

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

$$x = \sum_{n=m_x}^{\infty} \frac{d_n}{10^n} \quad (1.1)$$

that do not repeat. Thus the repeating decimals $1/2 = 0.50000\dots$ and $1/3 = 0.\bar{3} \equiv 0.33333\dots$ are rational, while $\pi = 3.141592654\dots$ 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

$$z = x + iy. \quad (1.2)$$

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

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

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

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

The polar representation of $z = x + iy$ is

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

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

$$r = |z| = \sqrt{x^2 + y^2} \quad (1.6)$$

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

$$\theta = \arctan(y/x). \quad (1.7)$$

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 $\text{atan2}(y, x)$ 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 $z = x + iy$

$$z^* = \bar{z} = x - iy. \quad (1.8)$$

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

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

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

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

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

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

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

$$f(\theta_1, \theta_2) = a + b\theta_1 + c\theta_2 + d\theta_1\theta_2 \quad (1.12)$$

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 1-dimensional array (a_1, a_2, \dots, 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 2-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 i th row and the k th column.

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

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n) \quad (1.13)$$

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

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

One can multiply arrays by numbers: Thus z times the 3-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

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

or an $m \times n$ matrix with entries $(ba)_{ki} = b_k a_i$. If a and b are complex, then one also can form the outer products $(\bar{a} b)_{ik} = \bar{a}_i b_k$, $(\bar{b} a)_{ki} = \bar{b}_k a_i$, and $(\bar{b} \bar{a})_{ki} = \bar{b}_k \bar{a}_i$. The outer product of a matrix a_{ik} and a 3-dimensional array $b_{j\ell m}$ is a five-dimensional array

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

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

$$(a, b) = \langle a|b \rangle = a \cdot b = (a_1, \dots, a_n) \cdot (b_1, \dots, b_n) = a_1 b_1 + \dots + a_n b_n. \quad (1.17)$$

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

$$(a, b) = \langle a|b \rangle = \bar{a} \cdot b = (\bar{a}_1, \dots, \bar{a}_n) \cdot (b_1, \dots, b_n) = \bar{a}_1 b_1 + \dots + \bar{a}_n b_n \quad (1.18)$$

or as its complex conjugate

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

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 i th component is

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

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

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

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

1.3 Matrices

Matrices are 2-dimensional arrays.

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

$$\text{Tr } a = \text{tr } a = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^n a_{ii}. \quad (1.22)$$

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

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

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

$$\text{Tr}(a b c \dots z) = \text{Tr}(b c \dots z a) = \text{Tr}(c \dots z a b) = \dots \quad (1.24)$$

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

$$(a^T)_{ij} = a_{ji}. \quad (1.25)$$

Mathematicians often use a prime to mean transpose, as in $a' = a^T$, but physicists tend to use primes to label different objects or to indicate differentiation. One may show that transposition inverts the order of multiplication

$$(a b)^T = b^T a^T. \quad (1.26)$$

A matrix that is equal to its transpose

$$a = a^T \quad (1.27)$$

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

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

One may show that

$$(a b)^\dagger = b^\dagger a^\dagger. \quad (1.29)$$

A matrix that is equal to its adjoint

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

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

$$a = a^\dagger \quad (1.31)$$

(Charles Hermite 1822–1901).

Example 1.1 (The Pauli matrices) All three of Pauli's matrices

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

are hermitian (Wolfgang Pauli 1900–1958).

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

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

is real for all complex n -tuples v .

The **Kronecker delta** δ_{ik} is defined to be unity if $i = k$ and zero if $i \neq k$

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

(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

$$a^{-1} a = a a^{-1} = I \quad (1.35)$$

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

$$U^\dagger U = U U^\dagger = I \quad (1.36)$$

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

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

$$O^T O = O O^T = I. \quad (1.37)$$

Orthogonal matrices are square.

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

$$A \geq 0 \quad (1.38)$$

if for all complex vectors V the quadratic form

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

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

$$\langle V|A|V \rangle > 0 \quad (1.40)$$

for all nonzero vectors $|V\rangle$.

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

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad (1.41)$$

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

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (1.42)$$

can represent the number i since

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I. \quad (1.43)$$

The 2×2 matrix

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad (1.44)$$

can represent a Grassmann number since

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0. \quad (1.45)$$

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

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

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 operators, satisfy the anticommutation relations

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

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

$$W = xU + yV \quad (1.48)$$

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

$$U = u_1 B_1 + u_2 B_2 + \dots + u_n B_n \quad (1.49)$$

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, \dots, 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

$$H = nW + mN \quad (1.50)$$

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 2-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

$$z = x1 + yi = x + iy. \quad (1.51)$$

The complex numbers are a 2-dimensional vector space over the real numbers, and the vectors 1 and i are a basis for it.

The complex numbers also form a 1-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 1-dimensional vector space is the set of all $z = z1$ for arbitrary complex z .

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

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \quad (1.52)$$

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 \mathbf{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 $\hat{\mathbf{x}}$

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

The components of the vector \mathbf{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 3-dimensional space is the set of all linear combinations

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \quad (1.54)$$

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 3-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 $x = y = 0$

1.4 Vectors

9

$$a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} \quad (1.55)$$

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 $[a, b]$

$$f(x) = \sum_i z_i f_i(x) \quad (1.56)$$

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

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle \quad (1.57)$$

also is a possible state of the system.

A harmonic oscillator in its k th 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 \mathbf{p} and spin σ is in a state represented by a vector $|\mathbf{p}, \sigma\rangle$. 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

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

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

$$A(bV + cW) = bA(V) + cA(W) = bAV + cAW. \quad (1.59)$$

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

$$AB_i = a_{1i}B_1 + a_{2i}B_2 + \cdots + a_{ni}B_n = \sum_{k=1}^n a_{ki}B_k \quad (1.60)$$

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 $V = u_1B_1 + u_2B_2 + \cdots + u_nB_n$ in S then is

$$AV = A \sum_{i=1}^n u_i B_i = \sum_{i=1}^n u_i AB_i = \sum_{i,k=1}^n u_i a_{ki} B_k = \sum_{i,k=1}^n a_{ki} u_i B_k. \quad (1.61)$$

So the k th component u'_k of the vector $V' = AV$ is

$$u'_k = a_{k1}u_1 + a_{k2}u_2 + \cdots + a_{kn}u_n = \sum_{i=1}^n a_{ki}u_i. \quad (1.62)$$

Thus the column vector u' of the components u'_k of the vector $V' = AV$ is the product $u' = au$ 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

$$B'_i = \sum_{\ell=1}^n u_{\ell i} B_\ell \quad (1.63)$$

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

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