Physical Mathematics

KEVIN CAHILL

CAMBRIDGE

CAMBRIDGE

more information – www.cambridge.org/9781107005211

Physical Mathematics

Unique in its clarity, examples, and range, *Physical Mathematics* explains as simply as possible the mathematics that graduate students and professional physicists need in their courses and research. The author illustrates the mathematics with numerous physical examples drawn from contemporary research. In addition to basic subjects such as linear algebra, Fourier analysis, complex variables, differential equations, and Bessel functions, this textbook covers topics such as the singular-value decomposition, Lie algebras, the tensors and forms of general relativity, the central limit theorem and Kolmogorov test of statistics, the Monte Carlo methods of experimental and theoretical physics, the renormalization group of condensed-matter physics, and the functional derivatives and Feynman path integrals of quantum field theory. Solutions to exercises are available for instructors at www.cambridge.org/cahill

Kevin Cahill is Professor of Physics and Astronomy at the University of New Mexico. He has done research at NIST, Saclay, Ecole Polytechnique, Orsay, Harvard, NIH, LBL, and SLAC, and has worked in quantum optics, quantum field theory, lattice gauge theory, and biophysics. *Physical Mathematics* is based on courses taught by the author at the University of New Mexico and at Fudan University in Shanghai.

Physical Mathematics

KEVIN CAHILL
University of New Mexico

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107005211

First published 2013

Printed and bound in the United Kingdom by the MPG Books Group

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Cahill, Kevin, 1941–, author.
Physical mathematics / Kevin Cahill, University of New Mexico.
pages cm
ISBN 978-1-107-00521-1 (hardback)
1. Mathematical physics. I. Title.
QC20.C24 2012
530.15–dc23
2012036027

ISBN 978-1-107-00521-1 Hardback

For Ginette, Mike, Sean, Peter, Mia, and James,
and in honor of Muntadhar al-Zaidi.

Contents

Preface

To the students: you will find some physics crammed in amongst the mathematics. Don't let the physics bother you. As you study the math, you'll learn some physics without extra effort. The physics is a freebie. I have tried to explain the math you need for physics and have left out the rest.

To the professors: the book is for students who also are taking mechanics, electrodynamics, quantum mechanics, and statistical mechanics nearly simultaneously and who soon may use probability or path integrals in their research. Linear algebra and Fourier analysis are the keys to physics, so the book starts with them, but you may prefer to skip the algebra or postpone the Fourier analysis. The book is intended to support a one- or two-semester course for graduate students or advanced undergraduates. The first seven, eight, or nine chapters fit in one semester, the others in a second. A list of errata is maintained at panda.unm.edu/cahill, and solutions to all the exercises are available for instructors at www.cambridge.org/cahill.

Several friends – Susan Atlas, Bernard Becker, Steven Boyd, Robert Burckel, Sean Cahill, Colston Chandler, Vageli Coutsias, David Dunlap, Daniel Finley, Franco Giuliani, Roy Glauber, Pablo Gondolo, Igor Gorelov, Jiaxing Hong, Fang Huang, Dinesh Loomba, Yin Luo, Lei Ma, Michael Malik, Kent Morrison, Sudhakar Prasad, Randy Reeder, Dmitri Sergatskov, and David Waxman – have given me valuable advice. Students have helped with questions, ideas, and corrections, especially Thomas Beechem, Marie Cahill, Chris Cesare, Yihong Cheng, Charles Cherqui, Robert Cordwell, Amo-Kwao Godwin, Aram Gragossian, Aaron Hankin, Kangbo Hao, Tiffany Hayes, Yiran Hu, Shanshan Huang, Tyler Keating, Joshua Koch, Zilong Li, Miao Lin, ZuMou Lin, Sheng Liu, Yue Liu, Ben Oliker, Boleszek Osinski, Ravi Raghunathan, Akash Rakholia, Xingyue Tian, Toby Tolley, Jiqun Tu, Christopher Vergien, Weizhen Wang, George Wendelberger, Xukun Xu, Huimin Yang, Zhou Yang, Daniel Young, Mengzhen Zhang, Lu Zheng, Lingjun Zhou, and Daniel Zirzow.

1

Linear algebra

1.1 Numbers

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Grassmann numbers θ_i are **anticommuting** numbers, that is, the **anti-commutator** of any two Grassmann numbers vanishes

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

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

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

(Hermann Grassmann, 1809–1877).

1.2 Arrays

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

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

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

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

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

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

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

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

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

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

or an $m \times n$ matrix with entries $(ba)_{ki} = b_k a_i$. If a and b are complex, then one also can form the outer products $(\overline{a}\, b)_{ik} = \overline{a_i}\, b_k$, $(\overline{b}\, a)_{ki} = \overline{b_k}\, a_i$, 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

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

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

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

or as its complex conjugate

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

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

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

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

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

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

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

1.3 Matrices

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

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

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

The trace of two matrices is independent of their order

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

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

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

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

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

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

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

A matrix that is equal to its transpose

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

is **symmetric**.

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

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

One may show that

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

A matrix that is equal to its adjoint

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

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

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

Example 1.1 (The Pauli matrices)

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

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

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

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

is real for all complex n-tuples v.

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

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

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

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

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

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

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

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

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

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

Orthogonal matrices are square.

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

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

if for all complex vectors V the quadratic form

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

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

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

for all nonzero vectors $|V\rangle$.

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

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

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

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

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

can represent the number i since

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

The 2×2 matrix

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

can represent a Grassmann number since

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

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

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

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

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

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

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

1.4 Vectors

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

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

also is a vector in the vector space S.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

form a vector space. □

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

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

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

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

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

also is a possible state of the system. □

1.5 Linear operators

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

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

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

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

The square matrix a_{ki} **represents** the linear operator A in the B_k basis. The effect of A on any vector $U = u_1 B_1 + u_2 B_2 + \cdots + u_N B_N$ in S then is

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

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

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

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

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

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

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

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

then the new basis vectors B'_i are given by

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

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

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

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

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

in matrix notation.

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

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

represent the basis vectors B_1 and B_2, then the matrix

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

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

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

then the vectors

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

would represent B_1 and B_2, and the matrix

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

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

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

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

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

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

in T.

1.6 Inner products

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

that is nonnegative when the two vectors are the same

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

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

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

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

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

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

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

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

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

that is nonnegative when the matrices are the same

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

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

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

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

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

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

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

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

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

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

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

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

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

1.7 The Cauchy–Schwarz inequality

For any two vectors f and g, the Schwarz inequality

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

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

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

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

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

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

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

Taking the square-root of each side, we get

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

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

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

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

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

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

and it implies (Exercise 1.16) that

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

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

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

and it implies (exercise 1.17) that

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

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

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

of the Schwarz inequality. □

1.8 Linear independence and completeness

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

$$c_1 V_1 + \cdots + c_N V_N = 0 \tag{1.101}$$

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

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

A set of N linearly independent vectors $V_1, V_2, \ldots, V_N$ that is maximal in a vector space S can represent any vector U in the space S as a linear combination of its vectors, $U = u_1 V_1 + \cdots + u_N V_N$. 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 $c, c_1, \ldots, c_N$, not all zero, that make the sum

$$c\,U + c_1 V_1 + \cdots + c_N V_N = 0 \tag{1.102}$$

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

$$U = -\frac{1}{c}\,(c_1 V_1 + \cdots + c_N V_N) = u_1 V_1 + \cdots + u_N V_N \tag{1.103}$$

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

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

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

1.9 Dimension of a vector space

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

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

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

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

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

Thus the linear combination

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

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

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

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

1.10 Orthonormal vectors

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

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

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

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

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

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

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

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

The normalized vector U_2 then is

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

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

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

and also to U_2

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

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

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

The normalized vector U_3 then is

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

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

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

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

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

The normalized vector then is

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

A basis is more convenient if its vectors are orthonormal.

1.11 Outer products

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

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

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

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

it follows that A is linear.

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

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

and in that basis A is the matrix with entries

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

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

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

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

then their outer product is the matrix

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

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

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

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

then their outer products are

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

as well as

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

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

1.12 Dirac notation

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

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

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

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

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

The original expansion (1.128) then must be

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

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

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

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

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

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

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

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

$$\langle f|g\rangle = \langle g|f\rangle^* \tag{1.134}$$
$$\langle f|z_1 g_1 + z_2 g_2\rangle = z_1\langle f|g_1\rangle + z_2\langle f|g_2\rangle \tag{1.135}$$
$$\langle z_1 f_1 + z_2 f_2|g\rangle = z_1^*\langle f_1|g\rangle + z_2^*\langle f_2|g\rangle \tag{1.136}$$
$$\langle f|f\rangle \ge 0 \quad \text{and} \quad \langle f|f\rangle = 0 \iff f = 0. \tag{1.137}$$

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

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

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

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

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

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

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