

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

© K. Cahill 2013

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

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

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

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

Contents

Preface

page xvii

1	Linear algebra	1
1.1	Numbers	1
1.2	Arrays	2
1.3	Matrices	4
1.4	Vectors	7
1.5	Linear operators	9
1.6	Inner products	11
1.7	The Cauchy–Schwarz inequality	14
1.8	Linear independence and completeness	15
1.9	Dimension of a vector space	16
1.10	Orthonormal vectors	16
1.11	Outer products	18
1.12	Dirac notation	19
1.13	The adjoint of an operator	22
1.14	Self-adjoint or hermitian linear operators	23
1.15	Real, symmetric linear operators	23
1.16	Unitary operators	24
1.17	Hilbert space	25
1.18	Antiunitary, antilinear operators	26
1.19	Symmetry in quantum mechanics	26
1.20	Determinants	27
1.21	Systems of linear equations	34
1.22	Linear least squares	34
1.23	Lagrange multipliers	35
1.24	Eigenvectors	37

CONTENTS

1.25	Eigenvectors of a square matrix	38
1.26	A matrix obeys its characteristic equation	41
1.27	Functions of matrices	43
1.28	Hermitian matrices	45
1.29	Normal matrices	50
1.30	Compatible normal matrices	52
1.31	The singular-value decomposition	55
1.32	The Moore–Penrose pseudoinverse	63
1.33	The rank of a matrix	65
1.34	Software	66
1.35	The tensor/direct product	66
1.36	Density operators	69
1.37	Correlation functions	69
	Exercises	71
2	Fourier series	75
2.1	Complex Fourier series	75
2.2	The interval	77
2.3	Where to put the $2\pi s$	77
2.4	Real Fourier series for real functions	79
2.5	Stretched intervals	83
2.6	Fourier series in several variables	84
2.7	How Fourier series converge	84
2.8	Quantum-mechanical examples	89
2.9	Dirac notation	96
2.10	Dirac’s delta function	97
2.11	The harmonic oscillator	101
2.12	Nonrelativistic strings	103
2.13	Periodic boundary conditions	103
	Exercises	105
3	Fourier and Laplace transforms	108
3.1	The Fourier transform	108
3.2	The Fourier transform of a real function	111
3.3	Dirac, Parseval, and Poisson	112
3.4	Fourier derivatives and integrals	115
3.5	Fourier transforms in several dimensions	119
3.6	Convolutions	121
3.7	The Fourier transform of a convolution	123
3.8	Fourier transforms and Green’s functions	124
3.9	Laplace transforms	125
3.10	Derivatives and integrals of Laplace transforms	127

CONTENTS

3.11	Laplace transforms and differential equations	128
3.12	Inversion of Laplace transforms	129
3.13	Application to differential equations	129
	Exercises	134
4	Infinite series	136
4.1	Convergence	136
4.2	Tests of convergence	137
4.3	Convergent series of functions	138
4.4	Power series	139
4.5	Factorials and the gamma function	141
4.6	Taylor series	145
4.7	Fourier series as power series	146
4.8	The binomial series and theorem	147
4.9	Logarithmic series	148
4.10	Dirichlet series and the zeta function	149
4.11	Bernoulli numbers and polynomials	151
4.12	Asymptotic series	152
4.13	Some electrostatic problems	154
4.14	Infinite products	157
	Exercises	158
5	Complex-variable theory	160
5.1	Analytic functions	160
5.2	Cauchy's integral theorem	161
5.3	Cauchy's integral formula	165
5.4	The Cauchy–Riemann conditions	169
5.5	Harmonic functions	170
5.6	Taylor series for analytic functions	171
5.7	Cauchy's inequality	173
5.8	Liouville's theorem	173
5.9	The fundamental theorem of algebra	174
5.10	Laurent series	174
5.11	Singularities	177
5.12	Analytic continuation	179
5.13	The calculus of residues	180
5.14	Ghost contours	182
5.15	Logarithms and cuts	193
5.16	Powers and roots	194
5.17	Conformal mapping	197
5.18	Cauchy's principal value	198
5.19	Dispersion relations	205

CONTENTS

5.20	Kramers–Kronig relations	207
5.21	Phase and group velocities	208
5.22	The method of steepest descent	210
5.23	The Abel–Plana formula and the Casimir effect	212
5.24	Applications to string theory	217
	Exercises	219
6	Differential equations	223
6.1	Ordinary linear differential equations	223
6.2	Linear partial differential equations	225
6.3	Notation for derivatives	226
6.4	Gradient, divergence, and curl	228
6.5	Separable partial differential equations	230
6.6	Wave equations	233
6.7	First-order differential equations	235
6.8	Separable first-order differential equations	235
6.9	Hidden separability	238
6.10	Exact first-order differential equations	238
6.11	The meaning of exactness	240
6.12	Integrating factors	242
6.13	Homogeneous functions	243
6.14	The virial theorem	243
6.15	Homogeneous first-order ordinary differential equations	245
6.16	Linear first-order ordinary differential equations	246
6.17	Systems of differential equations	248
6.18	Singular points of second-order ordinary differential equations	250
6.19	Frobenius’s series solutions	251
6.20	Fuch’s theorem	253
6.21	Even and odd differential operators	254
6.22	Wronski’s determinant	255
6.23	A second solution	255
6.24	Why not three solutions?	257
6.25	Boundary conditions	258
6.26	A variational problem	259
6.27	Self-adjoint differential operators	260
6.28	Self-adjoint differential systems	262
6.29	Making operators formally self adjoint	264
6.30	Wronskians of self-adjoint operators	265
6.31	First-order self-adjoint differential operators	266
6.32	A constrained variational problem	267

CONTENTS

6.33	Eigenfunctions and eigenvalues of self-adjoint systems	273
6.34	Unboundedness of eigenvalues	275
6.35	Completeness of eigenfunctions	277
6.36	The inequalities of Bessel and Schwarz	284
6.37	Green's functions	284
6.38	Eigenfunctions and Green's functions	287
6.39	Green's functions in one dimension	288
6.40	Nonlinear differential equations	289
	Exercises	293
7	Integral equations	296
7.1	Fredholm integral equations	297
7.2	Volterra integral equations	297
7.3	Implications of linearity	298
7.4	Numerical solutions	299
7.5	Integral transformations	301
	Exercises	304
8	Legendre functions	305
8.1	The Legendre polynomials	305
8.2	The Rodrigues formula	306
8.3	The generating function	308
8.4	Legendre's differential equation	309
8.5	Recurrence relations	311
8.6	Special values of Legendre's polynomials	312
8.7	Schlaefli's integral	313
8.8	Orthogonal polynomials	313
8.9	The azimuthally symmetric Laplacian	315
8.10	Laplacian in two dimensions	316
8.11	The Laplacian in spherical coordinates	317
8.12	The associated Legendre functions/polynomials	317
8.13	Spherical harmonics	319
	Exercises	323
9	Bessel functions	325
9.1	Bessel functions of the first kind	325
9.2	Spherical Bessel functions of the first kind	335
9.3	Bessel functions of the second kind	341
9.4	Spherical Bessel functions of the second kind	343
	Further reading	345
	Exercises	345

10	Group theory	348
10.1	What is a group?	348
10.2	Representations of groups	350
10.3	Representations acting in Hilbert space	351
10.4	Subgroups	353
10.5	Cosets	354
10.6	Morphisms	354
10.7	Schur's lemma	355
10.8	Characters	356
10.9	Tensor products	357
10.10	Finite groups	358
10.11	The regular representation	359
10.12	Properties of finite groups	360
10.13	Permutations	360
10.14	Compact and noncompact Lie groups	361
10.15	Lie algebra	361
10.16	The rotation group	366
10.17	The Lie algebra and representations of $SU(2)$	368
10.18	The defining representation of $SU(2)$	371
10.19	The Jacobi identity	374
10.20	The adjoint representation	374
10.21	Casimir operators	375
10.22	Tensor operators for the rotation group	376
10.23	Simple and semisimple Lie algebras	376
10.24	$SU(3)$	377
10.25	$SU(3)$ and quarks	378
10.26	Cartan subalgebra	379
10.27	Quaternions	379
10.28	The symplectic group $Sp(2n)$	381
10.29	Compact simple Lie groups	383
10.30	Group integration	384
10.31	The Lorentz group	386
10.32	Two-dimensional representations of the Lorentz group	389
10.33	The Dirac representation of the Lorentz group	393
10.34	The Poincaré group	395
	Further reading	396
	Exercises	397
11	Tensors and local symmetries	400
11.1	Points and coordinates	400
11.2	Scalars	401
11.3	Contravariant vectors	401

CONTENTS

11.4	Covariant vectors	402
11.5	Euclidean space in euclidean coordinates	402
11.6	Summation conventions	404
11.7	Minkowski space	405
11.8	Lorentz transformations	407
11.9	Special relativity	408
11.10	Kinematics	410
11.11	Electrodynamics	411
11.12	Tensors	414
11.13	Differential forms	416
11.14	Tensor equations	419
11.15	The quotient theorem	420
11.16	The metric tensor	421
11.17	A basic axiom	422
11.18	The contravariant metric tensor	422
11.19	Raising and lowering indices	423
11.20	Orthogonal coordinates in euclidean n -space	423
11.21	Polar coordinates	424
11.22	Cylindrical coordinates	425
11.23	Spherical coordinates	425
11.24	The gradient of a scalar field	426
11.25	Levi-Civita's tensor	427
11.26	The Hodge star	428
11.27	Derivatives and affine connections	431
11.28	Parallel transport	433
11.29	Notations for derivatives	433
11.30	Covariant derivatives	434
11.31	The covariant curl	435
11.32	Covariant derivatives and antisymmetry	436
11.33	Affine connection and metric tensor	436
11.34	Covariant derivative of the metric tensor	437
11.35	Divergence of a contravariant vector	438
11.36	The covariant Laplacian	441
11.37	The principle of stationary action	443
11.38	A particle in a gravitational field	446
11.39	The principle of equivalence	447
11.40	Weak, static gravitational fields	449
11.41	Gravitational time dilation	449
11.42	Curvature	451
11.43	Einstein's equations	453
11.44	The action of general relativity	455
11.45	Standard form	455

CONTENTS

11.46	Schwarzschild's solution	456
11.47	Black holes	456
11.48	Cosmology	457
11.49	Model cosmologies	463
11.50	Yang–Mills theory	469
11.51	Gauge theory and vectors	471
11.52	Geometry	474
	Further reading	475
	Exercises	475
12	Forms	479
12.1	Exterior forms	479
12.2	Differential forms	481
12.3	Exterior differentiation	486
12.4	Integration of forms	491
12.5	Are closed forms exact?	496
12.6	Complex differential forms	498
12.7	Frobenius's theorem	499
	Further reading	500
	Exercises	500
13	Probability and statistics	502
13.1	Probability and Thomas Bayes	502
13.2	Mean and variance	505
13.3	The binomial distribution	508
13.4	The Poisson distribution	511
13.5	The Gaussian distribution	512
13.6	The error function erf	515
13.7	The Maxwell–Boltzmann distribution	518
13.8	Diffusion	519
13.9	Langevin's theory of brownian motion	520
13.10	The Einstein–Nernst relation	523
13.11	Fluctuation and dissipation	524
13.12	Characteristic and moment-generating functions	528
13.13	Fat tails	530
13.14	The central limit theorem and Jarl Lindeberg	532
13.15	Random-number generators	537
13.16	Illustration of the central limit theorem	538
13.17	Measurements, estimators, and Friedrich Bessel	543
13.18	Information and Ronald Fisher	546
13.19	Maximum likelihood	550
13.20	Karl Pearson's chi-squared statistic	551

13.21	Kolmogorov's test	554
	Further reading	560
	Exercises	560
14	Monte Carlo methods	563
14.1	The Monte Carlo method	563
14.2	Numerical integration	563
14.3	Applications to experiments	566
14.4	Statistical mechanics	572
14.5	Solving arbitrary problems	575
14.6	Evolution	576
	Further reading	577
	Exercises	577
15	Functional derivatives	578
15.1	Functionals	578
15.2	Functional derivatives	578
15.3	Higher-order functional derivatives	581
15.4	Functional Taylor series	582
15.5	Functional differential equations	583
	Exercises	585
16	Path integrals	586
16.1	Path integrals and classical physics	586
16.2	Gaussian integrals	586
16.3	Path integrals in imaginary time	588
16.4	Path integrals in real time	590
16.5	Path integral for a free particle	593
16.6	Free particle in imaginary time	595
16.7	Harmonic oscillator in real time	595
16.8	Harmonic oscillator in imaginary time	597
16.9	Euclidean correlation functions	599
16.10	Finite-temperature field theory	600
16.11	Real-time field theory	603
16.12	Perturbation theory	605
16.13	Application to quantum electrodynamics	609
16.14	Fermionic path integrals	613
16.15	Application to nonabelian gauge theories	619
16.16	The Faddeev–Popov trick	620
16.17	Ghosts	622
	Further reading	624
	Exercises	624

CONTENTS

17	The renormalization group	626
17.1	The renormalization group in quantum field theory	626
17.2	The renormalization group in lattice field theory	630
17.3	The renormalization group in condensed-matter physics	632
	Exercises	634
18	Chaos and fractals	635
18.1	Chaos	635
18.2	Attractors	639
18.3	Fractals	639
	Further reading	642
	Exercises	642
19	Strings	643
19.1	The infinities of quantum field theory	643
19.2	The Nambu–Goto string action	643
19.3	Regge trajectories	646
19.4	Quantized strings	647
19.5	D-branes	647
19.6	String–string scattering	648
19.7	Riemann surfaces and moduli	649
	Further reading	650
	Exercises	650
	<i>References</i>	651
	<i>Index</i>	656

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 Coutsiias, David Dunlap, Daniel Finley, Franco Giuliani, Roy Glauber, Pablo Gondolo, Igor Gorelov, Jiaying 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.

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 the Europe until the seventeenth century.

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

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

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

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

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

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

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

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

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

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 $\theta_1\theta_2 \dots \theta_N$. 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, \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 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 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 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

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

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 three-dimensional array $b_{j\ell m}$ is a five-dimensional array

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

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

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

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 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 $b' = ab$ 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 ab is a matrix with entries

$$(ab)_{i\ell} = a_{i1} b_{1\ell} + \dots + a_{in} b_{n\ell}. \quad (1.21)$$

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

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

The trace of two matrices is independent of their order

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

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

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

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

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

Some mathematicians use a prime to mean transpose, as in $a' = a^\top$, but physicists tend to use primes to label different objects or to indicate differentiation. One may show that

$$(ab)^\top = b^\top a^\top. \quad (1.26)$$

A matrix that is equal to its transpose

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

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

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

One may show that

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

A matrix that is equal to its adjoint

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

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

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

Example 1.1 (The Pauli matrices)

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

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

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 $i \neq k$ (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

$$a^{-1} a = a a^{-1} = I \quad (1.34)$$

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

$$U^\dagger U = U U^\dagger = I \quad (1.35)$$

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

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

$$O^T O = O O^T = I. \quad (1.36)$$

Orthogonal matrices are square.

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

$$A \geq 0 \quad (1.37)$$

if for all complex vectors V the quadratic form

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

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. \square

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

$$W = xU + yV \quad (1.47)$$

also is a vector in the vector space S .

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

$$U = u_1B_1 + u_2B_2 + \dots + u_NB_N \quad (1.48)$$

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

$$H = nW + mN \quad (1.49)$$

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. \square

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

$$z = x1 + yi = x + iy. \quad (1.50)$$

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 . \square

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

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \quad (1.51)$$

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

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

So 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.** \square

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

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

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

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. \square

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 $x = y = 0$

$$a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} \quad (1.54)$$

form a vector space. \square

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

$$f(x) = \sum_i z_i f_i(x) \quad (1.55)$$

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

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

$$|\psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle \quad (1.56)$$

also is a possible state of the system. \square

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

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

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

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

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

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

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

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

Thus the column vector u' of the components u'_k of the vector $U' = AU$ 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 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

$$B_k = \sum_{\ell=1}^N u_{\ell k} B'_\ell \quad (1.61)$$

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

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

then the new basis vectors B'_i are given by

$$B'_i = \sum_{k=1}^N u_{ki}^{-1} B_k. \quad (1.63)$$

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

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

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**

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

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

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

represent the basis vectors B_1 and B_2 , then the matrix

$$a = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad (1.67)$$

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

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

then the vectors

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

would represent B_1 and B_2 , and the matrix

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

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

$$AB_i = \sum_{k=1}^M a_{ki} C_k \quad (1.71)$$

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

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

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

$$(f, g) = (g, f)^* \quad (1.73)$$

$$(f, z g + w h) = z (f, g) + w (f, h) \quad (1.74)$$

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

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

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

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)

$$(f, f) \geq 0. \quad (1.77)$$

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

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

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**

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

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)

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

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

$$\|f - g\|. \quad (1.81)$$

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

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

that is nonnegative when the two vectors are the same

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

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**. \square

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

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

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

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

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

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

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

that is nonnegative when the matrices are the same

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

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

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

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

The squared norm $\|f\|^2$ of a function $f(x)$ is

$$\|f\|^2 = \int_a^b dx |f(x)|^2. \quad (1.89)$$

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. \square

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

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

$$\mathbf{p} \cdot \mathbf{p} = \mathbf{p} \cdot \mathbf{p} - E^2/c^2 = -c^2m^2 \leq 0. \quad (1.90)$$

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). \square

1.7 The Cauchy–Schwarz inequality

For any two vectors f and g , the Schwarz inequality

$$(f, f)(g, g) \geq |(f, g)|^2 \quad (1.91)$$

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

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

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)

$$(f, f)(g, g) \geq |(f, g)|^2. \quad (1.93)$$

Taking the square-root of each side, we get

$$\|f\| \|g\| \geq |(f, g)|. \quad (1.94)$$

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

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

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

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

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

and it implies (Exercise 1.16) that

$$\|\mathbf{x}\| + \|\mathbf{y}\| \geq \|\mathbf{x} + \mathbf{y}\|. \quad (1.97)$$

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

$$(\mathbf{u}^* \cdot \mathbf{u}) (\mathbf{v}^* \cdot \mathbf{v}) \geq |\mathbf{u}^* \cdot \mathbf{v}|^2 = (\mathbf{u}^* \cdot \mathbf{u}) (\mathbf{v}^* \cdot \mathbf{v}) \cos^2 \theta \quad (1.98)$$

and it implies (exercise 1.17) that

$$\|\mathbf{u}\| + \|\mathbf{v}\| \geq \|\mathbf{u} + \mathbf{v}\|. \quad (1.99)$$

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

$$\int_a^b dx |f(x)|^2 \int_a^b dx |g(x)|^2 \geq \left| \int_a^b dx f^*(x)g(x) \right|^2 \quad (1.100)$$

of the Schwarz inequality. □

1.8 Linear independence and completeness

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

$$c_1 V_1 + \dots + c_N V_N = 0 \quad (1.101)$$

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, \dots, 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, $U = u_1 V_1 + \dots + u_N V_N$. 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, \dots, c_N , not all zero, that make the sum

$$c U + c_1 V_1 + \dots + c_N V_N = 0 \quad (1.102)$$

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

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

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

$U = u_1 V_1 + \dots + u_N V_N$ 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, \dots, V_N and W_1, \dots, 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 U s 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, \dots, 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

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

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

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

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

$$u_2 = V_2 - (U_1, V_2) U_1. \quad (1.110)$$

The normalized vector U_2 then is

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

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

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

and also to U_2

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

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

$$u_3 = V_3 - (U_1, V_3) U_1 - (U_2, V_3) U_2. \quad (1.114)$$

The normalized vector U_3 then is

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

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

$$u_k = V_k + \sum_{i=1}^{k-1} c_{ik} U_i \quad (1.116)$$

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

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

The normalized vector then is

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

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)

$$Ah = f(g, h). \quad (1.119)$$

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

$$A(zh + we) = f(g, zh + we) = zf(g, h) + wf(g, e) = zAh + wAe \quad (1.120)$$

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

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

and in that basis A is the matrix with entries

$$A_{ij} = f_i g_j^*. \quad (1.122)$$

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

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

then their outer product is the matrix

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

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

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

then their outer products are

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

as well as

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

Students should feel free to write down their own examples. \square

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 $Ah = f(g, h)$ 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

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

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

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

The original expansion (1.128) then must be

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

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

$$I = \sum_{n=1}^N |n\rangle \langle n|. \quad (1.131)$$

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

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

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

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

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

$$\langle f|g\rangle = \langle g|f\rangle^* \quad (1.134)$$

$$\langle f|z_1g_1 + z_2g_2\rangle = z_1\langle f|g_1\rangle + z_2\langle f|g_2\rangle \quad (1.135)$$

$$\langle z_1f_1 + z_2f_2|g\rangle = z_1^*\langle f_1|g\rangle + z_2^*\langle f_2|g\rangle \quad (1.136)$$

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

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. \square

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

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

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

$$|U_k\rangle = \frac{|u_k\rangle}{\sqrt{\langle u_k|u_k\rangle}}. \quad (1.139)$$

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

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