equiv |n,\ell,m\rangle \otimes |\sigma\rangle \equiv |n,\ell,m\rangle|\sigma\rangle \tag{1.451}$$

represent a hydrogen atom including the spin of its electron. These separable states are eigenvectors of all four operators H, $\boldsymbol{L}^2$, L_3, and S_3:

$$\begin{aligned} H|n,\ell,m,\sigma\rangle &= E_n|n,\ell,m,\sigma\rangle \quad \boldsymbol{L}^2|n,\ell,m,\sigma\rangle = \hbar^2\ell(\ell+1)|n,\ell,m,\sigma\rangle \\ L_3|n,\ell,m,\sigma\rangle &= \hbar m|n,\ell,m,\sigma\rangle \quad S_3|n,\ell,m,\sigma\rangle = \sigma\tfrac{1}{2}\hbar|n,\ell,m,\sigma\rangle. \end{aligned} \tag{1.452}$$

Suitable linear combinations of these states are eigenstates 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. Many of these states are entangled. □

Example 1.62 (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, $S_a = \frac{1}{2}\hbar\,\sigma_a$, the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.453}$$

which are both hermitian and unitary. They map the basis vectors

$$|+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.454}$$

to $\sigma_1|\pm\rangle = |\mp\rangle$, $\sigma_2|\pm\rangle = \pm i|\mp\rangle$, and $\sigma_3|\pm\rangle = \pm|\pm\rangle$.

Suppose two spin operators $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ act on two spin-one-half systems with states $|\pm\rangle_1$ that are eigenstates of $S_3^{(1)}$ and states $|\pm\rangle_2$ that are eigenstates of $S_3^{(2)}$

$$S_3^{(1)}|\pm\rangle_1 = \pm\tfrac{1}{2}\hbar|\pm\rangle_1 \quad \text{and} \quad S_3^{(2)}|\pm\rangle_2 = \pm\tfrac{1}{2}\hbar|\pm\rangle_2. \tag{1.455}$$

Then the tensor-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 $S_3^{(2)}$

$$S_3^{(1)}|\pm, s_2\rangle = \pm\tfrac{1}{2}\hbar\,|+, s_2\rangle \quad \text{and} \quad S_3^{(2)}|s_1,\pm\rangle = \pm\tfrac{1}{2}\hbar\,|s_1,\pm\rangle. \tag{1.456}$$

These states also are eigenstates of the third component of the spin operator of the combined system

$$S_3 = S_3^{(1)} + S_3^{(2)} \quad \text{that is} \quad S_3|s_1,s_2\rangle = \tfrac{1}{2}\hbar\,(s_1+s_2)\,|s_1,s_2\rangle. \tag{1.457}$$

Thus $S_3|+,+\rangle = \hbar|+,+\rangle$, and $S_3|-,-\rangle = -\hbar|-,-\rangle$, while $S_3|+,-\rangle = 0$ and $S_3|-,+\rangle = 0$.

Using the notation (1.454), we can compute the effect of the operator $\boldsymbol{S}^2$ on the state $|++\rangle$. We find for S_1^2

$$\begin{aligned} S_1^2|++\rangle &= \left(S_1^{(1)} + S_1^{(2)}\right)^2 |++\rangle = \frac{\hbar^2}{4}\left(\sigma_1^{(1)} + \sigma_1^{(2)}\right)^2 |++\rangle \\ &= \tfrac{1}{2}\hbar^2 \left(1 + \sigma_1^{(1)}\sigma_1^{(2)}\right)|++\rangle = \tfrac{1}{2}\hbar^2 \left(|++\rangle + \sigma_1^{(1)}|+\rangle\sigma_1^{(2)}|+\rangle\right) \\ &= \tfrac{1}{2}\hbar^2 \left(|++\rangle + |--\rangle\right) \end{aligned} \tag{1.458}$$

and leave S_2^2 and S_3^2 to exercise 1.36. □

Example 1.63 (Entangled states) A neutral pion π^0 has zero angular momentum and negative parity. Its mass is 135 MeV/c^2 and 99% of them decay into two photons with a mean lifetime of 8.5×10^{-17} s. A π^0 at rest decays into two photons moving in opposite directions along the same axis, and the spins of the photons must be either parallel to their momenta $|+,+\rangle$, positive helicity, or antiparallel to their momenta $|-,-\rangle$, negative helicity. Parity reverses helicity, and so the state of negative parity and zero angular momentum is

$$|\gamma,\gamma\rangle = \frac{1}{\sqrt{2}}\Big(|+,+\rangle - |-,-\rangle\Big). \tag{1.459}$$

The two photons have the same helicity. If the helicity of one photon is measured to be positive, then a measurement of the other photon will show it to have positive helicity. The state is entangled.

One π^0 in 17 million will decay into a positron and an electron in a state of zero angular momentum. The spin part of the final state is

$$|e^+,e^-\rangle = \frac{1}{\sqrt{2}}\Big(|+,-\rangle - |-,+\rangle\Big). \tag{1.460}$$

If the spin along any axis of one of the electrons is measured to be positive, then a measurement of the spin of the other electron along the same axis will be negative. The state is entangled. □

1.35 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 $\{|k\rangle\}$ is any complete set of orthonormal states

$$I = \sum_k |k\rangle\langle k| \tag{1.461}$$

then the probability that the system is in the state $|k\rangle$ is

$$p_k = \langle k|\rho|k\rangle = \mathrm{Tr}\left(\rho|k\rangle\langle k|\right). \tag{1.462}$$

Since $\mathrm{Tr}\rho = 1$, the sum of these probabilities is unity

$$\sum_k p_k = \sum_k \langle k|\rho|k\rangle = \mathrm{Tr}\left(\rho \sum_k |k\rangle\langle k|\right) = \mathrm{Tr}\left(\rho I\right) = \mathrm{Tr}\rho = 1. \tag{1.463}$$

A system that is measured to be in a state $|k\rangle$ cannot simultaneously be measured to be in an orthogonal state $|\ell\rangle$. The probabilities sum to unity because the system must be in some state.

Since the density operator ρ is hermitian and positive, it has a complete, orthonormal set of eigenvectors $|k\rangle$ all of which have nonnegative eigenvalues ρ_k

$$\rho|k\rangle = \rho_k|k\rangle. \tag{1.464}$$

They afford for it an expansion in their outer products

$$\rho = \sum_k \rho_k |k\rangle\langle k| \tag{1.465}$$

each weighted by the probability ρ_k that the system is in the state $|k\rangle$.

A system composed of two systems, one with basis kets $|i\rangle$ and the other with basis kets $|k\rangle$, has basis states $|i,k\rangle = |i\rangle|k\rangle$ and can be described by the density operator

$$\rho = \sum_{ijk\ell} |i,k\rangle\langle i,k|\rho|j,\ell\rangle\langle j,\ell|. \tag{1.466}$$

The density operator for the first system is the trace of ρ over the states $|k\rangle$ of the second system

$$\rho_1 = \sum_k \langle k|\rho|k\rangle = \sum_{ijk} |i\rangle\langle i,k|\rho|j,k\rangle\langle j| \tag{1.467}$$

and similarly the density operator for the second system is the trace of ρ over the states $|i\rangle$ of the first system

$$\rho_2 = \sum_i \langle i|\rho|i\rangle = \sum_{jk\ell} |k\rangle\langle i,k|\rho|i,\ell\rangle\langle \ell|. \tag{1.468}$$

Classical entropy is an **extensive** quantity like volume, mass, and energy. The classical entropy of a composite system is the sum of the classical entropies of its parts. But quantum entropy $S = -k\mathrm{Tr}(\rho \log \rho)$ is not necessarily extensive. The quantum entropy of an entangled system can be less than the sum of the quantum entropies of its parts. The quantum entropy of each of the eigenstates $|\gamma, \gamma\rangle$ and $|e^+, e^-\rangle$ of example 1.63 is zero, but the sum of the quantum entropies of their parts is in both cases $2k \log 2$.

1.36 Schmidt decomposition

Suppose $|\psi\rangle$ is an arbitrary vector in the tensor product of the vector spaces B and C

$$|\psi\rangle = \sum_{i=1}^{n} \sum_{k=1}^{m} A_{ik} \, |i\rangle \otimes |k\rangle. \tag{1.469}$$

The arbitrary matrix A has a singular-value decomposition (1.396)

$$A_{ik} = \sum_{\ell=1}^{\min(n,m)} U_{i\ell} \, S_\ell \, V^\dagger_{\ell k}. \tag{1.470}$$

In terms of this SVD, the vector $|\psi\rangle$ is

$$\begin{aligned} |\psi\rangle &= \sum_{i=1}^{n} \sum_{k=1}^{m} \sum_{\ell=1}^{\min(n,m)} U_{i\ell} \, S_\ell \, V^\dagger_{\ell k} \, |i\rangle \otimes |k\rangle \\ &= \sum_{\ell=1}^{\min(n,m)} S_\ell \, |U, \ell\rangle \otimes |V^\dagger, \ell\rangle \end{aligned} \tag{1.471}$$

where the state $|U, \ell\rangle$ is

$$|U, \ell\rangle = \sum_{i=1}^{n} U_{i\ell} \, |i\rangle \tag{1.472}$$

and is in the vector space B, and the state $|V^\dagger, \ell\rangle$ is

$$|V^\dagger, \ell\rangle = \sum_{k=1}^{m} V^\dagger_{\ell k} \, |k\rangle \tag{1.473}$$

and is in the vector space C. The states $|U, \ell\rangle$ and $|V^\dagger, \ell\rangle$ are orthonormal

$$\langle U, \ell | U, \ell' \rangle = \delta_{\ell\ell'} \quad \text{and} \quad \langle V^\dagger, \ell | V^\dagger, \ell' \rangle = \delta_{\ell\ell'} \tag{1.474}$$

because the matrices U and V are unitary.

The outer product of a tensor-product state (1.469) is a pure-state density operator

$$\rho = |\psi\rangle\langle\psi| = \sum_{\ell\ell'} S_\ell S_{\ell'} (|U,\ell\rangle \otimes |V^\dagger,\ell\rangle)(\langle U,\ell'| \otimes \langle V^\dagger,\ell'|). \tag{1.475}$$

Taking the trace over a complete set of orthonormal states in B or C, we get the density operator ρ in the spaces B or C

$$\begin{aligned} \rho_B &= \mathrm{Tr}_C(\rho) = \sum_{\ell''} \langle V^\dagger,\ell''|\rho|V^\dagger,\ell''\rangle = \sum_\ell S_\ell^2 \, |U,\ell\rangle\langle U,\ell| \\ \rho_C &= \mathrm{Tr}_B(\rho) = \sum_{\ell''} \langle U,\ell''|\rho|U,\ell''\rangle = \sum_\ell S_\ell^2 \, |V^\dagger,\ell\rangle\langle V^\dagger,\ell|. \end{aligned} \tag{1.476}$$

The density operators ρ_B and ρ_C have the same eigenvalues and therefore the same von Neumann entropy

$$\begin{aligned} s(\rho_B) &= -k\,\mathrm{Tr}_B(\rho_B \log \rho_B) = -k \sum_\ell S_\ell^2 \, \log(S_\ell^2) \\ s(\rho_C) &= -k\,\mathrm{Tr}_C(\rho_C \log \rho_C) = -k \sum_\ell S_\ell^2 \, \log(S_\ell^2). \end{aligned} \tag{1.477}$$

The nonzero, positive singular values S_ℓ are called Schmidt coefficients. The number of them is the Schmidt rank or Schmidt number of the state $|\psi\rangle$. The state $|\psi\rangle$ is entangled if and only if its Schmidt rank is greater than unity.

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

$$(f,g) \equiv \langle f|\rho|g\rangle \tag{1.478}$$

for $g = f$ is nonnegative, $(f,f) = \langle f|\rho|f\rangle \geq 0$, and satisfies the other conditions (1.78, 1.79, & 1.81) for a Schwarz inner product.

The second Schwarz inner product applies to operators A and B and is defined (Titulaer and Glauber, 1965) as

$$(A,B) = \mathrm{Tr}\left(\rho A^\dagger B\right) = \mathrm{Tr}\left(B\rho A^\dagger\right) = \mathrm{Tr}\left(A^\dagger B\rho\right). \tag{1.479}$$

This inner product is nonnegative when $A = B$ and obeys the other rules (1.78, 1.79, & 1.81) for a Schwarz inner product.

These two degenerate inner products are not inner products in the strict

sense of (1.78–1.84), but they are Schwarz inner products, and so (1.98–1.99) they satisfy the Schwarz inequality (1.99)

$$(f,f)(g,g) \geq |(f,g)|^2. \tag{1.480}$$

Applied to the first, vector, Schwarz inner product (1.478), the Schwarz inequality gives

$$\langle f|\rho|f\rangle\langle g|\rho|g\rangle \geq |\langle f|\rho|g\rangle|^2 \tag{1.481}$$

which is a useful property of density matrices. Application of the Schwarz inequality to the second, operator, Schwarz inner product (1.479) gives (Titulaer and Glauber, 1965)

$$\mathrm{Tr}\left(\rho A^\dagger A\right)\mathrm{Tr}\left(\rho B^\dagger B\right) \geq \left|\mathrm{Tr}\left(\rho A^\dagger B\right)\right|^2. \tag{1.482}$$

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 $E_i^{(-)}(x) = (E_i^{(+)}(x))^\dagger$

$$E_i(x) = E_i^{(+)}(x) + E_i^{(-)}(x). \tag{1.483}$$

Glauber has defined the first-order correlation function $G_{ij}^{(1)}(x,y)$ as (Glauber, 1963b)

$$G_{ij}^{(1)}(x,y) = \mathrm{Tr}\left(\rho E_i^{(-)}(x)E_j^{(+)}(y)\right) \tag{1.484}$$

or in terms of the operator inner product (1.479) as

$$G_{ij}^{(1)}(x,y) = \left(E_i^{(+)}(x), E_j^{(+)}(y)\right). \tag{1.485}$$

By setting $A = E_i^{(+)}(x)$ and $B = E_j^{(+)}(y)$ in the Schwarz inequality (1.482), we find that the correlation function $G_{ij}^{(1)}(x,y)$ is bounded by

$$\left|G_{ij}^{(1)}(x,y)\right|^2 \leq G_{ii}^{(1)}(x,x)G_{jj}^{(1)}(y,y). \tag{1.486}$$

Interference fringes are sharpest when this inequality is saturated

$$\left|G_{ij}^{(1)}(x,y)\right|^2 = G_{ii}^{(1)}(x,x)G_{jj}^{(1)}(y,y) \tag{1.487}$$

which can occur only if the correlation function $G_{ij}^{(1)}(x,y)$ factorizes (Titulaer and Glauber, 1965)

$$G_{ij}^{(1)}(x,y) = \mathcal{E}_i^*(x)\mathcal{E}_j(y) \tag{1.488}$$

as it does when the density operator is an outer product of coherent states

$$\rho = |\{\alpha_k\}\rangle\langle\{\alpha_k\}| \tag{1.489}$$

which are eigenstates of $E_i^{(+)}(x)$ with eigenvalue $\mathcal{E}_i(x)$ (Glauber, 1963b,a)

$$E_i^{(+)}(x)|\{\alpha_k\}\rangle = \mathcal{E}_i(x)|\{\alpha_k\}\rangle. \tag{1.490}$$

The higher-order correlation functions

$$G^{(n)}_{i_1\dots i_{2n}}(x_1\dots x_{2n}) = \mathrm{Tr}\left(\rho E^{(-)}_{i_1}(x_1)\dots E^{(-)}_{i_n}(x_n)E^{(+)}_{i_{n+1}}(x_{n+1})\dots E^{(+)}_{i_{2n}}(x_n)\right) \tag{1.491}$$

satisfy similar inequalities (Glauber, 1963b) which also follow from the Schwarz inequality (1.482). (Roy Jay Glauber, 1925–2019)

1.38 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 1.64 (Rank) The 3×4 matrix

$$A = \begin{pmatrix} 1 & 0 & 1 & -2 \\ 2 & 2 & 0 & 2 \\ 4 & 3 & 1 & 1 \end{pmatrix} \tag{1.492}$$

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, $2r_1 + 3r_2 - 2r_3 = 0$, 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. □

1.39 Software

High-quality software for virtually all numerical problems in linear algebra are available in the linear-algebra package Lapack. Lapack codes in Fortran and C++ are available at `netlib.org/lapack/` and at `math.nist.gov/tnt/`. Apple's Xcode command `-framework accelerate` links this software into gnu

executables. The Basic Linear Algebra Subprograms (BLAS) on which Lapack is based are also available in Java at `icl.cs.utk.edu/f2j/` and at `math.nist.gov/javanumerics/`.

Matlab solves a wide variety of numerical problems. A free GNU version is available at `gnu.org/software/octave/`. Maple and Mathematica are good commercial programs for numerical and symbolic problems. Python (`python.org`), Scientific Python (`scipy.org`), and Sage (`sagemath.org`) are websites of free software of broad applicability. Maxima, xMaxima, and wxMaxima (`maxima.sourceforge.net`) are free Lisp programs that excel at computer algebra. Intel gives software to students and teachers (`software.intel.com`).

Exercises

1.1 What is the most general function of three Grassmann numbers $\theta_1, \theta_2, \theta_3$?

1.2 Derive the cyclicity (1.24) of the trace from Eq. (1.23).

1.3 Show that $(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$, which is Eq.(1.26).

1.4 Show that a real hermitian matrix is symmetric.

1.5 Show that $(AB)^\dagger = B^\dagger A^\dagger$, which is Eq.(1.29).

1.6 Show that the matrix (1.41) 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 (1.46) 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 (1.46) and their adjoints $a_i^\dagger = \theta_i^\dagger$ satisfy the anticommutation relations (1.47) of the creation and annihilation operators for a system with two fermionic states.

1.9 Derive (1.66) from (1.63–1.65).

1.10 Fill in the steps leading to the formulas (1.74) for the vectors b'_1 and b'_2 and the formula (1.75) for the matrix a'.

1.11 Show that the antilinearity (1.81) of the inner product follows from its first two properties (1.78 & 1.79).

1.12 Show that the Minkowski product $(x, y) = \boldsymbol{x} \cdot \boldsymbol{y} - x^0 y^0$ of two 4-vectors x and y is an inner product obeying the rules (1.78, 1.79, and 1.84).

1.13 Show that if $f = 0$, then the linearity (1.79) of the inner product implies that (f, f) and (g, f) vanish.

1.14 Show that the condition (1.80) of being positive definite implies nondegeneracy (1.84).

1.15 Show that the nonnegativity (1.82) of the Schwarz inner product implies the condition (1.83). 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 (1.103) follows from the Schwarz inequality (1.102).

1.17 Show that the inequality (1.105) follows from the Schwarz inequality (1.104).

1.18 Use the Gram-Schmidt method to find orthonormal linear combinations of the three vectors

$$\boldsymbol{s_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{s_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{s_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{1.493}$$

1.19 Now use the Gram-Schmidt method to find orthonormal linear combinations of the same three vectors but in a different order

$$\boldsymbol{s'_1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{s'_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{s'_3} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{1.494}$$

Did you get the same orthonormal vectors as in the previous exercise?

1.20 Derive the linearity (1.125) of the outer product from its definition (1.124).

1.21 Show that a linear operator A that is represented by a hermitian matrix (1.167) in an orthonormal basis satisfies $(g, A\, f) = (A\, g, f)$.

1.22 Show that a unitary operator maps one orthonormal basis into another.

1.23 Show that the integral (1.186) defines a unitary operator that maps the state $|x'\rangle$ to the state $|x' + a\rangle$.

1.24 For the 2×2 matrices

$$A = \begin{pmatrix} 1 & 2 \\ 3 & -4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & -1 \\ 4 & -3 \end{pmatrix} \tag{1.495}$$

verify equations (1.220–1.222).

1.25 Derive the least-squares solution (1.248) 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 (1.286)?

1.28 Use (1.297) to derive expression (1.298) 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 Use the characteristic equation of exercise 1.29 to derive identities (1.301) and (1.302) for the 3×3 real orthogonal matrix $\exp(-i\boldsymbol{\theta}\cdot\boldsymbol{J})$.

1.31 Show that the sum of the eigenvalues of a normal antisymmetric matrix vanishes.

1.32 Consider the 2×3 matrix A

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -3 & 0 & 1 \end{pmatrix}. \tag{1.496}$$

Perform the singular value decomposition $A = USV^{\mathsf{T}}$, where V^{T} the transpose of V. Use Matlab or another program to find the singular values and the real orthogonal matrices U and V.

1.33 Consider the 6×9 matrix A with elements $A_{j,k} = x + x^j + i(y - y^k)$ in which $x = 1.1$ and $y = 1.02$. Use Matlab or another program to find the singular values, and the first left and right singular vectors.

1.34 Show that the totally antisymmetric Levi-Civita symbol ϵ_{ijk} where $\epsilon_{123} = 1$ satisfies the useful relation

$$\sum_{i=1}^{3} \epsilon_{ijk}\,\epsilon_{inm} = \delta_{jn}\,\delta_{km} - \delta_{jm}\,\delta_{kn}. \tag{1.497}$$

1.35 Consider the hamiltonian $H = \frac{1}{2}\hbar\omega\sigma_3$ where σ_3 is defined in (1.453). The entropy S of this system at temperature T is $S = -k\mathrm{Tr}\,[\rho\ln(\rho)]$ in which the density operator ρ is

$$\rho = \frac{e^{-H/(kT)}}{\mathrm{Tr}\left[e^{-H/(kT)}\right]}. \tag{1.498}$$

Find expressions for the density operator ρ and its entropy S.

1.36 Use example 1.62 to find the action of the operator $\boldsymbol{S}^2 = \left(\boldsymbol{S}^{(1)} + \boldsymbol{S}^{(2)}\right)^2$ 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 $|t, u, v\rangle \equiv (a_1^\dagger)^t (a_2^\dagger)^u (a_3^\dagger)^v |0, 0, 0\rangle$. We can represent them by eight 8-vectors each of which has seven 0's with a 1 in position $4t + 2u + v + 1$. 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 (1.478) is degenerate because it can violate (1.84) for certain density operators and certain pairs of states.

1.39 Show that the Schwarz inner product (1.479) is degenerate because it can violate (1.84) 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

$$E_i^{(+)}(x) = \sum_k a_k \, \mathcal{E}_i^{(+)}(k) \, e^{i(kx - \omega t)}. \tag{1.499}$$

Show that $|\{\alpha_k\}\rangle$ is an eigenstate of $E_i^{(+)}(x)$ as in (1.490) and find its eigenvalue $\mathcal{E}_i(x)$.

1.41 Show that if X is a nondefective, nonsingular square matrix, then the variation of the logarithm of its determinant is $\delta \ln(\det X) = \mathrm{Tr}(X^{-1}\delta X)$.