

Compact Lie Groups

Richard Koch

September 11, 2025

Contents

1	Preliminaries	8
1.1	Introduction	8
1.2	Killing's 1890 Paper	9
1.3	Structure of Lie Algebra of a Compact Group	10
1.4	Compact Groups and Killing's Classification	11
1.5	Irreducible Representations of Compact Lie Groups	11
2	Short Course in Representation Theory	12
2.1	Introduction	12
2.2	Decomposing Representations into Sums of Irreducible Representations . .	14
2.3	Integration and the Fundamental Theorem of Finite Group Representation Theory	16
2.4	Intermission: $U(n) \subset GL(n, R)$	17
2.5	Unitary Equivalence of Unitary Representations	18
2.6	Irreducible Representations of Abelian Groups	19
2.7	Orthogonality Relations	20
2.8	Representation Coefficients are Complete in $L(G)$	22
2.9	Intermission: Fourier Series	24
2.10	Group Characters	25
3	Representations over $R, C, \mathcal{H}$	29
3.1	Quaternionic Representations	29
3.2	Going Right and Going Left	30
3.3	The Type of a Representation	32
3.4	Small Group Examples	33
3.5	Three Types of Real Irreducible Representations	36
3.6	Three Types of Quaternionic Representations	37
3.7	Three Types of Complex Irreducible Representations	37
3.8	Another Way To Move Real Representations Right	41
3.9	Irreducible Representations over R and C	43
3.10	Irreducible Representations over $\mathcal{H}$	46

4	Representations of Lie Groups and Lie Algebras	50
4.1	Lie Group Representations	50
4.2	Representations of Finite and Compact Groups	52
4.3	Representations of Lie Algebras	53
5	Invariant Integrals on Lie Groups	55
5.1	Left Invariant Metrics, Right Invariant Metrics	55
5.2	Left and Right Invariant Integrals	56
6	Metrics That Are Both Left and Right Invariant	59
6.1	Existence	59
6.2	Consequences	60
6.3	Lie Algebras of Compact Groups	62
6.4	Lie Algebras over $\mathbb{C}$	64
6.5	Connection between the Compact and Complex Cases	65
6.6	Curvature and Myers' Theorem	66
6.7	Weyl's Theorem on Complete Reducibility	68
7	Maximal Tori	69
7.1	Definition; Conjugacy Theorem	69
7.2	Uniqueness of Maximal Tori up to Conjugacy	70
7.3	The Weyl Group	72
7.4	Two Proofs of the Conjugacy Theorem	73
8	Roots	77
8.1	Introduction	77
8.2	Irreducible Representations of Abelian Groups	77
8.3	The Action of a Maximal Torus on $\mathcal{G}$	78
8.4	The Many Faces of $\mathcal{T}$	79
8.5	Irreducible Complex Representations of T	80
8.6	Dimension ≤ 3	81
8.7	$su(2) \subset \mathcal{G}$	82
8.8	$Sp(1), SU(2), SO(3), SO(4)$, and $SL(2, \mathbb{R})$	84
8.9	Irreducible Representations of $sl(2, \mathbb{R})$, $su(2)$, and $so(3)$	87
8.10	Irreducible Representations of $SL(2, \mathbb{R})$, $SU(2)$, and $SO(3)$	89
8.11	Irreducible Representations of $SO(3)$	91
8.12	Miscellaneous Remarks on Representations of $SO(3)$	98
9	Classification of Root Systems	114
9.1	Outline of Chapter and Cautionary Note	114
9.2	Brackets of Root Spaces	115
9.3	Complex Root Spaces are One Dimensional	116

9.4	Another Application of $su(2)$ Representation Theory	118
9.5	Easy Consequences of Root Results	119
9.6	Restrictions on Angles and Lengths of Root Vectors	120
9.7	Pictures	120
9.8	Simple Roots	123
9.9	Dynkin Diagrams	125
9.10	Dynkin Diagrams Determine All Roots	130
9.11	More	131
10	Classification of Compact Lie Algebras	134
10.1	Intermission One	134
10.2	Intermission Two	135
10.3	The Isomorphism Theorem	137
10.4	The Compact Case	139
11	$SU(n)$	143
11.1	A_n is $su(n+1)$	143
12	$SO(n)$	147
12.1	B_n is $so(2n+1)$; D_n is $so(2n)$	147
12.2	Special Cases	152
13	$Sp(n)$	153
13.1	Symplectic Maps	153
13.2	Lie Algebra of $Sp(n)$	155
13.3	$sp(n)$ is C_n	156
13.4	C_n Interpreted Using Quaternions	158
14	Irreducible Complex Representations	162
14.1	Weights	162
14.2	Highest Weight	165
15	Kaleidoscopes	169
15.1	Aspirations	169
15.2	The Root Lattice and the Weight Lattice	171
15.3	Lattices for $su(2)$ and $su(3)$	175
15.4	Root, Weight, Fundamental Region	182
15.5	Weyl Groups as Reflection Groups	186
15.6	Weyl Chambers	188
16	Conjugacy and the Weyl Group	191
16.1	The Main Theorem	191

16.2	Reflections in $N(T)/T$	191
16.3	$N(T)/T$ Is Simply Transitive on Weyl Chambers	193
17	Stiefel Diagrams and Torus Lattices	195
17.1	Regular Elements in G	195
17.2	The Stiefel Diagram	197
17.3	$Aut_0(\mathcal{G})$	200
17.4	The Lattice Associated With $Aut_0(\mathcal{G})$	202
17.5	Tricky Points	203
17.6	Torus Lattices in General	204
17.7	Looking Backward	207
17.8	The Affine Weyl Group	209
18	$SU(2)$ and $SU(3)$	213
18.1	Structure of Characters	213
18.2	$SU(2)$	214
18.3	$SU(3)$	215
18.4	The Eightfold Way, Part 1	227
18.5	The Eightfold Way, Part 2	230
18.6	The Eight Fold Way, Part 3	232
18.7	The Eight Fold Way, Part 4	235
19	The Peter-Weyl Theorem	237
19.1	Getting Our Hands on Representations	238
19.2	The Weierstrass Approximation Theorem	239
19.3	The Stone-Weierstrass Theorem	242
19.4	Hilbert Space	246
19.5	Completion of Inner Product Spaces	248
19.6	Elementary Results in Hilbert Space Theory	250
19.7	Hilbert Space Bases	254
19.8	Application To Fourier Series	256
19.9	Integral Operators	258
19.10	Compact Operators	260
19.11	Equiuniform Continuity	264
19.12	Compact Operators in $L^2(G)$	268
19.13	The Peter-Weyl Theorem	269
19.14	A Basis for $L^2(G)$	272
19.15	A Basis for the Space of Class Functions	274
19.16	Matrix Groups	277
19.17	Hilbert Space Representations	278

20 Weyl's Integral Formula	280
20.1 The Weyl Integral Formula	280
20.2 The Case $SU(2)$	281
20.3 Proof of the Weyl Integral Formula, Part 1	283
20.4 Proof of the Weyl Integral Formula, Part 2	284
20.5 Proof of the Weyl Integral Formula, Part 3	285
20.6 Proof of the Weyl Integral Formula, Part 4	288
20.7 Intermission	290
20.8 Completion of the Proof:	291
21 The Weyl Character Formula	295
21.1 Alternating Functions on $\mathcal{T}$	295
21.2 A Formula for δ	299
21.3 The Weyl Character Formula	301
21.4 The Weyl Dimension Formula	304
21.5 Applications of the Dimension Formula	305
21.6 Dimensions of Irreducible Representations of G_2	311
21.7 The Weyl Character Formula and $SU(3)$	313
21.8 A Preliminary Computer Program	319
21.9 A Program To Compute Characters of Representations of $SU(3)$	328
21.10 The Code to Compute Characters of $SU(3)$	333
21.11 The Code to Draw Weight Diagrams	338
22 Characters of $SO(5)$	339
22.1 Introduction	339
22.2 Initial Steps for $SO(5)$	339
22.3 Examples	342
23 Characters for G_2	348
23.1 Introduction	348
23.2 Other Software	355
24 The Symmetric Group	356
24.1 Introduction	356
24.2 Permutation Groups	357
24.3 The Sign of a Permutation	363
25 Irreducible Representations of $\mathcal{S}_n$	366
25.1 From Diagram to Representation	366
25.2 A Preliminary Observation	368
25.3 Key Lemmas	370
25.4 Irreducibility	371

25.5	Replacing the Canonical Tableau with Another Tableau	372
25.6	More Lemmas	373
25.7	Different Diagrams Give Inequivalent Representations	374
25.8	A Formula for $c_\lambda \cdot c_\lambda$	375
25.9	Interchanging a_λ and b_λ	375
26	Specht Modules	376
26.1	Introduction	376
26.2	Specht Modules	376
26.3	A Lower Bound for the Dimension of Representations	381
27	The Robinson-Schensted Correspondence	383
27.1	History of the Robinson-Schensted Correspondence	383
27.2	Proof of the Robinson-Schensted Correspondence	385
28	The Hook Formula	391
28.1	The Hook Formula and the Dimension of V_λ	391
28.2	Frightening Exercises	396
29	Examples	397
29.1	$\mathcal{S}_3$	397
29.2	$\mathcal{S}_4$ and $\mathcal{S}_5$	399
29.3	The Tetrahedron	400
29.4	Remaining Cases for $\mathcal{S}_5$	402
29.5	Real and Complex Representations of $\mathcal{S}_n$	403
30	Induced Representations	404
30.1	Induced Representations	404
30.2	Subgroups $H \subset G$ of Index Two	408
30.3	Irreducible Representations of the Alternating Group	410
30.4	The Icosahedron and Dodecahedron	412
31	$SU(n)$	416
31.1	The Setup	416
31.2	Example: The Case of Symmetric Tensors	417
31.3	Other Examples	418
31.4	Irreducible Representations of $\mathcal{S}_k$ and the Structure of W	419
31.5	$\text{Hom}_{\mathcal{S}_k}[V_1 \otimes \dots \otimes V_m, V_1 \otimes \dots \otimes V_m]$	422
31.6	A Mental Picture of All of This	423
31.7	A Theorem of Schur and Weyl	424
31.8	Consequences of the Schur-Weyl Theorem for $GL(n, C)$	427
31.9	Schur-Weyl for $SL(n, C)$, $GL(n, R)$, $SL(n, R)$, $U(n)$, and $SU(n)$	427

31.10	A Basis for the Image of C_λ	429
31.11	A Calculation Given Without Proof	431
31.12	Highest Weights	432
31.13	Examples of Irreducible Representations of $SU(n)$	437
31.14	Weyl's Unitary Trick	438
31.15	Representations of $GL(n, R)$ and $GL(n, C)$	440
31.16	That Pesky Assumption	442

Chapter 1

Preliminaries

1.1 Introduction

These notes are about Lie groups which are compact, and about the irreducible representations of these groups. In our sketch of the results, we will assume that the groups are simply connected, although the notes cover the general case.

As it happens, these groups have been completely classified. Each such group can be written

$$G_1 \times G_2 \times \dots \times G_k$$

where the G_i are compact and simply connected and cannot be further decomposed. The G_i are unique up to order. Each G_i is either the set of real isometries of R^n or the set of complex isometries of C^n , or the set of quaternionic isometries of H^n , or one of five exceptional groups G_2, F_4, E_6, E_7 , or E_8 .

We have to be slightly cautious. Denote the group of real isometries of R^n by $O(n)$, the group of complex isometries of C^n by $U(n)$, and the group of quaternionic isometries of H^n by $Sp(n)$. We call these the *orthogonal group*, the *unitary group*, and the *symplectic group*. In the first two cases (but not the third) we can define a determinant $D : O(n) \rightarrow R$ and $D : U(n) \rightarrow C$. In the classification, we must replace G by the kernel of this mapping. Indeed $O(n)$ is not connected, but the kernel $SO(n)$ is. $U(n)$ is not simply connected because its fundamental group is Z , but the kernel $SU(n)$ is simply connected.

We must also “doctor” our groups by replacing each G by its universal cover. The groups $SU(n)$ and $Sp(n)$ are simply connected, but $SO(n)$ has fundamental group Z_2 for $n \geq 3$, so we must use its 2-fold cover, $Spin(n)$.

The most elementary compact group is S^1 , but its universal cover is not compact. So while this group will play a big role in our notes, it is ignored in this introduction.

Another minor complication is that there are some low dimensional isomorphisms. We have $Sp(1) \cong SU(2) \cong Spin(3)$. Also $SO(4) \sim SO(3) \times SO(3)$ in the sense that their universal covers are isomorphic.

So the G_i are $SU(n)$ for $n \geq 2$, $Sp(n)$ for $n \geq 2$, $Spin(n)$ for $n \geq 5$, G_2, F_4, E_6, E_7, E_8 .

1.2 Killing's 1890 Paper

Wilhelm Karl Joseph Killing was born in 1847. He wrote his dissertation under Weierstrass and Kummer at Berlin in 1872, and then became a professor at the seminary college Collegium Hosianum in Braunsberg, Prussia. To get this position, he took holy orders. He eventually became rector of the college and chair of the town council before taking up a position at the University of Munster, where he had been an undergraduate, in 1892.

Killing invented Lie algebras independently of Sophus Lie. The University library at Collegium Hosianum did not subscribe to the journal where Lie published his results, so Killing was unaware of Lie's work. But Lie became aware of Killing and wrote of Killing's central papers: "the results in these papers are of two sorts; some are due to me, and the others are wrong."

In a series of four papers which appeared in 1888, 1888, 1889, and 1890, Killing classified all simple Lie algebras over the complex numbers. There are gaps in his arguments, and Killing was extremely modest about his work. That may have been because Killing had originally hoped to classify all real Lie algebras. For whatever reason, he decided not to publish his results until Engels convinced him otherwise. After the 1890 paper, he returned to geometry, and never wrote more about Lie algebras.

Because of his modesty, Killing would probably have been astonished had he known that at the 100th anniversary of these papers, A. John Coleman would publish an article in the *Mathematical Intelligencer*, vol. 11, no. 3, in which he claimed that Killing's paper was "The Most Important Mathematical Research Paper Ever Published." But there is justification in this judgment. Several ideas were involved in the classification, and each seems to have generated a new field of research.

Shortly after Killing published these papers, a young Elie Cartan in France learned of his results and wrote his PhD thesis refining Killing's proof and filling in gaps in the argument. In particular, Killing's results implied that there are five exceptional algebras, but it is Cartan who first proved their existence. Cartan became one of the great mathematicians of the twentieth century and without a doubt its greatest geometer, and much of his work is an extension of the work on Killing's paper.

The theorem proved by Killing states that there are four infinite families of simple Lie algebras over the complex numbers, which he called A_n, B_n, C_n, D_n , and five exceptional

simple algebras which he called G_2, F_4, E_6, E_7, E_8 . But the connection of this result to compact Lie groups came later.

1.3 Structure of Lie Algebra of a Compact Group

If G is a compact Lie group, it has a Riemannian metric which is left and right invariant. To get a left invariant metric, select an arbitrary inner product on T_e and extend it to all of G by left translation. To get a right invariant metric, extend the inner product on T_e by right translation. To obtain a metric which is both left and right invariant, the initial inner product we select on T_e must have a special property: if we translate T_e to T_g by left translation, and then back to T_e by right translation, the resulting map must be an isometry. But this map equals $Ad(g) : T_e \rightarrow T_e$. So we must choose an Ad invariant inner product on T_e , and this is possible by selecting an arbitrary inner product and then averaging over G via $\int_G \langle Ad(g)X, Ad(g)Y \rangle$.

Recall that the derivative of $Ad(g)$ in the direction X is $ad(X)$. So differentiating the equation

$$\langle Ad(g)Y, Ad(g)Z \rangle = \langle Y, Z \rangle$$

in the direction X gives $\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0$. It follows that the Lie algebra $\mathfrak{g}$ of G has an inner product satisfying this equation.

Let $\mathcal{C} \subset \mathfrak{g}$ be the center of $\mathfrak{g}$, that is, the set of all Y such that $[X, Y] = 0$ for all X . It is easy to prove that $\mathfrak{g} = \mathcal{C} \oplus \mathcal{C}^\perp$ where both of these are ideals. The first is abelian, and therefore the Lie algebra of R^n and T^n , where T is the torus. In particular, it is the Lie algebra of a compact Lie group.

We claim that the Killing form of $\mathcal{C}^\perp$ is negative-definite, and essentially equal to the negative of the invariant metric $\langle, \rangle$. Recall that by definition the Killing form of a Lie algebra is $K(X, Y) = \text{tr}(ad(X)ad(Y))$. To prove the claim, let $X_1, \dots, X_k$ be an orthonormal basis of $\mathcal{C}^\perp$. Then

$$\langle ad(X)ad(X)X_i, X_i \rangle = - \langle ad(X)X_i, ad(X)X_i \rangle = - \langle [X, X_i], [X, X_i] \rangle$$

and so $\text{tr}(ad(X)ad(X)) = - \sum_i ||[X, X_i]||^2$. This expression is less than or equal to zero. If it is zero, then $[X, X_i] = 0$ for all basis vectors, and thus X is in both $\mathcal{C}$ and $\mathcal{C}^\perp$ and so zero.

Thus we have proved that the Lie algebra of a compact Lie group must be $R^n \oplus \mathfrak{g}$ where the Killing form on $\mathfrak{g}$ is negative-definite. Conversely as we will see later, it is easy to prove that every such Lie algebra is the Lie algebra of a compact Lie group.

1.4 Compact Groups and Killing's Classification

If L is a real Lie algebra, we can form a complex Lie algebra $L \otimes C$; to do this, select a basis $X_1, \dots, X_n$ for L and let $L \otimes C$ be the set of linear combinations of these vectors with complex coefficients. If we do this with g obtained above, we get a complex Lie algebra whose Killing form is non-degenerate in the sense that if $X \neq 0$, there is a Y such that $K(X, Y) \neq 0$. But an important step in Killing's classification states that a complex Lie algebra is semisimple (that is, a sum of simple algebras) if and only if its Killing form is non-degenerate. So $g \otimes C$ must be an algebra on Killing's list.

But more is true. The standard classification of complex simple algebras produces a specific basis for each algebra; all structure constants for this basis are real. Using that basis, it is possible to construct a real form of the algebra with negative definite Killing form. So every algebra on Killing's list comes from at least one g .

There is a possible complication. It can happen that two different real Lie algebras L_1 and L_2 lead to isomorphic complex algebras $L_1 \oplus C$ and $L_2 \oplus C$. For instance, $so(3)$ is the Lie algebra of $SO(3)$, which is compact, and $sl(2, R)$ is the Lie algebra of $SL(2, R)$, which is not compact, but $so(3) \otimes C$ and $sl(2, R) \otimes C$ are isomorphic.

However, it can be proved that the real forms of a complex simple Lie algebra with negative-definite Killing form are isomorphic to each other. It follows that simple real g with negative-definite Killing form are in one-to-one correspondence with Killing's $A_n, B_n, C_n, D_n, G_2, F_4, E_6, E_7, E_8$.

In these notes, we will classify the Lie algebras of compact Lie groups directly, without referring to Killing's classification. The beginning steps of the two approaches are different. Toward the end we will be repeating arguments which appear in the complex case and in particular, both approaches produce the same root spaces and the same Dynkin diagrams.

1.5 Irreducible Representations of Compact Lie Groups

In 1930, Hermann Weyl wrote an amazing series of papers in which he extended the theory of group representations from finite groups to compact groups. In these papers, Weyl completely classified all irreducible representations of the compact Lie groups. We will obtain this classification in these notes.

Chapter 2

A Short Course in the Representation Theory of Finite Groups

2.1 Introduction

Let V be a finite dimensional complex vector space. Define $GL(V)$ to be the set of invertible linear transformations from V to V .

Definition 1 *Let G be a finite group. A group representation ρ on V is a group homomorphism $\rho : G \rightarrow GL(V)$.*

Remark: Once we pick a basis for V , each element of $GL(V)$ becomes an $n \times n$ complex matrix. A group representation assigns such a matrix $\rho(g)$ to each $g \in G$, such that group elements and matrices multiply the same way.

Remark: The matrices just mentioned depend on the basis. If we choose a new basis, we get a new set of matrices and a new representation. But in some sense, these representations are really the same. On the matrix level, a new basis corresponds to an invertible matrix A , and the new matrices have the form $A\rho(g)A^{-1}$.

Definition 2 *Two representations ρ of G on V and ψ of G on W are equivalent if there is a linear isomorphism $A : V \rightarrow W$ such that $\psi(g) = A\rho(g)A^{-1}$ for each $g \in G$.*

Definition 3 *Let $V = V_1 \oplus V_2$ and let $\rho : G \rightarrow GL(V_1)$ and $\psi : G \rightarrow GL(V_2)$ be representations of G on V_1 and V_2 . Then $\rho \oplus \psi$ is the representation of G on V given by $(\rho \oplus \psi)(g) = \rho(g) \oplus \psi(g)$.*

Remark: Suppose we have a basis for V_1 and a basis for V_2 . Then representations are essentially matrices and the sum operation amounts to a way to construct new representations once some are known by the natural operation show below:

$$\begin{pmatrix} (\rho) & 0 \\ 0 & (\psi) \end{pmatrix}$$

We are going to show that every representation of a finite G is equivalent to a direct sum of smaller pieces which cannot be further reduced. These pieces are called *irreducible representations* and the theory reduces to finding them.

But there is a problem. If a representation is a direct sum, and then we change the basis, the new matrices will usually not have the simple structure illustrated above. So it is difficult to recognize direct sums. A simple example makes this problem clear.

Example: Let us find all representations of the group with two elements Z_2 . We must map 0 to the identity matrix, and 1 to a matrix A whose square is the identity. So our theory reduces to finding all matrices A with $A^2 = I$.

It is possible to try this by direct calculation, but we wind up with a giant mess. There are *lots* of possibilities, even in the 2×2 case. For example, it is easy to check that the following matrix has square I :

$$\begin{pmatrix} 11 & -4 \\ 30 & -11 \end{pmatrix}$$

How did I find this example? I started with an easy matrix with square I , namely $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and then changed coordinates using a matrix with integer coefficients carefully chosen to have determinant 1 so its inverse would also have integer coefficients:

$$\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} 11 & -4 \\ 30 & -11 \end{pmatrix}$$

But perhaps we should work in an even more elementary way. Start with $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and write down the equations on a, b, c, d making $A^2 = I$. This leads to a whole world of trouble.

But if we proceed more abstractly, we easily find all representations of Z_2 . Suppose $A : V \rightarrow V$ has square I . Define

$$V_1 = \{v \in V \mid Av = v\} \qquad V_2 = \{v \in V \mid Av = -v\}$$

The intersection of these subspaces is clearly $\{0\}$. Their union is all of V because for any $v \in V$ we can write

$$v = \frac{v + A(v)}{2} + \frac{v - A(v)}{2}$$

So $V = V_1 \oplus V_2$ and $A = I$ on V_1 and $A = -1$ on V_2 . If we pick bases for V_1 and V_2 , the matrix of A becomes

$$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & -1 & \\ & & & & \ddots \\ & & & & & -1 \end{pmatrix}$$

It follows that every representation of Z_2 is equivalent to a direct sum of just two irreducible representations, χ_0 and χ_1 . Here $\chi_0(g) = 1$ for all g , while $\chi_2(g) = 1$ when $g = 0$ and -1 when $g = 1$.

2.2 Decomposing Representations into Sums of Irreducible Representations

Definition 4 A representation ρ of G on V is indecomposable if it is not possible to write $V = V_1 \oplus V_2$ with both subspaces non-zero and $\rho(V_1) \subseteq V_1, \rho(V_2) \subseteq V_2$.

Definition 5 A representation ρ of G on V is irreducible if it is not possible to find a non-trivial subspace $V_1 \subset V$ such that $\rho(V_1) \subseteq V_1$.

Remark: Notice that irreducible matrices are certainly indecomposable. In matrix language, a representation is indecomposable if it cannot be put in the following form by changing bases:

$$\begin{pmatrix} (\rho) & 0 \\ 0 & (\psi) \end{pmatrix}$$

It is irreducible if it cannot be put in the following form by changing bases:

$$\begin{pmatrix} (\rho) & 0 \\ (\star) & (\psi) \end{pmatrix}$$

The following pair of theorems explain why we introduce both terms.

Theorem 1 Every representation is equivalent to a representation which is a direct sum of indecomposable representations.

Proof: By induction on the dimension of V . In the proof of the induction step, we are done if V is indecomposable. Otherwise we can find non-trivial V_1 and V_2 with $V = V_1 \oplus V_2$

and each invariant under ρ . By induction, each can be decomposed into indecomposable spaces. QED

Theorem 2 Suppose ρ on V is the direct sum $\rho_1 \oplus \dots \oplus \rho_k$ on $V_1 \oplus \dots \oplus V_k$ where each ρ_i on V_i is irreducible. Then the decomposition is unique in the sense that if $\rho = \psi_1 \oplus \dots \oplus \psi_m$ on $W_1 \oplus \dots \oplus W_m$ and the ψ_j are irreducible, then $k = m$ and after applying a permutation to the ψ_i and W_i , we find that ρ_i and ψ_i are equivalent.

Proof: Consider the map $T_{ij} : V_i \rightarrow V_1 \oplus \dots \oplus V_k = W_1 \oplus \dots \oplus W_m \rightarrow W_j$. Here the first map is inclusion and the last map is projection. Notice that $T_{ij} \circ \rho_i(g) = \psi_j(g) \circ T_{ij}$ for all $g \in G$.

Definition: In general, if ρ is a representation of G on V and ψ is a representation of G on W , a linear transformation $T : V \rightarrow W$ is called an *intertwining operator* if $T \circ \rho(g) = \psi(g) \circ T$ for all $g \in G$.

Lemma 1 (Schur's Lemma) If V and W are irreducible representations and $T : V \rightarrow W$ is an intertwining operator, then either T is identically zero or else T is an equivalence between the two representations.

Proof: An easy calculation shows that $\text{Ker}(T) \subset V$ and $\text{Im}(T) \subset W$ are invariant subspaces. Hence $\text{Ker}(T) = (0)$ or $\text{Ker}(T) = V$ and $\text{Im}(T) = (0)$ or $\text{Im}(T) = W$. The lemma follows immediately. QED.

Proof of the theorem. By the lemma, each T_{ij} is either zero or else an isomorphism. Consider the first irreducible representation ρ_1 in V . Rearrange the other representations so all representations equivalent to ρ_1 come first, and let $V_1 \subset V$ be the sum of these representation spaces. Let V_2 be the sum of all other representation spaces. Find all representations ψ_i in W that are equivalent to ρ_1 and rearrange the representations of W so that these ψ_i come first. Let $W_1 \subset W$ be the sum of these representation spaces and let W_2 be the sum of the remaining spaces. If there are no ψ_i equivalent to ρ_1 , then W_1 will be zero.

Notice that $V = V_1 \oplus V_2$ and $W = W_1 \oplus W_2$.

The T_{ij} may be rearranged by these steps, but remain the same maps. Notice that each T_{ij} either maps V_1 to W_1 and is zero everywhere else, or maps V_2 to W_2 and is zero everywhere else. Since T is an isomorphism, the sum of the T_{ij} mapping V_1 to W_1 must be an isomorphism, so these spaces have the same dimension and therefore each contains the same number of representations equivalent to ρ_1 .

Now restrict to V_2 and W_2 and continue. QED.

Remark: If we can prove that the irreducible representations of G are exactly the same as the indecomposable representations of G , then existence and uniqueness of decompositions

of representations into irreducible pieces will follow. This is the subject of the next section of this short course.

2.3 Integration and the Fundamental Theorem of Finite Group Representation Theory

Definition 6 Suppose G is a finite group. Then $L(G)$ is the set of all complex valued functions on G . This notation is chosen to remind users of L_1 or L_2 in analysis.

If $f \in L(G)$, define

$$\int f(g) dg = \frac{1}{|G|} \sum_{g \in G} f(g)$$

Definition 7 Let f be a function on G and $g_1 \in G$. Then $L_{g_1}f$, called the left translation of f by g_1 , is defined by

$$(L_{g_1}f(g)) = f(g_1^{-1}g)$$

Remark: This formula contains g_1^{-1} for the same reason that in high school algebra we shift a function right by D by subtracting D from the variable.

Remark: Similarly, we define right translation by $R_{g_1}f$ at g is $f(gg_1)$.

Theorem 3 The integral defined above is left-invariant and right-invariant in the sense that

$$\int L_{g_1}f dg = \int f dg = \int R_{g_1}f dg$$

Theorem 4 Let G be a finite group. Then every representation of G can be written as a sum of irreducible representations, and these irreducible summands are unique up to order and equivalence .

Proof: This theorem is an immediate consequence of the results in the previous paragraph and the very important

Theorem 5 If G is a finite group and representations are defined for vector spaces over the reals, the complex numbers, or the quaternions, then a representation is irreducible if and only if it is decomposable.

Proof: The idea of the proof is very simple. We will define an inner product $\langle \dots \rangle$ on the vector space such that $\langle \rho(g)v, \rho(g)w \rangle = \langle v, w \rangle$ for all $g \in G$ and $v, w \in V$. Suppose a representation has an invariant subspace $W \subset V$. Write $V = W \oplus W^\perp$ using the inner product. Then $W^\perp$ is also invariant because if $w \in W^\perp$ and $v \in W$, then

$$\langle v, \rho(g)w \rangle = \langle \rho(g^{-1})v, \rho(g^{-1})\rho(g)w \rangle = \langle \rho(g^{-1})v, w \rangle = 0$$

Next we prove that $\langle \dots \rangle$ exists. Begin with choosing an arbitrary inner product $\langle \dots \rangle$ on V . Recall that $\langle v, w \rangle = \langle w, v \rangle$ in the real case, and $\langle v, w \rangle = \overline{\langle w, v \rangle}$ in the complex and quaternionic cases. Now define a new inner product $\langle\langle v, w \rangle\rangle$ by

$$\langle\langle v, w \rangle\rangle = \int \langle \rho(g)v, \rho(g)w \rangle dg$$

This works because

$$\langle\langle \rho(g_1)v, \rho(g_1)w \rangle\rangle = \int \langle \rho(g)\rho(g_1)v, \rho(g)\rho(g_1)w \rangle = \int \langle \rho(gg_1)v, \rho(gg_1)w \rangle$$

and by right invariance, this equals

$$\int \langle \rho(g)v, \rho(g)w \rangle = \langle\langle v, w \rangle\rangle$$

2.4 Intermission: $U(n) \subset GL(n, \mathbb{C})$

Recall that $U(n)$ is the set of all complex matrices U satisfying $\overline{U}^T U = I$. This is the group of all linear transformations preserving the standard Hermitian inner product on $\mathbb{C}^n$.

Recall also that a matrix H is Hermitian if $\overline{H}^T = H$. A fundamental theorem of linear algebra states that if H is Hermitian, there is an orthonormal basis consisting of eigenvectors of H , and all eigenvalues are real.

Theorem 6 *Every element A of $GL(n, \mathbb{C})$ can be written uniquely as UH where $U \in U(n)$ and H is a Hermitian matrix with positive real eigenvalues.*

Proof: It took imagination to guess that this theorem might be true. But once it is written down, the theorem essentially suggests its proof. Indeed if the result is true, then

$$\overline{A}^T A = \overline{H}^T \overline{U}^T U H = H^2$$

We now attempt to find such an H . Notice that $\overline{A}^T A$ is Hermitian, so an orthonormal basis exists making it diagonal. If v is an eigenvector with eigenvalue λ , then

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle \overline{A}^T A v, v \rangle = \langle A v, A v \rangle = \|A v\|^2$$

It follows that eigenvalues are real and non-negative. Since $A \in GL(n, \mathbb{C})$, it is nonsingular and eigenvalues are real and positive. Using the same orthonormal basis, and replacing each λ by its positive square root, we obtain a non-singular Hermitian matrix H and $\overline{A}^T A = H^2$.

Write $U = AH^{-1}$. Then $A = UH$ and $\bar{U}^T U = (\bar{H}^T)^{-1} \bar{A}^T AH^{-1} = HH^2H^{-1} = I$. This proves existence of the decomposition; to finish we must prove the representation unique. The first line of the proof shows that H^2 is unique; if H is unique, then U is unique and we are done.

There is something to prove, because matrices can have many square roots. But in our case, H^2 is Hermitian and so diagonalizable. Write the underlying vector space V as $V_1 \oplus \dots \oplus V_k$ where H^2 is constant on each V_i and these constants are different. The constants are the distinct eigenvalues of H^2 , all positive. Then H is also constant on each V_i and these constants are the positive square roots of the eigenvalues of H^2 , so all distinct. Therefore the matrix for H in this representation of V is unique.

Remark: Let $\mathcal{H}$ be the set of $n \times n$ Hermitian matrices, not necessarily nonsingular. The set of such matrices is a linear subspace of C^{n^2} and thus contractible. The exponential map given by $H \rightarrow I + H + \frac{1}{2!}H^2 + \dots$ sends $\mathcal{H}$ diffeomorphically onto the set of Hermitian matrices with positive eigenvalues, so this set is also contractible. It follows that $U(n)$ and $GL(n, C)$ have the same homotopy type, so all topological invariants of $GL(n, C)$ can be determined by working with the compact group $U(n)$.

It has been proved that every connected Lie group G has a maximal compact subgroup, which is unique up to conjugacy (Cartan-Iwasawa-Malcev theorem). It is also true that G is a product $K \times R^k$ of such a compact subgroup K with a Euclidean space (this is not a group product, but only a topological product). So the algebraic topology of Lie groups in general is essentially the same as the algebraic topology of compact Lie groups. This is one reason that the study of compact Lie groups is central in mathematics.

2.5 Unitary Equivalence of Unitary Representations

According to the principal theorem of section 2.3, if ρ is a complex representation of a finite group G on V , then there is an inner product on V such that each $\rho(g)$ preserves the inner product. Once we have an inner product, we can find an orthonormal basis and thus assume that we have the standard inner product. Then each ρ is a unitary map, i.e., $\langle \rho(g)v, \rho(g)w \rangle = \langle v, w \rangle$ for all $v, w \in V$. This is equivalent to saying that each matrix $U = \rho(g)$ is unitary, so $\bar{U}^T U = I$.

Theorem 7 *Suppose ρ and ψ are unitary representations on V and W . If ρ and ψ are equivalent, then they are equivalent via a unitary transformation. In other words, without loss of generality, we can assume that equivalences preserve the inner product.*

Proof: Suppose $T : V \rightarrow W$ is an isomorphism which satisfies $T\rho(g) = \psi(g)T$. Using the result of the previous section, write $T = UH$, so $UH\rho(g) = \psi(g)UH$. Suppose $H\rho(g) = \rho(g)H$ for all g . Then $U\rho(g)H = \psi(g)UH$; multiplying by H^{-1} , we obtain $U\rho(g) = \psi(g)U$

so the representations are unitarily equivalent.

Notice that $T\rho(g^{-1}) = \psi(g^{-1})T$. From now on, write $\overline{A}^T = A^*$ for any matrix A . If ρ and ψ are unitary representations, this gives $T\rho(g)^* = \psi(g)^*T$ and taking adjoints, $\rho(g)T^* = T^*\psi(g)$. Thus $\rho(g)T^*T = T^*\psi(g)T = T^*T\rho(g)$. But $T^*T = H^2$, so $\rho(g)H^2 = H^2\rho(g)$.

The final step of the argument is easy, but a little strange. Find a polynomial with real coefficients such that $P(\lambda_i) = \sqrt{\lambda_i}$ for all eigenvalues of $T^*T = H^2$. Since T^*T commutes with all $\rho(g)$, so does $P(T^*T)$. Hence $P(H^2)$ commutes with all $\rho(g)$. If we were to change the basis, then $\rho(g)$ and T^*T would change, but $P(T^*T)$ would change in the same way, and thus still commute with all $\rho(g)$. But one such basis change makes H^2 diagonal. In that case $P(H^2)$ just acts on the diagonal elements, converting them to the diagonal elements for H . So H also commutes with all $\rho(g)$. QED.

2.6 Irreducible Representations of Abelian Groups

All results so far hold for vector spaces over the reals, the complex numbers, and the quaternions. But a deeper form of Schur's lemma, proved below, requires complex numbers, and all our results until section 2.11 assume we are over the complex numbers.

Lemma 2 (Schur) *Let ρ be an irreducible representation of G on a complex vector space, and let T be an intertwining operator of V with itself. Then there is a constant λ such that $T(v) = \lambda v$ for all $v \in V$.*

Proof: Let λ be an eigenvalue for an intertwining operator T . Then $T - \lambda$ is also an intertwining operator, and it is not an isomorphism because it maps the nonzero eigenvector to zero. So the weak form of Schur's lemma says that $T - \lambda$ is zero. QED.

Theorem 8 *Every irreducible complex representation of an abelian group is one dimensional.*

Proof: Since G is abelian, $\rho(gg_1) = \rho(g)\rho(g_1) = \rho(g_1)\rho(g)$. So each $\rho(g)$ is an intertwining operator and thus $\rho(g) = \lambda(g)I$. This is only irreducible if the dimension of I is one. QED.

Remark: Notice that a one dimensional representation of G is the same thing as a group homomorphism $\varphi : G \rightarrow C^*$. We claim the image is in S^1 . If $|\varphi(g)| > 1$, then $|\varphi(g^k)|$ grows monotonically with k , which is impossible since it is periodic. Similarly $|\varphi(g)| < 1$ is impossible.

Remark: If $G = G_1 \times \dots \times G_k$ and $\varphi : G \rightarrow S^1$ is a group homomorphism, let φ_i be φ restricted to G_i . Then clearly every one dimensional representation has the form

$$g_1 \times \dots \times g_k \rightarrow \varphi_1(g_1) \cdot \dots \cdot \varphi_k(g_k)$$

where $\varphi_i : G_i \rightarrow S^1$. Since every finite abelian group has the form

$$Z_{n_1} \times \dots \times Z_{n_k}$$

the irreducible representations of all finite abelian groups will be known as soon as we find all homomorphisms $\varphi : Z_n \rightarrow S^1$. But it is easy to see that these are exactly

$$\varphi_k(g) = e^{\frac{2\pi i}{n} kg} \quad 0 \leq k < n$$

2.7 Orthogonality Relations

Remark: From now on, whenever we have a representation ρ on a vector space V , we assume that V has an invariant inner product. When we write a representation in matrix form, we assume we have chosen an orthonormal basis for V . Then each representation preserves this inner product and thus has unitary matrices. So $\overline{\rho(g)}^T \rho(g) = I$, or equivalently $\rho(g^{-1}) = \overline{\rho(g)}^T$.

Theorem 9 *Let $(a_{ij}(g))$ be matrices for an irreducible representation of G . Then*

$$\int a_{ij}(g) \overline{a_{kl}(g)} dg = \frac{1}{\dim V} \delta_{ik} \delta_{jl}$$

In other words, each component is a function in $L(G)$ of norm $\frac{1}{\sqrt{\dim V}}$, and distinct components are orthogonal.

Theorem 10 *Let $(a_{ij}(g))$ and $(b_{ij}(g))$ be matrices for two inequivalent irreducible representations of G . Then*

$$\int a_{ij}(g) \overline{b_{kl}(g)} dg = 0$$

In other words, the components of two distinct irreducible representations are orthogonal

Proof: The idea behind the proof is simple. Pick any linear transformation $T : V \rightarrow W$. By averaging over G , we can make T an intertwining operator. By Schur's lemma, there are sharp restrictions on intertwining operators. The theorem follows easily. Here are the details.

Let ρ on V and τ on W be irreducible unitary representations, which are either equal or else inequivalent. Let $E : V \rightarrow W$ be a transformation E_{ij} , a matrix with 1 at the ij th entry and zero elsewhere. Consider

$$T = \int \rho(g) E_{ij} \tau(g^{-1}) dg$$

Notice that for each fixed g_1 we have

$$T = \int \rho(g_1 g) E_{ij} \tau((g_1 g)^{-1}) dg = \rho(g_1) T \tau(g_1^{-1})$$

or equivalently $\rho(g_1)T = T \tau(g_1)$. So T is an intertwining operator and by Schur's lemma, $T = 0$ if ρ and τ are inequivalent, and $T = \lambda I$ for some constant λ if $\tau = \rho$.

Consider the inequivalent case first. In this case the mn entry of T is zero, so

$$\int \rho(g)_{mi} \tau(g^{-1})_{jn} dg = \int \rho(g)_{mi} \overline{\tau(g)}_{nj} dg = 0$$

This proves the second theorem.

Now consider the first case. If $m \neq n$, the mn entry of T is zero, so

$$\int \rho(g)_{mi} \rho(g^{-1})_{jn} dg = \int \rho(g)_{mi} \overline{\rho(g)}_{nj} dg = 0$$

This proves most of the first theorem, except the situation when $m = n$ and i and j may or may not be equal.

Fix $m = n$. Then for fixed i and j , all terms below are equal:

$$\int \rho(g)_{mi} \rho(g^{-1})_{jm} dg$$

Write these functions in reverse order:

$$\int \rho(g^{-1})_{jm} \rho(g)_{mi} dg = \lambda$$

Sum these equal terms over m , getting

$$\int \rho(g^{-1}g)_{ji} dg = (\dim V)\lambda = \int \rho(id)_{ij} dg = \int \delta_{ij} dg = \delta_{ij}$$

So $(\dim V)\lambda = \delta_{ij}$ and $\lambda = \frac{\delta_{ij}}{\dim V}$.

Therefore

$$\int \rho(g)_{mi} \rho(g^{-1})_{jm} dg = \int \rho(g)_{mi} \overline{\rho(g)}_{mj} dg = \frac{\delta_{ij}}{\dim V}$$

QED.

Corollary 1 Suppose G has inequivalent irreducible representations of dimensions $d_1, \dots, d_k$. Then

$$\sum d_i^2 \leq |G|$$

Corollary 2 *A finite group of order $|G|$ has at most $|G|$ inequivalent irreducible representations. The maximum value occurs if and only if G is abelian.*

Proof: The vector space $L(G)$, which clearly has dimension $|G|$, has at least $\sum d_i^2$ orthogonal vectors, and orthogonal vectors are linearly independent.

If G has $|G|$ inequivalent representations of dimension d_i , then $\sum d_i^2$ can only equal $|G|$ if each $d_i = 1$. So each irreducible representation is one-dimensional and takes commutators to the identity. Therefore every representation takes commutators to zero.

Note that G acts on $L(G)$ by left translation, and only the identity element maps to I. Hence all commutators in G are the identity, so G is abelian. QED.

2.8 Representation Coefficients are Complete in $L(G)$

Theorem 11 *The group G acts on $L(G)$ by left translation $(L_{g_1}f)(g) = f(g_1^{-1}g)$. This representation preserves the inner product $\langle f_1, f_2 \rangle = \int f_1(g)\overline{f_2(g)} dg$.*

- *In the expression of this representation as a direct sum of irreducible representations, every irreducible representation of G occurs. Indeed, if such a representation has dimension d , it occurs d times.*
- *The set $\{a_{ij}\}$ of all coefficients of all irreducible representations up to equivalence is an orthogonal basis of $L(G)$, easily normalized.*
- *If the irreducible representations of G have dimensions $d_1, \dots, d_k$, then $\sum d_i^2 = |G|$.*

Proof: An irreducible component of L_{g_1} is represented by a matrix $(a_{ij}(g_1))$. This matrix depends on a basis $b_1, \dots, b_d$ for a representation subspace of $L(G)$, where the b_i are functions $b_i(g)$ on G . The proofs of the items in the theorem are essentially a matter of keeping the roles of g_1 and g straight.

We begin, then, by choosing a basis $b_1(g), \dots, b_d(g)$ for an irreducible subspace of $L(G)$. This basis must be orthonormal, so $\int b_i(g)\overline{b_j(g)} dg = \delta_{ij}$.

We then have two ways to write the representation by left translation:

$$\begin{aligned} L(g_1)(\sum c_j b_j(g)) &= \sum c_j b_j(g_1^{-1}g) \\ L(g_1)(\sum c_j b_j(g)) &= \sum_i \left(\sum_j a_{ij}(g_1) c_j \right) b_i(g) \end{aligned}$$

Comparing the right sides, we must have

$$b_j(g_1^{-1}g) = \sum_i a_{ij}(g_1) b_i(g)$$

Replace g_1^{-1} by g_1 and noting that $a_{ij}(g_1^{-1}) = \overline{a_{ji}(g_1)}$ because our representations are unitary:

$$b_j(g_1g) = \sum_i \overline{a_{ji}(g_1)} b_i(g)$$

Setting $g = e$, we find that

$$b_j(g_1) = \sum_i \overline{a_{ji}(g_1)} b_i(e)$$

The final formula is a clue which leads to the proof of the theorem. Suppose we fix a *column* of the representation matrix $\bar{a}$, say the i th column. The entries of this column are functions in $L(G)$: $\overline{a_{1i}}(g_1), \dots, \overline{a_{di}}(g_1)$. These functions are orthogonal in $L(G)$ and all have the same length $\frac{1}{d}$, so multiplying each by the same constant gives an orthonormal basis of a subspace of $L(G)$ that we will call V_i . The calculations above show that V_i is invariant and irreducible and the representation of G on this space is given by $a_{ij}(g)$.

If we pick a different column we obtain a different invariant subspace, and indeed a subspace orthogonal to the original by the orthogonality relations. The representation formed by any column is still given by $a_{ij}(g)$. So this representation occurs d times in the decomposition of $L(G)$ and in this way the columns generate $V_1 \oplus \dots \oplus V_d \subseteq L(G)$.

Both of these assertions are proved by our original calculation. To be sure, we started backward by trying to find the form of an invariant, irreducible subspace W with basis $b_1(g), \dots, b_d(g)$. Our calculation let us to a matrix (a_{ij}) for an irreducible representation. But this representation $a(g)$ occurs multiple times in $L(G)$, so we don't find a unique invariant subspace. Indeed, the general theory states that such a W would have the form $\{\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_d v_d \mid \lambda_i \text{ are constants and } v_i \in V_i\}$. In our final formula, these $\lambda_i = b_i(e)$. Indeed, in the special case that $b_j(g_1) = \overline{a_{ji}(g_1)}$ we have $b_j(e) = \delta_{ij}$.

We can apply the same calculations to other irreducible representations of G , which lead to basis functions orthogonal to the above functions by the orthogonality relations.

Suppose we have accounted for all irreducible representations of G by the above method, but still do not have the complete $L(G)$. Since $L(G)$ can be written as a direct sum of irreducible representations, we can find a separate irreducible invariant subspace orthogonal to all of the spaces produced above. Select an orthogonal basis for this space and apply the calculation at the start of the proof to this basis. This calculation gives irreducible representation matrices. The matrices define a representation *equivalent* to one of the representations already used. So there is a change of basis which gives one of the matrices we already used. But this change of basis replaces the b_j with linear combinations of the b_j , and so the b_j remain orthogonal to the earlier constructions. And yet they are linear combinations of the earlier column entries, a contradiction.

Example: The smallest non-abelian group is D_3 , the third dihedral group or symmetry group of an equilateral triangle. This group has an identity, two other rotations, and

three reflections. One irreducible representation is the identity character, χ_0 , mapping everything to 1. Another is the determinant character χ_1 , mapping the rotations to 1 and the reflections to -1 . A final representation is the two dimensional representation ρ given by mapping each symmetry to its 2×2 matrix. This representation is irreducible, because otherwise it would be a sum of one-dimensional representations and would thus map all rotations to the identity. So D_3 has representations χ_0, χ_1, ρ of dimensions 1, 1, 2. Notice that $1^2 + 1^2 + 2^2 = 6$.

In this particular case, the representation matrices are all real even though we applied a theorem from complex representation theory. Usually irreducible complex representations have complex entries.

2.9 Intermission: Fourier Series

Although this chapter is about finite groups, it is useful to think ahead to the Lie group case. The simplest compact Lie group is the circle S^1 . Think of this object as the set of complex numbers of absolute value one. It is a group under complex multiplication.

When we consider representations of Lie groups, we always require that $\rho : G \rightarrow U(n)$ be continuous. As in the finite theory, every irreducible representation of a compact Abelian Lie group is one dimensional. Since $U(1)$ is the circle, we study group homomorphisms $S^1 \rightarrow S^1$.

According to elementary Lie theory, such a homomorphism is determined by the corresponding homomorphism of Lie algebras: $\rho_* : R \rightarrow R$. The only such linear maps are $r \rightarrow \lambda r$ for a fixed real λ . The corresponding map $S^1 \rightarrow S^1$ sends $\theta \rightarrow \lambda\theta$. This map sends multiples of 2π to multiples of 2π if and only if λ is an integer. Consequently the irreducible representations of S^1 , i.e., group homomorphisms $S^1 \rightarrow S^1$, are $\theta \rightarrow n\theta$ or $e^{i\theta} \rightarrow e^{in\theta}$ or $z \rightarrow z^n$.

The main theorem of the previous section says that the matrix entries of the irreducible representations of a finite group form a basis for the space $L^2(G)$ of all complex-valued functions on G . A similar theorem for the circle would assert that the $e^{in\theta}$ form a basis for the set of all reasonable functions on S^1 . Thus if f is a periodic function with period 2π , we should be able to write

$$f(\theta) = \sum_{-\infty}^{\infty} c_n e^{in\theta}$$

And sure enough, that is exactly the central claim of Fourier series theory. Notice, however, that there are additional technical details: exactly what sorts of functions are allowed? what kind of convergence is intended? etc?

We will later answer these questions, not just for S^1 but for any compact Lie group, and

obtain a similar theorem in general.

But there is more. We defined an integral over a finite group,

$$\int f(g) dg = \frac{1}{|G|} \sum_{g \in G} f(g)$$

The analogue for S^1 would be

$$\int f(g) dg = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) d\theta$$

The analogue of the orthogonality relations would state that

$$\frac{1}{2\pi} \int_0^{2\pi} e^{im\theta} \overline{e^{in\theta}} d\theta = \delta_{mn}$$

and sure enough this is a central result at the start of a Fourier Series course. Using this result, we obtain a formula for the coefficients of the series $f(\theta) = \sum_{-\infty}^{\infty} c_n e^{in\theta}$:

$$c_n = \langle f, e^{in\theta} \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) \overline{e^{in\theta}} d\theta$$

One easily checks that the analogue of this formula holds in the case $L^2(G)$ for a finite group.

2.10 Group Characters

We now come to a crucial part of the theory. If ρ is a representation, we currently have no general method to determine whether it is irreducible. And if not, we currently have no general method to decompose it into its irreducible pieces. And finally, we currently have no method to determine the number of irreducible representations of a finite G . Character theory will fill in these blanks.

Let ρ be a representation of G . Then the associated character in $L(G)$ is the function

$$\chi(g) = \text{tr}(\rho(g))$$

where tr stands for the trace of a matrix.

Theorem 12

- *Equivalent representations have the same character. (The converse is also true and will be proved shortly.)*
- *Characters are constant on conjugacy classes of G .*

Proof: Since $\text{tr}(AB) = \text{tr}(BA)$, $\text{tr } M\rho(g)M^{-1} = \text{tr } \rho(g)$. Similarly

$$\text{tr } \rho(g_1 g g_1^{-1}) = \text{tr } \rho(g_1) \rho(g) \rho(g_1^{-1}) = \text{tr } \rho(g)$$

Example: Consider the dihedral group. The conjugacy classes are $A = \{e\}$, $B = \{\text{the two rotations}\}$, and $C = \{\text{the three reflections}\}$. There are also three irreducible characters, $\chi_0 = \text{the identity}$, $\chi_1 = \text{det}$, and $\rho = \text{the natural two dimensional representation}$. We can then encode the character data in a *character table*:

Class	χ_0	χ_1	ρ
A	1	1	2
B	1	1	-1
C	1	-1	0

Notice, incidentally, that the column vectors, which give the characters, are orthonormal in $L(G)$. This actually holds in general:

Theorem 13 *The characters form an orthonormal subset of $L(G)$*

Proof: If (a_{ij}) is irreducible, then $\text{tr } \rho = \sum_i a_{ii}$. So the norm of this character is

$$\int \sum_i a_{ii} \overline{\sum_j a_{jj}} dg = \int \sum_i a_{ii} \overline{a_{ii}} dg + \int \sum_{i \neq j} a_{ii} \overline{a_{jj}} dg$$

By the orthogonality relations, this equals

$$\sum_i \int a_{ii} \overline{a_{ii}} dg = \sum_i \frac{1}{\dim V} = \frac{\dim V}{\dim V} = 1$$

If (a_{ij}) and (b_{ij}) are inequivalent irreducible representations

$$\int \text{tr}(a) \overline{\text{tr}(b)} dg = \sum_{ij} \int a_{ii} \overline{b_{jj}} dg = 0$$

by the orthogonality relations.

Remark: If ρ is an arbitrary representation with character χ , then χ determines whether or not ρ is irreducible. If in addition the character table is known, χ determines the decomposition of ρ into irreducible representations. Indeed

Theorem 14

- Suppose that a representation ρ with character χ can be decomposed as a sum of distinct irreducible representations $\rho_1, \dots, \rho_k$. Suppose each ρ_i occurs d_i times in this decomposition. Then $\int \chi \overline{\chi} dg = \sum d_i^2$
- In particular, ρ is irreducible if and only if $\int \chi \overline{\chi} dg = 1$
- Suppose ρ with character χ can be decomposed into irreducible representations. Let ρ_i be irreducible, with character χ_i . Then the number of times χ_i occurs in the representation ρ is $\int \chi \overline{\chi_i} dg$
- In particular, two arbitrary representations ρ and ψ , with characters χ and ϕ , are equivalent if and only if $\chi = \phi$.

Proof: Clearly $\chi = \sum d_i \chi_i$ where χ_i is the character of ρ_i . So

$$\int \chi \overline{\chi} dg = \sum_{i,j} d_i d_j \int \chi_i \overline{\chi_j} dg = \sum_i d_i^2 \int \chi_i \overline{\chi_i} dg = \sum d_i^2$$

The second item then follows immediately.

If $\rho = \sum d_i \rho_i$, then $\chi = \sum d_i \chi_i$ and

$$\int \chi \overline{\chi_i} dg = \int \sum_j d_j \chi_j \overline{\chi_i} dg = \sum_j d_j \int \chi_j \overline{\chi_i} dg = d_i$$

The final result follows since equivalent representations have the same characters, and conversely if $\chi = \phi$, then $\int \chi \overline{\chi_i} dg = \int \phi \overline{\chi_i} dg$ so each irreducible representation occurs in ρ and ψ the same number of times, so ρ and ψ are equivalent. QED.

Remark: The results above show that if $\rho_1, \dots, \rho_k$ are the irreducible representations of G , then $\chi_1, \dots, \chi_k$ are orthonormal elements of $L(G)$ which are constant on conjugacy classes.

Theorem 15 *The characters $\chi_1, \dots, \chi_k$ form an orthonormal basis for the subspace of $L(G)$ consisting of functions which are constant on conjugacy classes.*

Corollary 3 *The number of irreducible representations of G is equal to the number of conjugacy classes in G .*

Warning: This proof does not canonically identify conjugacy classes with irreducible representations. Instead, the two sets are dual to each other. So it is a hopeless task to start with a conjugacy class and attempt to construct an irreducible representation.

Proof: If $f \in L(G)$ and ρ is a representation of G on V , define $T_f(\rho) = \int \rho(g) \overline{f(g)} dg$. This is a linear transformation $V \rightarrow V$.

Lemma 3 *If f is a class function then for any ρ , $T_f(\rho)$ is an intertwining operator for ρ .*

Proof: If f is a class function, then

$$\rho(h)T_f(\rho) = \int \rho(h)\rho(g)\overline{f(g)} dg = \int \rho(hg)\overline{f(g)} dg$$

Replacing g by $h^{-1}gh$ in the integral and using the class invariance of f , this equals

$$\int \rho(hh^{-1}gh)\overline{f(h^{-1}gh)} dg = \int \rho(g)\overline{f(g)}\rho(h) dg = T_f(\rho)\rho(h)$$

QED.

Main proof, continued: If the characters of the irreducible representations do not generate the full space of class functions in $L(G)$, then we can write this full space as the space generated by characters of irreducible representations plus its orthogonal complement. So suppose f is in this orthogonal complement.

For any irreducible ρ , $T_f(\rho)$ is an intertwining operator and thus by Schur's lemma equals λI for a constant λ . So the trace of $T_f(\rho)$ equals

$$(\dim \rho)\lambda = \int \text{tr}(\rho(g)) \overline{f(g)} dg$$

Since f is orthogonal to all traces of irreducible representations, $\lambda = 0$.

It follows that $\int \rho(g)\overline{f(g)} dg = 0$ for irreducible ρ , and thus for sums of irreducible ρ , and thus for all representations ρ . So this holds for left translation acting on $L(G)$. But the function δ which equals 1 at e and 0 elsewhere is in $L(G)$ and on this element

$$T_f(L_g)(\delta)|_g = \int (L_{g_1}\delta)(g)\overline{f(g_1)} dg_1 = 0 = \int \delta(g_1^{-1}g)\overline{f(g_1)} dg_1$$

The final answer is the zero function on G , but it's value at g_1 is $\overline{f(g_1)}$. So f is identically zero, a contradiction. QED.

Chapter 3

Real, Complex, and Quaternionic Representations

3.1 Quaternionic Representations

We have been discussing *complex* representations of a group G . But many of our results hold for *real* and *quaternionic* representations as well. In the end, we will discover that a list of all complex irreducible representations of G yields corresponding lists of all real irreducible representations and all quaternionic irreducible representations of G . This chapter explains the theory.

Since the quaternions are not commutative, special care is required when discussing their representations, and we'll provide that discussion now. Denote the quaternions by $\mathcal{H}$.

If V is a finite dimensional quaternionic vector space, we define a basis as usual. Given a basis, V is isomorphic to the set of all n -tuples $(q_1, \dots, q_n)$ where the q_i are quaternions. We scalar multiply vectors, and we will multiply them by matrices, but it turns out that we need to perform these operations on opposite sides of the vectors to make the theory work. From now on, we scalar multiply from the right, and we multiply by matrices from the left.

Therefore, define a quaternionic vector space to be an additive abelian group V with a scalar operation $(v, \lambda) \rightarrow v\lambda$ by quaternions λ , such that $v(\lambda_1\lambda_2) = (v\lambda_1)\lambda_2$. We could continue to write scalars on the left if we wanted, but then we would have the clumsy requirement $(\lambda_1\lambda_2)v = \lambda_2(\lambda_1v)$.

A *linear transformation* between quaternionic vector spaces is a map $f : V \rightarrow W$ such that $f(v_1 + v_2) = f(v_1) + f(v_2)$ and $f(v\lambda) = f(v)\lambda$.

Let us write this in matrix language. For simplicity, consider the case where $f : V \rightarrow V$ and the same basis is used for the domain and range. Let $e_1, \dots, e_n$ be a basis of such a V . If f is a linear transformation, we can write $f(e_i) = \sum e_j a_{ji}$. Let v be an arbitrary vector and write it using the basis as $v = \sum e_i q_i$. Then

$$f(v) = f\left(\sum e_i q_i\right) = \sum f(e_i) q_i = \sum e_j a_{ji} q_i$$

As usual, we can write the matrix A of f by letting the i th column be the coordinates of $f(e_i)$ and then this result becomes the standard matrix product

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$$

So, as predicted, writing scalars on the right forces us to multiply matrices from the left. The reader can easily check that if f has matrix A and g has matrix B , then the standard matrix product AB is the matrix for $f(g(v))$.

A quaternionic representation of G assigns to each $g \in G$ a quaternionic linear transformation $\rho(g)$ such that $\rho(g_1 g_2) = \rho(g_1) \rho(g_2)$, i.e., $\rho(g_1 g_2)v = \rho(g_1)(\rho(g_2)v)$, or equivalently assigns to each g an $n \times n$ quaternionic matrix $\rho(g)$ such that $\rho(g_1 g_2) = \rho(g_1) \rho(g_2)$.

3.2 Going Right and Going Left

We are going to define maps

$$\{\text{real representations}\} \rightarrow \{\text{complex representations}\} \rightarrow \{\text{quaternionic representations}\}$$

and maps

$$\{\text{real representations}\} \leftarrow \{\text{complex representations}\} \leftarrow \{\text{quaternionic representations}\}$$

We will call the first set of maps “going right” and the second set of maps “going left”. None of these maps necessarily preserve irreducibility.

Definition of Going Right: To go right, we extend the scalars acting on V , so a real V is pushed to $V \otimes C$ and a complex V is pushed to $V \otimes \mathcal{H}$. The representations ρ then extend linearly to the new spaces. So if $e_1, \dots, e_n$ is a real basis of V , the same set is a complex basis of $V \otimes C$, and if $e_1, \dots, e_n$ is a complex basis of V , it becomes a quaternionic basis of $V \otimes \mathcal{H}$. With these bases, the representation matrices are unchanged under going right. So if we start with matrices with real coefficients, we’ll get the same real matrices on $V \otimes C$ and if we start with complex matrices on a complex V , we will still have complex

matrices on $V \otimes \mathcal{H}$. Notice that $\dim_R V = \dim_C V \otimes C$, and similarly for the quaternionic push.

It might be useful to say a little more about $V \otimes \mathcal{H}$. We are thinking of $\mathcal{H}$ as $C \oplus Cj$, so a typical element of the extension is $\sum e_i c_i + \sum e_i d_i j$ where $c_i, d_i \in C$. Notice that $k = ij$ so $vk = (vi)j$. If we push a real space right twice, then a typical element has the form $\sum e_s(u_s + iv_s) + \sum e_s(w_s + ix_s)j = \sum e_s u_s + \sum e_s i v_s + \sum e_s w_s j + \sum e_s x_s k$ where $u_s, v_s, w_s, x_s \in R$.

Definition of Going Left: To go left, we keep the same set V and the same linear maps $\rho(g)$ and forget some scalars. So pushing a complex V left results in the same vectors V , but now we are only allowed to multiply by real scalars. Similarly pushing a quaternionic V left results in the same V , but now we are only allowed to scalar multiply by complex numbers.

We are hardly doing anything, and yet on the matrix level we are doing something more complicated. That is because if $e_1, \dots, e_n$ is a complex basis for V and we push left, then the real basis is $e_1, ie_1, e_2, ie_2, \dots, e_n, ie_n$. The representation matrix then changes from $\rho_{ij}(g) = a_{ij}(g) + ib_{ij}(g)$ to twice as large a matrix of the form

$$\begin{pmatrix} \ddots & & & \\ & a_{ij}(g) & -b_{ij}(g) & \\ & b_{ij}(g) & a_{ij}(g) & \\ & & & \ddots \end{pmatrix}$$

Similarly if we push a quaternionic V right twice, a quaternionic basis $e_1, \dots, e_n$ becomes a real basis $\dots, e_s, e_s i, e_s j, e_s k$ and a quaternionic matrix $\rho_{ij}(g) = a_{ij}(g) + b_{ij}i(g) + c_{ij}j(g) + d_{ij}k(g)$ becomes the matrix array

$$\begin{pmatrix} \ddots & & & & \\ & a_{ij}(g) & -b_{ij}(g) & -c_{ij}(g) & -d_{ij}(g) \\ & b_{ij}(g) & a_{ij}(g) & -d_{ij}(g) & c_{ij}(g) \\ & c_{ij}(g) & d_{ij}(g) & a_{ij}(g) & -b_{ij}(g) \\ & d_{ij}(g) & -c_{ij}(g) & b_{ij}(g) & a_{ij}(g) \\ & & & & \ddots \end{pmatrix}$$

Notice that if we push a quaternionic V left, $\dim_C V = 2 \dim_{\mathcal{H}} V$ and if we push further left, $\dim_R V = 2 \dim_C V = 4 \dim_{\mathcal{H}} V$.

3.3 The Type of a Representation

From now on we imagine that we have a list of all real irreducible representations of G on the left, a list of all complex irreducible representations of G in the center, and a list of all quaternionic irreducible representations of G on the right. We also have maps called “going right” and “going left” between them, but we have to be careful because these maps may not preserve irreducibility. If you like, think of the maps as “going right and then decomposing into irreducible subrepresentations”, etc., but in the end this issue will not play an important role.

We are going to separate real irreducible representations into three types, called *real type*, *complex type*, and *quaternionic type*. For instance, we might say that ρ is a real representation of complex type. We use the same terminology for complex irreducible representations, and for quaternionic irreducible representations.

If we have a real irreducible representation of real type, then we can push it right once or twice. We will show that this gives a complex representation of real type, and a quaternionic representation of real type, and both are irreducible. We will also show that all complex or quaternionic representations of real type arise in this way.

If we have a quaternionic irreducible representation of quaternionic type, we can push it left once or twice. We will show that this gives a complex representation of quaternionic type, and a real representation of quaternionic type, and both are irreducible. All complex or real representations of quaternionic type arise in this way.

The situation for complex representations is more complicated because the complex numbers have an automorphism given by conjugation. It is useful to treat this automorphism pedantically. If V is a complex vector space, we define a new complex vector space $\bar{V}$ to be the set V with the same addition, but with scalar multiplication given by $\lambda \times v \rightarrow \bar{\lambda}v$. If ρ is a representation of G on V , the same map defines a representation $\bar{\rho}$ on $\bar{V}$, which is also irreducible. Since the maps $\rho(g)$ did not change, it is easy to jump to the conclusion that $\rho = \bar{\rho}$, but that would be wrong. Suppose $e_1, \dots, e_n$ is a basis for V . Then these vectors also give a basis for $\bar{V}$. But if $v = \lambda_1 e_1 + \dots + \lambda_n e_n$ in V , then $v = \bar{\lambda}_1 e_1 + \dots + \bar{\lambda}_n e_n$ in $\bar{V}$. In the same way the matrix coefficients $\rho_{ij}(g)$ in V become $\bar{\rho}_{ij}(g)$ in $\bar{V}$. In short, this pedantry is a fancy way to say that $\bar{\rho}$ is the new representation we obtain by conjugating all matrix entries for ρ .

If ρ is an irreducible complex representation of G , we say that ρ is of complex type exactly when ρ and $\bar{\rho}$ are not equivalent representations.

If all entries of $\rho_{ij}(g)$ are real when using a particular basis, then certainly ρ and $\bar{\rho}$ are isomorphic and thus not of complex type. But there are cases where ρ and $\bar{\rho}$ are isomorphic and yet no basis can be chosen to make ρ_{ij} real. Luckily, there is an easy way to see if a complex irreducible representation has complex type. If χ is the character of ρ , then $\bar{\chi}$ is the character of $\bar{\rho}$. So ρ and $\bar{\rho}$ are not equivalent if and only if $\chi \neq \bar{\chi}$, which happens exactly when χ is not real.

Notice that complex representations of complex type come in pairs. If we push such a representation left, we get a real representation of complex type, and this representation is irreducible. Both ρ and $\bar{\rho}$ become equivalent when pushed left. Every real representation of complex type has this form. Similarly if we push a complex representation of complex type right, we get a quaternionic representation of complex type, and this representation is irreducible. Both ρ and $\bar{\rho}$ become equivalent as quaternionic representations. All quaternionic representations of complex type are obtained in this way.

3.4 Small Group Examples

We will prove the above results shortly. It is surprising to find that the quaternions are part of the mix. We might have expected an easier theory that only involved real and complex representations. Let us study the smallest possible finite groups and see where this theory suggests that we need to add quaternions to the mix.

The theory of complex representations is straightforward for one-dimensional representations. In this case, representations are just homomorphisms $G \rightarrow S^1$, equivalence is the same as equality, and the homomorphism is also the character of the representation. So a homomorphism is of complex type unless it takes only real values, and in that case it is real. Note that all abelian G fit this situation.

Consider, for example, the group Z_n . This group has n irreducible representations, each determined by the image of $1 \in Z_n$. If $\xi = e^{\frac{2\pi i}{n}}$, these images are $1, \xi, \xi^2, \dots, \xi^{n-1}$. Hence all of these representations are of complex type except the trivial representation and (when n is even) the representation taking 1 to -1 .

The one or two exceptional representations are trivially real representations of Z_n . The remaining representations can be converted to real representations by selecting the real basis $1, i$ for C and writing each complex number as a two by two matrix. For instance, we convert ξ to

$$\begin{pmatrix} \cos(\frac{2\pi}{n}) & -\sin(\frac{2\pi}{n}) \\ \sin(\frac{2\pi}{n}) & \cos(\frac{2\pi}{n}) \end{pmatrix}$$

But in this case, reflection across the x -axis takes the representations which sends $1 \rightarrow \xi^k$ to the representation which takes it to $1 \rightarrow \xi^{n-k}$. In short, each pair of irreducible

representations of complex type generates one irreducible representation of real type.

Let us list other groups of small order. The only groups of orders 2 and 3 are Z_2 and Z_3 . There are two groups of order 4, Z_4 and $Z_2 \times Z_2$, but both are abelian. The only group of order 5 is Z_5 . There are two groups of order 6, Z_6 and the symmetry group of an equilateral triangle D_3 . This D_3 is the first non-abelian group, but we already computed its irreducible representations and character table. The only group of order 7 is Z_7 . In order 8, there are three abelian groups, $Z_8, Z_4 \times Z_2$, and $Z_2 \times Z_2 \times Z_2$, but this time there are also two non-abelian groups, D_4 and Q , that is, the symmetries of a square and the group of unit quaternions. The groups D_4 and Q will illustrate why quaternions are needed in the theory.

Representations of D_4 First, we determine the conjugacy classes. The two reflections dividing sides are conjugate and the two reflections through vertices are conjugate. Rotation by $\frac{\pi}{2}$ is conjugate to rotation by $\frac{3\pi}{2}$. Thus there are five conjugacy classes. The dimensions of the irreducible representations must satisfy $d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 = 8$. This equation has only one solution $(1, 1, 1, 1, 2)$.

We can guess these representations. Two one-dimensional representations are obvious, ρ_1 which is identically 1 and ρ_2 given by the determinant which maps the rotations to 1 and the reflections to -1 .

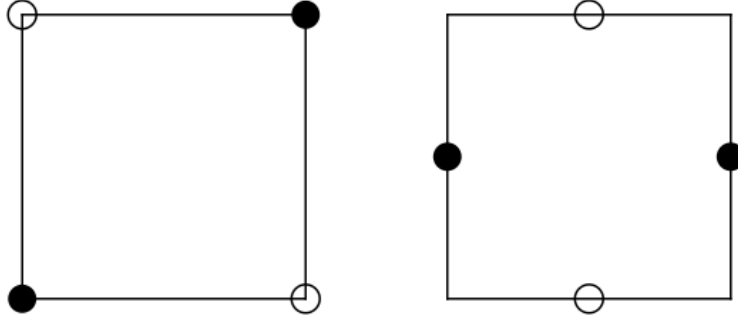


Figure 3.1: Decorating a Square

The above pictures suggest two other one-dimensional representations. In both cases a symmetry of the square either preserves or switches the set of white dots and the set of black dots. So the first picture defines ρ_3 , which assigns 1 to a symmetry that keeps the sets as they are and -1 to a symmetry that switches them. The second picture similarly defines ρ_4 . In more detail, both of these pictures assign -1 to rotations by $\frac{\pi}{2}$ and $\frac{3\pi}{2}$ and 1 to rotations by 0 and π . The first picture assigns 1 to reflections across lines through vertices and -1 to horizontal and vertical reflections. The second picture reverses these last two assignments.

Finally, let ρ_5 be the ordinary real 2-dimensional representation by affine motions in the plane. We can then construct the character table.

Class	χ_1	χ_2	χ_3	χ_4	χ_5
Identity	1	1	1	1	2
2-Click Rotation	1	1	1	1	-2
1-Click Rotations	1	1	-1	-1	0
Edge Reflections	1	-1	1	-1	0
Vertex Reflections	1	-1	-1	1	0

This table shows that the group has four one-dimensional representations, all real, and one two dimensional representation, also real.

Representations of the Group of Unit Quaternions: Again we determine the conjugacy classes. Since 1 and -1 commute with everything, each is a conjugacy class. Since $(-j)ij = -i$, i and $-i$ are conjugate. Similarly j and $-j$ are conjugate and k and $-k$ are conjugate. Thus there are five conjugacy classes, and as before the dimensions of the irreducible complex representations must be $(1, 1, 1, 1, 2)$.

Suppose that ρ is a one-dimensional representation. Then $\rho(i) = \pm 1$, $\rho(j) = \pm 1$, $\rho(k) = \pm 1$, and we easily deduce that there are four one dimensional representations, $\rho_1 =$ the identity, $\rho_2 = -1$ on i and j , $\rho_3 = -1$ on j and k , and $\rho_4 = -1$ on k and i .

The final irreducible representation over C is 2-dimensional, and the easy way to define it is to select a basis we'll call $1, j$ over C and let the representation multiply on the left by the appropriate signed quaternion in Q . The result is a linear combination of 1 and j ; recall that we write this combination with scalars on the right.

For example, k maps 1 to $0 \times 1 + i \times j$ and it maps j to $(-i) \times 1 + 0 \times j$ and thus

$$\rho_5(k) = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

The full representation can be read off from

$$\rho_5(-1) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}; \rho_5(i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}; \rho_5(j) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \rho_5(k) = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$$

The character table is then shown below:

Class	χ_1	χ_2	χ_3	χ_4	χ_5
1	1	1	1	1	2
-1	1	1	1	1	-2
i, -i	1	1	-1	-1	0
j, -j	1	-1	1	-1	0
k, -k	1	-1	-1	1	0

Notice that D_4 and Q have exactly the same character table, although the groups are not isomorphic. Since these character tables have no characters with complex values, none of these representations is of complex type. Four of the five representations are one dimensional real representations; if these representations are pushed right, they remain given by real matrices and remain one-dimensional, hence irreducible. So over $R, C, \mathcal{H}$ both groups have four irreducible representations of real type. But how about the final representation? For Q , this representation is not of complex type and yet the matrices defining the representation are not real. You can easily show that no basis will make them real. But Q has a trivial one-dimensional quaternionic representation. It is the representation which assigns to each unit quaternion $q \in Q$ the quaternionic matrix (q) . Our general theory claims that this induces the two-dimensional complex irreducible representation described earlier, and a four dimensional real irreducible representation.

3.5 Three Types of Real Irreducible Representations

We want to split real irreducible representations into three types, real type, complex type, and quaternionic type. Then we want to do the same thing for complex irreducible representations, and for quaternionic irreducible representations. Most of our work will be in the complex case, starting soon. The splits in the real and quaternionic cases are elegant and easy, so we give them first.

The simple version of Schur's lemma remains true in all three cases : if T is an intertwining operator between irreducible representations V and W , then either T is identically zero or T is an isomorphism.

We introduced a deeper version of Schur's lemma just before proving the orthogonality relations. That form stated that an intertwining operator T for an irreducible complex representation V has the form λI . The proof used the existence of an eigenvector for T , and therefore required the complex numbers. This is the only direct use of the complex numbers, but several of our later results depended on this deeper Schur lemma, including orthogonality relations and the fact that irreducible representations of abelian groups have dimension one.

There is a beautiful theorem for real representations which explains how far from the truth the hard form of Schur's lemma can stray. Notice that if T_1 and T_2 are intertwining

operators, $T_1 + T_2$, $T_1 \circ T_2$, and λT_1 are also.

Theorem 16 *Let V be an irreducible representation over R . Then the set of intertwining operators is a finite extension (skew) field over R , and thus is isomorphic to the real numbers, the complex numbers, or the quaternions.*

Proof: Proving that the set is a skew field is essentially trivial. The only finite dimensional skew fields over R are R, C, H . QED.

Definition 8 *Let ρ be an irreducible real representation. We say ρ has real type if its intertwining operators are isomorphic to R . We say ρ has complex type if its intertwining operators are isomorphic to C . We say ρ has quaternionic type if its intertwining operators are isomorphic to H .*

3.6 Three Types of Quaternionic Representations

Definition 9 *Let ρ be an irreducible quaternionic representation on V . We say ρ has real type if there is a basis of V for which all matrices of ρ are real. Otherwise, we say that ρ has complex type if there is a basis for V for which all matrices of ρ are complex. If neither of these are true, we say that ρ has quaternionic type.*

3.7 Three Types of Complex Irreducible Representations

Earlier we separated real and quaternionic irreducible representations into three types. But for complex irreducible representations, we only defined representations of complex type; these are irreducible representations such that $V \not\simeq \bar{V}$. In this section we study the case $V \simeq \bar{V}$ and separate these into types.

Suppose ρ is an irreducible representation on V and the resulting representations on V and $\bar{V}$ are equivalent. Then there is a map $J : V \rightarrow \bar{V}$ carrying one representation to the other. We can think of J as a map from V to V commuting with ρ , but this map is only linear over R ; it is anti-linear over C . It turns out that among the possible such maps, we can find a map which satisfies an additional condition: either $J^2 = I$ or $J^2 = -I$. If we know ρ , then only one of the two conditions can be satisfied. We will prove these results in this section. If $J^2 = I$, we say that ρ is of real type. If $J^2 = -I$, we say that ρ is of quaternionic type.

Warning: The result does *not* assert that the map J is unique. Also note that $J : V \rightarrow \bar{V}$, so it makes no sense to study eigenvectors of J . But $J^2 : V \rightarrow V$ and eigenvectors now make sense.

Theorem 17 *If ρ is an irreducible complex representation on V and its representations on V and $\bar{V}$ are equivalent, then there are isomorphisms $J : V \rightarrow \bar{V}$ which commute with ρ and are anti-linear over C . Among these are maps which also satisfy $J^2 = I$ or $J^2 = -I$. But only one of these conditions can be satisfied by any such map.*

Proof of theorem: As usual, we can introduce a Hermitian inner product $\langle v, w \rangle$ on V , where the expression is linear in v and anti-linear in w . Replacing this inner product by

$$\sum_{g \in G} \langle \rho(g)v, \rho(g)w \rangle$$

allows us to assume that the inner product is invariant under ρ .

Let V^* be the dual space to V , i.e., $\text{Hom}(V, C)$. Then V^* is canonically isomorphic to $\bar{V}$ as follows. If $w \in \bar{V}$, map w to the element $v \rightarrow \langle v, w \rangle$ in $\text{Hom}(V, C)$. The resulting map $\bar{V} \rightarrow V^*$ is complex linear because λw in $\bar{V}$ equals $\bar{\lambda}w$ in V , and

$$v \rightarrow \langle v, \bar{\lambda}w \rangle = \lambda \langle v, w \rangle$$

If ρ acts on V , it acts on $\text{Hom}(V, C)$, and $\bar{\rho}$ acts on $\bar{V}$ and all maps above preserve representations.

By assumption, $V \simeq \bar{V}$ as representation spaces. So

$$S^2(V) \oplus \Lambda^2(V) = V \otimes V \simeq V \otimes V^* \simeq \text{Hom}(V, V)$$

as representation spaces over C .

In particular, the subspaces of each vector space left fixed by every $\rho(g)$ are mapped to each other and thus all have the same dimension over C . The action of the group on $\text{Hom}(V, V)$ is $f \rightarrow \rho(g)^{-1}f(\rho(g)v)$, so the functions left fixed are intertwining operators. Since V is irreducible, this set is one-dimensional over C . It follows that either the set of invariant elements of $S^2(V)$ is one dimensional and set of invariant elements of $\Lambda^2(V) = 0$, or the set of invariant elements of $S^2(V) = 0$ and the set of invariant elements of $\Lambda^2(V)$ is one dimensional.

We now abandon the sequence of complex representations displayed above and start over with the space $V^* \otimes V^*$, on which ρ acts as a complex representation. This space is canonically isomorphic to the set of complex bilinear maps $B(v, w)$, and ρ acts on this space as a natural way. The decomposition $S^2(V) \oplus \Lambda^2(V)$ says that we can write an arbitrary B as a symmetric bilinear map plus a skew bilinear map. So either there is a nontrivial symmetric bilinear map $B(v, w)$ invariant under ρ , or else there is a nontrivial skew symmetric bilinear map $B(v, w)$ invariant under ρ , but not both.

We next claim that the nonzero $B(v, w)$ is non-degenerate, so if $v \neq 0$, then there is a w with $B(v, w) \neq 0$. Indeed, let $W \subset V$ be the set of all $v \in V$ such that $B(v, w) = 0$

for all $w \in V$. This is an invariant subspace of V because if $B(v, w) = 0$ for all w , then $B(\rho(g)v, w) = B(v, \rho^{-1}w) = 0$ for all w , and $\rho(g)v \in W$. Since V is irreducible, either $W = V$ and $B = 0$ or else $W = \{0\}$ and B is non-degenerate.

As earlier, select a ρ invariant Hermitian metric $\langle v, w \rangle$. Define J via $\langle v, Jw \rangle = B(v, w)$. Since $\langle \rho(g)v, \rho(g)w \rangle = \langle v, w \rangle$, $\langle v, Jw \rangle = B(v, w) = B(\rho(g)v, \rho(g)w) = \langle \rho(g)v, J\rho(g)w \rangle = \langle \rho(g^{-1})\rho(g)v, \rho(g^{-1})J\rho(g)w \rangle = \langle v, \rho(g^{-1})J\rho(g)w \rangle$ so $Jw = \rho(g^{-1})J\rho(g)w$ and $\rho(g)Jw = J\rho(g)w$. Also have $\langle v, J\lambda w \rangle = B(v, \lambda w) = \lambda B(v, w) = \lambda \langle v, Jw \rangle = \langle v, \bar{\lambda}Jw \rangle$. So $J\lambda w = \bar{\lambda}Jw$. So J is an anti-linear map from V to V , but this makes J^2 a complex linear map from V to V and so an intertwining operator for ρ . Since ρ is irreducible, $J^2 = \lambda I$ for some complex λ .

In the symmetric case, $\langle v, Jw \rangle = B(v, w) = B(w, v) = \langle w, Jv \rangle = \overline{\langle Jv, w \rangle}$. So $\langle v, J^2w \rangle = \langle J^2v, w \rangle$. Hence J^2 is Hermitian and its eigenvalues are real. Recall that $J^2 = \lambda I$, so λ is real. Moreover for any v we have $\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle J^2v, v \rangle = \overline{\langle Jv, Jv \rangle} = \langle Jv, Jv \rangle$, so λ is positive. Replace J by $J/\sqrt{\lambda}$; then $J^2 = I$.

In the skew symmetric case, $\langle v, Jw \rangle = B(v, w) = -B(w, v) = -\langle w, Jv \rangle = -\overline{\langle Jv, w \rangle}$. Then $\langle v, J^2w \rangle = -\overline{\langle Jv, Jw \rangle} = \langle J^2v, w \rangle$, so again J^2 is Hermitian and λ is real. This time $\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = -\langle J^2v, v \rangle = -\overline{\langle Jv, Jv \rangle} = -\langle Jv, Jv \rangle$, so λ is negative and we can replace J by $J/\sqrt{-\lambda}$ to obtain $J^2 = -I$.

QED.

Theorem 18 (Frobenius-Schur Indicator Theorem) *Let V be a complex irreducible representation with character χ . Then*

$$\int \chi(g^2) = \begin{cases} 1 & \text{if } V \cong \bar{V} \text{ and } \exists J \text{ with } J^2 = 1 \\ 0 & \text{if } V \not\cong \bar{V} \\ -1 & \text{if } V \cong \bar{V} \text{ and } \exists J \text{ with } J^2 = -1 \end{cases}$$

Lemma 4 *If ρ acts on V , then it acts on $V \otimes V = S^2(V) \oplus \Lambda^2(V)$, and we have*

$$\chi(g^2) = \chi(g)_{S^2} - \chi(g)_{\Lambda^2}$$

Proof: We can restrict attention to the abelian subgroup of G generated by g . Then ρ on V breaks into a sum of one-dimensional invariant subspaces V_i , and χ on V is the sum of χ on these subspaces. Moreover, χ on the subspaces is just ρ so $\chi(g^2) = \chi(g)^2$ and so forth.

Thus

$$\chi(g^2) = \sum \chi_{V_i}(g)^2 = \left(\sum \chi_{V_i}(g) \right)^2 - 2 \sum_{i < j} \chi_{V_i}(g) \chi_{V_j}(g)$$

The term $\sum_{i < j} \chi_{V_i}(g) \chi_{V_j}(g)$ is $\chi_{\Lambda^2(V)}(g)$ because

$$\Lambda^2(\sum V_i) = \sum_{i < j} \Lambda^0(V_i) \otimes \Lambda^2(V_j) + \sum_{i < j} \Lambda^1(V_i) \otimes \Lambda^1(V_j) + \sum_{i < j} \Lambda^2(V_i) \otimes \Lambda^0(V_j)$$

but Λ^2 of a one-dimensional space is zero. Moreover, evidently we have $(\sum \chi_{V_i}(g))^2 = \chi_{V \otimes V}(g)$ and this is $\chi_{S^2(V)}(g) + \chi_{\Lambda^2(V)}(g)$. We conclude that

$$\chi(g^2) = \chi_{S^2(V)}(g) - \chi_{\Lambda^2(V)}(g)$$

and so

$$\int \chi(g^2) = \int \chi_{S^2(V)}(g) - \int \chi_{\Lambda^2(V)}(g)$$

This completes the proof of the lemma.

Now replace V by V^* and consider the action of G on $V^* \otimes V^* = S^2(V^*) \oplus \Lambda^2(V^*)$. Recalling that V^* is isomorphic to $\overline{V}$, the characters in the previously displayed integrals are just conjugated. This makes no difference in the final result since the claim is that the integral is ± 1 or 0 . On the other hand, $\int \chi = \int \chi \overline{\chi_{id}}$, where χ_{id} is the character of the trivial identity representation. This integral gives the number of times id appears in the representation with character χ , and thus the dimension of the ρ invariant elements in V .

But $V^* \otimes V^* \cong (V \otimes V)^*$ is the space of bilinear forms $B(v, w)$. So the lemma asserts that the integral of our theorem equals the dimension of the invariant elements in $S^2(V)$ minus the dimension of the invariant elements in $\Lambda^2(V)$.

If $V \cong \overline{V}$, we proved that either there is a J with $J^2 = 1$ and $S^2(V^*)$ has exactly one non-trivial invariant element and $\Lambda^2(V^*)$ has none, or else there is a J with $J^2 = -1$ and the roles of these two spaces are reversed. So clearly our integral is 1 in the first case and -1 in the second.

On the other hand, $V^* \otimes V^*$ is canonically isomorphic to $Hom(V, V^*)$ and invariant elements in the first correspond to intertwining operators in the second. If $V \not\cong \overline{V}$, then V and V^* are not equivalent and all intertwining operators are zero. So there are no invariant elements and our integral is zero. QED.

Examples: Consider the groups D_4 and Q . In D_4 , all four reflections have square the identity. The rotations by 0 and π also have square the identity, and the rotations by $\frac{\pi}{2}$ and $\frac{3\pi}{2}$ have square a rotation by π . Looking back at the character table, $\int \chi(g^2) = \frac{1}{8} (2 \times 6 + (-2) \times 2) = 1$. So the 2-dimensional representation of D_4 has real type.

In Q , 1 and -1 have square 1 , and $\pm i, \pm j, \pm k$ have square -1 . So

$$\int \chi(g^2) = \frac{1}{8} (2 \times 1 + (-2) \times 6) = -1$$

Therefore the 2-dimensional representation of Q is of quaternionic type.

3.8 Another Way To Move Real Representations Right

Suppose ρ is an irreducible real representation on V . Then the set of intertwining operators is isomorphic to R, C , or $\mathcal{H}$. We can use these intertwining operators to extend the scalars acting on V to C when ρ has complex type, or $\mathcal{H}$ when ρ has quaternionic type. Notice that V does not change. However $\dim_C V = \frac{1}{2}\dim_R V$ and $\dim_{\mathcal{H}} V = \frac{1}{4}\dim_R V$ because a complex basis $e_1, \dots, e_n$ will induce a real basis $e_1, ie_1, \dots$, and a quaternionic basis $e_1, \dots, e_n$ will induce a real basis $e_1, e_1i, e_1j, e_1k, \dots$.

The new representations are still irreducible because if V has no nontrivial invariant *real* subspaces, it certainly has no such subspaces which are *complex* or *quaternionic*.

However, there is a complication: we only know the skew field up to an isomorphism preserving R . In the case when this skew field is isomorphic to C , for instance, there are two intertwining operators T satisfying $T^2 = -I$, namely T and $-T$, so there are two ways to extend the scalars to C . Thus we get two possible complex spaces where scalar multiplication by i in one of them is scalar multiplication by $-i$ in the other. Call these spaces V and $\bar{V}$, and notice that we can choose either one to play the role of V_C .

When V arises from a real representation of complex type, $V \not\cong \bar{V}$, because otherwise there would be an equivalence given by a complex anti-linear map $J : V \rightarrow \bar{V}$ commuting with ρ , so this J would be an intertwining operator which anticommutes with the intertwining operator T , contrary to the assumption that the intertwining operators are isomorphic to C .

So each irreducible real representation of complex type maps to a pair of inequivalent conjugate complex irreducible representations.

At first this complication seems much more likely in the quaternionic case, because $\mathcal{H}$ has infinitely many automorphisms. But we will show that all the quaternionic extensions are equivalent, essentially because quaternionic automorphisms are all inner.

Here are the details. Think of quaternions as ordered pairs (r, v) consisting of a real number and a three dimensional vector. Then

$$(r_1, v_1)(r_2, v_2) = (r_1r_2 - v_1 \cdot v_2, r_1v_2 + r_2v_1 + v_1 \times v_2)$$

If $\psi : \mathcal{H} \rightarrow \mathcal{H}$ is an automorphism fixing R , then $\psi(i^2) = -1 = \psi(i)^2$. If $\psi(i) = (r, v)$, then

$$(r, v)(r, v) = (r^2 - v \cdot v, 2rv) = (-1, 0)$$

so either $v = 0$ or $r = 0$. But $v = 0$ implies $r^2 - v \cdot v = r^2 = -1$, which is impossible. So $r = 0$. Similar calculations work for j and k , so ψ maps vectors to vectors and therefore $\psi(r, v) = (r, A(v))$ for some real linear $A : R^3 \rightarrow R^3$.

Then

$$A((0, v)(0, w)) = A(-v \cdot w, v \times w) = (-v \cdot w, v \times w) = (0, Av)(0, Aw) = (-Av \cdot Aw, Av \times Aw)$$

so that A preserves dot and cross products of vectors. Because A preserves dot products, A is an orthogonal map. Because it preserves cross products, it preserves orientation. So $A \in SO(3)$.

Let u be a unit quaternion, that is, a quaternion of length one. Then $q \rightarrow uq\bar{u}$ is clearly an automorphism. Such automorphisms preserve the real part of quaternions and map R^3 to R^3 . Indeed the set of unit quaternions is a group diffeomorphic to S^3 , and this operation maps that group to $SO(3)$ as a 2-fold covering. This is the classical construction of $Spin(3)$, the universal covering group of $SO(3)$. The upshot is that every automorphism of $\mathcal{H}$ has the form $q \rightarrow uq\bar{u}$ for some unit quaternion u .

Now suppose ρ is a real irreducible representation on V of quaternionic type. If we pick one fixed way to think of intertwining operators as elements of $\mathcal{H}$, then any other way will be given by the new scalar product $\hat{v}q = vuq\bar{u}$. Let us work out the coordinate formula for ρ with the initial scalar and with the new scalar.

Select a quaternionic basis $e_1, \dots, e_n$ for V . Then using the initial scalar product, every element of V is uniquely $v = e_1q_1 + \dots + e_nq_n$. We write this as a column vector

$$\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}$$

Write $\rho(e_i) = \sum e_j \rho_{ji}$ so the effect on basis vectors is given by the columns. Then $\rho(v) = \rho(\sum e_i q_i) = \sum \rho(e_i) q_i = \sum e_j \rho_{ji} q_i$ and in matrix form this is

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & \rho_{nn} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$$

We repeat these calculations using the new scalar product $vuq\bar{u}$. Below is v , written in terms of the older scalars on the left, and in terms of the new scalars on the right (remember that the new scalars will multiply by u on the left and $\bar{u}$ on the right, and then use these as old scalars):

$$v = e_1q_1 + \dots + e_nq_n = e_1\bar{u}q_1u + \dots + e_n\bar{u}q_nu$$

Similarly $\rho(e_i)$ is written in terms of old scalars on the left and new ones on the right:

$$\rho(e_i) = \sum e_j \rho_{ji} = \sum e_j \bar{u} \rho_{ji} u$$

Finally below is $\rho(v)$ written using old scalars and then again using new ones:

$$\begin{aligned}\rho(v) &= \rho\left(\sum e_i q_i\right) = \sum \rho(e_i) q_i = \sum e_j \rho_{ji} q_i = \\ \rho(v) &= \rho\left(\sum e_i \bar{u} q_i u\right) = \sum \rho(e_i) \bar{u} q_i u = \sum e_j \bar{u} \rho_{ji} u \bar{u} q_i u = \sum e_j \bar{u} \rho_{ji} q_i u\end{aligned}$$

So we can map vectors when use the old scalar product to vectors when using the new one by

$$\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \rightarrow \bar{u} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} u$$

and we can map matrices from the old to the new space by

$$\begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{pmatrix} \rightarrow \bar{u} \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{pmatrix} u$$

and thus the two representations are equivalent.

Remark: A related operation obtains a complex representation from an irreducible real representation of quaternionic type on V by extending the scalars to $\mathcal{H}$ using the intertwining operators, and then forgetting scalar multiplication by j and k . The resulting representation does not depend on the isomorphism between intertwining operators and $\mathcal{H}$ because it is a composition of two operators that preserve equivalence, extending scalars to $\mathcal{H}$ and then forgetting j and k .

3.9 Irreducible Representations over R and C

Theorem 19 *Consider the set $\mathcal{E}_R$ of equivalence classes of real irreducible representations, and the set $\mathcal{E}_C$ of equivalence classes of complex irreducible representations.*

- *It is possible to assign to each $\rho \in \mathcal{E}_R$ of real type a unique $\psi \in \mathcal{E}_C$ of real type.*
- *It is possible to assign to each $\rho \in \mathcal{E}_R$ of complex type a unique pair of conjugate $\psi_i \in \mathcal{E}_C$ of complex type.*
- *It is possible to assign to each $\rho \in \mathcal{E}_R$ of quaternionic type a unique $\psi \in \mathcal{E}_C$ of quaternionic type. This ψ comes from an irreducible quaternionic representation τ by just forgetting scalar multiplication by j and k .*
- *Every element $\psi \in \mathcal{E}_C$ comes from a $\rho \in \mathcal{E}_R$ by one of these mappings.*

- In the first case, $\dim_R \rho = \dim_C \psi$; in the second case $\dim_R \rho = 2 \dim_C \psi_i$ for each i ; in the third case, $\dim_R \rho = 2 \dim_C \psi = 4 \dim_{\mathcal{H}} \tau$.

Proof: To define the maps, recall that the set of intertwining operators of a real irreducible representation is isomorphic to R, C , or $\mathcal{H}$. In the real case, send ρ on V to ρ on $V \otimes C$.

In the complex case, make V a complex vector space by defining scalar multiplication by i to be an intertwining operator T satisfying $T^2 = -I$. There are exactly two of these T and $-T$, giving a pair of representations V and $\bar{V}$. In the previous section, we proved that both are irreducible over C and $V \not\simeq \bar{V}$. So our map gives a pair of irreducible, conjugate representations of complex type in this case.

In the quaternionic case, send ρ on V to ρ on V , but extend scalar multiplication to the quaternions by finding T_i, T_j, T_k playing the roles of i, j, k and letting them act on V . This gives a quaternionic representation; obtain the complex representation by forgetting T_j, T_k . We proved earlier that this representation does not depend on the choice of i, j, k and is irreducible. It is of quaternionic type because scalar multiplication by j is a real anti-complex linear map J from V to V such that $J^2 = -I$.

Thus the last two of the three maps give the right kind of resulting complex representation, and we only have to show that the first map also has this property.

We claim the space $V \otimes C$ is irreducible. Indeed the set of intertwining operators contains only real multiples of the identity, so the hard version of Schur's lemma is true, and consequently the full orthogonality relations hold for this real ρ . In particular, the real character has norm 1. But the complex extension has the same real matrices, and so the same character. Since this character has norm 1, the complex representation is irreducible.

If ρ has real type, we just discovered that ρ on $V \otimes C$ still has a real character, so $V \simeq \bar{V}$. For concreteness, we can think of $V \otimes C$ as the set of ordered pairs $\langle v_1, v_2 \rangle$ of elements in V , with scalar multiplication by C given by $i \langle v_1, v_2 \rangle = \langle -v_2, v_1 \rangle$. Thus in $\bar{V} \otimes \bar{C}$ we have $i \langle v_1, v_2 \rangle = \langle v_2, -v_1 \rangle$. Define $J : V \otimes C \rightarrow V \otimes C$ by $J \langle v_1, v_2 \rangle = \langle v_1, -v_2 \rangle$. This map is complex anti linear because $Ji \langle v_1, v_2 \rangle = J \langle -v_2, v_1 \rangle = \langle -v_2, -v_1 \rangle$ and $(-i)J \langle v_1, v_2 \rangle = (-i) \langle v_1, -v_2 \rangle = \langle v_2, v_1 \rangle$. The map preserves ρ because $J\rho \langle v_1, v_2 \rangle = J \langle \rho v_1, \rho v_2 \rangle = \langle \rho v_1, -\rho v_2 \rangle$ and $\rho J \langle v_1, v_2 \rangle = \rho \langle v_1, -v_2 \rangle = \langle \rho v_1, -\rho v_2 \rangle$. Finally, $J^2 = 1$. Thus the resulting irreducible complex representation has real type.

Completion of the Proof: The above steps prove the first three claims of the full theorem. Since the last claim is trivial, we need only prove the fourth claim by showing that we can go backward.

We do this first in the real case. So we start with an irreducible ρ on a complex vector space V , and a complex antilinear map $J : V \rightarrow V$ with $J^2 = I$. Define $V^+ = \{v \in V \mid J(v) = v\}$

and $V^- = \{v \in V \mid J(v) = -v\}$. Both of these are real subspaces of V . Also $V = V^+ \oplus V^-$ because any v can be written $\frac{v+J(v)}{2} + \frac{v-J(v)}{2}$. Since $J(iv) = -i(J(v))$, multiplication by i maps $V^+ \rightarrow V^-$ and $V^- \rightarrow V^+$. Since our maps commute with ρ , the representation ρ leaves V^+ and V^- invariant. Pick a real basis of V^+ , $e_1, \dots, e_n$. Then $ie_1, \dots, ie_n$ is a real basis of V^- and so $e_1, \dots, e_n$ is a complex basis for V .

The real representation ρ on V^+ is irreducible, since if $W \subset V^+$ is a nontrivial real invariant subspace, then $W \oplus iW \subset V$ is a complex invariant subspace of V , but this is impossible because ρ is irreducible on V . Notice that our construction sending real irreducible representations of real type to complex representations sends ρ on V^+ to ρ on $V^+ \otimes C = V$ since $e_1, \dots, e_n$ is a complex basis of V . So the complex ρ on V comes from a real ρ on V^+ . We need only show that this ρ on V^+ has real type. If not, it is a complex or quaternionic type, so it has an intertwining operator I with $I^2 = -1$. This operator then induces an anti-complex $J : V \rightarrow V$ with $J^2 = -1$. But this is impossible because ρ on V cannot have both a J with $J^2 = 1$ and a J with $J^2 = -1$.

Next we start with a pair of complex irreducible representations ρ on V and $\bar{\rho}$ on $\bar{V}$, of complex type. By simply forgetting scalar multiplication by i , V becomes a real vector space V_R and ρ becomes a real representation on this space. In that case, multiplication by i is an intertwining operator on the real V_R , and when we use it to give V_R a complex structure, we return to the complex V that we started with. We only need to show that the real V_R is irreducible and its set of intertwining operators cannot be quaternionic. If its intertwining operators were quaternionic, then there would be a real intertwining operator J with $IJ = -JI$. This J would then be a complex antilinear map $J : V \rightarrow \bar{V}$, contradicting the assumption that our starting point were inequivalent representations on V and $\bar{V}$. So we are left with proving that the real ρ, V_R representation is irreducible. The proof requires a lemma which will also be used when discussing the quaternionic case:

Lemma 5 *Let ρ be an irreducible representation on a complex vector space V . Let V_R be the same set considered as a real vector space by forgetting scalar products by i , so that ρ becomes a (not necessarily irreducible) real representation on V_R . The resulting representation on $V_R \otimes C$ is equivalent to $V \oplus \bar{V}$.*

Proof of lemma: Think of $V_R \otimes C$ as the set of pairs (v_1, v_2) with $v_i \in V_R$. On this space, we have two ways to scalar multiply by i , and both commute with ρ . The first uses the existing scalar multiplication in V_R . Call this T_1 . The second is the standard way to make tensoring with C a complex vector space: $T_2((v_1, v_2)) = (-v_2, v_1)$. These methods agree on elements of the form $(v_1, -iv_1)$ and differ by a sign on (v_1, iv_1) . Each of these is a complex subspace of $V_R \otimes C$ and their sum is the entire space. The first space is isomorphic to V via the map $v \rightarrow (v, -iv)$ and the second is isomorphic to $\bar{V}$ via the map $v \rightarrow (v, iv)$.

Completion of the complex case of the theorem: Recall that we only need prove that V_R is irreducible.

If not, write $V_R = W_1 \oplus \dots \oplus W_k$ where the W_i are real subspaces on which ρ acts irreducibly. Then k must equal two because $V_R \otimes C = \sum W_i \otimes C = V \oplus \bar{V}$.

Notice that $W_1 \cap iW_1$ is a complex invariant subspace, and so all of W_i or $\{0\}$. A similar statement holds for W_2 . If $W_1 \cap iW_1 = W_1$ and $W_2 \cap iW_2 = W_2$, then both W_1 and W_2 are complex subspaces of $V_R = V$, contradicting the irreducibility of V . So at least one of the intersections, say $W_1 \cap iW_1$, is $\{0\}$. But iW_1 is invariant, so in that case $W_2 = iW_1$. Thus $V = W_1 \otimes C$ and $\bar{V} = iW_1 \otimes C$.

On the other hand, $W_1 \otimes C$ and $iW_1 \otimes C$ are isomorphic, because a typical element of the first space is $w_1 + iw_2$ where $w_1, w_2 \in W_1$, and a typical element of the second space is $iw_1 - w_2$ where $w_1, w_2 \in W_1$ and multiplication by i is a complex linear isomorphism. This preserves ρ because ρ acts separately on each component. But we are working on the case when V and $\bar{V}$ are not equivalent.

Completion of the Quaternionic Case: Finally, suppose we have an irreducible complex ρ on V and a complex anti-linear map $J : V \rightarrow V$ such that $J^2 = -1$. Clearly we can use this J to make V into a quaternionic vector space. The resulting quaternionic representation is irreducible. Indeed, V has no non-trivial complex invariant subspaces, so it certainly has no non-trivial quaternionic invariant subspaces.

Let V_R be the real representation formed by forgetting i, j, k as scalars. We will prove that ρ on V is irreducible. Suppose this for a moment. Notice that i, j, k become intertwining operators. So V_R must be of quadratic type. We can use these intertwining operators to make V_R into the quaternionic representation space just created.

Thus it suffices to prove that V_R is irreducible over R . If not, the reasoning in the previous case shows that we can write $V_R = W_1 \oplus W_2$ where $W_1 \otimes C = V$. Define a complex anti-linear map $L : V \rightarrow V$ to be the identity on W_1 and $-I$ on iW_1 ; this commutes with ρ . Then $L^2 = I$. This is a contradiction because we are in the third case and have a J with $J^2 = -I$.

3.10 Irreducible Representations over $\mathcal{H}$

Lemma 6 *Suppose ρ is an irreducible quaternionic representation on V . Let V_C denote the representation on the same space formed by forgetting scalar multiplication by j and k . Then either V_C is irreducible over C or else V_C has a nontrivial complex invariant subspace $\mathcal{W}$ such that V_C is equivalent to $\mathcal{W} \oplus \bar{\mathcal{W}}$.*

Proof: If V_C is not irreducible as a complex vector space under the action of ρ , it must contain a nontrivial irreducible complex subspace $\mathcal{W}$. Since J commutes with ρ , $\mathcal{W}j$ must also be invariant; it is a complex subspace since $(wj)i = (-wi)j$. But $\mathcal{W} \cap \mathcal{W}j$ is mapped to itself by j , so it is a quaternionic invariant subspace of V . It is not all of V , so it must

be 0 because V is irreducible as a quaternionic representation. So $\mathcal{W} \oplus \mathcal{W}j$ is a direct sum of complex irreducible subspaces inside V . This subspace is invariant under j and thus an invariant quaternionic subspace of an irreducible V . So it must be all of V .

Notice that $\mathcal{W}$ and $\mathcal{W}j$ are both complex subspaces of V which intersect only at 0. The map $J : \mathcal{W} \rightarrow \mathcal{W}j$ commutes with ρ because ρ is linear over the quaternions. It is also a complex isomorphism from $\overline{\mathcal{W}}$ to $\mathcal{W}j$ because $(w(-i)j) = (wj)i$.

We can identify $\mathcal{W}j$ with $\mathcal{W}$ as sets by mapping wj to w . But since $(wj)i = ((w(-i)j))$, this identification maps $\mathcal{W}j$ to $\overline{\mathcal{W}}$ as complex vector spaces. So $V = \mathcal{W} \oplus \overline{\mathcal{W}}$ as complex representation spaces.

Definition 10 *Let ρ be an irreducible quaternionic representation on V .*

- *If the complex representation obtained by forgetting scalar multiplication by j and k is not irreducible and $\mathcal{W} \simeq \overline{\mathcal{W}}$, the quaternionic ρ has real type.*
- *If the complex representation obtained by forgetting scalar multiplication by j and k is not irreducible and $\mathcal{W} \not\simeq \overline{\mathcal{W}}$, the quaternionic ρ has complex type.*
- *If the complex representation obtained by forgetting scalar multiplication by j and k is irreducible, the quaternionic ρ has quaternionic type.*

Theorem 20 *Every irreducible quaternionic representation is equivalent to exactly one of the following:*

- *the tensor product $V \otimes \mathcal{H}$ of a real irreducible representation of real type;*
- *a representation formed from a real irreducible representation of complex type V_R by first extending the scalars to C using intertwining operators to obtain V_C (there are two ways to do this, but either can be used in the next step) and then forming $V_C \otimes_C \mathcal{H}$; equivalently, a representation formed from a pair of conjugate irreducible representations of complex type by selecting one member of the pair, V_C , and forming $V_C \otimes_C \mathcal{H}$;*
- *a representation formed from a real irreducible representation of quaternionic type V by extending the scalars to H using the intertwining operators*

In the first case, $\dim_{\mathcal{H}} \rho = \dim_R \rho$. In the second case, $\dim_{\mathcal{H}} \rho = \dim_R \rho$. In the final case, $\dim_{\mathcal{H}} \rho = \frac{1}{4} \dim_R \rho$.

Proof of real case: If ρ has real type, we can forget multiplication by j and k and obtain a complex representation which decomposes into a conjugate pair of irreducible complex representations. The representations of this pair are equivalent, so each is a representative of one complex irreducible representation of real type. By the main theorem of the previous section, there is an irreducible real representation of real type on V_R such that $\mathcal{W}$ is equivalent to $V_R \otimes C$. Then ρ on V is equivalent to $V_R \otimes H = \mathcal{W} \oplus \mathcal{W}j$.

Proof of the complex case: If ρ on V has complex type, we can forget multiplication by j and k and obtain a complex representation which decomposes into a conjugate pair of irreducible complex representations. The representations of this pair are not equivalent, so each is an element on the list of irreducible complex representations of complex type. By the main theorem of the previous section, if we forget scalar multiplication by i , then $\mathcal{W}$ is an irreducible real representation of complex type, V_R . Hence it has an intertwining operator T such that T^2 and $(-T)^2$ are the identity. Using T in one case and $-T$ in the other, we make V_R into complex irreducible representations equivalent to $\mathcal{W}$ and $\overline{\mathcal{W}}$. But then $\mathcal{W} \otimes_C \mathcal{H} = \mathcal{W} \oplus \mathcal{W}j = V$.

Proof of the quaternionic case: If ρ on V has quaternionic type, we can forget multiplication by j and k and obtain a complex irreducible representation V_C . There is an automorphism of $\mathcal{H}$ taking $i \rightarrow -i$ and $j \rightarrow -j$ and $k \rightarrow k$; applying this automorphism gives an equivalent representation ρ . But then if we forget j and k , we obtain $\overline{V_C}$. Consequently $V_C \simeq \overline{V_C}$. Moreover, multiplication by j gives an anti-linear map ρ -invariant map $J : V_C \rightarrow V_C$ such that $J^2 = -I$. So V_C has quaternionic type. By the main theorem of the previous section, there is a unique associated irreducible real representation of ρ of quaternionic type. The intertwining operators of this real representation can be used to extend the scalars to $\mathcal{H}$ and when we then forget scalar multiplication by j and k , we get V_C . These operators require us to select an isomorphism between the intertwining operators and $\mathcal{H}$, but the choice makes no difference up to equivalence of representations. Thus V_C is formed from any of these representations by forgetting scalar multiplication by j and k .

There is an easy summary of this complicated chapter.

The irreducible representations over $\mathcal{H}$ are given by quaternionic matrices. By cleverly choosing a basis, it may be possible to make all of these matrices complex, or even all real. Thus there are three types of quaternionic irreducible representations: those that can be given by real matrices, those that can be given by complex matrices, and those that require quaternionic matrices.

A representation that requires quaternionic entries can be pushed to a complex representation by replacing each $q_{ij} = c_{ij} + d_{ij}k$ by

$$\begin{pmatrix} c_{ij} & -d_{ij} \\ d_{ij} & c_{ij} \end{pmatrix}$$

and can be pushed to a real representation by replacing each $q_{ij} = a_{ij} + b_{ij}i + c_{ij}j + d_{ij}k$

by

$$\begin{pmatrix} a_{ij} & -b_{ij} & -c_{ij} & -d_{ij} \\ b_{ij} & a_{ij} & d_{ij} & -c_{ij} \\ c_{ij} & -d_{ij} & a_{ij} & b_{ij} \\ d_{ij} & c_{ij} & -b_{ij} & a_{ij} \end{pmatrix}$$

These complex or real representations remain irreducible, so there is little advantage in introducing real or complex entries; representations requiring quaternionic entries are essentially quaternionic.

Just to be clear, then, representations requiring quaternionic matrices also appear in the lists of irreducible representations over R and C , giving the representations of quaternionic type in these lists.

Next consider irreducible representations over R . Each representation is given by real matrices. But some of these representations remain irreducible even if considered acting on C^n or $\mathcal{H}^n$. These are the representations of real type, and without changing anything they also appear in the lists of irreducible representations over C and $\mathcal{H}$, giving the representations of real type in these lists.

Finally, there are the representations of complex type in the three lists. These are obtained from complex irreducible representations with two special properties: the matrices cannot be made real by selecting a different basis, and the representation remains irreducible when the matrices act on $\mathcal{H}^n$. These representations exist unaltered on lists of irreducible representations over $\mathcal{H}$, and they appear on lists of irreducible representations over R in their real forms obtained by replacing each $c_{ij} = a_{ij} + ib_{ij}$ with

$$\begin{pmatrix} a_{ij} & -b_{ij} \\ b_{ij} & a_{ij} \end{pmatrix}$$

The double size real matrix representation is irreducible.

There is one complication, which occurs only in the complex case. If we select an irreducible representation of complex type, write its complex matrix representation, and then replace each complex entry by its conjugate, we get a new representation not equivalent to the original. But this new representation gives nothing new when considered a quaternionic representation, or when pushed to a real representation. Thus the essentially complex representations appear twice in lists of complex irreducible representations, but only once in lists of real or quaternionic irreducible representations.

Chapter 4

Representations of Lie Groups and Lie Algebras

4.1 Lie Group Representations

Definition 11 A (real, complex, quaternionic) representation of a Lie group G is a continuous group homomorphism $\varphi : G \rightarrow GL(V)$ where V is a finite dimensional real, complex, or quaternionic vector space.

Remark: By a general theorem in Lie theory, the continuous map ρ is automatically C^∞ .

Remark: A representation associates a matrix with each element of G , and the matrices multiply exactly as the group elements multiply. Our goal is to find *all* representations of a given group.

Definition 12 If φ_1 is a representation on V_1 and φ_2 is a representation on V_2 , then $\varphi_1 \oplus \varphi_2$ is the obvious representation on $V_1 \oplus V_2$, consisting of matrices with φ_1 in the top-left block, φ_2 in the bottom-right block, and 0 in the two remaining blocks.

Definition 13 Two representations φ_1 and φ_2 are equivalent if there is an isomorphism $\psi : V_1 \rightarrow V_2$ such that $\varphi_1 = \psi^{-1}\varphi_2\psi$.

Remark: In terms of matrices, φ_1 and φ_2 are equivalent if there is a non-singular B with $\varphi_1 = B^{-1}\varphi_2B$. Another way to say this is that φ_1 and φ_2 are equivalent if φ_1 gives the matrices using one basis of V and φ_2 gives these matrices using another basis.

Remark: The theory is complicated because we want to combine these ideas. If the matrices of φ have blocks down the diagonal, then φ is a direct sum of easier representations. But changing the basis can hide these blocks.

Remark: We are going to prove that every representation of a compact Lie group is equivalent to a unique sum of “irreducible” representations. This result will hold over R , C , and $\mathcal{H}$ and hold for finite groups, compact groups, and semisimple groups (these to be defined later). The theorem is *not* true for all Lie groups.

Definition 14 A representation φ of G on V is indecomposable if we cannot write $V = V_1 \oplus V_2$ with both subspaces nonzero, such that $\varphi(V_1) \subset V_1$ and $\varphi(V_2) \subset V_2$.

Definition 15 A representation φ of G on V is irreducible if the only subspaces V_1 satisfying $\varphi(V_1) \subseteq V_1$ are (0) and V .

Definition 16 Let ρ act on V and W . A linear transformation $T : V \rightarrow W$ is an intertwining operator if $T\rho_V(g) = \rho_W(g)T$ for all g .

Lemma 7 (Schur) If V and W are irreducible and inequivalent, the zero map is the only intertwining operator between them. If V and W are irreducible and equivalent and the ground field is C , every intertwining operator has the form λI for a complex constant λ .

Proof: Exactly the same as the finite group case.

Theorem 21 Let ρ be a representation of the Lie group G on V .

- The representation ρ can be written as a sum of indecomposable representations (possibly non-uniquely).
- If ρ can be written as a sum of irreducible representations, then any two such representations have the same number of terms, and after a rearrangement, the corresponding terms are equivalent.

Proof: Exactly the same as the finite group case.

Remark: Consequently, if a representation can be written as a sum of irreducible representations, then these irreducible representations are unique up to order. Unfortunately, some representations are indecomposable but not irreducible, and they can never be written as sums of irreducible representations. For example, consider the representation of R by

$$\varphi(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

An invariant subspace of R^2 would be generated by a nonzero (a, b) with $(a+tb, b) = \lambda(a, b)$. Looking at the second component, $\lambda = 1$ or $b = 0$. If $\lambda = 1$ then $(a+tb, b) = (a, b)$, so $tb = 0$ for all t and thus $b = 0$. Thus in all cases the only non-trivial invariant subspace is generated by $(1, 0)$ and has no complementary invariant subspace.

This problem cannot occur if G is finite or if G is compact.

Theorem 22 *Suppose G is finite or else a compact Lie group. Then every indecomposable representation is irreducible. Therefore every representation of G can be written as a direct sum of irreducible representations, which are then unique up to order.*

Remark: We proved this for finite groups, by introducing a very elementary integral over these groups:

$$\int_G f(g) = \frac{1}{|G|} \sum_{g \in G} f(g)$$

To extend this theorem to the case of continuous representations of a compact Lie group, we will define an integral over the group. The remaining details of the proof are unchanged. We only need the following properties:

- $\int_G f(g) dg$ is defined for all real valued continuous functions
- $\int_G$ is linear in f
- If $f \geq 0$ and f is not identically zero, then $\int_G f(g) dg$ is greater than zero
- $\int_G$ is left-invariant; if $L_g(f)(g_1) = f(gg_1)$, then $\int_G L(g)f(g_1) dg_1 = \int_G f(g_1) dg_1$
- $\int_G 1 dg = 1$

Remark: In the 1930's, Haar proved that there is a left-invariant integral on any locally-compact topological group. When the group is not compact, integration is defined only for continuous functions with compact support, and the final condition is omitted. It immediately follows that our fundamental theorem of representations holds for any compact topological group. We will prove existence of invariant integrals for compact Lie Groups in the first section of the next chapter, without using Haar's result.

Shortly after Haar proved his theorem, von Neumann proved that the Haar integral is unique up to a constant multiple. In the compact case, we can fix this multiple by requiring that $\int_G 1 dg = 1$ and then the integral is unique.

Remark: There is a parallel theory for right-invariant integrals. In the compact case, left and right invariant Haar integrals turn out to be equal, so we can assume that the integral is both left and right invariant.

4.2 Generalizing Representation Theory from Finite Groups to Compact Groups

Once we are able to integrate continuous functions over compact groups, almost everything in the finite group generalizes to the compact case with exactly the same proof. Thus in the complex case we can introduce an invariant Hermitian inner product so the representation matrices are all unitary. If two such representations are equivalent, then they are

equivalent by a unitary intertwining map. The orthogonality relations remain valid with the same proofs, and the theorems of character theory remain valid. The entire theory of representations over $R, C, \mathcal{H}$ holds, with the same proofs.

There are only two results in the finite case which do not generalize — or rather, remain true but require new proofs. The first states that the coefficients ρ_{ij} of irreducible representation matrices form an orthonormal basis for $L^2(G)$. These coefficients are certainly orthonormal; the difficult step is proving completeness. In the compact case, $L^2(G)$ is an infinite dimensional Hilbert space, and the result, while true, requires techniques in analysis.

The second result is similar. It says that the characters of G form an orthonormal basis of the subspace of $L^2(G)$ formed by functions which are constant on conjugacy classes. This is difficult for the same reason; this space is an infinite dimensional Hilbert space and techniques from analysis are required in the proof.

For a long time, we will not need these results. If we eventually require them, we will stop and give a proof.

4.3 Representations of Lie Algebras

Definition 17 *A representation of a Lie Algebra L on a finite dimensional real, complex, or quaternionic vector space V is an assignment to each $x \in L$ of a linear transformation $\varphi(x) : V \rightarrow V$ such that the assignment is linear over R and $\varphi[x, y] = [\varphi(x), \varphi(y)]$. Here the Lie bracket on matrices or linear transformations is defined by $[A, B] = AB - BA$.*

Remark: Suppose $\varphi : G \rightarrow GL(V)$ is a representation of a Lie group. Then $\varphi^* : \mathcal{G} \rightarrow gl(V)$ is a representation of the Lie algebra of G by elementary Lie theory. If G is connected, the map φ^* completely determines φ , again by Lie theory.

Theorem 23 *Suppose G is a connected Lie group. A representation φ of G is irreducible if and only if the corresponding representation φ^* is irreducible. A representation φ of G splits as a direct sum of invariant subspaces $V_1 \oplus \dots \oplus V_k$ if and only if the corresponding representation φ^* of $\mathcal{G}$ splits as the same sum of invariant subspaces.*

Proof: The proof is easy. It is helpful to remember the exponential map $\exp : \mathcal{G} \rightarrow G$, the fact that $\exp \circ \varphi^* = \varphi \circ \exp$, and that fact that $\exp$ is onto an open neighborhood of G and thus its image generates all of G .

Remark: There is just one complication. If $\psi : \mathcal{G} \rightarrow gl(V)$ is a representation of a Lie algebra associated to a Lie group G , this representation does not necessarily induce a representation of G . If G is simply-connected, however, ψ induces such a representation. In particular there is an associated Lie group representation of the universal cover $\overline{G}$.

For example, the Lie algebra $so(3)$ has one irreducible representation of each dimension d . But only the representations associated with even d induce representations of $SO(3)$. The remaining representations are “spinor representations” of $Spin(3)$.

Remark: It follows that if L is the Lie algebra of a compact, simply connected Lie group G , then the representation theory of L is exactly the same as the representation theory of G . Every representation of L is a sum of irreducible representations, and these representations are unique to within order and equivalence.

Chapter 5

Invariant Integrals on Lie Groups

5.1 Left Invariant Metrics, Right Invariant Metrics

Definition 18 Let G be an arbitrary Lie group. For each $g_1 \in G$, define $L_{g_1} : G \rightarrow G$ by $L_{g_1}(g) = g_1g$ and define $R_{g_1} : G \rightarrow G$ by $R_{g_1}(g) = gg_1$. We call these maps left translation and right translation. Notice that $L_{g_1g_2} = L_{g_1} \circ L_{g_2}$ and $R_{g_1g_2} = R_{g_2} \circ R_{g_1}$.

Definition 19 Let G be an arbitrary Lie group. A Riemannian metric on G is an assignment to each tangent space T_g of G of a positive definite inner product $\langle X, Y \rangle_g$ on T_g . This assignment must vary in a C^∞ manner: if X and Y are C^∞ vector fields, then $\langle X, Y \rangle$ is a C^∞ function on G .

Definition 20 A Riemannian metric on G is left invariant if whenever X and Y are tangent vectors at $g_1 \in G$ and $g \in G$, then $\langle X, Y \rangle_{g_1} = \langle L_g^*(X), L_g^*(Y) \rangle_{gg_1}$. Similarly, a Riemannian metric is right invariant if $\langle X, Y \rangle_{g_1} = \langle R_g^*(X), R_g^*(Y) \rangle_{g_1g}$.

Theorem 24 An arbitrary Lie group has a left-invariant Riemannian metric and a right-invariant Riemannian metric.

Proof: Select a positive-definite inner product $\langle X, Y \rangle$ on $T_e(G)$. If $g \in G$ and $X, Y \in T_g$, define $\langle X, Y \rangle_g = \langle L_{g^{-1}}^*X, L_{g^{-1}}^*Y \rangle_e$. Very simple algebra shows that this metric is left-invariant. Note that if X and Y are left-invariant vector fields, $\langle X, Y \rangle$ is a constant function on G .

We obtain a right-invariant metric in the same way.

5.2 Left and Right Invariant Integrals

We now use these left and right invariant metrics to define right and left invariant integrals of real-valued functions:

Definition 21 *Let G be a Lie group. A left-invariant integral on G is an assignment to each continuous real-valued function with compact support on G of a real number $\int_G f(g) dg$ with the following properties:*

- $\int_G$ is linear in f
- If $f \geq 0$ and f is not identically zero, then $\int_G f(g) dg$ is greater than zero
- $\int_G$ is left-invariant; if $L_g(f)(g_1) = f(gg_1)$, then $\int_G L(g)f(g_1) dg_1 = \int_G f(g_1) dg_1$

Using an obvious modification, we define right-invariant integrals on G .

Theorem 25 *If G is a Lie group, it has a left-invariant integral and a right-invariant integral.*

Lemma 8 *Every Lie group is orientable.*

Proof: Select a basis $X_1, \dots, X_n$ for $T_e(G)$. Let this basis define an orientation at the origin. Use left translation to extend each X_i to a left invariant vector field, and let the corresponding basis of $T_g(G)$ define the orientation at g .

Proof of theorem: It is easiest to give the proof "backward." Fix an orientation. Find a covering of G by oriented coordinate neighborhoods $\mathcal{U}_i$. Find a partition of unity φ_j subordinate to the $\mathcal{U}_i$. Recall that this means that each φ_j has compact support entirely contained in some $\mathcal{U}_i$, and every point has an open neighborhood on which only finitely many φ_j are non-zero, and $\sum \varphi_j = 1$. Define

$$\int_G f = \sum \int_{\mathcal{U}_i} \varphi_i f$$

To finish the argument, we must explain how to integrate a continuous function with compact support in a coordinate set, and show that such that the integral is invariant under coordinate changes, and under right translation.

Definition 22 *Let $g(X, Y)$ be a Riemannian metric on an orientable C^∞ manifold. Suppose $\mathcal{U}$ is a coordinate neighborhood and f is a continuous function with compact support inside $\mathcal{U}$. Define*

$$\int_{\mathcal{U}} f = \int \dots \int f(x_1, \dots, x_n) \sqrt{\det \left(g \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right) \right)} dx_1 \dots dx_n$$

Lemma 9 *This integral is independent of the choice of coordinates.*

Proof: Suppose $y_1, \dots, y_n$ are new coordinates, each a function of $x_1, \dots, x_n$. Then

$$\frac{\partial}{\partial x_i} = \sum \frac{\partial y_j}{\partial x_i} \frac{\partial}{\partial y_j}$$

and the above expression equals

$$\begin{aligned} & \int \dots \int f(y_i(x_1, \dots, x_n)) \sqrt{\det \left(g \left(\sum \frac{\partial y_k}{\partial x_i} \frac{\partial}{\partial x_i}, \sum \frac{\partial y_l}{\partial x_j} \frac{\partial}{\partial x_j} \right) \right)} dx_1 \dots dx_n \\ &= \int \dots \int f(y_i(x_1, \dots, x_n)) \sqrt{\det^2 \left(\frac{\partial y_k}{\partial x_i} \right)} \sqrt{\det \left(g \left(\frac{\partial}{\partial y_k}, \frac{\partial}{\partial y_l} \right) \right)} dx_1 \dots dx_n \\ &= \int \dots \int f(y_i(x_1, \dots, x_n)) \left| \det \left(\frac{\partial y_k}{\partial x_i} \right) \right| \sqrt{\det \left(g \left(\frac{\partial}{\partial y_k}, \frac{\partial}{\partial y_l} \right) \right)} dx_1 \dots dx_n \end{aligned}$$

If the first determinant is positive, we can remove the absolutely value signs and get

$$\begin{aligned} & \int \dots \int f(y_i(x_1, \dots, x_n)) \det \left(\frac{\partial y_k}{\partial x_i} \right) \sqrt{\det \left(g \left(\frac{\partial}{\partial y_k}, \frac{\partial}{\partial y_l} \right) \right)} dx_1 \dots dx_n \\ &= \int \dots \int f(y_1, \dots, y_n) \sqrt{\det \left(g \left(\frac{\partial}{\partial y_i}, \frac{\partial}{\partial y_j} \right) \right)} dy_1 \dots dy_n \end{aligned}$$

Lemma 10 *If g is left-invariant, this integral is right-invariant.*

Proof: Choose an orthonormal basis $e_1, \dots, e_n$ for $T_e(G)$. Then $L_g^*(e_i)$ is an orthonormal basis for $T_g(G)$. Let $x_1, \dots, x_n$ be a coordinate patch, and write

$$\frac{\partial}{\partial x_i} = \sum A_{ij}(x_1, \dots, x_n) L^*(e_j)$$

Then our integral is

$$\int \dots \int f \sqrt{\det \sum_k A_{ik} A_{jk}} dx_1 \dots dx_n$$

The sum is AA^T and so the integral is $\int \dots \int f |\det A| dx_1 \dots dx_n$.

Notice that L_{g_1} and R_{g_2} commute. It follows that R_g is an isometry of G with its left-invariant Riemannian metric. Therefore, it suffices to prove that

$$|\det A(x_1, \dots, x_n)| dx_1 \dots dx_n$$

is right invariant.

Therefore, fix g and consider R_g as a map from an open set $\mathcal{U} \subset G$ to a coordinate neighborhood $\mathcal{V} \subset G$. Let $x_1, \dots, x_n$ be coordinates in $\mathcal{V}$. The map R_g then introduces coordinates $y_1, \dots, y_n$ in $\mathcal{U}$. These coordinates are actually the same as the x_i , so we distinguish them by where expressions are evaluated. For y_i and $\mathcal{U}$ we evaluate at g_1 and for x_i and $\mathcal{V}$ we evaluate at g_1g .

In $\mathcal{V}$ we compute the matrix A via $A_{ij} = \langle \frac{\partial}{\partial x_i}, L^*(e_j) \rangle$, evaluating all this at g_1 . In $\mathcal{U}$ we compute the matrix A by $A_{ij} = \langle \frac{\partial}{\partial y_i}, L^*(e_j) \rangle$, evaluating at g_1g . Since R_g is an isometry, these expressions are equal except that one is evaluated at g_1 and the other at g_1g . So

$$\int_{\mathcal{U}} f |det A| dy_1 \dots dy_n = \int_{\mathcal{V}} f \circ R_g |det A| \circ R_g dx_1 \dots dx_n$$

Similarly, if we choose a right invariant metric, we get a left invariant integral.

Theorem 26 *Suppose G is compact. Then the integral defined above is both left and right invariant.*

Proof: We will show that our left invariant and right invariant integrals are equal, and thus they are both left and right invariant.

The first integral is $\int f |det A|$ and the second is $\int f |det B|$; in the first case A expresses $\frac{\partial}{\partial x_i}$ in terms of $L_g^*(X_i)$ and in the second case, B expresses $\frac{\partial}{\partial x_i}$ in terms of $R_g^*(X_i)$. Consequently, the map $R_{g^{-1}}^* B A^{-1} L_g^*$ maps X_i to X_i and thus is the identity, so $B A^{-1} = R_g \star L_{g^{-1}}^* = Ad(g)$. It follows that $|det B| = |det A| |det Ad(g)|$.

However $|det Ad(g)|$ is a group homomorphism from G to the multiplicative group of positive real numbers. Since G is compact, its image is compact. The only compact subgroup is $\{1\}$. So $|det A| = |det B|$ and the two integrals are equal. QED.

Remark: If f is continuous on G and not identically zero, the integral is clearly positive. If G is compact, we can clearly normalize the integral so $\int_G 1 dg = 1$.

Chapter 6

Metrics That Are Both Left and Right Invariant

6.1 Existence

We previously proved

Theorem 27 *An arbitrary Lie group has a left-invariant Riemannian metric and a right-invariant Riemannian metric.*

Proof: Select a positive-definite inner product $\langle X, Y \rangle$ on $T_e(G)$. If $g \in G$ and $X, Y \in T_g$, define $\langle X, Y \rangle_g = \langle L_{g^{-1}}^* X, L_{g^{-1}}^* Y \rangle_e$. Very simple algebra shows that this metric is left-invariant. Note that if X and Y are left-invariant vector fields, $\langle X, Y \rangle$ is a constant function on G .

Similar results hold if we extend to vector fields using right translation.

Theorem 28 *If G is compact, it has a Riemannian metric which is both left and right invariant.*

Proof: Let us determine when the left-invariant and right-invariant metrics are equal. Fix $g \in G$ and $X, Y \in T_g$. Then we want $\langle L_{g^{-1}}^* X, L_{g^{-1}}^* Y \rangle_e = \langle R_{g^{-1}}^* X, R_{g^{-1}}^* Y \rangle_e$. Write $X \in T_g$ as $L_g^* \tilde{X}$ and $Y = L_g^* \tilde{Y}$ where $\tilde{X}, \tilde{Y} \in T_e$. Then the previous equation reads

$$\langle \tilde{X}, \tilde{Y} \rangle_e = \langle R_{g^{-1}}^* L_g^* \tilde{X}, R_{g^{-1}}^* L_g^* \tilde{Y} \rangle_e$$

Recall a piece of standard notation. The map $g_1 \rightarrow gg_1g^{-1}$ is a group automorphism of G , and thus induces a Lie algebra automorphism $Ad_g : \mathcal{G} \rightarrow \mathcal{G}$. Thus we obtain a Lie group

homomorphism

$$Ad : G \rightarrow Aut(\mathcal{G})$$

which induces a corresponding Lie algebra homomorphism

$$ad : \mathcal{G} \rightarrow Hom(\mathcal{G}, \mathcal{G})$$

In the standard theory, it is proved that $ad_X(Y) = [X, Y]$.

The previous two paragraphs show that to get a metric which is both left and right invariant, we need to start with an inner product $\langle X, Y \rangle_e$ on $\mathcal{G}$ such that

$$\langle X, Y \rangle_e = \langle Ad_g(X), Ad_g(Y) \rangle_e$$

for all $X, Y \in \mathcal{G}$.

To obtain such an inner product, start with an arbitrary inner product $\langle\langle X, Y \rangle\rangle$ on $T_e(G)$ and form

$$\langle X, Y \rangle_e = \int_G \langle\langle Ad_g(X), Ad_g(Y) \rangle\rangle dg$$

. We must prove that the above expression is equal to

$$\begin{aligned} \langle Ad(g_1)X, Ad(g_1)Y \rangle &= \int_G \langle\langle Ad(g)Ad(g_1)X, Ad(g)Ad(g_1)Y \rangle\rangle dg \\ &= \int_G \langle\langle Ad(gg_1)X, Ad(gg_1)Y \rangle\rangle dg \end{aligned}$$

But this expression is

$$\int_G R_{g_1} \langle\langle Ad(g)X, Ad(g)Y \rangle\rangle dg$$

and our integral is right invariant.

Remark: In recent sections we have proved the existence of a left and right invariant integral on a compact Lie group, and the existence of a left and right invariant metric on such a group. But notice that there is a big difference in these results. Any left invariant integral is automatically both left and right invariant if the group is compact. But there are many, many left invariant metrics, and we had to work hard to construct one of them which is also right invariant.

6.2 Consequences

Remark: The basic differential geometry used below is from my notes from a Differential Geometry course at the University of Oregon. See also Milnor's wonderful chapter on differential geometry in his book on Morse Theory. Milnor's book is now on the internet; see <http://www.maths.ed.ac.uk/~aar/papers/milnmors.pdf>

Theorem 29 *Suppose G is connected. An inner product on $\mathcal{G} = T_e(G)$ induces a left and right invariant metric if and only if for all $X, Y, Z \in \mathcal{G}$*

$$\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0$$

Proof: Let $X(t)$ be the one-parameter subgroup generated by X and suppose

$$\langle \text{Ad}_{X(t)} Y, \text{Ad}_{X(t)} Z \rangle = \langle Y, Z \rangle$$

Differentiate with respect to t at $t = 0$. We obtain

$$\langle \frac{d}{dt} \text{Ad}_{X(t)} Y, Z \rangle + \langle Y, \frac{d}{dt} \text{Ad}_{X(t)} Z \rangle = 0$$

or equivalently

$$\langle \text{ad}_X(Y), Z \rangle + \langle Y, \text{ad}_X(Z) \rangle = 0$$

Since $\text{ad}_X(Y) = [X, Y]$, this is the desired equation.

Conversely, from the indicated equation we find that $\frac{d}{dt} \langle \text{Ad}_{X(t)} Y, \text{Ad}_{X(t)} Z \rangle = 0$. Since $\text{Ad}_{X(t_0+t)} Y = \text{Ad}_{X(t)} \text{Ad}_{X(t_0)} Y$, we conclude that the previous derivative is always zero, so $\langle \text{Ad}_{X(t)} Y, \text{Ad}_{X(t)} Z \rangle$ is constant. When $t = 0$ it equals $\langle Y, Z \rangle$, so it always equals $\langle Y, Z \rangle$. Thus the inner product is invariant under a neighborhood of the identity, and consequently under the entire group. QED.

Theorem 30 *Let $\nabla_X Y$ be the unique connection associated with a left and right invariant metric on a Lie group. If X and Y are left invariant, $\nabla_X Y = \frac{1}{2}[X, Y]$.*

Proof: From page 92 of the differential geometry notes we have

$$\begin{aligned} 2 \langle \nabla_X Y, Z \rangle &= X \langle Y, Z \rangle + Y \langle X, Z \rangle - Z \langle X, Y \rangle \\ &\quad - \langle X, [Y, Z] \rangle - \langle Y, [X, Z] \rangle + \langle Z, [X, Y] \rangle \end{aligned}$$

If all of these vector fields are left invariant, then for instance $X \langle Y, Z \rangle = 0$ because $\langle Y, Z \rangle$ is constant. So the formula reduces to

$$\begin{aligned} 2 \langle \nabla_X Y, Z \rangle &= - \langle X, [Y, Z] \rangle - \langle Y, [X, Z] \rangle + \langle Z, [X, Y] \rangle \\ &= \langle X, [Z, Y] \rangle + \langle Y, [Z, X] \rangle + \langle [X, Y], Z \rangle = \langle [X, Y], Z \rangle \end{aligned}$$

The result follows.

Theorem 31 *If a compact Lie group has a left and right invariant metric, then geodesics through the origin and one-parameter subgroups are the same thing.*

Proof: Recall that a curve $\gamma(t)$ is a geodesic if $\nabla_{\gamma'(t)}\gamma'(t) = 0$. If $X(t)$ is a one-parameter subgroup, its derivatives $X'(t)$ follow a left-invariant vector field X , and then $\nabla_{X'(t)}X'(t) = \frac{1}{2}[X, X] = 0$, so the one-parameter group is a geodesic. Conversely, let $\gamma(t)$ be a geodesic through the origin. Let $X = \gamma'(0)$ and let $X(t)$ be the one-parameter subgroup with this derivative. Then $X(t)$ is a geodesic with $X'(0) = \gamma'(0)$, but such geodesics are unique, so $X = \gamma$. QED.

Remark: A key theorem in Riemannian geometry is the Hopf-Rinow theorem. This result is proved in the differential geometry chapter of Milnor's Morse Theory. The theorem says (among other things) that if M is a connected Riemannian manifold and every geodesic segment can be extended to a geodesic defined for all time, then any two points on M can be joined by a geodesic.

If G is a compact, connected Lie group and $\langle \rangle$ is a left and right invariant metric, then geodesics through the origin are one-parameter groups, and thus can be extended to be defined for all t . We earlier explained that if $\langle \rangle$ is left-invariant, then each R_g is an isometry. It follows that the isometry group of a compact Lie group is transitive, and consequently any geodesic can be defined for all t . Consequently, any two points on G can be joined by a geodesic, and thus by a translate of a one-parameter group. So

Theorem 32 *If G is a compact connected Lie group, then every element of G is on some one-parameter subgroup through the origin. In other words, the exponential map $\exp : \mathcal{G} \rightarrow G$ is onto.*

6.3 Lie Algebras of Compact Groups

In this section, we will find a necessary and sufficient condition that a given Lie algebra is the Lie algebra of a compact Lie group.

Definition 23 *If L is a Lie algebra, the Killing Form on L is the expression*

$$K(X, Y) = \text{tr } \text{ad}(X)\text{ad}(Y)$$

Remark: If the ground field for L is complex, the Killing form is complex-valued. When L is a real Lie algebra, the Killing form is real.

Theorem 33 *Let G be a compact Lie group, with Lie algebra $\mathcal{G}$ and left and right invariant Riemannian metric $\langle \rangle$. Let $k = \{ X \in \mathcal{G} \mid \langle X, Y \rangle = 0 \text{ for all } Y \in \mathcal{G} \}$. Then k is an abelian ideal in $\mathcal{G}$. The orthogonal complement $\mathcal{H}$ is also an ideal and $\mathcal{G} = k \oplus \mathcal{H}$ as a direct sum of Lie algebras. Moreover, the Killing form on $\mathcal{H}$ is negative definite.*

Proof: If $X \in k$ and $Y, Z \in \mathcal{G}$, then $0 = \langle X, [Y, Z] \rangle = - \langle [Y, Z], X \rangle$ for all Z , so $[Y, X] \in k$. Thus k is an ideal.

If $X \in k$, $Y \in \mathcal{G}$ and $Z \in \mathcal{H}$, then $\langle X, [Y, Z] \rangle = -\langle [Y, X], Z \rangle = 0$, so the orthogonal complement is an ideal.

Finally, let $H_1, \dots, H_n$ be an orthonormal basis for $\mathcal{H}$. Let $X, Y \in \mathcal{H}$. Then

$$\begin{aligned} K(X, Y) &= \text{tr } \text{ad}(X)\text{ad}(Y) = \sum \langle \text{ad}(X)\text{ad}(Y)H_i, H_i \rangle = - \sum \langle \text{ad}(Y)H_i, \text{ad}(X)H_i \rangle \\ &= - \sum \langle [X, H_i], [Y, H_i] \rangle \end{aligned}$$

and in particular

$$K(X, X) = - \sum \| [X, H_i] \|^2$$

This expression is less than or equal to zero. It is zero only if $[X, H_i] = 0$ for all i , and thus $[X, Y] = 0$ for all $Y \in \mathcal{H}$, but then $[X, Y] = 0$ for all $Y \in \mathcal{G}$ and so $X \in k$ and so $X = 0$. QED.

Remark: The same reasoning easily gives the following result:

Theorem 34 *Let G be a compact Lie group with Lie algebra $\mathcal{G}$. Then*

$$\mathcal{G} = k \oplus \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_n$$

where this is a direct sum of ideals, k is abelian, and each $\mathcal{G}_i$ is a simple Lie algebra over the reals with negative -definite Killing form. Conversely, each such Lie algebra is the Lie algebra of some compact group.

Remark: Notice that k is the Lie algebra of $S^1 \times \dots \times S^1$, which is compact, and R^m , which is not compact.

Proof: The previous proof essentially shows that the Lie algebra of a compact group has the required form. We need only show that each such algebra is the Lie algebra of a compact group. This is obvious for k , so we need only prove it for $\mathcal{G}_i$. We will prove in the following paragraphs that $\mathcal{G}_i$ is the Lie algebra of the set of all Lie algebra automorphisms of $\mathcal{G}_i$. This set is a compact Lie group because a Lie algebra automorphism preserves the Killing form, so the automorphism group is a closed subgroup of $O(n)$ for some n , and thus compact.

We need only show that the automorphism group of a Lie algebra L with negative definite Killing form has Lie algebra L . If φ_t is a one-parameter group of automorphisms, then $[\varphi_t(X), \varphi_t(Y)] = [X, Y]$. Differentiating with respect to t , $[\psi_t^*(X), Y] + [X, \psi_t^*(Y)] = 0$. A linear transformation of L satisfying this condition is called a *derivation*. Conversely, if $D : L \rightarrow L$ is a derivation, then e^{tD} is easily proved to be an automorphism. It follows that the identity component of the group of automorphisms, $\text{Aut}_0(L)$, is a compact Lie group with Lie algebra the set of derivations of L .

There is a natural Lie algebra homomorphism $ad : L \rightarrow Der(L)$. The kernel of this map is zero because if $ad(X)Y = 0$ for all $Y \in L$, then the Killing form satisfies $K(X, Y) = 0$ and so $X = 0$. To finish the argument, we prove that every derivation of L is inner. Then $L = Der(L)$ is the Lie algebra of Aut_0 .

The group $Aut_0(L)$ acts on itself by conjugation, and thus acts on its Lie algebra, $Der(L)$. Consequently, it induces a representation of the Lie algebra of Aut_0 , $Der(L)$, on $Der(L)$. This action is $D(D_1) = [D, D_1]$.

Note that $ad(L) \subset Der(L)$ is an ideal. Indeed, if D is a derivation, then $[D, ad(X)]Y = D[X, Y] - [X, D(Y)] = [D(X), Y]$, so $[D, ad(X)] = adD(X)$. Since Aut_0 is compact, we can find a complementary invariant subspace Der_1 so $Der = ad(L) \oplus Der_1$.

Fix $D \in Der_1$ and apply the previous identity $[D, ad(X)] = ad(D(X))$. This is in $ad(L)$ since $ad(L)$ is invariant under bracket with Der . It is also in Der_1 since Der_1 is invariant under bracket with Der . Consequently it is zero. So $ad(D(X))$ is identically zero, and therefore $D(X) = 0$ for all X . So $Der_1 = 0$ and all derivations of L are inner.

6.4 Lie Algebras over C

We will not use the material in the next two sections elsewhere in these notes.

Readers may have taken a course on Lie algebras. It is likely that this course began with Lie algebras over an arbitrary field, but soon assumed that the ground field was C . For example, a crucial theorem in the subject, due to Lie, asserts that a Lie subalgebra $L \subset gl(V)$ is solvable if and only if there is a new basis of V such that each element of L is upper triangular.

The course probably ended with a complete classification of simple Lie algebras over the complex numbers. We summarize here the theory of Lie algebras over C .

Suppose L is a Lie algebra over C . The Jordan-Holder theorem holds for such algebras. It asserts that every maximal composition series for L has the same length and the same composition factors up to order. Recall that such a composition series is a sequence

$$L_0 \subset L_1 \subset L_2 \dots \subset L_n = L$$

in which each L_{i-1} is an ideal in L_i and each quotient L_i/L_{i-1} has no non-trivial ideals. Such quotients are either abelian and equal to C , or else non-abelian with no non-trivial ideals; the latter algebras are said to be *simple*.

It turns out that composition series can always be arranged so that the C quotients come first. Indeed, a Lie algebra is *solvable* if and only if all its composition quotients are C ; every Lie algebra over C has a unique maximal solvable ideal.

When we quotient out this ideal, we get a Lie algebra with only simple composition quotients. Such an algebra is said to be *semisimple*. A key result, due to Cartan, asserts that an algebra over C is semisimple if and only if its Killing form is nondegenerate. Using this result, every semisimple algebra is a direct sum of its simple quotients.

Finally, these simple algebras are completely classified. They are known by the letters a through g , as $a_n, b_n, c_n, d_n, e_6, e_7, e_8, f_4$, and g_2 .

Incidentally, classifying solvable algebras is essentially impossible. There are uncountable many of them, and even in low dimensions, there is little order in the various possibilities.

6.5 Connection between the Compact and Complex Cases

Suppose L is a Lie algebra over R . Then $L \otimes C$ is a Lie algebra over C , constructed by picking a basis over R and allowing the coordinates to be complex. The Killing form of $L \otimes C$ is the Killing form of L , extended in the obvious way. In particular, if the Killing form of L is negative-definite, then the Killing form of $L \otimes C$ is non-degenerate, so $L \otimes C$ is a direct sum of the known simple algebras over C . However, if L is a complex Lie Algebra, then in most cases there are several non-isomorphic real Lie algebras L_i such $L_i \otimes C = L$.

In section 3.3, we proved that the Lie algebra of a compact Lie group is $k \oplus \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_n$. It follows that compactness completely eliminates the very complicated solvable radical k unless it is abelian and thus easily understood.

Turning to the simple pieces, the key theorem asserts that tensoring with C sets up a one-to-one correspondence between simple real algebras of compact groups, and simple complex Lie algebras. So while a complex simple algebra has many real forms, exactly one of them up to isomorphism has negative-definite Killing form. Thus the complex simple algebras $a_n, b_n, c_n, d_n, e_6, e_7, e_8, f_4, g_2$ correspond to simple algebras of compact groups $A_n, B_n, C_n, D_n, E_6, E_7, E_8, F_4, G_2$.

This is proved in two stages. The classification theorem over C selects an abelian subalgebra $h \otimes C$ of L , and then selects additional basis vectors $\{e_\alpha, e_{-\alpha}\}$, the so-called root vectors. These root vectors come in pairs, as indicated. We find a new basis by choosing a real basis of ih and adding $\frac{e_\alpha + ie_{-\alpha}}{2}$ and $\frac{e_\alpha - ie_{-\alpha}}{2}$. A short calculation shows that all structure constants are real, so these basis vectors generate a real subalgebra. Another calculation shows that the Killing form is negative-definite on this algebra.

In particular, if L is simple over C , it has the form $L_R \otimes C$ where L_R is a real Lie algebra. Even this step is not true in greater generality, because there are complex Lie algebras with complex structure constants for which no basis exists giving real structure constants.

The more difficult result says that if $L_1 \otimes C$ and $L_2 \otimes C$ are isomorphic simple complex algebras with negative-definite Killing form, then L_1 and L_2 are isomorphic.

There is a final deep theorem. Suppose G is a compact Lie group whose Lie algebra $\mathcal{G}$ has negative-definite Killing form. It can be proved that G has finite fundamental group, and consequently the universal covering of G is still compact. It follows from this theorem that if G is an arbitrary compact Lie group, then the universal cover of G has the form

$$R^k \times G_1 \times \dots \times G_n$$

where the terms are unique up to order and the G_i are $A_n, B_n, C_n, D_n, E_6, E_7, E_8, F_4$, or G_2 . Incidentally, $A_n = SU(n)$, $B_n = Spin(2n)$, $C_n = Sp(n)$, and $D_n = Spin(2n+1)$.

6.6 Curvature and Myers' Theorem

Although the curvature tensor is computed using vector *fields*, in the end its value at p depends only on the values of these fields at p . Consequently, we can simplify calculations on a compact Lie group by assuming that all vector fields are left invariant.

In that case, we already know that $\nabla_X Y = \frac{1}{2}[X, Y]$. The curvature tensor is defined to be

$$R(X, Y, Z) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$$

For compact Lie groups, this becomes

$$\begin{aligned} & \frac{1}{4}[X, [Y, Z]] - \frac{1}{4}[Y, [X, Z]] - \frac{1}{2}[[X, Y], Z] \\ &= \frac{1}{4}[[X, Y], Z] + \frac{1}{4}[Y, [X, Z]] - \frac{1}{4}[[Y, X], Z] - \frac{1}{4}[X, [Y, Z]] - \frac{1}{2}[[X, Y], Z] \\ &= \frac{1}{4}([Y, [X, Z]] - [X, [Y, Z]]) = -\frac{1}{4}[[X, Y], Z] \end{aligned}$$

Remark: The sectional curvature of a 2-dimensional subspace of a tangent space is defined to be

$$\langle R(X, Y, Y), X \rangle$$

where $\{X, Y\}$ is an orthonormal basis of the subspace. In our case, this is

$$-\frac{1}{4} \langle [[X, Y], Y], X \rangle = \frac{1}{4} \langle [[Y, [X, Y]], X \rangle = -\frac{1}{4} \langle [X, Y], [Y, X] \rangle$$

An easy theorem shows that this is independent of the orthonormal basis.

Theorem 35 *If $\langle, \rangle$ is a left and right invariant inner product on a compact Lie group, the sectional curvature of a 2 dimensional plane through the origin with orthonormal basis X, Y is*

$$\frac{1}{4} \|[X, Y]\|^2$$

Remark: Notice that this expression is greater than or equal to zero, and zero exactly when the plane is an abelian subalgebra.

It is useful to define an "average sectional curvature."

Definition 24 *The Ricci Curvature tensor is defined by the following sum over an orthonormal basis of the tangent space:*

$$K(X, Y) = \text{tr} \left(Z \rightarrow R(X, Z)Y \right) = \sum_i \langle \nabla_X \nabla_{U_i} Y - \nabla_{U_i} \nabla_X Y - \nabla_{[X, U_i]} Y, U_i \rangle$$

Theorem 36 *If X has unit length, then $K(X, X)$ can be obtained by extending to an orthonormal basis $X, X_2, \dots, X_n$ and computing the sum of the sectional curvatures of the planes $(X, X_2), (X, X_3), \dots, (X, X_n)$*

Proof: An easy calculation.

In his book on Morse theory, Milnor proves the following

Theorem 37 (Myers) *If M is a Riemannian manifold of dimension n and $K(X, X) \geq \frac{n-1}{r^2}$, for all unit vectors X , then no geodesic of length greater than πr is minimal.*

Corollary 4 *If M is complete and satisfies the conditions of the previous theorem, then M is compact and the distance between any two points is at most πr .*

Proof: By Hopf-Rinow, any two points of a complete space can be joined by a geodesic whose length is the distance between them.

Corollary 5 *If M is a compact Riemannian manifold and $K(X, X)$ is positive for all non-zero X , then the fundamental group of M is finite and the universal cover $\bar{M}$ is also compact.*

Proof: If $K(X, X) > 0$ for all non-zero X at a point p , then $K(X, X)$ has a positive minimum on vectors at X with unit length. Since $K(X, X)$ is continuous and M is compact, $K(X, X)$ has a positive minimum as X runs over all unit tangent vectors on M . Since the universal cover is locally isometric to M , this remains true on the universal cover. Hence there is an r making the hypothesis of Myer's theorem true, and the result follows.

Theorem 38 *Suppose G is a compact Lie group with semi-simple Lie algebra. Then the fundamental group of G is finite and the universal cover of G is also compact.*

Remark: Recall that $\mathcal{G} = \mathfrak{k} \oplus \mathcal{G}_\infty \oplus \dots \oplus \mathcal{G}_\lambda$. According to the above theorem, each $\mathcal{G}_\lambda$ is the Lie algebra of a simply-connected compact group. Obviously this is false for the entire group if $\mathfrak{k} \neq \{0\}$ since then it is the Lie algebra of R^n for some $n > 0$.

Proof: From earlier calculations, the Ricci curvature $K(X, X)$ is positive on unit vectors unless X commutes with all elements of the Lie algebra. Such an X belongs to $\mathfrak{k}$.

6.7 Weyl's Theorem on Complete Reducibility

Hermann Weyl proved the following crucial theorem.

Theorem 39 (Weyl's Unitary Trick) *Let G be a connected Lie group with semisimple Lie algebra. Then if $\phi : G \rightarrow GL(V)$ is a representation of G on a complex vector space and $W \subset V$ is an invariant subspace, there is a complementary invariant subspace. Consequently, any finite dimensional complex representation of G is a sum of irreducible representations, and this sum is unique up to order.*

Remark: This is very surprising, since for easy non-compact groups like R the result is false, and there are many non-compact semisimple groups: $SL(n, R)$, etc.

Proof: Weyl's proof used an ingenious trick, reproduced below. Later on, the algebraists discovered a purely algebraic proof, but it isn't as much fun.

A complex representation of G induces a complex representation of the Lie algebra $\mathcal{G}$ of G ; note that $\mathcal{G}$ is a Lie algebra over R , so the representation map is real linear to $gl(n, C)$. We can tensor with C and extend the representation to get a complex-linear representation of the complex Lie algebra $\mathcal{G} \otimes C$. This Lie algebra is still semisimple since its Killing form is negative definite.

By a result quoted earlier, this complex Lie algebra has a compact form, i.e., there is a real Lie algebra $\mathcal{G}_R$ with negative-definite Killing form such that $\mathcal{G}_R \otimes C$ is isomorphic to $\mathcal{G} \otimes C$. Restricting our representation to $\mathcal{G}_R$ gives a real-linear map to $gl(n, C)$ which is a representation of $\mathcal{G}_R$. The Lie algebra $\mathcal{G}_R$ is the Lie algebra of a compact Lie group G_R . By Myer's theorem, the universal cover of G_R is still compact. So the algebra $\mathcal{G}_R$ is the Lie algebra of a compact, simply-connected group $\bar{G}_R$. Since this group is simply-connected, the Lie algebra representation comes from a Lie group representation of the group. Since the group is compact, the subset V has a complementary invariant subspace W . Tracing backward, the subspace W is invariant for the original representation as well. QED.

Chapter 7

Maximal Tori

7.1 Definition; Conjugacy Theorem

Recall that a Lie subgroup H of a Lie group G need not have the induced topology. If G is the torus R^2/Z^2 , and H is the image of a straight line through the origin, then this image has the induced topology only if it is a circle. In that case, the line goes through a second lattice point (m, n) and thus has rational slope $\frac{n}{m}$. But otherwise the line maps one-to-one to a Lie subgroup which is dense in the torus. As a Lie subgroup, this subgroup has the topology of a line rather than the induced topology.

Closed subgroups of Lie groups are easier to study. Each such group is itself a Lie group with the induced topology.

A *torus* in a Lie group G is a Lie subgroup isomorphic to the standard torus. Such a subgroup is automatically compact, hence closed, so it has the induced topology.

Definition 25 *A maximal torus in a compact group G is a Lie subgroup isomorphic to a torus, which is not a subgroup of a larger torus. Equivalently, it is a maximal connected abelian Lie subgroup of G .*

Remark: The second statement implies the first, because if H is abelian, so is its closure $\overline{H}$, but this closure is closed, hence compact abelian, hence a torus.

Remark: A maximal torus always exists. Indeed let X be a nonzero element of the Lie algebra of G . Then X generates a one-dimensional subalgebra, and by the fundamental theorem of Lie theory, there is a corresponding one-dimensional Lie subgroup. This group is either a circle or a line, hence abelian. Its closure is a torus. Thus there is at least one torus in G .

Each torus is a Lie subgroup with the induced topology, and has a corresponding Lie

subalgebra. An increasing chain of tori corresponds to an increasing chain of subalgebras, which must end because the dimensions of these subalgebras strictly increase.

Remark: Consider the group $U(n)$. It consists of complex matrices A such that $\bar{A}^T A = 1$. This set is clearly a closed subgroup of the full matrix group, hence a Lie group. The last condition implies that $\sum_i \bar{a}_{ik} a_{ik} = 1$ and thus $|a_{ik}| \leq 1$, so the subset is bounded and thus compact.

The diagonal matrices in this group have the form

$$\begin{pmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & e^{i\theta_n} \end{pmatrix}$$

and thus form a torus of dimension n . This group is a maximal torus because if A is any other element of $U(n)$, it has some nonzero off-diagonal element a_{jk} and the commutator of A with the diagonal element above has jk th element $(e^{i\theta_j} - e^{i\theta_k}) a_{jk}$ and this is nonzero if $\theta_j \neq \theta_k$.

In $U(n)$, every element is conjugate to a diagonal matrix. Indeed since $\bar{A}^T A = 1$, $\bar{A}^T$ and A commute, and this is the necessary and sufficient condition that A can be diagonalized by choosing a new orthonormal basis. This result generalizes and indeed

Theorem 40 *Let T be a maximal torus in a connected compact group G . Then every $g \in G$ is conjugate to an element of T .*

Remark: There are several beautiful ways to prove this result, but they all involve knowing some deep result about geometry or topology. So we postpone the proof until the end of this chapter and first draw important consequences from it.

7.2 Uniqueness of Maximal Tori up to Conjugacy

If T is a maximal torus of G and $g \in G$, then clearly $g^{-1}Tg$ is another maximal torus, since $g_1 \rightarrow g^{-1}g_1g$ is an automorphism of G which preserves everything.

The component groups G_i of a general compact simply-connected group $G = G_1 \times \dots \times G_n$ are all simple groups in the sense that they have no non-trivial connected normal subgroups. Their maximal tori cannot possibly be unique, because then the maximal torus would be normal. But maximal tori are *almost* unique:

Theorem 41 *Any two maximal tori of a compact group G are conjugate. In particular, they have the same dimension. This dimension is called the rank of G .*

Proof: The idea of the proof is easy. We will show that each torus T contains $t \in T$ whose powers are dense in T . Pick a second torus T_1 and apply theorem 1 to t and T_1 . Then there exists $g \in G$ with $g^{-1}tg \in T_1$ and thus $g^{-1}Tg \subseteq T_1$. Then $g^{-1}Tg$ must be all of T_1 , for if it is strictly smaller and $g^{-1}Tg < T_1$, then $T < gT_1g^{-1}$, contradicting the assumption that T is maximal.

To find t we can suppose that $T = R/Z \times \dots \times R/Z$. The topology of this space has a countable basis $\mathcal{U}_1, \mathcal{U}_2, \dots$. We will define a series of closed non-empty cubes $C_1 \supset C_2 \supset C_3 \supset \dots$ and positive integers $N_1, N_2, N_3, \dots$ such that $C_k^{N_k} \subset \mathcal{U}_k$. If so, we are done. Indeed, $\cap C_i$ is non-empty and thus contains some $t \in T$. Since $t \in C_k$, $t^{N_k} \in \mathcal{U}_k$ for all k . So each basis set contains a power of t , and thus these powers are dense.

We construct the C_k inductively. Choose C_1 any nonempty closed cube in $\mathcal{U}_1$ and choose $N_1 = 1$. Now suppose C_k and N_k are known. This C_k is a cube whose sides have length L ; choose an integer N such that $LN > 1$. Then the interior of NC_k is an open cube whose sides have length greater than 1, so this open cube has representatives for all points in the torus. In particular, the interior of C_k has a point p such that Np represents a point in $\mathcal{U}_{k+1}$. Choose C_{k+1} to be a closed cube inside C_k containing p , and such that NC_{k+1} only represents points in $\mathcal{U}_{k+1}$, and choose $N_{k+1} = N$. QED

Theorem 42 *Suppose G is compact and connected. A maximal torus in G is a maximal abelian subgroup of G .*

Remark: The converse is false, since the set of diagonal matrices in $SO(n)$ is maximal abelian, but contains only ± 1 on the diagonal and thus is not connected.

Proof: The proof falls into two pieces. In the first half, we ignore the connectivity of G .

Suppose T is a maximal torus and $a \in G$ commutes with every element of T . Let H be the closure of the subgroup of G generated by T and a . This group is abelian, so the identity component of its closure is a torus, T_1 . The group H/T_1 is compact and discrete, so finite.

Since every element of H is a limit of expressions of the form $a^k t$, every element of the discrete space H/T_1 is a limit of elements represented by powers of a , but limits in a discrete space are ultimately constant, a must generate the finite quotient. It follows that H/T_1 is isomorphic to Z_m for some m .

Find $t \in T_1$ such that T_1 is the closure of the set of powers of t . Then $a^m \in T_1$ and $ta^{-m} \in T_1$. Since T_1 is a torus, there is an $s \in T_1$ such that $s^m = ta^{-m}$ and thus $(sa)^m = t$. Define g to be sa . Then powers of g^m are dense in T_1 .

Notice that $g^{m+1} = g^m g = g^m sa = (g^m s)a$. Since the elements of the form $g^m s$ are dense in T_1 , the powers of g are also dense in the coset aT_1 . Similarly $g^{m+2} = g^m g^2 = g^m sasa =$

$(g^m ss)a^2$ and these elements are dense in the coset a^2T_1 . Etc. In short, powers of g are dense in H .

Now finally, assume that G is connected. Then a conjugate of g belongs to T , or equivalently g belongs to a conjugate of T . Since g generates all of H , all of H belongs to this conjugate of T . In particular, both T and a belong to the conjugate. Since the conjugate is also a maximal torus, it equals T and thus $a \in T$.

7.3 The Weyl Group

Fix a maximal torus $T \subseteq G$. By definition, the normalizer N of T is the set of all $g \in G$ such that $g^{-1}Tg \subseteq T$. Clearly this is a subgroup of G , which acts on T by conjugation. Note that $T \subseteq N$ and T acts trivially on T . We define the *Weyl group*

$$W = N/T$$

This group acts on T by conjugation.

Theorem 43 *Let T be a maximal torus in a compact, connected G .*

- *W acts faithfully on T , that is, if $w \in W$ acts as the identity, then $w = e$.*
- *W is a finite group*
- *Two elements $t_1, t_2 \in T$ are conjugate if and only if there exists $w \in W$ which maps t_1 to t_2 .*

Proof: Suppose $a \in W$ acts trivially. Then a commutes with every element of T , so the subgroup generated by a and T is abelian. Since T is maximal abelian, $a \in T$. So a represents the identity of W .

We prove that W is finite as follows. Note that N is a compact subgroup of G . Let N_0 be its connected component of the identity, which is a closed (and open) subgroup. Then N/N_0 is discrete and compact, hence finite. Clearly $T \subset N_0$. We will prove these equal.

The proof that $T = N_0$ is quite straightforward. The automorphism group of T acts discretely, since $T = R^n/Z^n$ and an automorphism of T induces an automorphism of R^n which maps Z^n to Z^n . If N_0 is larger than T , then its Lie algebra is also larger, so there is a one-parameter subgroup of N_0 not in T . The inner automorphisms of T induced by the elements of this one-parameter subgroup must vary continuously and thus not at all. So the elements of the one-parameter subgroup commute with T and thus belong to T , a contradiction.

Finally, we prove the last assertion. It suffices to prove that if $t_1, t_2 \in T$ are conjugate under G , then they are conjugate under $N(T)$. Let H be the set of all elements of G which

commute with t_2 . This set is a closed subgroup of G containing T . Suppose $gt_1g^{-1} = t_2$. Then H also contains gTg^{-1} because $gtg^{-1}t_2 = gtg^{-1}gt_1g^{-1} = gtt_1g^{-1} = gt_1tg^{-1} = gt_1g^{-1}gtg^{-1} = t_2gtg^{-1}$. It follows that T and gTg^{-1} are both maximal tori in the connected component H_0 of H . By the fundamental theorem on maximal tori, these tori are conjugate in H_0 . Consequently there exists $h \in H_0$ such that $hgTg^{-1}h^{-1} = T$. Note that h commutes with t_2 , so $hgt_1g^{-1}h^{-1} = ht_2h^{-1} = t_2$. Finally, $hg \in N$ since $(hg)T(ht)^{-1} = T$, so hg induces an element of W . QED.

7.4 Two Proofs of the Conjugacy Theorem

We now return to the central conjugacy theorem:

Theorem: *Let T be a maximal torus in a connected compact group G . Then every $g \in G$ is conjugate to an element of T .*

We will give one complete proof, and then sketch a proof using the Lefschetz Fixed Point formula.

Theorem 44 *Let G be a compact connected Lie group with Lie algebra $\mathcal{G}$. Let $X, Y \in \mathcal{G}$. Then there exists $g \in G$ such that $[Ad_g(X), Y] = 0$.*

Proof: Select a left and right invariant metric $\langle, \rangle$ on G and $\mathcal{G}$. Consider the map $f : G \rightarrow \mathbb{R}$ by $f(g) = \langle Ad_g(X), Y \rangle$. This function is continuous, so it has a maximum value at g_0 . Replace X by $Ad_{g_0}X$; then $\langle Ad_g(X), Y \rangle$ has a maximum at the identity in G .

If $Z \in \mathcal{G}$, then $\langle Ad(e^{tZ})(X), Y \rangle$ has a maximum at $t = 0$, so the derivative at $t = 0$ is zero. Therefore $\langle [Z, X], Y \rangle = 0$ and so $\langle [X, Z], Y \rangle = 0$. Since the inner product is left and right invariant, $\langle Z, [X, Y] \rangle = 0$. This holds for all Z , so $[X, Y] = 0$ as desired. QED.

Corollary 6 *Any two maximal tori are conjugate.*

Proof: Let T_1, T_2 be maximal tori. Find $\xi_1 \in T_1$ whose powers are dense in T_1 , and ξ_2 in T_2 whose powers are dense in T_2 . Every element of a torus is on a one-parameter subgroup of the torus, since the universal cover is $\mathbb{R}^n$ and the one-parameter subgroups of $\mathbb{R}^n$ are $t \rightarrow tX$. Choose one-parameter subgroups $X(t)$ and $Y(t)$ for T_1 and T_2 through ξ_1 and ξ_2 . Find g such that $[Ad_g(X), Y] = 0$.

Replace $X(t)$ by $gX(t)g^{-1}$ and ξ_1 by $g\xi_1g^{-1}$ and T_1 by gT_1g^{-1} . Then X and Y commute and generate a Lie algebra whose associated connected Lie subgroup is abelian and contains ξ_1 and ξ_2 . The closure of this group is compact, connected, and abelian. Since it contains ξ_1 and ξ_2 , it contains T_1 and T_2 . By maximality of these tori, it equals T_1 and T_2 . QED.

Theorem 45 *Let T be a maximal torus in a connected compact Lie group G . Then every $g \in G$ is conjugate to an element of T .*

Proof: The only difficult step is in this paragraph. We claim there is a one-parameter subgroup $X(t)$ of G containing g . Indeed, we proved earlier that one-parameter subgroups are the same things as geodesics through the origin for any left and right invariant metric. Since G is compact, it is complete, so any two points of G are connected by a geodesic by the theorem of Hopf and Rinow. So e and g are connected by a geodesic, and thus by a one-parameter subgroup.

Since g is on a one-parameter subgroup, it is in a connected abelian subgroup, and thus in a maximal torus T_1 . This T_1 is conjugate to T , and the conjugation carries g to an element of T . QED.

Remark: We now sketch another proof. This proof does not require knowing that every $g \in G$ is on a one-parameter subgroup. But once we know that $g \in G$ is conjugate to an element of T , this fact is easily proved, because it is true for T and equivalently its universal cover R^n . Indeed the one parameter subgroups of R^n have the form $\gamma(t) = tX$ for a vector X , and every point on R^n is on such a line. The corresponding result for T follows immediately. But if $g_1 g g_1^{-1} = X(t)$, then $g = g_1^{-1} X(t) g_1$.

Remark: The new proof depends on the Lefschetz Fixed Point Theorem. Consider the map $L(g) : G/T \rightarrow G/T$. We prove that this map has at least one fixed point. A fixed point is a coset $g_1 T$ such that $g g_1 T = g_1 T$, or equivalently a g_1 such that $g g_1 = g_1 t$. But then $g_1^{-1} g g_1$ belongs to T because it equals t .

Remark: Suppose M is a compact, oriented C^∞ manifold and $f : M \rightarrow M$ is continuous. Then f induces maps $f^* H^k(M, R) \leftarrow H^k(M, R)$ which depend only on the homotopy class of f . Define the *Lefschetz number*

$$L(f) = \sum_k (-1)^k \text{Tr}(f^* : H^k \leftarrow H^k)$$

Theorem 46 (Lefschetz Fixed Point Theorem) *Suppose $f : M \rightarrow M$ is C^∞ . Then f is homotopic to a C^∞ map with finitely many fixed points, and such that at each fixed point the map $f_* : T_p(M) \rightarrow T_p(M)$ satisfies $\det(I - f_*) \neq 0$. Moreover,*

$$L(f) = \sum_{\text{fixed points } p} \text{sgn}(\det(I - f_*))$$

In particular, if $L(f) \neq 0$ then f has a fixed point because otherwise the above sum over fixed points is zero.

Remark: A second set of my notes, titled *deRham Course*, gives a complete proof of this Lefschetz theorem.

Remark: We apply this theorem by noticing that $L(f)$ is unchanged if we replace f by a homotopic map. But L_g is clearly homotopic to L_e since G is connected. The Lefschetz

number of the identity map L_e is the Euler characteristic of M . To finish the proof, it suffices to show that the Euler characteristic of G/T is not zero. Indeed, more is true: G/T is a cell complex with cells only in even dimensions.

Remark: However, we can obtain the same result from the Fixed Point Theorem without computing the Euler characteristic of G/T . We apply this result to $L_g : G/T \rightarrow G/T$ as before. If this map has infinitely many fixed points, or degenerate fixed points, then there is nothing to prove. Otherwise we must rule out the possibility that f has no fixed points, and that will follow if $L(f) \neq 0$.

Our map L_g is homotopic to L_e , so $L(f)$ is the Euler characteristic of G/T as before. But L_g is also homotopic to L_{g_1} for any other $g_1 \in G$. The new proof selects an ingenious g_1 and then computes the Lefschetz number of L_{g_1} using the second version of the Lefschetz theorem, showing that it is not zero.

Fix a $t_1 \in T$ with the property that the closure of the powers of t_1 is all of T . We are going to determine the set of fixed points of $L_{t_1} : G/T \rightarrow G/T$. A fixed point is a coset gT such that $t_1 gT = gT$, or equivalently $g^{-1}t_1 g \in T$. From this, the conjugate of any power of t_1 is in T , and so $g^{-1}Tg \subset T$. But then $g \in N(T)$, the normalizer of T . Note that replacing g by gt for $t \in T$ does not change the coset, so the fixed points in G/T are cosets wT for $w \in W$, the Weyl group. Since W is finite, there are finitely many fixed points. (Incidentally, we proved that W is finite in a section assuming that maximal tori are conjugate. But a glance back at the proof shows that finiteness of W did not require conjugacy.)

We now claim that each of these fixed points nT produces the same value $\det(I - L_{t_1}^*)$. The map $R_n : G/T \rightarrow G/T$ sends gT to $gTn = gnn^{-1}Tn = gnT$. So $L_g R_n = R_n L_g : G/T \rightarrow G/T$. In particular, $L_{t_1} R_n = R_n L_{t_1}$ and so $R_{n^{-1}} L_{t_1} R_n = L_{t_1}$, where the final L_{t_1} acts on eT and the initial one acts on nT . Since $\det(I - L_{t_1}^*)$ is independent of coordinates, the claim follows.

Thus it suffices to compute on eT .

The Lie group T acts on G by conjugation. Thus T acts on $\mathcal{G}$ by Ad ; since $\mathcal{G}$ is a vector space, this is a representation. By compactness of G , $\mathcal{G}$ has a metric invariant under Ad . This situation will be discussed in detail at the start of the next chapter.

We quote the results here. Let k be the Lie algebra of T . Then $k^\perp$ is invariant under T , and Ad on $k^\perp$ splits as a direct sum of 2-dimensional irreducible real representations. Write this as $\sum \mathcal{R}_i$. For each $\mathcal{R}_i$, there is a group homomorphism $\phi : T \rightarrow R/Z = S^1$ which lifts to a nonzero real-linear ϕ from the Lie algebra of T to R , such that Ad restricted to T on $\mathcal{R}_i$ has the form

$$\begin{pmatrix} \cos 2\pi\phi(t_1) & -\sin 2\pi\phi(t_1) \\ \sin 2\pi\phi(t_1) & \cos 2\pi\phi(t_1) \end{pmatrix}$$

Now consider $L_{t_i} : G/T \rightarrow G/T$, which fixes eT . This map is also $L_{t_i} R_{t_i}^{-1} = \text{Ad}(t_i)$ since this new map sends gT to $t_i g T t_i^{-1} = t_i g T$. The tangent space of G/T is just the complement of the tangent space to T , and thus $\sum \mathcal{R}_i$. So $L_{t_i} = \text{Ad}(t_i)$ is the sum of the above rotations, and its derivative is the same sum. Thus

$$\det(I - L_{t_i}^*) = \prod \det \begin{pmatrix} 1 - \cos 2\pi\phi(t_1) & \sin 2\pi\phi(t_1) \\ -\sin 2\pi\phi(t_1) & 1 - \cos 2\pi\phi(t_1) \end{pmatrix} = \prod (2 - 2\cos 2\pi\phi(t))$$

Notice that the terms of this product are positive because $1 > \cos 2\pi\phi(t)$, unless $\phi(t_1) = 0$. But powers of t_1 are dense in T and ϕ is not identically zero on T , so this cannot happen. QED.

Remark: Since each $\text{sgn } \det(1 - L_{t_i}^*)$ equals 1, the Lefschetz number of our map is $|W|$, so this is the Euler characteristic of G/T .

Chapter 8

Roots

8.1 Introduction

We are going to classify from scratch the Lie algebras of compact groups.

If G is any Lie group, G acts on itself by conjugation. This induces a representation of G on its Lie algebra $\mathcal{G}$, $Ad : G \rightarrow GL(\mathcal{G})$, and this in turn induces a representation of Lie algebras $ad : \mathcal{G} \rightarrow Hom(\mathcal{G}, \mathcal{G})$.

If $H \subset G$ is a subgroup, we can restrict all of these operations to H . If we know the representation theory of H , we can use this theory to decompose $\mathcal{G}$ into irreducible invariant subspaces. We are going to apply this idea twice in this chapter and the next. We first apply it to a maximal torus $T \subset G$, relying on a knowledge of representation theory for abelian groups. This will break up $\mathcal{G}$ into the Lie algebra of T and a large number of two dimensional subspaces. Each two dimensional subspace can be made into a Lie subalgebra by adding an element of the Lie algebra of T , and all of these subalgebras are isomorphic to $su(2)$. We then completely classify representations of $su(2)$. In the following chapter, we use this classification to find significant restrictions on our earlier decomposition of $\mathcal{G}$. These restrictions will ultimately allow us to completely classify these $\mathcal{G}$.

8.2 Irreducible Representations of Abelian Groups

We are going to use earlier results from representation theory. The results we use are so easy that we'll restate and reprove them here.

Theorem 47 *Let $\varphi : G \rightarrow GL(V)$ be a representation of a Lie group G on a complex vector space V .*

- *A one-dimensional complex representation of G is the same thing as a continuous homomorphism $G \rightarrow C^*$. If G is compact, this is the same thing as a continuous homomorphism $G \rightarrow S^1$.*
- *If G is abelian, then every complex irreducible representation is one dimensional.*

Proof: The first result is obvious except for the final assertion. If G is compact and $\varphi : G \rightarrow C^*$ is a homomorphism and $|\varphi(g)| \neq 1$, then replacing g by g^{-1} if necessary, we conclude that $|\varphi(g)| > 1$ and $\varphi(g^n)$ is unbounded, contradicting compactness.

Suppose φ is an irreducible representation of G and $g, g_1 \in G$. Then $\varphi(gg_1) = \varphi(g)\varphi(g_1) = \varphi(g_1)\varphi(g)$ and so $\varphi(g) = \varphi(g_1)\varphi(g)\varphi^{-1}(g_1)$. Thus for each fixed g_1 , $\varphi(g_1)$ is an intertwining operator from V to itself. By Schur's lemma, this intertwining operator is a constant λ times the identity, where the constant depends on g_1 . So $\varphi(g_1) = \lambda(g_1)Id$. Then any subspace is invariant, so an irreducible representation must be one dimensional. QED.

8.3 The Action of a Maximal Torus on $\mathcal{G}$

Let $T \subset G$ be a maximal torus. Then T acts on G by conjugation and this induces a representation Ad of T on the Lie algebra $GL(\mathcal{G})$. This can be extended to a representation of T on the complexified algebra $GL(\mathcal{G} \otimes C)$. Since T is abelian, this representation splits as a sum of one dimensional complex representations. Each of these irreducible representations is given by a homomorphism $f : T \rightarrow S^1$ and a vector $e_f \in \mathcal{G} \otimes C$ with $Ad(t)e_f = f(t)e_f$.

Let us determine the homomorphisms $T \rightarrow S^1$, starting with the case $T = R/Z = S^1$. The universal cover of T is R and we have a commutative diagram as follows:

$$\begin{array}{ccc} R & \xrightarrow{\tilde{f}} & R \\ \downarrow e^{2\pi i r} & & \downarrow e^{2\pi i r} \\ T & \xrightarrow{f} & T \end{array}$$

The most general homomorphism $\tilde{f}$ is $\tilde{f}(r) = kr$ for some fixed k . To fit in this diagram, $\tilde{f}$ must map Z to Z , so k must be an integer. Thus the most general homomorphism is $f(e^{2\pi i r}) = e^{2\pi i kr}$ for a fixed integer k . If we identify an element of T with a complex $z \in S^1$, this becomes the easier $f(z) = z^k$.

In the general case, $T = S^1 \times \dots \times S^1$ and

$$f(z_1, \dots, z_n) = f(z_1, 1, \dots, 1) \dots f(1, 1, \dots, z_n) = z_1^{k_1} \cdot \dots \cdot z_n^{k_n}$$

If all $k_i = 0$, this is the trivial representation acting on some invariant subspace of $\mathcal{G}$. In this case, we claim that the subspace is in the complexification of the Lie algebra $\mathcal{T}$ of T . Indeed, the real and complex parts of a vector in this invariant subspace would be real vectors $v \in \mathcal{G}$ such that $Ad(t)(v) = v$ for all t in $\mathcal{T}$. If we pick a tangent vector X in the tangent space of the torus at the identity, and take any path $\phi(s)$ in T starting from the identity in the direction X , then $Ad(\phi(s))(v) = v$. Differentiating both sides with respect to s at $s = 0$ gives $ad(X)(v) = 0$, so $[X, v] = 0$. Since T is a maximal torus, we conclude that $v \in \mathcal{T}$.

All remaining one dimensional irreducible representations of Ad for $\mathcal{T}$ have complex type. Our formula for these characters is a little dangerous because it depends on selecting an isomorphism $T \simeq R^n/Z^n$. Our G has a left and right invariant metric which is inherited by both T and its Lie algebra, $\mathcal{T}$. If we select an orthonormal basis for $\mathcal{T}$, the lattice $\mathcal{L}$ corresponding to Z^n will usually not be rectangular. Notice that for *any* choice of coordinates giving an isomorphism $\mathcal{T} \simeq R^n$, the map f will have the form $f(t_1, \dots, t_n) = a_1 t_1 + \dots + a_n t_n$ where the a_i are real, and thus will belong to the real dual space of $\mathcal{T}$. The lattice condition then translates into the requirement that $f(\mathcal{L}) \subseteq Z$.

In summary, under the action of T on $\mathcal{G} \otimes C$, we can pick a basis for $\mathcal{T} \otimes C$ and extend it to a basis $\{e_f\}$ of $\mathcal{G} \otimes C$ such that T acts trivially on the basis of $\mathcal{T}$ and acts on the remaining elements via

$$Ad(t)(e_f) = e^{2\pi i f(t)} e_f$$

where f is not zero. The elements f of the dual space of $\mathcal{T}$ are called the *roots* of $\mathcal{G}$, and depend on the choice of T .

8.4 The Many Faces of $\mathcal{T}$

Our maximal torus T has a Lie algebra $\mathcal{T}$, which is contained in the Lie algebra $\mathcal{G}$ of the group G . This $\mathcal{T}$ plays many different roles in our theory, and it is important to keep them straight. For one thing, $\mathcal{T}$ is a standard Euclidean space R^k and even comes with an inner product inherited from the invariant inner product on $\mathcal{G}$. So we can calculate lengths of vectors and angles between them. We will soon be drawing pictures in this space; it will become as familiar as the ordinary plane in high school geometry.

The exponential mapping $\mathcal{T} \rightarrow T$ is a covering map in this abelian case, with kernel the lattice $\mathcal{L}$. So $\mathcal{T}$ is the universal covering group of T . Sometimes when we talk of elements of $\mathcal{T}$, we will imagine that they are in T .

The group T acts on G by conjugation $g \rightarrow tgt^{-1}$. We seldom study this map directly, but it induces a representation Ad of T on the Lie algebra $\mathcal{G}$. If $X \in T$, then $Ad(X)$ is an element of $GL(\mathcal{G})$ and thus an invertible linear transformation on $\mathcal{G}$.

Finally, Ad induces a corresponding map ad from the Lie algebra of T to the Lie algebra of $GL(\mathcal{G})$, that is, a map from $\mathcal{T}$ to homomorphisms from $\mathcal{G}$ to itself. If $X \in \mathcal{T}$, then $ad(X)$ is a linear transformation on $\mathcal{G}$ but not necessarily invertible.

It will be helpful to work out one piece of this puzzle now. We found that the action of T on $\mathcal{G} \otimes C$ breaks up as a sum of one-dimensional representations $Ad(t)(e_f) = e^{2\pi i f(t)} e_f$. Let us work out the corresponding equation for ad . To do this, we fix a non-zero vector X in the Lie algebra of T and form the one-parameter group $u \rightarrow uX$. Then we form the exponential map from the Lie algebra $\mathcal{T}$ to the Lie group T , but since we have identified $\mathcal{T}$ with the universal cover of T , this path is still $u \rightarrow uX$. We then let this act on $\mathcal{G}$ by Ad and the result is $Ad(uX)(e_f) = e^{2\pi i f(uX)} e_f = e^{2\pi i u f(X)} e_f$. Finally, we differentiate at $u = 0$, and obtain

$$ad(X)(e_f) = 2\pi i f(X) e_f$$

8.5 Irreducible Complex Representations of T

Suppose φ is a one-dimensional real representation of T . Then $\varphi(t) = \pm 1$, but T is connected, so φ is the identity representation.

The irreducible complex representations of $\mathcal{T}$ on $\mathcal{G} \otimes C$ have the form $e_f \rightarrow e^{2\pi i f} e_f$ where $e_f = u + iv$, $u, v \in \mathcal{G}$. Notice that $e^{2\pi i f} = \cos 2\pi f + i \sin 2\pi f$. We can select u and v as a basis of an invariant subspace of $\mathcal{G}$ under the real Ad , and then the matrix for the representation has columns given by the action on the basis vectors, and so

$$\begin{pmatrix} \cos 2\pi f(t) & \sin 2\pi f(t) \\ -\sin 2\pi f(t) & \cos 2\pi f(t) \end{pmatrix}$$

This is a two-dimensional irreducible real representation of T , since if it were reducible it would be a direct sum of one dimensional real representations, all of which are trivial, but the induced complex representation is not trivial. Notice that we can change f to $-f$ and simultaneously change u, v to $u, -v$ and get the same representation. Hence the two one-dimensional complex representations of T giving this representation are f and $-f$. Thus if f is a root, so is $-f$ and both give the same 2-dimensional real representation.

The derivative of this expression gives $ad(t)$ acting on $\mathcal{G}$:

$$\begin{pmatrix} 0 & 2\pi f(t) \\ -2\pi f(t) & 0 \end{pmatrix}$$

Said another way,

$$[t, u] = -2\pi f(t)v$$

$$[t, v] = 2\pi f(t)u$$

In summary of all of this, the maximal torus $\mathcal{T}$ acts by Ad on the Lie algebra $\mathcal{G}$ of G . This representation acts trivially on the Lie algebra of $\mathcal{T}$, and breaks up as a sum of two-dimensional real irreducible representations on the rest of $\mathcal{G}$. Each of the two-dimensional representations corresponds to a pair of roots, that is, non-zero linear transformations from the Lie algebra of $\mathcal{T}$ to R satisfying the condition $f(\mathcal{L}) \subseteq Z$.

8.6 Dimension ≤ 3

All of this has been fairly abstract. It is convenient to show that we know enough to find all Lie algebras of compact groups in dimension ≤ 3 . Select a maximal torus T . If this torus is not all of G , then the remaining Lie algebra is a sum of irreducible representations of dimension 2. The only possibility is that there is one such subspace. So the Lie algebra is R, R^2, R^3 , or $R \oplus V$ for a two dimensional invariant subspace V .

From now on, assume T has dimension 1 and G has dimension 3. By the previous section, $\mathcal{G}$ has a basis w, u, v with $[w, u] = -\lambda v$ and $[w, v] = \lambda u$. This $\lambda \neq 0$ since f^* is not trivial. Replace w by $\frac{2}{\lambda}w$, so $[w, u] = -2v$ and $[w, v] = 2u$.

Then $[w, [u, v]] = [[w, u], v] + [u, [w, v]] = 0$. By maximality of T , $[u, v]$ is in the Lie algebra of T and thus equals a multiple of w . Write $[u, v] = \tau w$.

Since G is compact, it has a left and right invariant metric. Then

$$\langle [u, v], w \rangle + \langle v, [u, w] \rangle = 0$$

so

$$\tau \langle w, w \rangle - \langle v, [w, u] \rangle = \tau \langle w, w \rangle - \langle v, -2v \rangle = \tau \langle w, w \rangle + 2 \langle v, v \rangle = 0.$$

Since $\langle w, w \rangle$ and $\langle v, v \rangle$ are positive, $\tau < 0$.

So $[w, u] = -2v$, $[w, v] = 2u$, and $[u, v] = \tau w$. If we replace u and v by γu and γv , we get the same equations, except that now $[\gamma u, \gamma v] = \gamma^2 \tau w$. Since $\tau < 0$, we can select γ so this constant is -2. Hence $\mathcal{G}$ has a basis $\{w, u, v\}$ with $[w, u] = -2v$, $[w, v] = 2u$, $[u, v] = -2w$.

This is the Lie algebra of $G = SU(2)$, and therefore $su(2)$ is the only Lie algebra of a non-abelian compact Lie group of dimension at most three. Indeed, $SU(2)$ is the set of all complex 2×2 matrices A such that $\bar{A}^T A = I$ and $\det(A) = 1$. Therefore the Lie algebra is the set of all complex 2×2 matrices A such that $\bar{A}^T = -A$ and $\text{tr}(A) = 0$. The general form of these matrices is

$$\begin{pmatrix} ih & \alpha + i\beta \\ -\alpha + i\beta & -ih \end{pmatrix}$$

A basis of this algebra is given by

$$w = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad U = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad V = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

A short calculation shows that this basis has the desired bracket products.

8.7 $su(2) \subset \mathcal{G}$

We return now to the general theory. Each two dimensional irreducible subspace of $\mathcal{G}$ under $Ad(T)$ is also a two dimensional irreducible subspace of the Lie algebra representation $ad(T)$ of $\mathcal{T}$ acting on $\mathcal{G}$. This representation is associated with a pair $\pm f$ of non-zero elements of the dual space of $\mathcal{T}$, called *roots*, satisfying $f(\mathcal{L}) \subset Z$. We can find a basis u, v for the two dimensional irreducible subspace so $ad(T)$ is given by

$$[t, u] = -2\pi f(t)v$$

$$[t, v] = 2\pi f(t)u$$

The Lie algebra $\mathcal{G}$ has an inner product $\langle X, Y \rangle$ invariant under $Ad(G)$. This induces an inner product on $\mathcal{T}$, and using this inner product, we have a canonical isomorphism between $\mathcal{T}$ and its dual space. From now on, we use this isomorphism to identify the roots f with elements t_f of $\mathcal{T}$ and thus with geometrical vectors in a Euclidean vector space. To be specific, $f(t) = \langle t_f, t \rangle$.

Theorem 48 *If u, v is the above basis of an irreducible two-dimensional representation associated with f and t_f , then*

$$[u, v] = -2\pi \langle v, v \rangle t_f$$

Moreover, u and v have the same length.

Proof: We have $\langle [u, v], t \rangle = -\langle v, [u, t] \rangle$ by Ad invariance, and this equals

$$\langle v, [t, u] \rangle = \langle v, -2\pi f(t)v \rangle = -2\pi \langle v, v \rangle f(t) = -2\pi \langle v, v \rangle \langle t_f, t \rangle$$

Therefore, $[u, v] + 2\pi \langle v, v \rangle t_f$ belongs to $\mathcal{T}$ and yet is perpendicular to every element $t \in \mathcal{T}$. Since $\langle \rangle$ is positive-definite on $\mathcal{T}$, we conclude that $[u, v] + 2\pi \langle v, v \rangle t_f = 0$ and

$$[u, v] = -2\pi \langle v, v \rangle t_f$$

If we repeat the same argument with u and v interchanged, then in the first displayed line of the proof, the expression $-2\pi f(t)v$ changes to $2\pi f(t)u$. So we get

$$[v, u] = 2\pi \langle u, u \rangle t_f$$

and find that $[v, u] - 2\pi \langle u, u \rangle f_t$ equals zero. Hence changing all signs,

$$[u, v] = -2\pi \langle u, u \rangle f_t$$

It follows that $\langle u, u \rangle = \langle v, v \rangle$ QED.

Theorem 49 *We have*

- $\|u\| = \|v\|$
- $\langle u, v \rangle = 0$

Proof: The first result was proved above. To prove the second, notice that

$$\langle \text{Ad}(t\alpha)u, \text{Ad}(t\alpha)u \rangle = \langle u, u \rangle$$

by invariance of the inner product under Ad . Differentiating with respect to t at $t = 0$ gives $\langle \text{ad}_\alpha u, u \rangle + \langle u, \text{ad}_\alpha u \rangle = 0$ and so $2\langle \text{ad}_\alpha u, u \rangle = 0$. So

$$\langle [\alpha, u], u \rangle = 0 = \langle -2\pi \langle \alpha, \alpha \rangle v, u \rangle$$

and so $\langle v, u \rangle = 0$.

Remark: It follows that $\mathcal{T}$ acts on the irreducible real subspace by rotation. The speed of the rotation is determined by $f(t) = \langle f_t, t \rangle$.

Theorem 50 *The set $\{u, v, t_f\}$ forms a Lie subalgebra of $\mathcal{G}$ isomorphic to $\mathfrak{su}(2)$.*

Proof: Since $SU(2)$ is the group of all matrices A such that $\bar{A}^T A = I$, the Lie algebra $\mathfrak{su}(2)$ consists of all 2×2 complex matrices A such that $A = -\bar{A}^T$. This real Lie algebra is three dimensional, and has a basis

$$w = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad U = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad V = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

The Lie brackets of these basis vectors are

$$[w, U] = -2V$$

$$[w, V] = 2U$$

$$[U, V] = -2w$$

The previous root material gives

$$[t_f, u] = -2\pi \langle t_f, t_f \rangle v$$

$$[t_f, v] = 2\pi \langle t_f, t_f \rangle u$$

$$[u, v] = -2\pi \langle u, u \rangle t_f = -2\pi \langle v, v \rangle t_f$$

In the first two equations, we can multiply both u and v by the same number without changing the equations. Since $\|u\| = \|v\|$, we can assume that $\|u\| = \|v\| = 1$. Then

$$\left[\frac{t_f}{\pi \|t_f\|^2}, u \right] = -2v$$

$$\left[\frac{t_f}{\pi \|t_f\|^2}, v \right] = 2u$$

$$[u, v] = -2\pi t_f = -2\pi^2 \|t_f\|^2 \frac{t_f}{\pi \|t_f\|^2}$$

Let

$$w = \frac{t_f}{\pi \|t_f\|^2}$$

$$U = \frac{u}{\pi \|t_f\|}$$

$$V = \frac{v}{\pi \|t_f\|}$$

These elements satisfy the bracket relations from our basis of $su(2)$. QED.

8.8 $Sp(1)$, $SU(2)$, $SO(3)$, $SO(4)$, and $SL(2, R)$

Recall that the norm of a quaternion is given by $\|a + bi + cj + dk\| = \sqrt{a^2 + b^2 + c^2 + d^2}$. The standard formula $\|q_1 q_2\| = \|q_1\| \|q_2\|$ remains true for quaternions. It follows that the quaternions of norm 1 form a group which is topologically S^3 . This group is denoted $Sp(1)$.

We can think of a quaternion as an ordered pair $(a, bi + cj + dk) = (r, v)$ where r is a number and v is a three dimensional vector. Then the product is given by $(r, v) \circ (s, w) = (rs - v \cdot w, v \times w)$. In particular $(r, v) \circ (r, -v) = (r^2 + v \cdot v, 0)$ and so $q\bar{q} = \|q\|^2$. It is easy to check that $\overline{q_1 q_2} = \bar{q}_2 \bar{q}_1$.

Let V be the set of quaternions of the form $(0, v)$. Notice that $V = \{q \mid \bar{q} = -q\}$. If $v \in V$ and $q \in Sp(1)$, then $qvq^{-1} = qv\bar{q}$ and so $\overline{qvq^{-1}} = \overline{qv\bar{q}} = -qv\bar{q}$. Thus the map $v \rightarrow qv\bar{q}$ maps

V to V . It clearly preserves norms and thus belongs to $SO(3)$. An easy check shows that $Sp(1) \rightarrow SO(3)$ is a group homomorphism.

If $qv\bar{q} = v$ for all v , then $qv = v\bar{q}^{-1} = vq$ for all v and thus q is a real number. Since $\|q\| = 1$, $q = \pm 1$. Hence the kernel of $Sp(1) \rightarrow SO(3)$ is Z_2 . Because both groups are three-dimensional and connected, this homomorphism is onto and $Sp(1)$ is a 2-fold covering group of $SO(3)$.

Suppose q_1 is a quaternion of norm 1. Then $q \rightarrow q_1q$ is a linear transformation $\mathcal{H} \rightarrow \mathcal{H}$ which preserves norms, and thus is an element of $SO(4)$. So we get a group homomorphism $Sp(1) \rightarrow SO(4)$. However this cannot be onto because $Sp(1)$ has dimension 3 and $SO(4)$ has dimension 6. Indeed, suppose we want to move the standard basis e_1, e_2, e_3, e_4 of R^4 to another orthonormal position. We can move e_1 to any unit vector in R^4 and thus to any point of S^3 . That accounts for 3 dimensions. We can then rotate e_2 to any unit vector in the three dimension space perpendicular to the new position of e_1 , and these positions form a two dimensional sphere, so that accounts for 2 more dimensions. Then e_3 has one final dimension of positions, and e_4 is fixed by orientation considerations. And $3 + 2 + 1 = 6$.

The remaining rotations of $\mathcal{H}$ have the form $q \rightarrow qq_2$ for a unit quaternion q_2 . These rotations are almost disjoint from the first set, since $q_1q = qq_2$ for all q implies $q_1 = q_2$ by setting $q = e$, and thus that $q_1q = qq_1$ for all q . Since q_1 has norm 1, $q_1 = \pm 1$.

Consequently, we obtain a map $Sp(1) \times Sp(1) \rightarrow SO(3)$; to make this a group homomorphism, we write it $(q_1, q_2) : q \rightarrow q_1qq_2^{-1}$. In particular, $Sp(1) \times Sp(1)$ is topologically $S^3 \times S^3$, and this becomes the universal cover or spin group for $SO(4)$. We could also divide out by the larger $Z_2 \times Z_2$ and get a 2-fold cover $SO(4) \rightarrow SO(3) \times SO(3)$. So on the Lie algebra level, $so(4) = so(3) \oplus so(3)$ and thus does not appear in our classification list as a separate compact Lie algebra.

The groups $SO(3)$, $SU(2)$, and $SL(2, R)$ are three dimensional and simple. They are the easiest non-abelian simple Lie groups. These three groups are closely related. $SU(2)$ is the 2-fold universal cover of $SO(3)$, and so their Lie algebras $so(3)$ and $su(2)$ are isomorphic. The Lie algebra $sl(2, R)$ of the third group is different because that group is not compact, but $su(2) \otimes C$ and $sl(2, R) \otimes C$ are isomorphic.

We stop to verify these assertions. The group $SO(3)$ consists of rotation matrices satisfying $A^T A = I$ and $\det(A) = 1$. The Lie algebra contains matrices satisfying $A^T = -A$. A basis of this algebra is given by infinitesimal rotations about the x, y , and z axes:

$$E_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad E_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

A brief calculation gives $[E_x, E_y] = E_z$, $[E_y, E_z] = E_x$, $[E_z, E_x] = E_y$.

We previously found a basis for $su(2)$ satisfying $[w, U] = -2V$, $[w, V] = 2U$, $[U, V] = -2w$. Let $\bar{w} = \frac{w}{2}$, $\bar{U} = \frac{U}{2}$, $\bar{V} = \frac{V}{2}$. Then $[\bar{w}, \bar{U}] = -\bar{V}$, $[\bar{w}, \bar{V}] = \bar{U}$, $[\bar{U}, \bar{V}] = -\bar{w}$. Rearranging gives the brackets of $so(3)$, making $su(2)$ isomorphic to $so(3)$:

$$[\bar{w}, \bar{V}] = \bar{U}$$

$$[\bar{V}, \bar{U}] = \bar{w}$$

$$[\bar{U}, \bar{w}] = \bar{V}$$

The group $SL(2, R)$ contains real matrices with determinant 1, so its Lie algebra $sl(2, R)$ contains real matrices of trace 0. A natural basis is

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad L_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad L_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

and

$$[H, L_+] = 2L_+ \quad [H, L_-] = -2L_- \quad [L_+, L_-] = H$$

Compare this with the basis of $su(2)$

$$w = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad U = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad V = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and notice that

$$w = iH \quad U = i(L_+ + L_-) \quad V = L_+ - L_-$$

and

$$H = -iw \quad L_+ = \frac{1}{2}(-iU + V) \quad L_- = \frac{1}{2}(-iU - V)$$

Therefore $su(2) \otimes C \simeq sl(2, R) \otimes C$.

Earlier we discovered that if f is a root with $e_f = u + iv$, then $\{u, v, t_f\}$ forms a subalgebra of $\mathcal{G}$ isomorphic to $su(2)$. It is convenient to find the corresponding basis vectors H, L_+, L_- for $su(2) \otimes C$ in $\mathcal{G} \otimes C$. We are going to use these equations extensively in the next chapter. But in that chapter, we will change the notation. Instead of writing f , we will write α for roots. And instead of writing t_f for the related element of $\mathcal{T}$, we will also write α for this element. So instead of writing $Ad(t)(e_f) = e^{2\pi i f(t)} e_f$ we will write

$$Ad(t)e_\alpha = e^{2\pi i \langle \alpha, t \rangle} e_\alpha$$

and instead of writing $[t, e_f] = 2\pi i f(t) e_f$ we will write

$$[t, e_\alpha] = 2\pi i \langle \alpha, t \rangle e_\alpha$$

and instead of writing $[t, u] = -2\pi f(t)v$, $[t, v] = 2\pi f(t)u$ we will write

$$[t, u] = -2\pi \langle \alpha, t \rangle v$$

$$[t, v] = 2\pi \langle \alpha, t \rangle u$$

Earlier we wrote formulas for w, U, V in terms of t_f, u, v . In old and new notations, these were

$$\begin{aligned} w &= \frac{t_f}{\pi \|t_f\|^2} = \frac{\alpha}{\pi \|\alpha\|^2} \\ U &= \frac{u}{\pi \|t_f\|} = \frac{u}{\pi \|\alpha\|} = \\ V &= \frac{v}{\pi \|t_f\|} = \frac{v}{\pi \|\alpha\|} \end{aligned}$$

Consequently

$$\begin{aligned} H &= -iw = \frac{-i}{\pi \|\alpha\|^2} \alpha \\ L_+ &= \frac{1}{2}(-iU + V) = \frac{1}{2\pi \|\alpha\|} (-iu + v) = \frac{-i}{2\pi \|\alpha\|} (u + iv) = \frac{-i}{2\pi \|\alpha\|} e_\alpha \\ L_- &= \frac{1}{2}(-iU - V) = \frac{1}{2\pi \|\alpha\|} (-iu - v) = \frac{-i}{2\pi \|\alpha\|} (u - iv) = \frac{-i}{2\pi \|\alpha\|} e_{-\alpha} \end{aligned}$$

8.9 Irreducible Representations of $sl(2, R)$, $su(2)$, and $so(3)$

It is a remarkable fact that Lie algebra representation theory on a complex vector space gives exactly the same results for the three algebras $so(3)$, $su(2)$, $sl(2, R)$. The reason is essentially Weyl's unitary trick. Each of these Lie algebras is a real Lie algebra of dimension three. A Lie algebra representation on V is a Lie algebra homomorphism $\mathcal{G} \rightarrow gl(V, n)$. This homomorphism is real linear, but since $gl(V, n)$ is a complex vector space, it can be trivially extended to a complex linear homomorphism $\mathcal{G} \otimes C \rightarrow gl(V, n)$. A complex subspace of V is invariant under the real representation if and only if it is invariant under the complex extension, a complex subspace is irreducible under the real representation if and only if it is invariant under the complex extension, and so forth. Since the tensor products of the three algebras with C are isomorphic, the entire theory is the same for all three algebras.

We have to be careful if we raise this assertion to the group level. Recall that a group homomorphism $\varphi : G_1 \rightarrow G_2$ induces a Lie algebra homomorphism $\mathcal{G}_1 \rightarrow \mathcal{G}_2$. If G_1 is connected, the group homomorphism is uniquely determined by the Lie algebra homomorphism. But a Lie algebra homomorphism $\mathcal{G}_1 \rightarrow \mathcal{G}_2$ is guaranteed to induce a Lie group homomorphism $\varphi : G_1 \rightarrow G_2$ only if G_1 is simply connected. In the general case, we can

construct the group homomorphism from the universal cover $\tilde{G}_1$ to G_2 ; this drops to a homomorphism from G_1 to G_2 only if it takes the kernel of the map $\pi : \tilde{G}_1 \rightarrow G_1$ to the identity.

In the case of our particular three groups, $SU(2)$ is simply connected. Thus any Lie algebra representation of $\mathfrak{su}(2)$ induces a representation of $SU(2)$. Since $SU(2)$ is compact, this Lie group representation splits as a sum of irreducible representations. That implies that the Lie algebra representation of $\mathfrak{su}(2)$ also splits, and consequently any Lie algebra representation of $\mathfrak{su}(2)$, $\mathfrak{sl}(2, R)$, or $\mathfrak{so}(3)$ splits since these algebras are isomorphic.

It is easiest to classify the irreducible representations of $\mathfrak{sl}(2, R)$, so that is what we will do. The resulting theory is one of the most beautiful in all mathematics. Our goal is to prove that $\mathfrak{sl}(2, R)$ has exactly one irreducible representation of dimension n for each $n = 1, 2, 3, \dots$

Suppose we have an irreducible representation φ of $\mathfrak{sl}(2, R)$ on a complex vector space V . The algebra $\mathfrak{sl}(2, R)$ has a basis H, L_+, L_- and the representation produces three linear transformations of V , which we will also call H, L_+, L_- . Since φ preserves brackets, these transformations satisfy $[H, L_+] = 2L_+$, $[H, L_-] = -2L_-$, and $[L_+, L_-] = H$.

Suppose v is an eigenvector of H with eigenvalue λ . Then

$$H(L_+v) = [H, L_+]v - L_+Hv = 2L_+v + L_+\lambda v = (\lambda + 2)L_+v$$

So L_+v is also an eigenvector of H with eigenvalue $\lambda + 2$. For this reason, L_+ is called a *ladder operator* in the literature.

Similarly

$$H(L_-v) = [H, L_-]v - L_-Hv = -2L_-v + L_-\lambda v = (\lambda - 2)L_-v$$

So L_- is also a ladder operator, this time *decreasing* the eigenvalue by 2.

Start with a non-zero eigenvector v and apply the ladder operator L_+ over and over. The resulting vectors have distinct eigenvalues, so they are linearly independent as long as they are non-zero. Eventually $L_+^k v \neq 0$ but $L_+^{k+1} v = 0$. Select $L_+^k v$ as a new starting vector. This time $v \neq 0$ but $L_+ v = 0$.

Now apply the decreasing ladder operator. This again gives a sequence of vectors with distinct eigenvalues, until one of the vectors in the sequence is zero. Suppose the sequence is $L_-^n v, L_-^{n-1} v, \dots, v$. These vectors are linearly independent and the space they generate is invariant under L_- and H . We will prove by induction that it is also invariant under L_+ . If so, then since V is irreducible we conclude that the vectors generate the entire space V and thus form a basis for V .

Notice first that $L_+ v = 0$ by construction. Also

$$L_+(L_-v) = [L_+, L_-]v + L_-L_+v = Hv = \lambda v$$

We now prove by induction that L_+ of $L_-^k v$ is a multiple of $L_-^{k-1} v$.

$$L_+ L_-^k v = [L_+, L_-] L_-^{k-1} v + L_- L_+ L_-^{k-1} v = H L_-^{k-1} v + L_- L_+ L_-^{k-1} v$$

Obviously the last equation proves the claim.

But we can do better. Define μ_k by $L_+ L_-^k v = \mu_k L_-^{k-1} v$. Then $\mu_0 = 0$ and $\mu_1 = \lambda$ and $\mu_k = (\lambda - 2(k-1)) + \mu_{k-1}$. So $\mu_0 = 0, \mu_1 = \lambda, \mu_2 = 2\lambda - 2, \mu_3 = 3\lambda - (2+4), \mu_4 = 4\lambda - (2+4+6)$, and in general $\mu_k = k\lambda - 2(1 + 2 + \dots + k-1) = k\lambda - k(k-1) = k(\lambda - k + 1)$.

However, $L_-^{n+1} v = 0$ so $L_+ L_-^{n+1} v = \mu_n L_-^n v = 0$. Since we are assuming that $L_-^n v \neq 0$, we must have $\mu_n = n(\lambda - n + 1) = 0$ and thus $\lambda = n - 1$. Notice that our V has dimension $n + 1$. So if the dimension of V is 1, then the eigenvalues of H are 0. If the dimension of V is 2, then the eigenvalues of H are 1, -1 . If the dimension of V is 3, then the eigenvalues of H are 2, 0, -2 . And so forth.

Now that we know λ , we know the three matrices H, L_+, L_- completely, and thus there is at most one irreducible representation of each dimension 1, 2, 3, $\dots$. We leave it to the reader to show that the three matrices just obtained have the correct commutation relations; this result is already implicit in our calculations of the effects of the operators on the canonical basis. Finally, our representation is irreducible, for if w were a non-zero vector in an invariant subspace, then by applying the positive ladder operator enough times we would find that $L_+^k w$ is a non-zero multiple of v , so v would belong to the invariant subspace, and then so would $L_-^k v$ for all k .

8.10 Irreducible Representations of $SL(2, R)$, $SU(2)$, and $SO(3)$

Notice that the group $GL(n, C)$ acts on $V = C^n$ by $v \rightarrow Av$, and similarly acts on $V \otimes V \otimes \dots \otimes V$ by $v_1 \otimes v_2 \otimes \dots \otimes v_k \rightarrow A(v_1) \otimes A(v_2) \otimes \dots \otimes A(v_k)$. So any matrix subgroup of $GL(n, C)$ acts in a similar way. If we set $V = C^2$, we obtain representations of $SU(2)$, $SL(2, R)$, and $SL(2, C)$. We will ignore the last group from now on because $SL(2, C)$ is a complex manifold and we are dealing with Lie groups that are real manifolds.

There is an obvious decomposition of $V \otimes V = \Lambda^2(V) \oplus S^2(V)$ where the first subspace consists of skew-symmetric tensor products and the second contains symmetric tensor products. If e_1, e_2 is the standard basis of V , the first space has basis $e_1 \wedge e_2$ and the second has basis $e_1 \odot e_1, e_1 \odot e_2, e_2 \odot e_2$. Tensor products of more copies of V have fancier decompositions, but we only need $S^n(V)$, the set of symmetric tensors $v_1 \odot \dots \odot v_k$. These various subspaces are invariant under $GL(n, C)$ and thus under $SU(2)$ and $SL(2, R)$.

The resulting representation on $\Lambda^2(V)$ is the one dimensional determinant representation, but elements of $SU(2)$ and $SL(2, R)$ have determinant one, so it is the trivial representation. The representations on V and $S^2(V)$ turn out to be irreducible with dimensions 2 and 3,

and the representations on $S^n(V)$ for $n \geq 3$ are also irreducible, of dimensions 4, 5, 6, These give all irreducible representations of $SU(2)$ and $SL(2, R)$.

Let us verify this for $SL(2, R)$. We form the one-parameter group in $SL(2, R)$ generated by the three basis vectors. A brief calculation shows that H, L_+, L_- generate

$$\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$$

All three act trivially on $\Lambda^2(V)$ and as H, L_+, L_- on V . On $S^n(V)$ for $n \geq 2$, we let $f_0 = e_2 \odot \dots \odot e_2, f_1 = e_2 \odot \dots \odot e_2 \odot e_1, \dots, f_n = e_1 \odot \dots \odot e_1$. This is a basis of $S^n(V)$ of size $n + 1$ on which $\exp(tH)$ acts diagonally as $e^{-nt}, e^{-(n-2)t}, \dots, e^{nt}$. Differentiating at $t = 0$ shows that the Lie algebra element H acts as $-n, -(n-2), \dots, n$ on these basis elements.

Similarly $\exp(tL_+)$ acts on these vectors by sending $f_0 \rightarrow (te_1 + e_2) \odot \dots \odot (te_1 + e_2)$, and $f_n \rightarrow f^n$, so taking the derivative at $t = 0$ shows that the Lie algebra element L_+ sends $f_0 \rightarrow nf_1, f_1 \rightarrow (n-1)f_2, \dots, f_{n-1} \rightarrow f_n$ and $f_n \rightarrow 0$. The same calculations using $\exp(L_-)$ and L_- show that L_- sends $f_n \rightarrow nf_{n-1}, f_{n-1} \rightarrow (n-1)f_{n-2}, \dots, f_1 \rightarrow f_0, f_0 \rightarrow 0$.

If we set $F_0 = f_0, F_1 = nf_1, F_2 = n(n-1)f_2, \dots, F_n = n!f_n$, we find that $L_+ : F_i \rightarrow F_{i+1}$ and $F_n \rightarrow 0$. Moreover,

$$L_-(F_k) = L_-(n(n-1) \dots (n-k+1)f_k) = n(n-1) \dots (n-k+1)kf_{k-1} = (n-k+1)kF_{k-1}$$

Notice that $(n-k+1)k = k(n+1-k) = \mu_k$. So we have reproduced the bracket conditions we obtained when we classified irreducible representations of $sl(2, R)$.

It immediately follows that the irreducible representations of $SU(2)$ also arise as $S^n(C^2)$ symmetric vectors.

While the group $SU(2)$ is simply connected, $SL(2, R)$ is not simply connected. Indeed the circle group $SO(2) \subset SL(2, R)$ and the induced map for fundamental groups is an isomorphism. So $SL(2, R)$ has a universal cover which covers each point a countable number of times.

On the other hand, we know all irreducible representations of $sl(2, R)$ and they all arise from representations of $SL(2, R)$. Since every representation is a direct sum of irreducible representations, every Lie algebra homomorphism from $sl(2, R)$ to $gl(n, C)$ induces a Lie group homomorphism $SL(2, R) \rightarrow GL(n, C)$. It follows that every homomorphism from the universal covering group of $SL(2, R)$ to $GL(n, C)$ maps the entire subgroup if the universal cover which covers $I \in SL(2, R)$ to the identity matrix. Consequently the universal cover of $SL(2, R)$ is an example of a Lie group which is not a group of matrices.

We will later show that every compact Lie group can be obtained as a group of matrices.

We earlier proved that the Lie algebras $su(2)$ and $so(3)$ are isomorphic. Choose an isomorphism $\varphi^* : su(2) \rightarrow so(3)$. Since $SU(2)$ is simply connected, this isomorphism is induced by a Lie group homomorphism $\varphi : SU(2) \rightarrow SO(3)$. Since φ^* is an isomorphism, the homomorphism φ is a local diffeomorphism near e , and so onto an open neighborhood of $e \in SO(3)$. Since $SO(3)$ is connected, this neighborhood generates the entire group, so φ is onto.

Suppose u is in the kernel of φ . Then $\varphi(g) = \varphi(ugu^{-1})$. But φ is a local diffeomorphism near the identity element because φ^* is an isomorphism, so $g = ugu^{-1}$ for g in a neighborhood of the identity, and therefore for all g . So u is in the center of $SU(2)$. We claim the center of this group is $\pm I$ and therefore $SO(3) = SU(2)/\pm I$. Indeed it is easy to check that the following matrices are in $SU(2)$:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

If a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ commutes with the left matrix above, $a = d$ and $b = -c$. If it commutes with the right matrix above, $a = d$ and $b = c$. So a matrix in the center must be aI for some a . Since the determinant of the matrix is one, $a = \pm 1$.

The irreducible algebra representations of $so(3)$ are exactly the same as the irreducible algebra representations of $su(2)$, and each of these induces a group representation of $SU(2)$. These representations drop down from $SU(2)$ to $SO(3) = SU(2)/\pm I$ exactly when $-I$ is mapped to the identity. But $-I$ acts on a symmetric vector with n entries $e_1 \odot \dots \odot e_1 \odot e_2 \odot \dots \odot e_2$ as $(-1)^n$. This vector space has dimension $n+1$. So this representation drops down to a representation of $SO(3)$ if and only if n is even and thus if and only if the dimension is odd. It follows that $SO(3)$ has one irreducible representation in each odd dimension $1, 3, 5, 7, \dots$

8.11 Irreducible Representations of $SO(3)$

We will now construct these irreducible representations in a completely different way. Consider the vector space $\mathcal{C}^\infty(S^2)$ of complex-valued functions C^∞ functions on S^2 . The group $SO(3)$ acts on S^2 and thus acts on the function space via $A : f(p) \rightarrow f(A^{-1}p)$. Here we use A^{-1} rather than A for exactly the same reason that in high school we write $f(x-a)$ when we want to translate a function to the *right* by a . We will prove that $\mathcal{C}^\infty$ is a direct sum of irreducible subspaces of functions $V_1 \oplus V_3 \oplus V_5 \oplus \dots$, where $\dim V_i = i$. So each irreducible representation of $SO(3)$ appears exactly once in this decomposition. The functions in V_i are called *spherical functions* on S^2 .

The basic idea of this decomposition is very simple: each V_k will contain all homogeneous polynomials of degree $\frac{k-1}{2}$ in x, y, z . Thus V_1 contains the constant functions and has basis

$\{1\}$. Similarly V_3 contains linear functions and has basis $\{x, y, z\}$. But something tricky happens at the next stage because on S^2 the function that is constantly 1 can also be written $x^2 + y^2 + z^2$. This function should not belong to V_5 , and we can achieve that by taking as basis the functions $\{x^2, y^2, xy, xz, yz\}$ and omitting z^2 from the list. However, if we do that, then V_5 will not be invariant under $SO(3)$ because a rotation takes x^2 to z^2 . Instead the proper step is to take as basis $\{x^2 - y^2, y^2 - z^2, xy, xz, yz\}$. At the next stage V_7 we don't want to include $(x^2 + y^2 + z^2)V_3 = \{(x^2 + y^2 + z^2)x, (x^2 + y^2 + z^2)y, (x^2 + y^2 + z^2)z\}$. This time it is not at all clear what to omit and preserve invariance under $SO(3)$. One correct choice turns out to be the basis $\{x^3 - 3xy^2, 3x^2y - y^3, x^3 - 3xz^2, 3x^2z - z^3, y^3 - 3yz^2, 3y^2z - z^3, xyz\}$. It is not clear how to do the modification in general without a new idea.

Let us introduce the Laplacian $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. Since this operator is invariant under $SO(3)$, it is not surprising that it might play a role in our theory. We can write the operator in spherical coordinates, where it becomes

$$\nabla^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}$$

Think of the first term of this formula as giving the radial component of the Laplacian, and the remaining terms as giving the spherical component of the Laplacian:

$$\nabla^2 f = \nabla_r^2 f + \frac{1}{r^2} \nabla_{S^2}^2 f$$

If we write a homogeneous polynomial of degree n in spherical coordinates, it becomes

$$P(x, y, z) = r^n Q(\theta, \phi)$$

and

$$\nabla^2 P = n(n+1)r^{n-2}Q + r^{n-2}\nabla_{S^2}^2 Q$$

so P is harmonic in R^3 if and only if Q is an eigenvector of the Laplacian on the sphere:

$$\nabla_{S^2}^2 Q = -n(n+1)Q$$

Note that in this formula, Q stands for a function which comes from a homogeneous polynomial, not an arbitrary function on the sphere.

Let $\mathcal{P}_n$ be the set of all homogeneous polynomials of degree n on R^n and let $\mathcal{H}_n$ stand for the set of elements of $\mathcal{P}_n$ which satisfy $\nabla^2 P = 0$. Equivalently $\mathcal{H}_n$ is the set of spherical parts Q on S^2 of homogeneous polynomials of degree n which are eigenfunctions of the spherical Laplacian with eigenvalue $-n(n+1)$.

Theorem 51 *Introduce an inner product on $\mathcal{P}_n$ by declaring that distinct monomials are orthogonal and the length of $x^k y^l z^m$ is $k!l!m!$. Define a map $T : \mathcal{P}_n \rightarrow \mathcal{P}_{n+2}$ by $TP(x, y, z) = (x^2 + y^2 + z^2)P(x, y, z)$. Then we have an orthogonal decomposition*

$$\mathcal{P}_{n+2} = T(\mathcal{P}_n) \oplus \mathcal{H}_{n+2}$$

Remark: Notice that this theorem is a precise form of the calculation we did at the start of this section. It asserts that the representation spaces should be selected so each element is an eigenvalue of the spherical Laplacian, or equivalently is perpendicular to earlier elements.

Remark: The dimension of the set of homogeneous polynomials in x and y of degree m is clearly $m + 1$ since x can have any exponent from 0 to m . Therefore the dimension of $\mathcal{P}_n$, the homogeneous polynomials in x, y, z of degree n , is $1 + 2 + \dots + (n + 1) = \frac{(n+1)(n+2)}{2}$. The theorem then asserts that

$$\dim \mathcal{H}_n = \dim \mathcal{P}_n - \dim \mathcal{P}_{n-2} = \frac{(n+1)(n+2)}{2} - \frac{(n-1)n}{2} = 2n + 1$$

So the $\mathcal{H}_n$ have the correct dimensions to be irreducible representation spaces: $1, 3, 5, 7, \dots$

Remark: The theorem implies that

$$\mathcal{P}_n = \mathcal{H}_n \oplus (x^2 + y^2 + z^2)\mathcal{H}_{n-2} \oplus (x^2 + y^2 + z^2)^2\mathcal{H}_{n-4} \oplus \dots$$

So if we restrict all homogeneous polynomials of degree n to S^2 , we get a sum of representation spaces. Notice this sum contains roughly half of the spaces up to $\mathcal{H}_n$.

Proof, step 1: If V and W are complex vector spaces, each with a Hermitian inner product, and $T : V \rightarrow W$ is linear, define $T^* : W \rightarrow V$ by $\langle Tv, w \rangle_W = \langle v, T^*w \rangle_V$. If T^* exists satisfying this equation, it is unique. Indeed if we had two candidates, then their difference would be a candidate for $T = 0$. If A is this difference, then $\langle v, Aw \rangle$ would be identically zero for all v , so $Aw = 0$ for all w and thus $A = 0$.

We now prove existence. The idea of the proof is very easy. If we are given $T : V \rightarrow W$, we can define the dual map $\tilde{T} : V^* \leftarrow W^*$. On the other hand, the Hermitian inner products on V and W give antilinear isomorphisms $V^* \rightarrow V$ and $W^* \rightarrow W$. Combining these maps gives a linear map $W \rightarrow W^* \rightarrow V^* \rightarrow V$, which is the desired transpose T^* .

Here are details. We define the dual map by

$$\left(\tilde{T}(\varphi)\right)(v) = \varphi(T(v))$$

If we assume the Hermitian inner product is linear in the first variable and anti-linear in the second variable, we define the anti-isomorphism $V \rightarrow V^*$ by

$$v(w) = \langle w, v \rangle$$

Putting these together, our map T^* is defined on $w \in W$ by first obtaining the element in W^* defined by

$$w_1 \rightarrow \langle w_1, w \rangle_W$$

and mapping it to an element of V^* by

$$v \rightarrow \langle Tv, w \rangle_W$$

and then finding the element $T^*w \in V$ which is carried to this element by $V \rightarrow V^*$:

$$T^*w \rightarrow \langle v, T^*w \rangle_V = \langle Tv, w \rangle_W$$

Proof, step 2: Again assume that V and W are complex vector spaces, each with a Hermitian inner product, and $T : V \rightarrow W$ is linear. Then

$$W = T(V) \oplus \text{Ker}(T^*)$$

Indeed it is easy to prove that for any subspace $X \subset W$ we have $W = X \oplus X^\perp$. Apply this to $T(V) \subset W$ to get $W = T(V) \oplus T(V)^\perp$. But the equation $\langle Tv, w \rangle_W = \langle v, T^*w \rangle_V$ shows that $w \in T(V)^\perp$ if and only if $T^*w = 0$.

Proof, step 3: Recall $T : \mathcal{P}_n \rightarrow \mathcal{P}_{n+2}$ by $T(P) = (x^2 + y^2 + z^2)P$. We now claim that ∇^2 and T are adjoint. If so, the previous result shows that $\mathcal{P}_n = T(\mathcal{P}_{n-2}) \oplus \mathcal{H}_n$ and we are done.

We must prove that

$$\langle x^a y^b z^c, T(x^d y^e z^f) \rangle = \langle \nabla^2(x^a y^b z^c), x^d y^e z^f \rangle$$

The left side of this equation is $\langle x^a y^b z^c, x^{d+2} y^e z^f + x^d y^{e+2} z^f + x^d y^e z^{f+2} \rangle$ and this is zero except in three cases:

$$\begin{array}{llll} d+2=a & e=b & f=c & a!b!c! \\ d=a & e+2=b & f=c & a!b!c! \\ d=a & e=b & f+2=c & a!b!c! \end{array}$$

The right side of this equation is $\langle a(a-1)x^{a-2}y^b z^c + b(b-1)x^a y^{b-2} z^c + c(c-1)x^a y^b z^{c-2}, x^d y^e z^f \rangle$ and this is zero except in three cases:

$$\begin{array}{llll} d=a-2 & e=b & f=c & a(a-1)d!e!f! \\ d=a & e=b-2 & f=c & b(b-1)d!e!f! \\ d=a & e=b & f=c-2 & c(c-1)d!e!f! \end{array}$$

Clearly these are equal. QED.

Remark: To finish the theory we'd like to show that each $\mathcal{H}_n$ is invariant and irreducible under the action of $SO(3)$. Invariance is easy. Rotating by A replaces each of x, y, z by a linear combination of these elements, and so preserves homogeneous polynomials of degree

n . The Laplacian is invariant, so if $\nabla^2 f = 0$ and $A \in SO(3)$, then $\nabla^2(f \circ A^{-1}) = 0$. Since $\mathcal{H}_n$ is the set of functions satisfying both conditions, it is invariant under $SO(3)$.

Finally, we prove that $\mathcal{H}_n$ gives the unique irreducible representation of dimension $2n + 1$. The argument is unexpected. We will show that when $SO(3)$ acts on functions, the elements H, L_+, L_- coming from $so(3)$ and $sl(2, R)$ become differential operators, and a simple expression involving them equals ∇^2 . Consequently, functions in an irreducible space of dimension $2k + 1$ must be eigenfunctions of the spherical Laplacian with eigenvalue $-k(k + 1)$. If $\mathcal{H}_n$ is not irreducible, it must be a sum of irreducible subspaces of smaller dimension. Every function belonging to such a smaller subspace would be an eigenfunction of the spherical Laplacian with eigenvalue $-k(k + 1)$, but we know that every element in $\mathcal{H}_n$ is an eigenfunction of the spherical Laplacian with eigenvalue $-n(n + 1)$. So a decomposition into smaller irreducible subspaces is impossible.

Here are the details. Recall that $so(3)$ has basis E_x, E_y, E_z . For example, E_z is an infinitesimal rotation about the z -axis, which generates the one-parameter subgroup

$$\gamma_z(t) = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If f is a C^∞ function on R^3 , $f \circ \gamma^{-1}(t) = f(\cos(t)x - \sin(t)y, \sin(t)x + \cos(t)y, z)$ and the derivative of this expression at $t = 0$ gives $E_z(f)$. So $E_z(f) = -y \frac{\partial f}{\partial x} + x \frac{\partial f}{\partial y}$ and $E_x(f) = -z \frac{\partial f}{\partial y} + y \frac{\partial f}{\partial z}$ and $E_y(f) = -x \frac{\partial f}{\partial z} + z \frac{\partial f}{\partial x}$. We will work with these as operators:

$$E_x = -z \frac{\partial}{\partial y} + y \frac{\partial}{\partial z}$$

$$E_y = -x \frac{\partial}{\partial z} + z \frac{\partial}{\partial x}$$

$$E_z = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}$$

At the end of section 8.8, we wrote the generators of $su(2)$ in terms of E_x, E_y, E_z and then wrote the generators of $sl(2, R)$ in terms of the generators of $su(2)$. Those equations give

$$H = -2iE_z$$

$$L_+ = (E_x - iE_y)$$

$$L_- = (-E_x - iE_y)$$

Notice that $L_+L_- = (E_x - iE_y)(-E_x - iE_y) = -E_x^2 - E_y^2 - E_z^2 + E_z^2 - i(E_xE_y - E_yE_x)$. So $L_+L_- = -(E_x^2 + E_y^2 + E_z^2) - \frac{H^2}{4} + \frac{H}{2}$ and therefore

$$(E_x^2 + E_y^2 + E_z^2) = -L_+L_- - \frac{H}{2} \left(\frac{H}{2} - 1 \right)$$

The operator $E_x^2 + E_y^2 + E_z^2$ plays the central role in the remainder of this section. Later it was generalized to all Lie algebras by Hendrik Casimir; this generalization is called the Casimir element.

We claim that $E_x^2 + E_y^2 + E_z^2$ commutes with E_x, E_y , and E_z . By symmetry, it suffices to show that it commutes with E_x . In the calculation which follows, we will apply the two formulas below:

$$E_yE_x = E_xE_y - E_z$$

$$E_zE_x = E_xE_z + E_y$$

Here is the calculation

$$\begin{aligned} (E_x^2 + E_y^2 + E_z^2)E_x &= \\ E_xE_x^2 + E_y(E_yE_x) + E_z(E_zE_x) &= \\ E_xE_x^2 + E_y(E_xE_y - E_z) + E_z(E_xE_z + E_y) &= \\ E_xE_x^2 + (E_yE_x)E_y - E_yE_z + (E_zE_x)E_z + E_zE_y &= \\ E_xE_x^2 + (E_xE_y - E_z)E_y - E_yE_z + (E_xE_z + E_y)E_z + E_zE_y &= \\ E_x(E_x^2 + E_y^2 + E_z^2) - E_zE_y - E_yE_z + E_yE_z + E_zE_y &= \\ E_x(E_x^2 + E_y^2 + E_z^2) \end{aligned}$$

Now suppose that $V \subset C^\infty(S^2)$ is an irreducible invariant subspace under the action of $SO(3)$. Then V is invariant under $E_x^2 + E_y^2 + E_z^2$ and operator commutes with the action of every element of $so(3)$. So it is an intertwining operator, and consequently is constant on V . If the dimension of V is $n = 2k + 1$, then the eigenvalues of H are $-2k, -2k + 2, \dots, 2k$ and the eigenvector associated with $-2k$ vanishes under L_- , so applying this constant operator to that lowest eigenvector and using the right side of the formula at the top of the page gives

$$0 - \left(\frac{-2k}{2} \right) \left(\frac{-2k}{2} - 1 \right) = -k(k + 1)$$

Therefore, $E_x^2 + E_y^2 + E_z^2$ is constantly $-k(k + 1)$ on an irreducible $V \subset C^\infty(S^2)$ of dimension $2k + 1$.

Finally, we show that $E_x^2 + E_y^2 + E_z^2$ is the spherical Laplacian, connecting our work with the discussion of the $\mathcal{H}_n$. This involves a straightforward but messy calculation. We start with

$$E_z^2 = \left(-y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}\right) \left(-y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}\right) = y^2 \frac{\partial^2}{\partial x^2} + x^2 \frac{\partial^2}{\partial y^2} - x \frac{\partial}{\partial x} - y \frac{\partial}{\partial y} - 2xy \frac{\partial^2}{\partial x \partial y}$$

The other two terms can be obtained by symmetry, so we immediately write

$$E_x^2 + E_y^2 + E_z^2 = (x^2 + y^2 + z^2) \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) - \left(x^2 \frac{\partial^2}{\partial x^2} + y^2 \frac{\partial^2}{\partial y^2} + z^2 \frac{\partial^2}{\partial z^2}\right) - 2 \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right) - 2 \left(xy \frac{\partial^2}{\partial x \partial y} + yz \frac{\partial^2}{\partial y \partial z} + zx \frac{\partial^2}{\partial z \partial x}\right)$$

We work on the bottom row of terms, which are related to

$$\left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)^2 = \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right) \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right) = \left(x^2 \frac{\partial^2}{\partial x^2} + y^2 \frac{\partial^2}{\partial y^2} + z^2 \frac{\partial^2}{\partial z^2}\right) + 2 \left(xy \frac{\partial^2}{\partial x \partial y} + yz \frac{\partial^2}{\partial y \partial z} + zx \frac{\partial^2}{\partial z \partial x}\right) + \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)$$

So

$$E_x^2 + E_y^2 + E_z^2 = (x^2 + y^2 + z^2) \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) - \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)^2 - \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)$$

If $f(x, y, z)$ is a function, we can write it in spherical coordinates as $f(r \cos \theta \cos \phi, r \sin \theta \cos \phi, r \sin \phi)$.

So

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \cos \theta \cos \phi + \frac{\partial f}{\partial y} \sin \theta \cos \phi + \frac{\partial f}{\partial z} \sin \phi = \frac{1}{r} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)$$

Therefore

$$\left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right)^2 + \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} + z \frac{\partial}{\partial z}\right) = \left(r \frac{\partial}{\partial r}\right)^2 + \left(r \frac{\partial}{\partial r}\right) = r^2 \frac{\partial^2}{\partial r^2} + 2r \frac{\partial}{\partial r}$$

So

$$E_x^2 + E_y^2 + E_z^2 = r^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) - r^2 \frac{\partial^2}{\partial r^2} - 2r \frac{\partial}{\partial r} = r^2 \nabla^2 - \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r}\right)$$

Earlier we wrote

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2} \nabla_{S^2}^2$$

Substituting this into the previous formula gives our final result, the amazing

$$E_x^2 + E_y^2 + E_z^2 = \nabla_{S^2}^2$$

Theorem 52 $\mathcal{H}_n$ is invariant and irreducible under $SO(3)$. It is the unique irreducible representation of dimension $2n + 1$.

Proof: The operator ∇^2 vanishes on $\mathcal{H}_n$, but it is a sum of two terms, a radial term and a spherical term. All elements of $\mathcal{H}_n$ have the same radial term, so we easily computed that $E_x^2 + E_y^2 + E_z^2 = \nabla_{S^2}^2$ is constantly $-n(n + 1)$ on $\mathcal{H}_n$.

Since $\mathcal{H}_n$ is invariant under $SO(3)$, so it can be written as a direct sum of irreducible representations. On those representations of dimension $2k + 1$, the value of the operator is $-k(k + 1)$ by previous results. So the only representation occurring in the decomposition must be the representation of dimension $2n + 1$. This is the dimension of $\mathcal{H}_n$, so $\mathcal{H}_n$ itself must be irreducible. QED.

8.12 Miscellaneous Remarks on Representations of $SO(3)$

We earlier found bases for the first few $\mathcal{H}_n$. This process can be done systematically. Let us start with $\mathcal{H}_2$. A homogeneous polynomial $Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz$ is in $\mathcal{H}_n$ just in case its Laplacian is zero, so $2A + 2B + 2C = 0$. This gives

$$\mathcal{H}_2 = \{x^2 - y^2, y^2 - z^2, xy, xz, yz\}$$

We'll carry this one more step, although much more can be said. A basis for $\mathcal{P}_3$ is

$$\{x^3, y^3, z^3, x^2y, y^2z, z^2x, xy^2, yz^2, zx^2, xyz\}$$

This space has dimension 10, but $\mathcal{H}_3$ has dimension 7. The Laplacian of x^3 is $6x$, the Laplacian of x^2y is $2y$, and the Laplacian of xyz is 0. So the Laplacian of a general element $Ax^3 + By^3 + Cz^3 + Dx^2y + Ey^2z + Fz^2x + Gxy^2 + Hyz^2 + Izx^2 + Jxyz$ in $\mathcal{P}_3$ is

$$6xA + 6yB + 6zC + 2yD + 2zE + 2xF + 2xG + 2yH + 2zI$$

This is harmonic if

$$6A + 2F + 2G = 0$$

$$6B + 2D + 2H = 0$$

$$6C + 2E + 2I = 0$$

These equations are symmetrical, so let us deal with the first one. It says $3A = -F - G$. This sets one condition on the variables (A, F, G) and we get two independent solutions $(1, -3, 0)$ and $(1, 0, -3)$. So we get two elements $x^3 - 3xz^2$ and $x^3 - 3xy^2$. By symmetry, we also get the pair $y^3 - 3yz^2$ and $y^3 - 3yx^2$ and the pair $z^3 - 3zx^2$ and $z^3 - 3zy^2$. There is no condition on J , so we get a basis we had written earlier,

$$\mathcal{H}_3 = \{x^3 - 3xz^2, x^3 - 3xy^2, y^3 - 3yz^2, y^3 - 3yx^2, z^3 - 3zx^2, z^3 - 3zy^2, xyz\}$$

The physicists often use a completely different approach to this theory by starting with partial differential equations. It is fun to sketch this approach and notice the moment when it merges with representation theory.

In quantum theory, classical equations can be converted to quantum equations by replacing the observable quantities with operators. The functions being operated on are often written $\psi(x, y, z)$. The observable giving the x coordinate of a point is replaced with the operator “multiplication by x ”, so $\psi \rightarrow x\psi$. The observable giving the x momentum of a point is replaced with the operator $-i\hbar \frac{\partial}{\partial x}$, so $\psi \rightarrow -i\hbar \frac{\partial \psi}{\partial x}$. Consider, for example, an electron with mass m in a hydrogen atom. The total energy of this election is given by adding its kinetic energy to its potential energy, so

$$\frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + V(x, y, z) = E$$

Replacing observables with operators gives

$$\frac{1}{2m} \left((-i\hbar \frac{\partial}{\partial x})^2 + (-i\hbar \frac{\partial}{\partial y})^2 + (-i\hbar \frac{\partial}{\partial z})^2 \right) + V = E$$

or

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi + V\psi = E\psi$$

This is the famous Schrodinger equation. Here $V(x, y, z)$ is the potential energy due to the nucleus and E is the energy level. These energies are discrete because the physicists require that $\int \int \int |\psi|^2 = 1$ and only special values of E allow that. When Schrodinger first solved this equation, the mathematics predicted the known discrete values of E in this case and many other cases, and that was the first strong evidence for the equation. But for several years, the physicists argued about the physical significance of the function ψ . The current interpretation is that $|\psi|^2$ gives the probability of finding an election at a given spot. In the theory it is very important that ψ take complex values, and surprising that the differential equation involves ψ rather than $|\psi|$. But we leave these consequences to the physicists and turn to the matter of solving the equation.

According to legend, Schrodinger went on a skiing trip and let his friends ski while he calculated in the chalet. With him was a copy of Hilbert and Courant’s book *Methods of Mathematical Physics*.

We write the equation in the form

$$\nabla^2 \psi - \frac{2m}{\hbar^2} (V - E) \psi = 0$$

Because of the symmetry of the hydrogen atom, we work in spherical coordinates, and use the fact that V depends only on r . So

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} - \frac{2m}{\hbar^2} (V(r) - E) \psi = 0$$

We now try to separate the solution into the form $\psi(r, \theta, \phi) = R(r)Y(\theta, \phi)$. Then

$$\left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) - \frac{2m}{\hbar^2} (V(r) - E) R \right] Y + R \left[\frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \right] = 0$$

Multiplying both sides by r^2 gives

$$- \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) - \frac{2mr^2}{\hbar^2} (V(r) - E) R \right] Y = R \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \right]$$

If we divide both sides by RY , then the first term involves only r and the second term involves only θ and ϕ . So both terms must equal a constant K .

Let us consider the second half. It becomes a differential equation

$$\left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \right] = KY$$

This equation has already come up in our work. We usually write the equation

$$\nabla_{S^2}^2 Y = KY$$

We proved that $C^\infty(S^2)$ can be written as a sum of $SO(3)$ irreducible subspaces of dimension $2l+1$ for $l = 0, 1, 2, \dots$, and on the l th subspace, $\mathcal{H}_l$, $K = -l(l+1)$. We conclude that the constant must have the form $-l(l+1)$ and the Y must be elements of $\mathcal{H}_l$. We have switched to using l here because the physicists use that letter.

The physicists sometimes place atoms in an electric field. This reduces the symmetry group from $SO(3)$ to $SO(2)$, and the representation $\mathcal{H}$ then breaks into $2l+1$ one-dimensional representations indexed by eigenvalues of $H_z : -2l, -(2l-2), \dots, 2l$. The physicists divide these values by two and call them m , so $m = -l, -(l-1), \dots, l$. This gives two quantum numbers, l and m , for the angular quantum number.

Schrodinger did not have representation theory at his disposal, so he solved the spherical Laplacian directly. This led to various Legendre polynomials and expressions in ϕ which can be found in physics books. These expressions via Legendre polynomials provide another useful way to study spherical functions; indeed the entire representation theory of $SO(3)$ can be developed directly from this point of view, since the solutions of $\nabla_{S^2}^2 f = \lambda f$ are null for most λ and an irreducible representation space when $\lambda = -l(l+1)$.

For completeness, let us also deal with the radial equation. For $V(r)$ we will use the coulomb potential $V(r) = -\frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r}$. The factor $4\pi\epsilon$ depends on electrical units and becomes 1 for some unit systems. The Schrodinger equation gives a good approximation for some other atoms, particular those with one electron in the outer shell, and in that case Z counts the number of protons in the nucleus. Another slight complication is that in hydrogen, for

instance, both the proton and the electron rotate around a common center of mass, and we should account for that by using the reduced mass of the electron, $\frac{m_1 m_2}{m_1 + m_2}$, rather than its actual mass. Since the proton is so much heavier than the electron, this correction is very small.

$$\frac{1}{R} \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) - \frac{2mr^2}{\hbar^2} \left(-\frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r} - E \right) \right] - l(l+1) = 0$$

Rewriting slightly and noticing that everything depends only on r :

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \left[\frac{2mr^2}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{r} + \frac{2mr^2}{\hbar^2} E - l(l+1) \right] = 0$$

and so

$$\frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \left[\frac{2mr^2}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{r} + \frac{2mr^2}{\hbar^2} E - l(l+1) \right] R = 0$$

Let $R(r) = \frac{u(r)}{r}$. Then $\frac{dR}{dr} = \frac{ru' - u}{r^2}$ and $r^2 \frac{dR}{dr} = ru' - u$ and $\frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) = ru'' + u' - u' = ru''$. So the differential equation becomes

$$\frac{d^2 u}{dr^2} + \left[\frac{2mr^2}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{r} + \frac{2mr^2}{\hbar^2} E - l(l+1) \right] \frac{u(r)}{r^2} = 0$$

or

$$\frac{d^2 u}{dr^2} + \left[\frac{2m}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{r} + \frac{2m}{\hbar^2} E - \frac{l(l+1)}{r^2} \right] u(r) = 0$$

It is possible to solve this equation directly, but more illuminating to guess the form of the solution. Suppose first that r is very large. Then several terms vanish and the equation becomes approximately

$$\frac{d^2 u}{dr^2} + \left(\frac{2mE}{\hbar^2} \right) u = 0$$

All solutions of this equation are linear combinations of $u(r) = e^{\pm \frac{\sqrt{-2mE}}{\hbar} r}$.

At this point, physicists are smarter than mathematicians and argue as follow. Suppose the electron is far away from the proton. Then it hardly notices the proton and thus should have energy close to zero. What happens as the electron approaches the proton? It is well-known that the electron is then attracted to the proton and becomes bound to it. So the energy of the electron should become negative because it takes positive energy to pry it away. Classically the electron is attracted more and more strongly and should crash into the proton, but we hope that quantum mechanics and the Schrodinger equation will save us from that catastrophe. The upshot is that $E < 0$ and thus the square root is a real number. We must take the negative of this exponent to make E go to zero at infinity.

In short, we suspect that $u(r) \sim e^{-\sqrt{\frac{2m(-E)}{\hbar^2}}r}$ for large r . To avoid writing all of these constants over and over, we will define σ by $\left(\frac{\sigma}{2}\right)^2 = \frac{-2mE}{\hbar^2}$. So $u(r) \sim e^{-\frac{\sigma r}{2}}$. With this abbreviation, our equation becomes

$$\frac{d^2u}{dr^2} + \left[\frac{2m}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{r} - \frac{\sigma^2}{4} - \frac{l(l+1)}{r^2} \right] u(r) = 0$$

Next we rescale the variable r to make $\sigma = 1$. Qualitatively this is a very minor change. We let $x = \sigma r$; then $r = \frac{x}{\sigma}$ and u becomes a function of x via $u(r(x))$, so $\frac{du}{dx} = \frac{du}{dr} \frac{dr}{dx} = \frac{du}{dx} \frac{1}{\sigma}$ and $\frac{d^2u}{dx^2} = \frac{d^2u}{dr^2} \frac{dr}{dx} \frac{1}{\sigma} = \frac{d^2u}{dr^2} \frac{1}{\sigma^2}$. So our equation becomes

$$\sigma^2 \frac{d^2u}{dx^2} + \left[\frac{2m\sigma}{\hbar^2 4\pi\epsilon_0} \frac{Ze^2}{x} - \frac{\sigma^2}{4} - \frac{l(l+1)\sigma^2}{x^2} \right] u(x) = 0$$

and dividing both sides by σ^2 gives

$$\frac{d^2u}{dx^2} + \left[\frac{2m}{\hbar^2 4\pi\epsilon_0\sigma} \frac{Ze^2}{x} - \frac{1}{4} - \frac{l(l+1)}{x^2} \right] u(x) = 0$$

Note that we now expect $u(x) \sim e^{-\frac{x}{2}}$ for large x . Therefore, let us write $u(x) = P(x)e^{-\frac{x}{2}}$. Then $u' = P'e^{-\frac{x}{2}} - \frac{1}{2}Pe^{-\frac{x}{2}}$ and $u'' = P''e^{-\frac{x}{2}} - P'e^{-\frac{x}{2}} + \frac{1}{4}Pe^{-\frac{x}{2}}$. When we substitute this into our equation, the exponents will cancel and we get

$$P'' - P' + \frac{1}{4}P + \left[\frac{2m}{\hbar^2 4\pi\epsilon_0\sigma} \frac{Ze^2}{x} - \frac{1}{4} - \frac{l(l+1)}{x^2} \right] P = 0$$

So two terms cancel and the final equation is

$$\frac{d^2P}{dx^2} - \frac{dP}{dx} + \left[\frac{2m}{\hbar^2 4\pi\epsilon_0\sigma} \frac{Ze^2}{x} - \frac{l(l+1)}{x^2} \right] P = 0$$

Let us try to solve this equation with a power series. The first nonzero term of that series will be a constant times x^k . Putting this into the differential equation and writing only the lowest terms gives $k(k-1)x^{k-2} - l(l+1)x^{k-2} = 0$, so $k(k-1) = l(l+1)$ or $k^2 - k = l^2 + l$ or $k^2 - l^2 = k + l$ or $(k-l)(k+l) = (k+l)$, so $k = l + 1$. Let us consequently write $P(x) = x^{l+1}Q(x)$ so that Q is a power series which starts with a constant term.

Suppose, for example, that $l = 0$. Then the series starts with x . But notice that $P(x) = x$ actually solves the equation, provided $-1 + \frac{2mZe^2}{\hbar^2 4\pi\epsilon_0\sigma} = 0$, or $\hbar^2 2\pi\epsilon_0\sigma = mZe^2$, or $\sigma = \frac{mZe^2}{\hbar^2 2\pi\epsilon_0}$. Recall, however, that $\left(\frac{\sigma}{2}\right)^2 = -\frac{2mE}{\hbar^2}$. So this is really a condition on E , the only unknown quantity in the equation. It says that

$$E = -\frac{\hbar^2}{2m} \left(\frac{mZe^2}{\hbar^2 4\pi\epsilon_0} \right)^2 = \frac{m}{2\hbar^2} \left(\frac{e^2}{4\pi\epsilon_0} \right)^2$$

and for this E we can select $u(x) = xe^{-\frac{x}{2}}$.

It is encouraging that in a special case we could actually solve our equation, so we continue. Since we know the power series starts with x^{l+1} , we write $P(x) = x^{l+1}Q(x)$ and rewrite the equation in terms of Q . We have $P' = (l+1)x^lQ + x^{l+1}Q'$ and $P'' = l(l+1)x^{l-1}Q + 2(l+1)x^lQ' + x^{l+1}Q''$. So

$$\begin{aligned} & \left(l(l+1)x^{l-1}Q + 2(l+1)x^lQ' + x^{l+1}Q'' \right) - \left((l+1)x^lQ + x^{l+1}Q' \right) + \\ & \left[\frac{2m}{\hbar^2 4\pi\epsilon_0\sigma} \frac{Ze^2}{x} - \frac{l(l+1)}{x^2} \right] x^{l+1}Q = 0 \end{aligned}$$

and this simplifies after dividing out by x^l to

$$xQ'' + [2(l+1) - x]Q' + \left[\frac{2mZe^2}{\hbar^2 4\pi\epsilon_0\sigma} - (l+1) \right] Q = 0$$

Notice that the coefficient of Q is just a constant which we will call A . Assume that $Q(x) = \sum_{k \geq 0} a_k x^k$ where $a_0 \neq 0$. Then

$$\sum_{k \geq 2} k(k-1)a_k x^{k-1} + \sum_{k \geq 1} 2(l+1)ka_k x^{k-1} - \sum_{k \geq 1} ka_k x^k + A \sum_{k \geq 0} a_k x^k = 0$$

or

$$\sum_{k \geq 1} k(k+1)a_{k+1}x^k + \sum_{k \geq 0} 2(l+1)(k+1)a_{k+1}x^k - \sum_{k \geq 1} ka_k x^k + A \sum_{k \geq 0} a_k x^k = 0$$

So

$$\begin{aligned} k(k+1)a_{k+1} + 2(l+1)(k+1)a_{k+1} - ka_k + Aa_k &= 0 \quad \text{for } k \geq 1 \\ 2(l+1)a_1 + Aa_0 &= 0 \end{aligned}$$

This simplifies to

$$\begin{aligned} & \left[k(k+1) + 2(l+1)(k+1) \right] a_{k+1} = (k-A)a_k \quad \text{for } k \geq 1 \\ & a_1 = -\frac{A}{2(l+1)}a_0 \end{aligned}$$

Now notice that if we multiply a_0 by a constant, we multiply a_1 by the same constant, and thus all higher a_k by the same constant. This is to be expected since our differential equation is linear. So we may as well take $a_0 = 1$.

Next we observe that a_0 determines all higher a_k , so we get exactly one solution. Our differential equation has order two, but the other solution is singular at the origin, since

the equation starts xQ'' . We discard that solution and deal with the solution at hand. Finally, notice that if $A = 0$, then $a_1 = a_2 + \dots = 0$ and we get the solution $Q = 1$ and $P = x$ discovered earlier.

One other crucial fact stands out that was hidden away until now. If A is a non-negative integer, say $A = n_r$, then $a_{n_r+1} = a_{n_r+2} = \dots$ all vanish and the solution is a polynomial of degree n_r . These are exactly the solutions we are looking for; otherwise the solution is not normalizable (we leave that last assertion to the physicists).

The integer n_r is called the *radial quantum number* by the physicists. Note that $n_r \geq 0$.

If n_r is an integer, then $\frac{2mZe^2}{\hbar^2 4\pi\epsilon_0\sigma} = n$ is also an integer. This integer is called the *principal quantum number* by the physicists. Note that $n - (l + 1) = n_r$. Since $n_r \geq 0$, $n \geq (l + 1)$ and so $n > l$. In the expression defining n , all of the items are physical constants except σ , where is related to the energy. Consequently we can solve for E in terms of n . We have

$$\frac{2mZe^2}{\hbar^2 4\pi\epsilon_0\sigma} = n \quad \text{and} \quad \left(\frac{\sigma}{2}\right)^2 = \frac{-2mE}{\hbar^2}$$

So

$$\sigma = \frac{2mZe^2}{\hbar^2 4\pi\epsilon_0 n} \quad \text{and} \quad E = -\frac{\hbar^2 \sigma^2}{8m}$$

and therefore

$$E = -\frac{\hbar^2}{2m} \left(\frac{2mZe^2}{\hbar^2 4\pi\epsilon_0 n} \right)^2 = -\frac{2m}{\hbar^2} \left(\frac{Ze^2}{4\pi\epsilon n} \right)^2$$

Notice that these energies are all negative. The lowest energy occurs when $n = 1$; as n increases, the energy goes to zero. This is reasonable because larger and larger n produce electrons with more and more energy until finally they can escape the atom and become free electrons when $E = 0$.

In quantum mechanical systems, energy is quantized, so for each new system, physicists ask for the possible values of energy. In the hydrogen atom, this is completely determined by the integer n , so calling it the *principal quantum number* makes sense.

Returning to our power series solution of the radial equation, we have

$$a_{k+1} = -\frac{A - k}{(k + 1)(k + 2l + 2)} a_k$$

so

$$\begin{aligned} a_1 &= -\frac{A}{(2l + 2)} a_0 \\ a_2 &= \frac{A - 1}{2(2l + 3)} \frac{A}{(2l + 2)} a_0 \end{aligned}$$

$$a_3 = -\frac{A-2}{3(2l+4)} \frac{A-1}{2(2l+3)} \frac{A}{(2l+2)} a_0$$

$$a_4 = \frac{A-3}{4(2l+5)} \frac{A-2}{3(2l+4)} \frac{A-1}{2(2l+3)} \frac{A}{(2l+2)} a_0$$

From this, it is easy to write the general case:

$$a_k = (-1)^k \frac{A!(2l+1)!}{k!(2l+k+1)!(A-k)!} x^k$$

so

$$Q(x) = \sum_{k=0}^{n_r} (-1)^k \frac{n_r!(2l+1)!}{k!(2l+k+1)!(n_r-k)!} x^k$$

When the physicists first did this calculation, they discovered that all of the differential equations involved had already been studied in other contexts. Rather than giving the solutions new names to fit the application to the hydrogen atom, they used the old names, allowing their followers to discover useful results in older books.

In particular, if j and k are non-negative integers, the differential equation

$$x \frac{d^2 L_j^k}{dx^2} + (1-x+k) \frac{dL_j^k}{dx} + jL_j^k = 0$$

had been studied under the name *associated Laguerre equation* and its solutions L_j^k were called *associated Laguerre polynomials*. If we let $k = 2l+1$ and $j = n-(l+1)$ for l, n integers with $l \geq 0$ and $n > l$, we find that the coefficient of $\frac{dL_j^k}{dx}$ is $1-x+(2l+1) = 2(l+1)-x$ and the coefficient of L_j^k is $j = n-(l+1)$ and Laguerre's equation is exactly our equation for Q . So our solution is $L_{n-(l+1)}^{2l+1}(x)$.

Consequently $P(x) = x^{l+1} L_{n-(l+1)}^{2l+1}(x)$ and $u(x) = e^{-\frac{x}{2}} x^{l+1} L_{n-(l+1)}^{2l+1}(x)$. Recall, however, that $R(x) = \frac{u(x)}{x}$. So

$$R(x) = e^{-\frac{x}{2}} x^l L_{n-(l+1)}^{2l+1}(x)$$

Finally, we made a substitution $x = \sigma r$ which I will not undo here.

There is one other feature of the theory I will not discuss here. In representation theory, we have the orthogonality relations making the various polynomials involved in angular momentum orthogonal and normalizable. It turns out that eigenfunctions associated to different eigenvalues in quantum mechanics are also orthogonal. This leads to the theory of orthogonal polynomials, and the various Laguerre functions are an essential part of that theory. Orthogonality in representation theory will be central to our theory soon enough.

We end this discussion of the radial quantum number with two pictures from a Stack Exchange discussion of this number, at

<https://physics.stackexchange.com/questions/373623/what-is-the-radial-quantum-number-n-r>.

The first shows the radial function for $n = 1, 2, 3, 4$ with $l = 0$ and the second shows these for $l = 2$:

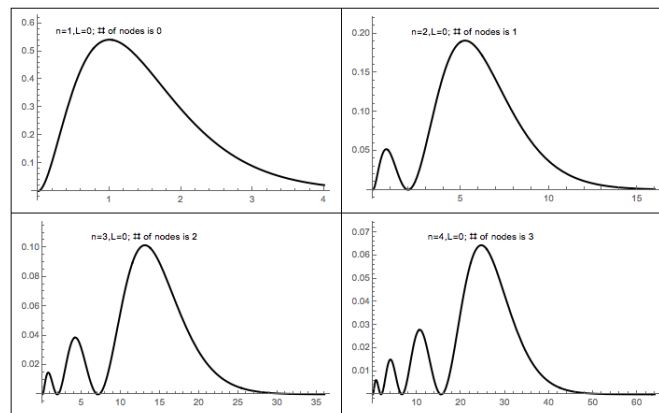


Figure 8.1: $n = 1, 2, 3, 4; l = 0$

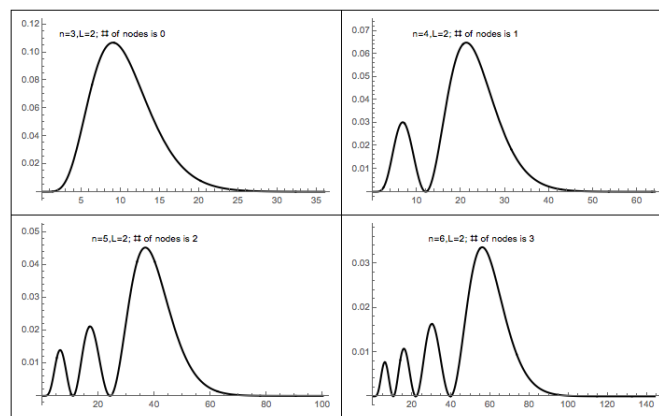


Figure 8.2: $n = 3, 4, 5, 6; l = 2$

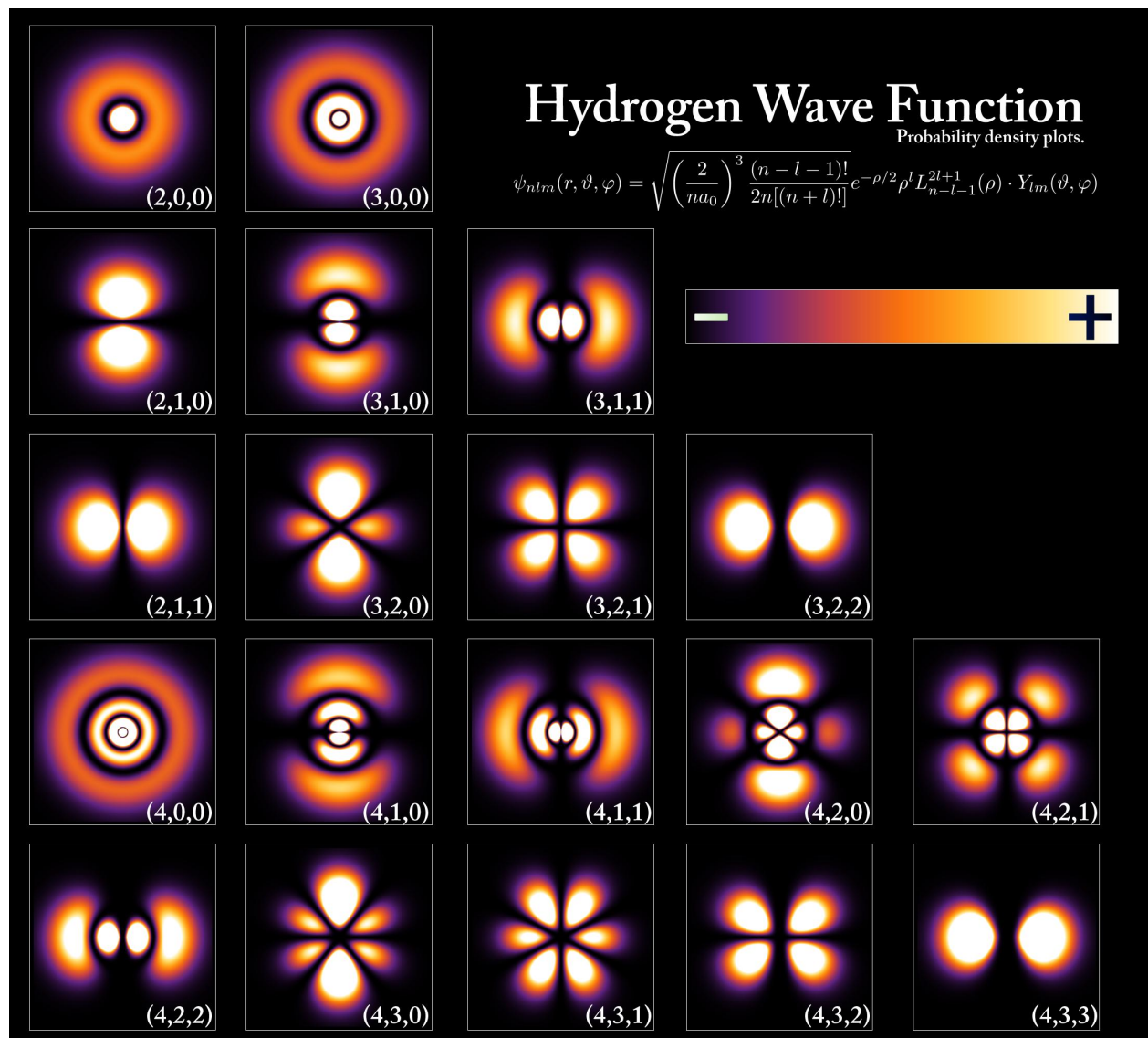
Before the Schrodinger equation was discovered in 1925, Bohr had proposed a model of the atom with a nucleus at the center and electrons in orbits about the nucleus. These electrons lived in shells: at most two electrons could occupy the first shell, at most 8 could occupy the second shell, at most 18 could occupy the third shell, and so forth. One piece of evidence for this was the periodic table of elements. Other evidence came from the study of spectral lines: when an element is heated, it glows, and a spectrogram shows that this glow consists of very specific frequencies of light. Physicists argued that electrons could jump from shell to shell, giving up energy in the process. Bohr predicted the energy of the electrons in each shell, and this idea then agreed with the various energies in these spectrums.

All of this was confirmed by Schrodinger, together with the Pauli exclusion principle. Each electron has a spin, s , which can take only two values. So there are four quantum numbers: n , l , m , s , and Pauli argued that no two distinct electrons can have the same four values.

The principal quantum number n corresponds to the energy of the electron and thus determines the shell. Recall that $l < n$ and m can take any integer value from $-l$ to l . So m has $2l + 1$ values, and l can take any value between 0 and $n - 1$. The total number of possibilities in the n th shell is therefore

$$2\left[1 + 3 + 5 + \dots + (2(n-1) + 1)\right] = 2\left[1 + 3 + 5 + \dots + (2n-1)\right]$$

Notice that $1 = 1^1$, $1 + 3 = 2^2$, $1 + 3 + 5 = 3^2$, $1 + 3 + 5 + 7 = 4^2$. Thus the number of electrons in the n th shell is $2n^2$. Since $n = 1, 2, 3, \dots$, the shells can have 2, 8, 18, $\dots$ electrons, as proposed by Bohr.

Figure 8.3: $n = 3, 4, 5, 6$; $l = 2$

Periodic Table of the Elements

1
IA
H
Hydrogen
1.008

3
IIA
Li
Lithium
6.94

11
IIIA
Na
Sodium
22.990

19
IIA
K
Potassium
39.098

37
IIA
Rb
Rubidium
85.468

55
IIA
Cs
Cesium
132.905

87
IIA
Fr
Francium
223

2
IIA
Be
Beryllium
9.012

12
IIA
Mg
Magnesium
24.305

20
IIA
Ca
Calcium
40.078

38
IIA
Sr
Strontium
87.62

56
IIA
Ba
Barium
137.327

88
IIA
Ra
Radium
226

13
IIIA
B
Boron
10.81

21
IIIB
Sc
Scandium
44.956

39
IIIB
Y
Yttrium
88.906

57-71
IIIB
Lanthanides

89-103
IIIB
Actinides

14
IIIA
C
Carbon
12.011

22
IIIB
Ti
Titanium
47.867

40
IIIB
Zr
Zirconium
91.224

72
IIIB
Hf
Hafnium
178.49

104
IIIB
Rf
Rutherfordium
261

15
IIIA
N
Nitrogen
14.007

23
IIIB
V
Vanadium
50.942

41
IIIB
Nb
Niobium
92.906

73
IIIB
Ta
Tantalum
180.948

105
IIIB
Db
Dubnium
262

16
IIIA
O
Oxygen
15.999

24
IIIB
Cr
Chromium
51.996

42
IIIB
Mo
Molybdenum
95.94

74
IIIB
W
Tungsten
183.84

106
IIIB
Sg
Seaborgium
266

17
IIIA
F
Fluorine
18.998

25
IIIB
Mn
Manganese
54.938

43
IIIB
Tc
Technetium
98

75
IIIB
Re
Rhenium
186.21

107
IIIB
Bh
Bohrium
264

18
IIIA
Ar
Argon
39.948

26
IIIB
Fe
Iron
55.845

44
IIIB
Ru
Ruthenium
101.07

76
IIIB
Os
Osmium
190.23

108
IIIB
Hs
Hassium
277

1
IIIA
Al
Aluminum
26.982

9
IIIB
Co
Cobalt
58.933

17
IIIB
Ni
Nickel
58.693

25
IIIB
Cu
Copper
63.546

33
IIIB
Zn
Zinc
65.38

41
IIIB
Ga
Gallium
69.723

49
IIIB
Ag
Silver
107.87

57
IIIB
Cd
Cadmium
112.41

65
IIIB
In
Indium
114.82

73
IIIB
Sn
Tin
118.71

81
IIIB
Sb
Antimony
121.76

89
IIIB
Te
Tellurium
127.6

97
IIIB
I
Iodine
126.90

105
IIIB
Xe
Xenon
131.29

113
IIIB
Ba
Barium
137.327

121
IIIB
La
Lanthanum
138.905

129
IIIB
Pr
Praseodymium
140.908

137
IIIB
Nd
Neodymium
144.24

145
IIIB
Pm
Promethium
145

153
IIIB
Sm
Samarium
150.36

161
IIIB
Eu
Europium
151.964

169
IIIB
Gd
Gadolinium
157.25

177
IIIB
Tb
Terbium
158.925

185
IIIB
Dy
Dysprosium
162.50

193
IIIB
Ho
Holmium
164.930

201
IIIB
Er
Erbium
167.259

209
IIIB
Tm
Thulium
168.934

217
IIIB
Yb
Ytterbium
173.054

225
IIIB
Lu
Lutetium
174.967

233
IIIB
Ac
Actinium
227

241
IIIB
Th
Thorium
232.04

249
IIIB
Pa
Protactinium
231.04

257
IIIB
U
Uranium
238.03

265
IIIB
Np
Neptunium
237

273
IIIB
Pu
Plutonium
244

281
IIIB
Am
Americium
243

289
IIIB
Cm
Curium
247

297
IIIB
Bk
Berkelium
247

305
IIIB
Cf
Californium
251

313
IIIB
Es
Einsteinium
252

321
IIIB
Fm
Fermium
257

329
IIIB
Md
Mendelevium
258

337
IIIB
No
Nobelium
259

345
IIIB
Lr
Lawrencium
262

Atomic Number →

← Symbol

← Atomic Weight

← Name

← Electrons per shell

State of matter (color of name)
GAS LIQUID SOLID UNKNOWN

Subcategory in the metal-metalloid-nonmetal trend (color of background)
Alkali metals Alkaline earth metals Transition metals Lanthanides Actinides Post-transition metals Reactive nonmetals Noble gases

Unknown chemical properties

Figure 8.4: Periodic Table

The periodic table of elements was proposed by Dimitri Mendeleev in 1869 and revised in the following years. This is before the discovery of quantum mechanics and Schrodinger's work. Yet the solutions of the Schrodinger equation explain many features of the table.

The table is shown above. The rows of the table correspond to the principal quantum number n and thus the shells of Bohr's model. Below the main table are two additional rows; these rows should be in the table on the last two rows, starting at the blank boxes in the third column. This would make the table impractically wide, so they are traditionally placed below the main table. Notice that the number of items on these rows is

$$2, 8, 8, 18, 18, 32, 32$$

and thus of the form $2n^2$. It is puzzling, however, that most of these numbers are repeated, which does not happen with the Schrodinger quantum numbers. We'll explain that.

Without the exclusion principle, the connection between Schrodinger's equation and the table would be tenuous. Luckily, the exclusion principle is a universal law throughout quantum field theory. It was discovered slightly after Schrodinger's equation. First chemists noticed that each quantum triple n, l, m seemed to occur for exactly two electrons in each atom. The discovery of spin explained the "two" and immediately implied the exclusion principle.

When we apply the theory to atoms with multiple electrons, we imagine inserting these electrons one by one and assume that each electron selects an energy as low as possible, modulo the exclusion principle. That is why the various shells are filled one by one.

Schrodinger's equation is for hydrogen only, because it describes a single electron. The corresponding equations for multiple electrons are too difficult to solve. Thus applying the quantum numbers n, l, m, s to electrons in general atoms is an approximation made by assuming that we can totally ignore the interaction of these electrons. Since electrons are very light, this interaction is weak enough to make the approximation useful. However, this is one crucial place that it is misleading.

In hydrogen, the quantum number n completely determines energy. The numbers l and m have no influence. But in multiple electron atoms, the extra energy from electron interaction makes electrons with higher values of l more energetic than those with l equal or close to zero. Sometimes this makes so much difference that electrons in the n th shell with large l have higher energy than electrons in the $n + 1$ st shell with low l . Therefore when we insert electrons into the atom one by one, higher shells are sometimes started before lower shells are complete.

Incidentally, this complication does not affect the m quantum number, which only fixes orientation of the electron distribution and does not contribute to energy considerations. Consequently, as soon as the lowest empty spot reaches a fixed n and l , the m values $-l, -l + 1, \dots, l$ are filled one by one. The chemists discovered a beautiful rule describing this filling known as Hund's rule. Once an (n, l, m) value is filled by one electron with a given spin, all possible other m values are filled with electrons of the same spin. Once these m 's are completely filled, the same values are filled again with electrons of opposite spin.

Let us return to the influence of l on energy. The chemists discovered an easy rule of thumb which covers almost all cases, called the “aufbau principle”. It is illustrated in the following diagram. Here the rows give the n quantum numbers, labeled from 1 through 8. In each row, the possible l values are shown, starting from 0 and ending at $n - 1$. Incidentally, the chemists label these s, p, d, f rather than $0, 1, 2, 3$. Then the dotted lines at 45 degrees describe the order of the energy levels from lowest to highest.

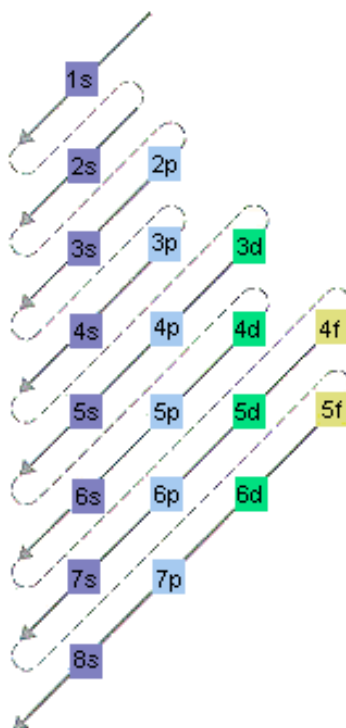


Figure 8.5: Aufbau Principle

Therefore, according to the diagram, the lowest energy is $n = 1$ followed by $n = 2$ with $l = 0$, and then $l = 1$, followed by $n = 3$ with $l = 0$ and $l = 1$. But then $n = 4$ with $l = 0$ has lower energy than $n = 3$ and $l = 2$.

Let us stop here and see how this is reflected in the table. Recall that each time we reach a new value of l , we expect to see $2l + 1$ boxes, corresponding to different values of m , followed by $2l + 1$ boxes with the same m but opposite spin. So $l = 0$ will just have two boxes, but $l = 1$ will have 6 boxes and $l = 2$ will have 10 boxes and $l = 3$ will have 14 boxes

So at the first level, we expect two boxes for $l = 0$. Sure enough, there are two boxes, one at extreme left and one at extreme right. At the second level, we expect two boxes

for $l = 0$ and 6 boxes for $l = 1$. Sure enough, there are two boxes at the left and then six boxes at the right.

Now look at the third level. Here is what happens as we fill in from the left.

$$(3, 0, m) \rightarrow (3, 1, m) \rightarrow (4, 0, m) \rightarrow (3, 2, m)$$

So we expect 2 boxes, followed by 6 boxes, followed by 2 boxes on the next level, followed by 10 boxes.

Notice that the 2 and then 6 boxes completely fill out the third row. So all atoms on that row are accounted for, but the third shell is not yet complete. There are still spaces for 10 more electrons. These places are filled in on the following row, as the elements just below the large gap in the middle, mostly blue boxes ending with a purple one. All of these elements have principal quantum number 3, so we might expect them on the third row, but in fact they are filled in after we already have two electrons in a higher shell.

One way to think of this is that if you want to find electrons with quantum numbers $(3, 2, m)$, there are such electrons in Sc through Zn, that is, in Scandium, Titanium, Vanadium, Chromium, Manganese, Iron, Cobalt, Nickel, Copper, and Zinc. Indeed if you want to find all possible electrons with principal quantum number 3, then Zinc is the first element containing all possibilities. However, Zinc also has some electrons with principal quantum number 4. Or to say it another way, Potassium and Calcium contain some electrons with principal quantum number 4, but no electrons with quantum numbers $(3, 2, m)$.

Let us carry this out for just one more level. The complete story starting with $n = 4$ and $l = 0$ according to the aufbau diagram is

$$(4, 0, m) \rightarrow (3, 2, m) \rightarrow (4, 1, m) \rightarrow (5, 0, m) \rightarrow (4, 2, m) \rightarrow (5, 1, m) \rightarrow (6, 0, m) \rightarrow (4, 3, m)$$

These will be 2 boxes, followed by 10 boxes, followed by 6 boxes followed by 2 boxes followed by 10 boxes followed by 6 boxes followed by 2 boxes followed by 14 boxes.

We already saw the first part of this story. We start on the fourth row and fill in the first two boxes, followed by the next 10. These ten boxes have $n = 3$, so they fill in an earlier shell. Then we fill in the last 6 boxes in the row. However, these are not all possible electrons in the fourth shell. We then start on the fifth row of the periodic table, filling in the first two boxes. Then we fill in the next 10 boxes on this fifth row, but these electrons only have $n = 4$. We then fill in 6 more boxes, completing the fifth row of the periodic table. But not only is the fifth shell incomplete, even the fourth shell isn't complete. We next fill in the first two elements of the 6th row, followed by the remaining 14 elements of the fourth shell. Notice that these 14 elements are on the first extra row below the table, so the final electrons added for each of these elements only have $m = 4$ even though they are on the sixth row of the periodic table.

Incidentally, the aufbau diagram is only a rule of thumb, and sometimes it gives the wrong result as determined by experiment. In the first 40 elements, it only makes two mistakes, for chromium and copper.

Chapter 9

Classification of Root Systems

9.1 Outline of Chapter and Cautionary Note

In the previous chapter, we introduced the *roots* of a compact Lie group. In this chapter, we study the full set of roots. By reinterpreting our theory, we can think of the roots as non-zero vectors in a Euclidean vector space, that is, a finite dimensional real vector space with inner product. This Euclidean space is just the Lie algebra of a maximal torus, with a left and right invariant metric as inner product, and the roots turn out to be a beautiful symmetric set of arrows in this space. Pictures of all possible such root systems in the two dimensional case are given later in the chapter.

If we select an arrow and reflect the picture in the hyperplane perpendicular to this arrow, this reflection maps roots to roots. So reflections induced by roots generate a finite group of symmetries, which turns out to be the Weyl group introduced in an earlier chapter.

The reflection groups obtained from compact Lie groups almost exhaust the set of all finite groups generated by reflections. The theory of such groups is the theory of kaleidoscopes, since a kaleidoscope produces beautiful symmetric patterns using only a small set of mirrors.

If we have a root system in R^k and another in R^l , we can consider their union as a root system in R^{k+l} . This new system is *reducible* as a union of orthogonal pieces; a root system which cannot be decomposed in this way is called *irreducible*. It is trivial to show that every root system is uniquely formed as a union of irreducible pieces.

We will then classify all irreducible root systems. Each such system gives rise to a *Dynkin diagram* and our classification lists all possible Dynkin diagrams.

We earlier proved that a real Lie algebra is the Lie algebra of a compact Lie group if

and only if it has the form $R^k \oplus \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_n$. Our main classification theorem states that the root systems of these $\mathcal{G}_i$ are irreducible and thus correspond to Dynkin diagrams. Moreover, we will prove that the diagram completely determines the Lie algebra $\mathcal{G}_i$.

Our Dynkin diagrams will be labeled $a_n, b_n, c_n, d_n, g_2, f_4, e_6, e_7, e_8$. In a later chapter we will show that the a_n, b_n, c_n, d_n come from the special unitary groups, the orthogonal groups, and the symplectic groups. We will probably skip the proof that g_2, f_4, e_6, e_7, e_8 also come from the Lie algebras of compact groups.

In this chapter we sometimes refer to compact Lie groups rather than compact Lie algebras. The corresponding classification theory for compact groups states that every compact Lie group has universal cover uniquely of the form $R^k \times G_1 \times \dots \times G_n$ where the G_i are compact simply connected Lie groups. These G_i are the universal covers of $SU(n), SO(n), Sp(n)$ and five exceptional groups G_2, F_4, E_6, E_7, E_8 . In particular, every compact, simply-connected Lie group has the form $G_1 \times \dots \times G_n$.

The cautionary note is that we will sometimes talk about compact Lie groups without assuming them simply connected. These groups yield the same Lie algebras, but sometimes results at the algebra level don't quite apply at the group level. For example, it will turn out that $so(4) = so(3) \oplus so(3)$; this is the only such "decomposition" of a rotation group. It does not follow that $SO(4) = SO(3) \times SO(3)$. But if we write universal covers, then $Spin(4)$ is indeed $Spin(3) \times Spin(3)$. We always have $SO(n) = Spin(n)/\pm I$. In the special case $n = 4$, the center of $Spin(3) \times Spin(3)$ is $Z_2 \times Z_2$, but $SO(4)$ is obtained from $Spin(3) \times Spin(3)$ by dividing out $\{0 \times 0, 1 \times 1\}$ rather than the full center of all four elements. Details of all of this will be provided later.

9.2 Brackets of Root Spaces

So far, we only studied one root at a time, calling it f . Our torus can be identified with its Lie algebra modulo a lattice, so $T = \mathcal{T}/\mathcal{L}$. Recall that f is a non-zero map $f : \mathcal{T} \rightarrow R$ such that $f(\mathcal{L}) \subset Z$.

When more than one root is being studied, it is conventional to name roots $\alpha, \beta, \dots$ rather than $f, g, \dots$ as done earlier, because g has other meanings. We will adopt this new convention.

We will usually write root spaces as two-dimensional real subspaces of $\mathcal{G}$. But from time to time it is useful to think of them as one-dimensional complex subspaces of $\mathcal{G} \otimes C$. Recall the connection between these ideas.

If α is a root of G , we define $E_\alpha \subset \mathcal{G} \otimes C$ to be the set of all complex vectors e_α such that whenever $t \in T$ we have

$$Ad(t)e_\alpha = e^{2\pi i \alpha(t)} e_\alpha$$

or equivalently whenever $X \in \mathcal{T}$ we have

$$[X, e_\alpha] = 2\pi i \alpha(X) e_\alpha$$

If α is a root associated with e_α , then $-\alpha$ is a root associated with $\overline{e_\alpha}$ because conjugating the above equation gives

$$Ad(t)\overline{e_\alpha} = e^{-2\pi i \alpha(t)} \overline{e_\alpha}$$

If we write $e_\alpha = u + iv$ and $e^{2\pi i \alpha(t)} = \cos(2\pi \alpha(t)) + i \sin(2\pi \alpha(t))$, we obtain our previous two-dimensional representation in $\mathcal{G}$, and then α and $-\alpha$ give the same two-dimensional representation.

Theorem 53

$$[E_\alpha, E_\beta] \subset E_{\alpha+\beta}$$

Proof: If $X \in \mathcal{T}$,

$$\begin{aligned} [X, [e_\alpha, e_\beta]] &= [[X, e_\alpha], e_\beta] + [e_\alpha, [X, e_\beta]] = [2\pi i \alpha(X) e_\alpha, e_\beta] + [e_\alpha, 2\pi i \beta(X) e_\beta] \\ &= 2\pi i (\alpha(X) + \beta(X)) [e_\alpha, e_\beta] \end{aligned}$$

Note: In particular, if $\alpha + \beta$ is not a root, then $[e_\alpha, e_\beta] = 0$.

Note: We cannot conclude that $[E_\alpha, E_\alpha] = 0$ because we do not yet know that E_α has dimension one. This will eventually turn out to be true. It will also turn out to be true that if α is a root, then 2α is not a root.

Note: We know that if α is a root, so is $-\alpha$. The above proof shows that $[E_\alpha, E_{-\alpha}] \subset \mathcal{T} \otimes C$, since its vectors commute with all $X \in \mathcal{T}$.

9.3 Complex Root Spaces are One Dimensional

We have proved that $su(2) \subset \mathcal{G}$ and therefore the Lie algebra $su(2)$ acts via ad on $\mathcal{G} \otimes C$. We will now use the representation theory of $su(2)$ to understand $\mathcal{G}$.

We often employ an ingenious trick. We fix a root α , and notice that $u + iv \in E_\alpha$ and $u - iv \in E_{-\alpha}$. Then we select a subspace $V \subset \mathcal{G} \otimes C$ which is a sum of *some* root spaces, $V = \sum E_\beta$. We will make certain that $E_{\alpha+\beta}$ is in the sum if $\alpha + \beta$ is a root, and that $E_{-\alpha+\beta}$ is in the sum if $-\alpha + \beta$ is a root. It will then follow that V is invariant under $ad(u + iv)$ since $[E_\alpha, E_\beta] \subset E_{\alpha+\beta}$. Similarly V is invariant under $ad(u - iv)$. Therefore it is invariant under $ad(u)$ and $ad(v)$ separately. It is surely invariant under $ad(t_\alpha)$ since every element of E_β is an eigenvector of all elements of $\mathcal{T}$. We'll conclude that $su(2)$ acts in the space, and therefore that $sl(2, R)$ acts on the space. We'll then apply the much easier representation theory of $sl(2, R)$.

Finally, a word about our choice of H . In the section of the previous chapter on the many faces of $\mathcal{T}$, we determined that $\mathcal{T}$ acts on $\mathcal{G} \otimes \mathcal{C}$ by $ad(X)e_f = 2\pi i f(X)e_f$, or in the notation of this chapter, $ad(X)e_\alpha = 2\pi i \alpha(X)e_\alpha = 2\pi i \langle t_\alpha, X \rangle e_\alpha$. Later we defined $H = -iw = -i \frac{t_\alpha}{\pi \|t_\alpha\|^2}$. So $ad(H) = \frac{-i}{\pi \|t_\alpha\|^2} ad(t_\alpha)$. and

$$ad(H)e_\alpha = [H, e_\alpha] = \frac{-i}{\pi \|t_\alpha\|^2} [ad(t_\alpha), e_\alpha] = \frac{-i}{\pi \|t_\alpha\|^2} (2\pi i \langle t_\alpha, t_\alpha \rangle e_\alpha) = 2e_\alpha$$

Similarly, for any $e_\beta \in E_\beta$

$$ad(H)e_\beta = [H, e_\beta] = \frac{-i}{\pi \|t_\alpha\|^2} [ad(t_\alpha), e_\beta] = \frac{-i}{\pi \|t_\alpha\|^2} (2\pi i \langle t_\beta, t_\alpha \rangle e_\beta) = 2 \frac{\langle t_\alpha, t_\beta \rangle}{\|t_\alpha\|^2} e_\beta$$

Theorem 54 *Let α be a root for $\mathcal{G}$ and $\mathcal{T}$. Then the only real multiples of α that are roots are α and $-\alpha$, and the corresponding complex root spaces have complex dimension one.*

Apply the representation theory of $sl(2, R)$ to $V = \sum_r E_{r\alpha}$ where E_α is the set of all complex vectors w such that $Ad_t(w) = e^{2\pi i \alpha(t)} w$. Let the sum run over all real r such that $r\alpha$ is a root. Include $E_0 = \mathcal{T} \otimes \mathcal{C}$ in this sum. Notice that the trick applies, so V is invariant under $sl(2, R)$.

Consider $\mathcal{T}$. The element $t_\alpha \in \mathcal{T}$; let $\mathcal{T}_0$ be its orthogonal complement with respect to the invariant inner product $\langle \cdot, \cdot \rangle$. Recall that $[t, u] = -2\pi \alpha(t)v = -2\pi \langle t_\alpha, t \rangle v$. Therefore if $t \in \mathcal{T}_0$ we have $[u, t] = -[t, u] = 0$. Similarly $[v, t] = 0$ and so $[e_\alpha, t] = 0$ and $[\bar{e}_\alpha, t] = 0$. Of course $[t_\alpha, t] = 0$. It follows that $su(2)$ acts trivially on $\mathcal{T}_0 \otimes \mathcal{C}$, which is thus a sum of one-dimensional invariant subspaces. The complementary invariant subspace is $\sum_{r \neq 0} E_{rf} \oplus Ct_\alpha$, and from now on we work with it.

This space is invariant under $sl(2, R)$, so it is a direct sum of irreducible representation spaces. From the representation theory for $sl(2, R)$ we know that the eigenvalues of H on any irreducible subspace are integers separated by 2 and symmetric about the origin, for instance $-4, -2, 0, 2, 4$ or $-5, -3, -1, 1, 3, 5$. In particular, the complex subspace with basis $e_{-\alpha}, t_\alpha, e_\alpha$ is irreducible, and in the irreducible representation of dimension 3 the operator H has eigenvalues $-2, 0, 2$.

If E_α had a second element linearly independent of e_α , say d_α , then this element would generate an irreducible representation containing elements from *at least* $E_{-\alpha}, Ct_\alpha, E_\alpha$ with eigenvalues $-2, 0, 2$. However, the ladder operators would move these elements up and down starting from the middle element t_α and we already know that when applied to t_α we only get $e_{-\alpha}, t_\alpha, e_\alpha$. So all root spaces have dimension one.

Our initial formulas for $[H, e_\beta] = 2 \frac{\langle t_\alpha, t_\beta \rangle}{\|t_\alpha\|^2} e_\beta$ shows that if $r\alpha$ is a root, then $2r$ is an eigenvalue of H and therefore $2r$ is an integer. If $2r$ is an *even* integer which is positive, then the ladder operators will produce *at least* all eigenvalues in the set

$$\{-2r, -2r + 2, \dots, 0, \dots, 2r - 2, 2r\}$$

but the only choice for the eigenvector with eigenvalue 0 is t_α and so this chain can only have three elements. The case when $2r < 0$ is handled similarly.

We have proved that if α is a root, then 2α is not a root. It follows that $\frac{1}{2}\alpha$ cannot be a root, since then α would not be a root.

So $r = -1$ or $r = 1$ or $r = \pm\frac{1}{2}, \pm\frac{3}{2}, \dots$. Recall that if $r\alpha$ is a root, then $2r$ is an eigenvalue of H . So if one of these fractions is a root, then it generates an irreducible representation whose H has at least eigenvalues $\{-1, 1\}, \{-3, -1, 1, 3\}, \{-5, -3, -1, 1, 3, 5\}$, etc. In particular 1 is an eigenvalue and thus $\frac{1}{2}\alpha$ is a root, contradicting an earlier assertion. QED.

9.4 Another Application of $su(2)$ Representation Theory

Theorem 55 *Let α and β be roots of $\mathcal{G}$ with respect to $\mathcal{T}$. Then $\beta + k\alpha$ is a root exactly for a range of integers $p \leq k \leq q$, where $p \leq 0$ and $q \geq 0$. Moreover,*

$$\frac{2 \langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} = -(p + q)$$

Proof: First we handle the trivial case when β is a multiple of α . By the previous theorem we have $\beta = \pm\alpha$. If $\beta = \alpha$, the string is $\beta - 2\alpha = -\alpha, \beta - \alpha = 0, \beta = \alpha$, so $p = -2$ and $q = 0$ and $-(p + q) = 2 = 2 \frac{\langle t_\alpha, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle}$. The case $\beta = -\alpha$ is similar.

In all other cases, let $V = \sum E_{\beta+k\alpha}$ where the sum is over roots of the form $\beta + k\alpha$. Notice that this vector space contains no non-zero elements of $\mathcal{T}$. Act on V using the $su(2)$ associated with α . Thus up to constants $L_+ = e_\alpha$ and $L_- = \bar{e}_\alpha$. Moreover $H = -i \frac{t_\alpha}{\pi \langle t_\alpha, t_\alpha \rangle}$. Clearly our space is invariant under these operators and breaks up into a sum of irreducible representations. The eigenvalues of each irreducible cannot contain 0 because V does not intersect $\mathcal{T}$, and so have the form $\dots, -1, 1, \dots$. Since the $E_{\beta+k\alpha}$ are one-dimensional, the eigenvalues -1 and 1 occur in only one of the irreducible representations, so the full space is irreducible and thus the $\beta + k\alpha$ which are roots form a single string containing β . This string must then have the form $\beta + p\alpha, \dots, \beta + q\alpha$ with $p \leq 0$ and $q \geq 0$.

Recall that if $u + iv \in E_f$ then $[t, u] = -2\pi f(t)v$ and $[t, v] = 2\pi f(t)u$. So

$$[H, u + iv] = \left[-i \frac{t_\alpha}{\pi \|t_\alpha\|^2}, u + iv \right] = -i \frac{f(t_\alpha)}{\pi \|t_\alpha\|^2} (-2\pi v + 2\pi iu) = \frac{2f(t_\alpha)}{\|t_\alpha\|^2} (u + iv)$$

In the special case $f = \beta + k\alpha$ and $u + iv \in E_{\beta+k\alpha}$, the resulting eigenvalue is

$$2 \frac{\beta(t_\alpha) + k\alpha(t_\alpha)}{\langle t_\alpha, t_\alpha \rangle} = 2 \frac{\langle t_\beta, t_\alpha \rangle + k \langle t_\alpha, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} = \frac{2 \langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} + 2k$$

But for an irreducible representation, the lowest and highest eigenvalues are equal up to sign, so

$$\frac{2 \langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} + 2p = - \left(\frac{2 \langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} + 2q \right)$$

and so

$$\frac{4 \langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} = -2(p + q)$$

and the result in the theorem follows. QED.

9.5 Easy Consequences of Root Results

Theorem 56 *If α, β , and $\alpha + \beta$ are roots, $[E_\alpha, E_\beta] = E_{\alpha+\beta}$.*

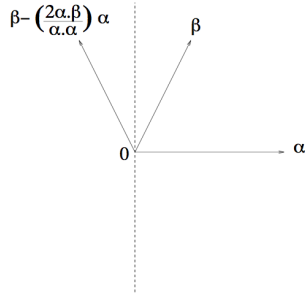
All of these spaces are one-dimensional, so the claim is that $[e_\alpha, e_\beta] \neq 0$. Consider the chain

$$\beta + p\alpha, \dots, \beta, \beta + \alpha, \dots, \beta + q\alpha$$

Since $\beta + \alpha$ is a root, it must be in the chain. The raising ladder operator is essentially bracket with e_α , and this is never zero within the chain. QED.

Theorem 57 *If α and β are roots, then the expression below is also a root*

$$\beta - 2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} \alpha$$



Proof: By the main theorem in the previous section, our expression equals $\beta + (p + q)\alpha$ and this is in the previous chain because $p \leq (p + q) \leq q$. QED.

Remark: The above result is a formula between roots, which live in $\mathcal{T}^*$. But the map $\mathcal{T}^* \rightarrow \mathcal{T}$ is linear, so the formula also holds for root vectors. Thus if t_α and t_β are root vectors, so is

$$t_\beta - 2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} t_\alpha$$

This is a very significant result because it has a geometric interpretation. The map reflects t_β across the hyperplane perpendicular to t_α . See the diagram above. It is easy to verify that $v \rightarrow v - 2 \frac{\langle t_\alpha, v \rangle}{\langle t_\alpha, t_\alpha \rangle} v$ is such a reflection, because if v is perpendicular to t_α then the map is $v \rightarrow v$, and if $v = \lambda t_\alpha$, then the map is $v \rightarrow \lambda v - 2\lambda v = -\lambda v$.

Remark: Consider the subspace of $\mathcal{T}$ generated by the roots t_α . The root vectors yield a finite number of vectors in this vector space. Each reflection perpendicular to a root permutes the roots. It follows that these reflections generate a finite group W , which acts on the roots. Eventually we will prove that this group is the Weyl group introduced earlier. In the 1930's, the Canadian mathematician Coxeter classified all finite subgroups of the orthogonal group which are generated by reflections. It turns out that his classification is almost the same as our classification of compact Lie algebras.

Groups generated by reflections are very closely related to kaleidoscopes, since a kaleidoscope is formed by a small set of mirrors (usually two) and yet generates a pattern with a large symmetry group, often the dihedral group D_n . So in a sense, we have discovered that every compact Lie group has a hidden kaleidoscope inside it, which almost completely determines the compact group.

9.6 Restrictions on Angles and Lengths of Root Vectors

Remark: We proved in the previous sections that $2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle \langle t_\beta, t_\beta \rangle}$ is an integer. It follows that

$$2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} 2 \frac{\langle t_\beta, t_\alpha \rangle}{\langle t_\beta, t_\beta \rangle} = 4 \left\langle \frac{t_\alpha}{\|t_\alpha\|}, \frac{t_\beta}{\|t_\beta\|} \right\rangle^2$$

is an integer. So if θ is the angle between t_α and t_β , then $4 \cos^2 \theta$ is an integer. Therefore $4 \cos^2 \theta = 4, 3, 2, 1, 0$ and $\cos \theta = \pm 1, \pm \frac{\sqrt{3}}{2}, \pm \frac{1}{\sqrt{2}}, \pm \frac{1}{2}, 0$. If we are allowed to replace vectors by their negatives, the angle between root vectors in degrees is 0 or 180, 30 or 150, 45 or 135, 60 or 120, or 90.

If the angle is not 0 or 90 degrees, we can divide the two expressions rather than multiplying and compare the lengths of these root vectors. Taking absolute values of terms, each term of the product is a positive integer. If $\theta = 30$ or 150 , one term is 1 and the other is 3, so the ratio of longer to shorter lengths is $\sqrt{3}$. If $\theta = 45$ or 135 , one term is 1 and the other is 2, so the ratio of longer to shorter lengths is $\sqrt{2}$. If $\theta = 60$ or 120 , both terms are 1 and the vectors have equal length.

9.7 Pictures

There are four possible root systems in two dimensions. The first is an orthogonal product of the one possible system in one dimension, which has two opposite roots. This one

dimensional system corresponds to $su(2)$, and it turns out that orthogonal products of root systems come from direct sums of the corresponding Lie groups. So this first system corresponds to $su(2) \oplus su(2)$. Since the pieces are orthogonal, there is no restriction on their lengths. We could choose one left and right invariant metric on the first group, and another on the second, and this would lead to different lengths in the two pieces.

These are pictures in $\mathcal{T}$; the roots are labeled α, β , but these are really t_α and t_β .

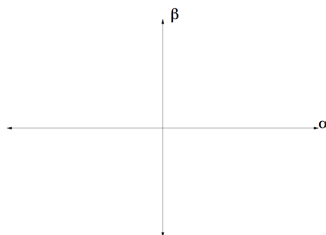


Figure 9.1: $su(2) \oplus su(2)$

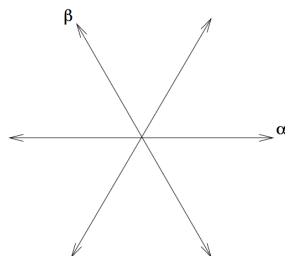
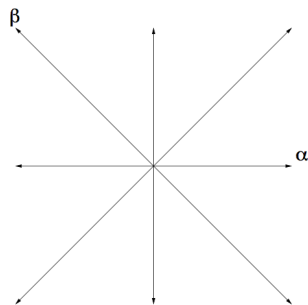
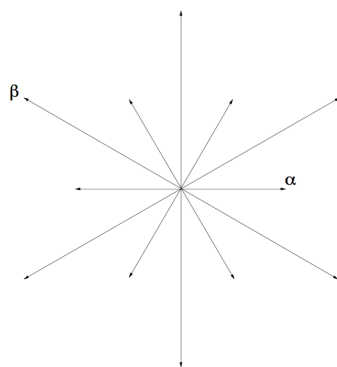


Figure 9.2: $su(3)$

Figure 9.3: $so(5)$ Figure 9.4: g_2

The next system, $su(3)$ is the simplest non-product. All the vectors have equal length, which is forced by the angle between these vectors and the theorem of the previous section.

The next root system corresponds to $so(5)$, since $so(3)$ and $so(4)$ are related to earlier systems. Indeed $so(3) = su(2)$ and $so(4) = so(3) \oplus so(3) = su(2) \oplus su(2)$. Notice the string of roots $\beta, \beta + \alpha, \beta + 2\alpha$.

The final system is more complicated, and corresponds to one of the exceptional groups. Notice the string of roots $\beta, \beta + \alpha, \beta + 2\alpha, \beta + 3\alpha$.

9.8 Simple Roots

We have been making a distinction between roots $\alpha \in \mathcal{T}^*$ and the associated $t_\alpha \in \mathcal{T}$. The time has come to simplify the notation. From now on we will identify α and t_α and suppose that roots always belong to $\mathcal{T}$. In particular, an expression involving inner products, like $\langle t_\alpha, t_\beta \rangle$ will now be written $\langle \alpha, \beta \rangle$. When roots were first introduced, we wrote $Ad(t)(u + iv) = e^{2\pi i f(t)}(u + iv)$. Then we began calling roots α instead of f , and now we would write the formula

$$Ad(t)(u + iv) = e^{2\pi i \langle \alpha, t \rangle}(u + iv)$$

This convention already occurred in the pictures in the previous section, where we wrote α and β instead of t_α and t_β .

In each picture in the previous section, two roots α and β are labeled. Notice that every other root in the picture is a linear combination of these roots *with integer coefficients*. Moreover, these coefficients are either *all nonnegative* or else *all nonpositive*.

In this section, we show that every root system has a similar basis.

Suppose, then, that we have the root system from a compact Lie group, and suppose these roots generate a real vector space W . Pick a hyperplane in W which does not contain any roots. This plane divides the complement into two pieces. Arbitrarily choose one of these pieces and call the set P of roots in this piece the *positive roots*. Their negatives are then in the other set, which we call the *negative roots*.

Choose an orthonormal basis $w_1, w_2, \dots, w_k$ for W such that $w_1, w_2, \dots, w_{k-1}$ is a basis for the chosen hyperplane and w_k points toward the positive side of this hyperplane. Since there are only finitely many roots, there are a finite number of positive w_k components for positive roots.

In the two dimensional pictures from the previous section, this hyperplane will be a line, and α and β are on the positive portion of this line.

Call a positive root *simple* if it cannot be written as a sum of two other positive roots. In the pictures from the previous section, check that α and β are simple, and no other positive root is simple.

We claim that simple roots exist, and indeed that every positive root is a sum of simple roots with non-negative integer coefficients. To prove this, consider the w_k component of the positive roots. This component has only finitely many levels. We prove the assertion by induction on level, starting at the lowest level. Each root in the lowest level is simple, since if it is a sum of two positive roots, each of these roots would have lower level.

Now consider roots in the second layer. Some of these roots may be simple. The rest are sums of two roots at the lower level, and thus linear combinations of simple roots with non-negative integer coefficients.

Continue the process. Either a root in the third layer is simple, or else it is a sum of two roots in lower levels, and all roots in lower levels are sums of simple roots with non-negative integer coefficients. Continue. QED.

Theorem 58 *If $\alpha_1, \dots, \alpha_n$ are simple, then*

- *these roots are linearly independent*
- *every positive root is uniquely $\sum n_i \alpha_i$ where the n_i are non-negative integers*
- *every negative root is uniquely $\sum n_i \alpha_i$ where the n_i are non-positive integers*

Remark: Suppose for a moment that we have proved this theorem. Notice that an easy consequence is that the list of roots in the previous section is complete. Indeed, there must be two simple roots α and β and the angle between them must be 30, 45, 60, or 90 degrees. This determines the ratios of the lengths of the vectors, and then the other positive roots. The last of these conclusions requires a little work left to the reader.

Lemma 11 *If α and β are distinct simple roots, then $\langle \alpha, \beta \rangle \leq 0$.*

Proof: Notice that $\beta - \alpha$ is not a root, for if it were a positive root, then $\beta = (\beta - \alpha) + \alpha$, a contradiction, and if it were negative then $\alpha = (\alpha - \beta) + \beta$, another contradiction. Therefore in the string below

$$\beta + p\alpha, \dots, \beta, \beta + \alpha, \dots, \beta + q\alpha$$

we have no $\beta - \alpha$ and so $p = 0$. Therefore $2 \frac{\langle t_\beta, t_\alpha \rangle}{\langle t_\alpha, t_\alpha \rangle} = -(p + q) = -q$ is less than or equal to zero, so $\langle t_\beta, t_\alpha \rangle \leq 0$.

Lemma 12 *The simple roots are linearly independent over R .*

Proof: Suppose $\sum r_i \gamma_i = 0$ where the γ_i are simple roots. Let $\alpha = \sum_{r_i > 0} r_i \gamma_i$ and $\beta = \sum_{r_i < 0} (-r_i) \gamma_i$, so $\alpha - \beta = 0$. Notice that both α and β are positive vectors, and notice that $\langle \alpha, \beta \rangle \leq 0$, since both vectors have positive coefficients and $\langle \gamma_i, \gamma_j \rangle \leq 0$ for $i \neq j$.

Then $0 = \langle 0, 0 \rangle = \langle \alpha - \beta, \alpha - \beta \rangle = \langle \alpha, \alpha \rangle - 2\langle \alpha, \beta \rangle + \langle \beta, \beta \rangle$ is greater than or equal to $\langle \alpha, \alpha \rangle + \langle \beta, \beta \rangle$. So $\alpha = \beta = 0$. But a linear combination of positive roots with positive coefficients is positive unless all coefficients are zero. So all the coefficients of our original linear expression are zero.

Completion: Clearly the full theorem follows from these results.

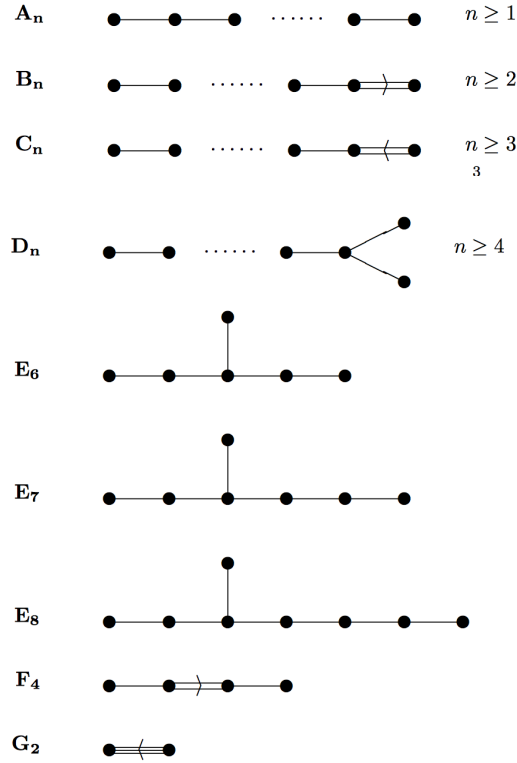
9.9 Dynkin Diagrams

Eugene Dynkin is a Russian mathematician who was asked to write an expository paper on Lie algebras while still a graduate student. He wrote an elegant treatment often read today. In particular, he invented a convenient diagram to describe systems of simple roots. Dynkin eventually came to the United States, and became professor of statistics at Cornell. He died in 2014.

Start with a collection of simple roots. For each root, make a solid dot on a plane. If α and β are simple roots, we have proved that their inner product is less than or equal to zero and that the angle between them is 150, 135, 120, or 90. Connect the corresponding dots by a line if the angle is 120, by two lines if the angle is 135, and by three lines if the angle is 150. In the latter two cases, the roots have different lengths; draw an arrow on the lines from the longer root to the shorter root. Do not connect perpendicular roots.

A Dynkin diagram can be split into connected components. We will classify these.

Theorem 59 *The following is a complete list of connected admissible Dynkin diagrams. The labels come from Cartan's treatment. The associated algebras for the infinite families are $A_n = su(n+1)$, $B_n = so(2n+1)$, $C_n = sp(2n)$, $D_n = so(2n)$*



Proof: We will eliminate possible diagrams until only a few are left. We eliminate diagrams by assigning positive real numbers to some nodes and zero to the rest. This assignment is equivalent to defining an element $v = \sum r_i \alpha_i$. We then calculate $v \cdot v$. If the result is zero or negative, the diagram is impossible because $v \neq 0$. Indeed no diagram can contain a sub-diagram equal to the nodes attached to nonzero r_j . Notice that $\sum r_i \alpha_i = \sum (r_i \|\alpha_i\|) \frac{\alpha_i}{\|\alpha_i\|}$. So we can assume simple roots have length one and ignore arrows.

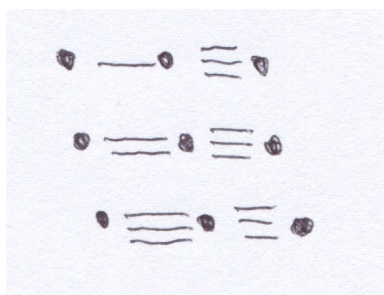
We can also ignore extra edges that may connect roots with $r_i \neq 0$, since these edges only contribute negative values of the cosine and thus make the final result even more negative.

Lemma 13 *The follow subdiagrams are impossible.*

Proof: Attach $r_1 = \frac{1}{2}, r_2 = 1, r_3 = \frac{\sqrt{3}}{2}$ to the three nodes in the first diagram. The resulting value for $v \cdot v$ is

$$r_1^2 + r_2^2 + r_3^2 + 2 \left(-\frac{1}{2} \right) r_1 r_2 + 2 \left(\frac{-\sqrt{3}}{2} \right) r_2 r_3 = \frac{1}{4} + 1 + \frac{3}{4} - \frac{1}{2} - \frac{3}{2} = 2 - 2 = 0$$

So the first diagram is impossible. The second diagram changes $-\frac{1}{2}$ to $-\frac{1}{\sqrt{2}}$ in this expression, making the sum negative, so it is also impossible. The final diagram makes the sum



even more negative. QED.

Corollary 7 *The only connected admissible Dynkin diagram with a three-edge line is G_2 .*

Lemma 14 *An admissible Dynkin diagram contains no cycles.*

Proof: Consider the following diagram and attach $r_i = 1$ to each cycle. Then the resulting sum is

$$\sum r_i^2 + 2 \sum_{i < n} \left(-\frac{1}{2} r_i r_{i+1} \right) + 2 \left(-\frac{1}{2} \right) r_n r_1 = n - (n-1) - 1 = 0$$

Adding additional lines between some nodes just makes the negative terms more negative. QED



Corollary 8 *An admissible Dynkin diagram must be a tree.*

Lemma 15 *A connected admissible Dynkin diagram cannot contain two edges with double lines.*



Proof: It suffices to show that the above diagram is impossible. Attach $\frac{1}{\sqrt{2}}$ to the vertices at both ends and 1 to remaining vertices. The resulting sum is

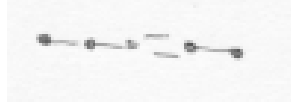
$$\left(\frac{1}{\sqrt{2}} \right)^2 + n + \left(\frac{1}{\sqrt{2}} \right)^2 - 2 \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} - 2(n-1) \frac{1}{2} - 2 \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} = (n+1) - 1 - (n-1) - 1 = 0$$

Lemma 16 *A connected admissible diagram cannot contain an edge with two sides and a node connected to at least three edges.*



Proof: This time we will give a slight indication of the method used to assign numbers to vertices. Attach a to the two left nodes, b to the right node, and 1 to the middle nodes. The contribution of these middle nodes to the sum is thus $n - (n - 1) \left(2 \cdot 1 \cdot 1 \cdot \frac{1}{2} \right) = 1$. The additional contribution from the left nodes is $a^2 + a^2 - 2(2 \cdot a \cdot 1 \cdot \frac{1}{2}) = 2(a^2 - a)$. The additional contribution from the right node is $b^2 - 2 \cdot 1 \cdot b \cdot \frac{1}{\sqrt{2}} = b^2 - \sqrt{2}b$. The minimum value of $2(a^2 - a)$ occurs when the derivative vanishes, so when $a = \frac{1}{2}$. This minimum is $-\frac{1}{2}$. The minimum value of $b^2 - \sqrt{2}b$ occurs when the derivative vanishes, so when $b = \frac{1}{\sqrt{2}}$. This minimum is also $-\frac{1}{2}$. So if we assign $\frac{1}{2}$ to the left nodes, $\frac{1}{\sqrt{2}}$ to the right node, and 1 to the remaining nodes, the value is $-\frac{1}{2} + 1 - \frac{1}{2} = 0$.

Lemma 17 *The diagram below is not admissible.*



Proof: From left to right, attach 1, 2, 3, $2\sqrt{2}$, $\sqrt{2}$. The sum is

$$1 + 4 + 9 + 8 + 2 - 2 \cdot 1 \cdot 2 \cdot \frac{1}{2} - 2 \cdot 2 \cdot 3 \cdot \frac{1}{2} - 2 \cdot 3 \cdot 2\sqrt{2} \cdot \frac{1}{\sqrt{2}} - 2 \cdot 2\sqrt{2} \cdot \sqrt{2} \cdot \frac{1}{2} = 0$$

Discussion: Suppose an admissible diagram has a double edge (we know it has only one). If this side has no edges attached to either end, it is B_2 . If it has edges attached to only one side, these edges must form a simple chain, so the diagram is B_n or C_n .

Suppose it has edges attached to both ends. Neither of these edges can start a chain of length greater than one by a previous lemma, and neither can split by the previous lemma. So the diagram is F_4 .

Discussion: From now on, all diagrams contain edges with one line.

If a diagram contains a single chain, it is A_n . Otherwise, at least one node must be attached to three or more sides. There cannot be two such nodes by the following lemma. This lemma also rules out the possibility that a node has more than three sides.

Lemma 18 *The diagram below is not admissible.*



Proof: Attach $\frac{1}{2}$ to the four nodes at the ends, and 1 to every other node. Then we get

$$\frac{1}{4} + \frac{1}{4} + n + \frac{1}{4} + \frac{1}{4} - 2 \left(\frac{1}{2} \right) \left(\frac{1}{2} \right) - 2 \left(\frac{1}{2} \right) \left(\frac{1}{2} \right) - 2(n-1) \left(\frac{1}{2} \right) - 2 \left(\frac{1}{2} \right) \left(\frac{1}{2} \right) - 2 \left(\frac{1}{2} \right) \left(\frac{1}{2} \right) = 0$$

Discussion: Otherwise our diagram must be a three-sided node, each attached to “single line” chains. If two of these chains have length one, we have D_n . The case where none of the chains has length one is ruled out by the following lemma.

Lemma 19 *The diagram below is not admissible.*

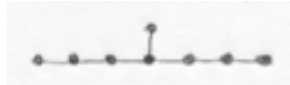


Proof: Attach 3 to the central node, attach 2 to the nodes closest to it, and attach 1 to the three end nodes. Then

$$1 + 1 + 1 + 4 + 4 + 4 + 9 - 3 \left(2 \cdot 1 \cdot 2 \cdot \frac{1}{2} \right) - 3 \left(2 \cdot 2 \cdot 3 \cdot \frac{1}{2} \right) = 0$$

Discussion: Otherwise our diagram is a three-sided node attached to three chains. One chain has length one and the other chains have length at least two. The case where both remaining chains have length at least three is ruled out by the following lemma.

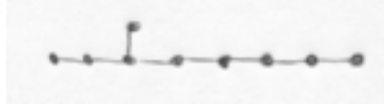
Lemma 20 *The diagram below is not admissible.*



Proof: Attach 1, 2, 3, 4, 3, 2, 1 to the horizontal chain segment, and 2 to the node above them. Then

$$2(1 + 4 + 9) + 16 + 4 - 2 \left(2 \cdot 1 \cdot 2 \cdot \frac{1}{2} \right) - 2 \left(2 \cdot 2 \cdot 3 \cdot \frac{1}{2} \right) - 2 \left(2 \cdot 3 \cdot 4 \cdot \frac{1}{2} \right) - 2 \cdot 4 \cdot 2 \cdot \frac{1}{2} = 0$$

Discussion: Otherwise one chain has length one, one chain has length two, and the final chain has length two or more. The cases where the final chain has length 2, 3, or 4 give E_6, E_7, E_8 . All remaining cases are ruled out by the following lemma.



Lemma 21 *The diagram below is not admissible.*

Proof: Attach 2, 4, 6, 5, 4, 3, 2, 1 to the horizontal chain from left to right, and attach 3 to the node above them. Then

$$4+16+36+25+16+9+4+1+9-2\cdot\frac{1}{2}(2\cdot 4+4\cdot 6+6\cdot 5+5\cdot 4+4\cdot 3+3\cdot 2+2\cdot 1+6\cdot 3)=0$$

QED.

9.10 Dynkin Diagrams Determine All Roots

Lemma 22 *Suppose β is a positive non-simple root. Then there is a simple root α such that $\beta - \alpha$ is a root.*

Proof: Suppose not. Then for any simple α , $\beta - \alpha$ is not a root. So the α string through β must be $\beta, \beta + \alpha, \dots, \beta + q\alpha$ and $2\frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} = -q$ is less than or equal to zero, so $\langle \beta, \alpha \rangle \leq 0$. But β is positive, so it is a sum of simple roots with positive coefficients, and this leads to the following contradiction:

$$\langle \beta, \beta \rangle = \langle \beta, m_1\alpha_1 + \dots + m_k\alpha_k \rangle \leq 0$$

Lemma 23 *Every positive root can be written in the form $\alpha_1 + \dots + \alpha_n$, where the α_i are not necessarily distinct simple roots, such that for each j , $\alpha_1 + \dots + \alpha_j$ is a root.*

Proof: Every positive root is uniquely a linear sum of simple roots with non-negative integer coefficients. The sum of these coefficients is called the *height* of the root. We prove the result by induction on height, noting that roots of height one are themselves simple.

Let β have height $m + 1$. Then there is a simple α such that $\beta - \alpha$ is a root, and $\beta = (\beta - \alpha) + \alpha$. If $\beta - \alpha$ is positive, then it has height m and we are done by induction. Otherwise $\beta - \alpha$ is negative and $\alpha - \beta$ is positive. so $\alpha = (\alpha - \beta) + \beta$ is a sum of positive roots and α is not simple. Contradiction.

Lemma 24 *If β is a positive, non-simple root, and α is simple, then the α -string through β contains only positive roots.*

Proof: Suppose this string is $\beta - p\alpha, \dots, \beta, \dots, \beta + q\alpha$. If $\beta + n\alpha$ is in the string and $n \geq 0$, then $\beta + n\alpha$ is positive because β and α are positive.

Suppose n is negative. If $\beta + n\alpha$ is negative, follow the string upward until there is a negative m with $\beta + m\alpha$ negative, but $\beta + (m+1)\alpha$ positive. Then $\alpha = -(\beta + m\alpha) + (\beta + (m+1)\alpha)$ is a sum of two positive roots, which is impossible because α is simple. QED.

Theorem 60 *Given simple roots and their Dynkin diagram, it is possible to reconstruct all roots.*

Proof: It suffices to construct the positive roots. If α is positive, it equals $\alpha_1 + \dots + \alpha_k$ where the α_i are not necessarily distinct, but are all simple, and where each $\alpha_1 + \dots + \alpha_i$ is always a root. We therefore construct the roots by height. Suppose we know all positive roots of height less than or equal to k . A root of height $k+1$ must equal $\beta + \alpha$ where β is a known root of height k and α is a simple root. Form the chain $\beta + p\alpha, \dots, \beta, \beta + \alpha, \dots, \beta + q\alpha$. All of the roots left of β are positive of height less than m , so we can check which exist and compute p . But we can compute $2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}$ from the Dynkin diagram, and thus q . This allows us to determine if $\beta + \alpha$ is a root. QED.

Theorem 61 *If the Dynkin diagram of $\mathcal{G}$ can be broken into two disconnected pieces, then $\mathcal{G} = \mathcal{G}_1 \oplus \mathcal{G}_2$ where each $\mathcal{G}_i$ corresponds to one of the pieces.*

Proof: The simple roots in one piece are perpendicular to those in the other. Since all roots can be written as linear combinations of simple roots, the full roots for one diagram are perpendicular to those for the other diagram. From the full set of roots, we can form the e_α of one candidate $\mathcal{G}_1$ and the e_β for the other candidate $\mathcal{G}_2$. then $[e_\alpha, e_\beta] = 0$ since if α is in the full root set generated by one diagram and β is in the full root space generated by the other diagram, then $\alpha + \beta$ is not a root. So the root vectors separate into two sets. Similarly $\mathcal{T}$ separates into a subspace generated by simple roots in one diagram, a subspace generated by simple roots in the other diagram, and the orthogonal complement of these two spaces. Thus we obtain $\mathcal{G}_1$ and $\mathcal{G}_2$ and an abelian torus from the leftover piece. QED.

Remark: Note that the previous theorem applies to the Lie algebra level, but not necessarily to the Lie group level. It does work at the Lie group level if we assume our Lie groups are simply connected.

9.11 More

At the beginning of these notes, we proved that a compact Lie group has a left and right invariant metric. An immediate consequence is that the universal cover of a compact Lie group is isomorphic to $R^k \times G_1 \times \dots \times G_k$ where each G_i is compact with a left and right invariant metric and cannot be further decomposed.

In the last section, we discovered that once several choices have been made, each G_i produces a connected Dynkin diagram. We are about to prove that the G_i are completely

classified by their Dynkin diagrams.

The first barrier to this result is that several choices had to be made to obtain the Dynkin diagram: the choice of left and right invariant metric, the choice of maximal torus, and the choice of a hyperplane in the root space selecting positive roots. We must show that the resulting Dynkin diagram is independent of these choices. Otherwise, one group might lead to several Dynkin diagrams.

Let's first discuss the left and right invariant metric.

Theorem 62 *Suppose a compact Lie group G acts irreducibly on a real vector space V . Then any two invariant inner products on V are equal up to a positive real multiple.*

Proof: If the inner products are $\langle \rangle$ and $\langle \langle \rangle \rangle$, we claim there is an operator A such that $\langle \langle x, y \rangle \rangle = \langle Ax, y \rangle$. Fix x and consider the map $V \rightarrow V$ by $y \rightarrow \langle \langle x, y \rangle \rangle$. This is an element of V^* but $\langle \rangle$ induces a canonical isomorphism from V to V^* given by $y \rightarrow \langle Ax, y \rangle$ for some unique Ax . As x varies, it is easy to show that $x \rightarrow Ax$ is linear.

For each g we have $\langle \langle T_g x, T_g y \rangle \rangle = \langle \langle x, y \rangle \rangle$ so

$$\langle AT_g x, T_g y \rangle = \langle Ax, y \rangle = \langle T_g Ax, T_g y \rangle$$

It follows that $AT_g = T_g A$, so A is an intertwining operator. Moreover,

$$\langle x, A^T y \rangle = \langle Ax, y \rangle = \langle \langle x, y \rangle \rangle = \langle \langle y, x \rangle \rangle = \langle Ay, x \rangle = \langle x, Ay \rangle$$

so $A = A^T$. This A has a real eigenvalue, λ . Then $A - \lambda I$ is an intertwining operator which is not an isomorphism, and thus is identically zero. So $A = \lambda I$. Clear, then, $\lambda > 0$ since both $\langle \rangle$ and $\langle \langle \rangle \rangle$ are positive definite. QED

Theorem 63 *If G_i is a compact group and $\mathcal{G}$ cannot be decomposed, then the representation Ad of G on $\mathcal{G}$ is irreducible, and thus a left and right invariant metric on G is unique up to a constant multiple.*

Proof: Suppose $\mathcal{G} = \mathcal{G}_1 \oplus \mathcal{G}_2$ is a decomposition of $\mathcal{G}$ into invariant subspaces. Then if $X \in \mathcal{G}$ and $Y \in \mathcal{G}_i$, we have $[X, Y] = ad(X)(Y)$, and by invariance this belongs to $\mathcal{G}_i$ for both i . Hence our decomposition is a sum of ideals, which is not possible because we are assuming that $\mathcal{G}_i$ cannot be decomposed.

Remark: Thus the Dynkin diagram assigned to such a G does not depend on the invariant metric.

We proved that maximal tori are equivalent under conjugation. Conjugation preserves roots. Hence the choice of maximal torus need not concern us. Thus the only choice that matters is the hyperplane that determines which roots are positive. It is possible to prove that this choice is invariant up to conjugation, but we can avoid that as follows.

Notice that the rank of G , that is, the dimension of the maximal torus, and the dimension of G are both invariants. On the other hand, these invariants can be read from the Dynkin diagram. The rank is the number of nodes on the diagram. We proved that the roots can be reconstructed from the diagram, and the dimension of G equals its rank plus the number of roots. Consequently, if a G_i could lead to several Dynkin diagrams, both diagrams would have the same number of nodes and the same number of roots.

In the letters $A_n, B_n, C_n, D_n, G_2, F_4, E_6, E_7, E_8$ assigned to these diagrams, the number n is the rank. The table below shows the rank and the dimension; we will compute the dimension for the classical cases later.

A_n	$su(n+1)$	$(n+1)^2 - 1 = n(n+2)$
B_n	$so(2n+1)$	$\frac{(2n)(2n+1)}{2} = n(2n+1)$
C_n	$sp(n)$	$n(2n+1)$
D_n	$so(2n)$	$\frac{(2n-1)(2n)}{2} = n(2n-1)$
G_2		14
F_4		52
E_6		78
E_7		133
E_8		248

Recall that A_n, B_n, C_n, D_n are defined for $n \geq 1, n \geq 2, n \geq 3, n \geq 4$. A short calculation shows that the only cases when a group might produce two different diagrams are $B_n = C_n$ and $B_6 = C_6 = E_6$.

All roots of E_6 have the same length. In the cases B_n and C_n , root vectors have only two possible lengths, called *short* roots and *long* roots. The number of short roots of B_n is $2n$ and the number of short roots of C_n is $2n(n-1)$. These numbers can be deduced from the Dynkin diagram. Recall that any two maximal tori are conjugate, and this conjugacy maps roots of one torus to roots of the other. Thus the distinction between short and long roots depends only on the invariant metric, which we proved is unique up to a multiplicative constant.

Therefore, isomorphic Lie algebras yield the same Dynkin diagram.

Chapter 10

Classification of Compact Lie Algebras

10.1 Intermission One

The two intermission sections which follow contain no results needed for our theory; they are here to give context to the results in this chapter.

A Lie algebra L over a field K is *simple* if it has no ideals except $\{0\}$ and L and it is not the trivial one-dimensional algebra with zero bracket.

One of the key theorems of Sophis Lie states that if G is a Lie group with Lie algebra $\mathcal{G}$, then there is a one-to-one correspondence between connected Lie subgroups of G and Lie subalgebras of $\mathcal{G}$. In particular, normal connected Lie subgroups correspond to ideals of the subalgebra. So simple Lie algebras correspond to Lie groups with no non-trivial connected normal subgroups.

This theorem of Sophis Lie requires that we define the notion of a Lie subgroup with great care: the definition does not require that the subgroup have the induced topology. For instance, if G is the torus $R^2/Z \times Z$, the Lie algebra is R^2 and straight lines through the origin with angle an irrational multiple of π correspond to subgroups of the torus which wind around it infinitely often. These subgroups are dense in the torus, but are given the topology of a line.

Because of the importance of simple groups, a lot of work has been done classifying simple Lie algebras; the classification is known over both the complex and real numbers. Usually the classification over C is done first. The classification over R then proceeds by taking a real simple Lie algebra L and forming $L \otimes C$; it turns out that this complex Lie algebra is either simple or else a direct sum of two simple algebras. Continuing, it turns out that

every simple complex algebra is $L \otimes C$ for at least two real algebras, one compact and one not compact. Sometimes there are many real L inducing the same complex algebra. If L is simple and compact over R , then $L \otimes C$ is simple, and this sets up a one-to-one correspondence between real compact simple Lie algebras and complex simple Lie algebras.

It is useful to know something about the complex classification. It ends with exactly the same root space decomposition, Dynkin diagram, etc., that we found starting with a compact L and forming $L \otimes C$. But it begins by replacing two key ideas of our development with algebraic alternatives. We started with a maximal torus in G inducing an abelian subalgebra $\mathcal{T} \subset L$ in our real algebra and $\mathcal{T} \otimes C \subset L \otimes C$. This $\mathcal{T} \otimes C$ is a maximal abelian subalgebra of the complex algebra, but in the complex case not all maximal abelian subalgebras are isomorphic. So instead, the complex classification defines a *Cartan subalgebra* to be a nilpotent subalgebra which is its own normalizer. Such a subalgebra turns out to be a maximal nilpotent subalgebra, and a slightly tricky argument proves that a Cartan subalgebra always exists. Later, it is proved that such subalgebras are abelian. In this way, the particular maximal abelian subalgebras are found which replace the torus algebras.

The second idea that needs replacement is the invariant metric $\langle X, Y \rangle$. In the complex case, this is replaced with the Killing form $K(X, Y) = \text{tr}(adXadY)$. It is possible that $K(X, X) = 0$ for a nonzero X , but a crucial theorem, requiring several preliminary results, shows that this form is non-degenerate on a complex simple L , and that is enough for many of our arguments.

10.2 Intermission Two

We are classifying compact real algebras with connected Dynkin diagrams, and our beginning results show that such algebras are simple. A consequence is that they have many automorphisms. Indeed in section 6.3 we proved that the automorphism group of a Lie algebra L with negative definite Killing form is a Lie group with Lie algebra L , and the identity component $G = \text{Aut}_0(L)$ of this group is a compact Lie group with Lie algebra L . Each inner automorphism $g \rightarrow g_1 g g_1^{-1}$ in G induces $Ad(g) : L \rightarrow L$.

The existence of this large set of automorphisms makes it difficult to break $\mathcal{G} = L$ apart and find discrete structures inside which serve to classify L . Glancing back at our work so far shows how we avoided this difficulty. We first discussed maximal tori $T \subset G$ and the corresponding subalgebras $\mathcal{T} \subset \mathcal{G} = L$. These objects are not unique because automorphisms of L transform one to another, but we proved that any two maximal tori are equivalent under an automorphism. This allowed us to fix $T \subset G$ and $\mathcal{T} \subset \mathcal{G} = L$.

From this point on, only automorphisms which map T to T interest us. Recall that the normalizer N of T is the set of $g \in G$ such that $gTg^{-1} \subset T$. Clearly $g \in T$ has this

property and we defined the Weyl group to be the quotient N/T and proved that it is a finite group. We will study it in detail later on, but for now we take this as evidence that most automorphisms preserving T are inner automorphisms by some $t \in T$. We must pay attention to these automorphisms as we complete the classification.

Since T is abelian, these automorphisms act trivially on T and we need only consider their action on the remaining portion of G . This immediately led us to the root decomposition of our Lie algebra, so

$$L \otimes C = (\mathcal{T} \otimes C) \oplus \sum L_\alpha$$

These root spaces are one-dimensional and $[L_\alpha, L_\beta] \subset L_{\alpha+\beta}$.

While the root spaces are determined by these considerations, the basis vectors for these spaces are not determined. Suppose for a moment that we randomly choose non-zero elements $e_\alpha \in L_\alpha$ for each α . Suppose we also choose an ordering of the roots and thus a basis of $L \otimes C$ formed by the simple roots $\alpha_1, \dots, \alpha_k$. Then the α_i and the various e_α form a basis of L and the Lie algebra is completely determined by structure constants

$$[e_\alpha, e_\beta] = N_{\alpha\beta} e_{\alpha+\beta}$$

These constants are zero if $\alpha + \beta$ is not a root.

We can make this a little more explicit. Note that $[e_\alpha, e_{-\alpha}]$ is a multiple of α . Let us agree to choose these basis elements so that for all roots, $[e_\alpha, e_{-\alpha}] = \alpha$. Then all brackets between elements of $\mathcal{T} \otimes C$ and the e_α are determined, as are brackets between e_α and $e_{-\alpha}$.

But this still leaves a large amount of ambiguity, caused again by our automorphisms. If we apply $Ad(t)$, the entire root space is fixed, but the e_α are mapped to $\sigma_\alpha e_\alpha$. Our condition on root pairs forces $\sigma_\alpha \sigma_{-\alpha} = 1$, but otherwise these new basis vectors are other reasonable choices. With the new choices, the structure constants change:

$$[\sigma_\alpha e_\alpha, \sigma_\beta e_\beta] = \sigma_\alpha \sigma_\beta N_{\alpha\beta} e_{\alpha+\beta} = \left(\frac{\sigma_\alpha \sigma_\beta}{\sigma_{\alpha+\beta}} N_{\alpha\beta} \right) (\sigma_{\alpha+\beta} e_{\alpha+\beta})$$

Of course the conditions on the roots put restrictions on the renormalizations σ induced by automorphisms. But it is not obvious at first how to choose appropriate normalizations, and it is not obvious that the resulting structure constants can be determined from the Dynkin diagrams.

Luckily, this is a beautiful theorem about all of this, to be stated and proved in the next section.

10.3 The Isomorphism Theorem

Theorem 64 *Let L_1 and L_2 be compact Lie algebras with the same connected Dynkin diagram. Divide the roots into positive and negative roots in L_1 and L_2 and then select simple roots $\alpha_1, \dots, \alpha_k$ in L_1 and corresponding simple roots $\hat{\alpha}_1, \dots, \hat{\alpha}_k$ in L_2 . Choose this correspondence so α_i corresponds to the same node in the Dynkin diagrams for both algebras. The A_n, D_n , and E_6 diagrams have symmetries which allow this correspondence to be made in more than one way; select one. Clearly there is a map $f : \mathcal{T}_1 \rightarrow \mathcal{T}_2$ which is an isometry up to a positive scalar, sends α_i in L_1 to $\hat{\alpha}_i$ in L_2 , and maps the roots of L_1 isomorphically to the roots of L_2 .*

For each simple root α of $\mathcal{T}_1$, choose a basis vector $x_\alpha \in E_\alpha \subset L_1 \otimes C$, and choose a basis vector $\tilde{x}_\alpha \in \tilde{E}_\alpha \subset L_2 \otimes C$ similarly. Then there is a unique Lie algebra isomorphism $\varphi : L_1 \otimes C \rightarrow L_2 \otimes C$ which equals f on $\mathcal{T}_1$ and maps x_{α_i} to $\tilde{x}_{\hat{\alpha}_i}$ for each simple α_i .

Remark: This theorem says that the complex algebra $L \otimes C$ is completely determined by the Dynkin diagram. We will later prove that the compact real algebra is also determined by the Dynkin diagram.

As stated, the theorem only applies to complex algebras coming from a compact real algebra. But the proof doesn't really use this fact and could be used unchanged in a classification of all complex simple Lie algebras.

At the end of the previous section, we worried about renormalizing the basis vectors of all E_α and thus changing all of the structure constants, but the theorem says that we have far less freedom, and normalizing only basis vectors for simple roots determines everything else.

I first saw the proof we will give in Terrance Tao's blog on Lie Algebras. Later I found the same proof in lecture notes by Kiyoshi Igusa, and will follow his proof. See http://people.brandeis.edu/~igusa/Math223aF11/Math223a_2011F.html

Proof: We first claim that the one-dimensional spaces E_{α_i} for the simple roots generate the algebra L_+ formed by taking the sum of the E_α over all positive roots. Indeed every positive root can be written uniquely as $\alpha = \sum k_i \alpha_i$ where this sum is over a subset of simple roots and the k_i are positive integers. We prove the result by induction on $\sum k_i$. If this sum is one, we are dealing with the root space of a simple root and the result is obvious. Otherwise suppose we are given E_α as above, and the result is true for any positive root with a smaller sum than $\sum k_i$. But we proved earlier that $\alpha = \beta + \alpha_i$ where β is a smaller positive root and α_i is simple. So $E_\alpha = [E_\beta, E_{\alpha_i}]$. But $E_\beta \in L_+$ by induction and so $E_\alpha \subset L_+$.

A similar argument shows that the $E_{-\alpha_i}$ generate the subalgebra L_- formed by the sum of the E_α over negative roots.

Putting the two results together, it follows that L is generated by the E_{α_i} and $E_{-\alpha_i}$ taken over simple roots. Indeed, this sum contains E_{α_i} and $E_{-\alpha_i}$ and thus contains their bracket, which consists of all complex multiples of α .

In the theorem, we select generators $x_{\alpha_i} \in E_{\alpha_i}$ for each simple root of L_1 . For any choice of generator $y_{\alpha_i} \in E_{-\alpha_i}$ we know from results in section 8.8 that $[x_{\alpha_i}, y_{\alpha_i}]$ is a complex multiple of α . Since x_{α_i} is fixed, there is a unique choice of y_{α_i} such that this bracket is exactly α . Select these $y_{\alpha_i} \in L_1$ and select in a similar way $\hat{y}_{\hat{\alpha}_i} \in L_2$. Then the conjectured isomorphism from L_1 to L_2 would have to map y_{α_i} to $\hat{y}_{\hat{\alpha}_i}$ because $[x_{\alpha_i}, y_{\alpha_i}] = \alpha_i$ in L_1 and there is a similar equation in L_2 and the map takes x to $\hat{x}$ and α to $\hat{\alpha}$ and thus must take y to $\hat{y}$.

Since L_1 is generated by the x_{α_i} and y_{α_i} , it follows that the isomorphism φ is unique if it exists at all.

Proof, continued: Consider the subalgebra $D \subset L_1 \oplus L_2$ generated by all $\bar{x}_{\alpha_i} = (x_{\alpha_i}, \hat{x}_{\hat{\alpha}_i})$ and all $\bar{y}_{\alpha_i} = (y_{\alpha_i}, \hat{y}_{\hat{\alpha}_i})$. We are going to prove that this subalgebra is the graph of an isomorphism φ from L_1 to L_2 , completing the proof. Notice that there are obvious projection maps from D to L_1 and to L_2 . Clearly the previous steps of the proof show that these maps are onto. It will suffice if each is one-to-one. This will be true provided $D \cap L_1 = \{0\}$ and $D \cap L_2 = \{0\}$, for then the kernels of the two projection maps are trivial.

Proof, next step: A positive root β for L_1 is maximal if $\beta + \alpha_i$ is not a root for any simple α_i . Clearly such a root exists (actually it is unique, but we do not need this fact). Let $\hat{\beta}$ be the corresponding root for L_2 . Select generators $z \in E_{\beta}$ and $\hat{z} \in E_{\hat{\beta}}$. We will use these choices to construct another helper subset of $L_1 \oplus L_2$, but not to construct our isomorphism.

Let M be the subspace of $L_1 \oplus L_2$ containing $\bar{z} = (z, \hat{z})$ and all linear combinations of expressions of the form

$$[\bar{y}_{\alpha_{i_1}} [\bar{y}_{\alpha_{i_2}} \cdots [\bar{y}_{\alpha_{i_k}}, \bar{z}] \cdots]]$$

Notice that bracketing with $\bar{y}_{\alpha_i}$ decreases the root β , so the intersection of $E_{\beta} \oplus E_{\hat{\beta}}$ with M is one dimensional, generated by $\bar{z}$.

Proof, continued: We now claim that $[D, M] \subset M$. By construction, M is preserved by bracketing with the $\bar{y}_{\alpha_i}$. We must prove it is also preserved by bracketing with the $\bar{x}_{\alpha_i}$. This is, in a sense, the heart of the entire proof. We prove this by induction on the number of $\bar{y}_{\alpha_i}$ terms in the bracket. If there are no such terms we are done because β is a maximal root, so $[\bar{x}_{\alpha_i}, \bar{z}] = 0$.

To prove the induction step, consider

$$[\bar{x}_{\alpha_j} [\bar{y}_{\alpha_{i_1}} [\bar{y}_{\alpha_{i_2}} \cdots [\bar{y}_{\alpha_{i_k}}, \bar{z}] \cdots]]]$$

If $j \neq i_1$, then $[\bar{x}_{\alpha_j}, \bar{y}_{\alpha_{i_1}}] = 0$ because the difference of two different simple roots is not a root, so we can commute the first two terms and obtain

$$[\bar{y}_{\alpha_{i_1}} [\bar{x}_{\alpha_j} [\bar{y}_{\alpha_{i_2}} \cdots [\bar{y}_{\alpha_{i_k}}, \bar{z}] \cdots]]]$$

The induction hypothesis then shows that this expression is in M .

But if $j = i_1$, then $[\bar{x}_{\alpha_{i_1}}, \bar{y}_{\alpha_{i_1}}] = (\alpha_{i_1}, \hat{\alpha}_{i_1})$ and so our expression equals

$$[\bar{y}_{\alpha_{i_1}} [\bar{x}_{\alpha_{i_1}} [\bar{y}_{\alpha_{i_2}} \cdots [\bar{y}_{\alpha_{i_k}}, \bar{z}] \cdots]]] + [(\alpha_{i_1}, \hat{\alpha}_{i_1}) [\bar{y}_{\alpha_{i_2}} \cdots [\bar{y}_{\alpha_{i_k}}, \bar{z}] \cdots]]]$$

The first of these terms is in M by induction and the second term is in M because $(\alpha_{i_1}, \hat{\alpha}_{i_1})$ acts on any term in M by multiplying both factors by the same constant.

Proof, continued: The two previous steps are used only once in the proof, to prove that $D \neq L_1 \oplus L_2$. Indeed, if $D = L_1 \oplus L_2$, then $[D, M] \subset M$ would prove that M is an ideal in $L_1 \oplus L_2$. Since L_1 and L_2 are simple, the only ideals are $\{0\}$, L_1 , L_2 , and $L_1 \oplus L_2$. But $(z, \hat{z}) \in M$ has non-zero components in L_1 and L_2 , so the only possibility is that M is everything. However $E_\beta \oplus E_{\hat{\beta}}$ has dimension two and only contains one element of M up to scalar multiples.

Proof, concluded: Finally we prove that $D \cap L_1 = \{0\}$ and $D \cap L_2 = \{0\}$. The proofs are the same, so we only study the first case. Suppose $(w, 0)$ is in D for some non-zero element of L_1 . Since $[\bar{x}_{\alpha_i}, (w, 0)] = ([x_{\alpha_i}, w], 0)$ and similarly for $\bar{y}_{\alpha_i}$ and these elements generate all of L_1 , $[D, (w, 0)] = L_1$. It would follow that D contains L_1 and thus also contains L_2 and thus equals $L_1 \oplus L_2$, contradicting the previous step.

10.4 The Compact Case

The goal of this section is to prove an analogue of the isomorphism theorem for compact Lie Algebras:

Theorem 65 *Suppose L_1 and L_2 are compact Lie algebras with the same Dynkin diagram. Then L_1 and L_2 are isomorphic.*

Remark: Suppose L_1 is a Lie algebra with negative definite Killing form. Form the complex algebra $L = L_1 \otimes \mathbb{C}$. Let $\sigma : L \rightarrow L$ be the map $\sigma(X + iY) = X - iY$. It is easy to show that

- $\sigma : L \rightarrow L$ is real linear.
- $\sigma(\lambda X) = \bar{\lambda}\sigma(X)$.
- $[\sigma(X), \sigma(Y)] = \sigma([X, Y])$.
- $\sigma^2 = id$
- If K is the Killing form on L , then $-K(X, \sigma(Y))$ is a Hermitian inner product on L

Remark on the third condition: This condition is equivalent to the statement that the structure constants are real if we start with a real basis of the complexified algebra. When classifying semisimple complex algebras, there are arguments that show directly that we can obtain real structure constants (or even integer constants) if we carefully choose a basis, but this result is obvious if we start with a real algebra and complexify.

Remark on the last condition: We proved long ago that there is an Ad -invariant inner product on the Lie algebra of a compact Lie group. Using that result, we proved that this Lie algebra has the form $R^k \oplus L$ where L has negative definite Killing form, and that L is the sum of simple ideals. Later we proved that on each simple ideal, the Ad -invariant inner product is unique up to a scalar constant.

The advantage of using the Ad -invariant inner product is that it also exists on the abelian component of the Lie algebra. But now that we are dealing with Dynkin diagrams and the semisimple portion of the algebra, we might as well replace this inner product with the negative of the Killing form. One advantage is that it removes the ambiguity of scale, since now the inner product is unique.

At one point earlier in these notes, we extended the invariant inner product to a Hermitian inner product on the complexified Lie algebra, and that is exactly what the last item above does when we use the negative of the Killing form as our inner product.

Remark: Conversely, if these conditions hold, it is easy to show that $\{X \in L | \sigma(X) = X\}$ is a Lie algebra L_1 with negative definite Killing form such that $L_1 \otimes C$ is isomorphic to L . We leave the easy details to the reader.

If we start with a real compact algebra, then describing the algebra as a subset of $L \otimes C$ is easy. But there are those pesky automorphisms φ of the complex algebra, and once we apply φ to a real subalgebra, it is suddenly hard to identify. Luckily, the effect on σ is easy, so when dealing with several real compact subalgebras of a complex algebra, it is much better to deal with their σ 's than the algebras themselves.

Lemma 25 *Suppose L is a real Lie algebra with negative definite Killing form, and $\sigma : L \otimes C \rightarrow L \otimes C$ is its associated conjugation map. Let φ be an automorphism of $L \otimes C$. Then the conjugation map for $\varphi(L)$ is $\varphi \circ \sigma \circ \varphi^{-1}$.*

Proof: If $X \in \varphi(L)$, then $\varphi^{-1}(X)$ is in L and thus fixed by σ , so

$$\varphi \circ \sigma \circ \varphi^{-1}(X) = \varphi \circ \varphi^{-1}(X) = X$$

The remaining details are left to the reader.

Remark: Now suppose that L_1 and L_2 are compact Lie algebras with the same connected Dynkin diagram. Form $L_1 \otimes C$ and $L_2 \otimes C$. By the theorem of the previous section, these complex Lie algebras are isomorphic. So we can apply this isomorphism and assume that L_1 and L_2 belong to the same complex Lie algebra, and are associated with conjugations σ_1 and σ_2 . We are going to prove that there is an isomorphism φ of the complex algebra such that $\varphi \circ \sigma_2 \circ \varphi^{-1} = \sigma_1$. It will follow that the two real Lie algebras are isomorphic.

There are two steps in the remaining argument. The first step shows that by replacing σ_2 by $\varphi \circ \sigma_2 \circ \varphi^{-1}$, we can make σ_1 and σ_2 commute. The second step shows that in this case $\sigma_1 = \sigma_2$.

Lemma 26 *If σ_1 and σ_2 are conjugations in a semisimple complex Lie algebra, there is an automorphism φ of the complex algebra such that $\varphi \circ \sigma_1 \circ \varphi^{-1}$ and σ_2 commute.*

Proof: Consider the Hermitian inner product $\langle X, Y \rangle = -K(X, \sigma_1(Y))$ on the complex algebra. Let $A = \sigma_2 \sigma_1$ and notice that A is an automorphism of the complex algebra and thus preserves the Killing form. So

$$\begin{aligned} \langle AX, Y \rangle &= -K(\sigma_2 \sigma_1 X, \sigma_1 Y) = -K(X, \sigma_1(\sigma_2 \sigma_1 Y)) \\ &= -K(X, \sigma_1 AY) = \langle X, AY \rangle \end{aligned}$$

Therefore A is a Hermitian operator and there is a basis $X_1, \dots, X_n$ of the complex algebra with $AX_i = \lambda_i X_i$ for real λ_i .

Let $B = A^2$. Since all eigenvalues of B are positive, we can define B^t for any real t by $B^t X_i = (\lambda_i^2)^t X_i$. We now claim that B^t is an automorphism of the complex algebra for all t . Indeed our complex algebra has complex structure constants defining the bracket: $[X_i, X_j] = \sum c_{ijk} X_k$. Since A is an automorphism, $[AX_i, AX_j] = \sum c_{ijk} AX_k$, and thus $\lambda_i \lambda_j \sum c_{ijk} X_k = \sum \lambda_k c_{ijk} X_k$. So $\lambda_i \lambda_j = \lambda_k$ unless $c_{ijk} = 0$. It follows that $(\lambda_i^2)^t (\lambda_j^2)^t = (\lambda_k^2)^t$ unless $c_{ijk} = 0$ and so running the argument backwards gives $[B^t X_i, B^t X_j] = \sum c_{ijk} B^t X_k$.

Therefore $\hat{\sigma}_1 = B^t \circ \sigma_1 \circ (B^t)^{-1}$ is a conjugation for any t . We will show that this conjugation commutes with σ_2 when $t = \frac{1}{4}$.

Notice that

$$\sigma_1 A \sigma_1 = \sigma_1 (\sigma_2 \sigma_1) \sigma_1 = \sigma_1 \sigma_2 (\sigma_1)^2 = \sigma_1 \sigma_2 = A^{-1}$$

and so $\sigma_1 B \sigma_1 = (\sigma_1 A \sigma_1)(\sigma_1 A \sigma_1) = A^{-2} = B^{-1}$. Therefore

$$\sigma_1 B = B^{-1} \sigma_1$$

From this, we can conclude that

$$\sigma_1 B^t = B^{-t} \sigma_1$$

Indeed $\sigma_1 B X_i = B^{-1} \sigma_1 X$ for all X . If X is in the λ_i^2 eigenspace of B we have $\lambda_i^2 \sigma_1 X = B^{-1} \sigma_1 X$ or

$$B \sigma_1 X = (\lambda_i^2)^{-1} \sigma_1 X$$

So σ_1 maps the λ_i^2 eigenspace of B to the $(\lambda_i^2)^{-1}$ eigenspace isomorphically. These are also the $(\lambda_i^2)^t$ and $(\lambda_i^2)^{-t}$ eigenspaces of B^t and the required equation follows.

Now we are ready to compare $\hat{\sigma}_1 \sigma_2$ and $\sigma_2 \hat{\sigma}_1$. We first do this for any fixed t , and then for the special case $t = \frac{1}{4}$. Recall that conjugations are their own inverses.

We have

$$\begin{aligned} \hat{\sigma}_1 \sigma_2 &= B^t \sigma_1 B^{-t} \sigma_2 = [\sigma_2 (B^t \sigma_1 B^{-t})]^{-1} = [(\sigma_2 \sigma_1) (\sigma_1 B^t \sigma_1 B^{-t})]^{-1} \\ &= (\text{because } \sigma_1 B^t = B^{-t} \sigma_1) [A (B^{-t} \sigma_1 \sigma_1 B^{-t})]^{-1} = B^{2t} A^{-1} \end{aligned}$$

and we have

$$\begin{aligned} \sigma_2 \hat{\sigma}_1 &= \sigma_2 (B^t \sigma_1 B^{-t}) = \sigma_2 \sigma_1 (\sigma_1 B^t \sigma_1 B^{-t}) \\ &= (\text{because } \sigma_1 B^t = B^{-t} \sigma_1) A (B^{-t} \sigma_1 \sigma_1 B^{-t}) = AB^{-2t} \end{aligned}$$

In particular, if $t = \frac{1}{4}$ we have

$$\begin{aligned} \hat{\sigma}_1 \sigma_2 &= B^{1/2} A^{-1} \\ \sigma_2 \hat{\sigma}_1 &= AB^{-1/2} \end{aligned}$$

On the eigenvector X_i , $\hat{\sigma}_1 \sigma_2$ equals $(\lambda_i^2)^{1/2} \lambda_i^{-1} X_i = \text{sgn}(\lambda_i) X_i$. Similarly $\sigma_2 \hat{\sigma}_1$ equals $\lambda_i (\lambda_i^2)^{-1/2} X_i = \text{sgn}(\lambda_i) X_i$. So $\hat{\sigma}_1 \sigma_2 = \sigma_2 \hat{\sigma}_1$. QED.

Lemma 27 Suppose σ_1 and σ_2 are conjugations in a semisimple complex Lie algebra which commute. Then $\sigma_1 = \sigma_2$.

Proof: As before, consider the Hermitian inner product $\langle X, Y \rangle = -K(X, \sigma_1(Y))$ and let $A = \sigma_2 \sigma_1$; then A is self-adjoint and thus can be diagonalized so $A(X_i) = \lambda_i X_i$. But

$$A^{-1} = \sigma_1^{-1} \sigma_2^{-1} = \sigma_1 \sigma_2 = \sigma_2 \sigma_1 = A$$

So the λ_i are ± 1 .

On the other hand,

$$\langle AX, Y \rangle = \langle X, AY \rangle = -K(X, \sigma_2 \sigma_1 \sigma_1 Y) = -K(X, \sigma_2 Y)$$

Since $-K(X, \sigma_2 Y)$ is also a Hermitian inner product, -1 cannot be an eigenvector of A because $Ae = -e$ would imply that $\langle -e, e \rangle$ is negative and yet $\langle -e, e \rangle = \langle Ae, e \rangle = -K(e, \sigma_2 e) > 0$. So A is the identity and $\sigma_2 \sigma_1 = I$, or $\sigma_2 = \sigma_1^{-1} = \sigma_1$. QED.

Chapter 11

$SU(n)$

Recall that $U(n)$, the *unitary group*, is the group of all complex linear transformations $A : C^n \rightarrow C^n$ which preserve the standard Hermitian inner product $\langle v, w \rangle = \sum v_i \overline{w_i}$. At the matrix level it is the group of $n \times n$ matrices which satisfy $\overline{A}^T A = I$. This equation implies that $\overline{\det(A)} \det(A) = 1$ and thus that $|\det(A)| = 1$.

The determinant is a group homomorphism $\det : U(n) \rightarrow S^1$ whose kernel is the group $SU(n)$, the *special unitary group* of all matrices satisfying $\overline{A}^T A = I$ and $\det(A) = 1$. The Lie algebra of this group, $su(n)$, is thus the set of all $n \times n$ matrices such that $-\overline{A}^T = A$ and $\text{tr}(A) = 0$. On the Lie algebra level, $u(n) = R \oplus su(n)$.

Both $U(n)$ and $SU(n)$ are connected; $SU(n)$ is simply connected.

11.1 A_n is $su(n+1)$

For convenience, we work with $SU(n+1)$ and $su(n+1)$ because these objects will have Dynkin diagrams of type A_n . We claim that the set of diagonal matrices form a maximal torus. Notice that these diagonal entries have absolute value 1 and thus belong to S^1 . As a group, the diagonal matrices are isomorphic to $S^1 \times \dots \times S^1$ for n copies of the circle, because the first n entries are arbitrary and then the last one is fixed by the condition that the matrix have determinant one. We will prove that this is a maximal torus in just a moment. The corresponding Lie algebra consists of matrices of the form

$$t = \begin{pmatrix} it_1 & 0 & \dots & 0 \\ 0 & it_2 & \dots & 0 \\ \dots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & it_{n+1} \end{pmatrix}$$

such that $\sum t_i = 0$.

The remaining elements of the Lie algebra are linear combinations of the following matrices, with zeros on the diagonal and nonzero entries only in the ij and ji spots.

$$u_{ij} = \begin{pmatrix} 0 & i_{ij} \\ i_{ji} & 0 \end{pmatrix} \quad v_{ij} = \begin{pmatrix} 0 & 1_{ij} \\ -1_{ji} & 0 \end{pmatrix}$$

Notice that $[t, i_{ij}] = i(t_i - t_j)i_{ij} = -(t_i - t_j)1_{ij}$ and $[t, i_{ji}] = i(t_j - t_i)i_{ji} = (t_i - t_j)1_{ji}$, so $[t, u_{ij}] = -v_{ij}$ and $[t, v_{ij}] = (t_i - t_j)u_{ij}$. Compare these with the “formulas”

$$[t, u] = -2\pi f(t)v$$

$$[t, v] = 2\pi f(t)u$$

$$[u, v] = -2\pi < v, v > it_f$$

To prove that the diagonal matrices form a maximal torus, it suffices to show that its Lie algebra is a maximal abelian subalgebra. If not, we can extend the algebra by adding a commuting element of the form

$$\sum_{i < j} A_{ij} u_{ij} + \sum_{i < j} B_{ij} v_{ij}$$

Note that when we bracket this element with an arbitrary diagonal t , we obtain

$$-\sum_{i < j} (t_i - t_j) A_{ij} v_{ij} + \sum_{i < j} (t_i - t_j) B_{ij} u_{ij}$$

Fix $i < j$ and let $t_i = 1, t_j = -1$ and all other $t_k = 0$. Then the displayed sum may have many terms, but the terms involving u_{ij} and v_{ij} will be $-2A_{ij}v_{ij} + 2B_{ij}u_{ij}$. So $A_{ij} = B_{ij} = 0$. This must hold for all $i < j$, so the diagonal elements form a *maximal* abelian subalgebra.

We conclude that $f(t) = \frac{1}{\pi}(t_i - t_j)$ is a root for $1 \leq i < j \leq (n+1)$. This comes from $\mathcal{G}$ itself, rather than $\mathcal{G} \otimes C$, so the full set of roots are these and their negatives. Let these be the positive roots. Ignore the factor $\frac{1}{\pi}$. Notice that $t_i - t_j = (t_i - t_{i+1}) + (t_{i+1} - t_{i+2}) + \dots + (t_{j-1} - t_j)$. Consequently this root is only simple if $j = i + 1$, so the simple roots are $(t_1 - t_2), (t_2 - t_3), \dots, (t_n - t_{n+1})$. Call these $\alpha_1, \alpha_2, \dots, \alpha_n$. Notice that the $\frac{1}{\pi}$ factors cancel in expressions of the form

$$2 \frac{\langle \alpha_i, \alpha_{i+1} \rangle}{\langle \alpha_i, \alpha_i \rangle} = -(p + q)$$

Notice that $\alpha_i - \alpha_{i+1}$ is not a root, but $\alpha_i + \alpha_{i+1}$ is a root, and $\alpha_i + 2\alpha_{i+1}$ is not a root. So in this expression, $p = 0$ and $q = 1$. Interchanging i and j , we discover that $p = 1$ and $q = 0$. Putting the two results together shows that

$$\left\langle \frac{\alpha_i}{\|\alpha_i\|}, \frac{\alpha_{i+1}}{\|\alpha_{i+1}\|} \right\rangle^2 = \frac{1}{4}$$

so the angle between these roots is 120 degrees. If i and j differ by more than one, similar arguments show that the angle is 90 degrees. Consequently the Dynkin diagram is A_n .

Remark: Notice that we determined the diagram without a specific knowledge of the invariant metric. It is usually not necessary to compute the metric specifically. Let's do it for $SU(n+1)$ for fun.

We already proved that an invariant metric is unique up to a positive constant. Notice that $X, Y \rightarrow -\text{tr}(XY)$ is invariant for $\mathfrak{su}(n+1)$. Indeed

$$-\text{tr}([Z, X]Y) = -\text{tr}(ZXY - XZY) = -\text{tr}(XYZ - XYZ) = -(-\text{tr}(X, [Z, Y]))$$

so this trace is invariant. It is real because

$$\begin{aligned} \overline{-\text{tr}(XY)} &= -\text{tr}(\overline{XY}) = -\text{tr}((-X^T)(-Y^T)) = -\text{tr}(X^T Y^T) \\ &= -\text{tr}((YX)^T) = -\text{tr}(YX) = -\text{tr}(XY) \end{aligned}$$

Finally, we claim $-\text{tr}(XY)$ is positive-definite on $\mathfrak{su}(n+1)$. If $A \in \mathfrak{su}(n+1)$, then $\exp(tA)$ is a one-parameter subgroup, so this group is conjugate to a one-parameter subgroup of diagonal matrices, and the Lie algebra generator is conjugate to a one-parameter generator which is diagonal. But $\text{tr}(g^{-1}XYg) = \text{tr}(XYgg^{-1}) = \text{tr}(XY)$. A typical diagonal element has purely imaginary entries it_i on the diagonal, so $\text{tr}(XX) = -\sum t_i^2$ and we are done.

Remark: Let us restrict this inner product to the Lie algebra of the maximal torus. A basis of this Lie algebra contains elements e_i which are equal to i at the i th spot, to $-i$ at the $(n+1)$ st spot, and to zero elsewhere. Clearly $\text{tr}(e_i e_i) = -1 - 1 = -2$. Thus our invariant metric satisfies $\langle e_i, e_i \rangle = 2$. Next consider $\langle e_i, e_j \rangle$ where $i \neq j$. Clearly $\text{tr}(e_i e_j) = -1$ and so $\langle e_i, e_j \rangle = 1$.

There is a way to make this reasonable. The root lattice of $\mathfrak{su}(n+1)$ is hexagonal, built of equilateral triangles. The easy way to get an equilateral triangle is to take the hyperplane of all $(x_1, \dots, x_{n+1})$ with $\sum x_i = 0$. The standard orthonormal basis $e_1, \dots, e_{n+1}$ of $\mathbb{R}^n$ gives a basis $e_i - e_{n+1}$ of this hyperplane, for $1 \leq i \leq n$. Then $\langle e_i - e_{n+1}, e_j - e_{n+1} \rangle = 1$ and $\langle e_i - e_{n+1}, e_i - e_{n+1} \rangle = 2$.

Remark: Incidentally, every element of $SU(n)$ is conjugate to an element of the maximal torus, and thus to a diagonal element of complex numbers of absolute value one. Thus if A is any isometry of C^n , we can choose a new orthonormal basis such that A just rotates the basis vectors about themselves.

This fact can be proved directly. If $\langle X, Y \rangle$ is a Hermitian inner product on a complex vector space V and $A : V \rightarrow V$, then there is a new orthonormal basis such consisting of eigenvectors of A if and only if A commutes with $\overline{A}^T$. In our case $\overline{A}^T A = I$, so $\overline{A}^T$ is the inverse of A and thus certainly commutes with it.

Chapter 12

$SO(n)$

Recall that $O(n)$, the *orthogonal group*, is the group of all real linear transformations $A : R^n \rightarrow R^n$ which preserve the standard inner product $\langle v, w \rangle = \sum v_i w_i$. At the matrix level it is the group of $n \times n$ matrices which satisfy $A^T A = I$. This equation implies that $\det(A)^2 = 1$ and thus that $\det(A) = \pm 1$.

The determinant is a group homomorphism $\det : O(n) \rightarrow \pm 1$ whose kernel is the group $SO(n)$, the *special orthogonal group* of all matrices satisfying $A^T A = I$ and $\det(A) = 1$. The Lie algebra of this group, $so(n)$, is thus the set of all $n \times n$ matrices such that $A^T = -A$. This group is connected, and has fundamental group Z_2 for $n \geq 3$.

12.1 B_n is $so(2n+1)$; D_n is $so(2n)$

This time, we claim that one choice for maximal torus consists of 2×2 rotation matrices down the diagonal, as in

$$\begin{pmatrix} \begin{pmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{pmatrix} & & & \\ & \begin{pmatrix} \cos \theta_2 & -\sin \theta_2 \\ \sin \theta_2 & \cos \theta_2 \end{pmatrix} & & \\ & & \ddots & \\ & & & \begin{pmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{pmatrix} \end{pmatrix}$$

Notice that the set of such matrices is a group $S^1 \times \dots \times S^1$ with n entries, and thus a torus. This choice works for $so(2n)$, but must be adjusted by adding an extra 1 on the diagonal in the odd case $so(2n+1)$. For this reason, the even and odd cases are different, giving rise to two sequences B_n and D_n .

We will prove that this is maximal abelian in just a moment. A consequence will be that every element of $SO(2n+1)$ and $SO(2n)$ is conjugate to an element in the torus. So every rotation of R^3 has an axis. Every rotation in R^4 can be obtained by choosing two perpendicular planes which only intersect at the origin, and rotating independently in the two planes. Every rotation in R^5 can be obtained by choosing two such perpendicular planes and an axis perpendicular to both, rotating independently in the two planes, and leaving the axis fixed.

The Lie algebra of our torus consists of matrices of the form

$$\begin{pmatrix} \begin{pmatrix} 0 & -t_1 \\ t_1 & 0 \end{pmatrix} & & & \\ & \begin{pmatrix} 0 & -t_2 \\ t_2 & 0 \end{pmatrix} & & \\ & & \ddots & \\ & & & \begin{pmatrix} 0 & -t_n \\ t_n & 0 \end{pmatrix} \end{pmatrix}$$

In the case $so(2n+1)$ we must add an extra 0 in the last diagonal entry.

The region above these diagonal entries in the Lie algebra can be divided into 2×2 submatrices J_{ij} for $i < j$. In the odd case, there is an additional column at the end containing one dimensional entries K_i . These 2×2 matrices J and one dimensional entries K can be expanded to become entries in $so(2n)$ or $so(2n+1)$ by adding appropriate elements below the diagonal. For instance, J_{ij} above the diagonal should also contain $-J_{ij}^T = -J_{ji}$ below the diagonal, and K_i in the last column should become $-K_i$ in the bottom row. We use the same letters to define these extended elements.

To understand brackets, it is useful to consider a special case. Consider the bracket below.

$$\left[\begin{pmatrix} \begin{pmatrix} 0 & -t_1 \\ t_1 & 0 \end{pmatrix} & \\ & \begin{pmatrix} 0 & -t_2 \\ t_2 & 0 \end{pmatrix} \end{pmatrix}, \begin{pmatrix} \begin{pmatrix} -a & -c \\ -b & -d \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{pmatrix} \right]$$

To compute this, we need only compute the bottom left result, and we find that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow \begin{pmatrix} -t_1b - t_2c & t_1a - t_2d \\ -t_1d + t_2a & t_1c + t_2b \end{pmatrix}$$

In particular

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow -(t_1 - t_2) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \rightarrow (t_1 - t_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and also

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \rightarrow -(t_1 + t_2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow (t_1 + t_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Compare these equations with the generic equations

$$[t, u] = -v \text{ and } [t, v] = u$$

We conclude that $t_1 - t_2$ and $t_1 + t_2$ are roots, up to a factor of $\frac{1}{\pi}$ which we will ignore. Clearly this calculation for $so(4)$ generalizes to $so(2n)$ and shows that $t_i - t_j$ and $t_i + t_j$ are roots for $i < j$. We can choose an ordering making these the positive roots and their negatives the negative roots. No element which is a sum of elements in these root spaces can belong to the maximal torus, since bracketing with elements of the torus preserve the individual root spaces, and for each element inside a root space we can find t_i making the bracket non-zero. It follows immediately that our choice of torus is a maximal torus for $so(2n)$.

Since all off-diagonal skew-symmetric matrices are sums of the basis vectors selected above, the roots just displayed form all possible roots.

Note that each $t_i - t_j$ is a sum of $t_i - t_{i+1}$ terms, so only the $t_i - t_{i+1}$ could be simple roots. Next observe that every positive root is a positive integral sum of $t_i - t_{i+1}$ and $t_{n-1} + t_n$. Indeed every $t_i - t_j$ has this form and $t_j + t_n$ has the form because

$$(t_j - t_{j+1}) + (t_{j+1} - t_{j+2}) + \dots + (t_{n-2} - t_{n-1}) + (t_{n-1} + t_n) = t_j + t_n$$

and $t_i + t_j$ has the form because

$$(t_i - t_j) + (t_j - t_n) + (t_j + t_n) = t_i + t_j$$

Since we have exactly the right number of positive roots to form a basis and every positive root is a positive sum of these roots, it follows that they are all simple.

From now on we let $\alpha_i = t_i - t_{i+1}$ for $1 \leq i < n$ and $\alpha_n = t_{n-1} + t_n$.

Let us compute the Dynkin diagram. As before, we do this using the formula

$$2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} = -(p + q)$$

where $\beta + k\alpha$ is a root for exactly $p \leq k \leq q$ where $q \leq 0$ and $p \geq 0$. At this moment, we do not know $< >$, but it is clear that the factor $\frac{1}{\pi}$ cancels in the formula.

The difference of two distinct simple roots is never a root, so $p = 0$ in our calculations. Consider α_i and α_j for $i < j < n$. Then $\alpha_i = (t_i - t_{i+1})$ and $\alpha_j = (t_j - t_{j+1})$ and the sum of these elements is not a root unless $j = i + 1$. So usually $q = 0$ and the roots are orthogonal, and in particular not connected in the Dynkin diagram. If $j = i + 1$, then $\alpha_i + \alpha_{i+1} = t_i - t_{i+1}$ is a root, but $\alpha_i + 2\alpha_{i+1} = (t_i + t_{i+1} - t_{i+2})$ is not a root. So $q = 1$ and the above expression is -1 . It follows that

$$\frac{\langle t_{\alpha_i}, t_{\alpha_{i+1}} \rangle^2}{\|\alpha_i\|^2 \|\alpha_{i+1}\|^2} = \frac{1}{4}$$

and consequently the angle between these roots is 120 degrees. So $\alpha_1, \dots, \alpha_{n-1}$ are connected in the same way that $\alpha_1, \dots, \alpha_n$ are connected for $su(n+1)$.

It remains to study the connection of α_n with the remaining simple roots. Note that $\alpha_{n-1} + \alpha_n = (t_{n-1} - t_n) + (t_{n-1} + t_n) = 2t_{n-1}$ is not a root, so α_{n-1} and α_n are perpendicular and thus not connected. However,

$$\alpha_{n-2} + \alpha_n = (t_{n-2} - t_{n-1}) + (t_{n-1} + t_n) = t_{n-2} + t_n$$

is a root. On the other hand, $\alpha_{n-2} + 2\alpha_n$ is clearly not a root. So the angle between α_{n-2} and α_n is 120 degrees and these roots are connected by one line. Finally, if $i < (n-2)$, then $\alpha_i + \alpha_n = (t_i - t_j) + (t_{n-1} + t_n)$ is not a root and these nodes are not connected.

Turn back to our Dynkin diagrams. We have obtained the diagram for D_n .

Just as before, the expression $-tr(XY)$ is invariant under ad on $so(2n)$ and negative definite because on the maximal torus its value for $(t_1, \dots, t_n)$ is $-\sum t_i^2$, and for other elements we use the fact that every element is conjugate to an element on the maximal torus. Details are as in the discussion of $su(n)$. This time we get the completely natural inner product on the Lie algebra of the maximal torus.

Remark: We now turn to $so(2n+1)$. The maximal torus is unchanged, and our analysis of root spaces for $t_i - t_j$ and $t_i + t_j$ is unchanged. But this time the linear span of these root spaces does not give elements of $so(2n+1)$ with entries in the last row or last column. Thus we need to account for additional elements in the Lie algebra whose last row is $(a_1, \dots, a_{2n}, 0)$ and whose last column is $(-a_1, -a_2, \dots, -a_{2n}, 0)$.

To get an idea of the bracket of torus elements with these elements, consider the bracket

$$\left[\begin{pmatrix} 0 & -t_1 & 0 \\ t_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & -a_1 \\ 0 & 0 & -a_2 \\ a_1 & a_2 & 0 \end{pmatrix} \right] = \begin{pmatrix} 0 & 0 & t_1 a_2 \\ 0 & 0 & -t_1 a_1 \\ -t_1 a_2 & t_1 a_1 & 0 \end{pmatrix}$$

Let

$$u = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

Then

$$[t, u] = -tv \text{ and } [t, v] = tu$$

Generalizing, it is clear that bracketing an element whose last row is $(a_1, a_2, \dots, a_{2n-1}, a_{2n}, 0)$ with an element of the torus algebra gives $(t_1 a_2, -t_1 a_1, t_2 a_4, -t_2 a_3, \dots)$. So we get additional roots $t_1, t_2, \dots, t_n$.

Let

$$\alpha_1 = t_1 - t_2, \dots, \alpha_{n-1} = t_{n-1} - t_n, \alpha_n = t_n$$

We will show that every positive root is a positive integer combination of these. Since we know that we have n positive roots, these must be our simple roots.

Indeed positive sums of the $t_i - t_{i+1}$ give all $t_i - t_j$ as before. In particular, we get all $t_i - t_n$ and the sum of this with t_n is t_i and the sum with t_n again is $t_i + t_n$. But in our discussion of $so(2n)$ we showed that from the $t_i - t_j$ and the $t_i + t_n$ we can obtain all $t_i + t_j$.

The analysis of angles between the various $t_i - t_{i+1}$ proceeds exactly as before. Notice that $(t_i - t_{i+1}) + (t_n)$ is not a root unless $i = n - 1$. So t_n is only linked to $t_{n-1} - t_n$. This time, the sum $t_{n-1} - t_n$ with t_n is t_{n-1} and the sum with two copies of t_n is $t_{n-1} + t_n$. So $p = 0$ and $q = 2$ and

$$2 \frac{\langle t_{\alpha_{n-1}}, t_{\alpha_n} \rangle}{\langle t_{\alpha_n}, t_{\alpha_n} \rangle} = -2$$

and therefore

$$\frac{\langle t_{\alpha_{n-1}}, t_{\alpha_n} \rangle^2}{\|t_{\alpha_{n-1}}\|^2 \|t_{\alpha_n}\|^2} = \frac{1}{2}$$

So the angle between these two roots is 135 degrees, and in the Dynkin diagram there are two lines between these roots.

The Dynkin arrow goes from the large root to the small root. Which of our two roots is larger? We have

$$2 \frac{\langle t_{\alpha_{n-1}}, t_{\alpha_n} \rangle}{\langle t_{\alpha_n}, t_{\alpha_n} \rangle} = -2$$

and

$$2 \frac{\langle t_{\alpha_{n-1}}, t_{\alpha_n} \rangle}{\langle t_{\alpha_{n-1}}, t_{\alpha_{n-1}} \rangle} = -1$$

because in the first case we are adding multiples of α_n until we don't get a root, and in the second case we are adding multiples of α_{n-1} until we don't get a root, and the first

happens twice but the second happens only once. So clearly the larger root is α_{n-1} . Thus the arrow should go from α_{n-1} to α_n . This gives the B_n diagram.

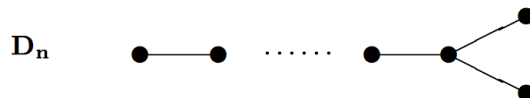
Finally, just as before the expression $-tr(XY)$ gives the Killing form on $so(2n)$ and $so(2n+1)$ and in particular on the algebra of the torus, the norm of each element is $\sum t_i^2$ and the geometry is the expected one. Details are exactly as for $su(n)$ and thus safely left to the reader.

12.2 Special Cases

The Dynkin Diagrams of A_1, B_1 , and C_1 all have a single dot. So $su(2), so(3)$, and $sp(1)$ have the same diagram and are isomorphic. We will say more about this when we discuss the symplectic group associated with C_1 . For now, note that the Lie group $SU(2)$ is the universal cover of $SO(3)$. This is a 2-fold cover.

The Dynkin Diagram of B_2 has two dots, and an arrow pointing from one to the other. This is also the diagram of C_2 , and indeed $so(5)$ and $sp(2)$ are isomorphic. We'll say more about this when we discuss the symplectic group.

Recall that the Dynkin Diagram of $so(2n)$ is



Here n should be at least 2 because $SO(2) = S^1$ is abelian and has no Dynkin diagram.

When $n = 2$, the diagram has two dots which are not connected. This equals the diagram of $A_1 \oplus A_1$. And indeed, $so(4) = su(2) \oplus su(2) = so(3) \oplus so(3) = sp(1) \oplus sp(1)$. The group $SU(2) \times SU(2)$ is topologically $S^3 \times S^3$. Its center is $Z_2 \times Z_2$. When we divide out the subgroup $\{(0,0), (1,1)\}$, we get $SO(4)$. If we divide out the entire center, we get $SO(3) \times SO(3)$. More details will be given later.

When $n = 3$, the diagram has three connected dots and equals A_3 . So $so(6)$ and $su(3)$ have the same Dynkin diagram, and indeed $SU(3)$ is the universal cover of $SO(6)$.

Finally when $n = 4$, we obtain a Dynkin Diagram with a central node and three spokes. The symmetry group of this diagram is the dihedral group of an equilateral triangle, a group with six elements. So $SO(8)$ has an unusually large number of outer automorphisms, corresponding to these symmetries of the Dynkin diagram.

Chapter 13

$Sp(n)$

Let $\mathcal{H}$ be the quaternions. By definition the *symplectic group* is the group $Sp(n)$ of all quaternionic isomorphisms of $\mathcal{H}^n$ preserving the quaternionic inner product, Here by definition,

$$\langle v, w \rangle = \langle (p_1, \dots, p_n), (q_1, \dots, q_n) \rangle = \sum \bar{p}_i q_i$$

where the p_i and q_i are quaternions and the conjugate of $q = a_0 + a_1i + a_2j + a_3k$ is

$$\bar{q} = a_0 - a_1i - a_2j - a_3k$$

The Lie algebra of this group is denoted $sp(n)$. We will describe the exact structure of matrices in these groups and algebras shortly.

Until this moment all of our inner products were linear in the first term and conjugate linear in the second term. But in the quaternionic case, if we want matrices to act on vectors from the left, then scalars must act on vectors from the right. The definition $\langle v, w \rangle = \sum \bar{p}_i q_i$ has the property that $\langle v, w\lambda \rangle = \langle v, w \rangle \lambda$ as desired. Note that $\langle v, w \rangle = \overline{\langle w, v \rangle}$ so that

$$\langle v\lambda, w \rangle = \overline{\langle w, v\lambda \rangle} = \overline{\langle w, v \rangle \lambda} = \bar{\lambda} \langle v, w \rangle$$

13.1 Symplectic Maps

Defining $Sp(n)$ as the group of quaternions gives a nice symmetry to the classification of compact groups, but it is not the original way of looking at this group. The older approach may be more useful in applications, so we discuss it first.

The orthogonal and unitary groups are groups of linear transformations preserving a symmetric or Hermitian inner product. But there is another object which occurs often in applications: a skew-symmetric form.

Definition 26 A non-singular skew-form on a vector space V defined over R or C is an assignment to any pair of vectors v and w of a scalar $\Omega(v, w)$ in R or C such that

- $\Omega(v, w)$ is linear in each variable if the other is held fixed
- $\Omega(v, w) = -\Omega(w, v)$
- if $v \neq 0$, there is a w such that $\Omega(v, w) \neq 0$

Remark: There is a beautiful standard canonical form for such skew-forms.

Theorem 66 Let Ω be a non-singular skew form on V , a vector space over R or C . Then the dimension of V is even, and V has a basis $e_1, \dots, e_n, f_1, \dots, f_n$ such that $\Omega(e_i, e_j) = \Omega(f_i, f_j) = 0$ and $\Omega(e_i, f_j) = \delta_{ij}$.

Proof: We prove this by induction on the dimension of V . Pick a non-zero $e_1 \in V$. Since Ω is non-singular, there is an f_1 with $\Omega(e_1, f_1) = 1$. These vectors must be linearly independent. Let $\{e_1, f_1\}$ be the subspace of V generated by these elements, and let $V^\perp = \{v \in V \mid \Omega(v, e_1) = \Omega(v, f_1) = 0\}$. We claim that $V = \{e_1, f_1\} \oplus V^\perp$ and that Ω restricted to $V^\perp$ is still a non-singular skew form. If so, we are done by induction.

To prove the direct sum statement, notice that for any $v \in V$ we can write

$$v = \left(\Omega(v, f_1)e_1 - \Omega(v, e_1)f_1 \right) \oplus \left(v - \Omega(v, f_1)e_1 + \Omega(v, e_1)f_1 \right)$$

QED.

Let J be the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

Notice that the skew-symmetric form defined by the canonical basis from the previous theorem is given by

$$\Omega(v, w) = v^T J w$$

A map $M : V \rightarrow V$ is said to be *symplectic* if it preserves the canonical skew-symmetric form. This happens if and only if $M^T J M = J$ because

$$\Omega(Mv, Mw) = (Mv)^T J Mw = v^T M^T J Mw$$

$$\Omega(v, w) = v^T J w$$

The set of all such matrices forms a group, called the real symplectic group $Sp(2n, R)$ or the complex symplectic group $Sp(2n, C)$ depending on the scalars for the vector space.

It turns out that the Lie algebra of $Sp(2n, C)$ is the complex simple Lie algebra C_n . The Lie algebra of $Sp(2n, R)$ is a real form of $Sp(2n, C)$, but unfortunately is not compact. To

get a compact group, we need to bound the sizes of our matrices, which can be done by making the linear transformations preserve the length of vectors. So one way to describe the group $Sp(n)$ is that it is $U(2n) \cap Sp(2n, C)$. This description emphasizes the symplectic or skew-symmetric nature of the group.

13.2 Lie Algebra of $Sp(n)$

Let us write the resulting Lie algebra down very concretely. Then we will show that its Dynkin diagram is C_n . Finally we will discuss the quaternionic definition of $Sp(n)$ and show that we get exactly the same matrices. This will complete the symplectic story.

A matrix M is in $Sp(2n, C)$ if $M^T J M = J$. To find the Lie algebra, imagine that $M(t)$ is a path starting at the origin and differentiate: $M'^T J + J M' = 0$. Consequently, the Lie algebra $sp(2n, C)$ consists of $2n \times 2n$ complex matrices M such that $M^T J = -J M$.

Write the matrices of the Lie algebra $sp(2n, C)$ as blocks of $n \times n$ complex matrices

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

Then

$$\begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

and we find that the Lie algebra matrices have the following form, with $B = B^T$ and $C = C^T$:

$$\begin{pmatrix} A & B \\ C & -A^T \end{pmatrix}$$

Recall, however, that $Sp(n) = U(2n) \cap Sp(2n, C)$. It follows that $sp(n) = u(2n) \cap sp(2n, C)$. The algebra $u(2n)$ contains matrices M such that $M = -\overline{M}^T$. Hence the matrices in $sp(n)$ satisfy

$$\begin{pmatrix} A & B \\ C & -A^T \end{pmatrix} = - \begin{pmatrix} \overline{A}^T & \overline{C}^T \\ \overline{B}^T & -\overline{A} \end{pmatrix}$$

It follows that $sp(n)$ contains all complex matrices of the following form, where $A = -\overline{A}^T$ and $B = B^T$:

$$\begin{pmatrix} A & B \\ -\overline{B} & \overline{A} \end{pmatrix}$$

13.3 $sp(n)$ is C_n

There is an obvious candidate for the maximal torus of $Sp(n)$:

$$\begin{pmatrix} e^{it_1} & & & & & \\ & e^{it_2} & & & & \\ & & \ddots & & & \\ & & & e^{it_n} & & \\ & & & & e^{-it_1} & \\ & & & & & e^{-it_2} \\ & & & & & & \ddots \\ & & & & & & & e^{-it_n} \end{pmatrix}$$

Since every element of $Sp(n)$ is conjugate to such an element, we find that if $M \in Sp(n)$, we can find a conjugate element that rotates a given e_i by θ and rotates the associated f_i by $-\theta$. Note that $e^{i\theta}e^{-\theta} = 1$, so the skew-symmetric form is preserved by this map.

The corresponding elements in the Lie algebra of the maximal torus have the form

$$\begin{pmatrix} it_1 & & & & & \\ & it_2 & & & & \\ & & \ddots & & & \\ & & & it_n & & \\ & & & & -it_1 & \\ & & & & & -it_2 \\ & & & & & & \ddots \\ & & & & & & & -it_n \end{pmatrix}$$

Since the A portion in the top-left of our matrices belong to $u(n)$, and the A portion in the bottom-right just mirrors what happens in the top-left, we can find many roots by copying what we did with $su(n+1)$. This gives us roots $(t_i - t_j)$ for $1 \leq i < j \leq n$.

The remaining root spaces must belong to B in the top-right, and be mirrored by the corresponding $-\bar{B}$ in the bottom-left. The calculation below shows how brackets work in this case; in the last line we use the fact that $\bar{T} = -T$.

$$\begin{aligned} & \begin{pmatrix} T & 0 \\ 0 & -T \end{pmatrix} \begin{pmatrix} 0 & B \\ -\bar{B} & 0 \end{pmatrix} - \begin{pmatrix} 0 & B \\ -\bar{B} & 0 \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & -T \end{pmatrix} \\ &= \begin{pmatrix} 0 & TB \\ T\bar{B} & 0 \end{pmatrix} - \begin{pmatrix} 0 & -TB \\ -\bar{B}T & 0 \end{pmatrix} = \begin{pmatrix} 0 & BT+TB \\ -\bar{B}T+T\bar{B} & 0 \end{pmatrix} \end{aligned}$$

Since $B^T = B$, our root spaces are generating by choosing a B with only one non-zero element on or above the diagonal, and then mirroring below the diagonal and also in the $-\overline{B}$ at bottom-left. If the nonzero element is in the diagonal spot b_{jj} , then the action sends this element to $i \ 2t_j b_{jj}$. So if we let u be the element of this type with $b_{jj} = i$ and we let v be the element of this type with $b_{jj} = 1$, we find that $[T, u] = -2t_j v$ and $[T, v] = 2t_j u$. Thus $2t_j$ is a root.

But if $j < k$ and the only non-zero element above or on the diagonal is b_{jk} , then the action sends this element to $i(t_j + t_k)b_{jk}$. If we let u be the element of this type with $b_{jk} = i$ and v be the element with $b_{jk} = 1$, then $[T, u] = -(t_j + t_k)v$ and $[T, v] = (t_j + t_k)u$. So $t_j + t_j$ is a root.

This gives us roots $t_j - t_k$ and $t_j + t_k$ for $j < k$ and $2t_i$.

We claim we can choose related simple roots $t_j - t_{j+1}$ and $2t_n$. Indeed as done several times earlier, linear combinations of these with positive integer coefficients give all $t_j - t_k$ with $j < k$. Also $(t_j - t_n) + (2t_n) = t_j + t_n$. Then $(t_{n-1} - t_n) + (t_{n-1} + t_n) = 2t_{n-1}$. Repeat the argument to get all $t_j + t_{n-1}$ and then $2t_{n-2}$. In the end, all positive roots can be obtained.

Finally we must determine the Dynkin diagram. Recall again our technique. The difference of two distinct simple roots is not a root, so if α and β are simple roots, then $\beta + k\alpha$ will be a root for $0 \leq k \leq q$ and

$$2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} = -q$$

Consider first the roots $t_1 - t_2, t_2 - t_3, \dots, t_{n-1} - t_n$. If we pick non-adjacent roots, then $(t_j - t_{j+1}) + (t_k - t_{k+1})$ is not a root. So $q = 0$ and these roots are perpendicular. However if we pick adjacent roots, then $(t_j - t_{j+1}) + (t_{j+1} - t_{j+2}) = t_j - t_{j+2}$ is a root, but $(t_j - t_{j+1}) + 2(t_{j+1} - t_{j+2}) = t_j + t_{j+1} - 2t_{j+2}$ is not a root, so $q = 1$. If we work in the opposite order, $(t_{j+1} - t_{j+2}) + (t_j - t_{j+1}) = t_j - t_{j+2}$ is a root, but $(t_{j+1} - t_{j+2}) + 2(t_j - t_{j+1}) = 2t_j - t_{j+1} - 2t_{j+2}$ is not a root, so $q = 1$

So

$$2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} 2 \frac{\langle t_\beta, t_\alpha \rangle}{\langle t_\beta, t_\beta \rangle} = (-1)(-1) = 1$$

and

$$\frac{\langle t_\alpha, t_\beta \rangle^2}{||t_\alpha||^2 ||t_\beta||^2} = \frac{1}{4}$$

So if θ is the angle between these vectors, $\cos^2 \theta + \frac{1}{4}$ and $\cos \theta = \pm \frac{1}{2}$ and the angle is 120 degrees and the nodes are connected by one link.

We have one other simple root, $2t_n$. This is not connected to any simple root except possibly $t_{n-1} - t_n$ because, for example, $(t_{n-2} - t_{n-1} + 2t_n)$ is not a root.

On the other hand, if $\beta = t_{n-1} - t_n$ and $\alpha = 2t_n$, then $\beta + \alpha = t_{n-1} + t_n$ is a root, but $\beta + 2\alpha = t_{n-1} - t_n + 4t_n$ is not a root. So $q = 1$. But if we reverse these roots, letting $\beta = 2t_n$ and $\alpha = t_{n-1} - t_n$, then their sum is $t_{n-1} + t_n$ and $\beta + 2\alpha = 2t_{n-1}$ is a root. On the other hand, $\beta + 3\alpha$ is not a root. So this time $q = 2$. Thus

$$\frac{\langle t_\alpha, t_\beta \rangle^2}{\|t_\alpha\|^2 \|t_\beta\|^2} = \frac{1}{2}$$

and $\cos^2 \theta = \frac{1}{2}$ so $\theta = 90 + 45 = 135$ degrees. Thus these roots are connected by two links.

Finally we have to determine which is the longer root. Sticking with our original selection of $\beta = t_{n-1} - t_n$ and $\alpha = 2t_n$, we have

$$2 \frac{\langle t_\alpha, t_\beta \rangle}{\langle t_\alpha, t_\alpha \rangle} = -1 \quad \text{and} \quad 2 \frac{\langle t_\beta, t_\alpha \rangle}{\langle t_\beta, t_\beta \rangle} = -2$$

So α is longer and the arrow should point from α to β . This is, indeed, the situation for C_n .

13.4 C_n Interpreted Using Quaternions

We will now obtain the same group using quaternions. An element of $\mathcal{H}^n$ can be written uniquely in the form $v = v_1 + jv_2$, where v_1 and v_2 are complex vectors. We are deliberately writing j on the left here, although we will continue to scalar multiply from the right. Thus we are thinking of $\mathcal{H}^n$ as $C_n \oplus C_n$.

Similarly a matrix M acting on $\mathcal{H}^n$ is a matrix of quaternions and can be written uniquely as $M = A + jB$ where A and B are $n \times n$ complex matrices. Since we scalar multiply from the right, this matrix acts from the left. We are deliberately writing the j on the left side.

Notice that

$$(A + jB)(v_1 + jv_2) = Av_1 + Ajv_2 + jBv_1 + jBjv_2 = (Av_1 - \bar{B}v_2) + j(Bv_1 + \bar{A}v_2)$$

Consequently, multiplication by a quaternionic matrix takes the form

$$\begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Let $\langle v, w \rangle$ be the quaternionic inner product on $\mathcal{H}^n$ introduced at the start of this chapter. Write $\langle v, w \rangle = \langle\langle v, w \rangle\rangle + j\Omega(v, w)$ where $\langle\langle v, w \rangle\rangle$ is complex and Ω is complex.

Since $\langle w, v \rangle = \overline{\langle v, w \rangle}$, we have

$$\begin{aligned} \langle \langle w, v \rangle \rangle + j\Omega(w, v) &= \overline{\langle \langle v, w \rangle \rangle + j\Omega(v, w)} \\ &= \overline{\langle \langle v, w \rangle \rangle} + \overline{\Omega(v, w)}(-j) = \overline{\langle \langle v, w \rangle \rangle} + j\Omega(v, w) \end{aligned}$$

We conclude that

$$\langle \langle w, v \rangle \rangle = \overline{\langle \langle v, w \rangle \rangle} \quad \text{and} \quad \Omega(w, v) = -\Omega(v, w)$$

So $\langle \langle v, w \rangle \rangle$ is a complex Hermitian inner product on

$$C_n \oplus C_n$$

and Ω is a skew-symmetry form on this space. Note that Ω is C -linear because $\langle v, wi \rangle = \langle v, w \rangle i$ and so

$$\langle \langle v, wi \rangle \rangle + j\Omega(v, wi) = \langle \langle v, w \rangle \rangle i + j\Omega(v, w)i$$

Remark: This explains, then, how the quaternionic inner product gives rise to a skew-symmetric form.

Remark: If $\langle v, w \rangle$ is a quaternionic inner product, we define $\|v\| = \langle v, v \rangle$. If $v = (q_1, \dots, q_n)$, then $\langle v, v \rangle = \sum \bar{q}_i q_i = \sum \|q_i\|^2$. Since $\|a_0 + a_1i + a_2j + a_3k\|^2 = a_0^2 + a_1^2 + a_2^2 + a_3^2$, the length of vectors in $\mathcal{H}^n$ is exactly the standard length on R^{4n} .

This same result holds if $\langle v, w \rangle$ is a Hermitian inner product on C^n because $\|a + bi\|^2 = a^2 + b^2$.

For real inner products, it is well known that a linear transformation which preserves distances automatically preserves the inner product. Indeed

$$\langle v + w, v + w \rangle = \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle$$

so

$$\langle v, w \rangle = \frac{1}{2} (\|v + w\|^2 - \|v\|^2 - \|w\|^2)$$

This result is also true over the complex numbers. This time

$$\begin{aligned} \|v + w\|^2 &= \|v\|^2 + \langle v, w \rangle + \langle w, v \rangle + \|w\|^2 \\ \|v - w\|^2 &= \|v\|^2 - \langle v, w \rangle - \langle w, v \rangle + \|w\|^2 \\ \|v + w\|^2 - \|v - w\|^2 &= 2\langle v, w \rangle + 2\langle w, v \rangle \end{aligned}$$

and

$$\begin{aligned}
||v + iw||^2 &= ||v||^2 + i \langle v, w \rangle - i \langle w, v \rangle + ||w||^2 \\
||v - iw||^2 &= ||v||^2 - i \langle v, w \rangle + i \langle w, v \rangle + ||w||^2 \\
||v + iw||^2 - ||v - iw||^2 &= 2i \langle v, w \rangle - 2i \langle w, v \rangle
\end{aligned}$$

and we can get $\langle v, w \rangle + \langle w, v \rangle$ as a combination of distances from the first set and $\langle v, w \rangle - \langle w, v \rangle$ as a combination of distances from the second set.

If we forget that we can scalar multiply by j and think of $\mathcal{H}^n$ as $C^n \oplus C^n$, then $Sp(n)$ is a group of complex linear transformations which preserve the lengths of vectors. But $U(2n)$ is another group of complex linear transformations which preserve the lengths of vectors. These lengths are the same, so both of these groups also preserve the associated inner products $\langle \langle v, w \rangle \rangle$. So the only difference between these groups is that elements of $Sp(n)$ also respect scalar multiplication by j and preserve Ω .

Lemma 28 *If $M : \mathcal{H}^n \rightarrow \mathcal{H}^n$ is linear over the quaternions and preserves the complex Hermitian inner product, then it automatically preserves Ω .*

Proof: We have $\langle v, wj \rangle = \langle v, w \rangle j$, so

$$\langle \langle v, wj \rangle \rangle + j\Omega(v, wj) = \langle \langle v, w \rangle \rangle j + j\Omega(v, w)j = \langle \langle v, w \rangle \rangle j - \overline{\Omega(v, w)}$$

It follows that

$$\Omega(v, w) = -\overline{\langle \langle v, wj \rangle \rangle}$$

Then

$$\begin{aligned}
\Omega(Mv, Mw) &= -\overline{\langle \langle Mv, M(w)j \rangle \rangle} = -\overline{\langle \langle Mv, M(wj) \rangle \rangle} \\
&= -\overline{\langle \langle v, wj \rangle \rangle} = \Omega(v, w)
\end{aligned}$$

QED.

Remark: We want to intersect $u(2n)$ with the Lie algebra of quaternionic-linear maps preserving the quaternionic inner-product. By the lemma, this is the intersection of $u(2n)$ with the Lie algebra of quaternionic-linear maps preserving $\langle \langle v, w \rangle \rangle$. In turn, this is the intersection of $u(2n)$ with the Lie algebra of quaternionic-linear maps, since any element of this intersection is in $u(2n)$ and thus preserves $\langle \langle v, w \rangle \rangle$. The Lie algebra of $u(2n)$ consists of all

$$M = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

such that $-\overline{M}^T = M$, and thus the set of all matrices with the following form:

$$\begin{pmatrix} A & -\overline{B}^T \\ B & D \end{pmatrix}$$

with $A = -\overline{A}^T$ and $D = -\overline{D}^T$. By an earlier result, the Lie algebra of quaternionic linear maps is the set of all

$$\begin{pmatrix} A & -\overline{B} \\ B & \overline{A} \end{pmatrix}$$

The intersection is thus the set of all

$$\begin{pmatrix} A & -\overline{B} \\ B & \overline{A} \end{pmatrix}$$

such that $A = -\overline{A}^T$ and $B = B^T$.

But this is exactly the set of matrices for $u(2n) \cap sp(2n, C)$ obtained in section 13.2. Consequently the description of $sp(n)$ as distance-preserving symplectic maps, and as quaternionic maps which preserve a quaternionic inner product give the same algebra.

Chapter 14

Irreducible Complex Representations of Compact Groups

14.1 Weights

In this chapter, G is a compact simply-connected Lie group with connected Dynkin diagram, and $\mathcal{G}$ is its Lie algebra. Suppose φ is a complex irreducible representation of G . Then φ induces a corresponding irreducible Lie algebra representation of $\mathcal{G}$. By elementary Lie theory there is a one-to-one correspondence between irreducible representations of G and the corresponding irreducible representations of $\mathcal{G}$.

If G is not simply connected, some Lie algebra representations may not induce Lie group representations. This aspect of the theory for Lie groups that are not simply connected will be considered in a later chapter. In this chapter, we begin the classification of all irreducible representations of $\mathcal{G}$ and thus of simply-connected G .

We are deliberately choosing to study representations on a *complex* vector space. The theory of chapter two applies without change to compact Lie groups, and allows us to also classify real and quaternionic irreducible representations, but the heart of the theory lies in the complex case.

Restrict φ to the maximal torus T of G . Because V is a complex space, this restriction splits into irreducible representations which are all one-dimensional. The universal covering group of T is canonically isomorphic to the Lie algebra $\mathcal{T}$ of T , so there exist linear maps $w : \mathcal{T} \rightarrow \mathbb{R}$ and vectors $e_w \in V$ such that $\varphi(t)e_w = e^{2\pi i w(t)}e_w$. We call these w the *weights*

of the representation. Let φ^* be the induced Lie algebra representation. Note that

$$\varphi^*(t)(e_w) = 2\pi i w(t)e_w.$$

Earlier in these lectures we studied the special case of the Ad representation of G on $\mathcal{G}$, where the weights are called roots. We first denoted these by $f : \mathcal{T} \rightarrow R$, and later by α . As we will see shortly, these roots still play a significant role in representation theory; therefore we deliberately denote the new objects by w to make clear that $e_w \in V$ while $e_\alpha \in \mathcal{G} \otimes C$.

Suppose α is a root for $\mathcal{G}$ and let $e_\alpha = u + iv$ be a root vector. Recall that this root vector lives in $\mathcal{G} \otimes C$, but both u and v live in $\mathcal{G}$. We get

$$\varphi^*([t, e_\alpha])e_w = [\varphi^*(t), \varphi^*(e_\alpha)]e_w = \varphi^*(t)\varphi^*(e_\alpha)e_w - \varphi^*(e_\alpha)\varphi^*(t)e_w$$

and so

$$2\pi i \alpha(t)\varphi^*(e_\alpha)(e_w) = \varphi^*(t)\varphi^*(e_\alpha)e_w - 2\pi i w(t)\varphi^*(e_\alpha)e_w$$

and so

$$\varphi^*(t)\varphi^*(e_\alpha)e_w = 2\pi i(\alpha + w)(t)\varphi^*(e_\alpha)e_w$$

It follows that either $\varphi^*(e_\alpha)e_w = 0$ or else it is a weight vector with weight $\alpha + w$. A similar conclusion follows using $e_{-\alpha}$.

Let us pause and clarify some steps taken rather casually in the previous paragraph. We started with a real Lie algebra $\mathcal{G}$ and a real Lie homomorphism $\varphi^* : \mathcal{G} \rightarrow gl(V, C)$. Since the image is an algebra over C , we can trivially extend this map to a *complex* Lie homomorphism $\mathcal{G} \otimes C \rightarrow gl(V, C)$. We will often do this without comment in the future; an advantage is that the e_α belong to the domain.

We sometimes apply an easy form of Weyl's Unitary Trick. For instance, $\mathcal{G}$ has many real subalgebras isomorphic to $su(2)$. So if we have a Lie algebra representation of $\mathcal{G}$, we can restrict it to $su(2)$ and analyze the decomposition of V into irreducible subspaces under this subgroup. But what we often do instead is to find the equivalent basis for $sl(2, R)$ in $\mathcal{G} \otimes C$, extend to this algebra, and then restrict to the real algebra $sl(2, R)$ and decompose this representation. Why is this legal, and how do we argue that we get exactly the same decomposition of V ?

A complex decomposition of V into complex subspaces will be invariant under $su(2)$ if and only if it is invariant under $su(2) \otimes C$, and these subspaces will be irreducible under $su(2)$ if and only if they are irreducible under $su(2) \otimes C$. A similar comment holds for $sl(2, R)$. But φ^* on $su(2)$ is induced by φ on a compact subgroup of G locally isomorphic to $SU(2)$, so every complex representation space V can be decomposed as a sum of irreducible invariant subspaces. This decomposition also works for φ^* on $su(2)$ and thus for φ^* extended to $su(2) \otimes C = sl(2, R) \otimes C$, and so to the restriction of φ^* to $sl(2, R)$.

Resuming our adventure, we can start with a root α and form the subalgebra $su(2)$ generated by $\{u, v, \alpha\}$. We can then study the representation of the corresponding $sl(2, R)$ and decompose V under it. We have to be somewhat careful. Complex root spaces are one-dimensional, but complex weight spaces can be higher dimensional. If we start with a weight vector and apply $\varphi^*(e_{-\alpha})$ and $\varphi^*(e_\alpha)$, we can get a string of weights $w + p\alpha, \dots, w, \dots, w + q\alpha$ for $p \leq 0$ and $q \geq 0$ by looking at weight vectors $\varphi^*(e_{-\alpha})^k e_w$ and $\varphi^*(e_\alpha)^l e_w$, but it can happen that $\varphi^*(e_\alpha)^{q+1} e_w$ is zero and yet $w + (q+1)\alpha$ is a weight. That is because a second linearly independent e_w can lead to a longer chain.

Let us carry out this program. Earlier we discovered that the $sl(2, R)$ in $su(2) \otimes C$ has a basis

$$\begin{aligned} H &= \frac{-i}{\pi \|\alpha\|^2} \alpha \\ L_+ &= \frac{-i}{2\pi \|\alpha\|} e_\alpha \\ L_- &= \frac{-i}{2\pi \|\alpha\|} e_{-\alpha} \end{aligned}$$

An irreducible representation is formed by acting on e_w with $\varphi^*(e_\alpha)$ to get elements in weight spaces with weights $w, w + \alpha, \dots, w + q\alpha$, where all of these elements are non-zero, but the next one is zero, and then acting on e_w with $\varphi^*(-\alpha)$ to get elements in weight spaces with weights $w, w - \alpha, \dots, w - (p)\alpha$, where all of these elements are non-zero but the next one is zero. These vectors then form an irreducible representation of $sl(2, R)$ of dimension $n + 1$.

This chain of vectors are eigenvectors of H with eigenvalues

$$\begin{aligned} &0 \\ &-1, 1 \\ &-2, 0, 2 \\ &-3, -1, 1, 3 \end{aligned}$$

for dimensions 1, 2, 3, 4, etc.

Theorem 67 *Let φ be a representation and let e_w be a weight vector. Suppose $q \geq 0$ and $\varphi^*(e_\alpha)^q e_w \neq 0$ but $\varphi^*(e_\alpha)^{q+1} = 0$. Suppose $p \leq 0$ and suppose $\varphi^*(e_{-\alpha})^{-p} e_w \neq 0$ but $\varphi^*(e_{-\alpha})^{-p-1} e_w = 0$. Then $p \leq 0$ and $q \geq 0$ and $w + p\alpha, w + (p-1)\alpha, \dots, w, w + \alpha, \dots, w + q\alpha$ are weights. Moreover*

$$\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} = -(p+q) \qquad w - \frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} \alpha \text{ is a weight}$$

Proof: We have a chain of weight vectors. The top weight vector is an eigenvector of $\varphi^*(H)$ with eigenvalue n and the lowest weight vector is an eigenvector of $\varphi^*(H)$ with eigenvalue $-n$. Thus the sum of these eigenvalues is 0. But $H = -i \frac{\alpha}{\langle \alpha, \alpha \rangle \pi}$ and

$$\varphi\left(-i \frac{\alpha}{\langle \alpha, \alpha \rangle \pi}\right) \varphi(e_\alpha)^q e_w = 2\pi i \left(\frac{-i}{\langle \alpha, \alpha \rangle \pi}\right) (w + q\alpha)(\alpha) \varphi(e_\alpha)^q e_w = \left(\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} + 2q\right) \varphi(e_\alpha)^q e_w$$

Similarly this expression applied to $\varphi(e_{-\alpha})^{-p} e_w$ is

$$\left(\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} + 2p\right) \varphi(e_{-\alpha})^q e_w$$

Setting one eigenvalue equal to the negative of the other gives

$$\left(\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} + 2q\right) = -\left(\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} + 2p\right)$$

or

$$\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} = -(p + q)$$

The second result is implied by the first, because p is negative and q is positive and thus $p \leq p + q \leq q$, so $w + (p + q)\alpha$ is a weight, so $w - \frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} \alpha$ is a weight.

14.2 Highest Weight

Definition 27 Let φ be a representation on V of a compact Lie group with connected Dynkin diagram. A weight w is a highest weight if $w + \alpha$ is not a weight for positive roots α .

Theorem 68 Every such representation has a highest weight.

Proof: Select a weight w . Select simple roots $\alpha_1, \dots, \alpha_k$. Then every positive root can be written uniquely as $\sum m_i \alpha_i$ where the m_i are non-negative integers. If two weights are different, their weight vectors are linearly independent, so there are only finitely many weights, and in particular only finitely many $w + \sum m_i \alpha_i$ which are weights. Choose a weight of this form with $\sum m_i$ as large as possible. Then w is a highest weight, since if $w + \alpha$ is a weight for some positive alpha, the alpha would be $\sum n_i \alpha_i$ with $n_i \geq 0$ and at least one $n_i > 0$ and we would have a contradiction.

Warning: For general representations, the highest weight may not be unique.

Theorem 69 If w is a highest weight, $w(\alpha) \geq 0$ for all positive roots.

Proof: Assume $w(\alpha) < 0$. By an earlier result, $w - \frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} \alpha$ is a weight. Here the coefficient of α is an integer, and the string $w, w + \alpha, w + 2\alpha, \dots$ contains this integer. So $w + \alpha$ is a weight.

Theorem 70 *Suppose φ is an irreducible complex representation on V of a compact Lie group with connected Dynkin diagram. Then φ has a unique highest weight. The eigenspace for this highest weight has dimension one. If $\alpha_1, \dots, \alpha_k$ are the simple roots, then $\frac{2w(\alpha_i)}{\langle \alpha_i, \alpha_i \rangle}$ is a non-negative integer for each i . These integers completely determine w .*

Proof: Let e_w be a weight vector for a highest weight. The main step of the proof constructs V from e_w .

Consider all vectors of the form $\varphi(e_{-\alpha_i})e_w$ where the α_i are simple positive roots. Some of these will be zero and can be ignored. Repeat the process, adding additional vectors of the form $\varphi(e_{-\alpha_j})\varphi(e_{-\alpha_i})e_w$. Repeat until no new vectors are added to the process. Let W be the span of all these vectors. Notice that W is invariant under $\mathcal{T}$ since we added weight vectors. We will prove that W is invariant under all $\varphi(e_\alpha)$ and $\varphi(e_{-\alpha})$. If so, $W = V$ because V is irreducible.

By a previous result, a positive root α can be written $\alpha_1 + \alpha_2 + \dots + \alpha_k$ where the α_i are simple (but not necessarily distinct) and for each j , $\alpha_1 + \dots + \alpha_j$ is a root. We also proved earlier that when α, β , and $\alpha + \beta$ are roots we have $[E_\alpha, E_\beta] = E_{\alpha+\beta}$. Since α_1, α_2 , and $\alpha_1 + \alpha_2$ are roots, $[e_{\alpha_2}, e_{\alpha_1}] = \lambda e_{\alpha_1+\alpha_2}$ where $\lambda \neq 0$. In general, then, up to a non-zero constant $e_\alpha = [e_{\alpha_j}, \dots [e_{\alpha_2}, e_{\alpha_1}] \dots]$. A similar result holds for negative roots, written as a sum of negatives of simple roots.

It immediately follows that W is invariant under $\varphi(e_\alpha)$ for arbitrary negative roots. To show it invariant under general, we only need prove that it is invariant under $\varphi(e_\alpha)$ for α a positive *simple* root.

Notice first that $\varphi(e_\alpha)e_w = 0$ since w is a highest weight, so $w + \alpha$ is not a weight. Next consider $\varphi(e_\alpha)(\varphi(e_{-\beta})e_w)$ where α and β are simple positive roots. This equals

$$\varphi([e_\alpha, e_{-\beta}])e_w + \varphi(e_{-\beta})\varphi(e_\alpha)e_w$$

The last term is zero because $\varphi(e_\alpha)e_w = 0$. If $\beta = \alpha$, the bracket in the first term is in $\mathcal{T}$ and thus the first term preserves e_w . Otherwise the first term is zero because $\alpha - \beta$ is not a root. Indeed, if it is a positive root then $\alpha = (\alpha - \beta) + \beta$ and so α is not simple. A similar argument shows that $\alpha - \beta$ cannot be a negative root.

We finish the invariance argument by induction on the number of $\varphi(e_{-\alpha_i})$ needed to create the terms in W . Therefore suppose $v_1 \in W$ is invariant under all $e_{-\alpha}$ for simple α and consider $\varphi(e_\beta)v_1$ where β is simple. Then for α simple we have

$$\varphi(e_\alpha)\varphi(e_{-\beta})v_1 = [\varphi(e_\alpha), \varphi(e_{-\beta})]v_1 - \varphi(e_{-\beta})\varphi(e_\alpha)v_1$$

The last term is in W by induction and the construction of W . The first term is in W because it is zero unless $\alpha = \beta$ by the reasoning used in the initial step of the proof.

From here, the key result of the theorem is easy. Suppose w and w' are different highest weights, with weight vectors e_w and $e_{w'}$. From a knowledge of the construction of V we know that $e_{w'}$ is in V , and thus $w' = w - \beta$ where β is a linear combination of positive simple roots with positive integer coefficients. So β is a positive vector. For the same reason, $w = w' - \gamma$ where γ is a positive vector. So $\beta = w - w'$ and $\gamma = w' - w = -\beta$. Contradiction, since both vectors are positive.

It follows that V is generated by e_w and elements of smaller weight spaces, so the w weight space has dimension one in V .

QED.

Remark: Consider the Dynkin diagram of our compact group G and let φ be an irreducible complex representation of G . The last part of the previous theorem says that the highest root assigns a non-negative integer to each node of the diagram. These integers completely determine the highest root.

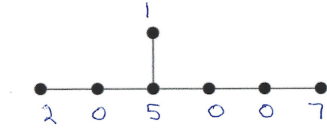


Figure 14.1: One Irreducible Representation of E_7

Theorem 71 Suppose φ_1 and φ_2 are irreducible complex representation on V_1 and V_2 of a compact Lie group with connected Dynkin diagram. Suppose they have the same highest weight. Then there is an isomorphism $\psi : V_1 \rightarrow V_2$ taking φ_1 to φ_2 , so the two representations are equivalent.

Proof: The proof is essentially the same as the proof of theorem 64, showing that two compact Lie algebras with the same connected Dynkin diagram are isomorphic.

Consider $\varphi_1 \oplus \varphi_2$ acting on $V_1 \oplus V_2$. Let v_1 and v_2 be weight vectors in V_1 and V_2 for the highest weights of φ_1 and φ_2 and consider $v_1 \oplus v_2$. Let $W \subset V_1 \oplus V_2$ be the subspace generated by acting on $v_1 \oplus v_2$ by repeated negative simple roots:

$$\dots (\varphi_1 \oplus \varphi_2)(e_{-\alpha_j})(\varphi_1 \oplus \varphi_2)(e_{-\alpha_i})(v_1 \oplus v_2)$$

The subspace W is clearly invariant under the action of $e_{-\alpha}$ for arbitrary negative roots, and of the torus. It is also invariant under the action of the e_{α} for arbitrary positive roots by essentially the argument used in the proof of Theorem 70. Moreover, this argument shows that the weight space generated by $v_1 \oplus v_2$ is one dimensional.

We are going to show that W is the graph of an isomorphism from V_1 to V_2 intertwining φ_1 and φ_2 . Notice that the obvious projection maps $V_1 \oplus V_2 \rightarrow V_1$ and $V_1 \oplus V_2 \rightarrow V_2$ induce maps $W \rightarrow V_1$ and $W \rightarrow V_2$ which are onto. For instance, the image of $W \rightarrow V_1$ is an invariant subspace containing v_1 and thus is all of V_1 . These two maps clearly intertwine $\varphi_1 \oplus \varphi_2$ with φ_1 and with φ_2 . If both maps are one-to-one, then we are done and the desired isomorphism is the inverse of $W \rightarrow V_1$ followed by the map $W \rightarrow V_2$.

The argument that these maps are one-to-one is the same for V_1 and V_2 , so we give it for V_1 . If $W \rightarrow V_1$ is not one-to-one, then the kernel of this map is the set of all $w_1 \oplus w_2 \in W$ with $w_1 = 0$. Thus it equals $W \cap V_2$ and since both of these are invariant, so is the intersection. But V_2 is irreducible, so this kernel is either zero or else all of V_2 . In the second case, $0 \times v_2 \in W$, but $v_1 \times v_2 \in W$, so the weight space containing $v_1 \oplus v_2$ must have dimension greater than 1, contradicting an earlier conclusion. QED.

Remark: There is one final step in the proof that highest weights completely classify irreducible representations of compact simple Lie groups. That step shows that every highest weight, i.e., numbering system on the nodes of the Dynkin diagram, comes from an irreducible representation. The final step is considerably more difficult than the results we just proved, but can be proved in many different ways. We will give a proof first given by Hermann Weyl.

The classification of irreducible representations of simple Lie algebras using highest weights was first stated and proved by Elie Cartan in 1913.

Chapter 15

Kaleidoscopes, Lattices, and the Weyl Group

15.1 Aspirations

The ultimate goal of the remaining chapters is to prove the existence theorem: every highest weight comes from an irreducible representation.

Our proof will follow the proof given by Herman Weyl, who obtained the theorem from a deep study of the characters of these irreducible representations. Since characters are constant on conjugacy classes, it suffices to restrict these characters to the maximal torus. The universal covering of this torus is R^k , so ultimately the characters will be functions on R^k .

Weyl first determined the rough form of these characters; they turn out to be finite Laurent series in simple exponentials. Since irreducible characters have a unique highest weight, these Laurent polynomials have a unique term of highest positive degree. Earlier we defined the Weyl group to be the finite group $N(T)/T$ which acts on T , and proved that two elements of a torus are conjugate if and only if they are equivalent under this group. Since characters are constant on conjugacy classes, the characters must be invariant under the Weyl group. We will study this group in detail shortly.

Finally, the characters must be orthogonal, but here we need to be cautious. Orthogonality is defined by integrating over the entire group G , not over the torus T . Weyl computed a weight function $w(t)$ on T , with the property that

$$\int_G f(g) \, dg = \int_T f(t) w(t) \, dt$$

for functions f constant on conjugacy classes. Thus the characters must be orthogonal on T with respect to this weight function.

Then Weyl showed that these simple properties completely determine the characters; it is possible to inductively compute the characters from their highest weights using these properties. This computation makes sense even if a highest weight doesn't correspond to an irreducible representation. In a sense, then, Weyl computed the characters of all irreducible representations, even those that do not exist.

Recall that for finite groups, the characters of the irreducible representations give a basis for the space of functions on G which are constant on conjugacy classes. When we generalize to compact groups, this space of functions is infinite dimensional; using invariant integration, it becomes a Hilbert Space. Weyl managed to prove the same theorem in the general compact case. The proof uses analysis, but very little from the elaborate classification theorem.

But then if some highest weight did not come from an irreducible representation, Weyl's construction of the weights would give an element of $L^2(G)$ which is orthogonal to this basis, a contradiction. Therefore, Weyl obtained an (indirect) proof of the existence theorem for irreducible representations.

Weyl found an explicit formula for the characters, and we will prove that *Weyl Character Formula* in these notes. If χ is a character, $\chi(e)$ is the dimension of the representation, so from the character formula Weyl could deduce the *Weyl Dimension Formula* giving the dimension of each irreducible representation associated with a particular highest weight. We will also prove that result.

There is a completely different proof of existence, by directly constructing all irreducible representations, computing their highest weights, and showing that the resulting weights exhaust the possibilities. About ten years after his paper on character theory, Weyl carried out this program for the classical groups (but not the exceptional groups), writing about them in a famous book named *The Classical Groups*. We will discuss some of these results at the end of these notes.

Along the path toward these high points, we will obtain many other interesting and important results.

15.2 The Root Lattice and the Weight Lattice

Recall that a *lattice* in a finite dimensional real vector space V is a subgroup $L \subset V$ such that

- the vectors in L generate V
- there is an open neighborhood of the origin $\mathcal{U} \subset V$ such that $\mathcal{U} \cap L = \{0\}$

According to the *fundamental theorem of lattice theory*, there is a basis $e_1, \dots, e_n$ of V consisting of lattice vectors, such that $L = \{m_1 e_1 + \dots + m_n e_n \mid m_i \in \mathbb{Z}\}$. Often V comes with an inner product, and in that case the e_i are usually not orthogonal.

Remark: Suppose $\alpha_1, \dots, \alpha_n$ are simple roots. We proved that these vectors are linearly independent in $\mathcal{T}$ and that every root has the form $\sum n_i \alpha_i$ where the n_i are integers. It follows that the α_i form a lattice basis for a lattice in $\mathcal{T}$.

Definition 28 *The lattice in $\mathcal{T}$ generated by the roots is called the root lattice and denoted L_R . The simple roots provide a lattice basis for this lattice.*

We know that $\mathcal{T}$ has a natural positive definite metric, but we have to be careful with it. For instance, the α_i above are not orthonormal.

Remark: When we discussed weights earlier, we thought of them as linear maps $w : \mathcal{T} \rightarrow \mathbb{R}$, and we proved that all of our weights had the property that whenever α is a root, we have $\frac{2w(\alpha)}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}$. The standard trick allows us to identify these weights with vectors in $\mathcal{T}$ which we will also call w , so $w(t) = \langle w, t \rangle$. We are interested in $w \in \mathcal{T}$ such that $\frac{2\langle w, \alpha \rangle}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}$. Thus we define

Definition 29 *A weight for a compact Lie group is a vector $w \in \mathcal{T}$ with the property that whenever α is a root we have*

$$\frac{2 \langle w, \alpha \rangle}{\langle \alpha, \alpha \rangle} \in \mathbb{Z}$$

The weights form a lattice known as the weight lattice L_W .

Remark: The results which follow will prove that L_W is a lattice. Notice that roots certainly satisfy the weight condition, so trivially $L_R \subset L_W$.

Remark: At the moment, our root system comes from a compact Lie group. We are going to "abstract" the central ideas with a short list of axioms, so root systems don't have to come from compact groups. But these abstract root systems will still have simple roots, Dynkin diagrams, and lattices. The Dynkin diagrams are exactly those already discussed, so the compact groups are "really there" but can be ignored.

Definition 30 *An abstract root system is a finite collection of nonzero vectors in a real vector space with positive definite inner product V , such that*

- The vectors generate V
- If α is a root, the only real multiples of α which are roots are $\pm\alpha$.
- If α and β are roots, $\frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}$ is an integer
- If α and β are roots, so is $\beta - \frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}\alpha$

Note that the roots associated to a compact Lie group satisfy all of these properties.

Remark: We claim that the construction of simple roots works for any abstract root system. Introduce a hyperplane in V and thus separate the elements of this abstract system into positive and negative roots. Call a positive root "simple" if it is not a sum of two positive roots. Then every root is a sum of simple roots with integer coefficients. This much is easy.

Now we prove that the simple roots in our abstract system are linearly independent. We proved this for ordinary roots from a lemma stating that whenever α and β are simple roots, $\langle\alpha,\beta\rangle \leq 0$. We need only prove this lemma in general.

If α and β are unequal positive simple roots, then

$$\frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle} \cdot \frac{2\langle\beta,\alpha\rangle}{\langle\beta,\beta\rangle} = \left(2 \frac{\langle\alpha,\beta\rangle}{\|\alpha\| \|\beta\|}\right)^2 \leq 4$$

If the number on the right is 4, then α and β are colinear, but both are positive, so root axioms give $\alpha = \beta$. Since we are assuming $\alpha \neq \beta$, the number on the right is 0, 1, 2, 3. If it is zero, $\langle\alpha,\beta\rangle = 0$ and we are done.

Otherwise the integers $\frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}$ and $\frac{2\langle\beta,\alpha\rangle}{\langle\beta,\beta\rangle}$ are $(-1, -1), (-1, -2), (-2, -1), (-1, -3), (-3, -1)$ or else one of these with both terms positive. In the negative cases, $\langle\alpha,\beta\rangle < 0$ and we are done. In each of the positive cases, one of the two terms is 1. For instance, suppose $\frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle} = 1$. Then

$$\beta - \frac{2\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}\alpha = \beta - \alpha$$

is a root. If it is positive, then $\beta = (\beta - \alpha) + \alpha$ and β is not simple. If it is negative, then $\alpha = -(\beta - \alpha) + \beta$ and α is not simple. QED.

Remark: It follows that if $\alpha_1, \dots, \alpha_n$ are the simple roots of an abstract root system, then every root is $m_1\alpha_1 + \dots + m_n\alpha_n$ and consequently belongs to the lattice generated by the α_i . It also follows that an abstract root system has a Dynkin diagram, which must be one of the diagrams we listed earlier.

Remark: Once we have an abstract root system, we can construct a new abstract root system from it, called the *associated coroot system*, by replacing each α by $\frac{2}{\langle\alpha,\alpha\rangle}\alpha$. Indeed,

the first two conditions for coroots are obvious. Condition 3 for coroots holds because

$$\frac{2 \left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \frac{2\beta}{\langle \beta, \beta \rangle} \right\rangle}{\left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \frac{2\alpha}{\langle \alpha, \alpha \rangle} \right\rangle} = \frac{2 \langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \frac{\langle \alpha, \alpha \rangle \langle \alpha, \alpha \rangle}{\langle \alpha, \alpha \rangle \langle \beta, \beta \rangle} = 2 \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}$$

is an integer. The last condition holds for coroots because

$$\frac{2\beta}{\langle \beta, \beta \rangle} - \frac{2 \left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \frac{2\beta}{\langle \beta, \beta \rangle} \right\rangle}{\left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \frac{2\alpha}{\langle \alpha, \alpha \rangle} \right\rangle} \frac{2\alpha}{\langle \alpha, \alpha \rangle} = \frac{2\beta}{\langle \beta, \beta \rangle} - 2 \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} \frac{2\alpha}{\langle \alpha, \alpha \rangle}$$

is equal to

$$\frac{2 \left(\beta - \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha \right)}{\langle \beta, \beta \rangle} = \frac{2 \left(\beta - \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha \right)}{\langle \beta - \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha, \beta - \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha \rangle}$$

This is the coroot associated with $\beta - \frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha$. QED.

Remark: It follows that the coroots also generate a lattice, called the *coroot lattice*. Notice that the coroot of a coroot is the original root. Thus the coroot construction is a sort of "duality".

Remark: There is a different and unrelated kind of duality that applies to lattices:

Definition 31 Let $L \subset V$ be a lattice in a vector space with inner product. The dual lattice is the set

$$\{ v \in V \mid \langle v, w \rangle \in \mathbb{Z} \text{ whenever } w \in L \}$$

Lemma 29 The dual lattice is a lattice.

Proof: Temporarily consider the dual space V^* and its corresponding dual basis φ_j where $\varphi_j(e_i) = \delta_{ij}$. The inner product on V defines an isomorphism $\psi : V \rightarrow V^*$ by

$$\psi(v)(w) = \langle v, w \rangle$$

Select vectors $f_j \in V$ which map to the φ_j by this isomorphism. Then

$$\langle f_j, e_i \rangle = \delta_{ij}$$

It immediately follows that the f_j are in the dual lattice and indeed form a lattice basis for this dual lattice. QED.

Remark: Look back at the definition of the weight lattice and notice that we have proved that it is a lattice. Indeed

Theorem 72 Consider the root system for a compact Lie group and form the corresponding coroot system and its lattice. The dual of this coroot lattice is the weight lattice of the Lie group.

Remark: We want to investigate the coroot lattice in more detail. The following key result is important but slightly tricky to prove:

Theorem 73 If $\alpha_1, \dots, \alpha_n$ are the simple roots in an abstract root system, then the corresponding coroots, $\left\{ \frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle} \right\}$, are the simple roots for the coroot system, and thus form a basis for the coroot lattice.

Corollary 9 An element $w \in \mathcal{T}$ belongs to the root lattice if and only if whenever α_i is a simple root we have

$$\frac{2 \langle w, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \in \mathbb{Z}$$

Proof of theorem: Note that a hyperplane in V separating the roots into positive and negative sets also separates the coroots. In particular, α is positive if and only if $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$ is positive. Both the simple roots and the simple coroots form a basis of V . So to prove the theorem, it suffices to prove that if α_i is a simple root, then $\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle}$ is a simple coroot. If this is false, then $\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle}$ is a sum of two positive coroots, so there are positive roots β_1, β_2 such that

$$\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle} = \frac{2\beta_1}{\langle \beta_1, \beta_1 \rangle} + \frac{2\beta_2}{\langle \beta_2, \beta_2 \rangle}$$

This implies that α_i is a linear combination, with non-negative rational coefficients, of β_1 and β_2 . But β_1 and β_2 are both linear combinations of the α_j with non-negative integer coefficients. Since the α_j are a basis of V , they are linearly independent over $\mathbb{R}$. It follows that the expressions for β_1 and β_2 in terms of the α_j can only involve α_i . But no positive multiple of α_i is a root except α_i itself, so $\beta_1 = \beta_2 = \alpha_i$, a contradiction.

Remark: The theorem we have just proved allows us to find the relationship between the Dynkin diagrams of a root space and its associated coroot space. If α and β are distinct simple roots, $\frac{2\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}$ is a nonpositive integer, and so

$$\left(\frac{2 \langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \right) \left(\frac{2 \langle \beta, \alpha \rangle}{\langle \beta, \beta \rangle} \right) = 4 \frac{\langle \alpha, \beta \rangle^2}{\|\alpha\|^2 \|\beta\|^2} = 4 \cos^2 \theta$$

is an integer. This integer can only be $k = 0, 1, 2, 3$, since otherwise $\theta = 90$ degrees and α and β are equal. We draw k lines joining α and β . If $k > 1$, then one of α and β is longer than the other one, and we draw an arrow from the long root to the short root.

Earlier we proved that

$$2 \left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \frac{2\beta}{\langle \beta, \beta \rangle} \right\rangle = 2 \frac{\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle}$$

and it follows that two coroots are joined by the same number of lines in the coroot Dynkin diagram that the corresponding roots were joined in the root Dynkin diagram.

On the other hand, the length of $\frac{2\alpha}{\langle\alpha,\alpha\rangle}$ is $\frac{2}{\|\alpha\|}$, so if α is longer than β , then the coroot corresponding to α is shorter than the coroot corresponding to β . So the coroot Dynkin diagram is the same as the root Dynkin diagram except that arrows are reversed when they occur.

Therefore the diagrams are the same for types A_n, D_n, E_6, E_7, E_8 , and the diagrams are isomorphic for G_2 and F_4 , but the coroot diagrams for B_n and C_n are the root diagrams for C_n and B_n .

Remark: Suppose we multiply the inner product on $\mathcal{T}$ by a constant and thus replace each root α by $\lambda\alpha$. The procedure described in this section still allows us to find the weight lattice. Indeed each corresponding coroot $\frac{2\alpha}{\langle\alpha,\alpha\rangle}$ will be multiplied by $\frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$. Suppose the inner product of $v \in V$ with these renormalized coroots is always an integer. Then for every one of the original coroots γ , $\langle v, \frac{\gamma}{\lambda} \rangle \in \mathbb{Z}$, so $\langle \frac{v}{\lambda}, \gamma \rangle \in \mathbb{Z}$, and thus $\frac{v}{\lambda}$ is a weight for the correctly normalized coroots and correctly normalized roots, and $\lambda(\frac{v}{\lambda}) = v$ represents a weight in the world where all roots and weights are multiplied by the same λ .

15.3 Lattices for $su(2)$ and $su(3)$

The Case $SU(2)$: The maximal torus of $SU(2)$ is one dimensional, so $\mathcal{T}$ is just a line. There are only two roots, $\pm\alpha$, where α is on the positive axis. The root lattice contains all integer multiples of this α .

The weight lattice contains vectors w such that $\frac{2\langle w, \alpha \rangle}{\langle \alpha, \alpha \rangle}$ is an integer. Notice that α belongs to this lattice since if we set $w = \alpha$, we get 2. But then $w = \frac{\alpha}{2}$ is in the weight lattice and the weight lattice contains all integer multiples of this w .

The root and weight lattices are pictured below. The solid black dots mark the root lattice vectors; we must add to these the small squares to get the weight lattice.

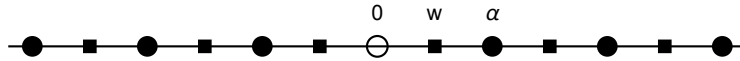


Figure 15.1: Root and Weight Lattices for A_1

The Case $SU(3)$: Earlier we discovered that the Lie algebra of the maximal torus for $SU(3)$ consists of purely imaginary diagonal matrices with trace 0. The positive roots have the form $f_1(t_1, t_2, t_3) = \frac{1}{\pi}(t_2 - t_1)$, $f_2(t_1, t_2, t_3) = \frac{1}{\pi}(t_3 - t_2)$, and $f_3(t_1, t_2, t_3) = \frac{1}{\pi}(t_3 - t_1)$. Note that $f_3 = f_1 + f_2$.

We also showed that the Killing form on matrices A, B is $\text{tr}(AB)$. In particular, for purely imaginary diagonal matrices, this is the negative of the standard inner product on R^3 , so this standard inner product is the invariant metric up to a constant factor. Since the Lie algebra has matrices with trace zero, we can write

$$\mathcal{T} = \{(t_1, t_2, t_3) \mid t_1 + t_2 + t_3 = 0\}$$

Consider the vectors $v_1 = (-1, 1, 0)$ and $v_2 = (-1, -1, 2)$. They form a basis of $\mathcal{T}$. We use Gram-Schmidt to get an orthonormal basis.

$$e_1 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$$

$$e_2 = \left(-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)$$

Note that $f_1(e_1) = \frac{\sqrt{2}}{\pi}$ and $f_1(e_2) = 0$. Notice that $f_2(e_1) = -\frac{1}{\pi\sqrt{2}}$ and $f_2(e_2) = \frac{3}{\pi\sqrt{6}} = \frac{\sqrt{3}}{\pi\sqrt{2}}$. The invariant inner product is only determined up to a positive constant, so let us multiply f_1 and f_2 by $\frac{\pi}{\sqrt{2}}$. Then $f_1(e_1) = 1$ and $f_2(e_2) = 0$, so $f_1 = e_1^*$. Also $f_2(e_1) = -\frac{1}{2}$ and $f_2(e_2) = \frac{\sqrt{3}}{2}$. Thus $f_2 = -\frac{1}{2}e_1^* + \frac{\sqrt{3}}{2}e_2^*$. Under our identification of $\mathcal{T}^*$ with $\mathcal{T}$, the roots f_1 and f_2 become the roots $\alpha = (1, 0)$ and $\beta = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$.

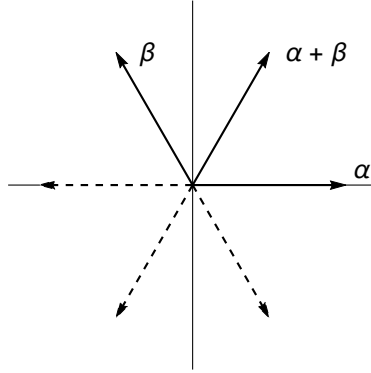
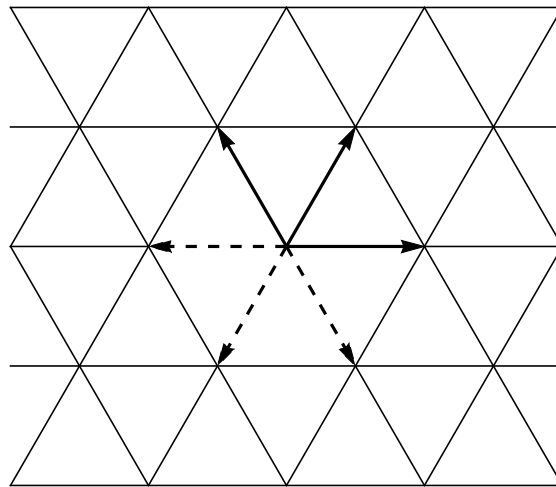
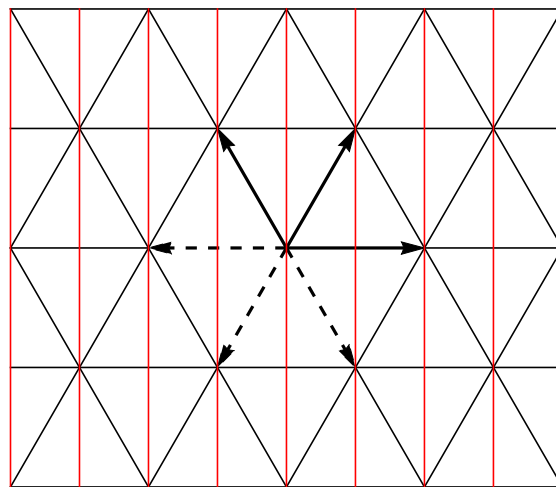


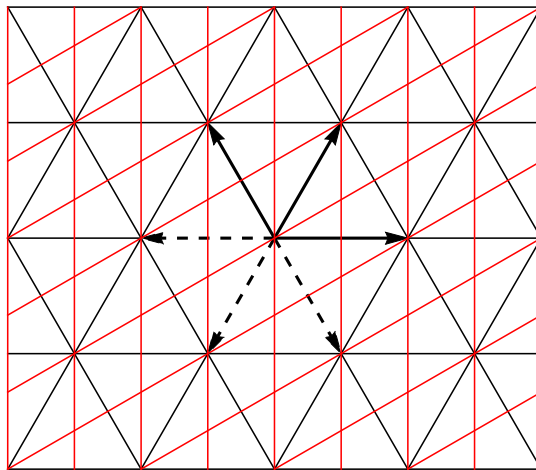
Figure 15.2: Roots for A_2

These roots clearly generate the hexagonal lattice L_R in $\mathcal{T}$. Notice that the simple roots form a lattice basis for this lattice.

Figure 15.3: Root Lattice for A_2

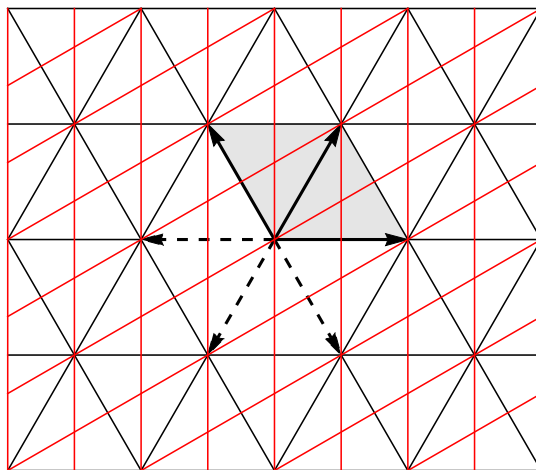
These roots have length 1. So the coroot lattice replaces each α by $\frac{2\alpha}{\langle \alpha, \alpha \rangle} = 2\alpha$ and the weight lattice consists of all vectors which have integer inner products with all such 2α . By our earlier theory, it suffices to check this for the simple roots, and thus for α and β in the diagram. We begin by drawing red lines of points whose dot product with 2α is an integer. After that, we add red lines of points whose dot product with β is an integer. Intersection points of these red lines are elements of the weight lattice. Notice that this lattice contains the root lattice.

Figure 15.4: Dot Product with 2α Is Integer

Figure 15.5: Dot Product with 2α or 2β Is Integer

Remark: It is useful to mention a theorem which will be proved later. It turns out that L_W/L_R is isomorphic to the center of the simply-connected compact Lie group G with the given root system. Recall that all compact groups with universal cover G can be obtained as quotients G/H where H is a subgroup of this center. Thus the root and weight diagrams contain an implicit classification of all Lie groups with the given Lie algebra.

Consider, for instance, $SU(3)$. In the figure below, the fundamental region of the root lattice generated by the simple roots is shaded.

Figure 15.6: Fundamental Domain of L_R

The fundamental domain contains two weight vectors in its interior and four more on its boundary. But the four boundary vectors are equivalent, so L_W/L_R must be isomorphic to Z_3 . And indeed, the center of $SU(3)$ consists of constant diagonal matrices λI . Since the determinant must equal 1, $\lambda^3 = 1$ and λ is a complex third root of unity. These define Z_3 .

Consider next the shaded region in the diagram below. This shaded region is called the *fundamental Weyl chamber* of the root diagram. Notice that its walls are the hyperplanes perpendicular to the fundamental roots of the diagram. Let α and β be the fundamental roots. Concentrate first on the vertical wall. This line contains points whose inner product with 2α is zero, and the vertical lines to the right of this line contain points whose inner products with 2α are positive integers. Next look at the diagonal wall. This line contains points whose inner product with 2β is zero, and the lines parallel to and above this line contain points whose inner products with 2β are positive integers. Since the fundamental Weyl chamber contains all points right of the vertical line and above the diagonal line, it contains exactly all weight lattice points whose inner products with both 2α and 2β are non-negative integers. We proved that every irreducible representation of our simply connected G is associated with a unique highest weight in this Weyl chamber. (We have yet to prove that every weight in the chamber comes from such a representation.)

Incidentally, the weight lattice is not a lattice directly associated to a root system, so it has features that don't fit earlier patterns in these notes. For instance, the canonical generators of this lattice meet in an acute angle rather than an obtuse angle.

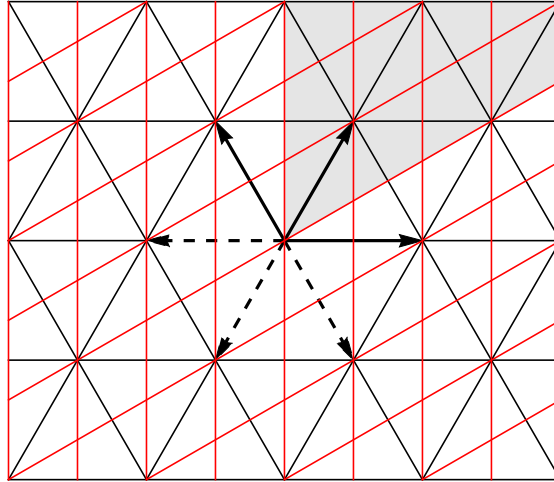


Figure 15.7: Weyl Chamber

Remark: We now come to the connection between this theory and kaleidoscopes, which will soon turn out to be a crucial part of the story.

If α is a root in an arbitrary root diagram, and if β is a second root, we proved that $\beta - 2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha$ is again a root. More generally, consider the map from $\mathcal{T}$ to $\mathcal{T}$ given by

$$v \rightarrow v - 2\frac{\langle v, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha$$

Notice that if v is in the hyperplane perpendicular to α , then v maps to v . But if $v = \lambda\alpha$ for some real λ , then v maps to $v - 2\lambda\alpha = -v$. Hence this map is just reflection across the hyperplane perpendicular to α .

It follows that the set of all products of reflections associated with roots preserves the full set of roots. Therefore, the set of all such products is a finite group. We are going to call this group the *Weyl group* because we will later prove that it is canonically isomorphic to the Weyl group introduced earlier in these notes. But for the moment, we only study this group from the new point of view.

In particular, the Weyl group is a finite group of isometries of $\mathcal{T}$ generated by reflections. In the 1930's, the Canadian geometer Coxeter classified all finite groups of isometries of R^n generated by reflections. It turns out that, with only a small number of exceptions, all arise in this way from the Lie theory. The most importance of these exceptions are the dihedral groups D_n of symmetries of an n -sided regular polygon. All are generated by reflections. But only D_1, D_2, D_3, D_4 , and D_6 come from the Lie theory. The remaining D_n do not preserve any lattice, while the Weyl groups preserve both L_R and L_W .

We will prove that every Weyl group is generated by the reflections across simple roots. The walls of the Weyl chamber are these reflection hyperplanes. The images of the fundamental Weyl chamber under these reflections and their products are other regions called *Weyl chambers*. We will prove that the Weyl group is simply transitive on Weyl chambers, i.e., there is a one-to-one correspondence between these chambers and the elements of the Weyl group.

All of this is closely related to the kaleidoscopes often used as toys by children. Such a kaleidoscope is shown on the next page. Notice that the key element inside is a bent piece of metal that provides two mirrors. The pattern seen by looking in this kaleidoscope is the pattern in the region between these mirrors, and reflected images of this region that appear to fill out an entire regular polygon.



Figure 15.8: Kaleidoscopes



Figure 15.9: Kaleidoscopes

15.4 All Root Lattices, Weight Lattices, and Fundamental Regions in Dimension Two

Roots $SU(2) \times SU(2)$:

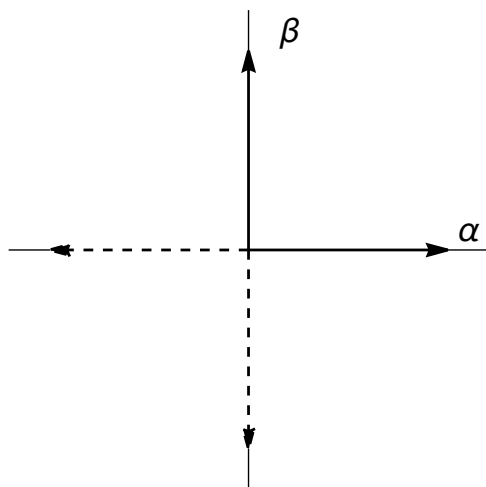


Figure 15.10: $A_1 \times A_1$

Roots $SU(3)$:

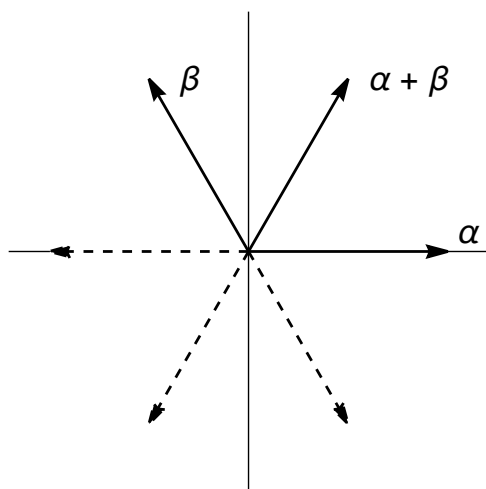


Figure 15.11: A_2

Roots $SO(5)$

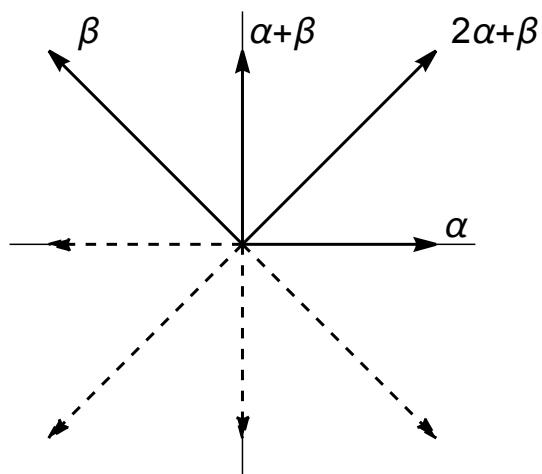


Figure 15.12: B_2

Roots G_2

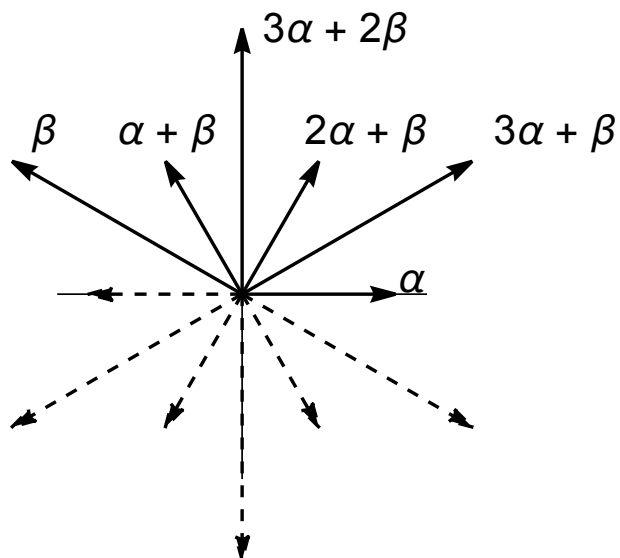


Figure 15.13: G_2

Lattices $SU(2) \times SU(2)$:

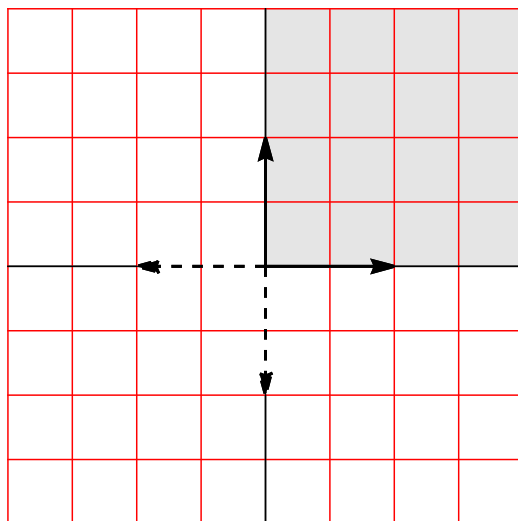


Figure 15.14: $A_1 \times A_1$

Lattices $SU(3)$:

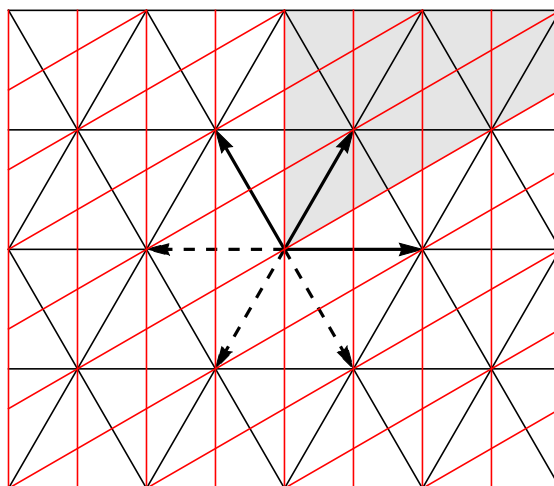


Figure 15.15: A_2

Lattices $SO(5)$:

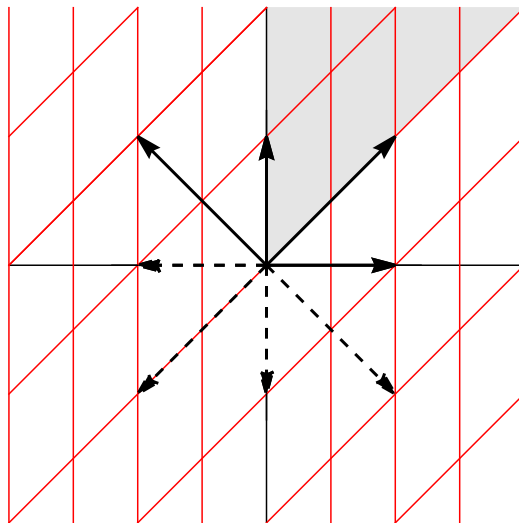


Figure 15.16: B_2

Lattices G_2 :

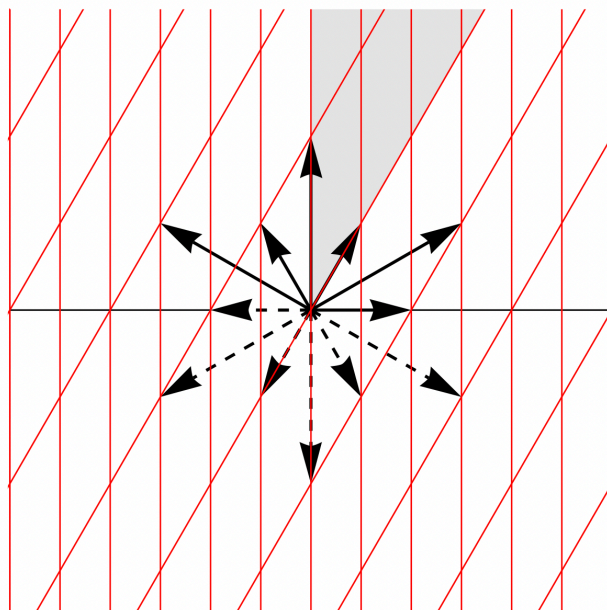


Figure 15.17: G_2

15.5 Weyl Groups as Reflection Groups

We now have two definitions of the Weyl group: as a group generated by reflections, and as the group $N(T)/T$ which determines the conjugacy classes of G in T . Both definitions lead to important theorems required by our later theory. In the rest of this chapter, we prove those results which follow from the definition as a group generated by reflections. In the following chapter, we prove those results which depend on the conjugacy class definition.

If α is a root, let σ_α be the associated reflection, $\sigma_\alpha(v) = v - \frac{2\langle v, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha$.

Theorem 74 *Let $\alpha_1, \dots, \alpha_n$ be the simple roots. The Weyl group is generated by the σ_{α_i} .*

Proof:

Lemma 30 *If α is simple, then σ_α changes the sign of α and maps all other positive roots to positive roots.*

Proof: Let β be a positive root not equal to α , and write $\beta = m_1\alpha + m_2\alpha_2 + \dots + m_n\alpha_n$, where α and the α_i are simple. Then $m_i \geq 0$ are integers; we can assume that $m_2 > 0$. We have

$$\sigma_\alpha(\beta) = \beta - \frac{2\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle} \alpha = \left(m_1 - \frac{2\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle}\right) \alpha + m_2\alpha_2 + \dots + m_n\alpha_n$$

Since $m_2 > 0$, all the coefficients are positive and $\sigma_\alpha(\beta)$ is positive.

Lemma 31 *If β is a root, there is a simple root α which is carried to β by a sequence of σ_{α_i} associated with simple roots.*

Proof: If a sequence of reflections maps α to β , that same sequence maps $-\alpha$ to $-\beta$. Therefore may assume β positive, for if not, find such an α and a sequence mapping α to $-\beta$ and then first map α to $-\alpha$ by σ_α and then apply the sequence.

Every positive β has the form $\sum m_i\alpha_i$ with m_i non-negative integers. Call $\sum m_i$ the *height* of β . We prove by induction on height that a sequence of reflections across simple roots maps β to some simple root. Inverting the sequence then proves the lemma.

If the height is 1 then β is simple and we are done.

Working by induction, we can suppose β is positive with height at least two. In particular, β is not simple.

We claim that we can find a simple root α such that $\langle \beta, \alpha \rangle > 0$. Otherwise we would have $\langle \beta, \beta \rangle = \sum m_i \langle \beta, \alpha_i \rangle \leq 0$, which is impossible because $\beta \neq 0$.

By the previous lemma, $\sigma_\alpha(\beta)$ is a positive root, since $\beta \neq \alpha$. This root equals

$$\beta - \frac{2\langle\beta, \alpha\rangle}{\langle\alpha, \alpha\rangle}\alpha$$

The coefficient $-\frac{2\langle\beta, \alpha\rangle}{\langle\alpha, \alpha\rangle}$ is negative, and thus this element has smaller height than the height of β . So by induction, a sequence of σ_{α_i} for α_i simple map $\sigma_\alpha(\beta)$ to some α_j . The inductive step is an immediate consequence.

Corollary 10 *If the Dynkin diagram of a root system is connected and has no multiple links, then all roots have the same length. If the Dynkin diagram of a root system is connected and has a multiple link, then there are exactly two possible lengths of roots.*

Proof: According to the Dynkin diagram, all simple roots have the same length if there are no multiple links, and otherwise simple roots have one of two possible lengths. The corollary follows since every root is an image of a simple root under the Weyl group.

Lemma 32 *If w is in the Weyl group and β is a root, $w\sigma_\beta w^{-1} = \sigma_{w(\beta)}$.*

Proof: The map $w\sigma_\beta w^{-1}$ maps $w(\beta)$ to $-w(\beta)$ and maps an element perpendicular to $w(\beta)$ to itself.

Proof of main theorem: Let β be an arbitrary root. Find a simple root α and a chain of simple reflections w such that $\beta = w(\alpha)$. Then $w^{-1}\beta = \alpha$ and by the previous lemma

$$w^{-1}\sigma_\beta w = \sigma_{w^{-1}(\beta)} = \sigma_\alpha$$

So

$$\sigma_\beta = w\sigma_\alpha w^{-1}$$

Hence reflection across any root equals a product of reflections across simple roots, proving the theorem. QED.

Remark: Every element in the Weyl group W is thus a product of reflections across simple roots. The *length* of an element of W is the shortest such expression. (This shortest such expression need not be unique.) There is a very pretty theorem about the length of an element w :

Theorem 75 *The length of $w \in W$ is the number of positive roots which are mapped to negative roots by w .*

Proof: We proved that if α is simple, then σ_α maps exactly one positive root to a negative root. It follows that a product of k simple reflections can map at most k positive roots to negative roots. To finish the proof, we assume that $w = \sigma_0\sigma_1\ldots\sigma_k$ and yet maps fewer than $k+1$ positive roots to negative roots, and then we prove that the expression is not as short as possible. We will work by induction on k .

We can thus assume that $\sigma_1\sigma_2\ldots\sigma_k$ is optimal and maps k positive roots to negative roots. The final map, $\sigma_0 = \sigma_\alpha$, must have mapped some negative root back to a positive root, and that negative root would have to be $-\alpha$.

We conclude the last k terms must have started with a positive β which became $-\alpha$ in the end, and that it must have switched sign just once. So the “word” $\sigma_1\ldots\sigma_k$ is a product $g\sigma_i h$ where all terms in h kept images of β positive, and eventually it mapped to α_i and then $\sigma_i = \sigma_{\alpha_i}$ made it negative, and then g mapped this negative element to other negative elements and eventually to $-\alpha$. So

$$\sigma_1\ldots\sigma_k(\beta) = g\sigma_i h(\beta) = g\sigma_i(\alpha_i) = g(-\alpha_i) = -\alpha$$

We claim that consequently $w = \sigma_\alpha g\sigma_i h$ can be rewritten $w = gh$, and thus was not optimal in its original form. To prove this, we need only show that $\sigma_\alpha g\sigma_i = g$ or $\sigma_\alpha g\sigma_i g^{-1} = e$. But $g\sigma_i g^{-1} = \sigma_{g(\alpha_i)} = \sigma_\alpha$, QED.

Theorem 76 *Let Δ denote the unordered set of simple roots. The only element of the Weyl group which maps Δ to itself is the identity.*

Proof: If w preserves Δ as a set, then w maps every positive root $\sum m_i \alpha_i$ to a positive root $\sum m_i w(\alpha_i)$ and thus the reduced word for w has length zero and $w = e$.

15.6 Weyl Chambers

Definition 32 *The reflection planes perpendicular to the roots of a root system separate the open complement of these planes in $\mathcal{T}$ into finitely many open, convex, arcwise connected components called Weyl chambers. The fundamental Weyl chamber is the chamber of all $t \in \mathcal{T}$ such that $\langle \alpha_i, t \rangle > 0$ for all simple roots α_i .*

Remark: Earlier we stated that there is a one-to-one correspondence between weight vectors in the fundamental Weyl chamber and irreducible representations of $\mathcal{G}$ (the existence portion of this claim has not yet been proved). Now that we have the precise definition of this Weyl chamber before us, we see that the correct statement is “there is a one-to-one correspondence between weight vectors in the *closure* of the fundamental Weyl chamber and irreducible representations of $\mathcal{G}$.” Indeed, highest weight vectors satisfy $\langle v, \alpha_i \rangle \geq 0$ for the simple roots α_i .

Theorem 77

- *The fundamental Weyl chamber, as defined above, is a Weyl chamber; it is non-empty, open, connected, contained in the complement of the hyperplanes, and not contained in a larger connected open set in this complement*
- *The Weyl group acts transitively on Weyl chambers*
- *Every Weyl chamber is the fundamental Weyl chamber for exactly one unordered set Δ of simple roots*
- *The Weyl group acts simply transitively on Weyl chambers and on unordered sets Δ of simple roots*
- *The number of elements of the Weyl group is exactly the number of Weyl chambers, and exactly the number of simple root systems.*

Remark: This theorem might be called “the fundamental theorem of kaleidoscopes.” According to the theorem, a kaleidoscope created by mirrors perpendicular to the simple roots of a root system will show repeated copies of the pattern that exists in the fundamental region. These copies will fit together and fill all space. Or if the patterns only extend a bounded distance from the origin, their reflections will fill a regular polygon. This works in any dimension. For instance, it is easy to make cone-like objects in R^3 bounded by three inwardly pointing mirrors. Looking into the opening of the cone reveals many copies of your face. If the mirrors are selected correctly, these copies will form the outside of a tetrahedron, cube, octahedron, dodecahedron or icosahedron. The pattern will be repeated $|W|$ times, completely filling space with a pattern symmetric under W .

Proof: Let $\alpha_1, \dots, \alpha_n$ be the simple roots. These form a basis for $\mathcal{T}$. The resulting dual basis $\alpha_1^*, \dots, \alpha_n^*$ lives in the dual space of $\mathcal{T}$, which is canonically isomorphic to $\mathcal{T}$ using the invariant inner product. Let $t \in \mathcal{T}$ be the image of $\sum \alpha_j^*$ under this isomorphism. Then

$$\langle t, \alpha_i \rangle = \left(\sum \alpha_j^* \right) (\alpha_i) = 1 > 0$$

for all i , so t belongs to the fundamental Weyl chamber, which is therefore non-empty.

This chamber is certainly open. If p, q belong to the chamber, then for all t satisfying $0 \leq t \leq 1$ we have

$$\langle tp + (1-t)q, \alpha_i \rangle = t \langle p, \alpha_i \rangle + (1-t) \langle q, \alpha_i \rangle$$

and this is the formula for a parameterized interval in R connecting two positive numbers, so it is always positive. Therefore, the fundamental chamber is convex and so arcwise connected.

The fundamental chamber does not intersect any hyperplane perpendicular to a root α . Indeed, such a root is either positive or negative and the argument is the same for both

cases. For instance, if α is positive then $\alpha = \sum m_i \alpha_i$ where $m_i \geq 0$ and some $m_i \neq 0$ and then when t is in the fundamental chamber, $\langle t, \alpha \rangle = \sum m_i \langle t, \alpha_i \rangle > 0$.

If a hyperplane is removed from V , the complement has two open sets, both of which are convex. Since the intersection of convex sets is convex, any Weyl chamber is open and convex, hence arcwise connected.

Returning to the fundamental Weyl chamber, suppose it is contained in a larger connected open set in the complement after the various hyperplanes are removed. This set would be arcwise connected, so we could find a continuous curve starting at a point p in the fundamental Weyl chamber and ending at a point q the larger open set. Let $\gamma(t)$ be this curve. Then $\langle \gamma(t), \alpha_i \rangle$ is never zero because we are in the complement of the hyperplanes, and $\langle \gamma(t), \alpha_i \rangle$ is positive at p and thus positive at q . This holds for all i , so q is also in the fundamental Weyl chamber.

Proof, continued: Now we prove that the Weyl group acts transitively on Weyl chambers. Notice that a reflection hyperplane has codimension 1 and thus splits $\mathcal{T}$ into two pieces. On the other hand, intersections of two hyperplanes have codimension 2, and removing a finite number of such intersections leaves an open connected complement.

Thus if C_1 and C_2 are Weyl chambers, we can find a continuous and differentiable path $\gamma(u)$ from $c_1 \in C_1$ to $c_2 \in C_2$ which may hit reflection hyperplanes, but misses all intersections of two or more such hyperplanes. We can assume that this path hits hyperplanes at only a finite number of u_i , and actually crosses from one side to the other at these u_i . Each crossing corresponds to a reflection σ_{u_i} . Our path starts at C_1 , crosses to $\sigma_{u_1}C_1$, and then to $\sigma_{u_2}\sigma_{u_1}C_1$ and eventually to $C_2 = \sigma_{u_k} \dots \sigma_{u_2}\sigma_{u_1}C_1$. QED.

Proof, concluded: Since the Weyl group acts transitively on Weyl chambers, and acts effectively on sets Δ (that is, at most one element of the Weyl group maps one Δ to another), the remaining items all follow from the assertion that there is a one-to-one correspondence between Weyl chambers and sets Δ of simple roots. This is easy. Given a set of simple roots Δ , map this set to the corresponding fundamental Weyl chamber of t such that $\langle t, \alpha_i \rangle > 0$ for roots $\alpha_i \in \Delta$. Conversely, given a Weyl chamber, C_1 , map this to the set Δ_1 of roots α_i whose reflections form the boundary of C_1 .

To complete the argument, we need only show that this is a set of simple roots for *some* choice of positive roots. But we started with a definition of positivity and a resulting set Δ of simple roots. We have proved that $C_1 = w(C)$ where C is the fundamental Weyl chamber corresponding to Δ . We know that w is an isometry, so it takes the dividing line between positive roots as originally defined to a new dividing line, and takes the original simple roots α_i to new simple roots $w(\alpha_i)$ and takes the original positive roots $\sum m_i \alpha_i$ to new positive roots $\sum m_i w(\alpha)_i$. QED.

Chapter 16

Conjugacy and the Weyl Group

16.1 The Main Theorem

We initially defined the Weyl group to be $N(T)/T$. Here the group $N(T)$ consists of those elements of G which map T back to T under conjugation. The group T is a subgroup and $W = N(T)/T$ acts effectively, so only the identity leaves all points of T fixed. We proved that every element of G is conjugate to an element of T , and that two such elements are conjugate in G if and only if they are conjugate under W . So the set T/W equals the set of conjugacy classes of G .

In the previous chapter, we defined W to be the group generated by reflections across the planes perpendicular to the roots. Using this definition, we proved many important properties of W .

In both cases, the Weyl group was defined as a group of transformations on a specific set. The original definition involved maps from T to T . The second definition involved maps from $\mathcal{T}$ to $\mathcal{T}$. But there is an obvious map from the first set to the second set, since each element of the first type of Weyl group is an automorphism $\psi : T \rightarrow T$ and thus induces a map $\psi^* : \mathcal{T} \rightarrow \mathcal{T}$.

Theorem 78 *If G is connected, the above map sends elements of the first type of Weyl group to elements of the second type of Weyl group, and is an isomorphism between these groups.*

16.2 Reflections in $N(T)/T$

We begin the proof of the main theorem. In the middle of the proof, we require a little lemma which seems totally irrelevant at the moment. This lemma appears again in the

next section. We'll get it out of the way first.

Lemma 33 *Let H be a commutative Lie group, T_1 a torus in H , and suppose H/T_1 is isomorphic to Z_m . Then there is an element $h \in H$ such that the closure of the set of powers of h is all of H .*

Proof of lemma: Since these groups are abelian, we write them additively. Let n be a generator of H/T_1 . Then $m n \in T_1$, say t . But t/m is in T_1 and $m(n - \frac{t}{m}) = mn - t = 0$. So without loss of generality, we can assume that $m n = 0$.

Since T_1 is a torus, we can find $t_1 \in T_1$ such that the multiples of t_1 are dense in T_1 . Select $t_2 \in T$ such that $m t_2 = t_1$. Let $h = t_2 + n$. Then $m h = m t_2 + m n = t_1$. So the closure of the set of multiples of h contains both T_1 and n and thus equals all of H . QED.

Proof of theorem: Let α be a root with complex root vector $u + iv$. Then $\{u, v, t_\alpha\}$ generate a subalgebra of $\mathcal{G}$ isomorphic to $su(2)$. By the fundamental theorem of Lie theory, there is a corresponding Lie subgroup of G . This Lie subgroup must be isomorphic to either $SU(2)$ or $SO(3) = SU(2)/\pm 1$. We are interested in the action of this subgroup on G by conjugation, and on the induced action on $\mathcal{G}$. In our case, the $\pm I$ act trivially, so we can work with $SU(2)$ without loss of generality.

The Lie algebra $su(2)$ consists of all matrices

$$\begin{pmatrix} it & \alpha + i\beta \\ -\alpha + i\beta & -it \end{pmatrix}$$

In particular, $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is in the Lie algebra and has exponent $\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$. Let this element act on $SU(2)$ by conjugation. Then Ad of it acts on $su(2)$ by conjugation, giving

$$\begin{aligned} & \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} it & \alpha + i\beta \\ -\alpha + i\beta & -it \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \\ &= \cos(2\theta) \begin{pmatrix} it & i\beta \\ i\beta & -it \end{pmatrix} + \sin(2\theta) \begin{pmatrix} i\beta & -it \\ -it & -i\beta \end{pmatrix} + \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix} \end{aligned}$$

Several features of this result deserve comments. First, the conjugating rotation commutes with $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, so the final result does not modify the α term. Second, the final rotation is by 2θ , so $\theta = \pi$ corresponds to conjugating by $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, which commutes with everything, and thus the induced conjugation is the identity when $\theta = \pi$. This is another indication that it does not matter if the subgroup is $SU(2)$ or $SO(3)$.

This conjugation does not preserve the one dimensional maximal torus for $SU(2)$, since the Lie algebra parameter t is rotated to the $t - \beta$ plane. However, there is one special case

when the t component is preserved, and that is when $\theta = \frac{\pi}{2}$. In that case, $\begin{pmatrix} it & 0 \\ 0 & -it \end{pmatrix}$ is mapped to $\begin{pmatrix} -it & 0 \\ 0 & it \end{pmatrix}$.

So far, we have spoken of $SU(2)$ acting by conjugation on itself. But of course it acts by conjugation on all of G , and thus Ad acts on all of $\mathcal{G}$. Let us determine what this action does to an element of T that is perpendicular to α . Recall that

$$[t, u] = -2\pi f(t)v$$

$$[t, v] = 2\pi f(t)u$$

where $f(t) = \langle \alpha, t \rangle$. Thus if t is perpendicular to α we have $[u, t] = 0$ and $[v, t] = 0$. And since $\alpha \in T$, we also have $[\alpha, t] = 0$. We conclude that the entire Lie algebra $su(2)$ acts trivially on t . This action is induced by $SU(2)$ acting on G by conjugation. So this action fixes our $t \in \mathcal{T}$.

To summarize, the rotation matrices $\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$ form a one-parameter subgroup of $SU(2)$, which maps to a subgroup of G . This one parameter group acts on G by conjugation, and so it acts in $\mathcal{G}$. The subspace $\mathcal{T}$ rotates around one axis under this action, while all other axes are fixed, so the maximal torus T is rotated to other maximal tori. But when $\theta = \frac{\pi}{2}$, the maximal torus T is mapped to itself and thus $\mathcal{T}$ is mapped to itself. This particular map is a reflection across the hyperplane perpendicular to α .

Our conclusion is that the conjugation map above when $\theta = \frac{\pi}{2}$ preserves T and thus belongs to $N(T)$, and induces reflection across α . Thus the image of the Weyl group acting on T contains all reflections across roots, and thus all maps in the group generated by these reflections. This proves half of the theorem.

16.3 $N(T)/T$ Is Simply Transitive on Weyl Chambers

We continue the proof. Each element ψ of the Weyl group on T acts by conjugation and thus preserves the invariant metric. Consequently ψ^* is an isometry from $\mathcal{T}$ to itself. This map clearly maps roots to roots. Consequently, it maps the hyperplane perpendicular to α to the hyperplane perpendicular to $\psi^*(\alpha)$. If $\mathcal{C}$ is an open Weyl chamber, it is open and connected and does not intersect any root hyperplane, so $\psi^*(\mathcal{C})$ is open and connected and does not intersect any root hyperplane and thus must be inside some other Weyl chamber. Applying this reasoning to the inverse of ψ^* , we find that each ψ^* permutes the open Weyl chambers.

We know that the Weyl group generated by reflections is simply transitive on Weyl chambers. So to complete the proof, it suffices to show that the image of the Weyl group acting

on T is also simply transitive. Since reflections are in the image, it suffices to prove that an element of $N(T)/T$ which maps the fundamental Weyl chamber back to itself must be the identity.

So suppose ψ is in $N(T)/T$ and $\psi^*(\mathcal{C}) = \mathcal{C}$ where $\mathcal{C}$ is the fundamental Weyl chamber. Since $N(T)/T$ is finite, $\psi^k = 1$ for some k . Fix $c \in \mathcal{C}$ and form $\bar{c} = \frac{1}{k} \sum (\psi^*)^i(c)$. Each element of the sum is in $\mathcal{C}$, so the sum is in $\mathcal{C}$ because $\mathcal{C}$ is convex. Clearly ψ^* fixes $\bar{c}$. Since ψ^* is linear, it fixes the entire line joining the origin to c . It is enough to prove that such a ψ is the identity map on T .

Definition 33 *A point in $\mathcal{T}$ is called a singular point if it covers a point in T which belongs to at least two distinct maximal tori.*

Lemma 34 *Suppose $\psi \neq 1$ is in the Weyl group on T . If ψ^* fixes $X \in \mathcal{T}$ and $X \neq 0$, then it fixes every point on the line containing 0 and X , and every point on this line is a singular point. Therefore X cannot be in any open Weyl chamber.*

Proof: Each ψ^* is linear, so if it fixes a point X , then it fixes the entire line through X . Suppose $n \in N(T)$ is not in T and suppose $n^*(X) = X$ for some nonzero $X \in \mathcal{T}$. Let K be the one-parameter subgroup through X , i.e., the line through both the origin and X , and notice that n fixes this entire line. Thus the subgroup of G generated by n and K is abelian; let H be the closure of this subgroup and let H_0 be the connected component of the identity in H . Then H_0 is a torus containing K and n generates H/H_0 . Since H is also compact, n has finite order m in H/H_0 .

We can apply the lemma at the start of the proof to $H_0 \subset H$. According to the lemma, there is an element $h \in H$ whose powers generate a subgroup of H with closure H . Since all of this action occurs in G , h belongs to some maximal torus $\hat{T}$ of G . This torus contains n , and it contains K and in particular the line through X . But we are assuming that $\psi \neq I$ and so $n \notin T$. It follows that the line from the origin to X is in the image of points in T which are contained in two distinct maximal tori, so all elements of the line are singular.

The final statement of the proof depends on a detailed study of singular points in $\mathcal{T}$. It will be clear from these results that if an entire line from the origin to a nonzero X only contains singular points, then X cannot be in the open fundamental Weyl chamber.

But the study of these singular points leads to important extra matters and deserves a separate chapter.

Chapter 17

Stiefel Diagrams and Torus Lattices

17.1 Regular Elements in G

Definition 34 An element of $g \in G$ is regular if it is contained in a unique maximal torus. Otherwise the element is singular.

Lemma 35 Let G be connected and suppose $g \in G$. Let $N(g)$ be the set of all elements of G which commute with g . Then $N(g)$ is a closed subgroup of G . The identity component of this group, $N_0(g)$, is the union of all maximal tori containing g .

Proof: If T is a maximal torus containing g , then every element of T certainly commutes with g . Moreover, T is arcwise connected. So $T \subset N_0(g)$.

Conversely, suppose n and g commute and $n \in N_0(g) \subset N(g)$. Our job is to find a maximal torus of G containing n and g .

Clearly $N_0(g)$ is a closed subgroup of G , so it is compact and connected and thus n belongs to a maximal torus T_1 of $N_0(g)$. Let H be the closure of the subgroup of G generated by g and T_1 . This H is a compact abelian subgroup of G . So its identity component, H_0 , is a torus. Since T_1 is connected, it belongs to H_0 . So g must generate H/H_0 and thus must have finite order in this quotient group.

Apply lemma 33 of the previous section. This lemma states that we can find an element $t \in H$ such that the closure of the powers of t gives all of H . Find a maximal torus $\hat{T}$ of G containing t . This maximal torus contains g and T_1 . Since $n \in T_1$, it contains g and n . QED.

Lemma 36 Recall that the rank of G is the dimension of a maximal torus. Since all

maximal tori are conjugate, this definition does not depend on the choice of maximal torus. An element $g \in G$ is regular if and only if the dimension of $N_0(g)$ is rank G . In all other cases, the dimension of $N_0(g) > \text{rank } G$.

Proof: Select a maximal torus T with $g \in T$. Then $T \subset N(g)$ and so $T \subset N_0(g)$. Since $N_0(g)$ is the union of the maximal tori containing g , g is regular if and only if $T = N_0(g)$, if and only if $\dim T = \dim N_0(g)$. Otherwise the dimension of $N_0(g)$ is larger. QED.

Remark: We now come to the point of all of this. Select $t \in T$. Notice that $N_0(t)$ is a closed connected subgroup of G and its dimension equals the dimension of the corresponding Lie algebra $\mathcal{N} \subset \mathcal{G}$. The algebra $\mathcal{T} \subset \mathcal{N}$ because $\mathcal{T}$ is the Lie algebra of T and all of the elements of T commute with t . The dimension of $\mathcal{T}$ is the rank of G . Thus t is regular if and only if $\mathcal{N} = \mathcal{T}$.

Lemma 37 Fix $t \in T$. The Lie algebra $\mathcal{N}$ of $N_0(t)$ is the set of all $X \in \mathcal{G}$ such that $\text{Ad}(t)X = X$.

Proof:

Suppose that $gtg^{-1} = g$ where $g \in N_0(t)$. Since N_0 is connected, we can find a one-parameter group in $N_0(t)$ from the identity to g . Write this group as $\gamma(u)$ for $0 \leq u \leq 1$. Then $\gamma(u)t\gamma^{-1}(u) = t$ and so $\gamma(u)t = t\gamma(u)$. Each side of this formula is a path in G which starts at t , so if we differentiate with respect to u at $u = 0$, we get two tangent vectors to G at t . One of these is obtained by right translating $X = \gamma'(0)$ to t , and the other is formed by left translating X to t . It follows that if X is right translated to t , and then left translated back to the origin, the result is just X . But this map is just $\text{Ad}(t)(X)$ and so $\text{Ad}(t)(X) = X$.

Conversely, suppose $\text{Ad}(t)X = X$. Let $\gamma(u)$ be the one-parameter group through X . Then $t\gamma(u)t^{-1}$ is a one-parameter group and its derivative at the origin is $\text{Ad}(t)X = X$. So $t\gamma(u)t^{-1} = \gamma(u)$ for all u and thus $\gamma(u)$ and t commute for all u . Thus $\gamma(u) \in N_0(t)$ and so $X \in \mathcal{N}$. QED.

Lemma 38 Recall that T acts on $\mathcal{G}$ by conjugation. The subalgebra $\mathcal{N}$ is invariant under this representation.

Proof: Since T is abelian, $\text{Ad}(t_1)\text{Ad}(t) = \text{Ad}(t)\text{Ad}(t_1)$ for $t, t_1 \in T$. So if $\text{Ad}(t)(X) = X$, then $\text{Ad}(t)\text{Ad}(t_1)X = \text{Ad}(t_1)\text{Ad}(t)X = \text{Ad}(t_1)X$. QED.

Remark: Recall that under the action of T ,

$$\mathcal{G} = \mathcal{T} \oplus \sum_{\alpha > 0} \{u_\alpha, v_\alpha\}$$

The root spaces are not equivalent, so an invariant subspace containing $\mathcal{T}$ must have the

form

$$\mathcal{N} = \mathcal{T} \oplus \sum_{\text{selected } \alpha > 0} \{u_\alpha, v_\alpha\}$$

Remark: To complete this analysis, fix $t \in T$; we must determine whether $Ad(t)$ fixes u_α and v_α . But

$$Ad(t)(u + iv) = e^{2\pi i \langle \alpha, t \rangle} (u + iv)$$

Consequently, $\{u_\alpha, v_\alpha\}$ is invariant if and only if $\langle \alpha, t \rangle \in \mathbb{Z}$.

This proves

Theorem 79 *An element $t \in \mathcal{T}$ represents a singular element of T if and only if there exists at least one root α such that $\langle \alpha, t \rangle \in \mathbb{Z}$.*

Remark: We can now complete the proof that our two definitions of the Weyl group agree. Just before the end of that proof, we obtained a non-trivial line through the origin containing only singular points, and we claimed that our line could not intersect the open fundamental Weyl chamber. The singular points lie on hyperplanes of the form $\langle t, \alpha \rangle \in \mathbb{Z}$. Clearly these planes are bounded away from the origin unless they have the form $\langle t, \alpha \rangle = 0$, so if a line through the origin contains only singular points, then an open portion of the line near the origin must belong to a plane of the form $\langle t, \alpha \rangle = 0$, and therefore the entire line must belong to this hyperplane. But points in the fundamental region satisfy $\langle t, \alpha_i \rangle > 0$ for simple roots, and therefore $\langle t, \alpha \rangle > 0$ for all positive roots, and $\langle t, \alpha \rangle < 0$ for negative roots, and so cannot belong to hyperplanes through the origin. QED.

17.2 The Stiefel Diagram

Recall that we constructed the Weyl chambers in $\mathcal{T}$ by starting with simple roots $\alpha_1, \dots, \alpha_n$ for a root system, constructing the coroot system, and showing that $\frac{2\alpha_1}{\langle \alpha_1, \alpha_1 \rangle}, \dots, \frac{2\alpha_n}{\langle \alpha_n, \alpha_n \rangle}$ are the simple coroots for this system. We then drew the hyperplanes defined by

$$\left\langle t, \frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle} \right\rangle \in \mathbb{Z}$$

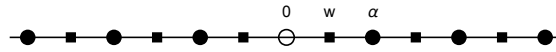
.

The *Stiefel Diagram* is a similar diagram in $\mathcal{T}$. This time we take *all* roots, ignore coroots, and draw the hyperplanes defined by

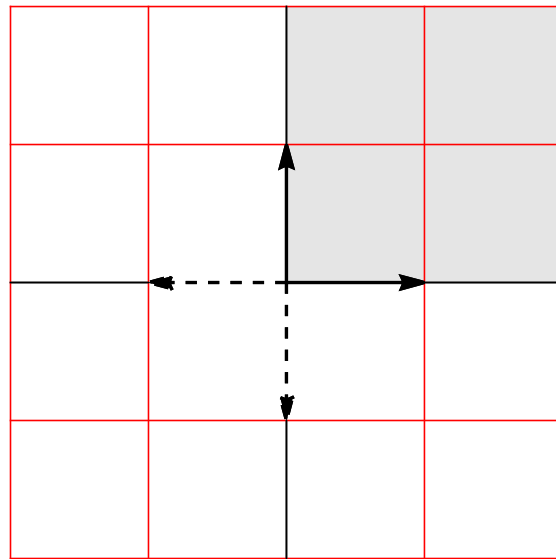
$$\langle t, \alpha \rangle \in \mathbb{Z}$$

A point represents a singular point in T if and only if it is on at least one of these hyperplanes.

$SU(2)$: The planes of the Stiefel diagram are perpendicular to the roots, and thus mere points. The Stiefel diagram is the set of large round points, which in this case is the root lattice. The weight lattice is also shown; it contains also the small square points.

Figure 17.1: Stiefel Diagram for A_1

$SU(2) \times SU(2)$:

Figure 17.2: Stiefel Diagram for $A_1 \oplus A_1$

$SU(3)$:

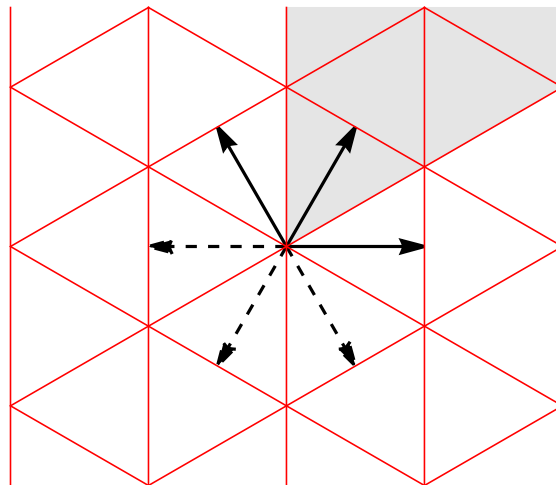


Figure 17.3: Stiefel Diagram for A_2

$SO(5)$:

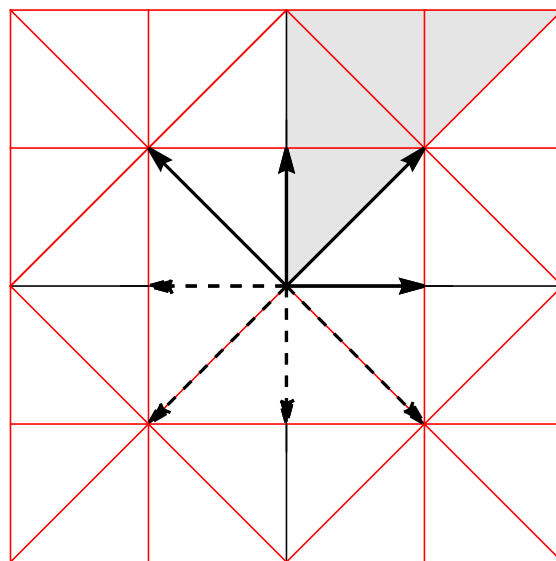


Figure 17.4: Stiefel Diagram for B_2

G_2 :

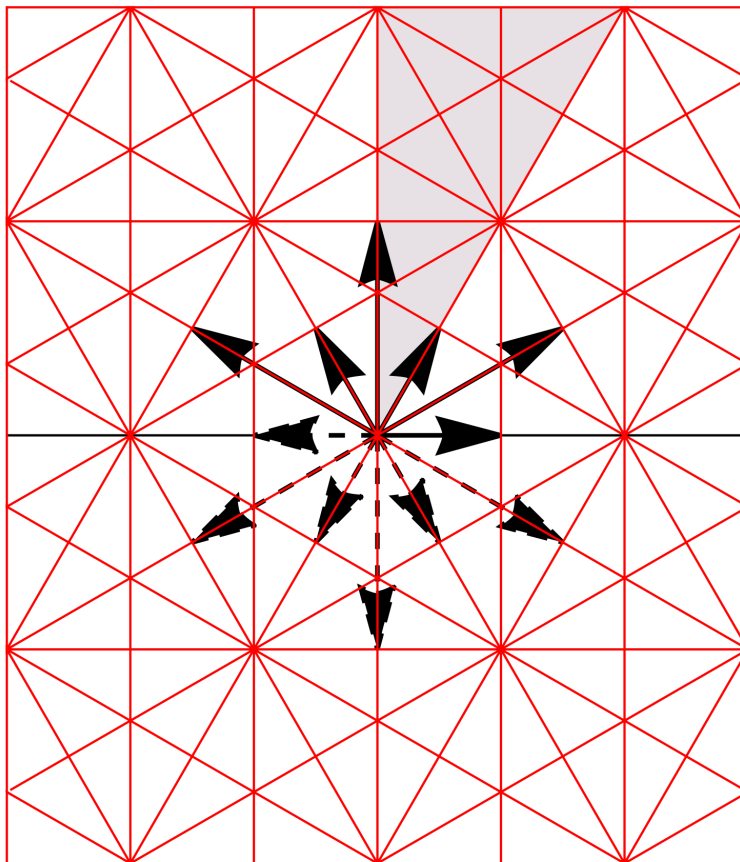


Figure 17.5: Stiefel Diagram for G_2

17.3 $Aut_0(\mathcal{G})$

Suppose G is a compact connected Lie group with a connected Dynkin diagram. The center C of G commutes with every element of G . This center is clearly closed, and thus a Lie subgroup. Its connected component C_0 would have a commutative Lie algebra $\mathcal{C} \subset \mathcal{G}$ whose bracket with everything is zero. Using the invariant metric on G we could write $\mathcal{G} = \mathcal{C} \oplus \mathcal{C}^\perp$. Our assumptions on G thus force C_0 to be trivial, and by compactness, C is finite.

It follows that G/C is locally isomorphic to G , has the same Lie algebra $\mathcal{G}$, and has trivial center. In this section we will deal with this group, so assume it is G from now on.

It is easy to identify the group: it is the connected component of the group of automorphisms of $\mathcal{G}$. Indeed, the automorphisms $\text{Aut}(\mathcal{G})$ of $\mathcal{G}$ form a subgroup of $GL(\mathcal{G})$ which is clearly closed in the full group and hence a Lie group. Each $g \in G$ induces an inner automorphism $g_1 \rightarrow g \circ g_1 \circ g^{-1}$ which induces an automorphism $\text{Ad}(g_1)$ of the Lie algebra of $\mathcal{G}$. In this way, we get a Lie homomorphism $G \rightarrow \text{Aut}(\mathcal{G})$. This map is one-to-one, for if $g \circ g_1 \circ g^{-1}$ induces the identity map on $\mathcal{G}$, then g is in the center of G and thus equals the identity. We claim that the map is onto the connected component of $\text{Aut}(\mathcal{G})$.

Indeed, if $\psi_t : \mathcal{G} \rightarrow \mathcal{G}$ is a parameterized path of automorphisms of $\mathcal{G}$ starting at the identity, then $\psi_t([X, Y]) = [\psi_t(X), \psi_t(Y)]$. Differentiating this equation at $t = 0$, and writing $\psi' = D$ gives $D([X, Y]) = [D(X), Y] + [X, D(Y)]$. Maps with this property are called *derivations*, and the Lie algebra of $\text{Aut}_0(\mathcal{G})$ is the set of these derivations in $\text{Hom}(\mathcal{G}, \mathcal{G})$. Inner automorphisms induce maps of the form $\text{ad}(X)$. To prove our claim, it suffices to prove

Lemma 39 *If $\mathcal{G}$ is the Lie algebra of a compact Lie group with no center, every derivation is inner.*

Proof: Let $\mathcal{D}$ be the set of derivations and $\mathcal{I}$ the set of all $\text{ad}(X)$. Then $\mathcal{I}$ is an ideal in $\mathcal{D}$ because

$$\begin{aligned} [\text{ad}(X), D](Y) &= \text{ad}(X)D(Y) - D([\text{ad}(X), Y]) = \\ &= [X, D(Y)] - [D(X), Y] - [X, D(Y)] = -\text{ad}(D(X))(Y). \end{aligned}$$

Forget that $\mathcal{D}$ is a set of maps and think of it as an abstract Lie algebra. Then it has a Killing form, K , which is probably degenerate. Computing that form would be a big mess, but luckily we don't need to do that.

Since $\mathcal{I}$ is an ideal in $\mathcal{D}$, if we restrict the Killing form on $\mathcal{D}$ to $\mathcal{I}$, we get the Killing form on $\mathcal{I}$. But $\text{ad} : \mathcal{G} \rightarrow \mathcal{I}$ is an isomorphism, and the Killing form on $\mathcal{G}$ is negative-definite. So the restriction of the Killing form on $\mathcal{D}$ to $\mathcal{I}$ is negative-definite. It is easy to see that

$$\mathcal{D} = \mathcal{I} + \mathcal{I}^\perp$$

We are not claiming that this is a direct sum. Indeed, let $i_1, \dots, i_m$ be an orthonormal basis of $\mathcal{I}$ with respect to the negative of the Killing form on $\mathcal{I}$. If $x \in \mathcal{D}$, write

$$X = \left(-\sum_j K(X, i_j) i_j \right) + \left(X + \sum_j K(X, i_j) i_j \right)$$

The first term is clearly in $\mathcal{I}$, and the second term is in $\mathcal{I}^\perp$ because

$$K(X + \sum_j K(X, i_j) i_j, i_k) = K(X, i_k) + \sum_j K(X, i_j) K(i_j, i_k) = 0$$

However, $\mathcal{I} \cap \mathcal{I}^\perp = \{0\}$ because the Killing form on $\mathcal{I}$ is negative definite. So the previous sum is a direct sum after all.

To complete the proof, we need only show that $\mathcal{I}^\perp = \{0\}$. Notice first that both $\mathcal{I}$ and $\mathcal{I}^\perp$ are ideals in $\mathcal{D}$. The first is an ideal by the first step of the proof, and the second is an ideal since the Killing form satisfies $K([X, Y], Z) + K(Y, [X, Z]) = 0$, so if $Y \in \mathcal{I}, Z \in \mathcal{D}, X \in \mathcal{D}$, then $K([X, Y], Z) = 0$ and so $K(Y, [X, Z]) = 0$. Therefore, if $D \in \mathcal{I}^\perp$ and $ad(X) \in \mathcal{I}$ we have $[D, ad(X)] \in \mathcal{I} \cap \mathcal{I}^\perp = \{0\}$, so whenever $X, Y \in \mathcal{G}$ we have

$$[D, ad(X)](Y) = D([X, Y]) - [X, D(Y)] =$$

$$[D(X), Y] + [X, D(Y)] - [X, D(Y)] = [D(X), Y] = 0$$

and therefore $D(X)$ is in the center of the Lie algebra and so zero. This holds for all X , so $D = 0$. QED.

17.4 The Lattice Associated With $Aut_0(\mathcal{G})$

Suppose that G is a compact Lie group with trivial center. Then G is associated with a Dynkin diagram. If this diagram is not connected, we can write $G = G_1 \times \dots \times G_k$ where each G_i has a connected Dynkin diagram and no center, and is the group of automorphisms of its associated Lie algebra.

Let T be a maximal torus in G . Then $\mathcal{T}$ is the universal covering group of T . The set of all points in $\mathcal{T}$ which map to the identity forms a lattice in $\mathcal{T}$, and our goal is to determine this lattice precisely.

Theorem 80 *If G is a connected compact Lie group with trivial center, the lattice of points in $\mathcal{T}$ which map to the identity in T is the set of all $t \in \mathcal{T}$ such that $\langle t, \alpha \rangle \in \mathbb{Z}$ for all roots α .*

Proof: The proof doesn't really assume that the Dynkin diagram is connected, but it is useful to imagine that case.

Of course $Ad(e) = I$. Conversely if $t \in T$ and $Ad(t) = I$, then conjugation by t induces the identity map on the Lie algebra and thus is the identity map on G , so t is in the center of G . We are assuming that this center is trivial, so $t = 0$.

When roots were first introduced, we discovered that $Ad(t)$ acts on the two-dimensional root space associated with α via

$$\begin{pmatrix} \cos 2\pi f_\alpha(t) & \sin 2\pi f_\alpha(t) \\ -\sin 2\pi f_\alpha(t) & \cos 2\pi f_\alpha(t) \end{pmatrix}$$

Since $\mathcal{G}$ is the direct sum of these two-dimensional subspaces and $\mathcal{T}$ and since t acts as the identity on $\mathcal{T}$, it follows that $Ad(t)$ is the identity if and only if $f_\alpha(t)$ is an integer for all α , or equivalently $\langle \alpha, t \rangle \in \mathbb{Z}$ for all α . QED.

Remark: The sets $\langle t, \alpha \rangle \in \mathbb{Z}$ are precisely the hyperplanes in the Stiefel diagram. Therefore the lattice consists of all points which lie on the intersection of hyperplanes associated with every root. These points are easily visible in our pictures of the Stiefel diagrams in the previous section.

17.5 Tricky Points

As we discuss some issues, it is useful to have a Stiefel diagram available. Below is the picture when the group is $SU(3)$ modulo its center.

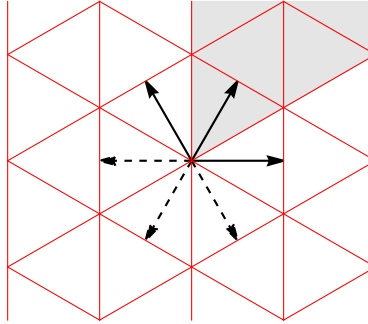


Figure 17.6: Stiefel Diagram for A_2

Question: Does the Stiefel diagram depend on the inner product $\langle , \rangle$ on $\mathcal{T}$?

Answer: Not really. We proved that this inner product is unique up to scalars when the Dynkin diagram is connected. The Stiefel diagram depends on the roots α , the points $t \in \mathcal{T}$, and the equations $\langle \alpha, t \rangle \in \mathbb{Z}$. The extra λ just rescales the diagram because

$$\langle \alpha, t \rangle = \lambda \left\langle \frac{\alpha}{\sqrt{\lambda}}, \frac{t}{\sqrt{\lambda}} \right\rangle$$

Question: Does the diagram depend on the scale used to draw the roots?

Answer: No. If we magnify all roots by λ then the diagram shrinks by $\frac{1}{\lambda}$ because

$$\langle \alpha, t \rangle = \left\langle \lambda \alpha, \frac{t}{\lambda} \right\rangle$$

Consequently we can draw tiny roots near the origin and get an expanded diagram, or larger roots and shrink the diagram. Authors tend to draw roots so they end at attractive points, like hyperplanes in the diagram or lattice points. This is irrelevant.

Question: Do all Stiefel diagrams look exactly the same?

Answer: Yes and no. After drawing all Stiefel diagrams possible in the plane earlier in these notes, I compared my pictures to pictures in *Lectures on Lie Groups* by J. Frank Adams. His pictures did not agree with mine. I spent a week trying to trace down the error.

For one thing, Adam's root vectors were longer than mine. But as we have seen, these lengths are irrelevant. Adams also marked only some points where all families of hyperplanes meet as lattice points. It looked like he was drawing more hyperplanes, perhaps twice as many, and then discarding some lattice point candidates.

Eventually I noticed that his diagram was labeled *Stiefel Diagram for $SU(3)$* and my drawing was labeled *Stiefel Diagram for A_2* . The problem is that A_2 refers to a Lie algebra, which is an algebra for several different groups. My diagram was for the group $SU(3)/Z_3$ obtained by dividing out the center of $SU(3)$.

The answer to the question is that all Stiefel diagrams have the same roots and the same hyperplanes, but possibly different lattices. The largest lattice, formed by all points where hyperplanes of all possible types intersect, occurs when the group has no center. Only some of these points are lattice points if the Lie algebra is the same but the group has a center. For these diagrams, it can appear that there are more hyperplanes, but actually there are fewer lattice points.

We discuss these alternate lattices in the next section.

17.6 Torus Lattices in General

The Lie algebra of a compact Lie group has the form

$$\mathcal{G} = R^k \oplus \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_k$$

We are currently dealing with groups with no initial R^k . These are called *semisimple* groups in the literature, so we adopt that term. If G is a connected compact semisimple group, it

has finite center. We proved that the universal cover of such a group is also compact. So that universal cover $\tilde{G}$ has a finite center C . All groups with the same Lie algebra are of the form $\tilde{G}/C'$ where C' is a subgroup of C . In particular, dividing by the entire C gives the groups studied in the previous section.

Theorem 81 *Let G be a compact semisimple Lie group which is simply connected. Let T be a maximal torus of G and let $\mathcal{T} \rightarrow T$ be the universal cover of the maximal torus. Then an element $t \in \mathcal{T}$ maps to the identity in G if and only if $w(t)$ is an integer for all weights in the weight lattice of T .*

Remark: The proof of this result is a little complicated, so we postpone the proof until the end of the section. The result has many beautiful consequences.

Remark: Since roots are weights, a point which meets the condition of the theorem certainly satisfies $\langle \alpha, t \rangle \in \mathbb{Z}$ for all roots, and thus belongs to a point in the lattice associated with the Stiefel diagram. But only some of these points will satisfy the extra condition imposed by additional weights.

Recall that if L is a lattice, the dual lattice L^* is the set of all t such that $\langle t, X \rangle \in \mathbb{Z}$ for all $X \in L$. Clearly the lattice associated with the Stiefel diagram is the dual of the root lattice. According to the theorem just stated, the lattice associated with the universal cover is the dual of the weight lattice. The root lattice is a subset of the weight lattice, $L_R \subset L_W$. So $L_R^* \supset L_W^*$, as asserted in the initial remark.

Remark: We proved that the weight lattice is the dual of the coroot lattice. It follows that the lattice associated with the universal cover is the coroot lattice. This lattice contains all coroots of ordinary roots. Notice that the coroot of α is a multiple of α . We can easily determine where this coroot ends in the Stiefel diagram because

$$\left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \alpha \right\rangle = 2$$

So the coroot arrow associated with α is obtained by moving in the direction of α until we reach the second perpendicular hyperplane it hits in the Stiefel Diagram. The coroots of the simple roots generate the coroot lattice, so it is very easy to draw the full lattice in our two dimensional pictures.

Corollary 11 L_R^*/L_W^* is isomorphic to the center of $\tilde{G}$.

Proof Let C be the center of $\tilde{G}$ and let $\mathcal{T}$ be its maximal torus. Clearly the map

$$\tilde{G} \rightarrow \tilde{G}/C = \text{Aut}_0(\mathcal{G})$$

maps C to the identity element and $\mathcal{T}$ to the maximal torus T of $\tilde{G}/C$ and thus induces maps

$$\mathcal{T} \subset \mathcal{G} \rightarrow \tilde{T} \rightarrow \tilde{T}/C = T$$

The inverse image of the identity on the right is L_R^* and the inverse image of $e \in \tilde{T}$ is L_W^* . It follows that these maps induce isomorphisms

$$L_R^*/L_W^* \rightarrow \tilde{T}/T = C$$

QED

Remark: It follows that this map induces isomorphisms between lattices $L_W^* \subseteq L \subseteq L_R^*$ and subgroups of C . So these maps also induce correspondences between connected Lie groups of the form $\tilde{G}/C'$ and lattices between L_W^* and L_R^* . Note that every connected Lie group with Lie algebra $\mathcal{G}$ has the form $\tilde{G}/C'$.

Proof of Main Theorem We are now ready to prove the main theorem. Our proof is going to use two crucial results which we have not yet proved, so we don't get to use this theorem until both results are proved. Luckily, the theorem we are proving isn't part of our machinery, but instead one of the culminating theorems of compact Lie group theory, allowing us to find the Stiefel diagram and torus lattice of an arbitrary Lie group, and classify all groups with Lie algebra $\mathcal{G}$ purely in terms of the Lie algebra.

Here are the two missing theorems:

- Peter-Weyl: If G is a compact topological group and $g \neq e \in G$, there is a continuous finite dimensional complex representation ρ of G with $\rho(e) \neq I$.
- Weyl: Recall that the simple roots $\alpha_1, \dots, \alpha_n$ of $\mathcal{G}$ form a basis of L_R and the corresponding coroot vectors $\frac{2\alpha_1}{\langle \alpha_1, \alpha_1 \rangle}, \dots, \frac{2\alpha_n}{\langle \alpha_n, \alpha_n \rangle}$ form a lattice basis for the coroot lattice. The dual lattice to this coroot lattice is the weight lattice, and the dual basis of the coroot lattice is a basis for this weight lattice. If w_i is one of these dual vectors, then, then

$$\left\langle w_i, \frac{2\alpha_j}{\langle \alpha_j, \alpha_j \rangle} \right\rangle = \delta_{ij}$$

This w_i is a candidate for the highest weight of an irreducible representation of $\mathcal{G}$. We have proved all of these results. According to Weyl, such a representation exists.

Proof Step One: We will show that if $w(t) \in Z$ for all $w \in L_W$, then t maps to $e \in G$. If $\varphi : G \rightarrow GL(V)$ is a complex representation of G , and one of the weights of φ is w with a complex weight vector e_w , we have

$$\varphi(t)e_w = e^{2\pi i w(t)} e_w$$

Since $w(t) = \langle w, t \rangle$, and all weights live in the weight lattice, and t is in the dual lattice, $w(t)$ is an integer and $\varphi(t)e_w = e_w$. Since the e_w form a basis for V , $\varphi(t) = I$. But the only element of G which is sent to the identity by every representation is the identity by the Peter-Weyl theorem. So t maps to the identity in G .

Proof Step Two: We will show that if t maps to $e \in G$, then $\langle w, t \rangle \in Z$ whenever w is a weight. Since the dual basis w_i mentioned earlier is a basis for the weight lattice, it suffices to prove that $\langle w_i, t \rangle \in Z$ for all i . Let φ be a representation of G with highest weight w_i . Then $\varphi(t) = I$ since e maps to the identity, so $\varphi(t)e_{w_i} = e_{w_i}$. But

$$\varphi(t)e_{w_i} = e^{2\pi i w_i(t)} e_{w_i}$$

and therefore $w_i(t) = \langle w_i, t \rangle \in Z$.

17.7 Looking Backward

Our study of compact Lie groups began with the creation of an invariant Riemannian metric on the group. This metric is unique up to multiplication by a positive scalar λ . (A minor aside: if $\mathcal{G} = \oplus(\mathcal{G}_i)$, we can multiply by different scalars on each piece.) Other mathematical theories possess a similar ambiguity; for example, Haar measure is unique up to a positive scalar.

It is reasonable to ask if there is a canonical way to normalize this metric. The correct answer is *no; the theory is completely independent of the choice of λ* . However, several tempting examples could mislead us. Haar measure on compact groups is often normalized by requiring that the measure of the entire group be one. If we did that, we would get a different normalization on $\mathcal{G}$ for each group with Lie algebra $\mathcal{G}$, a real mess. The algebraic approach to our theory uses the Killing form instead of the invariant metric, and this form is unique. But there is nothing special about the Killing form, and using it would tempt users to explicitly calculate it when studying examples, only to discover that the tedious calculation is irrelevant in the end. After we know more about the classification, and realize that there are at most two root lengths, we could normalize so all short roots have length one, or all long roots have length one. This is sometimes done to make certain calculations easier.

One of the lessons of modern mathematics is that if there is an essential ambiguity, it should be made an explicit part of the theory rather than hiding it under the rug. That is why Galois groups are important in algebraic number theory. We adopt that approach in this theory, hence the remaining paragraphs below.

We must cleanly separate discussions of $\mathcal{T}$ and of its dual $\mathcal{T}^*$ and avoid identifying these objects. For example, the entire theory of Stiefel diagrams is a theory about $\mathcal{T}$. The hyperplanes live in $\mathcal{T}$ and map to singular points in G . The dual of the root lattice lives in $\mathcal{T}$ and maps to the identity element of G when G has no center. The coroot lattice lives in $\mathcal{T}$ and maps to the identity when G is simply connected. In particular, the coroots live in $\mathcal{T}$.

On the other hand, the roots and weights live in $\mathcal{T}^*$. The root lattice and the weight lattice both live in $\mathcal{T}^*$.

Changing λ shrinks or magnifies both $\mathcal{T}$ and $\mathcal{T}^*$ and everything in them. If we do not assign units to distances, we always get the same picture in these spaces independent of λ .

However, the spaces are not independent because the elements of $\mathcal{T}^*$ act on $\mathcal{T}$ and these actions do not depend on λ . How can we symbolize that in our theory? Notice that the issue only comes up if we are dealing with a formula which contains elements of both $\mathcal{T}$ and $\mathcal{T}^*$.

Even before we introduce the inner product $\langle , \rangle$, we have a vector space $\mathcal{T}$ and a vector space $\mathcal{T}^*$ and each element of the second acts on the first, making the second vector space into the dual of the first. Then we add to the mix an inner product $\langle , \rangle$ on $\mathcal{T}$. This inner product induces a canonical isomorphism $\mathcal{T} \rightarrow \mathcal{T}^*$ and this isomorphism induces an inner product on $\mathcal{T}^*$. If we trace through these canonical operations, we find that the operation of $\mathcal{T}^*$ on $\mathcal{T}$ is now given by the simple formula below, where the α on the left belongs to the dual space and the α on the right is the corresponding element in $\mathcal{T}$.

$$\alpha(t) = \langle \alpha, t \rangle$$

The central question then becomes: if we replace $\langle , \rangle$ by $\lambda \langle , \rangle$, how does this formula change? The answer is very beautiful. The key formula $\alpha(t) = \langle \alpha, t \rangle$ is unchanged.

As a consequence, if we replace $\langle , \rangle$ by $\lambda \langle , \rangle$, then all lengths in $\mathcal{T}$ increase by $\sqrt{\lambda}$ and all lengths in $\mathcal{T}^*$ decrease by $\frac{1}{\sqrt{\lambda}}$. Moreover, the isomorphism $\mathcal{T}^* \rightarrow \mathcal{T}$ changes, so while roots in $\mathcal{T}^*$ remain the same, their images in $\mathcal{T}$ move. The proof is easy.

As long as we are looking at a picture of $\mathcal{T}$ or a picture of $\mathcal{T}^*$ and the picture has no scale, the picture will not change. We'll only notice a change if we drawn a picture containing elements from both spaces. Think of this as a picture with two layers, as provided by many computer drawing programs. When we modify by λ , one layer will expand while the other contracts. Inverting λ changes an expanding layer into a contracting layer and vice versa.

A purist might demand that we rewrite the notes and totally avoid identifying roots in $\mathcal{T}^*$ with vectors in $\mathcal{T}$. Of course this is possible, but it has two downsides. The first is that such an identification is often made in the literature, so avoiding it makes this literature hard to read. The second is that knowing the map $\mathcal{T}^* \rightarrow \mathcal{T}$ is really the same thing as knowing the induced inner product on $\mathcal{T}^*$, because that inner product is defined by mapping $\mathcal{T}^*$ to $\mathcal{T}$ and then applying the inner product on $\mathcal{T}$. So if we avoided the map, then we'd have to compute and utilize a second inner product, the one on $\mathcal{T}^*$, and that notion is more nebulous than simply mapping to a known geometry.

There is one mild surprise here: coroots live in $\mathcal{T}$. There is an alternate definition of a coroot that makes this clear from the beginning. Suppose $\alpha \in \mathcal{T}^*$ is a root with root vector

$u + iv$. Recall that $[u, v] \in \mathcal{T}$. This particular element depends on the choice of u and v , but the one dimensional subspace it generates is well-defined in $\mathcal{T}$. The coroot of α is the unique element t of this subspace satisfying $\alpha(t) = 2$.

On a more concrete level, the coroot of α is $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$. Notice that

$$\left\langle \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \beta \right\rangle = 2 \frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}$$

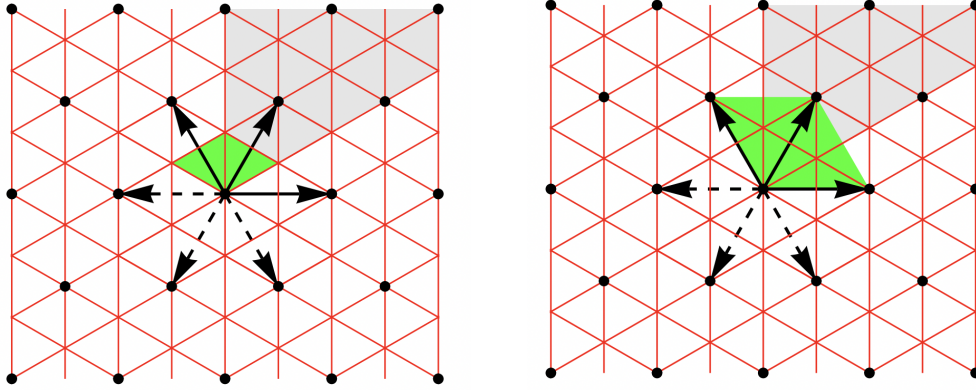
and the right side is independent of λ . So $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$ is in the dual space of $\mathcal{T}^*$, and thus in $\mathcal{T}$.

The new point of view cuts the plethora of lattices in these notes down to reasonable size. In $\mathcal{T}$ there are two natural lattices, L_R^* and $\{\text{coroots}\}$. These correspond to the smallest and largest groups with Lie algebra $\mathcal{G}$. In $\mathcal{T}^*$, there are two natural lattices L_R^{**} and $\{\text{coroots}\}^*$, that is, L_R and L_W . L_R is generated by the weights of the adjoint representation, which is defined on all groups with Lie algebra $\mathcal{G}$, while L_W is generated by the weights of all possible representations, including those only defined on the largest group $\tilde{G}$ with Lie algebra $\mathcal{G}$. So representation theory sets up a sort of duality between the first pair of lattices and the second pair of lattices.

17.8 The Affine Weyl Group

This additional material about the Stieffel Diagram will be used in the proof of the Weyl Integral Formula later on.

The *Affine Weyl Group* is the group generated by reflections across the hyperplanes that define the Stiefel diagram. Among these are the hyperplanes that contain the origin, and their reflections generate the ordinary Weyl Group. So the Affine Weyl Group contains the ordinary Weyl Group as a subgroup.

Figure 17.7: Fundamental Regions for $SU(3)/Z_3$ and $SU(3)$

The Affine Weyl Group also contains translations. If α is a root, then reflection across the hyperplane $\alpha(t) = 1$ followed by reflection across $\alpha(t) = 2$ is a translation that maps the origin to the coroot $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$. Recall that the coroots generate the coroot lattice, which is also the set of points in $\mathcal{T}$ which map to the identity in T if we are dealing with the simply connected Lie group associated with $\mathcal{G}$. It follows that every translation of the form $v \rightarrow v + t$ where t is in the coroot lattice belongs to the Affine Weyl Group. Call the subgroup of these particular translations T_{co} .

Lemma 40 *The subgroup T_{co} of the affine Weyl Group is normal, and the full Affine Weyl group is the semidirect product of this translation group with the ordinary Weyl group W .*

Proof: We claim that T_{co} is invariant under conjugation by elements of W . Indeed if $T(v) = v + d$, then $wT w^{-1}(v) = w(w^{-1}v + d) = v + wd$. Note that w maps coroots to coroots.

Next we introduce some temporary notation. Let $R_{\alpha,k}$ denote reflection across the hyperplane $\alpha(t) = k$, and let T_α denote translation by $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$. Notice that

$$R_{\alpha,k} = T_{k\alpha} R_{\alpha,0}$$

Indeed, a point on the hyperplane $\alpha(t) = k$ is moved by the reflection on the right to a point with $\alpha(t) = -k$ and then by the translation to a point with $\alpha(t) = -k + 2k = k$. Also the origin is moved by both sides to k times the α coroot.

It follows that every generating reflection can be written as a product Tw where $T \in T_{co}$ and $w \in W$. But then every element of the Affine Weyl group can be written this way since the set of such expressions is invariant under taking products. Indeed $(T_1 w_1)(T_2 w_2) = (T_1 w_1 T_2 w_1^{-1})(w_1 w_2)$. The first half of this expression is a product of two translations by the initial remark of the proof, and the second is a product of two elements of W ,

We claim that the previous representation is unique. Indeed if $T_1 w_1 = T_2 w_2$, then $T_2^{-1} T_1 = w_2 w_1^{-1}$, but the element on the right has finite order and the element on the left is either the identity or else has infinite order. If it is the identity, then both sides are the identity and $T_1 = T_2$, $w_1 = w_2$.

Notice that $(T_1 w_1)(T_2 w_2) = (T_1 w_1 T_2 w_1^{-1})(w_1 w_2)$ is a typical formula for a semi-direct product. This finishes the argument. QED.

Remark: Suppose t is a point in $\mathcal{T}$ which is not on any of the distinguished hyperplanes. Recall that such a point is called a regular point. For each α this point is between a family of parallel hyperplanes. Choose new coordinates $t_1, \dots, t_n$ such that the planes are parallel to the $t_1, \dots, t_{n-1}$ base and determined by values t_n that form a lattice in that direction. Then there are two closest planes to t , one above it and one below it, and the open region between these planes is convex.

Repeat this process for each of the finitely many roots, and take the intersection of these convex regions. In the end we obtain a bounded connected open set in $\mathcal{T}$ which is convex; the region contains only regular elements. The region is bounded by a finite number of hyperplanes. We call such a set an *alcove* of the diagram.

This process can be repeated starting with any regular $t \in \mathcal{T}$, so it shows that the set of regular elements of $\mathcal{T}$, which is the complement of an infinite number of hyperplanes, is a union of such connected, open, pre-compact, convex alcoves, each bounded by a finite number of hyperplanes.

Lemma 41 *The elements of the Affine Weyl Group map alcoves to alcoves. The group acts simply transitively on alcoves, so if two alcoves are given, there is a unique group element mapping the first to the second.*

Proof: The image of an alcove under the Affine Weyl group is a connected open set which intersects no defining hyperplane, and thus must be contained in a second alcove. Applying the argument to the inverse of the group element shows that this map is onto.

It is easy to see that the Affine Weyl group acts transitively on alcoves, since we can draw a path from any alcove to any other alcove which crosses boundary hyperplanes at non-vertices; this map easily produces a series of reflections leading from the first alcove to the second one.

Thus it suffices to prove that if two maps send an alcove $\mathcal{A}_1$ to the same alcove $\mathcal{A}_2$, the two maps are equal. It suffices to prove the special case that if an element of the Affine Weyl Group maps an alcove $\mathcal{A}$ back to itself, then the group element must be the identity. Since the group is transitive on alcoves, we can assume that $\mathcal{A}$ is a preferred alcove fixed in advance. Any Weyl chamber clearly has an alcove whose closure contains 0. This alcove is unique in that chamber, since two different alcoves would need to be separated by a

hyperplane through the origin, and there are none inside a Weyl chamber. So assume that $\mathcal{A}$ is the unique alcove containing 0 in the fundamental Weyl chamber.

If a non-trivial element of the affine Weyl group maps $\mathcal{A}$ back to itself, it equals Tw where $T \in T_{co}$ and $w \in W$. If Tw is not the identity, then neither T nor w can be the identity because both of these by themselves would move $\mathcal{A}$ to a different alcove. So $w(\mathcal{A})$ is an alcove in a different Weyl chamber whose boundary contains 0, and thus T is a non-trivial translation mapping this alcove back to the alcove in the fundamental Weyl chamber. Note that the maps in the affine Weyl group take vertices of alcoves to vertices of image alcoves, where by definition a vertex is a point where hyperplanes from all of the positive root families meet. Since T is a non-trivial translation, it must map the 0 vertex of the second alcove back to some other vertex of the alcove in the fundamental chamber. These other vertices belong to the weight lattice, but do not belong to the coroot lattice. This contradiction establishes the theorem. QED.

Chapter 18

Irreducible Representations of $SU(2)$ and $SU(3)$

18.1 Structure of Characters

Suppose φ is a representation of a compact connected Lie group G with connected Dynkin diagram. The character χ of φ lives on G and is invariant under conjugation. Since every element of G is conjugate to a element of T , χ is completely determined by its values on T . Since $\mathcal{T}$ is a covering space of T , χ is completely determined by its values on $\mathcal{T}$.

Recall that φ induces weights $w : \mathcal{T} \rightarrow R$ such that

$$\varphi(t)e_w = e^{2\pi i w(t)} e_w$$

The e_w form a complex basis for the representation space V , but the eigenspaces formed by the e_w may have dimension greater than one. Call this dimension d_w . Then

$$\chi(t) = \sum_{\text{weights of } \varphi} d_w e^{2\pi i w(t)}$$

Recall that the weight lattice is dual to the coroot lattice, which has lattice basis $\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle}$. The corresponding dual basis $w_1, \dots, w_n$ is a lattice basis for the weight lattice, and satisfies

$$\left\langle w_i, \frac{2\alpha_j}{\langle \alpha_j, \alpha_j \rangle} \right\rangle = \delta_{ij}$$

Any weight has the form $\sum a_j w_j$ where $a_j \in Z$, and

$$e^{2\pi i w(t)} = e^{2\pi i \sum a_j w_j(t)} = \prod \left(e^{2\pi i w_j(t)} \right)^{a_j}$$

If we define

$$\chi_j(t) = e^{2\pi i w_j(t)}$$

then

$$\chi(t) = \sum_w d_w \prod \chi_j^{a_j}$$

Therefore each character is a Laurent polynomial in the χ_j .

According to the highest weight theory, each irreducible representation of G has a unique highest weight in the closure of the fundamental Weyl chamber. This chamber equals the set of $\sum r_i w_i$ with $r_i \geq 0$, so the highest weight has the form $\sum a_j w_j$ where all a_j are non-negative integers. The dimension of the weight space associated with the highest weight has dimension one, and all other weights $\sum b_j w_j$ associated with φ have smaller sums $\sum b_j < \sum a_j$.

In this chapter we will verify this formula for $SU(2)$, and use it to find all irreducible representations of $SU(3)$.

18.2 $SU(2)$

The root and weight diagrams of $SU(2)$ are shown below.

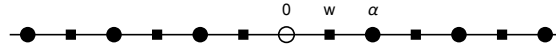


Figure 18.1: Stiefel Diagram for A_1

The highest weight can be any weight lattice element greater than or equal to zero, so $0, w, 2w, 3w, \dots$. Once we have the highest weight, all other weights are obtained by subtracting roots. We actually know the irreducible representations, so we can write down the complete weight list. Moreover in this special case all dimensions $d_w = 1$. From this information, it is easy to write the characters.

The following table lists the highest weights, the set of all weights, and the character for various representations. Note that there is only one weight basis vector which we call χ :

highest weight	:	weights	:	character
0	:	0	:	1
w	:	$-w, w$	:	$\chi^{-1} + \chi$
$2w$	:	$-\alpha, 0, \alpha$	:	$\chi^{-2} + 1 + \chi^2$
$3w$	:	$-3w, -w, w, 3w$	:	$\chi^{-3} + \chi^{-1} + \chi^1 + \chi^3$
$4w$	:	$-2\alpha, -\alpha, 0, \alpha, 2\alpha$	:	$\chi^{-4} + \chi^{-2} + 1 + \chi^2 + \chi^4$
$\dots$	:	$\dots$	:	$\dots$

Notice that half of these weight sets contain only roots; our theory says that these representations live on G/C , which is $SO(3)$ in our case. The remaining representations contain weights, and our theory says these representations do not live on all G associated with $\mathcal{G}$, but certainly live on $\tilde{G}$, which is $SU(2)$ in our case.

18.3 $SU(3)$

The root and weight diagrams of $SU(3)$ are shown below. The black dots give the root diagram, and intersection points of red lines give the weight lattice. The shaded region is the principal Weyl chamber.

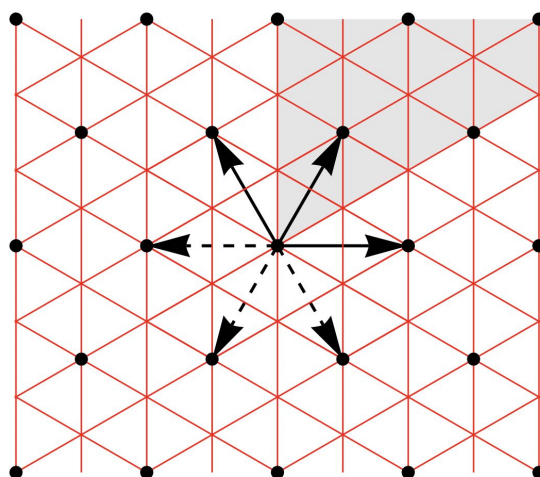


Figure 18.2: Stiefel Diagram for A_2

The Weyl group is D_3 , the symmetries of a triangle. It is easy to become confused by the many hexagons in the lattice, so take a moment to confirm that there are only three

lines perpendicular to the roots, and thus only three reflections in the resulting Weyl group.

The weights and roots live in the dual space to $\mathcal{T}$, so our diagram is a picture of that dual space. The arrows in the diagram are the roots, and the root lattice has a lattice basis given by the simple roots α and β where α points to the right and β points left and up.

There is a corresponding coroot lattice, not shown. In this particular case all of the roots have equal length, so informally the coroots can be identified with the roots. The coroots $\frac{2\alpha}{\langle\alpha,\alpha\rangle}$ and $\frac{2\beta}{\langle\beta,\beta\rangle}$ form a lattice basis of the coroot lattice. The weight lattice is dual to the coroot lattice and the basis dual to the previous basis forms a lattice basis for the weight lattice. Temporarily call this dual basis w_1, w_2 . We are going to identify these vectors in the picture.

For instance, consider w_1 . Then $\langle w_1, \frac{2\alpha}{\langle\alpha,\alpha\rangle} \rangle = 1$ and $\langle w_1, \frac{2\beta}{\langle\beta,\beta\rangle} \rangle = 0$. The vertical line through the origin consists of w whose inner product with $\frac{2\alpha}{\langle\alpha,\alpha\rangle}$ is zero, and the vertical red lines are places where this inner product is an integer. Notice that the inner product of α with its coroot is 2, both in the picture and algebraically; that is why the arrow α ends on the second vertical line. So w_1 must be somewhere on the first vertical red line. Since w_1 is perpendicular to the coroot associated with β , it is perpendicular to β and thus on the red line through the origin which slopes upward to the right. This is exactly the lower border of the fundamental Weyl chamber. Thus w_1 is the first weight lattice point on this chamber wall.

In the same way, w_2 is the first weight lattice point on the vertical fundamental Weyl chamber wall.

The possible highest weights are weight lattice points in the fundamental chamber, and thus exactly points with non-negative integer coordinates (h, k) in the coordinate system determined by w_1 and w_2 . To find such a point, move diagonally up by h units, and then vertically up by k units.

By the highest weight theorem, each irreducible representation is associated with a unique highest weight (h, k) in the fundamental Weyl chamber. It will turn out that this representation has dimension

$$\frac{(h+1)(k+1)(h+k+2)}{2}$$

Here are examples:

h	k	dimension
0	0	1
1	0	3
0	1	3
2	0	6
1	1	8
0	2	6
3	0	10
2	1	15
1	2	15
0	3	10
4	0	15
3	1	24
2	2	27
1	3	24
0	4	15

We can begin to guess some of these representations and their properties. Notice the symmetry between (h, k) and (k, h) . We can reasonably guess that these representations are dual. Since they aren't equivalent each would have complex type. The exceptions would be representations on the line (h, h) , which are of either real or quaternionic type. In the case of $SU(3)$, we will guess that all of these are of real type.

The representation $(1, 0)$ of dimension 3 is likely the group itself as a group of matrices acting on C^3 . The group $(1, 1)$ of real type has dimension 8, which is the dimension of the Lie algebra $su(2)$ and thus is likely the adjoint representation of G on its Lie algebra $su(2)$. Note that the highest weight of this representation is a root, so the representation drops to $\tilde{G}/C$. Moreover, there is a real basis for $\mathcal{G}$ and thus the matrices of the representation are all real in an appropriate basis.

The group $SU(3)$ acts naturally on a vector space V of dimension 3, so it also acts naturally on $V \otimes V$ of dimension 9. However this representation is not irreducible, since we can write $V \otimes V = \Lambda^2(V) + \text{Symm}(V)$. If e_1, e_2, e_3 is a basis of V , then $e_i \wedge e_j$ with $i < j$ is a basis of $\Lambda^2(V)$. Therefore this vector space has dimension 3. The symmetric tensors have as basis the six symmetric products of these basis vectors. The representation on $\Lambda^2(V)$ is probably equivalent to the natural action of $SU(3)$ on C^3 . But the two representations of dimension 6 are likely representations on symmetric polynomials, generalizing the representations of $SU(2)$.

It turns out that the full weight diagram for every irreducible representation of $SU(3)$ has

the following shape. Notice that this shape is invariant under the Weyl group D_3 . Here the entires labeled “a, b” are our “k, h”. The weight spaces on the outer rim have dimension 1, those in the next level have dimension 2, and so forth, until the six sided figures become triangles. From that point on, the dimensions of the weight spaces in the various triangles do not change.

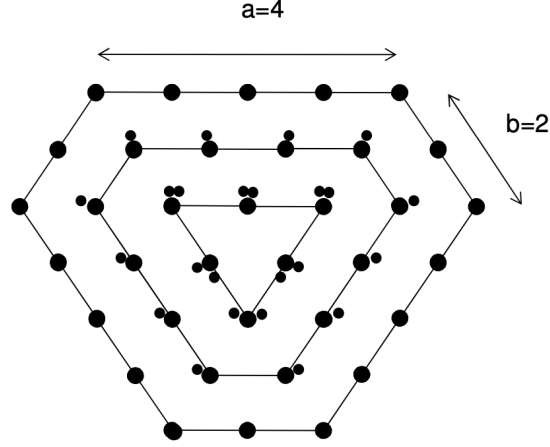


Figure 18.3: Typical Weight Diagram

By symmetry we can assume that $h \geq k$. If $h = k$ we have a perfect hexagon. The point (h, h) is one vertex of the hexagon and the other vertices are its images under D_3 . The various shrinking shells remain hexagons, with vertices $(h-1, h-1), (h-2, h-2), \dots$ and the final shell is the origin. We will later prove all this, but already we can obtain the dimension formula for this special case. According to that formula, the dimension of the representation is $\frac{(h+1)(h+1)(2h+2)}{2} = (h+1)^3$. Counting dots in the weight diagram with appropriate dimensions gives

$$6h + 2 \cdot 6(h-1) + 3 \cdot 6(h-2) + \dots + h \cdot 6((h-(h-1)) + (h+1))$$

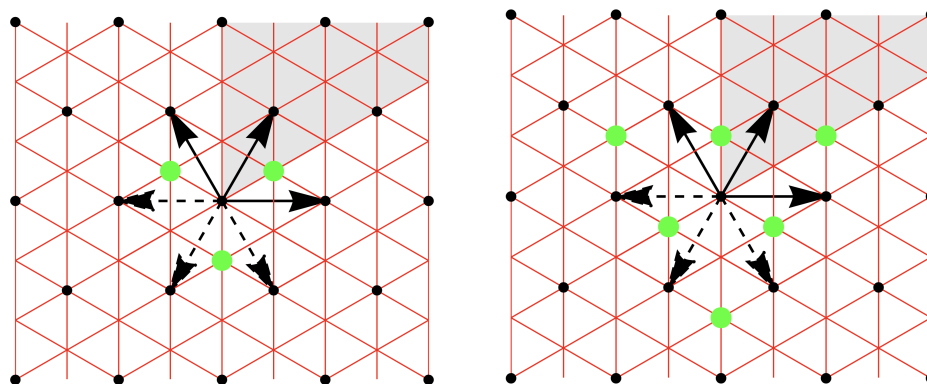
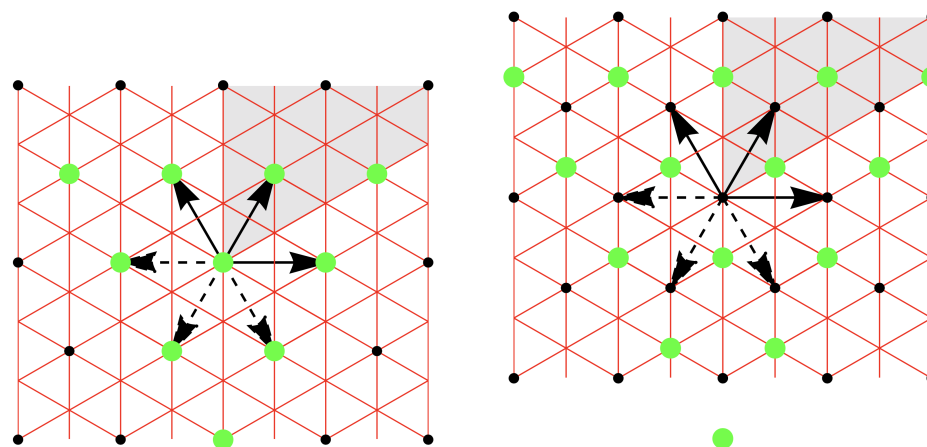
The first sum of terms is

$$6[h + 2(h-1) + 3(h-2) + \dots + h(1)] = 6 \sum_{k=1}^h k(h-k+1) = 6(h+1) \sum_{k=1}^h k - 6 \sum_{k=1}^h k^2$$

Since $\sum k = \frac{h(h+1)}{2}$ and $\sum k^2 = \frac{h(h+1)(2h+1)}{6}$ the complete sum is

$$6 \frac{h(h+1)^2}{2} - 6 \frac{h(h+1)(2h+1)}{6} + (h+1) = (h+1)[3h(h+1) - h(2h+1) + 1] = (h+1)^3$$

A second special case occurs when h or k is zero and the highest weight is on the boundary of the Weyl Chamber. In this case the boundary of the weight diagram is a triangle. It suffices to study the case $(h, 0)$ where one vertex of the triangle is $(h, 0)$; the other two are obtained by transforming using D_3 . The behavior of this case as the shells shrink is slightly tricky, so a few pictures help to show the general pattern. Below are the cases $h = 1, 2, 3, 4$:


 Figure 18.4: Triangles $(1, 0)$ and $(2, 0)$

 Figure 18.5: Triangles $(3, 0)$ and $(4, 0)$

These pictures reveal a significant feature of the triangular patterns; they have a periodicity of order 3. Every third $(h, 0)$ is a root vector, and for these diagrams the inner shell is the origin. The $h = 1$ and $h = 2$ cases show the other possible inner shells. Rather than

looking at shells, it pays to notice that the top row of the $(h, 0)$ triangle has $h + 1$ weights, the next row has h weights, and so forth until the final row has 1 weight. In the triangular case, all weight dimensions are one. Thus the dimension is

$$1 + 2 + 3 + \dots + h + (h + 1) = \frac{(h + 1)(h + 2)}{2} = \frac{(h + 1)(0 + 1)(h + 0 + 2)}{2}$$

We finally turn to the general case of (h, k) where $0 < h$ and $0 < k$ and $h \neq k$. In each of these cases, the boundary of the weight diagram is a lopsided hexagon with sides that have $h + 1$ and $k + 1$ dots in an alternating pattern. The interior weights are arranged in shells where the number of dots on these sides declines to $h, k, h - 1, k - 1$, etc. We are assuming that h and k are different. Say $h > k$. Then eventually we will reach the shell with dot numbers $h - k$ and 0. This shell is a triangle, and from then on the triangular pattern discussed earlier occurs. As for the dimensions of these weight spaces, as usual that boundary weight spaces have dimension 1. As the shells shrink through lopsided hexagons, the weight space dimension increases through 2, 3, etc. This holds until the inner shell is a triangle. From then on, the weight space dimension remains the same as we shrink to the ultimate inner triangle.

Knowing this pattern, we can prove the dimension formula in the remaining cases. The boundary lopsided hexagon has $3(h + k)$ weights and the next shell has $3(h + k - 2)$ weights, etc. so we get

$$3(h + k) + 2 \cdot 3(h + k - 2) + 3 \cdot 3(h + k - 4) + \dots + (k + 1) \cdot 3(h + k - 2k)$$

But this sum counts the first triangle, and it is better to count that with the remaining triangles. That first triangle, and all remaining triangles, have weight spaces with dimension $k + 1$ behaves like starting at the highest weight $(h - k, 0)$. By the previous calculation, this triangular piece should then contribute

$$(k + 1) \frac{(h - k + 1)(h - k + 2)}{2}$$

We must add this to

$$3[(h + k) + 2(h + k - 2) + 3(h + k - 4) + \dots + k(h + k - 2(k - 1))]$$

A typical term in this sum is $j(h + k - 2j + 2)$, so we form $\sum_{j=1}^k (h + k + 2)j - \sum_{j=1}^k 2j^2$. This equals

$$3 \left[(h + k + 2) \frac{k(k + 1)}{2} - 2 \frac{k(k + 1)(2k + 1)}{6} \right]$$

Adding the two contributions gives the final formula for dimension,

$$3 \left[(h + k + 2) \frac{k(k + 1)}{2} - 2 \frac{k(k + 1)(2k + 1)}{6} \right] + (k + 1) \frac{(h - k + 1)(h - k + 2)}{2}$$

This expression equals $\frac{k+1}{2}$ times $3k(h+k+2) - 2k(2k+1) + (h-k+1)(h-k+2)$ and this last expression is

$$\begin{aligned} 3hk + 3k^2 + 6k - 4k^2 - 2k + h^2 - hk + h - hk + k^2 - k + 2h - 2k + 2 \\ = hk + k + h^2 + 3h + 2 = (h+1)(h+k+2) \end{aligned}$$

and the final formula is

$$\frac{(h+1)(k+1)(h+k+2)}{2}$$

Remark: We will now prove *some* of our assertions on irreducible representations of $SU(3)$. We often refer to the *weight diagram* of a representation. This is a list or picture of the weights that occur in the representation, with numbers attached to each weight giving the dimension of the weight space.

Lemma 42 *The weight diagram of a representation is invariant under the action of the dihedral group D_3 .*

Proof: In the second section of chapter 16, we selected a root α of $\mathcal{G}$ and found three elements $\{t_\alpha, u, v\}$ in $\mathcal{G}$ as usual. These elements generated a subalgebra of $\mathcal{G}$ isomorphic to $su(2)$, and in the corresponding subgroup of G we found an element n such that $Ad(n)$ maps T back to itself, and thus maps $\mathcal{T}$ back to itself, and such that this map on $\mathcal{T}$ is reflection across the hyperplane perpendicular to α . We continue to work with this element n .

Suppose $\varphi(t)e_w = e^{2\pi i \langle w, t \rangle} e_w$. Then

$$\varphi(n t n^{-1}) \varphi(n) e_w = e^{2\pi i \langle w, t \rangle} \varphi(n) e_w$$

The map $\varphi(n) : V \rightarrow V$ induces an equivalence between the representation $g \rightarrow \varphi(g)$ and the representation $g \rightarrow \varphi(n g n^{-1}) = \varphi(n) \varphi(g) \varphi(n^{-1}) : V \rightarrow V$. According to the previous equation, $\varphi(n) e_w$ is a weight vector for this new representation.

In this weight equation, the term $e^{2\pi i \langle w, t \rangle}$ involves elements in the Lie algebra, and we need to recall how the Lie algebra becomes involved in the story. Notice that the new representation has the form

$$G \xrightarrow{g \rightarrow n g n^{-1}} G \xrightarrow{\varphi} GL(V)$$

and therefore induces the following maps on the Lie algebra level:

$$\mathcal{G} \xrightarrow{Ad(n)} \mathcal{G} \xrightarrow{\varphi^*} gl(V)$$

Thus $\varphi^* \circ Ad(n)$ is the Lie algebra representation associated with $\varphi(ngn^{-1})$, and we should replace $\langle w, t \rangle$ by $\langle Ad(n)w, Ad(n)t \rangle$ in the weight equation. These two expressions are equal because $Ad(n)$ is an isometry $\mathcal{T} \rightarrow \mathcal{T}$. We have

$$\varphi(ntn^{-1})\varphi(n)e_w = e^{2\pi i \langle Ad(n)(w), Ad(n)(t) \rangle} \varphi(n)e_w$$

This equation tells us that for the new representation $\varphi(ngn^{-1})$, $\varphi(n)e_w$ is a weight vector with weight $Ad(n)w$.

Recall that $Ad(n)$ maps roots to roots and therefore maps the coroot lattice to the coroot lattice. Therefore it maps the weight lattice to the weight lattice. The reflections $Ad(n)$ generate the entire Weyl group, so the same statements hold for the entire group.

Recall now that each irreducible representation of $\mathcal{G}$ has a unique maximal weight in the fundamental Weyl Chamber, and two representations are equivalent if and only if they have the same maximal weight. Since $\varphi(g)$ and $\varphi(ngn^{-1})$ are equivalent representations, they have the same maximal weights and therefore the same weights and weight spaces. So $Ad(n)$ maps the weight diagram of φ to itself. The same statement then holds for the group generated by the $Ad(n)$, that is, the entire Weyl group. QED.

Remark: The dihedral group D_3 contains three rotations. A fundamental domain for the subgroup containing these rotations is shown below. Every point in the plane can be mapped into the shaded region by one of the rotations.

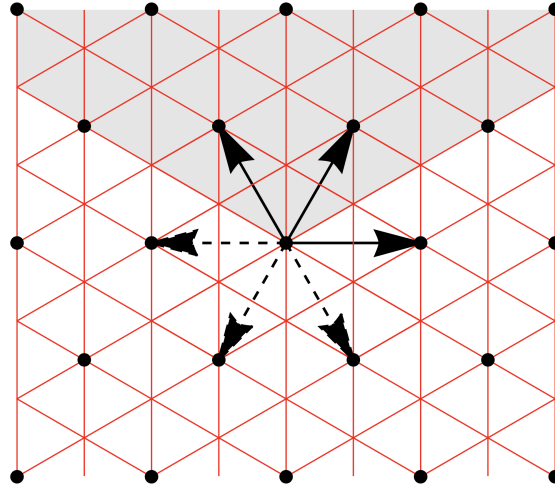
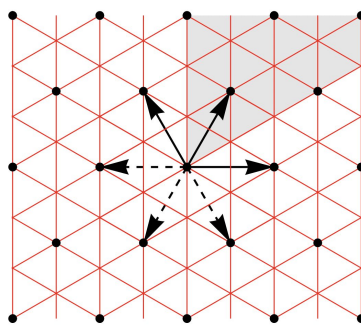


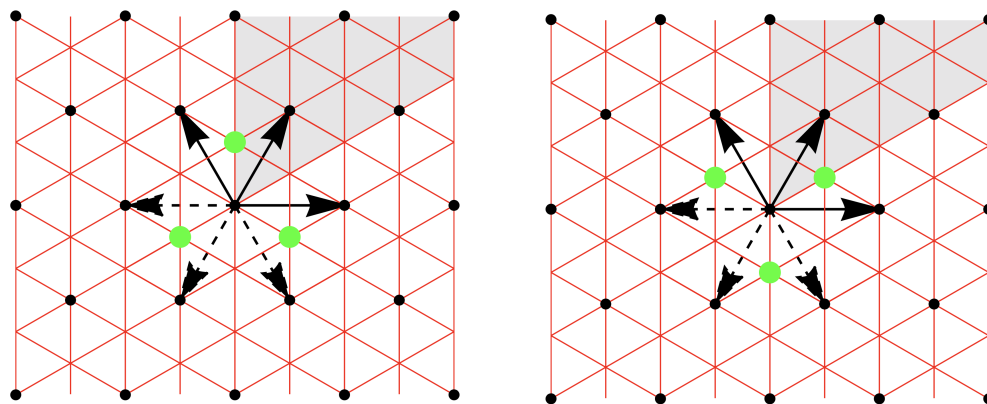
Figure 18.6: Stiefel Diagram for A_2

But the dihedral group also contains three reflections, across the three lines determined by the roots. So the fundamental domain for the full D_3 is only half as large, as shown

on the next page. Pay particular attention to points on the boundary of this region. The three rotations carry this fundamental region to closed regions which only intersect the fundamental region at the origin. The two reflections which produce regions adjacent to the fundamental region are reflections in the two walls of the fundamental region, and leave this wall fixed. Consequently no element of D_3 maps one wall of the region to the other wall, so all points in the closed region are inequivalent.

Figure 18.7: Stiefel Diagram for A_2

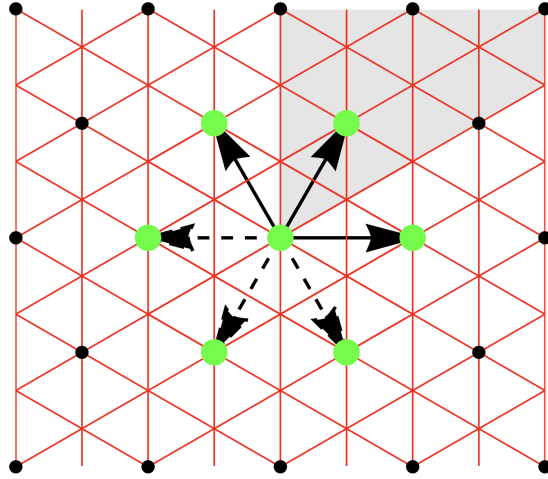
We are now ready to take baby steps toward finding the irreducible representations. The smallest representation is the identity, whose weight diagram contains only the origin. The next two weights in the fundamental region generate triangles when acted on by D_3 .

Figure 18.8: Smallest Representations of $SU(3)$

These two triangles must be the full weight diagrams for each of the two representations, for the following reasons:

- If a larger weight belonged to either diagram, then it could be moved into the fundamental Weyl chamber by an element of D_3 , and this element would be larger than the highest weight, contradicting the definition of a highest weight
- The origin cannot be a weight of either representation. Indeed we proved that all weight vectors are obtained from the highest weight w by acting on e_w with combinations of the $e_{-\alpha_i}$ over various chains of negative simple roots, and the corresponding weights will then all differ from w by negative integer combinations of simple roots.

Remark: Next consider the smallest interior lattice point in the fundamental Weyl chamber. Applying D_3 invariance gives the weights on the perimeter of the hexagon below. The highest weight space has dimension one, so each of these weight spaces also have dimension one. Recall that all other weights differ from w by an element of the root lattice. No larger weight can belong to the diagram because if one existed it could be moved to the fundamental Weyl chamber using D_3 , contradicting our choice of highest weight. The only weight inside the weight hexagon which differs from w by an element of the root lattice is the origin. We will show that it is a weight and that the dimension of this weight space is 1 or 2. The proof that the correct dimension is 2 will come later, as an application of the Weyl Character Formula.

Figure 18.9: Eight-Fold Way $SU(3)$

Consider the fundamental roots α and β , where α points right, and β points northwest. We know that a basis of the representation space is given by the $[e_{\gamma_1}, [e_{\gamma_2}, [\dots, e_w]]]$ where each γ_i is the negative of α or β . Notice carefully that all elements in this expression except w belong to $\mathcal{G}$ and the brackets are describing the action of $\mathcal{G}$ on the representation space V . So e_w and e_α are completely different animals. To avoid confusion, we will denote elements

of weight spaces in V by the letter f , so our highest weight is f_w . In our example, w is the element $\alpha + \beta$ in the root lattice, but we will continue to call this element f_w rather than $f_{\alpha+\beta}$.

Now $[e_{-\alpha}, f_w]$ is in the weight space associated with $w - \alpha = \beta$ and $[e_{-\beta}, f_w]$ is in the weight space associated with $w - \beta = \alpha$. So these elements are f_β and f_α . These are already in our diagram.

If we bracket f_β with $e_{-\alpha}$, we get an element in the weight space of $\beta - \alpha$, but this element is outside our hexagon and could be rotated into a higher element in the fundamental Weyl Chamber, contradicting our assertion that w is the highest weight. So this bracket must be zero. On the other hand, $[e_{-\beta}, f_\beta]$ would belong to the weight space of zero.

Similarly if we bracket f_α with $e_{-\beta}$, we get zero because the weight is too large, but $[e_{-\alpha}, f_\alpha]$ belongs to the weight space of zero.

At this point, the only “surviving” elements of the form $[e_{\gamma_1}, [e_{\gamma_2}, [\dots, f_w]]]$ are in the zero weight space and there are two of them, $[e_{-\alpha}, [e_{-\beta}, f_w]]$ and $[e_{-\beta}, [e_{-\alpha}, f_w]]$. If zero is not a weight, then the process would stop here and the diagram would not be invariant under D_3 . So zero is a weight and the weight space has dimension 1 or 2.

There is a reason to conjecture that the correct dimension is 2. The representation in question will have dimension 7 or 8. But we know an 8-dimensional representation of $SU(3)$, namely the adjoint representation on itself or rather $su(3) \otimes C$. There are six roots, providing six complex dimensions. In addition, the torus will provide 2 dimensions. Thus we expect the representation to have dimension 8, and the weight-space for 0 to be $\mathcal{T} \otimes C$.

But the rigorous proof will have to wait for the Weyl Character formula.

Remark: Before considering another case, it is useful to prove a general result. If a highest weight has the form (h, k) , it can be found by moving in the fundamental Weyl Chamber diagonally for h steps, and then vertically for k steps. Consider the chain of roots of the form $w + k\alpha$. Then $w + \alpha$ is higher than w in the chamber, or if outside the chamber, equivalent under D_3 to such a higher point. So $w + \alpha$ is not a weight. The chain thus has the form $w + p\alpha, w + (p+1)\alpha, \dots, w$ for some integer $p \leq 0$. All of these are weights of the representation, but the next bracket is zero. In chapter 14, we proved that $p = -\frac{\langle w, \alpha \rangle}{\langle \alpha, \alpha \rangle}$. Note that $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$ is the coroot associated to α , and the weight diagram is the dual of the coroot lattice, and (h, k) describes weight lattice points using the basis dual to the “simple coroot basis” for the coroot lattice. So $p = -h$. Consequently, if we start with w and bracket with $e_{-\alpha}$, we move left by the length of α in the root diagram, and we can do this exactly h times.

In exactly the same way, bracketing with $e_{-\beta}$ sends us diagonally down by the length of

β , and we can do this k times.

Remark: A specific example will help understand the significance of this. Consider the highest weight $(1, 2)$. Then we obtain the following weights in this manner:

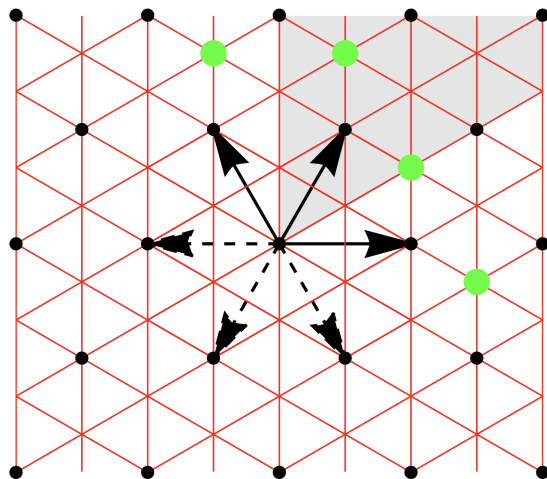


Figure 18.10: Case $(1, 2)$

As usual, we can map these to other weights using D_3 . Notice first that reflection across $\alpha + \beta$ preserves the picture. A simple argument given in a moment shows that this always holds. So we only need apply the rotations in D_3 , and we obtain the following picture:

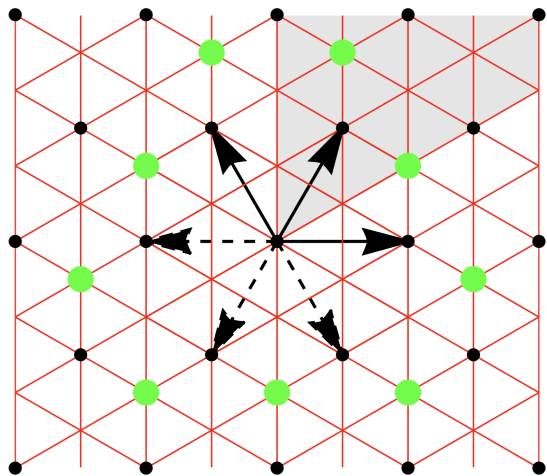


Figure 18.11: Case $(1, 2)$

It is easy to see that every weight vector outside this perimeter maps under D_3 to a higher weight than w . Therefore our result determines the perimeter of the weight diagram, not just in this special case, but for all representations of $SU(3)$.

We can push this still further. In our example, it would have sufficed to do the initial calculation for $w + k\alpha$, because that would have given the top two dots and reflection would have given the others. Let us repeat the same calculation for all of the weights w_i on the right side of the diagram. In each case $w_i + \alpha$ is outside the perimeter and thus not a weight. So in each case we get weights $w_i - k\alpha$ for positive integers $k = 0, 1, 2, \dots$ up to and including the first weight lattice coordinate of w_i . This coordinate just labels the vertical red line containing w_1 . So at the top of the diagram it is 1 and $k = 0, 1$ and there are two green dots. On the next line it is 2 and $k = 0, 1, 2$ and there are three green dots. On the next line it is 3 and $k = 0, 1, 2, 3$ and there are four red dots. On the final line it is 2 and $k = 0, 1, 2$ and there are three green dots. This little argument, however, produces all possible representation weights inside the diagram since all must differ by a root from w and thus by a root from w_1 .

This argument clearly works in general, and therefore gives all possible weights for any (h, k) . So only the dimensions of these weight spaces still need to be determined.

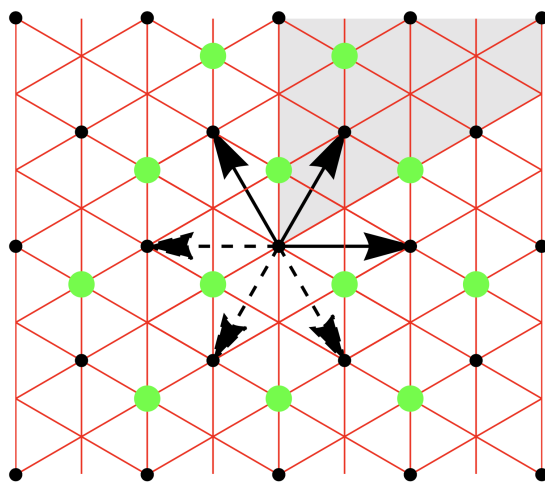


Figure 18.12: All Weights for Case (1, 2)

18.4 The Eightfold Way, Part 1

When I was a high school student in a small Kansas town in the 1950's, I discovered the magazine *Scientific American*. I still remember articles about the Godel incompleteness

theorem, and about non-Euclidean geometry.

I also read physics articles with care, particularly those on elementary particles. But the physicists began discovering more and more elementary particles, and elementary particle physics started to resemble biology. Eventually I gave up. As a result, I missed much of the amazing simplification that was to result in the *standard theory* of elementary particles during the 1970's.

In college I became aware of Hermann Weyl's book from 1928 *The Theory of Groups and Quantum Mechanics* in which Weyl described applications of representation theory to quantum mechanics. The general idea was that when a physical system has symmetries (think hydrogen atom and rotational invariance), the eigenfunctions will have degeneracies and the eigenspaces will be representation spaces for the symmetry group.

In the 1940's the representations of $SU(2)$ were used to describe the proton and neutron as a couple, because these particles seem to react equally to the strong force. This gradually led to the remarkable discoveries about representations of $SU(3)$ I am about to describe.

What I did not know in high school, and did not learn in college, is how the world of elementary particles changed from a small set containing electrons, protons, neutrons, the neutrino, and a couple of mesons called pions and muons to become instead a world cluttered with innumerable other mysterious particles. But recently I discovered the site <http://adsabs.harvard.edu/full/1990ICRC...10..113D> and a wonderful paper by I. V. Dorman written in 1990 which gives the history. Here are selections from that paper:

"The world of elementary particles seemed to be still rather simple in the late forties. Protons, neutrons, and electrons were constituents of the observed matter: photons were quanta of the electromagnetic field. Pions discovered early in 1947 by Powell were explained to be the exchange particles which produce nuclear forces. Pions and muons were united to form a meson family because their masses were alike and intermediate between the electron and nucleon masses. Given the antiparticle concept which appeared after Anderson discovered the positron, the only elementary particles whose discovery proved to be unexpected were muons and neutrinos.

"The above-mentioned relatively simple situation did not exist long. New particles with strange properties were discovered in the late forties as a result of studying the cosmic rays which are a natural source of high-energy particles.

"In the end of 1947, Rochester and Butler, who worked at the Cosmic Ray Laboratory of the University of Manchester under the guidance of Blackett, published two interesting pictures obtained in a Wilson cloud chamber. The first picture was obtained during the fall of 1946 and showed a Y-shaped fork on the right in the bottom part of the chamber. The fork was produced by tracks of two charged particles moving apart downwards from

the same point. The authors came to the conclusion that the fork arose from spontaneous disintegration of a heavy neutral particle into two charged pions or muons. The mass of the neutral particle was estimated to be about $1000 m_e$. (The mass of an electron is $.5 m_e$.)

“Rochester and Butler were quite aware of the importance of the picture, but hesitated to publish it until they had confirmation. On May 1947 a second picture was obtained, which shows the disintegration of a *charged* particle. They estimated that the lowest possible mass of the unstable particle was also about $1000 m_e$.

“The two pictures were the first manifestation of the complicated diversity of new particles. When Marshak visited Blackett’s laboratory in the fall of 1947, Rochester and Butler showed two excellent pictures of Y-particles to him, but he did not accept their interpretation and did not attract attention to the data when returning to the U.S. The next two years appeared to be poignant for the members of the Manchester group because other pictures of similar particles were not obtained. Eventually Rochester and Butler decided to carry their instruments to the high-mountain station Pic du Midi where the probability of new particles to be observed increased substantially.

“Meanwhile, having learnt about the discovery of Y-particles, the scientists from Anderson’s group at Cal Tech carried out relevant observations at high-mountain station Mt. Wilson and immediately obtained some promising results. On November 28, 1949, Anderson informed Blackett that his group had obtained some 30 cases of branched tracks and the interpretation of Rochester and Butler some two years before seemed to have been confirmed.

“In the summer of 1950, the instrument installed at Pic du Midi was put into operation and 53 pictures showed tracks of particles with masses between 1000 and $2000 m_e$. The number of the discovered particles increased very rapidly. Names of these particles at first caused confusion until a conference in France on July, 1953. Particles with masses above the pion mass but below the nucleon mass were called mesons, and particles with masses above the nucleon mass were called hyperons.

“The most striking thing was that, having been produced in collisions of high-energy particles as a result of strong (nuclear) interactions in times of 2^{-23} seconds, the new particles had lifetimes from 10^{-9} to 10^{-10} seconds, i.e., they disintegrated very slowly as a result of weak interactions. In other words, the process of their production was explicitly induced by the forces 10^{13} times as strong as the forces of interaction in the course of disintegration. Since the two processes seem to involve identical particles, the enormous difference in the rates of their production was regarded as strange. Because of this ‘strangeness’ in the behavior of the particles, they were called *strange particles*.

“Some theorists proposed an assumption in 1952 that could revolve the above paradox. Their basic idea was that strange particles appear only in groups of two or more particles

simultaneously and suffer strong interaction with each other. However, the particles interact weakly in the processes involving but a single strange particle, for example in the decay of a strange particle. This observation was then confirmed experimentally.

“The obvious question arose whether some law of conservation underlies that process. In 1953, Gell-Mann found the current solution for the problem. He proposed that a new quantum number, which he called ‘strangeness’, should be introduced, and he discovered the law of conservation of strangeness. The strangeness is zero for the conventional particles, nucleons, pions, etc. It differs from zero for hyperons and K-mesons. The strange-particle classification scheme introduced by Gell-Mann made it possible to coordinate the experimental evidence obtained by that time which needed being systematized.”

18.5 The Eightfold Way, Part 2

Skipping many developments after 1953, we come next to Gell-Mann’s 1961 paper on the 8-fold way. In this paper, Gell-Mann managed to arrange a large number of particles known at the time into diagrams given by the irreducible representations of $SU(3)$. To illustrate the idea, we reproduce three of his diagrams:

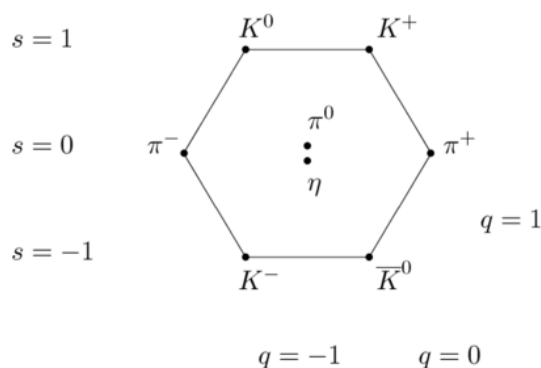


Figure 18.13: Meson Octet

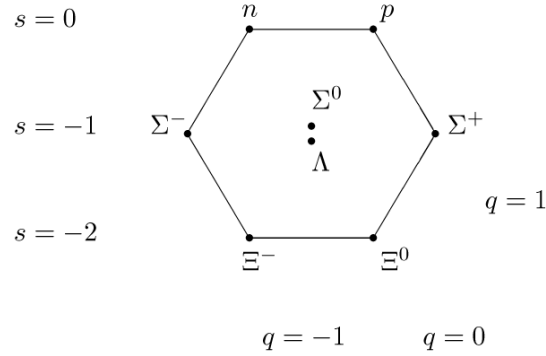


Figure 18.14: Baryon Octet

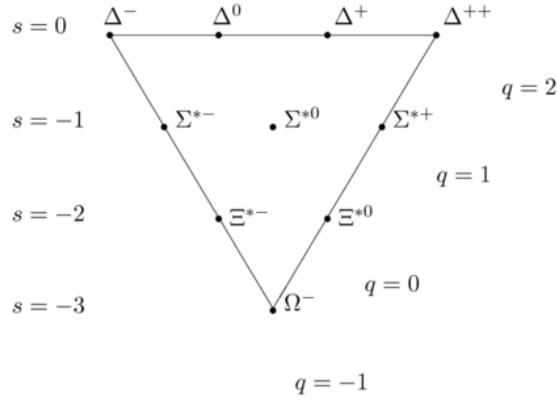


Figure 18.15: Baryon Decuplet

The first Octet organizes spin zero mesons. At first a single particle was placed in the center, but eventually two particles appear there. In that case, the vertices correspond to the $(1, 1)$ representation of $SU(3)$, whose weight space in the center has dimension two. Opposite vertices of the diagram correspond to particle-antiparticle pairs. The horizontal hyperplanes of the diagram (parallel to α in our notation) give lines of constant strangeness, with $s = 1$ at the top, $s = 0$ in the middle, and $s = -1$ at the bottom. The hyperplanes from lower right to upper left (parallel to β in our notation) give lines of constant charge, $q = -1$ at bottom left, $q = 0$ in the middle, and $q = 1$ on the right.

A similar diagram, at the top of this page, gives spin $\frac{1}{2}$ baryons, with the neutron and proton along the top. Horizontal and diagonal hyperplanes still gives lines of constant strangeness and charge.

The final diagram gives spin $\frac{3}{2}$ baryons. This time we have a 10 dimensional representation. When Gell-Mann published this picture, the bottom Ω^- particle had not yet been discovered, so Gell-Mann predicted its existence, strangeness, charge, spin, and approximate mass. A particle with these properties was found at Brookhaven in 1964.

Even before the 1961 paper was published, Gell-Mann lectured about it in India. An audience member asked if there were similar diagrams for the $(1, 0)$ and $(0, 1)$ representations of dimension three. As this person later recalled, Gell-Mann's answer was brief and not very convincing. The audience did not return to that issue.

However, around this time some physicists began to suspect that the proton and neutron were not really fundamental particles, but were built out of more elementary objects which some physicists called *partons*. In 1964, Gell-Mann proposed that these fundamental particles indeed existed and corresponded to the $(1, 0)$ and $(0, 1)$ representations. Each of these could describe three particles. One major difficulty which made Gell-Mann hesitate was that the theory predicted that these particles would have charges $-\frac{1}{3}$ and $\frac{2}{3}$, and that may have been the reason for his hesitation in India. Of course no experimenter had ever seen a particle with non-integer charge, but the theory began to predict that quarks would be essentially impossible to pull apart and examine separately. Gell-Mann discusses how he came to make the proposal in an article quoted in Part 3, below.

In 1967, the Stanford Linear Accelerator became operational, and it made possible the study of electron-nucleon scattering experiments. There is a wonderful document at <https://www.slac.stanford.edu/cgi-bin/getdoc/slac-pub-5724.pdf> that explains how these experiments gradually led to a confirmation of the quark model of the proton and neutron. The very first Stanford experiments suggested no unusual internal structure. But then a series of deeper experiments began to suggest that inside the proton there were very hard smaller objects. Several models of these objects existed, and it took until 1973 for the evidence to begin to favor the quark model, and until 1979 for physicists to generally accept the model.

18.6 The Eight Fold Way, Part 3

Here are quotes from the beginning of the Stanford article mentioned above, which give a flavor of both the difficulty and the excitement of these experiments:

“Just twenty years ago, physicists were beginning to realize that the protons and neutrons at the heart of the atomic nucleus are not elementary particles, after all. Instead, they appeared to be composed of curious pointlike objects called “quarks,” a name borrowed from a line in James Joyce’s novel, *Finnegins Wake*. First proposed in 1964 by Gell-Mann and Zweig, these particles had to have electrical charges equal to $1/3$ or $2/3$ that of an electron or proton. Extensive searches for particles with such fractional charge were

made during the rest of the decade in ordinary matter, in cosmic rays, and at high-energy accelerators, all without success. But surprise results from a series of electron scattering experiments, performed from 1967 through 1973 by a collaboration of scientists from the Massachusetts Institute of Technology (MIT) and the Stanford Linear Accelerator Center (SLAC), began to give direct evidence for the existence of quarks as real, physical entities. For their crucial contributions as leaders of these experiments, which fundamentally altered physicists' conception of matter, Jerome Friedman and Henry Kendall of MIT and Richard Taylor of SLAC were awarded the 1990 Nobel prize in physics.

“By the beginning of the 1960s, physicists had shown that protons and neutrons (known collectively as “nucleons”) had a finite size, as indicated by elastic electron-nucleon scattering experiments of Hofstadter and his Stanford coworkers, but the great majority considered these particles to be “soft” objects with only a diffuse internal structure. Along with pions, kaons and a host of other “hadrons” (particles that feel the effects of the strong nuclear force), they were thought by many to be all equally fundamental, composed of one another in what had been dubbed the “bootstrap model” of strongly interacting particles. Theories that tried to explain the growing variety of hadrons as combinations of a small set of fundamental entities were a definite minority until the MIT-SLAC experiments occurred.

“In 1961 Gell-Mann and Ne’eman introduced a scheme known as $SU(3)$ symmetry that allowed them to impose a measure of order on the burgeoning zoo of hadrons. In this scheme particles with the same spin are grouped together, as if they are just the various distinct states of one and the same entity, similar to the way the proton and neutron can be regarded as merely two different states of the nucleon. Particles with spin 0, like the pions and kaons, form a group of eight “mesons” called an octet, as do another group with spin 1; the proton and neutron are the lightest members of an octet of “baryons” with spin $-\frac{1}{2}$, and there is a group of ten spin $-\frac{3}{2}$ baryons known as a decuplet. In effect, Gell-Mann and Ne’eman did for physics what Mendeleev had done for chemistry, invent a “periodic table” of the hadrons. Using this approach, they even predicted new particles that were later discovered with appropriate properties, buttressing the faith of the physics community in $SU(3)$ symmetry as a correct representation of physical reality.

“In seeking a deeper explanation for the regularities of the $SU(3)$ classification scheme, Gell-Mann and Zweig invented quarks (1,2). In this approach there are three fundamental quarks dubbed “up” or u , “down” or d , and “strange” or s , and their antiparticles, the antiquarks. Mesons are built from a quark plus an antiquark, while baryons are composed of three quarks. The proton is a combination of two up quarks plus a down quark (written uud), for example, while the neutron is made of an up quark plus two downs (udd). By assigning a charge to the up quark of $+\frac{2}{3}e$ (where $-e$ is the charge on the electron) and $-\frac{1}{3}e$ to the other two, the charges on all the known mesons and baryons came out correctly. But the idea of fractional charges was fairly repulsive to physicists of the day. After several years of fruitless searches, most particle physicists agreed that although quarks

might be useful mathematical constructs, they had no innate physical reality as objects of experience.

“The first electron-proton scattering experiment at SLAC, in which electrons with energies up to 20 GeV recoiled elastically from the proton (that is, without breaking it up), gave no evidence for quark substructure. The cross section, or probability, for this process continued to plummet, approximately as the 12th power of the invariant momentum transfer from electron to proton, much as had been observed earlier in the decade at lower energies. This behavior was generally interpreted as evidence for a soft proton lacking any core; it was commonly thought that the existence of such a core would have slowed the rate at which the cross section decreased.

“In the next experiment, performed in late 1967 by the MIT-SLAC collaboration, electrons rebounded inelastically from protons; the energy imparted to the proton either kicked it into a higher energy excited state (such as one of the spin-3/2 baryons) or shattered it entirely. In the latter occurrence, known as “deep inelastic scattering,” the electron rebounded with much less energy. Theoretical analyses of deep inelastic electron-proton scattering made that year by Bjorken suggested that this process might indicate whether there were any constituents inside the proton, but his ideas were not well received initially by the particle physics community.

In the first inelastic experiment, which took place in the autumn of 1967, the 20 GeV spectrometer was used to measure electrons that rebounded from protons at an angle of 6 degrees. The raw counting rates were much higher than had been expected in the deep inelastic region, where the electron imparts most of its energy to the proton, but there was considerable disagreement among the MIT and SLAC physicists as to the proper interpretation of this effect. Electrons can radiate photons profusely as they recoil from a nucleus or pass through matter (in this case, the surrounding hydrogen and target walls); such an effect, which can lower their energy substantially, has to be removed from the raw data before one can assess the underlying physics. These “radiative corrections” were very time-consuming and fraught with uncertainties; they involved measuring cross sections over a large range of E and E' for each value of angle. After the experimental run was over, a computer program was used to deconvolute these data and obtain corrected cross sections at the same kinematics as measured.

When the radiative corrections were completed in the spring of 1968, it became clear that the high counting rates in the deep inelastic region were not due to radiative effects. A way to interpret this unexpected behavior was that the electrons were hitting some kind of hard core inside the target protons. In hindsight, such an observation paralleled the discovery of the atomic nucleus by Ernest Rutherford, in which the probability of large-angle alpha particle scattering from gold atoms was found to be far larger than had been anticipated based on J. J. Thomson’s “plum pudding” model of the atom. At the time, however, there were a few other possible interpretations of the inelastic electron scattering data that had

to be excluded before one could conclude that the MIT-SLAC group had found evidence for constituents inside the proton.”

I refer you to the link above for details of a large number of additional experiments which ultimately began to point in the direction of quarks.

18.7 The Eight Fold Way, Part 4

The history of Gell-Mann’s discovery of the role of $SU(3)$, and then of quarks, is interesting. It is described in Gell-Mann’s own words in Murray Gell-Mann: Selected Papers. There he writes:

“I had continued to play with Yang-Mills theory for the weak interaction all during 1958 and 1959. Early in 1959 I decided that Yang-Mills theory must also be relevant to the strong interaction. But it was very difficult to understand how the two could be related. Late in 1959, in Paris, I decided to find out what the possible generalizations of the Yang-Mills theory could be. Earlier authors had given the case of isotopic spin, what we would nowadays call $SU(2)$. Quantum electrodynamics was, of course, an example of a gauge theory using $U(1)$. We could readily conceive of a gauge theory involving a product of $SU(2)$ factors and $U(1)$ factors. The question was whether any other generalization existed. I worked and worked in my office in the College de France and finally I wrote down, as the necessary and sufficient conditions, the canonical relations

$$[F_i, F_j] = ic_{ijk}F_k$$

where the c_{ijk} are real and totally antisymmetric and the F_i are Hermitian charge operators. I had no idea what were the possible realizations of this formula.

“Every day I would have lunch with my French friends and drink wine, and afterwards I would come back and struggle with drowsiness in my office. I worked through the cases of three operators, four operators, five operators, six operators, and seven operators, trying to find algebras that did not correspond to what we now call products of $SU(2)$ factors and $U(1)$ factors. I got all the way up to seven dimensions and found none, of course. At that point I said ‘that’s enough.’ I did not have the strength after drinking all that wine to try eight dimensions.

“Unfortunately, I did not pay sufficient attention to the identity of one of my regular companions at lunch. It was Professor Serre, one of the world’s greatest experts on Lie algebras. I knew, of course, that Serre was a famous pure mathematician, but I did not know what his specialty was. It never occurred to me to ask him about my equations, and I doubt whether he would have given me the answer if I had asked him. Probably this canonical form for the commutation relations was far too explicit for Serre.”

Later in the same book, Gell-Mann writes

“Now we get to quarks. In early 1963, lecturing at M.I.T., on leave from Caltech, I tried to work out for my lectures the minimal set of fundamental hadronic entities. I found various sets of four objects, but no scheme of that kind looked particularly attractive. But then in March I went to Columbia on a visit, and at the Faculty Club there Bob Serber asked me why it was not possible to use my formula

$$3 \times 3 \times 3 = 1 + 8 + 10$$

to obtain the baryons. I replied that I had tried to get the baryons that way, but that the fundamental entities would then have fractional charges; I showed him on a paper napkin the fractional charges $\frac{2}{3}$ and $-\frac{1}{3}$. He said in effect ‘Oh well, I see why you do not do it then.’ But thinking about the matter afterwards, that day and the next morning, it occurred to me that if the bootstrap approach were correct, then any fundamental hadrons would have to be unobservable, incapable of coming out of the baryons and mesons to be seen individually, and that if they were unobservable, they might as well have fractional charge.”

Still later in the same article, Gell-Mann writes

“I dreaded philosophical discussions about whether particles could be considered real if they were permanently confined. While a colleague of mine falsely claims to have a doctor’s prescription forbidding him to engage in philosophical debates, I really do have one, given to me by a physician who was a student in one of my extension courses at U.C.L.A.

“On my first visit to Japan in the spring of 1964 I encountered the exactly contrary attitude of the group of Marxist theoretical physicists that included Sakata. Yukawa, who did not entirely share their views, had nevertheless helped to place many of these theorists in academic positions in the vicinity of his Institute in Kyoto. A meeting was arranged in my office with Taketani, Ohnuki Maki, and some others. All of these men were strongly opposed to my abstract approach, which they condemned as ‘bourgeois or revisionist idealism.’ They had missed the octet assignment of baryons in flavor $SU(3)$ on account of their *a priori* requirement of concreteness; presumably, they wanted concrete basic objects that could be explained to the masses. It was interesting to see these very intelligent theoretical physicists, working on exactly the right problems with suitable mathematical methods, missing right answers because of their fixed philosophical positions.”

Chapter 19

The Peter-Weyl Theorem

We are now close to the climax of these notes, the Weyl Character Formula. But before that final push, we take a break to explain mathematics of an entirely different character, using analysis rather than geometry or algebra. We will prove the Peter-Weyl theorem, which plays an important role in the proof of the Character Formula.

Fritz Peter lived from 1899 to 1949. He was a PhD student of Hermann Weyl, and together they proved the Peter-Weyl theorem, publishing it in 1927. Later Peter became the headmaster of a secondary school.

The Peter-Weyl theorem asserts that if G is a compact Lie group and $g \in G$ is not the identity, then there is an irreducible representation φ of G such that $\varphi(g) \neq 0$. The same result is true for any compact topological group, essentially by replacing integration over G with the Haar integral. But we only need the Lie group case.

The theorem has many important corollaries. It is used to show that every compact Lie group is isomorphic to a matrix group. The orthogonality relations, which we proved for finite groups, remain true for compact Lie groups. So suppose $\varphi_1, \varphi_2, \dots$ is an (infinite) list of irreducible matrix representations of a compact G . Then the matrix coefficients of these representations are orthonormal functions in $L^2(G)$. By the Peter-Weyl theorem, they form a Hilbert space basis for $L^2(G)$.

Long ago we proved that the characters of irreducible representations of a finite group form a basis for the set of all functions on G that are constant on conjugacy classes. The Peter-Weyl theorem will allow us to prove this for an arbitrary compact Lie group.

19.1 Getting Our Hands on Representations

Finding representations of an arbitrary compact group can be difficult, so proving the Peter-Weyl theorem by writing down a specific representation is a hopeless task. On the other hand, we don't need much information about our representation; if we cannot compute its dimension, nobody cares. So we should look for an abstract method which generates a lot of poorly understood representations.

One such method comes from the theory of linear partial differential equations. We saw an example of that when solving the Schrodinger equation for the hydrogen atom. If an equation is invariant under a group G , then its eigenspaces will also be invariant under G , giving an ample supply of representations. Unfortunately, the theory of linear partial differential equations is deep and difficult, and this approach is hard to carry out.

An important subfield of differential equation theory is about elliptic differential equations. If $\mathcal{D}f = g$ is such an equation on a compact manifold, a deep theorem asserts that the equation has a solution provided g satisfies a finite number of conditions and the solution is unique up to a kernel which is finite dimensional. So the operator $\mathcal{D}$ is almost an isomorphism. In the proof of this result, an approximate inverse to the operator is constructed, called the Green's function. This Green's function is an integral operator of the form $f \rightarrow \int K(x, y)f(y) dy$.

This suggests that it might be better to look for representations by studying the eigenspaces of integral operators. And that is exactly what Peter and Weyl did.

The theory of integral operators was well known in Germany in the 1920's because Hilbert had worked on the theory. Hilbert was motivated by a result in linear algebra. Suppose V is a finite dimensional complex vector space with a Hermitian inner product, and suppose $A : V \rightarrow V$ is a linear transformation. We say that A is *Hermitian* if $\langle Av, w \rangle = \langle v, Aw \rangle$ for all v and w . An easy result says that any Hermitian linear transformation can be diagonalized, and its eigenvectors will be real.

Hilbert worked on generalizing this result to the infinite dimensional case. He began with an infinite dimensional complex vector space V with a Hermitian inner product such that the space was complete with the metric induced by $\|v\|$. Nowadays, we call such spaces *Hilbert spaces*. He then studied Hermitian operators on these spaces. Unfortunately, in some cases these operators have no eigenvectors. But Hilbert concentrated on the case when V is a function space and A is given by an integral operator, and found an almost perfect analogue of the finite dimensional theorem.

A couple of years before the work of Peter and Weyl, a new form of quantum mechanics was invented in which Hermitian operators played a central role. In quantum mechanics, the states of a system correspond to vectors of length one in a Hilbert space V and each experiment corresponds to a Hermitian operator A on V . The possible outcomes of the

experiment are the eigenvalues of A . The Hilbert space must be complex to account for wave-like properties of matter, but the results of experiments are real because Hermitian operators have real eigenvalues.

So the technique used by Peter and Weyl was in the air at the time.

19.2 The Weierstrass Approximation Theorem

In 1885 when he was 70 years old, Weierstrass published a proof of the following theorem:

Theorem 82 *If $f(x)$ is a continuous real-valued function on $[0, 1]$ and $\epsilon > 0$, there is a polynomial $P(x)$ such that $|f(x) - P(x)| < \epsilon$ uniformly on $[0, 1]$.*

Proof: A wonderful proof of this result was discovered by Sergei Bernstein in 1911, based on probability theory. Like all great proofs, the basic idea is easy to state, and the entire proof can be constructed from scratch based on this idea.

Let us play a game. Fix a function $f(x)$ on $[0, 1]$, called the *payoff function*. This function is continuous, and may take both positive and negative values. When it is our turn to play the game, we toss a coin 100 times and compute the percentage of heads, $x = \frac{\text{heads}}{100}$. Then we pay the pot $f(x)$ dollars. If $f(x) < 0$, we take money out of the pot.

The game is enjoyable because sometimes we win money and sometimes we lose. But if we play many times, how much do we expect to win or lose per game on average? It is easy to guess the answer. We expect to get about half heads and half tails, so usually x should be about $\frac{1}{2}$ and our payoff should be about $f(\frac{1}{2})$.

This assumes that we are using a fair coin. If we have an unbalanced coin which gives a head with probability p and a tail with probability $1 - p$, then we expect that the ratio of heads will be about p and the payoff will be about $f(p)$.

If we only toss the coin 10 times, we expect a lot of variation in outcomes and our calculation of average return isn't very accurate. As we toss the coin more often, the accuracy should improve. When we convert these intuitive ideas into rigorous statements about probability, Weierstrauss' theorem will fall out. Let's carry this out.

Assume the probability of heads is p . If we toss the coin twice, the probability of HH is p^2 , the probability of HT is $p(1 - p)$, the probability of TH is $(1 - p)p$ and the probability of TT is $(1 - p)^2$. So the probability of two heads is p^2 , the probability of one head and one tail is $2p(1 - p)$, and the probability of two tails is $(1 - p)^2$. We easily generalize. If we toss n times, the probability of k heads is

$$\binom{n}{k} p^k (1 - p)^{n-k}$$

The *expected value* of our game is the average over all cases of the amount we win in each case times the probability of that case, and so

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}$$

Our intuition is that if n is large and we toss the coin many times in each game, then most of the time the percent of heads is close to p and the expected value is about $f(p)$. So

$$f(p) \sim \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}$$

Write x instead of p . On the left we have $f(x)$ and on the right we have a polynomial in x , and the almost equality is Weierstrauss' result. An advantage of this approach is that we have a specific equation for the approximating polynomials.

Of course we are not done. We still have to *prove* that the two sides are almost equal. This will follow from standard results in probability theory which we will now prove.

Notice that $1 = [p + (1-p)]^n = \sum \binom{n}{k} p^k (1-p)^{n-k}$ by the binomial theorem. In words, the probability of getting something is one.

Let us compute the expected value of the ratio $\frac{k}{n}$. It equals

$$\sum_{k=0}^n \frac{k}{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{k}{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = p \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

In the sum on the right, we can replace $k-1$ by j and sum from $j=0$ to $j=n-1$; notice also that $n-k = (n-1) - (k-1) = (n-1) - j$. So the sum equals 1. Thus the right side is p , and as we guess intuitively, the expected value of $\frac{k}{n}$ is p .

We will use a slightly different form of this equation: $\sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = pn$.

Now we want to compute the standard deviation of $\frac{k}{n}$ and p , which measures the probability that $\frac{k}{n}$ and p are substantially different. This is the sum

$$\begin{aligned} & \sum_{k=0}^n \left(\frac{k}{n} - p\right)^2 \binom{n}{k} p^k (1-p)^{n-k} = \\ & \frac{1}{n^2} \sum k^2 \binom{n}{k} p^k (1-p)^{n-k} - \frac{2p}{n} \sum k \binom{n}{k} p^k (1-p)^{n-k} + p^2 \sum \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Thus we need to find the sum when we multiply the individual probabilities by k^2 ; it is easier to find the sum when we multiply by $k(k-1)$ and we get

$$\sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = p^2 n(n-1) \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k}$$

As earlier, the last sum is just one and so

$$\sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = p^2 n(n-1)$$

and

$$\sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = p^2 n(n-1) + pn$$

We now have enough information to evaluate the sum giving the standard deviation. Putting earlier formulas together we obtain

$$\sum_{k=0}^n \left(\frac{k}{n} - p \right)^2 \binom{n}{k} p^k (1-p)^{n-k} = \frac{p^2 n(n-1) + pn}{n^2} - \frac{2p}{n} pn + p^2 = \frac{p(1-p)}{n}$$

As we expect, this standard deviation goes to zero as we toss the coin more often, so it becomes more and more unlikely that we get an unusual proportion of heads.

Notice that $p(1-p)$ is an inverted parabola which vanishes when $p = 0$ and $p = 1$ and is largest in the center when $p = \frac{1}{2}$. It follows that the right hand side is at most $\frac{1}{4n}$. On the left, we choose an arbitrary $\delta > 0$ and separate terms into those with $\left| \frac{k}{n} - p \right| < \delta$ and those with $\left| \frac{k}{n} - p \right| \geq \delta$. This gives

$$\sum_{\left| \frac{k}{n} - p \right| < \delta} \left(\frac{k}{n} - p \right)^2 \binom{n}{k} p^k (1-p)^{n-k} + \sum_{\left| \frac{k}{n} - p \right| \geq \delta} \left(\frac{k}{n} - p \right)^2 \binom{n}{k} p^k (1-p)^{n-k} \leq \frac{1}{4n}$$

We get an even smaller number on the left by ignoring the first term, replacing each $\left(\frac{k}{n} - p \right)^2$ by δ^2 and noticing that the sum of the remaining terms is just the probability that $\left| \frac{k}{n} - p \right| \geq \delta$. We have thus proved that

$$\text{the probability that } \left| \frac{k}{n} - p \right| \geq \delta \text{ is less than } \frac{1}{4n\delta^2}$$

When playing games with δ and ϵ , it pays to choose these in strict order. Thus suppose I am tossing coins and I want to be almost certain that $\frac{k}{n}$, which is the result that an experiment will produce, is within .01 of the result I expect to see, which is p . So I choose $\delta = .01$. In English I said I want to be “almost certain”. That role could be played by ϵ and if I only want one experiment in a million to be outside the expected bounds, I select $\epsilon = \frac{1}{1,000,000}$. I want $\frac{1}{4n\delta^2} < \frac{1}{1,000,000}$. Rearranging, $n > \frac{1,000,000}{4\delta^2} = \frac{1,000,000,000}{4}$. So if I toss

a fair coin this many times, having nothing else to do, then $\frac{k}{n}$ will be within .01 of $\frac{1}{2}$ except maybe one time in a million.

We are now ready to prove the Weierstrauss theorem. Since $\sum \binom{n}{k} p^k (1-p)^{n-k} = 1$, we have $f(p) = \sum f(p) \binom{n}{k} p^k (1-p)^{n-k} = 1$ and therefore

$$\begin{aligned} \left| f(p) - \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k} \right| &= \left| \sum_{k=0}^n \left(f(p) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} p^k (1-p)^{n-k} \right| \leq \\ &\sum_{|\frac{k}{n}-p|<\delta} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} + \sum_{|\frac{k}{n}-p|\geq\delta} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Suppose we are given $\epsilon > 0$. We will find N such that when $n > N$ the difference between f and the polynomial approximation to f is smaller than ϵ uniformly for all $p \in [0, 1]$.

Since f is continuous on $[0, 1]$, it is uniformly continuous there, so we can find a positive δ such that whenever $x, y \in [0, 1]$ and $|x - y| < \delta$ we have $|f(x) - f(y)| < \frac{\epsilon}{2}$. For this choice of δ , in the first displayed equation above, the expression in the absolute value is smaller than $\frac{\epsilon}{2}$. The remaining terms are all positive. We are only supposed to sum over selected k , but if we summed over all k we would get 1. So the actual sum is at most 1 and the first term is smaller than $\frac{\epsilon}{2}$.

Now consider the second term. This time the only control we have over the absolute value which starts this sum is that it is at most $2M$ where M is the maximum of f on $[0, 1]$. On the other hand, the sum of the remaining terms is the probability that $|\frac{k}{n} - p| \geq \delta$. By the earlier result on standard deviation, this probability is at most $\frac{1}{4n\delta^2}$. So the second term is at most $\frac{2M}{4n\delta^2}$. There is an N such that for all $n > N$ this term is smaller than $\frac{\epsilon}{2}$. The sum of these two terms is smaller than ϵ and we are done. QED.

19.3 The Stone-Weierstrass Theorem

It is natural to generalize the Weierstrauss theorem to higher dimensions. In the plane we suspect that any continuous $f(x, y)$ on a rectangle can be uniformly approximated by a polynomial $P(x, y)$. Perhaps this result is true of more complicated domains as well.

In 1937, the American mathematician Marshall H. Stone published an astonishing theorem which contained these generalizations and much more. Even more astonishing, his theorem was proved by “abstract nonsense” which contained only one difficult step. And the difficult step was a special case of the result proved by Weierstrass. This is the clearest case of “getting a whole lot more for free” that I know of in mathematics.

Stone’s father was Harlan Fiske Stone, chief justice of the U. S. Supreme Court from 1941 to 1946. When he proved the theorem, Stone was a professor of mathematics at Harvard.

After working for the government in World War II, Stone became chair of the mathematics department at Chicago, and made it one of the strongest centers of mathematics in the U.S.

Theorem 83 (Stone-Weierstrass) *Let X be a compact Hausdorff space, and let A be an algebra of continuous real-valued functions on X which contains all constant functions. Thus A is closed under sums, products, and scalar products by real numbers. Suppose A separates points, that is, whenever $p \neq q$ in X , there is an $f \in A$ with $f(p) \neq f(q)$. Then any continuous function on X can be uniformly approximated by an element of A .*

Remark: The algebra A is sort of a red herring. Suppose $f_1, \dots, f_k$ are continuous functions on X which separate points. We could form the algebra of all sums of products of the f_i and apply the theorem to conclude that any continuous function on X can be approximated by a polynomial $P(f_1, f_2, \dots, f_k)$. So “separation of points” is the essential requirement of the theorem.

Remark: In particular, the functions $x_1, x_2, \dots, x_n$ separate points on R^n , so any continuous function on a compact $K \subset R^n$ can be uniformly approximated by a polynomial $P(x_1, x_2, \dots, x_n)$.

Central Idea of the Proof (details follow): Rather than working with A , we immediately introduce its closure $\mathcal{C}$ in the topology of uniform convergence and work there. Suppose p and q are points in X . We can use the separation of points assumption to find an element of $\mathcal{C}$ which takes the values $f(p)$ and $f(q)$ at these points. This function is within ϵ of f in an open set that contains p and q . As we vary q , we get a series of open sets which cover X . So we can find a finite number of these functions, such that for any point of X at least one of the functions is close to f . We then form a new function whose value at each point is the maximum of the values of these finitely many functions at the point. This new function is $f(p)$ at p and never gets any lower than $f(x) - \epsilon$. A similar trick constructs a replacement which is also bounded above by $f(x) + \epsilon$.

This argument rests on proving that if $a(x), b(x) \in \mathcal{C}$, then $\max(a(x), b(x)) \in \mathcal{C}$. Since $\max(a, b) = \frac{(a+b) + |a-b|}{2}$, it suffices to prove the easier result that $f \in \mathcal{C}$ implies $|f| \in \mathcal{C}$. Since $|f| = \sqrt{f^2}$, it suffices to prove that if $f \geq 0$ is in $\mathcal{C}$, so is $\sqrt{f}$. This is a special case of the original Weierstrass theorem.

Proof of Stone-Weierstrass: Consider the set $\mathcal{C}$ of continuous functions on X which can be uniformly approximated by elements of A . This set is clearly closed under addition and scalar multiplication. We will show that it is closed under multiplication using

$$f(x)g(x) - a(x)b(x) = (f(x) - a(x))g(x) + a(x)(g(x) - b(x))$$

If $g \in \mathcal{C}$, then by compactness of X we can find a positive constant so $|g| < N$. If $\epsilon > 0$, find $a(x) \in A$ so $|f(x) - a(x)| < \frac{\epsilon}{2N}$. Then find a positive M such that $|a(x)| < M$, and find $b(x) \in A$ so $|g(x) - b(x)| < \frac{\epsilon}{2M}$. By the displayed equation, $|f(x)g(x) - a(x)b(x)| < \epsilon$.

If $f \in \mathcal{C}$ and $f \geq 0$, we claim that $\sqrt{f} \in \mathcal{C}$. Indeed consider $\sqrt{x}$ on $[0, 1]$ and let $\epsilon > 0$. By the original Weierstrass theorem, there is a polynomial $P(x)$ such that $|\sqrt{x} - P(x)| < \epsilon$ on $[0, 1]$. Find M such that $f < M$. Then $|\sqrt{\frac{f(x)}{M}} - P(\frac{f(x)}{M})| < \epsilon$. Since $f \in \mathcal{C}$, so is $\frac{f(x)}{M}$; since $\mathcal{C}$ is an algebra, $P(\frac{f(x)}{M}) \in \mathcal{C}$. So $\sqrt{\frac{f(x)}{M}} \in \mathcal{C}$ and its product by $\sqrt{M}$, $\sqrt{f(x)}$, is in $\mathcal{C}$.

If $f \in \mathcal{C}$, so is f^2 and therefore so is $\sqrt{f^2} = |f(x)|$. Therefore if $f(x), g(x) \in \mathcal{C}$, so are

$$\min(f, g) = \frac{(f + g) - |f - g|}{2} \quad \max(f, g) = \frac{(f + g) + |f - g|}{2}$$

From here the final steps are easy. Suppose f is a continuous real-valued function on X and $\epsilon > 0$. We will find a function $g \in \mathcal{C}$ with $|f(x) - g(x)| < \epsilon$. It will follow that f can be arbitrarily approximated by functions in $\mathcal{C}$ and thus $f \in \mathcal{C}$.

Suppose x_1 and x_2 are distinct points in X . By hypothesis we can find $a(x) \in A$ with $a(x_1) \neq a(x_2)$. We claim we can construct a new function $A(x) \in X$ with $A(x_1) = f(x_1)$ and $A(x_2) = f(x_2)$. Indeed we select $A(x) = ra(x) + s$ for real numbers r and s ; here we are using the hypothesis the constant functions belong to A . We must solve

$$ra(x_1) + b = f(x_1)$$

$$ra(x_2) + b = f(x_2)$$

Since $a(x_1) \neq a(x_2)$, we can solve these equations for r and s .

Call the functions just obtained $A_{x_1, x_2}(x)$. Fix x_1 for a moment and define

$$U_{x_1, x_2} = \{x \in X | A_{x_1, x_2}(x) < f(x) + \epsilon\}$$

This is an open subset of X containing x_1 and x_2 . As we vary x_2 , these opens sets cover X , so there is a finite subcover. This subcover is associated with a finite number of functions in $\mathcal{C}$. The minimum of these finitely many functions is in $\mathcal{C}$, is smaller than $f(x) + \epsilon$ everywhere, and equals $f(x_1)$ at x_1 . We have one such function for each x_1 . Call this function g_{x_1} .

Let

$$V_{x_1} = \{x \in X | g_{x_1}(x) > f(x) - \epsilon\}$$

This is an open neighborhood of x_1 . As x_1 varies, we obtain an open cover of X . Find a finite subcover and let $g(x)$ be the maximum of all g_{x_i} associated with this finite subcover. Then $g \in \mathcal{C}$. Note that $g(x) > f(x) - \epsilon$ for all x . Since each $g_{x_i}(x)$ is smaller than $f(x) + \epsilon$, $g(x) < f(x) + \epsilon$. So $|f(x) - g(x)| < \epsilon$ and we are done. QED.

Example: Let $X = S^1$ be the circle. Functions on this space can be identified with functions $f(\theta)$ defined on $[-\pi, \pi]$ which satisfy $f(-\pi) = f(\pi)$. Let A be the functions

$$1, \sin(\theta), \cos(\theta), \sin(2\theta), \cos(2\theta), \dots$$

and their finite linear combinations. Surprisingly, this set is an algebra, and thus closed under products. Indeed

$$\sin(k\theta + l\theta) = \sin(k\theta)\cos(l\theta) + \cos(k\theta)\sin(l\theta)$$

$$\sin(k\theta - l\theta) = \sin(k\theta)\cos(l\theta) - \cos(k\theta)\sin(l\theta)$$

and so

$$\sin(k\theta)\cos(l\theta) = \frac{1}{2}(\sin(k+l)\theta + \sin(k-l)\theta)$$

$$\cos(k\theta)\sin(l\theta) = \frac{1}{2}(\cos(k+l)\theta - \cos(k-l)\theta)$$

with similar formulas for $\sin(k\theta)\sin(l\theta)$ and $\cos(k\theta)\cos(l\theta)$. These functions separate points because $\cos\theta + i\sin\theta$ maps $[-\pi, \pi]$ to S^1 and so the two functions only agree on p, q for the equivalent pair $p = -\pi, q = \pi$.

By the Stone-Weierstrass theorem, any continuous function can be uniformly approximated by an expression of the form

$$\frac{a_0}{2} + \sum_{k=1}^n (a_k \cos k\theta + b_k \sin k\theta)$$

This is exactly what Fourier series look like. But there is a catch. We are not claiming that the a_k and b_k are the Fourier coefficients. Indeed there exist continuous functions whose Fourier series do not converge at certain points, and no theorem asserts that the Fourier series of a continuous function converges uniformly to f .

What we have shown is that we can achieve uniform approximation if we choose the coefficients some other way. We will soon come back to the connection of this result with Fourier theory.

Remark: It is far from true that an arbitrary continuous complex-valued function on a compact subset of C can be uniformly approximated by polynomials $P(z)$. Indeed, the uniform limit of holomorphic functions on an open set is holomorphic on that set, so $f(z)$ would have to be holomorphic on the interior of the compact subset. There is no need to pursue this topic here.

But there is an easy complex generalization of the Stone-Weierstrass theorem:

Theorem 84 *Let A be an algebra of continuous complex-valued functions defined on a compact Hausdorff space X . Suppose A contains all constant functions, and suppose $f(x) \in A$ implies $\overline{f(x)} \in A$. Then if A separates points, any continuous complex-valued function on X can be uniformly approximated by an element of A .*

Proof: If $f \in A$, then the real part of f , $\frac{f+\bar{f}}{2}$ is in A and similarly the imaginary part of f is in A . The theorem then follows easily by applying the real version of the theorem to the real-valued functions in A and to the purely imaginary functions in A . QED.

19.4 Hilbert Space

Definition 35 *A Hilbert Space is a complex vector space $\mathcal{H}$ together with a Hermitian inner product $\langle X, Y \rangle$ on the space, such that*

- *the metric space induced by the norm $\|X\| = \sqrt{\langle X, X \rangle}$ is complete*
- *the space has a countable dense subset*

Definition 36 *Let $L^2(\text{positive integers})$ be the set of all sequences $(c_1, c_2, \dots)$ such that $\sum_k |c_k|^2 < \infty$. Define an inner product on this space by*

$$\langle (c_1, c_2, \dots), (d_1, d_2, \dots) \rangle = \sum_k c_k \overline{d_k}$$

Theorem 85 *The inner product converges and makes $L^2(\text{positive integers})$ into a Hilbert space.*

Proof: Since $(|c| - |d|)^2 \geq 0$, we have $|c|^2 + |d|^2 \geq 2|c||d|$ and thus the sum defining the inner product converges. It is easy to see that the set of all $(p_1 + iq_1, p_2 + iq_2, \dots, p_n + iq_n, 0, 0, \dots)$ with $p_i, q_i \in \mathbb{Q}$ is a countable dense subset. Finally, the space is complete. To prove this, it is useful to think of the elements of L^2 as functions from the positive integers to $\mathbb{C}$ such that $\sum |X(k)|^2 < \infty$. Let $X_1, X_2, \dots$ be a Cauchy sequence of such functions. Clearly if we fix k then the complex numbers $X_i(k)$ form a Cauchy sequence. Let this sequence converge to $X(k)$. We will prove that $X \in L^2$ and that $X_n \rightarrow X$.

Before proving those results, notice that there is a uniform bound B for all $\|X_i\|$. Indeed find n_0 such that when $i, j \geq n_0$ we have $\|X_i - X_j\| < 1$. In particular $\|X_i - X_{n_0}\| < 1$ and so

$$\|X_i\| = \|X_i - X_{n_0} + X_{n_0}\| \leq \|X_i - X_{n_0}\| + \|X_{n_0}\| < 1 + \|X_{n_0}\|$$

for all $i \geq n_0$. By increasing this bound if necessary, we can make it also apply to the first $n_0 - 1$ elements.

To show that $X \in L^2$, notice that $\sum_{k=1}^N |X_i(k)|^2 \leq \sum_{k=1}^\infty |X_i(k)|^2 = \|X_i\|^2 \leq B^2$. So

$$\sum_{k=1}^N |X(k)|^2 = \lim_{i \rightarrow \infty} \sum_{k=1}^N |X_i(k)|^2 \leq B^2$$

This holds for all finite N and so $\|X\|^2 \leq B^2$.

A slight modification of this argument shows that $X_i \rightarrow X$. Suppose $\epsilon > 0$. Choose n_0 such that $i, j \geq n_0$ implies $\|X_i - X_j\| < \sqrt{\epsilon}$. Then $\sum_{k=1}^N |X_i(k) - X_j(k)|^2 \leq \|X_i - X_j\|^2 < \epsilon$, and

$$\sum_{k=1}^N |X(k) - X_j(k)|^2 = \lim_{i \rightarrow \infty} \sum_{k=1}^N |X_i(k) - X_j(k)|^2 < \epsilon$$

for any N , and so if $j > n_0$ we have $\|X - X_j\|^2 \leq \epsilon$. QED.

Example: Let $\mathcal{H}$ be the set of continuous complex-valued functions $f(x)$ defined on the interval $[-\pi, \pi]$ such that $f(-\pi) = f(\pi)$. Define a Hermitian inner product by

$$\langle f(x), g(x) \rangle = \int_0^1 f(x) \overline{g(x)} dx$$

The set of all $(p_j + iq_j) \cos kx$ and $(p_j + iq_j) \sin kx$ for p and q rational forms a countable dense subset by a previous application of the Stone-Weierstrass theorem. But this time, unfortunately, the space is not complete. For example, it is easy to find a sequence of continuous functions converging to a function with a jump at the origin. There are two standard ways to fix this. The first is to abstractly define the *completion* of any space satisfying all axioms except the completeness axiom, in the same way that the rational numbers can be completed to form the real numbers. Or we can extend the functions in $\mathcal{H}$ to be all Lebesgue integrable functions such that $\int_{-\pi}^{\pi} |f(x)|^2 dx < \infty$, and agree to identify two functions if they differ only on a set of measure zero. This space can be proved complete. In these notes, we will use the first of these approaches. We will never need the Lebesgue Integral.

Remark: The completeness axiom is automatically true when $\mathcal{H}$ is finite dimensional. Otherwise it is an essential requirement, and many results which follow depend on it.

Remark: The requirement that $\mathcal{H}$ have a countable dense subset is often omitted. This axiom will be true for all the Hilbert spaces used in these notes. If we assume the axiom, then two Hilbert spaces are isomorphic if and only if they have the same dimension, or said another way, *all infinite dimensional Hilbert spaces are isomorphic*. In particular, the two example Hilbert spaces discussed above, which seem completely different, are isomorphic.

Remark: In the case of continuous functions on $[-\pi, \pi]$, there are many inequivalent definitions of convergence. The statement $f_n(x) \rightarrow f(x)$ might mean *uniform convergence*, as it does in the Stone Weierstrass theorem. Or it might mean *pointwise convergence*, as it often does in beginning calculus. Or it might mean *L^2 convergence*, i.e., $\int_{-\pi}^{\pi} |f_n(x) - f(x)|^2 dx \rightarrow 0$, as it does in the Hilbert space example. When I was an undergraduate, I found it annoying that analysis used so many notions of convergence. Now I understand that part of the art is to find exactly the right notion of convergence for a given application. For now, notice that convergence in the Hilbert norm is not the same thing as uniform convergence. On the other hand, if f_n are continuous functions and $f_n(x) \rightarrow f(x)$ uniformly, then certainly they also converge in the L^2 norm.

19.5 Completion of Inner Product Spaces

Later we will consider the vector space $\mathcal{H}$ of all continuous functions on a compact Lie group, and define

$$\langle f, g \rangle = \int_G f(x) \overline{g(x)} dx$$

This is a pre-Hilbert space, but it is not complete. We will complete the space and apply theorems about Hilbert space. In this section, we define this completion process.

Suppose that V is a complex vector space with a Hermitian inner product. Introduce the metric $d(v, w) = \|v - w\|$. Define a space $\mathcal{H}$, called the *completion of V* , to be the set of all Cauchy sequences $(v_1, v_2, \dots)$ in V , but with an equivalence relation making two sequences v_i and w_i equivalent if $v_i - w_i \rightarrow 0$. Map $V \rightarrow \mathcal{H}$ by sending v to the sequence $(v, v, \dots)$. Define a Hermitian inner product in $\mathcal{H}$ by assigning to the sequences $v = (v_1, v_2, \dots)$ and $w = (w_1, w_2, \dots)$ the limit of $\langle v_i, w_i \rangle$.

Theorem 86 *The space $\mathcal{H}$ is a Hilbert space, and the map $V \rightarrow \mathcal{H}$ is a norm preserving injection. The image of this injection is dense in $\mathcal{H}$. If V has a countable dense subset, this subset is also a countable dense subset of $\mathcal{H}$. If V is already a Hilbert space, then $V \rightarrow \mathcal{H}$ is a norm preserving isomorphism.*

Proof: We'll describe the construction in a slightly different way. Consider the set $\mathcal{C}$ of all Cauchy sequences $(v_1, v_2, \dots)$ where the $v_i \in V$. Addition of sequences and multiplication by complex scalars preserve the Cauchy property, so this set is a vector space. The set of all elements $(v_1, v_2, \dots)$ such that $v_i \rightarrow 0$ is a subspace. Let $\mathcal{H}$ be the quotient space. The map $V \rightarrow \mathcal{C}$ defined by $v \rightarrow (v, v, \dots)$ is linear and induces a linear map $V \rightarrow \mathcal{H}$. This map is one-to-one because $(v, v, \dots)$ converges to v , which is zero only if $v = 0$.

If the original V was already complete, then every Cauchy sequence has a limit, so we can define a linear map $\mathcal{C} \rightarrow V$ by mapping each Cauchy sequence to its limit. This map sends the subspace of sequences converging to zero to 0, so it induces a map $\mathcal{H} \rightarrow V$. Clearly

the map $V \rightarrow \mathcal{H} \rightarrow V$ is the identity. The map $\mathcal{H} \rightarrow V \rightarrow \mathcal{H}$ is also the identity; it sends $(v_1, v_2, \dots)$ to the limit v and then to $(v, v, \dots)$, and the difference of these elements is $(v_1 - v, v_2 - v, \dots)$, which converges to zero and thus belongs to the subspace of sequences converging to zero. In short, if V is already complete then $V = \mathcal{H}$.

Suppose $(v_1, v_2, \dots)$ and $(w_1, w_2, \dots)$ are in $\mathcal{C}$. Then the sequence $\langle v_1, w_1 \rangle, \langle v_2, w_2 \rangle, \dots$ is a Cauchy sequence of complex numbers, and so has a limit. Indeed

$$\langle v_i, w_i \rangle - \langle v_j, w_j \rangle = \langle v_i - v_j, w_i \rangle + \langle v_j, w_i - w_j \rangle$$

so

$$\left| \langle v_i, w_i \rangle - \langle v_j, w_j \rangle \right| \leq \|v_i - v_j\| \|w_i\| + \|v_j\| \|w_i - w_j\|$$

If the v_i form a Cauchy sequence, there is an N such that when $i, j \geq N$ we have $\|v_i - v_j\| < 1$. So $\|v_i - v_N\| < 1$ and $\|v_i\| = \|v_i - v_N + v_N\| \leq \|v_i - v_N\| + \|v_N\| < 1 + \|v_N\|$. It follows that there is a common bound B for all $\|v_i\|$ and $\|w_i\|$ and so

$$\left| \langle v_i, w_i \rangle - \langle v_j, w_j \rangle \right| \leq B(\|v_i - v_j\| + \|w_i - w_j\|)$$

and the result immediately follows.

Define the inner product of $(v_1, v_2, \dots)$ and $(w_1, w_2, \dots)$ to be the limit of $\langle v_i, w_i \rangle$. This inner product passes to the quotient space because if $w_i \rightarrow 0$, then $\langle v_i, w_i \rangle$ converges to 0. Indeed $|\langle v_i, w_i \rangle| \leq \|v_i\| \|w_i\|$ and the $\|v_i\|$ are bounded while the $\|w_i\| \rightarrow 0$.

It is easy to show that this inner product on $\mathcal{H}$ has all the Hermitian properties. The only non-trivial case is showing that when $v \neq 0$ then $\langle v, v \rangle \neq 0$. But if the inner product of $(v_1, v_2, \dots)$ with itself is zero, then $\langle v_i, v_i \rangle \rightarrow 0$, so $\|v_i\| \rightarrow 0$, so $v_i \rightarrow 0$ and our element is in the set of all sequences which converge to zero.

The rest of the theorem follows once we know that V is dense in $\mathcal{H}$. Suppose $(v_1, v_2, \dots)$ is in $\mathcal{C}$. If we select a large N then $(v_N, v_N, \dots)$ belongs to V . The squared norm of the difference of these elements is the limit of $\|v_i - v_N\|^2$ as $i \rightarrow \infty$. By the Cauchy property, for any $\epsilon > 0$, N can be chosen so that all of these norms are smaller than ϵ . So any element of $\mathcal{H}$ is within ϵ of an element of V . QED.

19.6 Elementary Results in Hilbert Space Theory

Elementary results about norms:

Theorem 87 *In any space with a Hermitian inner product,*

- $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\operatorname{Re} \langle x, y \rangle$
- *Parallelogram Law:* $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$
- *Pythagorean Theorem:* if $x_1, \dots, x_n$ is a finite orthogonal set, $\|\sum x_i\|^2 = \sum \|x_i\|^2$

Proof: $\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + (\langle x, y \rangle + \langle y, x \rangle) + \langle y, y \rangle$. Notice that the middle term is $\langle x, y \rangle + \overline{\langle x, y \rangle} = 2\operatorname{Re} \langle x, y \rangle$.

We have $\|x + y\|^2 + \|x - y\|^2 = \|x\|^2 + 2\operatorname{Re} \langle x, y \rangle + \|y\|^2 + \|x\|^2 + 2\operatorname{Re} \langle x, -y \rangle + \|y\|^2$ and the two real parts cancel.

Finally $\langle \sum x_i, \sum x_j \rangle = \sum \langle x_i, x_j \rangle = \sum \langle x_i, x_i \rangle$. QED

Definition: Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces and $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ is a linear transformation. We say that A is *bounded* if there is a constant M such that $\|A(v)\| \leq M\|v\|$ for all $v \in \mathcal{H}_1$.

If both spaces are finite dimensional, A is given by a matrix and thus is clearly continuous. But in the infinite dimensional case, A is not necessarily continuous.

Theorem 88 *A linear transformation is continuous if and only if it is bounded.*

Proof: The transformation A is continuous if whenever $v_i \rightarrow v$, we have $A(v_i) \rightarrow A(v)$. The first statement is equivalent to $v_i - v \rightarrow 0$ and the second statement is equivalent to $A(v_i - v) \rightarrow 0$. In short, continuity at the origin implies continuity everywhere.

Suppose A is bounded. Then it is continuous at the origin, because $v_i \rightarrow 0$ implies $\|v_i\| \rightarrow 0$, so $\|A(v_i)\| \leq M\|v_i\| \rightarrow 0$ and so $A(v_i) \rightarrow 0$.

Conversely suppose A is continuous at the origin. Then for each $\epsilon > 0$ there is a $\delta > 0$ such that $\|v\| \leq \delta$ implies $\|A(v)\| \leq \epsilon$. If $v \neq 0$, the norm of $\frac{\delta v}{\|v\|}$ is equal to δ , so the norm of $A\left(\frac{\delta v}{\|v\|}\right) = \frac{\delta}{\|v\|}A(v)$ is less than ϵ and the norm of $A(v)$ is less than $\frac{\epsilon}{\delta}\|v\|$. QED.

Remark: In finite dimensions, every subspace $W \subset V$ is closed. This is false in infinite dimensions. For example, if V is an inner product space which is not complete, let $\mathcal{H}$ be its completion and notice that $V \subset \mathcal{H}$ is a subspace which is not equal to $\mathcal{H}$. But its closure is all of $\mathcal{H}$.

Theorem 89 *Let $\mathcal{H}$ be a Hilbert space and let $V \subset \mathcal{H}$ be a closed subspace. Then*

$$\mathcal{H} = V \oplus V^\perp$$

Proof: The statement $V \cap V^\perp = \{0\}$ is trivial. So it suffices to show that every element of $\mathcal{H}$ is a sum of elements in these subspaces.

Let $h \in \mathcal{H}$. We will prove that there is an element $v \in V$ that is closest to h . If so, suppose w is another element in V and consider the expression

$$\|h - v + tw\|^2 = \|h - v\|^2 + 2t \operatorname{Re} \langle h - v, w \rangle + t^2 \|w\|^2$$

This expression has a minimum at $t = 0$. Therefore its derivative there is zero. So $2 \operatorname{Re} \langle h - v, w \rangle = 0$. This holds for all $w \in V$, so it holds for iw , so

$$2 \operatorname{Re} \langle h - v, iw \rangle = 2 \operatorname{Im} \langle h - v, w \rangle = 0$$

and thus $\langle h - v, w \rangle = 0$ for all w . So $h - v \in V^\perp$ and $h = v + (h - v) \in V \oplus V^\perp$.

It remains to show that such a v exists. Let B be the greatest lower bound of all $\|h - v\|$ for $v \in V$. Then $\|h - v\| \geq B$ for all v , but for each positive n we can find $v_n \in V$ with $\|h - v_n\| < B + \frac{1}{n}$. This gives a sequence $v_n \in V$ with $\|h - v_n\| \rightarrow B$. We will show that v_n is a Cauchy sequence. If so, this sequence converges to some $v \in \mathcal{H}$ and $v \in V$ because V is closed, and $\|h - v\| = B$.

To conclude, apply the parallelogram law to $h - v_m$ and $h - v_n$. We have

$$2\|h - v_m\|^2 + 2\|h - v_n\|^2 = \|(h - v_m) + (h - v_n)\|^2 + \|v_m - v_n\|^2$$

The term just right of the equal sign is

$$\left\| 2 \left(h - \frac{v_m + v_n}{2} \right) \right\|^2 = 4 \left\| h - \frac{v_m + v_n}{2} \right\|^2$$

We conclude that

$$2\|h - v_m\|^2 + 2\|h - v_n\|^2 - 4 \left\| h - \frac{v_m + v_n}{2} \right\|^2 = \|v_m - v_n\|^2$$

Let us estimate the left hand side. The first and second terms are only slightly larger than $2B^2$ and $2B^2$, and the sum of these terms is only slightly larger than $4B^2$. Then we subtract a term which is at least $4B^2$ and probably larger. So the difference is only slightly larger than zero. It follows that our sequence is Cauchy.

If the reader is queasy, we can use respectable language. Suppose $\epsilon > 0$. Since $\|h - v_m\|$ converges to B , N exists such that $m, n > N$ implies $B^2 \leq \|h - v_m\|^2 \leq B^2 + \frac{\epsilon}{4}$. In the previous displayed equation, the term being subtracted is at least $4B^2$, and the left side is thus at most $2B^2 + \frac{\epsilon}{2} + 2B^2 + \frac{\epsilon}{2} - 4B^2 = \epsilon$. So $\epsilon \geq \|v_m - v_n\|^2$. QED.

Remark: Now suppose $\varphi : \mathcal{H} \rightarrow C$ is a continuous linear map. By definition, the set of such maps is the dual space of $\mathcal{H}$. It is easy to find such maps, because if $w \in \mathcal{H}$, then

$$v \rightarrow \langle v, w \rangle$$

is such a map.

Theorem 90 *If $\mathcal{H}$ is a Hilbert space, then every continuous linear map $\varphi : \mathcal{H} \rightarrow C$ has the form $v \rightarrow \langle v, w \rangle$ for some $w \in \mathcal{H}$.*

Proof: If $\varphi = 0$, then $\varphi(v) = \langle v, 0 \rangle$. Otherwise let $V = \text{Ker}(\varphi)$. Clearly V is a closed subspace of $\mathcal{H}$, so $\mathcal{H} = V \oplus V^\perp$. Clearly $\varphi : V^\perp \rightarrow C$ is an isomorphism. Select an element $w \in V^\perp$ of length one. We claim there is a λ such that $\varphi(v) = \langle v, \lambda w \rangle$. Both sides agree on V , and we need only choose λ so they agree on w . So $\varphi(w) = \langle w, \lambda w \rangle = \bar{\lambda}$ and we must choose $\lambda = \overline{\varphi(w)}$. QED.

Remark: Suppose $A : \mathcal{H} \rightarrow \mathcal{H}$ is a linear transformation. In the finite dimensional case, A is given by a matrix, and the conjugate transpose of this matrix defines another transformation $A^* : \mathcal{H} \rightarrow \mathcal{H}$. Moreover, for all v and w we have

$$\langle Av, w \rangle = \langle v, A^*w \rangle$$

In the infinite dimensional case, linear transformations do not necessarily possess nice matrix forms. However, we can still obtain A^* because

Theorem 91 *If $A : \mathcal{H} \rightarrow \mathcal{H}$ is a bounded linear transformation, there is a unique bounded linear transformation $A^* : \mathcal{H} \rightarrow \mathcal{H}$ such that for all v and w we have*

$$\langle Av, w \rangle = \langle v, A^*w \rangle$$

Proof: Fix w and consider the map $\varphi : \mathcal{H} \rightarrow C$ defined by $\varphi(v) = \langle Av, w \rangle$. This map is continuous and linear and thus has the form $\langle v, A^*w \rangle$ for a unique $A^*w \in \mathcal{H}$. Thus we get a unique map $\mathcal{H} \rightarrow \mathcal{H}$. This map is linear because $\langle Av, w_1 + w_2 \rangle = \langle Av, w_1 \rangle + \langle Av, w_2 \rangle$ and therefore $\langle v, A^*(w_1 + w_2) \rangle = \langle v, A^*(w_1) \rangle + \langle v, A^*(w_2) \rangle$ for all v . So $A^*(w_1 + w_2) = A^*(w_1) + A^*(w_2)$. Moreover, $\langle Av, \lambda w \rangle = \bar{\lambda} \langle Av, w \rangle$, so $\langle v, A^*(\lambda w) \rangle = \bar{\lambda} \langle v, A^*(w) \rangle = \langle v, \lambda A^*(w) \rangle$, so $A^*(\lambda w) = \lambda A^*(w)$.

Recall that A is bounded, so there is a B such that $\|Av\| \leq B\|v\|$. Using this, we will show that A^* is bounded. Indeed $|\langle Av, w \rangle| \leq \|Av\| \|w\| \leq B\|v\| \|w\|$ and so

$$|\langle v, A^*w \rangle| \leq B\|v\| \|w\|$$

Substitute $v = A^*w$ to conclude that $|\langle A^*w, A^*w \rangle| \leq B\|A^*w\| \|w\|$ and thus

$$\|A^*w\|^2 \leq B\|A^*w\| \|w\| \quad \text{and so} \quad \|A^*w\| \leq B\|w\|$$

QED.

Definition 37 *A bounded linear operator $A : \mathcal{H} \rightarrow \mathcal{H}$ is said to be Hermitian if $A = A^*$.*

Theorem 92 *If λ is an eigenvalue of a Hermitian operator A , then λ is real. If v and w are eigenvectors associated to unequal eigenvalues, then v and w are orthogonal.*

Proof: Suppose $Av = \lambda v$ for nonzero v . Then $\langle Av, v \rangle = \langle \lambda v, v \rangle = \lambda \|v\|^2$ but this is also $\langle v, A^*v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \|v\|^2$.

Now suppose $Av = \lambda v$ and $Aw = \tau w$ and A is Hermitian. Then $\langle Av, w \rangle = \langle v, Aw \rangle$ and so $\langle \lambda v, w \rangle = \langle v, \tau w \rangle$. Since both λ and τ are real, we conclude that $\lambda \langle v, w \rangle = \tau \langle v, w \rangle$. So either $\langle v, w \rangle = 0$ or else $\lambda = \tau$. QED.

Remark: A central theorem in linear algebra states the following:

Theorem 93 *Let $A : \mathcal{H} \rightarrow \mathcal{H}$ be a Hermitian operator on a finite dimensional Hilbert space. Then there is an orthonormal basis $e_1, \dots, e_n$ for $\mathcal{H}$ such that each e_i is an eigenvector of A with real eigenvalue.*

Proof: We prove this by induction on the dimension of $\mathcal{H}$; it is trivial in the one dimensional case. Start the induction step by finding an eigenvector of A , $Av = \lambda v$. This exists by the fundamental theorem of algebra, since our scalars are complex numbers. Let V be the set of all $v \in \mathcal{H}$ such that $Av = \lambda v$. This is clearly a subspace of $\mathcal{H}$, so $\mathcal{H} = V \oplus V^\perp$. We will prove that A maps $V^\perp$ back to itself. If so, we can apply induction to A on $V^\perp$ and the theorem immediately follows.

Therefore assume $w \in V^\perp$. Then for all $v \in V$ we have

$$\langle v, Aw \rangle = \langle Av, w \rangle = \langle \lambda v, w \rangle = \lambda \langle v, w \rangle = 0$$

So $Aw \in V^\perp$. QED.

Remark: This theorem is false in the infinite dimensional case. But everyone who has taken a quantum mechanics course will testify that the physicists pretend that it is true. In the quantum theory, vectors in $\mathcal{H}$ correspond to possible states of a physical system. Each experiment is associated with a Hermitian operator. The eigenvalues of this operator are the possible physical outcomes of the theory. In the quantum case, it is absolutely essential that $\mathcal{H}$ be a complex vector space, because that leads to interference and wave-like properties for particles. But the eigenvalues are all real, so the predictions of the theory are meaningful.

Remark: We will discover a special case when the theorem is true even in infinite dimensions. This special case is the key to the Peter-Weyl theorem. So keep the previous theorem in mind. It is a beacon that guides our work.

19.7 Hilbert Space Bases

Theorem 94 (Bessel's Inequality) *Let $\{e_\alpha\}$ be an arbitrary set of orthonormal vectors in a Hilbert space $\mathcal{H}$. Then for all $v \in \mathcal{H}$ we have*

$$\sum_{\alpha} |\langle v, e_\alpha \rangle|^2 \leq \|v\|^2$$

Proof: For any finite subset of the α , the expression below is greater than or equal to zero:

$$\begin{aligned} \|v - \sum_{\alpha} \langle v, e_\alpha \rangle e_\alpha\|^2 &= \|v\|^2 - 2 \sum_{\alpha} \overline{\langle v, e_\alpha \rangle} \langle v, e_\alpha \rangle + \sum_{\alpha, \beta} \overline{\langle v, e_\alpha \rangle} \langle v, e_\beta \rangle \langle e_\alpha, e_\beta \rangle \\ &= \|v\|^2 - \sum_{\alpha} |\langle v, e_\alpha \rangle|^2 \end{aligned}$$

For any positive constant k , only finitely many $|\langle v, e_\alpha \rangle|$ can be greater than k , else we could find a finite set of α contradicting the inequality. Consequently, at most countably many $\langle v, e_\alpha \rangle$ are non-zero. If we restrict to summing over these countably many terms, standard arguments imply that the inequality still holds for the full infinite sum. QED.

Definition 38 *An orthonormal basis for a Hilbert space is a maximal collection $\{e_\alpha\}$ of orthonormal vectors, in the sense that no additional element can be added to the collection.*

Theorem 95 *An orthonormal basis for a Hilbert space always exists.*

Proof: We will give two proofs. The first uses Zorn's lemma and gives little information about the size of the basis. The second is based on the condition that $\mathcal{H}$ have a countable dense subset, avoids Zorn's lemma, and shows that a countable basis exists.

Consider the collection of orthonormal subsets of $\mathcal{H}$, ordered by inclusion. Then any chain has an upper bound, for we can form the union of the elements of the chain. Indeed if e_α and e_β are in this union, then one comes first and both are in some orthonormal set, so the two elements have length one and are orthogonal. By Zorn's lemma, there is a maximal element. Done.

If $\mathcal{H}$ has a countable dense subset, enumerate this subset as $\{v_1, v_2, v_3, \dots\}$. Then apply the Gram-Schmidt process step by step. If $v_1 \neq 0$, the process replaces v_1 with a normalized vector in the same direction, e_1 . If $v_2, \dots, v_{k-1}$ are in the space generated by e_1 , ignore these terms. As soon as a vector v_k not dependent on e_1 occurs, apply the Gram-Schmidt process to e_1, v_k to obtain e_1, e_2 . Continue the process. In the end, either a finite or a countable infinite orthonormal set will be produced, and any v_i is in a subspace generated by finitely many of these countably many vectors. We claim the set is maximal. If not,

we can find a vector e of length one perpendicular to all of the e_i and thus to all of the v_i . But then $\|e - v_i\|^2 = \|e\|^2 + \|v_i\|^2 \geq 1$ so the v_i are not dense in $\mathcal{H}$. QED.

Remark: From now on we assume that $\mathcal{H}$ has a countably dense subset and thus that a countable basis exists. Any basis we mention from now on is required to be countable. Of course the basis is not unique, but we will discover that for many concrete examples we can find a very explicit basis.

Theorem 96 *Let $\mathcal{H}$ be a Hilbert Space, with basis $\{e_1, e_2, \dots\}$. This set can be finite or countable. Then*

- every $v \in \mathcal{H}$ can be written as a finite or convergent sum $v = \sum c_i e_i$
- the coefficients in this sum are unique; indeed $c_i = \langle v, e_i \rangle$
- $\sum |c_i|^2 = \|v\|^2$
- conversely if c_i are complex and $\sum |c_i|^2 < \infty$, there is a $v \in \mathcal{H}$ with this expansion
- the resulting map $\mathcal{H} \rightarrow L^2(\text{positive integers})$ is a Hilbert space isomorphism
- consequently, all infinite dimensional Hilbert spaces are isomorphic
- if $f \in \mathcal{H}$ then $\sum_{i=1}^N c_i e_i$ is the best approximation to f using these e_i , in the sense that for any other coefficients d_i we have

$$\|f - \sum_{i=1}^N c_i e_i\| \leq \|f - \sum_{i=1}^N d_i e_i\|$$

Proof: If $\sum_{i=1}^n c_i e_i \rightarrow v$, then $\langle \sum_{i=1}^n c_i e_i, e_j \rangle \rightarrow \langle v, e_j \rangle$, but by the orthonormal condition, the sum on the left is c_j for all large n , and thus $c_j = \langle v, e_j \rangle$.

Conversely if $c_i = \langle v, e_i \rangle$, then by Bessel's inequality, $\sum |c_i|^2 \leq \|v\|^2$ and consequently $\sum |c_i|^2$ converges. It follows that the finite sums of the form $\sum_{i=1}^N c_i e_i$ form a Cauchy sequence, and thus have a limit in $\mathcal{H}$. If this limit is not v , then $v - \sum_{i=1}^\infty c_i e_i$ is not zero and yet is perpendicular to all e_j . So the basis is not a maximal orthonormal set because the normalization of this difference could be added to the set.

So $v = \lim_{N \rightarrow \infty} \sum_{i=1}^N c_i e_i$ and consequently

$$\langle v, v \rangle = \lim_{N \rightarrow \infty} \left\langle \sum_{i=1}^N c_i e_i, \sum_{j=1}^N c_j e_j \right\rangle = \lim_{N \rightarrow \infty} \sum_{i=1}^N |c_i|^2$$

Notice that $f - \sum_{i=1}^N \langle f, e_i \rangle e_i$ is orthogonal to each e_j for $1 \leq j \leq N$. Consequently this sum is orthogonal to $\sum_{i=1}^N (d_j - c_j) e_j$ and so

$$\|f - \sum d_i e_i\|^2 = \left\| \left(f - \sum c_i e_i \right) + \left(\sum (d_i - c_i) e_i \right) \right\|^2 = \|f - \sum c_i e_i\|^2 + \left\| \sum (d_i - c_i) e_i \right\|^2$$

Everything else is trivial. QED.

19.8 Application To Fourier Series

Consider the vector space of continuous complex-valued functions $f(\theta)$ on the circle. Another way to think of this space is as the set of continuous functions on $[-\pi, \pi]$ which take the same values at the endpoints. Or think of it as the set of continuous functions on $\mathbb{R}$ which are periodic with period 2π . In all three cases, the inner product we will use is

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx$$

The vector space is an inner-product space but not a Hilbert space because it is not complete. Complete it and call the resulting Hilbert space $L^2(S^1)$.

A famous exercise at the beginning of any discussion of Fourier series asserts that the following functions are orthonormal: $\frac{1}{\sqrt{2}}, \cos k\theta$ for $k \geq 1$, $\sin k\theta$ for $k \geq 1$. What is not so clear is that this set is *complete*, that is, forms a basis for $L^2(S^1)$. We will prove that it is a basis; this proof is a nice illustration of the interplay between the Stone-Weierstrass theorem and Hilbert space theory. The technique we will use is often used to show completeness of orthonormal sets.

Once we know we have a basis, we can write every element $f \in L^2(S^1)$ as a sum of coefficients multiplying basis elements. It is conventional to denote the coefficients of $\cos k\theta$ by a_k and the coefficients of $\sin k\theta$ by b_k . Then we get the famous formulas

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos k\theta d\theta \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin k\theta d\theta$$

The constant term is $c_0 \frac{1}{\sqrt{2}}$ where $c_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \frac{1}{\sqrt{2}} d\theta$, or $\frac{a_0}{2}$ where $a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) d\theta$. Notice that this is a special case of the formula for a_k . So usually people write

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\theta + b_k \sin k\theta)$$

Notice carefully that this formula will hold in the Hilbert space, so convergence does not mean uniform convergence or pointwise convergence, but instead *convergence in the mean*.

This is actually the natural form of convergence for Fourier series, where the theorems become easy and straightforward. A typical Fourier series course goes on to prove trickier theorems about situations when convergence is pointwise, or uniform.

Theorem 97 *The above orthonormal set is complete, and thus forms a basis for the Hilbert space $L^2(S^1)$.*

Proof: It is inconvenient to have one notation for the cosine terms and another for the sin terms, so let us denote our orthonormal basis by $e_1(\theta), e_2(\theta)$, etc. Temporarily assume that f is continuous, rather than a general element of the completion. By the Stone-Weierstrass theorem, we can approximate f uniformly and arbitrarily closely by a finite linear combination of the basis vectors. So we can find coefficients d_i such that $\left|f(\theta) - \sum_{j=1}^N d_j e_j(\theta)\right| < \epsilon$. Therefore

$$\left\|f(\theta) - \sum_{j=1}^N d_j e_j(\theta)\right\|^2 = \frac{1}{\pi} \int_{-\pi}^{\pi} \left|f(\theta) - \sum_{j=1}^N d_j e_j(\theta)\right|^2 d\theta \leq 2\epsilon^2$$

By a previous theorem, the best approximation in the norm to f is the Fourier approximation, so

$$\left\|f(\theta) - \sum_{j=1}^N c_j e_j(\theta)\right\|^2 \leq \left\|f(\theta) - \sum_{j=1}^N d_j e_j(\theta)\right\|^2 \leq 2\epsilon^2$$

It follows that the Fourier series of f converges to f in the norm.

To finish the proof, we need to prove the same result for f in the completion $\mathcal{H}$. This will follow from our general basis theorem if we prove that the orthonormal set we have chosen is maximal. If not, we can find e in the completion $\mathcal{H}$ of length one and orthogonal to each of our e_j . It follows that $\langle f - \sum_{i=1}^N c_i e_i, e \rangle = \langle f, e \rangle$ is arbitrarily small, so $\langle f, e \rangle = 0$. Thus e is orthogonal to a dense subset of $\mathcal{H}$, and so orthogonal to all of $\mathcal{H}$, which is impossible. QED.

Remark: Fourier series are often written in terms of $\sum c_k e^{ik\theta}$ rather than as sines and cosines. Exactly the same argument shows that $e^{ik\theta}$ for $k \in \mathbb{Z}$ forms a basis for $L^2(S^1)$. This time the inner product should be $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) \overline{g(\theta)} d\theta$.

Remark: It follows that the completion of the space V of continuous function on S^1 is isomorphic to $L^2(\text{positive integers})$. This in spite of the fact that elements of V are determined by uncountably many numbers $f(x)$, while elements of L^2 are determined by countable many coordinates. In some sense, a continuous object over R has been made discrete over Z . The proof of this isomorphism requires a non-trivial detour through Fourier series; without that trick, we would be clueless how to proceed.

19.9 Integral Operators

We are now going to begin the study of linear transformations $A : \mathcal{H} \rightarrow \mathcal{H}$ and the duality we met in the previous section between continuous objects over R and discrete objects over Z will raise its head again. But this time the duality is not so easily handled, and a full understanding of these two worlds would require a much deeper course in Hilbert Space than these notes provide. This section has no concrete results and is here simply to provide a glimpse of the difficulties ahead.

Once bases are available, we can write linear transformations as (infinite) matrices. Indeed $A(e_i) = \sum a_{ji}e_j$ and $A(\sum c_i e_i) = \sum a_{ji}c_i e_j$ as usual. The matrix extends infinitely far to the right and down. Various sums in this matrix must converge, and it is tricky to write down the precise conditions that a matrix must satisfy. Matrices provide a way to think of linear transformations from the “discrete” point of view, but are used surprisingly little in the actual theory.

An analogy provides a way to think of linear transformations from the “continuous” point of view. Consider the space V of continuous complex valued functions on $[0, 1]$ with inner product

$$\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx$$

and let $\mathcal{H}$ be the completion of this space. We can imagine that this space has as basis the Dirac delta functions δ_x , where this function is zero except at x , and just infinite enough at x to have total integral one. Of course there are no such functions, but we’ll continue to dream. Then we could write vectors in V as linear combinations of these delta functions, but the linear combinations would be integrals rather than sums. Indeed, we might guess that the values of f are its coefficients in this linear combination and the combination is given by an integral

$$f = \int_0^1 f(x) \delta_x dx$$

If we continue thinking in these terms, then the matrix of a linear transformation would be given by a function $A(x, y)$ on $[0, 1] \times [0, 1]$ rather than an infinite matrix, and the formula for such a transformation would change from $A(v) = \sum a_{ij}c_j e_i$ to

$$A(f)(x) = \int_0^1 A(x, y) f(y) dy$$

Transformations of this form are called *integral transforms* and provide one way of thinking of the theory from the continuous point of view.

Our main goal is to generalize to infinite dimensions the theorem that a finite Hermitian matrix can be diagonalized. The full generalization is called the *spectral theorem* because

the eigenvalues of a transformation often give various spectra studied in physics. In quantum mechanics and elsewhere, physicists speak of the *discrete spectrum* and the *continuous spectrum*, and the spectral theorem has to deal with both sorts of eigenvalues.

The easiest example of an operator with a continuous spectrum is given by the Hilbert Space $\mathcal{H}$ obtained by completing the space V of all continuous complex-valued functions on $[0, 1]$, with operator

$$A(f) = xf(x)$$

This transformation is bounded because

$$\|xf(x)\|^2 = \int_0^1 x^2 |f(x)|^2 dx \leq \int_0^1 |f(x)|^2 dx = \|f\|^2$$

and therefore it extends to a bounded transformation from $\mathcal{H}$ to $\mathcal{H}$. Moreover $A = A^*$ because $\langle Af, g \rangle = \int_0^1 xf(x)g(x) dx$ and $\langle f, Ag \rangle = \int_0^1 f(x)g(x)x dx$ and $x = \bar{x}$. So the required equation holds on a dense subset of $\mathcal{H}$ and consequently on all of $\mathcal{H}$.

According to a generalized spectral theorem, then, this operator should be diagonalizable. But unfortunately, it has no eigenvectors at all. Indeed if $f \in V$ and $A(f) = \lambda f$, then $xf = \lambda f$ and so $(x - \lambda)f(x) = 0$. So $f(x) = 0$ except possibly when $x = \lambda$. By continuity, $f(\lambda) = 0$ and thus $f = 0$.

You might wonder whether an element of the completion could be an eigenvector. If we know what the completion actually is, it is easy to show that the answer is still no. The completion is the set of Lebesgue measurable functions on $[0, 1]$ such that $\int_0^1 |f(x)|^2 dx$ is finite, where this is the Lebesgue integral. In the completion, we identify two functions if they agree except on a set of measure zero. If such an f were an eigenvector of A , then $Af(x) = xf(x) = \lambda f(x)$ and so $(x - \lambda)f(x) = 0$. So $f = 0$ except possibly at $x = \lambda$ and thus $f = 0$ almost everywhere, so f represents $0 \in \mathcal{H}$.

No proof in these notes will require Lebesgue integration, but perhaps it is legal to use it for an example. However, there must be a proof that A has no eigenvectors in the completion which does not require knowing the completion explicitly. Can someone supply that proof?

In some sense this V does have a basis, namely the delta functions δ_x discussed earlier. Notice that δ_x is an eigenvector, since $x\delta_y$ is zero except at y itself, where it equals $y\delta_y$. None of this makes rigorous sense at the moment, but it suggests how the spectral theorem will deal with continuous parts of the spectrum and with operators which have no eigenvalues.

Luckily, there is a simple path out of these conundrums. This path would be too special to satisfy a Hilbert space expert, but it will give us exactly the tool we need. We will define a special type of bounded operator called a *compact operator*. And we will prove

that every compact Hermitian operator can be diagonalized in the sense that we can find a Hilbert space orthonormal basis e_i such that $A(e_i) = \lambda_i e_i$. The definition of a compact operator is technical and opaque, but we will prove that any operator of the form $Af(x) = \int_0^1 A(x, y)f(y) dy$ is compact if $A(x, y)$ is, say, continuous; such an operator is Hermitian if $A(y, x) = \overline{A(x, y)}$. These two results are all we need to prove the Peter-Weyl theorem.

19.10 Compact Operators

The n -dimensional sphere S^n is compact. There is a natural generalization of such spheres to real and complex Hilbert spaces: $\{h \in \mathcal{H} \mid \|h\| = 1\}$. Surprisingly, the sphere is not compact in infinite dimensions. Indeed, pick an orthonormal basis $e_1, e_2, \dots$, noticing that each element is in the sphere. When $i \neq j$, $\|e_i - e_j\|^2 = \|e_i\|^2 + \|e_j\|^2 = 2$ by Pythagorus, and thus $\|e_i - e_j\| = \sqrt{2}$. So no subsequence of this sequence can be a Cauchy sequence and thus no subsequence can converge.

Therefore, we must be extraordinarily careful when we discuss compact subsets of Hilbert Space. That is particularly true since continuous operators are bounded, and thus for any bounded domain $\mathcal{D}$, the set $A(\mathcal{D})$ will be bounded. It is all too easy to add “and therefore the closure of this set is closed and bounded and thus compact.” This is true in finite dimensions, but false in infinite dimensions.

Definition 39 *A bounded operator $A : \mathcal{H} \rightarrow \mathcal{H}$ is compact if whenever $\mathcal{D} \subset \mathcal{H}$ is a bounded set, the closure of $A(\mathcal{D})$ is compact.*

Remark When we apply this definition, we usually think of it in the following more practical way: if $d_1, d_2, d_3, \dots$ is a bounded sequence of elements of $\mathcal{H}$, then the sequence $A(d_1), A(d_2), A(d_3), \dots$ has a convergent subsequence.

Remark: If the image of A is finite dimensional, then A is certainly compact because the closure of $A(\mathcal{D})$ will be a closed and bounded set, and in finite dimensions such sets are compact.

Remark: We can get other examples by modifying this example slightly. For example, define $A : L^2(\text{positive integers}) \rightarrow L^2(\text{positive integers})$ by

$$A(c_1, c_2, c_3, \dots) = (c_1, \frac{1}{2}c_2, \frac{1}{3}c_3, \dots)$$

Suppose we had a bounded sequence of such vectors

$$\begin{aligned} d_1 &= (c_{11}, c_{12}, c_{13}, \dots) \\ d_2 &= (c_{21}, c_{22}, c_{23}, \dots) \\ d_3 &= (c_{31}, c_{32}, c_{33}, \dots) \\ &\dots = \dots \end{aligned}$$

Looking just at the first column, we could find a subsequence of vectors whose first columns converge to a number c_1 . We could find a subsequence of this subsequence whose second columns converge to a c_2 , and a subsequence of the subsequence of the subsequence whose third columns converge to a c_3 , etc. Now form a new subsequence by taking the first element of the first subsequence, and then the second element of the subsequence of a subsequence, and then the third element of the subsequence of the subsequence of the subsequence, etc. We get a subsequence of the original sequence so that every k^{th} column c_{ik} converges to c_i . We may as well assume that this was true of our original sequence.

Notice that $A(d_1), A(d_2), \dots$ have $\frac{c_{i2}}{2}$ in the second column and these terms converge to $\frac{c_2}{2}$. Similarly, these vectors have $\frac{c_{i3}}{3}$ in the third column and these entries converge to $\frac{c_3}{2}$. Etc.

The d_i are uniformly bounded by B , and so trivially each $|c_{ij}|$ is at most B . It follows that in the limit each $|\frac{c_n}{n}|$ is bounded by $\frac{B}{n}$. Thus $\sum |\frac{c_k}{k}|^2$ converges and the limiting vector belongs to $\mathcal{H}$. Call the limiting vector d . We claim $A(d_i) \rightarrow d$. Looking at the k^{th} component of $A(d_i)$ and d , the absolute value of their difference is $|\frac{c_{ik}}{k} - \frac{c_i}{k}|$. Since $|c_{ik}| \leq B$, $|c_i| \leq B$ and $|c_{ik} - c_i| \leq 2B$, this difference is at most $\frac{2B}{k}$. If we are given $\epsilon > 0$, we can find N such that $\sum_{k=N+1}^{\infty} (\frac{2B}{k})^2 < \frac{\epsilon}{2}$.

Hence

$$\|A(d_i) - d\|^2 \leq \left| \frac{c_{i1}}{1} - \frac{c_1}{1} \right|^2 + \dots + \left| \frac{c_{iN}}{N} - \frac{c_N}{N} \right|^2 + \frac{\epsilon}{2}$$

By choosing i large enough, the sum of the first N absolute values can be made less than $\frac{\epsilon}{2}$ and we are done. QED.

We now come to the main theorem for compact Hermitian operators:

Theorem 98 (The Spectral Theorem for Compact Hermitian Operators) *Let $\mathcal{H}$ be a Hilbert space and $A : \mathcal{H} \rightarrow \mathcal{H}$ be a compact Hermitian operator. Then there is an orthonormal basis $\{e_i\}$ of $\mathcal{H}$ consisting of eigenvectors of A with real eigenvalues λ_i . Moreover, $\lambda_i \rightarrow 0$.*

Remark: The condition $\lambda_i \rightarrow 0$ implies that the eigenspaces for non-zero eigenvalues are all finite dimensional, and that we can order the non-zero eigenvalues so $|\lambda_i| \geq |\lambda_{i+1}|$.

Proof:

If A is any bounded transformation, there is a $B > 0$ such that $\|Av\| \leq B\|v\|$. The greatest lower bound of all such B is called the *norm* of A and written $\|A\|$.

Lemma 43 *For any Hermitian A ,*

$$\|A\| = \sup\{|\langle Av, v \rangle| \mid \|v\| = 1\}$$

Proof of lemma: Let M be the sup in question. Note that $\overline{\langle Av, v \rangle} = \langle v, Av \rangle = \langle Av, v \rangle$, so $\langle Av, v \rangle$ is real. We have $|\langle Av, v \rangle| \leq \|Av\| \|v\| \leq \|A\| \|v\| \|v\|$. Since we assume $\|v\| = 1$,

$$|\langle Av, v \rangle| \leq \|A\|$$

and the sup of such terms is itself less than or equal to $\|A\|$. So $M \leq \|A\|$.

Now notice that for Hermitian A ,

$$\operatorname{Re}(\langle Av, w \rangle) = \frac{1}{4} \left(\langle A(v+w), v+w \rangle - \langle A(v-w), v-w \rangle \right)$$

Since $|\langle Av, v \rangle| \leq M\|v\|^2$

$$\operatorname{Re}(\langle Av, w \rangle) \leq \frac{1}{4} \left(M\|v+w\|^2 + M\|v-w\|^2 \right) = \frac{M}{4} (2\|v\|^2 + 2\|w\|^2)$$

by the Parallelogram law. Replace v by λv where λ has absolute value one to obtain

$$\operatorname{Re}(\lambda \langle Av, w \rangle) \leq \frac{M}{2} (\|v\|^2 + \|w\|^2)$$

The effect of λ is just to rotate the complex number $\langle Av, w \rangle$ in a circle; one of these values produces the real number $|\langle Av, w \rangle|$. We conclude that

$$|\langle Av, w \rangle| \leq \frac{M}{2} (\|v\|^2 + \|w\|^2)$$

Look at the special case when $w = \frac{\|v\|}{\|Av\|} Av$. In this case

$$\langle Av, w \rangle = \langle Av, Av \rangle \frac{\|v\|}{\|Av\|} = \|Av\| \|v\|$$

and

$$\frac{M}{2} (\|v\|^2 + \|w\|^2) = \frac{M}{2} \left(\|v\|^2 + \frac{\|v\|^2}{\|Av\|^2} \|Av\|^2 \right) = M\|v\|^2$$

and we obtain

$$\|Av\| \|v\| \leq M\|v\|^2 \quad \text{or} \quad \|Av\| \leq M\|v\|$$

Incidentally, this inequality also holds if $Av = 0$. So $\|A\| \leq M$, and using the first step, $\|A\| = M$. QED.

Proof of theorem, continued: We will now prove that A has an eigenvector. This is the key step of the proof.

Using the lemma, select a sequence v_n of vectors of length one such that

$$|\langle Av_n, v_n \rangle| \rightarrow \|A\|$$

Recall that A is Hermitian and $\langle Av, v \rangle$ is always real. So by picking a subsequence we can suppose that $\langle Av_n, v_n \rangle \rightarrow \lambda$ where $\lambda = \|A\|$ or $\lambda = -\|A\|$.

We have

$$\|Av_n - \lambda v_n\|^2 = \|Av_n\|^2 - 2\lambda \langle Av_n, v_n \rangle + \lambda^2 \|v_n\|^2 \leq \|A\|^2 \|v_n\|^2 - 2\lambda \langle Av_n, v_n \rangle + \lambda^2 \|v_n\|^2$$

But $\lambda = \pm\|A\|$ and $\|v_n\| = 1$ and $\langle Av_n, v_n \rangle \rightarrow \lambda$, so

$$\|Av_n - \lambda v_n\|^2 \leq 2\lambda^2 - 2\lambda \langle Av_n, v_n \rangle \rightarrow 0$$

In short,

$$Av_n - \lambda v_n \rightarrow 0$$

Now the grand moment. Since A is compact and v_n is bounded, the sequence Av_n has a convergent subsequence. Restricting to this subsequence, we can suppose that $Av_n \rightarrow w$. Then $\lambda v_n = Av_n - (Av_n - \lambda v_n) \rightarrow w$. By continuity of A , $A(\lambda v_n) \rightarrow A(w)$ but $A(\lambda v_n) = \lambda A(v_n) \rightarrow \lambda w$. So

$$A(w) = \lambda w$$

There is one final worry. Could $w = 0$? We show it could not as follows.

$$\|A\| = |\lambda| = \|\lambda v_n\| = \|Av_n - (Av_n - \lambda v_n)\| \leq \|Av_n\| + \|Av_n - \lambda v_n\| \rightarrow \|w\|$$

Hence w could only be zero if $\|A\| = 0$, but in that case $A = 0$ and every vector is an eigenvector.

Proof, continued: From here, the final steps are easy.

If v and w are eigenvectors with distinct eigenvalues, then they are orthogonal. Indeed $Av = \lambda v$ and $Aw = \tau w$ implies

$$\lambda \langle v, w \rangle = \langle \lambda v, w \rangle = \langle Av, w \rangle = \langle v, Aw \rangle = \bar{\tau} \langle v, w \rangle = \tau \langle v, w \rangle$$

so $(\lambda - \tau) \langle v, w \rangle = 0$ and the result follows.

For each eigenvalue λ of V , consider the eigenspace of all $v \in \mathcal{H}$ with $Av = \lambda v$. This is a closed subspace of $\mathcal{H}$ and thus itself a Hilbert space with a countable dense subset. For

each of these subspaces, choose an orthonormal basis, and notice that all of these basis vectors are eigenvectors of A with eigenvalue λ .

Now take the union of all of these bases as λ varies through eigenvalues. Any element of the resulting set is an eigenvector of length one, and any two elements are orthogonal. If this set is not maximal, it can be extended by adding additional vectors of length one perpendicular to the vectors of the existing set. We will then have a Hilbert space basis for $\mathcal{H}$ which can be divided into two pieces: the original eigenvectors and the new additional vectors. These pieces generate two spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ and $\mathcal{H}_1 \oplus \mathcal{H}_2 = \mathcal{H}$. Both closed because each is the orthogonal complement of the other. Since the first is generated by eigenvectors, $A(\mathcal{H}_1) \subset \mathcal{H}_1$. But then $A(\mathcal{H}_2) \subset \mathcal{H}_2$ since $\mathcal{H}_2$ consists of vectors perpendicular to $\mathcal{H}_1$ and $\langle v, Aw \rangle = \langle Av, w \rangle$.

If we restrict A to $\mathcal{H}_2$, we get a compact operator with no eigenvectors. This is impossible by an earlier step of the proof, so $\mathcal{H}_2$ is zero and we have a basis of eigenvectors.

If λ is an eigenvalue of A , consider the corresponding eigenspace of all v satisfying $Av = \lambda v$. This is a closed subspace of $\mathcal{H}$. If this subspace is infinite dimensional, then restricting A to the subspace gives an operator which is compact and yet identically λ . This is impossible unless $\lambda = 0$. So the eigenspace of every non-zero eigenvalue is finite dimensional.

Finally for every positive ϵ there are only finitely many eigenvalues with $|\lambda| \geq \epsilon$. Otherwise we can find infinitely many orthonormal vectors e_n with $\|Ae_n\| = \|\lambda e_n\| \geq \epsilon$. Then vectors e_n form a bounded set, so by compactness the Ae_n would have a convergent subsequence. But this is impossible because $\|Ae_m - Ae_n\|^2 = \|Ae_m\|^2 + \|Ae_n\|^2 \geq 2\epsilon^2$, so no subsequence can be Cauchy. QED.

19.11 Equiuniform Continuity

We now know the spectral theorem for compact operators, but the only example of a compact operator we have is already in diagonal form. We rectify that deficiency in this section. As usual, let V be the space of continuous functions on $[0, 1]$ with

$$\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx$$

Let $\mathcal{H}$ be its completion. Suppose $a(x, y)$ is continuous on $[0, 1] \times [0, 1]$ and define an operator

$$Af(x) = \int_0^1 a(x, y) dy$$

This operator is bounded and thus extends to $\mathcal{H}$. If $a(x, y) = \overline{a(y, x)}$, the operator is Hermitian. We are going to prove that A is compact.

Remark: The proof requires unusual but easy technical information. Require that a function $f(x)$ is *uniformly continuous* if for each $\epsilon > 0$ there is a $\delta > 0$ such that whenever $x, y \in [0, 1]$ and $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$. The key point is that the same delta works for the entire interval. A famous result says that a continuous function on $[0, 1]$ is always uniformly continuous there.

Suppose we have a family $\mathcal{F}$ of continuous functions on $[0, 1]$. This family can be finite or infinite and has no particular structure.

Definition 40 A family $\mathcal{F}$ is equiuniformly continuous if for each $\epsilon > 0$ there is a $\delta > 0$ such that whenever $x, y \in [0, 1]$ and $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$ simultaneously for all $f \in \mathcal{F}$.

Definition 41 Similarly we say that $\mathcal{F}$ is uniformly bounded if there is a $B > 0$ such that $|f(x)| \leq B$ for all $x \in [0, 1]$ and all $f \in \mathcal{F}$.

Remark: Here is the technical result we need:

Theorem 99 (Arzela-Ascoli Theorem) Suppose $\mathcal{F}$ is uniformly bounded and equiuniformly continuous. Let $f_1, f_2, \dots$ be a sequence in $\mathcal{F}$. Then there is a subsequence of this sequence which converges uniformly to a continuous function on $[0, 1]$.

Proof: Enumerate all rational numbers in $[0, 1]$ and call these points $q_1, q_2, q_3, \dots$. The numbers $f_i(q_1)$ are bounded, so we can find a subsequence of the f_i such that these numbers converge. Sticking just with this subsequence, the numbers $f_i(q_2)$ are bounded, so we can find a subsequence of the subsequence such that these numbers also converge. Etc. Now form a sequence by picking the first element of the first subsequence, and then the second element of the subsequence of this subsequence, and then the third element of the subsequence of the subsequence of this subsequence, etc. In this way a new subsequence of the original sequence is obtained such that $f_i(q)$ converges for each rational q . We can assume that this is the original sequence and proceed from here.

Suppose we are given $\epsilon > 0$. By equiuniform continuity, we can find $\delta > 0$ such that whenever $|x - y| < \delta$ we have $|f_i(x) - f_i(y)| < \frac{\epsilon}{3}$.

Fix an $x \in [0, 1]$ and find a rational q with $|x - q| < \frac{\delta}{2}$. Notice that if $|x - y| < \frac{\delta}{2}$, then $|y - q| < \delta$. Since $f_i(q)$ converge, so we can find N such that $i, j > N$ implies $|f_i(q) - f_j(q)| < \frac{\epsilon}{3}$. So

$$|f_i(y) - f_j(y)| \leq |f_i(y) - f_i(q)| + |f_i(q) - f_j(q)| + |f_j(q) - f_j(y)| < \epsilon$$

This inequality holds in the open neighborhood $\mathcal{U}_x = (x - \frac{\delta}{2}, x + \frac{\delta}{2})$ of x .

A finite number of such open sets cover $[0, 1]$, so selecting the maximum $\tilde{N}$ of the associated N gives $|f_i(x) - f_j(x)| < \epsilon$ whenever $i, j > \tilde{N}$. Consequently the sequence f_i is Cauchy

in the sup norm, and it is well known and easily proved that such a sequence converges uniformly to a continuous limiting function. QED.

Theorem 100 *The operator A described at the beginning of this section is compact. If v is an eigenvector with non-zero eigenvalue, then v is a continuous function on $[0, 1]$.*

Proof: Suppose first that $f_1, f_2, \dots$ is a sequence of continuous functions on $[0, 1]$ which are uniformly bounded in the Hilbert Space norm. Choose B so $\|f_i\| \leq B$. But for any f we have $|\langle f, 1 \rangle| \leq \|f\| \|1\|$ and consequently

$$\int_0^1 |f_i(x)| dx \leq \|f_i\| \leq B$$

We claim that the $A(f_i)$ are uniformly bounded. Indeed $a(x, y)$ is continuous on a compact set and so bounded by a number B_1 . So

$$|Af_i| = \left| \int_0^1 a(x, y) f_i(y) dy \right| \leq \int_0^1 |a(x, y) f_i(y)| dy \leq B_1 \int_0^1 |f_i(y)| dy \leq B_1 B.$$

Next we claim that the Af_i are equiuniformly continuous. Indeed

$$|Af_i(x_1) - Af_i(x_2)| = \left| \int_0^1 (a(x_1, y) - a(x_2, y)) f_i(y) dy \right|$$

Since a is continuous on $[0, 1] \times [0, 1]$, it is uniformly continuous there, so if $\epsilon > 0$, there is a $\delta > 0$ such that if $|x_1 - x_2| < \delta$ then $|a(x_1, y) - a(x_2, y)| < \epsilon$. The above equation then gives

$$|Af_i(x_1) - Af_i(x_2)| \leq \epsilon \int_0^1 |f_i(y)| dy \leq \epsilon B$$

Therefore we can apply the technical result about equiuniformly continuous families and conclude that there is a subsequence of f_i such that Af_i converges uniformly to a limiting f . Notice that

$$\|Af_i - f\|^2 = \int_0^1 |Af_i - f|^2 dx$$

then also converges to zero, so Af_i has a limit in the pre-Hilbert space V .

This proves that the compactness condition holds in V . By abstract nonsense, it then holds in $\mathcal{H}$. Indeed if v_i is a bounded sequence in $\mathcal{H}$, choose functions f_i in V such that for all n we have $\|f_n - v_n\| \leq \frac{1}{n}$. Then the sequence f_i is also bounded in V using the Hilbert space norm. So there is a subsequence for which f_i converges to f in norm, and $f \in V$. But then our sequence has a convergent subsequence, since for elements of the subsequence we have

$$\|v_i - f\| \leq \|v_i - f_i\| + \|f_i - f\| \rightarrow 0$$

Suppose $v \in \mathcal{H}$ and $Av = \lambda v$ with $\lambda \neq 0$. Find a sequence of continuous functions $\{f_i\}$ which converges to v in the Hilbert space norm. So $Af_i \rightarrow Av = \lambda v$, and therefore $\frac{1}{\lambda}Af_i \rightarrow v$.

We will prove that the Af_i are bounded in the sup norm, and equiuniformly continuous. It follows that a subsequence of Af_i converges uniformly to a limiting continuous function g . Replace the original sequence with this one. Then $\frac{1}{\lambda}Af_i$ also converges uniformly to $\frac{1}{\lambda}g$. But this sequence also converges to v , so $\frac{g}{\lambda} = v$ is a continuous function.

To complete the argument, we first show that the Af_i are uniformly bounded in the sup norm.

$$|Af_i(x)| = \left| \int_0^1 a(x, y) f_i(y) dy \right| \leq \int_0^1 |a(x, y)| |f_i(y)| dy$$

The function a is continuous and so bounded, say by B_1 . So

$$|Af_i(x)| \leq B_1 \int_0^1 |f_i(y)| dy = B_1 \|f_i\|_2 \leq B_1 \|f_i\|$$

Since $f_i \rightarrow v$ in norm, $\{\|f_i\|\}$ is bounded in the Hilbert space norm, and the above inequality then shows that the functions Af_i are bounded in the sup norm.

Equiuniform continuity is handled the same way.

$$|Af_i(x_1) - Af_i(x_2)| = \left| \int_0^1 (a(x_1, y) - a(x_2, y)) f_i(y) dy \right|$$

Since a is continuous, it is uniformly continuous, so if $\epsilon > 0$ we can find $\delta > 0$ such that when $|x_1 - x_2| < \delta$ we have $|a(x_1, y) - a(x_2, y)| < \epsilon$. So

$$|Af_i(x_1) - Af_i(x_2)| \leq \epsilon \int_0^1 |f_i(y)| dy = \epsilon \|f_i\|_2 \leq \epsilon \|f_i\|$$

Equiuniform continuity follows.

QED.

Remark: For example, $a(x, y) = (x - y)^2$ defines a compact Hermitian operator. It is amusing to find its eigenvalues and eigenvectors, which form a family of orthogonal polynomials.

19.12 Compact Operators in $L^2(G)$

The following section is very easy. We are going to show that the results of the previous section generalize with the same proofs when we replace $[0, 1]$ by a compact Lie group G , replace $\int_0^1$ by invariant integration on G , replace $a(x, y)$ by a continuous function on $G \times G$ and replace $\mathcal{H}$ by the completion of the space of complex-valued continuous functions on G with norm $\int |f(g)|^2 dg$.

There is one complication. Although our G has a Riemannian metric, we have never used the fact that manifolds with such metrics can be made into metric spaces. So we have no metric space structure on G and thus no notion of uniform continuity or equiuniform continuity. However, we can cover G by coordinate neighborhoods; in each such neighborhood uniform continuity and equiuniform continuity make sense. With a little more care, we can choose a finite number of coordinate systems, and in each one a compact rectangular box $[a_1, b_1] \times \dots \times [a_n, b_n]$ such that every point of G is mapped by a coordinate system to at least one of these boxes. Then we can call f on G uniformly continuous if its restriction to each box is uniformly continuous, and call a family $\mathcal{F}$ of functions on G equiuniformly continuous if its restriction to each box is equiuniformly continuous.

The general Arzela-Ascoli theorem for G then reduces to generalizing from $[0, 1]$ to a closed box, which we can safely leave to the reader. The final step to get the result on G proceeds as follows: start with a sequence, find a convergent subsequence on the first box, throw away other elements and assume the new sequence is the original sequence. Find a convergent subsequence on the second box, etc. In the end we will have a subsequence which converges uniformly on each closed box, and hence converges in the sup norm on G .

It remains to show that A on G is a compact operator. The first step of the proof requires that we start with a sequence of continuous functions which are uniformly bounded in the Hilbert Space norm, and show that they are also uniformly bounded in the norm $\int_G |f_i(g)| dg$. This is proved as earlier, since the constant function 1 belongs to the Hilbert space because G is compact. The remaining argument makes sense in coordinate boxes, and glueing these results together gives the result we are after.

Let us examine the final argument, that if v is an eigenvector with non-zero eigenvalue, it is given by a continuous function on G . As before, we start with a sequence of continuous functions f_i converging to v in the Hilbert space norm. As before, $\frac{1}{\lambda} Af_i \rightarrow v$. The proof in the special case that the Af_i are uniformly bounded in sup norm still works. When we prove that the Af_i are equiuniformly continuous, we can restrict to a rectangular coordinate box; that just makes each integral of absolute values smaller. So the previous argument still works.

Theorem 101 *Let V be the space of continuous complex valued functions on a Lie group G with inner product*

$$\langle f_1, f_2 \rangle = \int_G f_1(g) \overline{f_2(g)} dg$$

and let $\mathcal{H}$ be the completion of this space to a Hilbert space. Let $a(g_1, g_2)$ be a continuous function on G and suppose that $a(g_1, g_2) = \overline{a(g_2, g_1)}$. Define a linear transformation $V \rightarrow V$ by

$$Af(g_1) = \int_G a(g_1, g_2) f(g_2) dg_2$$

This transformation is bounded and thus induces a Hermitian operator on $\mathcal{H}$. The operator is compact and therefore $\mathcal{H}$ has a basis consisting of eigenvectors of A . If λ_i are the corresponding eigenvalues, $\lambda_i \rightarrow 0$. Finally, every eigenvector with nonzero eigenvalue is a continuous function.

19.13 The Peter-Weyl Theorem

The theorem in question is a powerful result with many significant consequences. Each of these consequences is sometimes known as the Peter-Weyl theorem. Take your pick.

Theorem 102 (Peter-Weyl) *Let G be a compact Lie group, and let $\hat{g} \neq e$ be an element of G . Then there is a finite dimensional representation φ of G such that $\varphi(\hat{g}) \neq I$.*

Proof: Let $a(g)$ be a continuous function on G with the property that $a(g^{-1}) = \overline{a(g)}$. Such functions exist; we could pick an arbitrary continuous $b(g)$ and let $a(g) = b(g)\overline{b(g^{-1})}$. Indeed then

$$a(g^{-1}) = b(g^{-1})\overline{b(g)} = \overline{b(g)}b(g^{-1}) = \overline{b(g)\overline{b(g^{-1})}} = \overline{a(g)}$$

Define $a(g_1, g_2) = a(g_1^{-1}g_2)$. Notice that

$$a(g_2, g_1) = a(g_2^{-1}g_1) = \overline{a((g_2^{-1}g_1)^{-1})} = \overline{a(g_1^{-1}g_2)} = \overline{a(g_1, g_2)}$$

It follows that A defines a compact Hermitian operator in the standard way.

Lemma 44 *If f is a continuous function on G then*

$$A(L_g(f)) = L_g A(f)$$

Proof:

$$A(L_g(f))(g_1) = \int_G a(g_1^{-1}g_2) f(g^{-1}g_2) dg_2$$

Replace g_2 by $g g_2$ everywhere in this formula; this is legal because our integral is invariant under left translation. We get

$$A(L_g(f)) = \int_G a(g_1^{-1}g g_2)f(g_2)dg_2 = \int_G a((g^{-1}g_1)^{-1}g_2)f(g_2)dg_2 = Af(g^1g_1) = L_g(Af)(g_1)$$

QED.

Continuation of the Proof of Peter-Weyl: In particular, suppose f is an eigenvector of A and $Af = \lambda f$ with a non-zero λ . Then

$$A(L_g f) = L_g(Af) = L_g(\lambda f) = \lambda(L_g f)$$

So L_g preserves the λ eigenspace of A and defines a representation of G on this space. Since A is compact, this is a finite dimensional representation. Using the spectral theorem of A , we get a large number of such finite dimensional representations.

Recall that $\hat{g} \neq e$. Find a continuous function f on G with the following properties:

- $f \geq 0$
- $f = 0$ in an open neighborhood of $\hat{g}$
- $f = 1$ in an open neighborhood of e

Notice that $(L_{\hat{g}}f)(\hat{g}) = f(\hat{g}^{-1}\hat{g}) = f(e) = 1$ and therefore $L_{\hat{g}}f \neq f$. Let f_1 be the continuous function $f - L_{\hat{g}}f$ and notice that this function is not zero.

Consider the λ eigenspace of A with orthonormal basis $e_1, \dots, e_k$. As shown earlier, $L_{\hat{g}}$ acts on this space. If this transformation is not the identity, we are done. Otherwise $L_{\hat{g}}e_i = e_i$ and by invariance of integration,

$$\langle f, e_i \rangle = \langle L_{\hat{g}}f, L_{\hat{g}}e_i \rangle = \langle L_{\hat{g}}f, e_i \rangle$$

so f and $L_{\hat{g}}f$ have the same expansion coefficients for basis vectors in the λ eigenspace and f_1 has zero expansion coefficients.

If $L_{\hat{g}}$ acts non-trivially on even one of the nonzero λ eigenspaces, we are done. Otherwise f_1 has an expansion that only involves e_i for which $Ae_i = 0$, and so $A(f_1) = 0$.

To finish the proof, we will show that if f_1 is a nonzero continuous function, we can find A which is not zero on this function. The rough idea of the proof is easy. We will select $a(g)$ as an approximation of the Dirac delta function. Then A approximates the identity map, and $Af_1 \sim f_1$ and certainly is nonzero.

Lemma 45 *Suppose $\epsilon > 0$. There is an open neighborhood $\mathcal{U}$ of $e \in G$ such that whenever $b \in \mathcal{U}$ we have $|f_1(gb) - f_1(g)| < \epsilon$ simultaneously for all g . Said another way, whenever $b = g_1^{-1}g_2 \in \mathcal{U}$ we have $|f_1(g_2) - f_1(g_1)| < \epsilon$.*

Remark: Earlier we used local coordinates to define uniform continuity on our Lie group G . The Peter-Weyl theorem is actually true for compact topological groups in general, and then coordinates are not available. There is a different way to define uniform continuity which works on all compact topological groups, and the lemma above hints at that approach.

Proof of Lemma: Fix g . By continuity, there is an open neighborhood $\mathcal{W}_g$ of g such that if $g_1 \in \mathcal{W}_g$ then $|f(g_1) - f(g)| < \frac{\epsilon}{2}$. Let $\mathcal{U}_g = g^{-1}\mathcal{W}_g$. Then equivalently, whenever $b \in \mathcal{U}_g$ we have $|f(gb) - f(g)| < \frac{\epsilon}{2}$.

Using continuity of multiplication, we can find a neighborhood $\mathcal{V}_g$ of the identity such that if $b_1, b_2 \in \mathcal{V}_g$ then $b_1 b_2 \in \mathcal{U}_g$. Since $g \in g\mathcal{V}_g$, the $g\mathcal{V}_g$ form an open cover of G and there is a finite subcover $g_1\mathcal{V}_{g_1}, \dots, g_k\mathcal{V}_{g_k}$. Form the neighborhood of e defined by $\mathcal{V} = \cap \mathcal{V}_{g_i}$.

If g is an arbitrary element of G , we can find i such that $g \in g_i\mathcal{V}_{g_i}$. So $g = g_i b_i$ where $b_i \in \mathcal{V}_{g_i}$ and therefore $|f(g) - f(g_i)| = |f(g_i b_i) - f(g_i)| < \frac{\epsilon}{2}$. If $b \in \mathcal{V}$, then $b \in \mathcal{V}_{g_i}$ and so $gb = g_i b_i b$, where $b_i b$ is the product of two elements in $\mathcal{V}_{g_i}$ and thus is in $\mathcal{U}_{g_i}$. So $|f(gb) - f(g_i)| = |f(g_i b_i b) - f(g_i)| < \frac{\epsilon}{2}$. Putting the two inequalities together, $|f(gb) - f(b)| < \epsilon$. QED.

Conclusion of the Proof of the Peter-Weyl Theorem: Choose $0 < \epsilon < \max|f_1|$. Find $\mathcal{U}$ as in the lemma above. Choose a coordinate neighborhood of e , and within it find an ϵ_1 ball about the origin whose points correspond to points in G which are in $\mathcal{U}$.

Choose a function $b(g)$ in this coordinate neighborhood with the following properties

- $b(g)$ is real
- $b(g) \geq 0$
- $b(g) = 0$ unless g is in the coordinate ball of radius ϵ_1 around the origin

Choose $a(g) = b(g)b(g^{-1})$. Notice that $a(g)$ also has these properties. We can multiply b by an arbitrary positive real number r , multiplying a by r^2 . Choose an r such that the integral of a over this ϵ_1 neighborhood (and thus the integral of a over all of G) is one.

As usual, select $a(g_1, g_2) = a(g_1^{-1}g_2)$. Note that

$$\int_G a(g_1^{-1}g_2)dg_2 = \int_G a(g_2)dg_2 = 1$$

by left invariance and so

$$\int_G a(g_1, g_2)f_1(g_1)dg_2 = f_1(g_1)$$

Consequently

$$A(f_1)(g_1) - f_1(g_1) = \int_G a(g_1, g_2)(f_1(g_2) - f_1(g_1))dg_2$$

and the absolute value of the difference of these functions is at most

$$\int a(g_1^{-1}g_2)|f_1(g_2) - f_1(g_1)|dg_2$$

But a vanishes unless $g_1^{-1}g_2$ is within ϵ_1 of the origin, and then $|f(g_1) - f(g_2)| < \epsilon$. So

$$|A(f_1)(g_1) - f_1(g_1)| \leq \int a(g_1^{-1}g_2)\epsilon = \epsilon$$

If $A(f_1) = 0$, then $|f_1(g_1)| \leq \epsilon < \max(|f_1|)$, a contradiction. QED.

19.14 A Basis for $L^2(G)$

The orthogonality relations play a big role in the theory of representations of finite groups, but surprisingly they have not been mentioned for Lie groups. That is about to change. From now on, we will assume that integration over G has been normalized so $\int_G 1 = 1$. We will also assume that all representation spaces have been assigned an invariant inner product, making all representation matrices unitary.

Theorem 103 (Orthogonality Relations I) *Let $(a_{ij}(g))$ be matrices for an irreducible representation of G . Then*

$$\int a_{ij}(g)\overline{a_{kl}(g)} dg = \frac{1}{\dim V}\delta_{ik}\delta_{jl}$$

In other words, each component is a function in $L^2(G)$ of norm $\frac{1}{\sqrt{\dim V}}$, and distinct components are orthogonal.

Theorem 104 (Orthogonality Relations II) *Let $(a_{ij}(g))$ and $(b_{ij}(g))$ be matrices for two inequivalent irreducible representations of G . Then*

$$\int a_{ij}(g)\overline{b_{kl}(g)} dg = 0$$

In other words, the components of two distinct irreducible representations are orthogonal

Proof: These theorems have been copied word for word from the chapter on representations of finite groups, and the proofs are exactly the same. QED.

Theorem 105 (Peter-Weyl Theorem; Second Version) *Let G be a compact Lie group. Select one irreducible representation from each equivalence class, and form the set of all matrix components, renormalizing by multiplying each component by $\sqrt{\dim V}$. Then the resulting continuous functions form a Hilbert Space basis for $L^2(G)$.*

Proof: Let $\mathcal{A}$ be the set of all finite linear combinations of these orthonormal elements. All elements of $\mathcal{A}$ are continuous functions on G . Notice the following points:

- If the coefficients of some matrix representation $\rho = (\rho_{ij})$ belong to $\mathcal{A}$, so do the coefficients of any equivalent matrix representation. Indeed the equivalent matrix is $B^{-1}\rho B$ for a constant matrix B , and so linear combinations of the ρ_{ij}
- If the coefficients of two representations ρ_1 and ρ_2 are in $\mathcal{A}$, so are the coefficients of $\rho_1 \oplus \rho_2$. This is obvious.
- The coefficients of *any* continuous representation of G are in $\mathcal{A}$. Indeed such a representation can be written as a sum of irreducible representations, and so is equivalent to a sum of the particular irreducible representations we used to construct $\mathcal{A}$.
- The algebra $\mathcal{A}$ is invariant under conjugation. Indeed if (ρ_{ij}) is a representation, so is $(\overline{\rho_{ij}})$.
- The identity function belongs to $\mathcal{A}$. Indeed one representation of G is the identity representation.
- The space $\mathcal{A}$ is closed under multiplication, and thus forms an algebra. Indeed if ρ_1 and ρ_2 are representations acting on V_1 and V_2 , so is $\rho_1 \otimes \rho_2$ acting on $V_1 \otimes V_2$. This space has basis $e_i \otimes f_j$ and $\rho_1 \otimes \rho_2$ sends this element to

$$\sum (\rho_1)_{ik} e_k \otimes \sum (\rho_2)_{jl} f_l = \sum (\rho_1)_{ik} (\rho_2)_{jl} e_k \otimes f_l$$

- The algebra $\mathcal{A}$ separates points. Indeed if $g_1 \neq g_2$, then $g_1^{-1}g_2 \neq 1$, so by the Peter-Weyl theorem there is a ρ with $\rho(g_1^{-1}g_2) \neq I$ and so $\rho(g_1)^{-1}\rho(g_2) \neq I$ and so $\rho(g_1) \neq \rho(g_2)$.

Therefore we can apply the Stone-Weierstrass theorem to conclude that any continuous function f on G can be uniformly approximated by a finite sum of the conjectured Hilbert Basis set, to arbitrarily small precision. If $|f_n - f|$ is arbitrarily small, then $\|f_n - f\|$ is also arbitrarily small. But if we replace the coefficients of the basis vectors in f_n used to make $|f_n - f|$ small by the “Fourier coefficients”, this is an even better approximation of $\|f_n - f\|$. It follows that the Hilbert Space series for f converges to f .

It follows that the conjectured Hilbert Space basis elements form a maximal orthonormal set and therefore a Hilbert Basis. Indeed if $h \in \mathcal{H}$ is orthogonal to all of our basis elements, then h is orthogonal to f for all continuous functions on G by the previous sum, so by continuity h is orthogonal to every element of $\mathcal{H}$, which is impossible. QED.

Remark: There is another way to interpret this result which is familiar from the finite group case. Consider the subspace of $\mathcal{H}$ formed by the coefficients of one particular irreducible representation ρ_{ij} . These functions form a basis for a finite dimensional subspace

of $\mathcal{H}$ which is invariant under left translation. Indeed $L_{g_1}\rho(g) = \rho(g_1^{-1}g) = \rho(g_1)^{-1}\rho(g)$ and so

$$\begin{aligned} L_{g_1} \left(\sum_i c_i \rho(g)_{ij} \right) &= \sum_i c_i \rho(g_1^{-1}g)_{ij} = \sum_i c_i \rho(g_1^{-1})_{ik} \rho(g)_{kj} = \\ &= \sum_i c_i \overline{\rho(g_1)}_{ik}^T \rho(g)_{kj} = \sum_k \left(\sum_i \overline{\rho(g_1)}_{ki} c_i \right) \rho(g)_{kj} \end{aligned}$$

This equation says we should fix j and thus fix a column of the representation matrix. The entries in this column are continuous functions in G and thus the column defines a finite dimensional subspace of $L^2(G)$ with the column entries as basis vectors. The left side of the full equation shows a typical element $\sum c_i \rho(g)_{ij}$ in this subspace, being acted upon by left translation. The right side of the equation says that the matrix for this action is $\overline{\rho(g_1)}$. So we began with the matrix for ρ and ended with $\dim \rho$ subspaces of $L^2(G)$, each invariant under left translation, where left translation is given by the conjugate of ρ . Therefore if ρ is any irreducible representation with dimension k , then $\bar{\rho}$ occurs in the decomposition of $L^2(G)$ exactly k times. Sometimes ρ and $\bar{\rho}$ are equivalent and sometimes not. But even if not, they are both irreducible and have the same dimension. Clearly it follows that if ρ is irreducible with dimension k , then ρ occurs in the decomposition of $L^2(G)$ exactly k times.

Theorem 106 *In the decomposition of $L^2(G)$ into irreducible subspaces, each irreducible representation G occurs, and indeed exactly as many times as the dimension of the representation.*

Remark: For finite groups, it immediately follows that $|G| = \sum_i d_i^2$, where d_i are the dimensions of the irreducible representations of G .

19.15 A Basis for the Space of Class Functions

A continuous function f on G is said to be a *class function* if it is constant on conjugacy classes, so $f(g_1^{-1}gg_1) = f(g)$ for all $g, g_1 \in G$. The set of class functions is a subspace of the space of continuous functions on G , and the closure of this set is a subspace of the space of L^2 functions called the space of L^2 class functions.

The characters of irreducible representations are class functions; they are orthonormal. Indeed the proof of this fact given earlier for finite groups holds word for word in the case of a compact Lie group.

Theorem 107 *The characters of the irreducible representations of G form a Hilbert Space basis for the space of class functions in $L^2(G)$*

Proof: Let f be a continuous class function. For each $\epsilon > 0$ there is a linear combination of the a_{ij} which uniformly approximates f . So $|f - \sum c_{ij}\varphi_{ij}| < \epsilon$. In this expression it is useful to group together the φ_{ij} which come from a specific irreducible representation. So we get

$$\left| f(g) - \sum_{V_m} \left(\sum c_{ij}\varphi_{ij} \right) \right| < \epsilon$$

where the interior sum is only over matrix entries from a specific irreducible representation.

Replace this statement by an average over conjugates:

$$\left| \int_G \left(f(g_1^{-1}gg_1) - \sum_{V_m} \left(\sum c_{ij}\varphi_{ij}(g_1^{-1}gg_1) \right) \right) dg_1 \right| < \epsilon$$

This follows from the previous statement since $\int_G 1 = 1$. Since f is a class function, this inequality can be written

$$\left| f(g) - \sum_{V_m} \left(\sum c_{ij} \int_G \varphi_{ij}(g_1^{-1}gg_1) dg_1 \right) \right| < \epsilon$$

For a while, let us work on the term from a particular irreducible representation. In the above expression for this term, φ is a representation on a vector space V and the integral gives a second representation on V by integrating each coefficient separately. Let us emphasize this point of view by writing

$$T(g) = \int_G \varphi(g_1^{-1}gg_1) dg_1$$

At the end of the proof, we will show that on each irreducible piece, $T(g)$ commutes with φ and thus is an intertwining operator. Hence $T(g)$ is $\lambda(g)$ on each piece. As we demonstrate next, this will prove the theorem.

Continue concentrating on the term in the expression from one of these irreducible pieces:

$$\sum c_{ij} \int_G \varphi_{ij}(g_1^{-1}gg_1) dg_1 = \sum c_{ij} T_{ij}$$

Since T is a diagonal matrix, we have

$$\sum_{i \neq j} c_{ij} \int_G \varphi_{ij}(g_1^{-1}gg_1) dg_1 = \sum_{i \neq j} c_{ij} T_{ij} = 0$$

We have

$$\int_G \varphi_{ii}(g_1^{-1}gg_1) dg_1 = \lambda(g)$$

and summing over i on the left gives the character χ of φ , so

$$\int_G \chi(g_1^{-1}gg_1) dg_1 = \lambda(g) \dim V$$

Since characters are class functions, this gives

$$\chi(g) = \lambda(g) \dim V$$

So

$$\sum_i c_{ii} \int_G \varphi_{ii}(g_1^{-1}gg_1) dg_1 = \sum_i c_{ii} T_{ii} = \sum_i c_{ii} \lambda(g) = \left(\sum_i c_{ii} \right) \frac{1}{\dim V} \chi(g)$$

Consequently going back to the beginning of the proof we have

$$\left| f(g) - \sum_{V_m} \left(\frac{\sum c_{ii}}{\dim V} \right) \chi(g) \right| < \epsilon$$

From here the proof has become canonical. Every class function can be approximately uniform closely by a linear sum of characters. This also gives a Hilbert Space norm approximation. But changing the coefficients to the Fourier coefficients is an even better Hilbert Space norm approximation. So the Fourier series converges. Thus the characters form a maximal orthonormal set, because any other orthonormal vector would be orthogonal to continuous class functions, and therefore to the completion of these functions, and therefore to the entire Hilbert space of class functions.

To complete the proof, we show that T on an irreducible subspace is an intertwining operator. Indeed

$$\begin{aligned} T\varphi(g_2) &= \int_G \varphi(g_1^{-1}gg_1) \varphi(g_2) dg_1 = \int_G \varphi(g_1^{-1}gg_1g_2) dg_1 = (\text{replace } g_1 \text{ by } g_1g_2^{-1}) \\ &= \int_G \varphi((g_1g_2^{-1})^{-1}gg_1) dg_1 = \int_G \varphi(g_2g_1^{-1}gg_1) dg_1 = \varphi(g_2)T \end{aligned}$$

QED.

Remark: In the finite group case, this theorem is used to prove that the number of irreducible representations of G equals the number of conjugacy classes of G . We will see that the theorem is equally important for Lie groups, and ultimately proves that every highest weight comes from an irreducible representation.

19.16 Matrix Groups

Theorem 108 *The universal cover of $SL(2, R)$ is not a matrix group.*

Proof: If the result is false, there is a representation φ for this universal cover which is one-to-one. This representation induces a Lie algebra representation of $sl(2, R)$. We have proved that every such representation is a direct sum of irreducible representations, and we have completely classified these irreducible representations. Each induces a representation of $SL(2, R)$. So to complete the proof, we need to show that the universal cover of $SL(2, R)$ is not just $SL(2, R)$ and this requires showing that the fundamental group of $SL(2, R)$ is not trivial.

The columns of each matrix in $SL(2, R)$ are linearly independent. Apply the Gram-Schmidt to these vectors to find orthonormal vectors. Notice that this step leaves rotation matrices

$$\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

fixed and maps any matrix to a rotation matrix. So we obtain maps $S^1 \rightarrow SL(2, R) \rightarrow S^1$ whose composition is the identity, and consequently the fundamental group of $SL(2, R)$ cannot be trivial. QED.

Theorem 109 *Every compact Lie group is isomorphic to a closed subgroup of $GL(n, C)$.*

Proof: Let $g \neq e$ be an element of the connected component of the identity of G . By the Peter-Weyl theorem, there is a representation φ of G which is nontrivial on g . Let K be the kernel of this representation and notice that K is a closed subgroup of G and hence a Lie group in the induced topology. The connected component of the identity in K does not contain g and thus cannot equal the connected component of the identity in G . So K is a Lie group of smaller dimension.

Let $g_1 \neq e$ be an element in the connected component of K . By the Peter-Weyl theorem, there is a representation ρ of G which is non-trivial on g_1 . Replace φ by $\varphi \oplus \rho$. Then the kernel of the new φ is a closed group $K_1 \subset K$. The connected component of the identity of K_1 does not contain g_1 and thus is not equal to the connected component of K . So K_1 has smaller dimension than the dimension of K . Continue the process until finally a group K_i of dimension zero occurs. At this point, φ is a representation whose kernel has no connected component and thus is finite.

For each $g \in K_i$ with $g \neq e$, there is a representation ψ_g of G which is not the identity on g . Replace φ by $\varphi \oplus \sum \oplus \psi_g$. QED.

19.17 Hilbert Space Representations

Definition 42 Let $\mathcal{H}$ be a Hilbert Space. A linear transformation $U : \mathcal{H} \rightarrow \mathcal{H}$ is unitary if $\langle Uv, Uw \rangle = \langle v, w \rangle$ for all $v, w \in \mathcal{H}$. Note that $\|Uv\| = \|v\|$, so unitary maps are continuous.

Remark: If U is unitary, then $\langle U^*Uv, w \rangle = \langle v, w \rangle$ for all u and w , and it easily follows that $U^*U = I$. Conversely, any such map is unitary.

Definition 43 Let G be a Lie group. A continuous representation of G on a Hilbert Space $\mathcal{H}$ is an assignment to each $g \in G$ of a unitary transformation $\varphi(g)$ such that

- $\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2)$
- for each fixed $v, w \in \mathcal{H}$, the map $g \rightarrow \langle \varphi(g)v, w \rangle$ is continuous

Definition 44 A Hilbert Space representation φ on $\mathcal{H}$ is irreducible if there is no closed subspace $\mathcal{K} \subset \mathcal{H}$ such that each $\varphi(g)$ maps $\mathcal{K}$ back to itself.

Remark: Note that if this is false, we can write $\mathcal{K} \oplus \mathcal{K}^\perp$ and the unitary condition then proves that both subspaces are invariant under φ .

Theorem 110 If G is compact, every irreducible representation φ of G is finite dimensional.

Proof: The idea of the proof is very easy. Suppose φ and ψ are two irreducible representations of G on $\mathcal{H}_1$ and $\mathcal{H}_2$. If $e_1, e_2, \dots$ is a Hilbert Space basis for $\mathcal{H}_1$ and $f_1, f_2, \dots$ is a Hilbert Space basis for $\mathcal{H}_2$, then $\langle \varphi(g)e_i, e_j \rangle$ is a typical matrix coefficient of φ and $\langle \psi(g)f_k, f_l \rangle$ is a typical matrix coefficient of ψ and both are continuous functions on G .

If both of these representations are finite dimensional and irreducible and the representations are not equivalent, we proved that these coefficients are orthogonal in $L^2(G)$, so

$$\int_G \langle \varphi(g)e_i, e_j \rangle \overline{\langle \psi(g)f_k, f_l \rangle} = 0$$

It turns out that this result is still true when φ is infinite dimensional and ψ is finite dimensional. But by the Peter-Weyl theorem, the ψ terms are dense in $L^2(G)$, so $\langle \varphi(g)e_i, e_j \rangle$ would be orthogonal to a dense subset and thus everything and therefore $\langle \varphi(g)e_i, e_j \rangle = 0$. This would hold for all e_i and e_j and so $\varphi(g) = 0$, a contradiction.

We now show the required orthogonality. We'll work slightly more generally. Let φ and ψ be representations on Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. At first we make no assumptions on dimensionality or irreducibility. Let $v, w \in \mathcal{H}_1$ and $v_1, w_1 \in \mathcal{H}_2$ and consider the

expression

$$\int \langle \varphi(g)v, w \rangle \overline{\langle \psi(g)v_1, w_1 \rangle} dg$$

The absolute value of this expression is at most the integral of the absolute values of the terms, which in turn is at most

$$\|v\| \|w\| \|v_1\| \|w_1\|$$

Fix v and v_1 . We claim that there is a bounded $T : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ such that the map has the form $\langle Tw, w_1 \rangle$. Indeed if we fix everything except w_1 and conjugate the value of the integral, then we get a bounded linear map from $\mathcal{H}_2 \rightarrow \mathbb{C}$. This map must have the form $w_1 \rightarrow \langle w_1, h \rangle$ for a unique $h \in \mathcal{H}_2$. Conjugating the integral and also this term, we find that the integral has the form $w_1 \rightarrow \langle h, w_1 \rangle$ for a unique $h \in \mathcal{H}_2$. This h depends on w , so call it Tw . Then the integral equals $\langle Tw, w_1 \rangle$. Since the integral is linear in w , T is linear in w . Finally, our initial inequality shows that T is a bounded transformation. In short

$$\langle Tw, w_1 \rangle = \int \langle \varphi(g)v, w \rangle \overline{\langle \psi(g)v_1, w_1 \rangle} dg$$

We claim that T intertwines φ and ψ ; indeed

$$\langle T\varphi(g_1)w, w_1 \rangle = \int \langle \varphi(g)v, \varphi(g_1)w \rangle \overline{\langle \psi(g)v_1, w_1 \rangle} dg = (\text{because each } \varphi \text{ is unitary})$$

$$\int \langle \varphi(g_1)^{-1}\varphi(g)v, w \rangle \overline{\langle \psi(g)v_1, w_1 \rangle} dg = (\text{by left invariance of integration})$$

$$\int \langle \varphi(g)v, w \rangle \overline{\langle \psi(g_1g)v_1, w_1 \rangle} dg = (\text{unitary } \psi)$$

$$\int \langle \varphi(g)v, w \rangle \overline{\langle \psi(g)v_1, \psi(g_1^{-1})w_1 \rangle} dg = \langle Tw, \psi(g_1)^{-1}w_1 \rangle = \langle \psi(g_1)Tw, w_1 \rangle$$

Now allow φ to be infinite dimensional, but assume that ψ is finite dimensional. Suppose both are irreducible. We have $T\varphi = \psi T$, so the kernel of T is a closed invariant subspace of $\mathcal{H}_1$ and the image of T is an invariant subspace of ψ . So either $T = 0$ or ψ is one-to-one, and either T is onto or $T = 0$. But T cannot be one-to-one and onto because $\mathcal{H}_1$ and $\mathcal{H}_2$ are not isomorphic. So $T = 0$ and the original integral is zero.

Remark: In general, the range of a continuous operator need not be closed. But in our case the image of T is finite dimensional and this problem does not arise.

Chapter 20

Weyl's Integral Formula

20.1 The Weyl Integral Formula

Let f be a class function. The values of f are completely determined by the values of f on a maximal torus T . We would like to compute $\int_G f(g) dg$ by integrating a related function $\int_T f(g)u(g) dg$ over T .

Theorem 111 *Let G be compact and connected, and let T be a maximal torus. If f is a class function on G , then*

$$\int_G f(g) dg = \int_T f(t)u(t) dt$$

where $u(t) = \delta(t)\overline{\delta(t)}/|W|$ and $|W|$ is the order of the Weyl Group. Here $\delta(t)$ is given by the following product over all roots of G :

$$\delta(t) = \prod_{\alpha > 0} (e^{2\pi i \alpha(t)} - 1)$$

Since only the expression $\delta\overline{\delta}$ matters and $\delta(t) = \prod_{\alpha > 0} e^{\pi i \alpha(t)} \prod_{\alpha > 0} (e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)})$, we can replace δ with

$$\delta(t) = \prod_{\alpha > 0} (e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)})$$

The proof will be given in a later section.

20.2 The Case $SU(2)$

Consider the group $SU(2)$, with maximal torus the set of diagonal matrices $A(t) = \begin{pmatrix} e^{2\pi it} & 0 \\ 0 & e^{-2\pi it} \end{pmatrix}$ for $t \in [0, 1]$. Then $Ad(A)$ is the action on the Lie algebra induced by conjugation and so

$$Ad(A) \begin{pmatrix} 0 & v \\ -\bar{v} & 0 \end{pmatrix} = \begin{pmatrix} e^{2\pi it} & 0 \\ 0 & e^{-2\pi it} \end{pmatrix} \begin{pmatrix} 0 & v \\ -\bar{v} & 0 \end{pmatrix} \begin{pmatrix} e^{-2\pi it} & 0 \\ 0 & e^{2\pi it} \end{pmatrix} =$$

$$\begin{pmatrix} 0 & e^{4\pi it} v \\ -e^{-4\pi it} \bar{v} & 0 \end{pmatrix} = e^{2\pi i(2t)} \begin{pmatrix} 0 & v \\ -\bar{v} & 0 \end{pmatrix}$$

and $\alpha(t) = 2t$. This and its negative are the roots of $SU(2)$. Notice that the 2 is expected because the weights multiply t by integers and the roots multiply integers by even integers; i.e., when we normalize to make the weight lattice Z , the root lattice becomes $2Z$.

Since the only positive root is $\alpha(t) = 2t$, $\delta = (e^{2\pi i(2t)} - 1) = (\cos(2\pi(2t)) - 1 + i \sin(2\pi(2t)))$ and

$$\delta \bar{\delta} = (\cos -1 + i \sin)(\cos -1 - i \sin) = (\cos -1)^2 + \sin^2 = 2 - 2 \cos(4\pi t)$$

The Weyl group has order 2 because there is only one reflection, so

$$u(t) = 1 - \cos 4\pi t$$

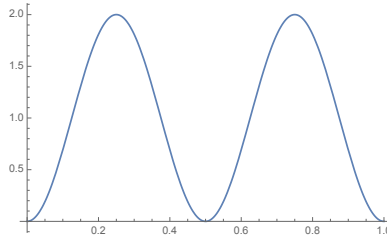


Figure 20.1: $u(t)$

This picture might be a little surprising, but it is easily explained. The matrices $A(t)$ for $t = 0$ and $t = \frac{1}{2}$ are $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ and these matrices commute with everything. So the conjugacy class of each matrix contains only one element. But the conjugacy classes of other elements of T are larger and contribute more weight to the total integral over G .

The characters of $SU(2)$ are $1, \chi + \chi^{-1}, \chi^2 + 1 + \chi^{-2}, \dots$ where $\chi(t) = e^{2\pi it}$. We are going to verify that these characters are orthonormal using the Weyl integral formula. This will

be surprisingly tricky, but the method we use gives insight into the general Weyl Character Formula.

First observe that $\frac{1}{2} \int_0^1 (\chi^k + \chi^{-k}) dt$ is zero if $k \neq 0$. Indeed

$$\int_0^1 (e^{\pi i k t} + e^{-\pi i k t}) dt = \left(\frac{e^{\pi i k t}}{\pi i k} + \frac{e^{-\pi i k t}}{-\pi i k} \right) \Big|_0^1 = \frac{1}{\pi i k} (e^{\pi i k} - e^{-\pi i k}) = \frac{e^{-\pi i k}}{\pi i k} (e^{2\pi i k} - 1) = 0$$

Next consider the functions $\chi^k - \chi^{-k}$. Using the inner product $\langle f, g \rangle = \frac{1}{2} \int_0^1 f(t) \overline{g(t)} dt$, we claim that these functions are orthonormal. They have norm one because

$$\frac{1}{2} \int_0^1 (\chi^k - \chi^{-k}) \overline{(\chi^k - \chi^{-k})} = \frac{1}{2} \int_0^1 (\chi^k - \chi^{-k})(\chi^{-k} - \chi^k) = \frac{1}{2} \int_0^1 (2 - \chi^{2k} - \chi^{-2k}) = 1$$

They are orthogonal because if $k \neq l$

$$\frac{1}{2} \int_0^1 (\chi^k - \chi^{-k}) \overline{(\chi^l - \chi^{-l})} = \frac{1}{2} \int_0^1 (\chi^k - \chi^{-k})(\chi^{-l} - \chi^l) = \frac{1}{2} \int_0^1 (\chi^{k-l} + \chi^{l-k} - \chi^{k+l} - \chi^{-k-l}) = 0$$

Recall $\delta(t) = \prod_{\alpha > 0} (e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)})$. In the case of $SU(2)$, this is $\chi - \chi^{-1}$.

Consider the functions $\frac{\chi^k - \chi^{-k}}{\delta}$. Neither the numerator nor the denominator is a class function because replacing α by $-\alpha$ changes the sign of both. However the quotient remains unchanged and is a class function. We claim the new functions are orthonormal with the standard inner product $\langle f, g \rangle = \int_{SU(2)} f \bar{g}$. Using the Weyl Integral Formula, we have

$$\int_0^1 \left(\frac{\chi^k - \chi^{-k}}{\delta} \right) \overline{\left(\frac{\chi^l - \chi^{-l}}{\delta} \right)} \frac{\delta \bar{\delta}}{2} dt = \frac{1}{2} \int_0^1 (\chi^k - \chi^{-k}) \overline{(\chi^l - \chi^{-l})} = \delta_{kl}$$

To complete the argument, observe that

$$(\chi - \chi^{-1})(\chi^{k-1} + \chi^{k-3} + \dots + \chi^{-(k-1)}) =$$

$$(\chi^k + \chi^{k-2} + \dots + \chi^{-k+2}) - (\chi^{k-2} + \chi^{k-4} + \dots + \chi^{-k}) = \chi^k - \chi^{-k}$$

$$\text{so } \frac{\chi^k - \chi^{-k}}{\delta} = \chi^{k-1} + \chi^{k-3} + \dots + \chi^{-(k+1)}.$$

Remark: All of this foreshadows what will happen in general. The expression $\chi^{k+1} - \chi^{-(k+1)}$ can be easily generalized to a function $h(t_1, \dots, t_n)$ given by a simple formula. This function changes sign under reflections by simple roots. The functions are orthonormal on T using the easy inner product on T . Consequently the functions $\frac{h}{\delta}$ are invariant under the Weyl group and orthonormal on G by the Weyl Integral Formula.

Weyl found a beautiful formula for these quotient functions. They turn out to be polynomials in the χ_i and their inverses. They are invariant under the Weyl group and orthonormal.

Each polynomial has a unique term of highest degree, whose coefficient is one. These are all properties of the characters of the irreducible representations, and they uniquely characterize these characters. Consequently, Weyl's formula gives the character of an irreducible representation in all cases when such a representation exists. By the Peter-Weyl theorem, the characters of irreducible representations form a basis for the Hilbert Space of class functions. Consequently each function from Weyl's formula must be the character of an irreducible representation, so each highest root comes from a representation.

All this will be proved in detail shortly. The main difficulty is proving the Weyl Integral formula, so we turn to it next with renewed vigor.

20.3 Proof of the Weyl Integral Formula, Part 1

Every element of G is conjugate to an element of T , so it is natural to start with the map

$$\pi : G \times T \rightarrow G$$

given by $(g, t) \rightarrow gtg^{-1}$. The left G supplies the conjugation and the right T supplies the source of the conjugates, and the map spreads these elements out over G . In particular, this map is onto.

Next we work on redundancies, trying to make the map one-to-one. The most important redundancy is conjugation by elements of T , which does nothing. So we can replace the G on the left by G/T , which is a perfectly acceptable compact manifold. Now we have

$$\pi : G/T \times T \rightarrow G$$

Both sides have the same dimension, so we have gone a long way toward our goal.

The second redundancy is the Weyl group. Suppose an element (g_1T, t_1) maps to g . Recall that the Weyl group is $W = N(T)/T$. All $t \in T$ conjugate to g are equivalent by an element of the Weyl group, so we can only replace t_1 by nt_1n^{-1} . Then $gt_1g^{-1} = (gn^{-1})(nt_1n^{-1})(gn^{-1})^{-1}$, so we must replace gT by $gn^{-1}T$. This is a well-defined operation on the quotient space G/T because $g_1 \sim g_2$ implies $g_1n^{-1} \sim g_2n^{-1}$. Indeed if $g_1 = g_2t_1$, then $g_1n^{-1} = g_2t_1n^{-1} = g_2n^{-1}nt_1n^{-1}$ but n preserves T so this last element is $g_2n^{-1}t_2$. Notice that $T \subset N(T)$, but when $n \in T$ the actions n on G/T and T are the identity. So the actions of n on G/T and T depend only on the element in W represented by n .

We have proved

Lemma 46 *Two elements (g_1T, t_1) and (g_2T, t_2) map to the same element of G if and only if there is an element $w \in W$ which maps g_1T to g_2T and t_1 to t_2 .*

20.4 Proof of the Weyl Integral Formula, Part 2

We have a map $\pi : G/T \times T \rightarrow G$. We are going to compute the determinant of π^* . This is the central calculation in the proof. We will discover that this determinant is often, but not always, non-zero, so π is often, but not always, a local diffeomorphism.

The Lie algebra of G at the identity has the form $\mathcal{T} \oplus \mathcal{T}^\perp = \mathcal{T} \oplus \mathcal{P}$ using the invariant inner product on G . Extend these vectors to left invariant vector fields on G . Then the left invariant vector fields induce a canonical isomorphism $T_g(G) = \mathcal{T} \oplus \mathcal{P}$. Use this isomorphism to define an inner product on $T_g(G)$ and notice that this inner product is left and right invariant because the inner product used on $\mathcal{G}$ was invariant. The left invariant fields also give a canonical isomorphism $T_{t_1}(T)$ with $\mathcal{T}$ and $T_{g_1T}(G/T)$ with $\mathcal{P}$. So the tangent space of $G/T \times T$ at (g_1T, t_1) is also canonically isomorphic to $\mathcal{T} \oplus \mathcal{P}$. This induces an invariant metric on $G/T \times T$. We will use these maps to interpret π^* as a map from $\mathcal{T} \oplus \mathcal{P}$ to itself.

Suppose, then, that $g_1t_1g_1^{-1} = g_2$. We want to compute π^* at (g_1, t_1) . If $\gamma(t)$ is a path entirely in T starting at the identity, then $(g_1, t_1\gamma(t))$ is a path in $G/T \times T$ starting at (g_1, t_1) and $g_1t_1\gamma(t)g_1^{-1}$ is a path in G starting at g_2 and π^* sends $\frac{d}{dt}(g_1, t_1\gamma(t))$ at $t=0$ to $\frac{d}{dt}g_1t_1\gamma(t)g_1^{-1} = \frac{d}{dt}(g_1t_1g_1^{-1})(g_1\gamma(t)g_1^{-1})$ at $t=0$. Write $\frac{d}{dt}\gamma(t) = X$, a tangent vector in $\mathcal{T}$. Then $\frac{d}{dt}(g_1, t_1\gamma(t))$ is the left translate of X to (g_1, t_1) . We identify this domain of π^* with $\mathcal{T} \oplus \mathcal{P}$ using left translation, so X represents an initial vector. Then $\pi^*(X)$ in $T_{g_2}(G)$ is represented by left translation via $L_{g_1t_1g_1^{-1}}^* = L_{g_2}^*$ of $\frac{d}{dt}g_1\gamma(t)g_1^{-1} = \text{Ad}(g_1)X$, another element of $\mathcal{T}$. Putting this altogether, the map π^* is, modulo all of these identifications, a map $\mathcal{G} \oplus \mathcal{P} \rightarrow \mathcal{G} \oplus \mathcal{P}$, and this map sends $X \oplus 0$ to $\text{Ad}(g_1)(X) \oplus 0$.

We compute the rest of our map the same way. This time we start with a vector $Y \in \mathcal{P}$ and find a path $\gamma(t)$ whose derivative is Y . Think of this as a path in G . Then $(g_1\gamma(t), t_1)$ is a corresponding path in $G/T \times T$ starting at (g_1, t_1) . Since it was obtained via left translation, it represents an element in $T(G/T \times T)$ identified with $0 \oplus Y \in \mathcal{T} \oplus \mathcal{P}$. The map π sends this to $g_1\gamma(t)t_1\gamma(t)^{-1}g_1^{-1}$, a path in G starting at g_2 . The derivative of this map at $t=0$ is

$$g_1\gamma'(0)t_1\gamma^{-1}(0)g_1^{-1} + g_1\gamma(0)t_1(\gamma^{-1})'(0)g_1^{-1} = g_1\gamma'(0)t_1g_1^{-1} + g_1t_1(\gamma^{-1})'(0)g_1^{-1}$$

Since $\gamma(t)\gamma(t)^{-1} = e$, $\gamma'(0) + (\gamma^{-1})'(0) = 0$, this can be rewritten

$$g_1t_1g_1^{-1}(g_1t_1^{-1}\gamma'(0)t_1g_1^{-1} - g_1\gamma'(0)g_1^{-1})$$

This is the left translation to g_2 of $\text{Ad}(g_1)\text{Ad}(t_1^{-1})Y - \text{Ad}(g_1)Y$.

Putting these results together, π^* is essentially the map $\mathcal{T} \oplus \mathcal{P}$ to itself given by

$$(X, Y) \rightarrow (\text{Ad}(g_1)X, \text{Ad}(g_1)\text{Ad}(t_1)^{-1})Y - \text{Ad}(g_1)Y)$$

Our vector space has a canonical inner product, and we want to know how π^* modifies volume. This is given by the absolute value of the determinant of π^* . Our inner product is invariant under Ad , so it suffices to find the determinant of the map

$$(X, Y) \rightarrow (X, Ad(t_1^{-1})Y - Y)$$

Hence we need the determinant of $Ad(t_1^{-1}) - I$ on the orthogonal complement of $\mathcal{T}$ in $\mathcal{G}$. To compute further, we need not find an orthonormal basis because the determinant does not depend on the basis.

Recall that $\mathcal{G}$ has a basis consisting of a basis for $\mathcal{T}$ and then for each pair of roots a u and v such that $Ad(t)(u + iv) = e^{2\pi i \alpha(t)}(u + iv)$. The matrix for this map is

$$\begin{pmatrix} \cos 2\pi i \alpha(t) & -\sin 2\pi i \alpha(t) \\ \sin 2\pi i \alpha(t) & \cos 2\pi i \alpha(t) \end{pmatrix}$$

Changing t to $-t_1$ and subtracting the identity, we have

$$\begin{pmatrix} \cos 2\pi i \alpha(t_1) - 1 & \sin 2\pi i \alpha(t_1) \\ -\sin 2\pi i \alpha(t_1) & \cos 2\pi i \alpha(t_1) - 1 \end{pmatrix}$$

and the determinant is $2 - 2 \cos 2\pi i \alpha(t)$. Note that this formula accounts for both α and $-\alpha$. Hence the full factor is

$$\prod_{\alpha > 0} (2 - 2 \cos 2\pi i \alpha(t))$$

For later applications it is important to rewrite this as

$$\prod_{\alpha > 0} (e^{2\pi i \alpha(t)} - 1) \overline{(e^{2\pi i \alpha(t)} - 1)}$$

It is easy to check that the product of the two terms shown is $2 - 2 \cos 2\pi i \alpha(t)$.

20.5 Proof of the Weyl Integral Formula, Part 3

Recall that an element of G is regular if it is contained in only one maximal torus. Regularity is preserved by conjugation, because if g belongs to T_1 and T_2 , then $g_1 g g_1^{-1}$ belongs to $g_1 T_1 g_1^{-1}$ and $g_1 T_2 g_1^{-1}$. Consequently our map

$$\pi : G/T \times T \rightarrow G$$

maps elements (g_1, t_1) with regular t_1 to regular elements of G and conversely.

The Lie algebra $\mathcal{T}$ is the universal covering group for T . In chapter 17 we proved that an element of $\mathcal{T}$ represents a singular element if and only if $\alpha(t) \in \mathbb{Z}$ for some root α .

So the regular elements are what remains in $\mathcal{T}$ after these hyperplanes are removed. Said another way, the regular elements correspond exactly to the alcoves in the Stiefel diagram in $\mathcal{T}$.

Below is a picture of $\mathcal{T}$ for A^2 , showing the roots, the hyperplanes, and the alcoves, as well as the Fundamental Weyl Chamber.

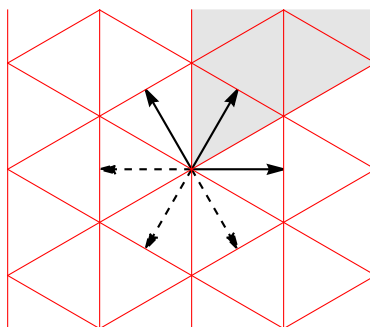


Figure 20.2: Alcoves for A^2

But this picture is lacking a fundamental piece of information. The algebra $\mathcal{T}$ also contains two lattices, the weight lattice and the dual of the coroot lattice. These lattices are pictured below. In the previous pictures, the roots are shown even though they really live in the dual space of $\mathcal{T}$. In the picture below, they have been replaced by the coroots, which do live in $\mathcal{T}$ itself. The weight lattice contains all points at intersections of the three families of lines, and the dual of the coroot lattice contains the black dots.

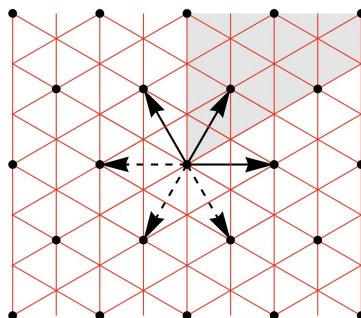


Figure 20.3: Alcoves for A^2

If $\hat{G}$ is the unique simply connected Lie group with Lie algebra $\mathcal{T}$, then the weight lattice correspond to all points mapping to the center of $\hat{G}$ and the dual of the coroot lattice corresponds to all points mapping to the identity of $\hat{G}$. Lattices between these two correspond to subgroups of the center, or in other words Lie groups with Lie algebra $\mathcal{T}$. In our case

then, the weight lattice comes from $SU(3)/Z_3$ and the root lattice comes from $SU(3)$ and there are no lattices between these two.

Once we know the lattices, we can shade a fundamental region of the torus $T \subset G$. This fundamental region is an alternate way to identify different groups with Lie algebra $\mathcal{T}$. There are many choices for fundamental region, but two such choices are shown below, for $SU(3)/Z_3$ and $SU(3)$.

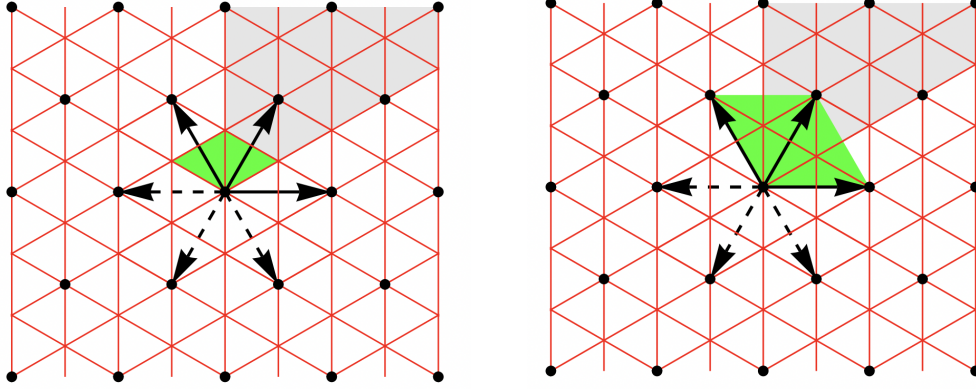


Figure 20.4: Fundamental Regions for $SU(3)/Z_3$ and $SU(3)$

The left side clearly shows that the torus $T \subset SU(3)/Z_3$ contains the images of two alcoves. These images do not intersect in T , and all other points in T are singular. The picture on the right requires more work. At first glance, it shows 12 half alcoves, and thus 6 alcoves. With more work we can see that the pieces can be dissembled and reassembled to show that these 6 alcoves are the images of 6 alcoves in $\mathcal{T}$, and thus divide $T \subset SU(3)$ into six non-overlapping open sets which cover the non-singular elements of T .

The Weyl group sends regular elements to regular elements, so if X in one alcove is mapped to $w(X)$ in a second alcove, then the entire alcove containing X is mapped in a one-to-one and onto manner to the alcove containing $w(X)$. But when we look at the images of the alcoves in T , a more interesting behavior is possible. The map sending an alcove in $\mathcal{T}$ to its image in T is always one-to-one and it follows that distinct alcoves in $\mathcal{T}$ either have the same image in T or else disjoint images in T .

It follows that when we look at the action of W on the alcove images in T , some elements of $w \in W$ may permute these images, while other elements may preserve them and move alcoves to themselves. It is easy to test this in our pictures since we understand exactly what the Weyl group does to $\mathcal{T}$ and we can determine images in T by translation.

For instance, the torus in $SU(3)/Z_3$ has exactly two alcove images and reflection generated by the horizontal roots interchanges them. It is then easy to guess that all three reflections

in W interchange the two alcoves, while the three rotations leave them setwise fixed. Since these alcoves have three sides, we conjecture that the three rotations rotate each of these alcoves by one-third of a full circle, leaving the center of the alcove fixed. It is easy to directly verify both guesses.

A slightly different story is told by $SU(3)$. This time T contains the images of six alcoves. One pair of such alcoves is shown directly in the green, the others have to be “reconstructed from partial triangles.” A little thought shows that the three rotations in W map this pair to a second and then a third pair of alcoves. The reflections, however, interchange pairs. Succinctly, the group W acts transitively on the six alcoves.

The picture simplifies when we take G/T into account. Said another way, we should pay attention to more than just decomposing T into conjugacy classes; we should also pay attention to the g needed to conjugate elements to produce a specific $g \in G$. For instance, consider the alcove for $SU(3)/Z_3$ which is preserved by three of the six elements of the Weyl group. This alcove could be divided into three sectors by drawing lines from the center of the alcove to the three vertices. Every element $g \in G$ with conjugates in this alcove will have three representatives in the alcove, but for each representative, only one $g_1 \in G/T$ conjugates the alcove element to g . However the center of the alcove is an exception; the associated g has only one representative in the alcove, but now there are three $g_1 \in G/T$ which conjugate this representative to g . So if we only look at T , then some elements have three representatives and some have only one, but if we look at $G/T \times T$ then every element has three representatives.

20.6 Proof of the Weyl Integral Formula, Part 4

The previous part contains many more ideas than we will actually need in the proof, but provide useful motivation.

Look back at the calculation of $\det(\pi^*)$ and notice that this determinant is nonzero at all regular points of T . We can restrict our previous map to these regular points, obtaining

$$G/T \times T_{reg} \rightarrow G_{reg}$$

Roughly speaking, it is enough to prove Weyl's Integral Formula for this map because the pieces that have been omitted correspond to hyperplanes in $\mathcal{T}$ and thus have measure zero, so they do not matter when we integrate. (We'll give a more precise argument later).

It turns out that after restricting to regular elements, the above map is a covering map with $|W|$ sheets. This greatly simplifies the arguments and the integral formulas, as we see shortly.

Lemma 47 *Each point in $G/T \times T_{reg}$ has an open neighborhood $\mathcal{U}$ with the following properties:*

- *The image $\pi(\mathcal{U}) \subset G_{reg}$ is an open set $\mathcal{W}$*
- *$\pi : \mathcal{U} \rightarrow \mathcal{W}$ is a diffeomorphism*
- *If $w \in W$ is an element of the Weyl group, $\pi : w(\mathcal{U}) \rightarrow \mathcal{W}$ is also a diffeomorphism*
- *The open sets $w(\mathcal{U})$ as w ranges over W are disjoint*
- *$\pi^{-1}(\mathcal{W}) = \cup_{w \in W} w(\mathcal{U})$*
- *$\pi^{-1}(\mathcal{W}) = \{p \in G/T \times T_{reg} \mid \pi(p) \in \mathcal{W}\}$*

Lemma 48 *Recall that $\pi : G/T \times T_{reg} \rightarrow G_{reg}$ is a well-defined map.*

- *The above map is a covering map, making the first space a covering space of the second space.*
- *Every point $p \in G_{reg}$ is covered by exactly $|W|$ points in $G/T \times T_{reg}$*

Proof: Clearly the second lemma follows from the first one, so it suffices to prove the first lemma. In the previous section we proved that $\det(\pi^*) \neq 0$ on regular elements. By the inverse function theorem, we can find open neighborhoods $\mathcal{U}$ and $\mathcal{W}$ in $G/T \times T$ and G such that our given point is in $\mathcal{U}$ and $\pi : \mathcal{U} \rightarrow \mathcal{W}$ is a diffeomorphism. Since the set of regular points in T is open, we can shrink this $\mathcal{U}$ so the second coordinates are all regular. The corresponding $\mathcal{W}$ shrinks to an open set, and since conjugacy preserves regularity, all of its points are also regular. This proves the first two assertions.

Elements $w \in W$ are represented by $n \in N(T)$. Recall that W acts in G/T by $w(gT) = gn^{-1}T$ and W acts on T by $w(t) = ntn^{-1}$. So W acts on $G/T \times T$ and we determined that $\pi(w(p)) = \pi(p)$ whenever $p \in G/T \times T$. Finally, W acts freely on G/T , so the only element of W leaving a point fixed is the identity. Indeed if $w(gT) = gT$, then $gn^{-1} = gt$ and canceling g gives $n = t$, and thus n represents the identity in W .

It follows that $\pi : \mathcal{U} \rightarrow \mathcal{W}$ and $\pi \circ w : \mathcal{U} \rightarrow \mathcal{W}$ are the same map, but the second map just replaces $\mathcal{U}$ with $w(\mathcal{U})$.

Suppose $p \in w_1(\mathcal{U}) \cap w_2(\mathcal{U})$. We want to prove that $w_1 = w_2$. It suffices to study the special case $p \in \mathcal{U} \cap w(\mathcal{U})$ and prove that $w = id$. So suppose that $p, q \in \mathcal{U}$ and $p = w(q)$. Therefore p and q map to the same point of G . Since π is one-to-one on $\mathcal{U}$, $p = q$. So $p = w(p)$. But W acts freely on G/T and thus on the first component of p . So $w = id$.

Now suppose (g, t) is in the inverse image of $\mathcal{W}$. Then there is an element $(g_1, t_1) \in \mathcal{U}$ mapping to the same element because $\mathcal{W}$ is defined as the image of $\mathcal{U}$. So $\pi(g, t) = \pi(g_1, t_1)$. By the last result in section 20.3, there exists $w \in W$ such that $(g, t) = (w(g_1), w(t_1))$. QED.

20.7 Intermission

In this section, we provide two examples to illustrate what we have just done. Nothing here is important in the final proof which comes next.

Example 1: Let $G = SO(3)$, whose maximal torus $T = SO(2)$ consists of rotations about the z -axis. The quotient $SO(3)/SO(2)$ is diffeomorphic to S^2 via the map which sends a rotation R to $R(v)$ where v is the unit vector in the z -direction.

The full group $SO(3)$ can be mapped to a three dimensional ball of radius $\frac{\pi}{2}$ about the origin in the following way. If $R \neq I$ is a rotation, it has an axis generated by a unit vector v . Using the right hand rule, R is rotation about this axis by an angle θ where $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$. Map R to the point θv . Note that this is ambiguous because we could have chosen $-v$ rather than v . But then the angle would have been $-\theta$ rather than θ and both calculations would give the same point θv . To complete the map, map I to the origin.

There is a problem for rotations by exactly $\frac{\pi}{2}$. In that special case, we could choose $\theta = \pm\frac{\pi}{2}$ because both angles give exactly the same rotation. Thus we would wind up with opposite points on the boundary of the ball. To fix this, we must identify opposite points on our ball, so that the image space is actually projective space P^3 .

In this picture, notice that our maximal torus maps to the z -axis from the bottom of the ball to its top. These are identified, so the maximal torus is indeed a circle. In the Stiefel diagram, the maximal torus is a straight line from the origin to the first coroot lattice point, and it passes one weight lattice point in the middle. So there are exactly two singular points in the maximal torus, which clearly are the identity and the boundary point.

The space $SO(3)/SO(2) \times T$ equals $S^2 \times [-\frac{\pi}{2}, \frac{\pi}{2}]$. And in our picture of $SO(3)$ as a ball, the ball is also an “onion” whose layers are spheres. So the mapping from $SO(3)/SO(2) \times T \rightarrow SO(3)$ is the map from the left side to a series of concentric spheres filling the ball.

There are several important things to observe about this map. The first is that the spheres are all the same size in $SO(3)/SO(2) \times T$, but have different sizes in $SO(3)$. This explains the factor $u(t)$ in the Weyl Integral Formula. The second point is that bad things happen at the singular points in T . The sphere attached to the origin is collapsed to a single point, and the sphere attached to the boundary point suddenly has opposite points identified. This is the reason that we avoid studying the map at singular points of T .

There is one other point to notice. Each of the concentric spheres hits T at two points, one above the origin and one below. Thus when we parameterize the ball by spheres, we can use either the positive portion of T or the negative portion of T to get all regular points of G . So G_{reg} is covered twice.

Example 2: Next we study the group $SU(3)/Z_3$. This time the group has dimension six,

so we can only picture the torus and its Stiefel diagram.

One purpose of this example is to issue a warning about the proof of the lemmas in the previous section. The Weyl group acts on T and (with some exceptions) the conjugacy class of $t \in T$ intersects T at $|W|$ points. Since every g is conjugate to an element of T , most $g \in G$ are conjugate to $|W|$ different elements of T . We might guess that the sheets correspond to these different conjugate elements, and consequently that when we consider a single sheet and $p \in \mathcal{U}$, then p is the only element in $\mathcal{U}$ that is conjugate to $\pi(p)$. But this is false.

In the case of $SU(3)/Z_3$ there are six fundamental alcoves in the Stiefel diagram, one in each Weyl chamber. If we project these alcoves to T via the covering map, we discover that we get only two image sets. The fundamental alcoves obtained by rotation of the alcove in the fundamental Weyl chamber map to one of these sets, and the remaining fundamental alcoves map to the other set.

By reflecting half of the images before projecting, we obtain six maps from the six fundamental alcoves to the image of the alcove in the fundamental Weyl chamber. These are the six covers of this open set which the theorem provides. So in this case each point in $\mathcal{W}$ is covered by 6 points in $G/T \times T$. However, if we only look at the T components of elements in $G/T \times T$, we find $3 \times 6 = 18$ values of t giving conjugates to g . An exception is the center of the image $\mathcal{W}$, which comes from only one point in each $\mathcal{U}$ and thus has only 6 values of the T coordinate giving its conjugate.

This paradoxical situation is resolved when we notice that $G/T \times T$ keeps track not only of conjugates but also of the elements in G which do the conjugating. For most $p \in \mathcal{W}$, once we know the t which conjugates to p then the conjugating element in G/T is unique. For the centers of the alcoves, there is only one element in the alcove which is a conjugate of p , but now there are three different gT which can perform the conjugation.

20.8 Completion of the Proof:

Lemma 49 *Suppose f is a class function on G . Then $f \circ \pi$ on $G/T \times T$ depends only on the coordinates in T .*

Proof: Suppose (g_1, t) and (g_2, t) are two points. Then $\pi(g_1, t) = g_1 t g_1^{-1}$ is conjugate to t , so $f \circ \pi(g_1, t) = f(t)$. Similarly $f \circ \pi(g_2, t) = f(t)$. QED.

Now suppose that f is a function on G_{reg} with compact support in an open $\mathcal{W}$ with the covering property. The inverse image of this set in $G/T \times T_{reg}$ is a disjoint union of $w(\mathcal{U})$ over $w \in W$, where $\pi : \mathcal{U} \rightarrow \mathcal{W}$ is a diffeomorphism. By the change of variables formula,

$$\int_{\mathcal{U}} f \circ \pi \det(\pi^*) = \int_{\mathcal{W}} f$$

and so

$$\int_{\pi^{-1}(\mathcal{W})} f \circ \pi \det(\pi^*) = |W| \int_{\mathcal{W}} f$$

Suppose next that f is an arbitrary function with compact support in G_{reg} . Using a partition of unity argument, we can find a finite number of continuous φ_i on G_{reg} such that each has compact support, each has values in $[0, 1]$, $\sum \varphi_i = 1$ on the support of f , and each φ_i has support in an open $\mathcal{W}$ with the covering property. Applying the previous result to each $\varphi_i f$ separately and summing gives

$$\int_{G/T \times T_{reg}} f \circ \pi \det(\pi^*) = |W| \int_{G_{exp}} f$$

Suppose next that f is a class function with compact support in G_{exp} . Then $f \circ w$ depends only on the t component of $G/T \times T$, and this is also true for $\det(\pi^*)$. So we can integrate by first integrating over G/T and then over T . The integrals over G/T are constant, say c , and we obtain

$$c \int_{T_{reg}} f \circ \pi \det(\pi^*) = |W| \int_{G_{exp}} f$$

Let $c \det(\pi^*) = u(t)$. We have

$$\int_{T_{reg}} f \circ \pi u(t) = |W| \int_{G_{exp}} f$$

This is very close to the Weyl Integral Formula. Our first task is to replace integrals over T_{reg} and G_{reg} with integrals over T and G and replace class functions with support in G_{reg} with class functions on all of G .

Let $\epsilon > 0$ be fixed. For each $p \in T_{reg}$, find an open neighborhood $\mathcal{U}_p$ of p which does not intersect the singular points, and find a C^∞ function on $\mathcal{U}_p$ with values in $[0, 1]$ whose support is inside $\mathcal{U}_p$ and which does not equal zero on a neighborhood $\mathcal{V}_p$ of p . For each singular point p of T , find an open neighborhood $\mathcal{U}_p$ of p such that every point of $\mathcal{U}_p$ is within ϵ of p , and find a C^∞ function on $\mathcal{U}_p$ with values in $[0, 1]$ whose support is inside $\mathcal{U}_p$ and which does not equal zero on a neighborhood $\mathcal{V}_p$ of p . Since T is compact, a finite number of these $\mathcal{V}_p$ cover T . Let φ be the sum of the functions associated with $p \in T_{reg}$ and let ψ be the sum of the remaining functions. Then the sum of these functions is never zero. Replace φ and ψ by $\frac{\varphi}{\varphi+\psi}$ and $\frac{\psi}{\varphi+\psi}$. The new functions sum to one and φ has support inside T_{reg} and ψ is only nonzero within ϵ of a singular point.

Replace $\varphi(t)$ by $\frac{1}{|W|} \sum_w \varphi(w(t))$ and replace ψ with a similar expression. Then φ and ψ are class functions on T and sum to 1 and φ has compact support in T_{reg} and $\psi(t)$ is non-zero only if t is within ϵ of a singular point.

Extend φ and ψ to be defined on all of G as follows: if $g \in G$, then g is conjugate to an element of T . Define $\varphi(g) = \varphi(t)$. This is independent of the choice of t because φ is a class function on T . Extend ψ similarly.

If f is a class function defined on all of G , then write $f = \varphi f + \psi f$. Then both terms are also class functions and the first has compact support in G_{reg} and the second is non-zero on g only if the distance from g to a singular point is smaller than ϵ .

Note that

$$\begin{aligned} \int_G f &= \int_G \varphi f + \int_G \psi f \\ c \int_T f \det(\pi^*) &= c \int_T \varphi f \det(\pi^*) + c \int_T \psi f \det(\pi^*) \end{aligned}$$

Since φf is a class function with compact support in G_{reg} , a previous result gives

$$c \int_T \varphi f \circ \pi \det(\pi^*) = |W| \int_G \varphi f$$

We would like to prove that

$$c \int_T f \circ \pi \det(\pi^*) = |W| \int_G f$$

To show this, we will argue that $\int_{G/T \times T} \psi f \det(\pi^*)$ and $\int_G \psi f$ are both small. These integrals are at most the maximum value of f times the area of the hyperplanes bounding G_{reg} and T_{reg} times the maximum of $\det(\pi^*)$ times 2ϵ . Here by “area of a hyperplane” we mean the $(n-1)$ -dimensional volume if n is the dimension of G , with a similar meaning for the case of T . All of these quantities are fixed except ϵ . By rechoosing φ and ψ , we can make ϵ go to zero and thus insure that the two desired sides of the Weyl Integral Formula are as close as we like.

We are almost, but not quite, done. We still have a constant c , which we want to prove equals 1. The constant is the volume of G/T ; notice that this manifold has a natural Riemannian structure which was inherited from G . To prove that $c = 1$, it is enough to prove that

$$\text{vol}(G/T) \times \text{vol}(T) = \text{vol}(G)$$

because we normalized metrics to make $\int_G 1 = 1$ and $\int_T 1 = 1$.

Notice that we have a natural map $\tau : G \rightarrow G/T$ given by $\tau(g) = gT$. By construction of G/T , each point $p \in G/T$ has an open neighborhood $\mathcal{U} \subset G/T$ and a map $r : \mathcal{U} \rightarrow G$ such that $\tau \circ r = \text{id}$. In other words, the map r selects a representative for each class gT in $\mathcal{U}$. We then get a natural diffeomorphism between $\tau^{-1}(\mathcal{U}) \subset G$ and $\mathcal{U} \times T \subset G/T \times T$ defined by mapping g to gT , finding the representative $r(gT)$ of this set, consequently writing

$g = r(gT)t_1$ and finally mapping g to $gT \times t_1 = gT \times r(gT)^{-1}g$. The inverse map starts with (gT, t) , finds $r(gT) \in G$, and maps (gT, t) to $r(gT)t$.

Both of these maps are easily seen to be isometries since the metric on G is left-invariant, etc. We conclude that $\text{vol}(\tau^{-1}(\mathcal{U})) = \text{vol}(\mathcal{U})\text{vol}(T)$. We normalized metrics so $\int_T 1 = 1$, and therefore $\text{vol}(\tau^{-1}(\mathcal{U})) = \text{vol}(\mathcal{U})$. Of course the left side of this equation is the volume of an open subset of G and the right side is the volume of an open subset of G/T and these two spaces have different dimensions. This doesn't matter; their volumes are equal.

Remark: Do not confuse our locally defined map $G/T \times T \rightarrow G$ with the earlier map $G/T \times T \rightarrow G$. That earlier map was given by conjugation, while the new map is much more straightforward. The earlier map was W to one, while the new map is one-to-one. The original map had singularities, while the new map does not. The original map did not preserve metrics, but the new map does.

The only problem is that the new map is only defined locally because $G \rightarrow G/T$ is a fibre bundle which has local sections, but no global section. So the final step of the argument is to cut G/T into pieces $\mathcal{U}_i$ and conclude that the volume of G/T is the sum of the volumes of the pieces, which is equal to the sum of the volumes of $\tau^{-1}(\mathcal{U}_i)$, and thus the volume of G . Since we normalized G to have volume 1 by requiring that $\int_G 1 = 1$, it follows that the volume of G/T is 1 and thus $c = 1$. QED.

This argument isn't quite rigorous because we cannot cut a connected space into disjoint open pieces, but the idea is clear. An overly anxious reader can use partitions of unity arguments or other tools to clean up the argument if desired.

Whew. QED.

Chapter 21

The Weyl Character Formula

21.1 Alternating Functions on $\mathcal{T}$

In this chapter we assume that G is a compact simply connected group. Thus the irreducible representations of G are the same as the irreducible representations of $\mathcal{G}$. In particular, T is the maximal torus of G , and thus $\mathcal{T}$ modulo the coroot lattice.

The previous chapter introduced $u(t)$, and the surprising fact that $u = \delta\bar{\delta}$; thus in some sense u has a natural square root. Although u is invariant under the Weyl group, δ is not. The Weyl group contains isometries of $\mathcal{T}$ and therefore any $w \in W$ satisfies $\det(w) = \pm 1$. If $\det(w) = 1$, $\delta(w(t)) = \delta(t)$, but otherwise $\delta(w(t)) = -\delta(t)$.

It is easy to prove this. Recall that the Weyl group is generated by reflections associated with simple roots. If we reflect across the hyperplane perpendicular to a simple α , then α maps to $-\alpha$, but all other positive roots remain positive. A glance at the formula for δ immediately reveals that it changes sign. It follows that every product of reflections across simple roots with an even number of terms has determinant 1 and leaves δ invariant and any product of reflections across simple roots with an odd number of terms has determinant -1 and changes the sign of δ . Since these reflections generate W , this rule applies to all elements of w . Since any reflection has determinant -1 , all reflections change the sign of δ .

Call a function on $\mathcal{T}$ *symmetric* if it is invariant under the Weyl group and *alternating* if it is invariant under elements in W with positive determinant and changes sign under elements with negative determinant.

Remark: In this chapter we find an explicit formula for the characters of G . We are guided by the known properties of these characters:

- The characters of irreducible representations of G have the following form, where the d_λ are integers and the λ are a finite number of elements from the weight lattice:

$$\chi(t) = \sum_{\text{weights } \lambda} d_\lambda e^{2\pi i \lambda(t)}$$

This formula is defined on $\mathcal{T}$, but since weight vectors in $\mathcal{T}^*$ take integer values on the lattice dual to the weight lattice, and since this dual lattice is contained in the coroot lattice, which consists of elements of $\mathcal{T}$ send to the identity of our G by the covering map, the formula induces a function on T .

- The characters are invariant under the Weyl group.
- The characters of irreducible representations are orthonormal using the inner product

$$\int_T f(t) \overline{g(t)} u(t) dt \quad \text{where} \quad u(t) = \frac{\delta \bar{\delta}}{|W|}$$

- The function $\chi(t)$ contains a unique highest weight in the fundamental Weyl chamber, and the coefficient of this element is $d_\lambda = 1$.

Definition 45 A character function is a complex-valued function on $\mathcal{T}$ given by a finite sum of the following form with real coefficients r_λ :

$$\sum_{\text{weights } \lambda} r_\lambda e^{2\pi i \lambda(t)}$$

Remark: Since character functions can be added and multiplied by real scalars, they form a vector space. The $e^{2\pi i \lambda(t)}$ are linearly independent and thus form a basis for this vector space. Indeed

Theorem 112 Let V be an arbitrary finite dimensional real vector space and assume that $f_1, \dots, f_k$ are distinct linear maps $V \rightarrow R$. Then $e^{2\pi i f_1}, \dots, e^{2\pi i f_k}$ are linearly independent over R .

Proof of theorem: We claim there is a ray from the origin in V such that $f_1, \dots, f_k$ are different numbers on this ray. Indeed if $f_i \neq f_j$, then the kernel of $f_i - f_j$ is not everything and thus is a hyperplane in V . Varying i and j , we obtain a finite number of hyperplanes, and so there is a non-zero vector which does not belong to any hyperplane.

Thus we can reduce to the special case when the dimension of V is one, and replace the f_i by a finite number of real numbers $r_1, \dots, r_k$. We must prove that the $e^{2\pi i r_j}$ are linearly

independent. Note that $\frac{d}{dt}e^{2\pi i r_j} = 2\pi i r_j e^{2\pi i r_j}$ and thus our functions are eigenfunctions of $\frac{d}{dt}$ with distinct eigenvalues. By an easy argument, they must be linearly independent. QED.

Remark: Instead of studying characters χ , it is convenient to study $\chi\delta$. The resulting functions are alternating rather than symmetric, but the advantage is that orthogonality is given by a much easier formula because

$$\int_T \chi_1 \overline{\chi_2} u(t) dt = \int_T \chi_1 \overline{\chi_2} \frac{\delta \overline{\delta}}{|W|} dt = \frac{1}{|W|} \int_T (\chi_1 \delta) \overline{(\chi_2 \delta)} dt$$

It is easy to produce alternating character functions, since one exists for each element of the weight lattice:

Definition 46 *If λ is a member of the weight lattice, let*

$$\chi_\lambda(t) = \sum_{w \in W} \det(w) e^{2\pi i \lambda(wt)}$$

Note that such a λ belongs to $\mathcal{T}^*$; it is a single point in this space.

Theorem 113 *The $\chi_\lambda(t)$ have the following properties:*

- $\chi_\lambda(w(t)) = \chi_{(\lambda \circ w)}(t) = \chi_{w^{-1}\lambda}(t)$
- $\chi_\lambda(t)$ is identically zero if and only if $\lambda \circ w = \lambda$ for some $w \neq e$ if and only if λ is in the wall of some Weyl chamber.
- If λ is a point in the weight lattice which is not in a wall of a Weyl chamber, the sum χ_λ contains $|W|$ distinct elements which are orthonormal using the standard inner product $\int_T f(t) \overline{g(t)} dt$ on T . There is one of these elements in each open Weyl chamber.
- The χ_λ associated with elements in the weight lattice in the interior of the fundamental Weyl chamber are orthonormal using the inner product $\frac{1}{|W|} \int_T f(t) \overline{g(t)} dt$ on T .

Proof: Recall that the Weyl group is a group of transformations of $\mathcal{T}$ generated by reflections across hyperplanes $\alpha(t) = 0$ for various roots α . So W acts on $\mathcal{T}$.

Roots and weights live in $\mathcal{T}^*$. We have a canonical isomorphism between this space and $\mathcal{T}$ given by $\varphi(t) = \langle \varphi, t \rangle$. If $w \in W$, we have $\langle \varphi, w^{-1}t \rangle = \langle w\varphi, ww^{-1}t \rangle = \langle w\varphi, t \rangle$. The conclusion is that W also acts on $\mathcal{T}^*$, and for consistency with the isomorphism between $\mathcal{T}^*$ and $\mathcal{T}$, this action is

$$w\varphi(t) = \varphi(w^{-1}t)$$

With this prelude, we are ready to prove the first statement of the theorem:

$$\begin{aligned}\chi_\lambda(w(t)) &= \sum_{w_1 \in W} \det(w_1) e^{2\pi i \lambda(w_1(w(t)))} = \sum_{w_1 \in W} \det(ww_1w^{-1}) e^{2\pi i \lambda((ww_1w^{-1})(w(t)))} = \\ &= \sum_{w_1 \in W} \det(w_1) e^{2\pi i \lambda((w_1w)(t))} = \sum_{w_1 \in W} \det(w_1) e^{2\pi i (\lambda \circ w)(w_1(t))}\end{aligned}$$

Proof continued: The Weyl group acts simply transitively on the open Weyl chambers. Therefore if $\lambda = \lambda \circ w = w^{-1}\lambda$ for $w \neq e$, then λ lies on the wall of a Weyl chamber and so there is a reflection $r \in W$ with $r(\lambda) = \lambda$ or equivalently $\lambda \circ r = \lambda$. Note that as w runs over W , so does rw . So

$$\begin{aligned}\chi_\lambda(t) &= \sum_w \det(w) e^{2\pi i \lambda(w(t))} = \sum_w \det(rw) e^{2\pi i \lambda(rw(t))} = \\ &= - \sum_w \det(w) e^{2\pi i (\lambda \circ r)(w(t))} = - \sum_w \det(w) e^{2\pi i \lambda(w(t))} = -\chi_\lambda(t)\end{aligned}$$

Now suppose conversely that $\chi_\lambda = 0$. We will prove that $\lambda \circ w = \lambda$ for some $w \neq e$. Otherwise the terms in the formula for χ_λ are all different, but these terms are linearly independent by the previous theorem.

Proof continued: We want to prove that

$$\int_T e^{2\pi i \lambda_j(t)} \overline{e^{2\pi i \lambda_k(t)}} dt = \int_T e^{2\pi i (\lambda_j - \lambda_k)(t)} dt = \delta_{jk}$$

If $j = k$, this is obvious because $\int_T 1 = 1$. Otherwise $\lambda_j - \lambda_k$ is a nonzero element of the weight lattice by previous parts of the lemma. Call this λ ; we want to prove that $\int_T e^{2\pi i \lambda(t)} dt = 0$.

This follows from the orthogonality relations applied to complex irreducible representations of the compact group T . Since T is abelian, all irreducible representations are abelian, and thus given by characters of T . At the very beginning of our study of roots, we found that all homomorphisms from T to S^1 have the form $e^{2\pi i \lambda(t)}$ where $\lambda : \mathcal{T} \rightarrow \mathbb{R}$ is linear and takes integer values on the lattice points of the lattice of points in $\mathcal{T}$ which map to $e \in T$. We are dealing with the simply connected compact group G , so that lattice is the root lattice. But weights even take integer values on the weight lattice, which is contained in the root lattice. So $e^{2\pi i \lambda(t)}$ is a non-trivial irreducible representation of T and its inner product with the identity representation is zero. QED.

Completion of Proof: If a $\chi_\lambda(t)$ is not zero, then it has a term in the interior of the fundamental Weyl Chamber. The other terms of the sum are obtained by applying elements of W to this initial term. The new terms belong to other Weyl chambers and each Weyl

chamber has exactly one of these terms. Terms with $\det(w) = 1$ are rotations of the original term, and terms with $\det(w) = -1$ are reflections of the original term.

For ease of notation, denote the characters in the sum by $\chi_i(t)$ for $1 \leq i \leq |W|$. Then

$$\begin{aligned} \frac{1}{|W|} \int_T \chi_\lambda \overline{\chi_\lambda} &= \frac{1}{|W|} \int_T \left(\sum \det(w_i) \chi_i \right) \overline{\left(\sum \det(w_j) \chi_j \right)} = \\ \frac{1}{|W|} \sum_i \det(w_i)^2 \int_T \chi_i \overline{\chi_i} &+ \frac{1}{|W|} \sum_{i \neq j} \det(w_i) \det(w_j) \int_T \chi_i \overline{\chi_j} = 1 + 0 = 1 \end{aligned}$$

When $\lambda_1 \neq \lambda_2$, the term from χ_{λ_1} and the term from χ_{λ_2} in the fundamental chamber are different, and it follows that the corresponding terms in other Weyl chambers are also different. When we form the inner product of a term from χ_{λ_1} and a term from χ_{λ_2} , the terms are certainly different and thus orthogonal if they are in different Weyl chambers, but even if the terms are in the same Weyl chamber, the terms are different and their inner product is zero. So χ_{λ_1} and χ_{λ_2} are orthogonal. QED.

Theorem 114 *Let V be the vector space of alternating characters on $\mathcal{T}$, and give V the inner product $\frac{1}{|W|} \int_T f(t) \overline{g(t)} dt$. Then every element of V can be written uniquely as a finite linear combination of the functions χ_λ as λ ranges over weights in the interior of the fundamental Weyl chamber. These basis functions are orthonormal.*

Proof: Much of this has been proved already. It is enough to show that the terms described generate the space. So suppose $f \in V$. For each term of the form $r_\lambda e^{2\pi i \lambda(t)}$ in f with λ in the interior of the fundamental Weyl chamber, form χ_λ and subtract $r_\lambda \chi_\lambda$ from f . It suffices to show that the remaining expression is zero. This expression contains no terms with weights from the interior of the fundamental Weyl Chamber and is invariant under the Weyl group up to sign, so it contains no terms with weights in any open Weyl chamber. Thus it only contains terms with λ in a wall of the Weyl chamber.

Repeat the argument for terms $r_\lambda e^{2\pi i \lambda(t)}$ with entries in the walls of the fundamental Weyl chamber. This time the terms being subtracted equal zero. Each term contains only one entry in a wall of the fundamental Weyl chamber, and when these terms are removed, the expression contains no terms in the closure of the fundamental Weyl chamber. Since f is invariant under W , it contains no terms at all. QED.

21.2 A Formula for δ

Definition 47 *The element $\delta \in \mathcal{T}^*$ is defined as*

$$\delta = \frac{\sum_{\alpha > 0} \alpha}{2}$$

Remark: Notice that whenever α is a root, $\frac{\alpha}{2}$ belongs to the weight lattice. Hence δ is a member of the weight lattice.

Remark: This notation may seem unfortunate since we have used δ in the previous section as one of the terms in the Weyl Integral Formula. However the old and new terms are related as follows:

Theorem 115 *Since δ belongs to the weight lattice, we can form*

$$\chi_\delta(t) = \sum_{w \in W} \det(w) e^{2\pi i \delta(t)}$$

Then

$$\chi_\delta(t) = \prod_{\alpha > 0} (e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)}) = \text{the Weyl Integral Term } \delta(t)$$

Proof: To avoid confusion, let us temporarily write $D(t) = \prod_{\alpha > 0} (e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)})$. One advantage of this expression is that it immediately shows that $D(t)$ is alternating. But to prove the result we need to rewrite this expression as

$$D(t) = \prod_{\alpha > 0} e^{\pi i \alpha(t)} \prod_{\alpha > 0} (1 - e^{-2\pi i \alpha(t)}) = e^{2\pi i \delta(t)} \prod_{\alpha > 0} (1 - e^{-2\pi i \alpha(t)})$$

The right hand side of the final equation shows that the weights which appear in the formula for $D(t)$ coming from the fundamental Weyl chamber are the weight δ and smaller weights, since in the formula we subtract positive roots from the weight $\delta(t)$. Moreover, the coefficient of $e^{2\pi i \delta(t)}$ in the expression for $D(t)$ is one. Thus we can subtract χ_δ from this formula and obtain an expression with only smaller weights. So $D(t) = \chi_\delta(t) + D_1(t)$ where D_1 is orthogonal to $\chi_\delta(t)$. It follows that $\|D(t)\|^2 = \|\chi_\delta(t)\|^2 + \|D_1(t)\|^2$.

To complete the proof, we will show that $\|D(t)\| = \|\chi_\delta\|$. It then follows that $\|D_1\| = 0$ and so D_1 is identically zero.

Apply the Weyl Integral Formula to the function on G which is identically one. This is a class function, and therefore

$$1 = \int_G 1 = \frac{1}{|W|} \int_T D(t) \overline{D(t)} dt = \|D(t)\|^2$$

On the other hand, $\delta(t)$ is a weight and therefore by the last part of theorem 113, we have

$$1 = \frac{1}{|W|} \int_T \chi_\delta \overline{\chi_\delta} = \|\chi_\delta\|^2$$

QED.

21.3 The Weyl Character Formula

Theorem 116 (The Weyl Character Formula) *By earlier results, every irreducible representation of G has a highest weight λ in the closure of the fundamental Weyl chamber, and this highest weight determines the representation up to equivalence. The character of this representation on T induces a function on $\mathcal{T}$. This function is*

$$\frac{\chi_{\lambda+\delta}}{\chi_\delta} = \frac{\sum_{w \in W} \det(w) e^{2\pi i(\lambda+\delta)(w(t))}}{\sum_{w \in W} \det(w) e^{2\pi i\delta(w(t))}}$$

Proof: Let $\Pi(t)$ be the character of the representation associated with λ . We will prove that $\Pi(t)\chi_\delta = \chi_{\lambda+\delta}$. Note first that the expression on the left is an alternating character, and thus a finite sum of χ_τ as τ ranges over weights in the interior of the fundamental Weyl chamber.

The character Π has a term $e^{2\pi i\lambda}$ where λ is a highest weight of the representation. The term χ_δ has a term $e^{2\pi i\delta}$. Consequently the product has a term $e^{2\pi i(\lambda+\delta)}$ with coefficient one. Note that $\lambda + \delta$ belongs to the interior of the fundamental Weyl chamber. It follows that $\chi_{\lambda+\delta}$ occurs in the finite sum of χ_τ giving the product, and with coefficient one.

The various χ_τ in this sum are orthonormal using the inner product $\frac{1}{|W|} \int_T f(t) \overline{g(t)} dt$. If we can prove that $\Pi(t)\chi_\delta(t)$ has norm one, then $\chi_{\lambda+\delta}$ is the only term in the finite sum representing the product, and we are done.

Thus we must compute $\frac{1}{|W|} \int_T (\Pi \chi_\delta) \overline{(\Pi \chi_\delta)} dt$. But we earlier proved that $\chi_\delta = \delta(t)$, the function which occurs in the $u(t)$ term for the Weyl Integral. So the expression we are computing is

$$\int_T \Pi(t) \overline{\Pi(t)} \frac{\delta(t) \overline{\delta(t)}}{|W|} dt = \int_T \Pi(t) \overline{\Pi(t)} u(t) dt = \int_G \Pi(g) \overline{\Pi(g)} dg$$

and this expression equals one because characters are orthonormal. QED.

Theorem 117 *Every highest weight comes from an irreducible representation of G .*

Proof: The idea of the proof is straightforward, but one detail is a little messy. Use the Weyl character formula to define the character of the missing representation associated to λ . Since the character formula is a quotient and the denominator is sometimes zero, it is necessary to show that this quotient can be continuously extended to all of $\mathcal{T}$. This will be proved at the end.

The quotient is invariant under W , so it defines a class function on all of G . We will prove that this class function has norm one and is orthogonal to the characters of irreducible representations. Once we prove this, we are done, because the Peter Weyl theorem asserts

that the characters of irreducible representations form a Hilbert space basis for the class functions on G . If our quotient did not come from an irreducible representation, it would be orthogonal to all characters of irreducible representations, and hence zero in the Hilbert space. So it certainly could not have norm one.

Call our quotient $\Pi(t)$ and let $\Pi_1(t)$ be either $\Pi(t)$ or else the character of an irreducible representation. Then

$$\begin{aligned} \int_G \Pi(g) \overline{\Pi_1(g)} dg &= \int_T \left(\frac{\chi_{\lambda+\delta}(t)}{\chi_\delta(t)} \right) \overline{\left(\frac{\chi_{\lambda_1+\delta}(t)}{\chi_\delta(t)} \right)} u(t) dt = \\ \frac{1}{|W|} \int_T \left(\frac{\chi_{\lambda+\delta}(t)}{\chi_\delta(t)} \right) \overline{\left(\frac{\chi_{\lambda_1+\delta}(t)}{\chi_\delta(t)} \right)} \delta(t) \overline{\delta(t)} dt &= \frac{1}{|W|} \int_T \chi_{\lambda+\delta}(t) \overline{\chi_{\lambda_1+\delta}(t)} dt \end{aligned}$$

But we earlier proved that the $\chi_\tau(t)$ are orthonormal in this inner product as τ runs over weights in the interior of the fundamental Weyl chamber.

Finally we return to the more difficult point of showing that $\frac{\chi_{\lambda+\delta}}{\chi_\delta}$ can be extended to a continuous function on T . We first examine the denominator

$$\chi_\delta(t) = \prod_{\alpha > 0} \left(e^{\pi i \alpha(t)} - e^{-\pi i \alpha(t)} \right) = \prod_{\alpha > 0} (2i \sin(\pi \alpha(t))) = (2i)^N \prod_{\alpha > 0} \sin(\pi \alpha(t))$$

where N is the number of positive roots. This function vanishes exactly when $\alpha(t)$ is an integer for some positive α , and so exactly on the hyperplanes defining the Stiefel diagram. If α is a positive root and n is an arbitrary integer, let $V_{\alpha,n}$ denote the hyperplane $\alpha(t) = n$. When written as a power series expansion in the coordinates of $\mathcal{T}$, $\chi_\delta(t)$ has order one if a point is on one hyperplane, order two if on two intersecting hyperplanes, etc.

To finish the argument, we need to show that the numerator also vanishes on these hyperplanes, and to an order equal to or greater than the order of $\chi_\delta(t)$. So consider

$$\chi_{\lambda+\delta}(t) = \sum_w \det(w) e^{2\pi i (\lambda+\delta)(w(t))}$$

Recall that $\lambda + \delta$ is an element of the weight lattice, and this lattice is dual to the coroot lattice. Recall also that the coroot lattice is exactly the set of elements of $\mathcal{T}$ sent to the identity by the covering map $\mathcal{T} \rightarrow T$, in the case that G is simply connected with maximal torus T . It follows that the function $e^{2\pi i (\lambda+\delta)(t)}$ equals 1 if t is in the coroot lattice, and thus is invariant under translations by elements of the coroot lattice. The coroot lattice is invariant under W , so the function $e^{2\pi i (\lambda+\delta)(w(t))}$ is also invariant under translations by coroot lattice points. Therefore the function $\chi_{\lambda+\delta}(t)$ is similarly invariant under these translations.

Suppose t is in the hyperplane $V_{\alpha,n}$. Let w be reflection across $V_{\alpha,0}$; it is an element in the Weyl group. Then

$$w(t) = t - \frac{2 \langle \alpha, t \rangle}{\langle \alpha, \alpha \rangle} \alpha = t - n \frac{2\alpha}{\langle \alpha, \alpha \rangle}$$

So $\chi_{\lambda+\delta}(w(t)) = \chi_{\lambda+\delta}(t)$ because $w(t)$ and t differ by a coroot translation, and on the other hand $\chi_{\lambda+\delta}(w(t)) = -\chi_{\lambda+\delta}(t)$ because $\det(w) = 1$. Since both are true, $\chi_{\lambda+\delta}$ vanishes on $V_{\alpha,n}$.

It follows that

$$\frac{\chi_{\lambda+\delta}(t)}{\sin \pi \alpha(t)}$$

is continuous over all hyperplanes parallel to the hyperplane perpendicular to α .

To complete the argument, we form

$$\frac{\chi_{\lambda+\delta}(t)}{\prod_{\alpha > 0} \sin \pi \alpha(t)}$$

This function is well-behaved except over points which belong to more than one hyperplane. We will argue that it is also well-behaved at these points.

Before giving details, let us understand the difficulty. When we divided $\chi_{\lambda+\delta}$ by $\sin \pi \alpha(t)$, we cancelled a zero of both functions on a family of hyperplanes associated with α . Perhaps we cancelled all of the zeros on this hyperplane, so when it intersects a second hyperplane in the family associated with root β , there is no zero in the numerator for the denominator to cancel, so dividing by $\sin \pi \beta(t)$ produces a singularity. The reason this does not happen, of course, is that the numerator has a higher order zero at intersections of hyperplanes. Therefore it is not possible to prove the theorem by simply determining where $\chi_{\lambda+\delta}$ vanishes. We must keep track of the order of vanishing by means of a power series or similar device.

Therefore let us consider a point p in the intersection of several hyperplanes, and expand $\chi_{\lambda+\delta}$ in a power series about p ; this is clearly possible because the terms of χ are simple analytic functions. We first analyze the step of dividing χ just by $\sin \pi \alpha(t)$. Choose coordinates near p , say $t_1, \dots, t_k$, so the coordinates at p are all zero and the last coordinate points in the direction of $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$. Our space is just $\mathcal{T}$, so these can be obtained from the standard linear coordinates by an affine transformation and a translation. Then $\alpha(t) = m + ct_k$ if we assume that $p \in V_{\alpha,m}$; here c is a non-zero constant. Expanding $\sin \pi \alpha(t)$ in a series, we have $\sin \pi \alpha(t) = 0 + \pi t_k + \dots$. Since χ vanishes on the hyperplane given by $t_k = 0$, every term of the expansion contains t_k . We can cancel one t_k from both numerator and denominator and obtain a quotient power series we will call $q(t)$ which converges near p .

The function $\sin \pi \alpha(t)$ vanishes on other hyperplanes associated with α , so dividing series might cause trouble at these other zeros. But we are only interested in the behavior of our series near p , so we assume it is only defined on an open neighborhood of p which does not intersect other hyperplanes in the α -series.

Note that $\frac{\chi_{\lambda+\delta}(t)}{\sin \pi \alpha(t)} = q(t)$ near p . In particular, $q(t)$ is zero at spots that χ is zero, except on the particular hyperplane where zeros are cancelled.

Now suppose β is a second positive root, and p also lies on a hyperplane in the β -series. Choose new coordinates as before near p , but this time let t_k point in the direction of $\frac{2\beta}{\langle \beta, \beta \rangle}$. Rewrite $q(t)$ as a series in these new coordinates. The hyperplane from the β -series containing p is given by $t_k = 0$ and by assumption $q(t)$ vanishes on this hyperplane except possibly on intersection points with the α hyperplane. But these intersection points have codimension 1 and thus $q(t)$ vanishes on a dense subset of the set where $t_k = 0$ and by continuity on the entire set. So in the expansion of $q(t)$ every term contains t_k . The argument is now exactly as before and we conclude that near p we can write $\frac{q(t)}{\sin \pi \beta(t)} = q_1(t)$ where q_1 vanishes where χ vanishes, except possibly on the α and β hyperplanes.

Continue the argument for remaining hyperplanes through p . QED.

21.4 The Weyl Dimension Formula

When $\chi(t)$ is the character of a representation, $\chi(0)$ is the trace of the identity map and thus the dimension of the representation. Consequently, the Weyl Character Formula allows us to determine the dimensions of the representations associated to highest weights λ . The argument requires care, because the denominator of the character formula vanishes at 0.

Theorem 118 (The Weyl Dimension Formula) *The dimension of the irreducible representation with highest weight λ is*

$$\prod_{\alpha > 0} \frac{\langle \alpha, \lambda + \delta \rangle}{\langle \alpha, \delta \rangle}$$

Remark: Notice that multiplying the inner product by a positive constant does not change this number, so the formula depends on angles between roots and weights, but not on lengths of these roots and weights.

Proof: Let $\delta = \frac{\sum_{\alpha > 0} \alpha}{2} \in \mathcal{T}^*$. Let $\hat{\delta}$ denote the corresponding element of $\mathcal{T}$, so for every element $\alpha \in \mathcal{T}^*$ we have $\alpha(\hat{\delta}) = \langle \alpha, \delta \rangle$. Consider the path $\gamma(s) = \hat{\delta}s$ in $\mathcal{T}$ defined for $s \geq 0$. Notice that this path is in the open fundamental Weyl chamber when $s > 0$.

Then

$$\begin{aligned}
\chi_{\lambda+\delta}(\hat{\delta}s) &= \sum_w \det(w) e^{2\pi i s(\lambda+\delta)(w(\hat{\delta}))} = \sum_w \det(w) e^{2\pi i s w^{-1}(\lambda+\delta)(\hat{\delta})} = \\
&= \sum_w \det(w) e^{2\pi i s \langle w^{-1}(\lambda+\delta), \delta \rangle} = \sum_w \det(w) e^{2\pi i s \langle \delta, w^{-1}(\lambda+\delta) \rangle} = \\
&= \sum_w \det(w) e^{2\pi i s \langle w(\delta), \lambda+\delta \rangle} = \sum_w \det(w) e^{2\pi i s w(\delta)(\hat{\lambda}+\hat{\delta})} = \\
&= \sum_w \det(w) e^{2\pi i s \delta(w^{-1}(\hat{\lambda}+\hat{\delta}))} = \sum_w \det(w) e^{2\pi i s \delta(w(\hat{\lambda}+\hat{\delta}))} = \\
&= \chi_{\delta}s(\hat{\lambda} + \hat{\delta}) = \prod_{\alpha > 0} \left(e^{\pi i \alpha(s(\hat{\lambda}+\hat{\delta}))} - e^{-\pi i \alpha(s(\hat{\lambda}+\hat{\delta}))} \right) = \\
&= \prod_{\alpha > 0} \left(e^{\pi i \langle \alpha, s(\lambda+\delta) \rangle} - e^{-\pi i \langle \alpha, s(\lambda+\delta) \rangle} \right) = \\
&= \prod_{\alpha > 0} (2\pi i < \alpha, \lambda + \delta > s + \text{terms in higher powers of } s) = \\
&= (2\pi i)^N s^N \prod_{\alpha > 0} < \alpha, \lambda + \delta > + \text{terms in higher powers of } s
\end{aligned}$$

In particular, when $\lambda = 0$ we obtain

$$\chi_{\delta}(\hat{\delta}s) = (2\pi i)^N s^N \prod_{\alpha > 0} < \alpha, \delta > + \text{terms in higher powers of } s$$

So

$$\Pi_{\lambda}(\hat{\delta}s) = \prod_{\alpha > 0} \frac{\langle \alpha, \lambda + \delta \rangle}{\langle \alpha, \delta \rangle} + \text{terms in } s^k \text{ for } k \geq 1$$

and taking the limit as $s \rightarrow 0$ gives the required result. QED.

21.5 Applications of the Dimension Formula

When applying the dimension formula, it is useful to think of it in a slightly different form:

Theorem 119 *Let λ be the highest weight of an irreducible representation of a simply-connected G . Then the dimension of this representation is*

$$\prod_{\alpha > 0} \frac{(\lambda + \delta) \left(\frac{2\alpha}{\langle \alpha, \alpha \rangle} \right)}{\delta \left(\frac{2\alpha}{\langle \alpha, \alpha \rangle} \right)}$$

Proof: This is our previous formula with extra factors of $\frac{2}{\langle \alpha, \alpha \rangle}$ in both numerator and denominator, but these extra factors cancel out.

Remark: This form is useful because λ and δ are weights in the closure of the fundamental Weyl chamber, and thus have coordinates $(k_1, k_2, \dots, k_m)$ where the k_i are non-negative integers. If λ is a weight with these coordinates, then

$$\lambda \left(\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle} \right) = k_i$$

for all simple roots α_i . Finally any positive root α is a sum of simple roots with integer coefficients:

$$\alpha = n_1\alpha_1 + \dots + n_m\alpha_m$$

and the coroots of the simple roots form a lattice basis for the coroot lattice, so

$$\frac{2\alpha}{\langle \alpha, \alpha \rangle} = p_1 \frac{2\alpha_1}{\langle \alpha_1, \alpha_1 \rangle} + \dots + p_m \frac{2\alpha_m}{\langle \alpha_m, \alpha_m \rangle}$$

The only warning needed is that the coefficients p_i are not necessarily equal to the n_i . This situation is important when the Dynkin diagram has roots of different lengths.

Using this formulation, no inner products need be computed and the only task is finding the integers p_i .

Remark: Recall that we are dealing with diagrams and lattices in both $\mathcal{T}^*$ and $\mathcal{T}$. Since the dimension formula involves roots and weights, it is the $\mathcal{T}^*$ diagram that concerns us. This diagram contains roots and the root lattice, the weight lattice, and the fundamental Weyl Chamber. Actual distances in the diagram are not important because the invariant metric on $\mathcal{T}^*$ is only defined up to a positive multiplicative scalar (when the Dynkin diagram is connected). This allows us to fix a size at random. When all weights have the same length, we will choose the scalar to make all weights have length one. When weights have two different lengths, we will choose the scalar so short weights have length one; then the long weights have a length which can be computed.

The simple roots $\alpha_1, \dots, \alpha_m$ form a lattice basis for the root lattice. However, the weight lattice is somewhat more complicated since it is determined by the coroots. These coroots live in $\mathcal{T}$ rather than $\mathcal{T}^*$. It is possible to draw them in $\mathcal{T}^*$, but we need to be aware that changing the scale of $\mathcal{T}^*$ will change the scale of $\mathcal{T}$ in a reciprocal manner, so magnifying the $\mathcal{T}^*$ diagram will shrink the corresponding coroots.

The easiest case is the case when all roots have the same length. If we normalize so this length is 1, then the coroots are just 2α . Recall that a weight $w \in \mathcal{T}^*$ satisfies the formula

$$w \left(\frac{2\alpha}{\langle \alpha, \alpha \rangle} \right) \in \mathbb{Z}$$

for all roots α . When all roots are normalized to have length one, this reduces to

$$w(2\alpha) \in Z$$

This formula is satisfied by all $w \in \mathcal{T}^*$ perpendicular to α , a hyperplane through the origin. It is also satisfied by parallel hyperplanes which pass through $\frac{\alpha}{2}, \alpha, \dots$. We obtain families of parallel hyperplanes], one for each root, and the weights lie on the points which belong to a hyperplane from each of these families.

A lattice basis for the weight lattice is given by weights w_i satisfying

$$w_i \left(\frac{2\alpha_j}{\langle \alpha_j, \alpha_j \rangle} \right) = \delta_{ij}$$

An arbitrary weight then has a coordinate expression $(k_1, k_2, \dots, k_m)$ in terms of this basis. For such an arbitrary weight,

$$w \left(\frac{2\alpha_i}{\langle \alpha_i, \alpha_i \rangle} \right)$$

is just k_i and can be computed on the diagram by finding the k_i th hyperplane in the family generated by α_i as we work our way out from the origin to w .

Finally we need to do this also for α which are not simple. This involves writing

$$\frac{2\alpha}{\langle \alpha, \alpha \rangle} = p_1 \frac{2\alpha_1}{\langle \alpha_1, \alpha_1 \rangle} + \dots + \frac{2\alpha_m}{\langle \alpha_m, \alpha_m \rangle}$$

and then using the previous calculation. In the case where all roots have the same length, this formula reduces to

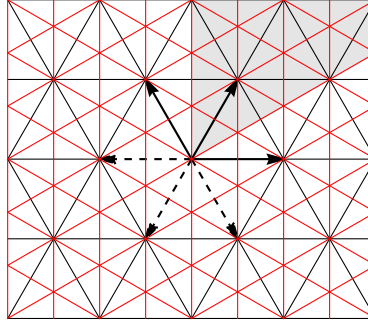
$$\alpha = p_1\alpha_1 + \dots + p_m\alpha_m$$

and thus can be read off directly from the root diagram without referring at all to the coroots. This suffices to handle the case $SU(3)$, which we do next.

Theorem 120 *The dimension of the irreducible representation of $SU(3)$ with highest weight (k, l) is*

$$\frac{(k+1)(l+1)(k+l+2)}{2}$$

Proof: The Weight diagram for this group is shown below.

Figure 21.1: Weight Diagram for A_2

The black arrows are the three roots, but since all roots have equal length, they are also the coroots up to length. Between the simple roots α and β is their sum, which is clearly the third root γ . Thus $\delta = \frac{\alpha+\beta+\gamma}{2} = \frac{2\gamma}{2} = \gamma$. Notice that the coordinates of δ are thus $(1, 1)$.

So $\lambda + \delta$ has coordinates $(k + 1, l + 1)$. When this element is evaluated on the coroot α , we get $k + 1$. When it is evaluated on the coroot β , we get $l + 1$. When it is evaluated on the coroot γ , which is the sum of the coroots α and β , we get $k + l + 2$.

Similar δ has coordinates $(1, 1)$ and when it is evaluated on the coroot α we get 1, when evaluated on the coroot β we get 1, and when evaluated on the coroot γ we get 2.

The final product is therefore $\frac{(k+1)(l+1)(k+l+2)}{1 \cdot 1 \cdot 2}$ as desired. QED.

Remark: We continue with the same calculation for the remaining rank 2 groups $B_2 = SO(5)$ and G_2 . Below is the Weight Diagram for B_2 . Recall that we are computing the dimensions of the irreducible representations of the Spin group which covers $SO(5)$, and some of these representations do not drop down to representations of $SO(5)$.

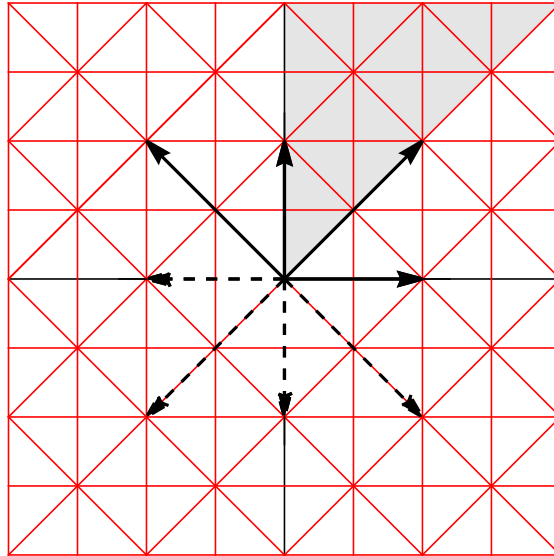
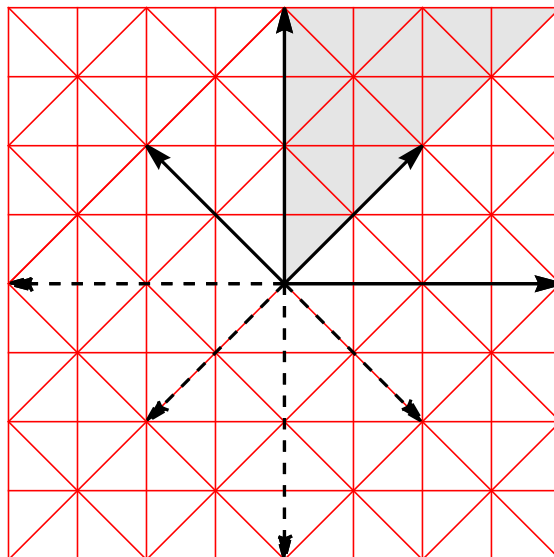
The diagram shows the roots as black arrows. We normalize so the short roots have length one. The corresponding coroots $\frac{2\alpha}{\langle \alpha, \alpha \rangle}$ then have length two. The long roots clearly have length $\sqrt{2}$ and the corresponding coroots thus are the same as the roots.

In the diagram below the weight diagram, the roots have been replaced by the corresponding coroots. But recall that the coroots really live in $\mathcal{T}$ rather than $\mathcal{T}^*$, so if we multiply the inner product by a factor, the weight diagram will increase while the coroots will shrink.

Finally, we have

$$\alpha \left(\frac{2\alpha}{\langle \alpha, \alpha \rangle} \right) = 2$$

So there should be red hyperplanes halfway along each root vectors, exactly as the weight diagram shows.

Figure 21.2: Weight Diagram for B_2 Figure 21.3: CoRoot Diagram for B_2

Let us use the coroot diagram to express each coroot in terms of the simple coroots. Denote coroots by adding a hat over the root name. The root pointing to the right is α , so the corresponding coroot is $\hat{\alpha}$. The root pointing to the upperleft is β , so the corresponding coroot is $\hat{\beta}$. Clearly the positive coroots are then $\hat{\alpha}, \hat{\alpha} + \hat{\beta}, \hat{\alpha} + 2\hat{\beta}, \hat{\beta}$.

Next concentrate on the roots bounding the fundamental Weyl chamber. Notice that the longer root pointing to the upper right contains a weight vector half-way along the root. By rights this should be $(1, 0)$. But the shorter root pointing straight up only has a weight vector at its end. By rights this should be $(0, 1)$. Let us verify that this is the correct reading of coordinates for points in the closed fundamental Weyl Chamber.

A point with second coordinate zero should be on the hyperplane perpendicular to β , and thus along the long boundary root. This root is $\beta + 2\alpha$ and we must find a multiple of this root whose value on $\hat{\alpha}$ is 1. This multiple is $\frac{1}{2}$ because

$$\begin{aligned} \left(\frac{\beta + 2\alpha}{2}\right)(\hat{\alpha}) &= \beta \left(\frac{\alpha}{\langle \alpha, \alpha \rangle} \right) + 2 = 2 + \frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} = \\ &2 + \frac{\|\beta\|}{\|\alpha\|} \left(\frac{-1}{\sqrt{2}} \right) = 2 - 1 = 1 \end{aligned}$$

Similarly a point with first coordinate zero should be on the hyperplane perpendicular to α and thus in the direction of the short boundary root. This root is $\alpha + \beta$. We want a multiple of this root which evaluates to 1 on $\hat{\beta}$. That multiple is one because

$$(\alpha + \beta)(\hat{\beta}) = (\alpha + \beta) \left(\frac{2\beta}{\langle \beta, \beta \rangle} \right) = \frac{2\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} + 2 = 2 \frac{\|\alpha\|}{\|\beta\|} \left(-\frac{1}{\sqrt{2}} \right) + 2 = -1 + 2 = 1$$

Finally, notice by direct inspection of the weight diagram that the half sum of the positive roots is the point in the fundamental Weyl chamber with coordinates $(1, 1)$.

Putting this altogether, we must evaluate $(k + 1, l + 1)$ on the positive coroots and take the product, and evaluate $(1, 1)$ on the positive coroots and take the product, and divide these products.

Theorem 121 *The dimension of the irreducible representation of B_2 with highest weight (k, l) is*

$$\frac{(k + 1)(l + 1)(k + l + 2)(k + 2l + 3)}{3!}$$

Proof: We know that $(k + 1, l + 1)$ on $\hat{\alpha}$ is $k + 1$ and on $\hat{\beta}$ is $l + 1$. So its values on the positive coroots $\hat{\alpha}, \hat{\alpha} + \hat{\beta}, \hat{\alpha} + 2\hat{\beta}, \hat{\beta}$ is $(k + 1), (k + 1 + l + 1), (k + 1 + 2(l + 1)), (l + 1)$ and the product of these terms is $(k + 1)(l + 1)(k + l + 2)(k + 2l + 3)$. When $k = l = 0$, the corresponding product is $1 \cdot 1 \cdot 2 \cdot 3 = 3!$.

Remark: Here is a table of these dimensions for small values of k and l :

k	l	dim
0	0	1
1	0	4
0	1	5
1	1	16
2	0	10
0	2	14
2	1	35
1	2	40
2	2	81

Remark: Clearly the 5-dimensional representation of $SO(5)$ is given by the group acting on R^5 and thus C^5 in the standard way.

The group $SO(5)$ has trivial center; its universal cover is $Spin(5)$, a 2-fold cover. These are the only connected Lie groups with Lie algebra $so(5)$. Consequently all of these irreducible representations are defined on $Spin(5)$, but some are not defined on $SO(5)$. In particular, the 4-dimensional representation is only defined on $Spin(5)$.

Let us say more about that four dimensional representation. The Dynkin diagrams for B_k and C_k are almost the same; the double arrow points in opposite directions in these diagrams. In particular, the Dynkin diagrams for B_2 and C_2 consist of a single double arrow and thus are isomorphic. So the algebras B_2 and C_2 are equal. We know that B_2 is the algebra $so(5)$ and C_2 is the algebra $sp(2)$.

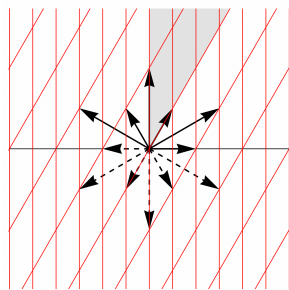
Earlier we proved that $sp(n)$ contains all complex matrices of the following form, where A and B are $n \times n$ matrices satisfying $A = -\overline{A}^T$ and $B = B^T$:

$$\begin{pmatrix} A & B \\ -\overline{B} & A \end{pmatrix}$$

Note that each of these is a 4×4 complex matrix, and thus we have a 4-dimensional complex representation of the Lie algebra $sp(2) \cong so(5)$. This is induced by a corresponding 4-dimensional representation of the group $Sp(2)$. This representation of $Sp(2)$ is just $Sp(2)$ itself acting as a set of matrices, and thus one-to-one. But $Sp(2)$ is simply-connected and thus equal to $Spin(5)$. Because the representation is one-to-one, it could not drop down to be a representation of $SO(5)$.

21.6 Dimensions of Irreducible Representations of G_2

The weight diagram for G_2 is shown below:

Figure 21.4: Weights and Roots for G_2

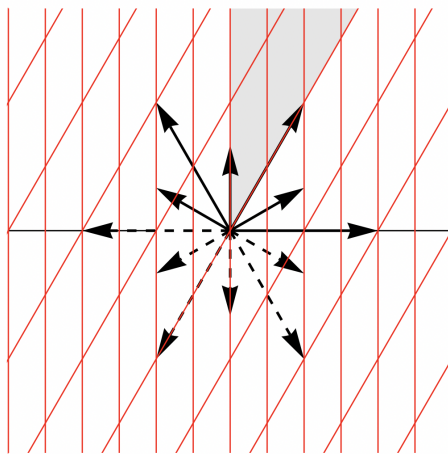
This time we have only shown the hyperplanes associated to the simple roots, which may lead to a diagram which is easier to understand.

Using the standard choices for α and β , the positive roots are

$$\alpha, \beta + 3\alpha, \beta + 2\alpha, 2\beta + 3\alpha, \beta + \alpha, \beta$$

The half sum of the positive roots is the point $(1, 1)$ in the fundamental Weyl chamber.

If the short roots have length 1, then the long root β defines a right triangle with base $\frac{3}{2}$ and height $\frac{\sqrt{3}}{2}$ and thus have length $\sqrt{3}$. The corresponding coroots are obtained by doubling the short roots and multiplying the long roots by $\frac{2}{3}$. This produces the diagram below:

Figure 21.5: Coroots for G_2

From this diagram, we can read off the coroots in terms of the coroot basis:

$$\hat{\alpha}, \hat{\alpha} + \hat{\beta}, 2\hat{\alpha} + 3\hat{\beta}, \hat{\alpha} + 2\hat{\beta}, \hat{\alpha} + 3\hat{\beta}, \hat{\beta}$$

Theorem 122 *The dimension of the irreducible representation of g_2 with highest weight (k, l) is*

$$\frac{(k+1)(l+1)(k+l+2)(k+2l+3)(k+3l+4)(2k+3l+5)}{5!}$$

Proof: Evaluating $(k+1, l+1)$ on the coroots gives

$$(k+1)(k+l+2)(2k+3l+5)(k+2l+3)(k+3l+4)(l+1)$$

Evaluating $(1, 1)$ on these same coroots gives $5!$ QED.

Remark: Here is a table of these dimensions for small values of k and l :

k	l	dim
0	0	1
1	0	7
0	1	14
1	1	64
2	0	27
0	2	77
2	1	189
1	2	286
2	2	729

Remark: Note that G_2 has dimension 14, so the 14-dimensional representation is the adjoint representation of the group on itself. This table shows that G_2 can be defined by a group of matrices of size 7. Indeed it turns out that G_2 is the automorphism group of the octonions, a non-associative normed algebra of dimension 8. These octonions were discovered within a year of Hamilton's discovery of the quaternions, by a friend of Hamilton's. The friend told Hamilton that he'd next try to find a 16 dimensional norm preserving product, but ultimately failed because no such animal exists. It is Hamilton who discovered that the octonions are not associative.

21.7 The Weyl Character Formula and $SU(3)$

Finally, we will compute the characters of some irreducible representations of $SU(3)$ using the character formula, and thus complete our discussion of these representations which we began earlier. We continually refer to the weight diagram for $SU(3)$ shown at the top of the next page.

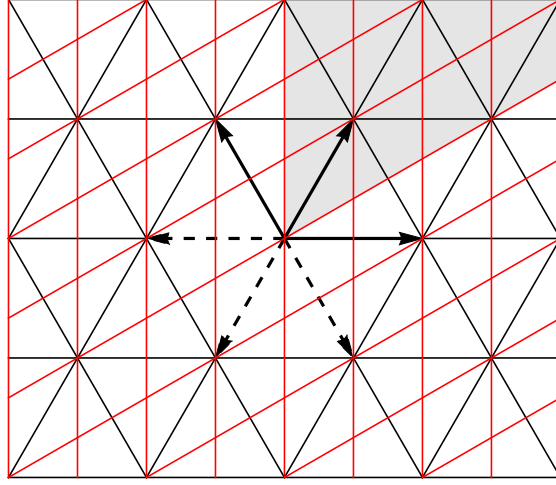


Figure 21.6: Weyl Chamber

Recall that the weight lattice points in the fundamental Weyl chamber have coordinates (k, l) where k and l are non-negative integers. Another way to say this is that the weight lattice has a basis given by $(1, 0)$ and $(0, 1)$. If these vectors are called w_1 and w_2 , define

$$\chi_1 = e^{2\pi i w_1(t)} \quad \text{and} \quad \chi_2 = e^{2\pi i w_2(t)}$$

An arbitrary weight has the form $w = (k, l)$ with $k, l \in \mathbb{Z}$ and the character attached to this weight is

$$\chi = e^{2\pi i w(t)} = \left(e^{2\pi i w_1(t)} \right)^k \left(e^{2\pi i w_2(t)} \right)^l = \chi_1^k \chi_2^l$$

It follows that every character of $SU(3)$ is a finite sum of the following form, where the a_{kl} are integers:

$$\sum_{k,l} a_{kl} \chi_1^k \chi_2^l$$

Since the k and l can be both positive and negative, we call such an expression a *Laurent polynomial*.

The Weyl group acts on these characters and we need to make this action explicit. It suffices to explain the action on χ_1 and χ_2 . Each of these basis vectors is moved to one of the six lattice points closest to the origin. Starting from $(1, 0)$ and moving counterclockwise, we can list these elements as

$$(1, 0), (0, 1), (-1, 1), (-1, 0), (0, -1), (1, -1) = \chi_1, \chi_2, \chi_1^{-1}\chi_2, \chi_1^{-1}, \chi_2^{-1}, \chi_1\chi_2^{-1}$$

Sums of the form

$$\sum_w \det(w) e^{2\pi i \lambda(w(t))}$$

played a special role in the proof of the character formula. We expect the following similar sums to appear in group characters:

$$\sum_w e^{2\pi i \lambda(w(t))}$$

This sum will have six entries unless λ is on the boundary of a Weyl chamber; in that special case the sum will only have three entries, each repeated twice. Instead of accepting this redundancy, we replace the sum by half its value in these cases.

Let us work out the first few of these explicitly. Note that the Weyl group has six elements: three are rotations S and three are reflections. The reflections can all be written SR where R is the reflection which leaves w_1 fixed. If we apply these elements to $w_1 = (1, 0)$, we only get three points: $(1, 0), (-1, 1), (0, -1)$. So the expression is

$$\chi_1 + \chi_1^{-1} \chi_2 + \chi_2^{-1}$$

Similarly $w_2 = (0, 1)$ maps to $(0, 1), (-1, 0), (1, -1)$ and thus to

$$\chi_2 + \chi_1^{-1} + \chi_1 \chi_2^{-1}$$

We now claim that these are the characters of the simplest non-trivial representations of $SU(3)$ as given by the Weyl character formula. To see this, we need only compute the numerator and denominator of the formula and divide. The numerators are the alternating characters attached to $(2, 1)$ and $(1, 2)$ and the denominator is the alternating character attached to $(1, 1)$. Let us first deal with the denominator. Its sum is given by three terms with positive coefficients given by rotation, and three terms with negative coefficients given by reflection. We easily read this off. Symbolically, we have

$$(1, 1) + (-2, 1) + (1, -2) - (-1, 2) - (-1, -1) - (2, -1) =$$

$$\chi_1 \chi_2 + \chi_1^{-2} \chi_2 + \chi_1 \chi_2^{-2} - \chi_1^{-1} \chi_2^2 - \chi_1^{-1} \chi_2^{-1} - \chi_1^2 \chi_2^{-1}$$

Ugh. Pretty messy. Maybe dividing by this will not be fun, so let's go backward. Let's multiply by $\chi_1 + \chi_1^{-1} \chi_2 + \chi_2^{-1}$ and see if we get the alternating element associated with $(2, 1)$.

We also abandon the χ characters, so instead of writing $\chi_1^2\chi_2$ we write $(2, 1)$. Multiplying the denominator by χ_1 , $\chi_1^{-1}\chi_2$, and χ_2^{-1} then gives the three lines below:

$$\begin{aligned} &(2, 1) + (-1, 1) + (2, -2) - (0, 2) - (0, -1) - (3, -1) \\ &(0, 2) + (-3, 2) + (0, -1) - (-2, 3) - (-2, 0) - (1, 0) \\ &(1, 0) + (-2, 0) + (1, -3) - (-1, 1) - (-1, -2) - (2, -2) \end{aligned}$$

Adding these terms produces a lot of cancellation and leaves six remaining terms:

$$(2, 1) - (3, -1) + (-3, 2) - (-2, 3) + (1, -3) - (-1, -2)$$

We hope that this is the alternating sum associated to $(2, 1)$. While we could determine this by looking at the weight picture, it is useful to have a more general algebraic formula. Let us produce matrices for the action of R and S . Note that R rotates $(1, 0)$ to $(-1, 1)$ and $(0, 1)$ to $(-1, 0)$. So it is given by the matrix

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

Note that S reflects $(1, 0)$ to $(1, 0)$ and $(0, 1)$ to $(1, -1)$. It is given by the matrix

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

The full Weyl group has three rotations I, R, R^2 and three reflections S, RS, R^2S ; their matrices are

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

Applying the first three matrices to $(2, 1)$ gives $(2, 1), (-3, 2), (1, -3)$, the three terms in our expression with coefficient 1. Applying the last three matrices to $(2, 1)$ gives $(3, -1), (-2, 3), (-1, -2)$ the three terms in our expression with coefficient -1 . So we have checked that the Weyl Character Formula gives the representation with highest weight $(1, 0)$ the character

$$\chi_1 + \chi_1^{-1}\chi_2 + \chi_2^{-1}$$

Whew!

Let us turn next to the representation with highest weight $(1, 1)$. We discussed this representation in chapter 18 and found its weights, which are the six images of $(1, 1)$ under the Weyl group together with the identity. But we could not determine the dimension of the identity weight space. Let us use the character formula to do that now. According to the character formula, the character is the alternating sum attached to $(2, 2)$ divided by

the alternating sum attached to $(1, 1)$, and we expect the quotient to be the regular sum attached to $(1, 1)$ plus a multiple of $(0, 0)$. We will do this calculation backward. So if we multiply the regular sum attached to $(1, 1)$ by the alternating sum attached to $(1, 1)$, we expect to get the alternating sum attached to $(2, 2)$ minus the alternating sum attached to $(1, 1)$ times a positive integer d . We believe that $d = 2$. Here we go.

By previous calculations, the denominator is

$$(1, 1) + (-2, 1) + (1, -2) - (-1, 2) - (-1, -1) - (2, -1)$$

The six images of $(1, 1)$ are obtained by multiplying $(1, 1)$ by the six Weyl matrices:

$$(1, 1), (-2, 1), (1, -2), (2, -1), (-1, 2), (-1, -1)$$

We multiply the denominator by each in turn and add, getting the six sums below:

$$\begin{aligned} & (2, 2) + (-1, 2) + (2, -1) - (0, 3) - (0, 0) - (3, 0) \\ & (-1, 2) + (-4, 2) + (-1, -1) - (-3, 3) - (-3, 0) - (0, 0) \\ & (2, -1) + (-1, -1) + (2, -4) - (0, 0) - (0, -3) - (3, -3) \\ & (0, 3) + (-3, 3) + (0, 0) - (-2, 4) - (-2, 1) - (1, 1) \\ & (0, 0) + (-3, 0) + (0, -3) - (-2, 1) - (-2, -2) - (1, -2) \\ & (3, 0) + (0, 0) + (3, -3) - (1, 1) - (1, -2) - (4, -2) \end{aligned}$$

Adding these elements, there is a lot of cancellation and we obtain

$$\begin{aligned} & (2, 2) + (-4, 2) + (2, -4) - (-2, -2) - (4, -2) - (-2, 4) \\ & -2\left((1, 1) + (-2, 1) + (1, -2) - (-1, 2) - (-1, -1) - (2, -1)\right) \end{aligned}$$

A quick check shows that the first element is the alternating sum attached to $(2, 2)$ and the second is the alternating sum attached to $(1, 1)$ with the expected coefficient of 2.

Remark: Let us start to work out one more example by hand, the representation with the following diagram:

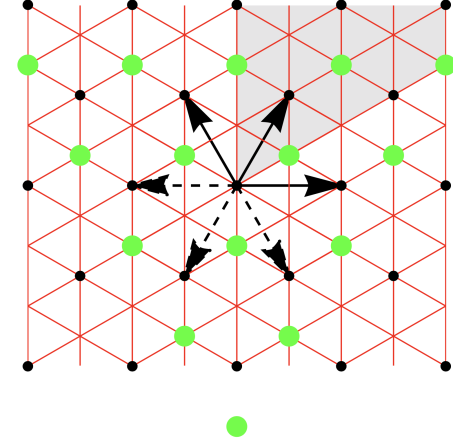


Figure 21.7: Triangle $(4, 0)$

This time we can guess the answer, assuming the assertion from an earlier chapter that in this triangular case each “shell” has multiplicity one. The highest root is $(4, 0)$; since it lies on a boundary of the fundamental Weyl chamber, its orbit under the Weyl group has three terms:

$$(4, 0) + (-4, 4) + (0, -4)$$

Another element in the outer shell is $(0, 2)$; its orbit under the Weyl group also has three terms:

$$(0, 2) + (-2, 0) + (2, -2)$$

Finally, the point $(2, 1)$ in the fundamental chamber has six elements in its orbit:

$$(2, 1) + (-2, 3) + (-3, 2) + (-1, -2) + (1, -3) + (3, -1)$$

The inner shell has $(0, 1)$ in the fundamental chamber and three elements in its orbit:

$$(1, 0) + (-1, 1) + (0, -1)$$

The sum of all of these terms should be given by dividing a numerator generated by $(4, 0) + (1, 1) = (5, 1)$, by a denominator generated by $(1, 1)$. The numerator should contain six terms, three with plus signs obtained by multiplying $(5, 1)$ by the matrices for I, R, R^2 , minus the three terms obtained by applying the matrices for S, RS, R^2S :

$$(5, 1) + (-6, 5) + (1, -6) - (6, -1) - (-5, 6) - (-1, -5)$$

and the denominator, generated by $(1, 1)$, should be

$$(1, 1) + (-2, 1) + (1, -2) - (2, -1) - (-1, 2) - (-1, -1)$$

From here, the check is easy but tedious. We must multiply each of the four listed regular sums by the alternating sum of the denominator and add. The result should be the alternating sum of the numerator. Rather than doing this by hand, we will write a computer program in the next section to perform this check. In a later section we will make the program more general so it can compute characters for all three rank 2 groups: $SU(3)$, $SO(5)$, G_2 .

21.8 A Preliminary Computer Program

We need a data structure which will represent a sum of the form $\sum a_{kl} \chi_1^k \chi_2^l$. This data can be described by an array $A(k, l)$ where $A(k, l)$ equals the element a_{kl} . We will limit the size of these character sums to

$$-100 \leq k, l \leq 100$$

We write the code in Mathematica, although it can easily be modified for other computer languages. Mathematica cannot deal with negative array indices, so our $A1(k, l)$ is defined for $0 < k, l < 202$ and we set $A(k, l) = A1(k + 101, l + 101)$.

The numerator and denominator of Weyl's formula need not be stored in such a data structure because both expressions have just six terms, which can be computed from an initial term by hard coding the six elements of the Weyl group in the program. For instance, in the last example from the previous section we have an alternating sum for $(4, 0) + (1, 1) = (5, 1)$ in the numerator, and an alternating sum for $(1, 1)$ in the denominator.

Similarly, there is no need to store the actual character which Weyl's formula gives us because we will construct this character piece by piece. Each piece will be invariant under the Weyl group, so we will actually find the weights and dimensions of these weight spaces for weights in the fundamental Weyl chamber,

Let's proceed with the implementation of the code.

```

(* Data structure for character-like sums *)
ClearAll[A1];
f[x_, y_] := 0;
A1 = Array[f, {202, 202}];

(* Weyl Group; thus w[[2]] is the second element of the group.
   The first three elements have determinant 1 and the remaining
   elements have determinant -1. *)
w1 = {{1, 0}, {0, 1}};
w2 = {{-1, -1}, {1, 0}};
w3 = {{0, 1}, {-1, -1}};
w4 = {{1, 1}, {0, -1}};
w5 = {{-1, 0}, {1, 1}};
w6 = {{0, -1}, {-1, 0}};
w = {w1, w2, w3, w4, w5, w6};
WeylOrder = 6;

(* This routine clears the A1 data structure. *)
ClearArray[ ] := Block[{i, j},
  For[i = -100, i <= 100, i++,
    For[j = -100, j <= 100, j++,
      A1[[i + 101, j + 101]] = 0]]];

(* This routine fills the A1 structure with the numerator of Weyl's formula for a
   highest weight (k, l). *)
WriteNumerator[k_, l_] := Block[{i, j, res, a, b, s1},
  For[i = 1, i <= WeylOrder, i++,
    res = w[[i]].{1 + k, 1 + l};
    a = res[[1]];
    b = res[[2]];
    If[(i <= (WeylOrder / 2)), s1 = 1, s1 = -1];
    A1[[a + 101, b + 101]] = s1];

```

```
(* This routine prints out the data in A1; each line of data contains
   the dimension of an element, followed by the element, as in
   1 - (3, 5). *)
ShowData[] := Block[ { i, j, dim },
  For[i = -100, i <= 100, i++,
    For[j = -100, j <= 100, j++,
      dim = A1[[i + 101, j + 101]];
      If[ (dim != 0),
        Print[ dim, " - (", i, ", ", j, ")"]]]];

(* The data in the data structure is always invariant under the Weyl group
   up to sign. Thus it is usually more illuminating to see the data just for
   weights in the fundamental Weyl chamber. *)
ShowFundamentalData[] := Block[ { i, j, dim },
  For[i = -100, i <= 100, i++,
    For[j = -100, j <= 100, j++,
      dim = A1[[i + 101, j + 101]];
      If[ ((dim != 0) && (i >= 0) && (j >= 0)),
        Print[ dim, " - (", i, ", ", j, ")"]]]];

(* Finally below are the key routines. This final routine starts with a weight
   (k, l) in the fundamental chamber, and extends it to a character-like element that
   is invariant under the Weyl group. This element has six elements for interior
   weights and three elements for weights on the boundary of the chamber.
   Then the routine multiplies this expression by the denominator of the Weyl
   formula and subtracts this from the A1 data. Each such term represents part
   of the quotient, and thus part of the representation character. The goal is to
   run this routine for each character weight in the fundamental chamber, thus giving
   the complete character. The calculation succeeded if the remaining A1 structure
   is empty. *)
```



```

RemoveHighWeightZero[ dimension_] := Block[{i, j, res, a, b, s1},
  For[i = 1, i <= WeylOrder, i++,
    res = w[[i]].{1, 1};
    a = res[[1]];
    b = res[[2]];
    If[(i <= (WeylOrder / 2)), s1 = 1, s1 = -1];
    A1[[a + 101, b + 101]] =
      A1[[a + 101, b + 101]] - s1 * dimension];

RemoveHighWeightNonZero[k_, l_, dimension_] := Block[{i, j, res, a, b, s1, s2, s3},
  For[i = 1, i <= WeylOrder, i++,
    For[j = 1, j <= WeylOrder, j++,
      res = (w[[i]].{1, 1}) + (w[[j]].{k, l});
      a = res[[1]];
      b = res[[2]];
      If[(i <= (WeylOrder / 2)), s1 = 1, s1 = -1];
      If[(j <= (WeylOrder / 2)), s2 = 1,
        If[(k == 0) || (l == 0),
          s2 = 0, s2 = 1]];
      s3 = s1 * s2;
      A1[[a + 101, b + 101]] = A1[[a + 101, b + 101]] - dimension * s3];
];

RemoveHighWeight[k_, l_, dimension_] :=
  If[ (k == 0) && (l == 0),
    RemoveHighWeightZero[dimension],
    RemoveHighWeightNonZero[k, l, dimension];];

```

Let us use this program to complete the calculation for the highest weight $(4, 0)$.

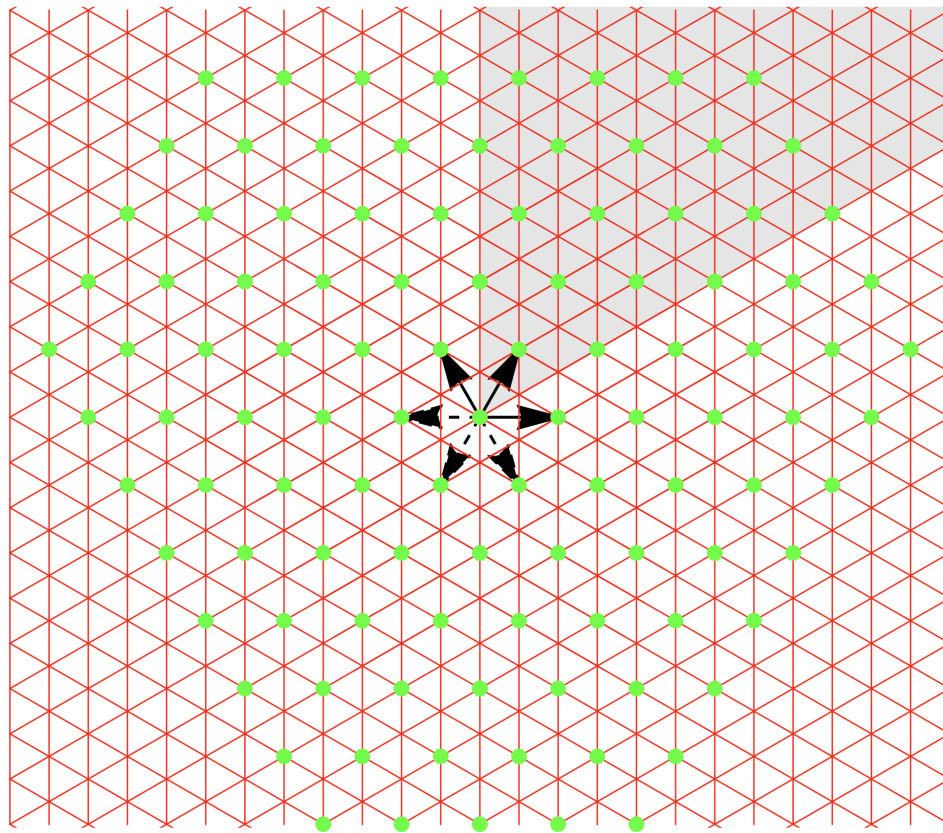
```
ClearArray[];
WriteNumerator[4, 0];
RemoveHighWeight[4, 0, 1];
RemoveHighWeight[2, 1, 1];
RemoveHighWeight[0, 2, 1];
RemoveHighWeight[1, 0, 1];
ShowData[];
```

The output of the final routine is empty, so if we divide the numerator of the Weyl formula by its denominator, the quotient is the Weyl-group invariant sum with weights in the fundamental chamber at $(4, 0)$, $(2, 1)$, $(0, 2)$, $(1, 0)$, all of dimension one. The first three give the outer ring and the final element gives the inner ring in our picture.

Our method is a little disappointing because we use it to check that a conjectured character satisfies Weyl's formula, but not apparently to deduce the character from the formula. This will be improved in the next section where we extend our program to compute any character of $SU(3)$ without first making guesses.

We end this section with a final example, the representation of $SU(3)$ with highest weight $(7, 4)$. We will learn some crucial facts from this example.

The next page shows the weight diagram for this representation. When I originally looked at this picture, I noticed that the outer three rings of the weight sets are lopsided hexagons. But I misinterpreted the fourth ring as a perfect hexagon and the remaining rings as smaller and smaller perfect hexagons until finally we reach a central point at the origin. The ends of the root vectors reinforced this incorrect interpretation, because they seemed to describe a perfect hexagon. Actually these points belong to a perfect triangle whose four vertices are omitted, so the sides have four dots rather than two. Let me continue with this misstep and show how the Weyl Character formula ultimately told me that something was wrong.

Figure 21.8: Representation with High Weight $(7, 4)$

If the various rings are never triangles, we claimed without proof earlier that the dimensions of weights in a ring increase by one for each subring. Let us assume this. We can read off the weights which lie in the fundamental chamber for each ring. This part of my calculation was valid even when I misunderstood the rings, and the results were

- Dimension 1: $(9, 0)$, $(8, 2)$, $(7, 4)$, $(5, 5)$, $(3, 6)$, $(1, 7)$
- Dimension 2: $(7, 1)$, $(6, 3)$, $(4, 4)$, $(2, 5)$, $(0, 6)$
- Dimension 3: $(3, 3)$, $(5, 2)$, $(6, 0)$, $(1, 4)$
- Dimension 4: $(4, 1)$, $(2, 2)$, $(0, 3)$
- Dimension 5: $(3, 0)$, $(1, 1)$
- Dimension 6: $(0, 0)$

Finally, we can use our program to check that these weights and dimensions give the

representation with highest weight $(7, 4)$. I'll do the calculation ring by ring. For each ring, I'll arrange the weights by the first sum of their coordinates from highest to lowest; in case of duplicate sums I'll use lexicographical order. This particular order turns out to be irrelevant in the end.

```

ClearArray[];
WriteNumerator[7, 4];

RemoveHighWeight[7, 4, 1];
RemoveHighWeight[8, 2, 1];
RemoveHighWeight[5, 5, 1];
RemoveHighWeight[9, 0, 1];
RemoveHighWeight[3, 6, 1];
RemoveHighWeight[1, 7, 1];

RemoveHighWeight[6, 3, 2];
RemoveHighWeight[7, 1, 2];
RemoveHighWeight[4, 4, 2];
RemoveHighWeight[2, 5, 2];
RemoveHighWeight[0, 6, 2];

RemoveHighWeight[5, 2, 3];
RemoveHighWeight[6, 0, 3];
RemoveHighWeight[3, 3, 3];
RemoveHighWeight[1, 4, 3];

RemoveHighWeight[4, 1, 4];
RemoveHighWeight[2, 2, 4];
RemoveHighWeight[0, 3, 4];

RemoveHighWeight[3, 0, 5];
RemoveHighWeight[1, 1, 5];

RemoveHighWeight[0, 0, 6];
ShowData[];

```

Unhappily, data remained at the end. So something was wrong.

To debug the problem, I added the line “ShowFundamentalData[]” after the code for each ring. Below is the illuminating result:

- After outer ring: dim -1: $(6, 3)$, dim 2: $(7, 4)$
- After 2nd ring: dim -2: $(5, 2)$, dim 3: $(6, 3)$
- After 3rd ring: dim -3: $(4, 1)$, dim 4: $(5, 2)$
- After 4th ring: dim 5: $(4, 1)$
- After 5th ring: dim 5: $(1, 1)$
- After final ring, the origin: dim 1: $(1, 1)$

After examining this data, my first conclusion was that something was wrong with the 4th ring. At that point, I printed out the weight diagram shown earlier in this writeup and hand drew the lines showing the various rings. This led to the discovery that the first four rings are lopsided hexagons, but the next two rings (counting the origin) are triangles.

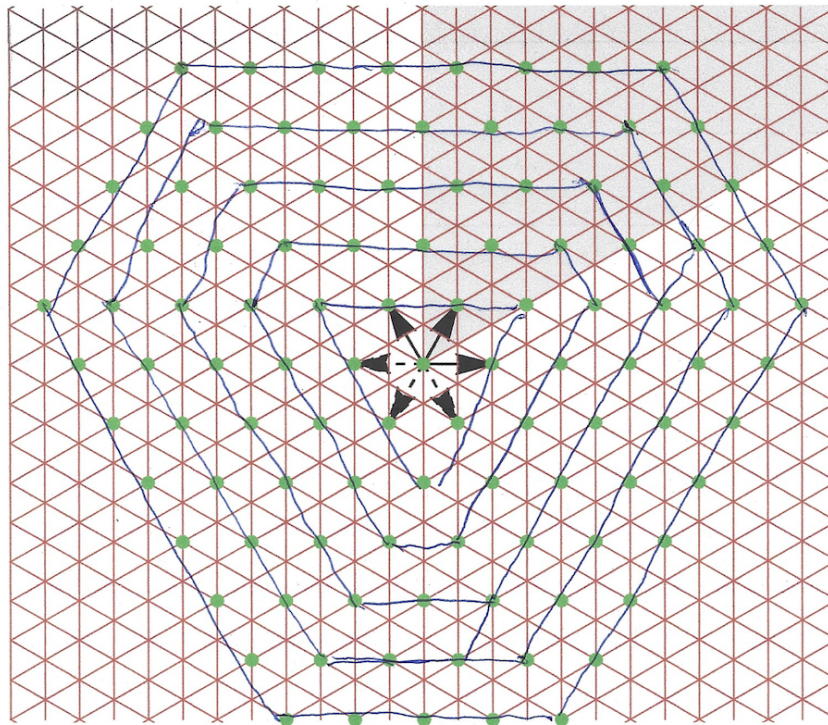


Figure 21.9: Representation with High Weight $(7, 4)$

To confirm this, I ran the following program:

```

ClearArray[];
WriteNumerator[7, 4];

RemoveHighWeight[7, 4, 1];
RemoveHighWeight[8, 2, 1];
RemoveHighWeight[5, 5, 1];
RemoveHighWeight[9, 0, 1];
RemoveHighWeight[3, 6, 1];
RemoveHighWeight[1, 7, 1];

RemoveHighWeight[6, 3, 2];
RemoveHighWeight[7, 1, 2];
RemoveHighWeight[4, 4, 2];
RemoveHighWeight[2, 5, 2];
RemoveHighWeight[0, 6, 2];

RemoveHighWeight[5, 2, 3];
RemoveHighWeight[6, 0, 3];
RemoveHighWeight[3, 3, 3];
RemoveHighWeight[1, 4, 3];

RemoveHighWeight[4, 1, 4];
RemoveHighWeight[2, 2, 4];
RemoveHighWeight[0, 3, 4];

RemoveHighWeight[3, 0, 5];
RemoveHighWeight[1, 1, 5];

RemoveHighWeight[0, 0, 5];
ShowData[];

```

No data was output, so the calculation is correct this time.

But we have learned more than that. Our experiments show that after the outer ring has been removed, the command “ShowFundamentalData” outputs exactly one weight with positive dimension. That dimension is the dimension of the next ring, and the actual weight is the highest weight in that ring plus $(1, 1)$. We can use this information to complete our program so it will compute the entire character associated with the highest weight. We’ll do that in the next section.

21.9 A Program To Compute Characters of Representations of $SU(3)$

We now describe two programs. One can compute the character of any representation of $SU(3)$ given its highest weight (k, l) , and the second can draw the weight diagram for this representation. The name of the first program is “CalculateCharacter[k, l]”. The second program is a block of Mathematica graphics code rather than a single procedure.

To explain how this works, let us try some simple examples, starting with the high weight $(10, 7)$. We execute

```
CalculateCharacter[6, 2];
```

and obtain the following output:

```
6 2 1
{{7,0},{0,5},{2,4},{4,3},{6,2}}
5 1 2
{{1,3},{3,2},{5,1}}
4 0 3
{{0,2},{2,1},{4,0}}
1 0 3
{{1,0}}
```

The top output line says that the highest weight in the outer ring is $(6, 2)$, and all weights in this ring have dimension 1. The weights from the Fundamental Weyl Chamber in this ring are listed on the second line.

A weight in the second ring is $(5, 1)$, and all weights in this ring have dimension 2. And so forth. This data is enough to list the entire character of the representation.

We now run an intermediate command, which can be one of the following-:

```
PrintAllWeights[];
PrintAllWeightsAndDimensions[];
```

The portion of the output of the first command is shown below. This time the command lists *all* weights, not just weights in the fundamental chamber. The weights are listed as graphic commands to draw a dot, and these graphic commands should be copied and pasted into the graphic block of code. Copy the list below the spot in the graphic code labeled

```
(* Dot Commands *)
```

To make the graphic code work correctly, the copy operation must be done as follows. Each Gdot command is a single line, so copying line by line would be tedious. But Mathematica

lists a second level of grouping on the right side of the page. Select that level and all commands will be copied at once. To copy, go to an appropriate menu in Mathematic, or select a contextual menu, and choose "Copy As Plain Text". No special instructions are needed to paste this into the graphic code.

```
Gdot[-8, 6], Gdot[-7, 4], Gdot[-7, 7], Gdot[-6, 2], Gdot[-6, 5], \
Gdot[-6, 8], Gdot[-5, 0], Gdot[-5, 3], Gdot[-5, 6], Gdot[-4, -2], \
Gdot[-4, 1], Gdot[-4, 4], Gdot[-4, 7], Gdot[-3, -4], Gdot[-3, -1], \
Gdot[-3, 2], Gdot[-3, 5], Gdot[-2, -6], Gdot[-2, -3], Gdot[-2, 0], \
Gdot[-2, 3], Gdot[-2, 6], Gdot[-1, -5], Gdot[-1, -2], ...
```

If we perform this copy/paste and then execute the graphic block, we obtain the diagram below.

If we execute `PrintAllWeightsAndDimensions[]` instead of `PrintAllWeights[]`, and copy and paste as above, the diagram will also show the dimension of each weight space beside the dot for the weight. An example is given in the second figure on the next page.

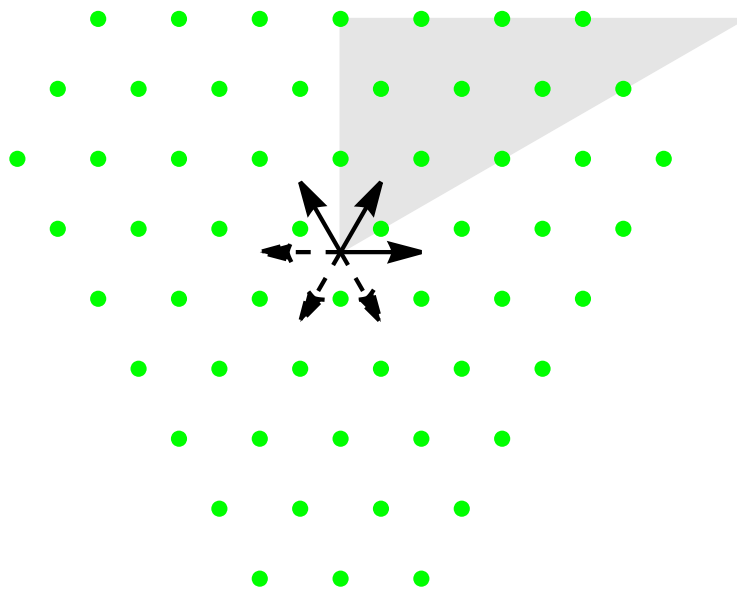
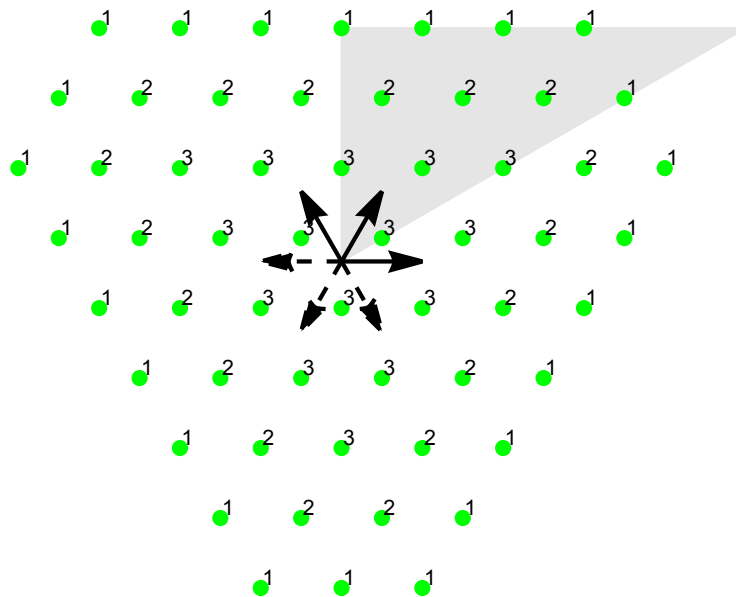


Figure 21.10: Weights for $(7, 2)$

If we execute `PrintAllWeightsAndDimensions[]` instead of `PrintAllWeights[]`, and copy and paste as above, the diagram will also show the dimension of each weight space beside the dot for the weight. The new version of the previous diagram is shown next.

Figure 21.11: Weights for $(7, 2)$

Notice that the first two rings are lopsided hexagons, while the third and remaining rings are triangles. As predicted earlier for $SU(3)$, the dimensions of the rings go up by one until reaching the triangle, and then then on all rings have the same dimensions.

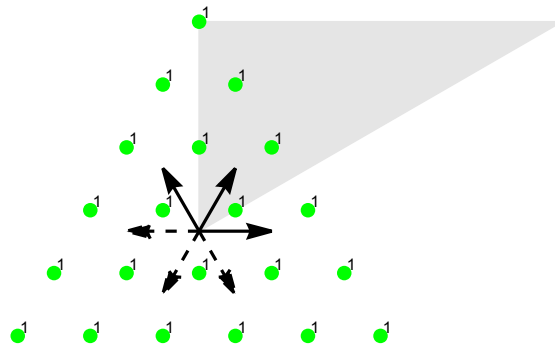
Let us try a few more examples. In each case, I list the command to construct a character and its output, and then show the weight graph.

High Weight $(0, 5)$:

```
CalculateCharacter[0, 5];
```

Output:

```
0 5 1
{{2,1},{1,3},{0,5}}
0 2 1
{{1,0},{0,2}}
```

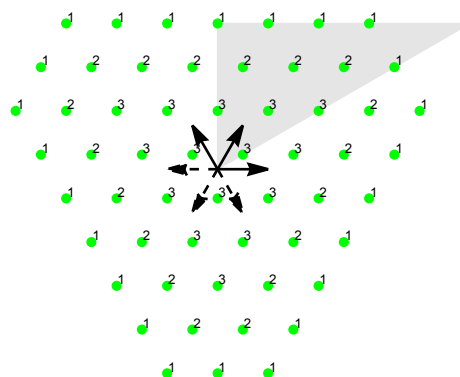
Figure 21.12: Weights for $(0, 5)$

High Weight $(6, 2)$:

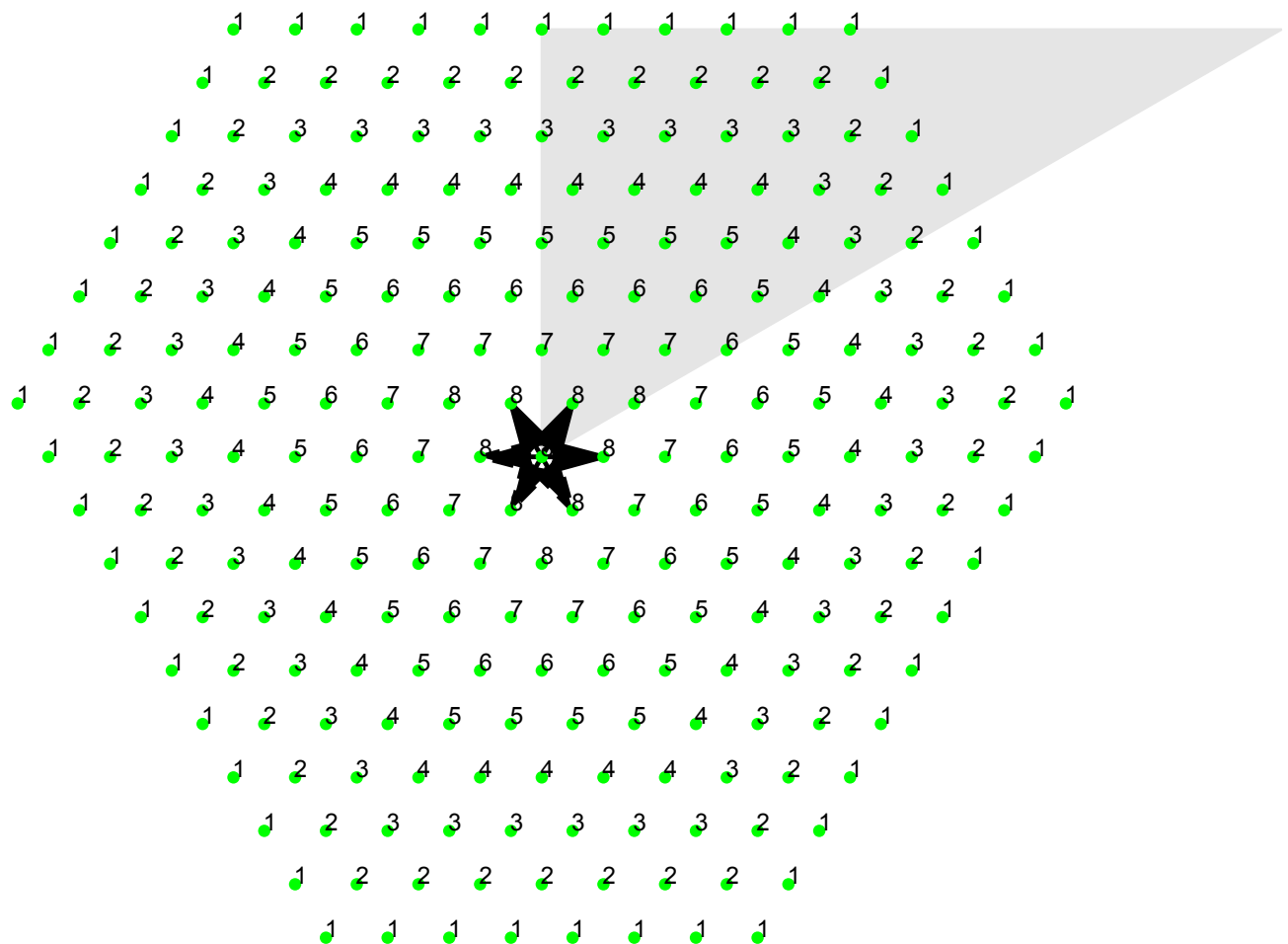
```
CalculateCharacter[6, 2];
```

Output:

```
6 2 1
adia {{7,0},{0,5},{2,4},{4,3},{6,2}}
5 1 2
{{1,3},{3,2},{5,1}}
4 0 3
{{0,2},{2,1},{4,0}}
1 0 3
{{1,0}}
```

Figure 21.13: Weights for $(6, 2)$

Finally we show the diagram for high weight $(10, 7)$:

Figure 21.14: Weights for $(10, 7)$

21.10 The Code to Compute Characters of $SU(3)$

We begin by repeating the earlier code with a few minor additions:

```

ClearAll[A1];
ClearAll[A2];
Clear[i];
Clear[j];
(* The two globals below transfer the highest weight to the graphic routine. *)
HighK = 1;
HighL = 1;
f[x_, y_] := 0;
A1 = Array[f, {202, 202}];
A2 = Array[f, {202, 202}];
(* The global below transfers the highest remaining element in the numerator
from one routine to another. *)
transferlist = {0, 0, 0};
(* The global below lists all weights. It is constructed during the calculation
of the character, and transfers the information to the graphics routine. *)
weightlist = { };

(* Weyl Group *)
w1 = {{1, 0}, {0, 1}};
w2 = {{-1, -1}, {1, 0}};
w3 = {{0, 1}, {-1, -1}};
w4 = {{1, 1}, {0, -1}};
w5 = {{-1, 0}, {1, 1}};
w6 = {{0, -1}, {-1, 0}};
w = {w1, w2, w3, w4, w5, w6};
WeylOrder = 6;
(* So w is the Weyl group and w[[2]] is the second element.
The first three elements have determinant one and the second
three have determinant -1. *)

(* Next we write the key routines. These routines form the symmetric
character generated by a weight (h, k) which contains six elements
when (h, k) is not on the boundary of a fundamental chamber and three
elements when it is a boundary element. It multiplies this character
by the denominator and subtracts the result from the array A1. *)

```

```

ClearArray := Block[{i, j},
  For[i = -100, i <= 100, i++,
    For[j = - 100, j <= 100, j++,
      A1[[i + 101, j + 101]] = 0;
      A2[[i + 101, j + 101]] = 0;
    ]];

WriteNumerator[k_, l_] := Block[{i, j, res, a, b, s1},
  For[i = 1, i <= WeylOrder, i++,
    res = w[[i]].{1 + k, 1 + l};
    a = res[[1]];
    b = res[[2]];
    If[(i <= (WeylOrder / 2)), s1 = 1, s1 = -1];
    A1[[a + 101, b + 101]] = s1;
  ];

RemoveHighWeightZero[ dimension_] := Block[{i, j, res, a, b, s1},
  For[i = 1, i <= WeylOrder, i++,
    res = w[[i]].{1, 1 };
    a = res[[1]];
    b = res[[2]];
    If[(i <= (WeylOrder / 2)), s1 = 1, s1 = -1];
    A1[[a + 101, b + 101]] =
      A1[[a + 101, b + 101]] - s1 * dimension;
  ];

RemoveHighWeightNonZero[k_, l_, dimension_] :=
Block[{i, j, res, a, b, s1, s2, s3},
  For[i = 1, i <= WeylOrder, i++,
    For[j = 1, j <= WeylOrder, j++,
      res = (w[[i]].{1, 1}) + (w[[j]].{k, l});
      a = res[[1]];
      b = res[[2]];
      If[(i <= (WeylOrder/2)), s1 = 1, s1 = -1];
      If[(j <= (WeylOrder/2)), s2 = 1,
        If[(k == 0) || (l == 0), s2 = 0, s2 = 1]];
      s3 = s1*s2;
      A1[[a + 101, b + 101]] = A1[[a + 101, b + 101]] - dimension*s3;
    ];];

```

```

For[j = 1, j <= WeylOrder, j++,
  res = (w[[j]] . {k, l});
  a = res[[1]];
  b = res[[2]];
  If[(j <= (WeylOrder/2)), s2 = 1,
    If[(k == 0) || (l == 0), s2 = 0, s2 = 1]];
  A2[[a + 101, b + 101]] = A2[[a + 101, b + 101]] + dimension * s2;
];];

RemoveHighWeight[k_, l_, dimension_] :=
  If[((k == 0) && (l == 0)),
    RemoveHighWeightZero[dimension],
    RemoveHighWeightNonZero[k, l, dimension];];

ListOuterRing[k_, l_] := Block[{alist, aweight},
  alist = { };
  aweight = {k, l};
  While[((aweight[[1]] >= 0) && (aweight[[2]] >= 0)),
    alist = Insert[alist, aweight, 1];
    aweight = {aweight[[1]] - 2, aweight[[2]] + 1};
  ];
  aweight = {k, l};
  aweight = {aweight[[1]] + 1, aweight[[2]] - 2};
  While[((aweight[[1]] >= 0) && (aweight[[2]] >= 0)),
    alist = Insert[alist, aweight, 1];
    aweight = {aweight[[1]] + 1, aweight[[2]] - 2};
  ];
  Return[alist];
];

RemoveOuterRing[k_, l_, dimension_] := Block[{alist, blist, i, j },
  alist = ListOuterRing[k, l];
  j = Length[alist];
  For[i = 1, i <= j, i++,
    a = alist[[i]][[1]];
    b = alist[[i]][[2]];
    RemoveHighWeight[a, b, dimension];];];

ShowData := Block[ { i, j, dim },
  For[i = -100, i <= 100, i++,
    For[j = -100, j <= 100, j++,

```

```

    dim = A1[[i + 101, j + 101]];
    If[ (dim != 0),
      Print[ dim, " - (", i, ", ", j, ")"];
    ];];];];

ShowFundamentalData := Block[ { i, j, dim, mylist },
  mylist = {0, 0, 0};
  For[i = -100, i <= 100, i++,
    For[j = -100, j <= 100, j++,
      dim = A1[[i + 101, j + 101]];
      If[ ((dim > 0) && (i >= 0) && (j >= 0)),
        mylist = {i, j, dim} ;
      ];
    ];
  ];
  transferlist = mylist;
];

(* Here is the crucial routine which computes characters. It lists the
   weights in the outer ring, adds them to "weightlist", and removes
   them the numerator in the standard way. Then it looks at the
   remaining numerator, finds the highest character and its dimension
   and uses that information to process the next ring. Etc. *)
CalculateCharacter[k_, l_] :=
  Block[{mylist, dim, newdim, a, b, templist},
    ClearArray;
    HighK = k;
    HighL = l;
    weightlist = {};
    WriteNumerator[k, l];
    Print[k, " ", l, " ", 1];
    templist = ListOuterRing[k, l];
    weightlist = Join[weightlist, templist];
    Print[templist ];
    RemoveOuterRing[k, l, 1];
    ShowFundamentalData;
    mylist = transferlist;
    dim = mylist[[3]];

    While[(dim > 0),
      a = mylist[[1]];

```

```

    b = mylist[[2]];
    dim = mylist[[3]];
    Print[a - 1, " ", b - 1, " ", dim];
    templist = ListOuterRing[a - 1, b - 1];
    weightlist = Join[weightlist, templist];
    Print[templist ];
    RemoveOuterRing[a - 1, b - 1, dim];
    ShowFundamentalData;
    mylist = transferlist;
    dim = mylist[[3]];
    ];
    ShowData;];

PrintAllWeights[ ] := Block[ {i, j, dim},
  For [i = -100, i <= 100, i++,
    For[j = - 100, j <= 100, j++,
      dim = A2[[i + 101, j + 101]];
      If[ (dim > 0),
        Print["Gdot[" , i, " ,", j, " ],"];
      ];];];];

PrintAllWeightsAndDimensions[] := Block[{i, j, dim},
  For [i = -100, i <= 100, i++,
    For[j = - 100, j <= 100, j++,
      dim = A2[[i + 101, j + 101]];
      If[ (dim > 0),
        Print["Gdot[" , i, " ,", j, " ],"];
      ];
    ];
  ];
  Print["RGBColor[0.0, 0.0, 0.0],"];
  For [i = -100, i <= 100, i++,
    For[j = - 100, j <= 100, j++,
      dim = A2[[i + 101, j + 101]];
      If[ (dim > 0),
        Print["Tdot[" , i, " ,", j, " ,", dim, " ],"];
      ];];];];

```


21.11 The Code to Draw Weight Diagrams

Here is the program to draw the weight diagram:

```

h = Sin[ Pi / 3];
Gdot[k_, l_] := Disk[{k , 1 Sqrt[ 1 + 4 h h / 9] + (k 2 h / 3)} , .2];
Tdot[k_, l_, dim_] := Text[dim,{k + .2 , 1 Sqrt[ 1 + 4 h h / 9] + (k 2 h / 3) + .2}];
grayheight = HighL Sqrt[ 1 + 4 h h / 9] + (HighK 2 h / 3);
Show[Graphics[{
  {GrayLevel[0.9],
    Polygon[{0, 0}, {0, grayheight}, {grayheight * Sqrt[3],
      grayheight}]}],
  AbsoluteThickness[2],
  Arrow[{0, 0}, {2, 0}],
  Arrow[{0, 0}, {2 Cos[ Pi/3], 2 Sin[ Pi/3]}],
  Arrow[{0, 0}, {2 Cos[ 2 Pi/3], 2 Sin[ 2 Pi/3]}],
  AbsoluteDashing[{7, 7}],
  Arrow[{0, 0}, {2 Cos[ 3 Pi/3], 2 Sin[ 3 Pi/3]}],
  Arrow[{0, 0}, {2 Cos[ 4 Pi/3], 2 Sin[ 4 Pi/3]}],
  Arrow[{0, 0}, {2 Cos[ 5 Pi/3], 2 Sin[ 5 Pi/3]}],
  AbsoluteThickness[1],
  AbsoluteDashing[0],

  RGBColor[0.0, 1.0, 0.0],
  (* Dot Commands *)

}]]

```

Chapter 22

Characters of $SO(5)$

22.1 Introduction

In this chapter we continue the story for rank 2 groups, by revising our program to handle $SO(5)$. A glance at the code shows that this should be easy, because we only have to revise the lines describing the Weyl group, and the lines finding the outer rim by adding the negatives of the basis vectors for the simple roots. In addition, the code which graphically displays weights contains formulas which convert from the coordinate system imposed by the Fundamental Weyl Chamber to standard Euclidean coordinates, and these lines must be revised.

On the other hand, we need to understand more clearly the case when not all roots have the same length.

Recall that $SO(5)$ is not simply connected, and instead has a two-fold universal covering group, the spinor group. So some of the representations we present exist for the spinor group but do not drop down to $SO(5)$.

22.2 Initial Steps for $SO(5)$

We can take $(1, 0)$ and $(-1, 1)$ as the simple roots. They form a basis of the root lattice, and their coroots form a basis for the coroot lattice. These coroots are $\frac{2(1,0)}{\langle(1,0),(1,0)\rangle} = (2, 0)$ and $\frac{2(-1,1)}{\langle(-1,1),(-1,1)\rangle} = \frac{2(-1,1)}{2} = (-1, 1)$. The weight lattice is dual to the lattice generated by these coroot vectors. Therefore (x, y) belongs to the weight lattice exactly when $2x \in Z$ and $-x + y \in Z$. Thus these lattice points are points where a line of the form $x = \frac{m}{2}$ meets a line of the form $y = x + n$. A basis of this lattice is $v_1 = (\frac{1}{2}, \frac{1}{2})$ and $v_2 = (0, 1)$. Recall that the roots and weights live in the same vector space, while the coroots live in the

dual of this space. Since the inner product is only determined up to a positive scalar, we can multiply both root and weight vectors by the same positive constant, but this changes nothing in the final formulas.

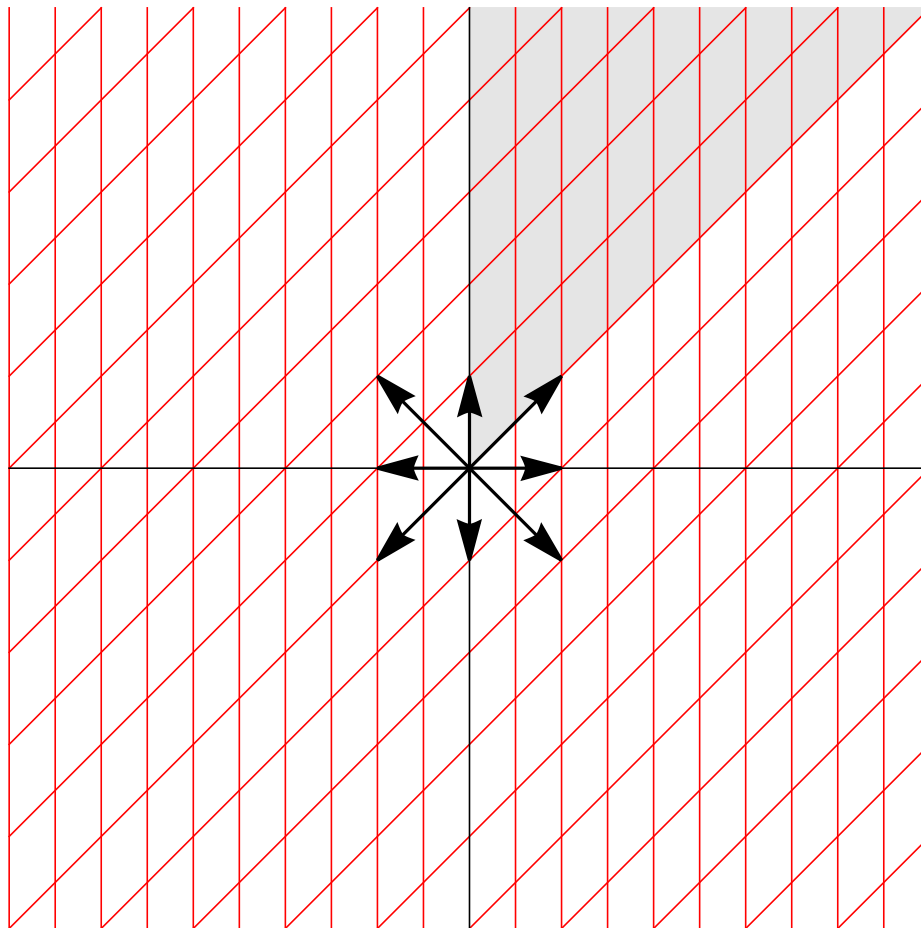


Figure 22.1: Weight Lattice for $SO(5)$

Given this diagram, we can read off the two translations by $-v_1$ and $-v_2$, when written in the coordinate system the diagram generates for the Fundamental Weyl Chamber. Translation by $-v_1$ moves one unit left, but in these coordinates that becomes $(-2, 1)$. Translation by $-v_2$ moves right and down, but in these coordinates that becomes $(2, -2)$.

Next we find the eight Weyl group elements. We want these matrices in the coordinates of the Fundamental Chamber. An easy way to find these matrices is to keep track of the four short roots, which in these coordinates are $(2, -1), (0, 1), (-2, 1), (0, -1)$. For example,

rotation by 90 degrees takes the first vector to the second one and the second to the third. A brief calculation shows that it is

$$\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}$$

The powers of this matrix then give the four rotations, which are

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

Reflection across the right horizontal arrow takes $(2, -1)$ to itself and $(0, 1)$ to $(0, -1)$, so this matrix is

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

So the remaining four elements of the Weyl group are the products of this matrix with the four rotations, or

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

These calculations are enough to modify the code which calculates the character of a representation. We can use the previous code, changing only the lines which define the Weyl Group to the following:

```
(*Weyl Group*)
w1={{1,0},{0,1}};
w2={{-1,-2},{1,1}};
w3={{-1,0},{0,-1}};
w4={{1,2},{-1,-1}};
w5={{1,0},{-1,-1}};
w6={{-1,-2},{0,1}};
w7={{-1,0},{1,1}};
w8={{1,2},{0,-1}};

w={w1,w2,w3,w4,w5,w6, w7, w8};
WeylOrder=8;
```

and changing the entire procedure ListOuterRing[k, l] to the following:

```

ListOuterRing[k_,l_]:=Block[{alist,aweight},alist={};
aweight={k,l};
While[((aweight[[1]]>=0)&&(aweight[[2]]>=0)),alist=Insert[alist,aweight,1];
aweight={aweight[[1]]-2,aweight[[2]]+1};];
aweight={k,l};
aweight={aweight[[1]]+2,aweight[[2]] - 2 };
While[((aweight[[1]]>=0)&&(aweight[[2]]>=0)),alist=Insert[alist,aweight,1];
aweight={aweight[[1]]+2,aweight[[2]]-2};];
Return[alist];];

```

The code to draw weight diagrams should be replaced by the following:

```

Gdot[k_,l_]:=Disk[{k/ 2,k/2 + 1}, .15];
Tdot[k_, l_, dim_] := Text[dim, {k/ 2 + .2, k/2 + 1 + .2}];
Show[Graphics[{
GrayLevel[0.9],
Polygon[{{0,0}, {0, HighK / 2 + HighL}, {HighK / 2 + HighL, HighK / 2 + HighL}}],
AbsoluteThickness[2],
RGBColor[0.0, 0.0, 0.0],
Arrow[{{0, 0}, {1, 0}}],
Arrow[{{0, 0}, {1, 1}}],
Arrow[{{0, 0}, {-1, 1}}],
Arrow[{{0, 0}, {0, 1}}],
Arrow[{{0, 0}, {-1, 0}}],
Arrow[{{0, 0}, {-1, -1}}],
Arrow[{{0, 0}, {1,- 1}}],
Arrow[{{0, 0}, {0, -1}}],

RGBColor[0.0,1.0,0.0],
(*Dot Commands*)

}]]

```

22.3 Examples

When playing with examples, it is convenient to save the following lines in a separate file and add the drawing code above to that file. To compute a new example, copy this material to the appropriate spot in a Mathematica document and proceed.

```

CalculateCharacter[k, l];
PrintAllWeightsAndDimensions[];

```

High Weight (2, 1):

```
CalculateCharacter[2, 1];
```

```
2 1 1
{{0,2},{2,1}}
2 0 2
{{0,1},{2,0}}
0 1 1
{{0,1}}
0 0 3
{{0,0}}
```

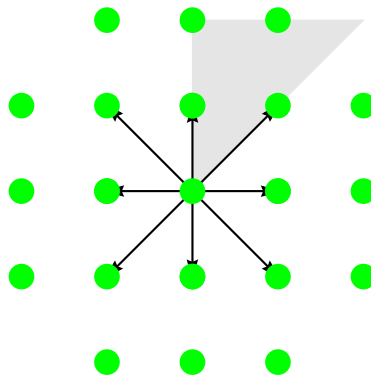
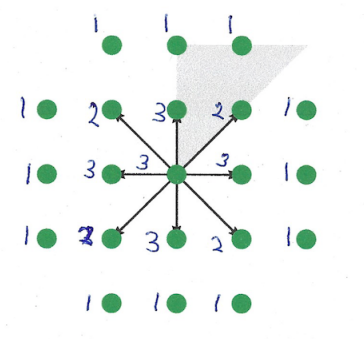


Figure 22.2: High Weight (2, 1)

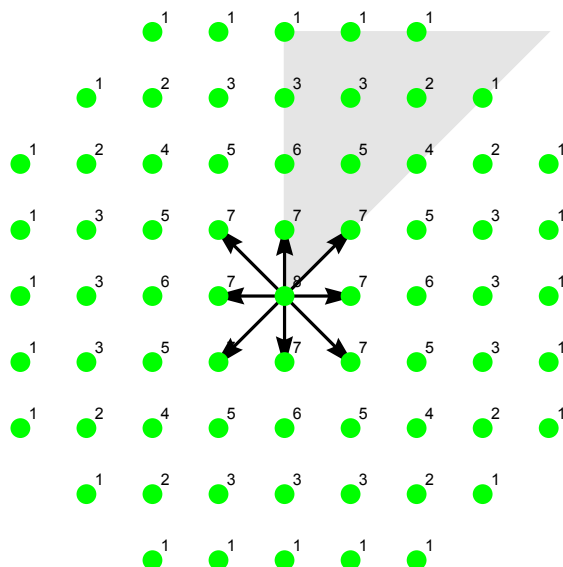
Remark: Look at the output of “CalculateCharacter” and notice that one weight, $(0, 1)$, is mentioned twice. On the the fourth output line it is assigned dimension 2, but on the sixth line it is assigned one more dimation. When I first noticed this, I spent a day trying to find the error in the code. However, the code is correct and it is no longer true that all weights in a ring have the same dimension, or that this dimension increases with easy regularity. Instead it is important to use the character formula to obtain the dimensions. I subsequently revised the code to automatically generate the dimension data and output it on the graph upon request. In this case, we get the following data.

Figure 22.3: Dimensions for $(3, 2)$

We earlier used the Weyl dimension formula to show that the dimension of the representation of $SO(5)$ with highest weight (k, l) is $\frac{(k+1)(l+1)(k+l+2)(k+2l+3)}{3!}$. When $(k, l) = (2, 1)$, this number is 35. An easy check shows that this is the dimension predicted by the above picture.

Remark: We now compute some other examples. When $k \neq l$, we get a lopsided octagon, whose rings eventually collapse to a square.

High Weight $(4, 2)$:

Figure 22.4: High Weight $(4, 2)$

Remark: When $k = l$, we get perfect hexagons.

High Weight (3, 3):

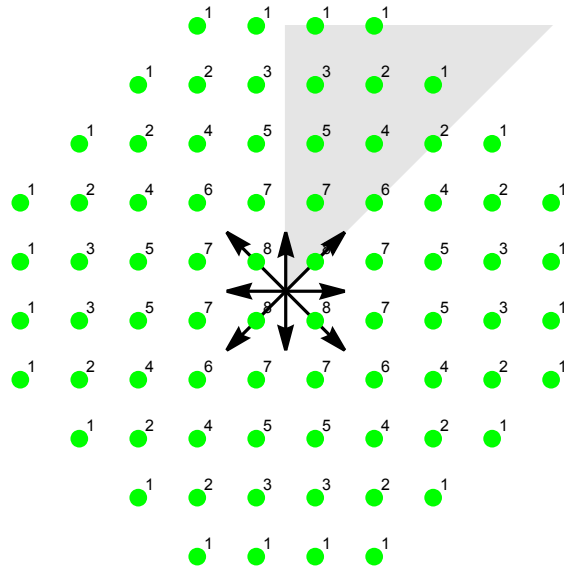


Figure 22.5: High Weight (3, 3)

Remark: When $k = 0$ or $l = 0$, we get squares. But the patterns depends on which variable is zero. Here is (4, 0):

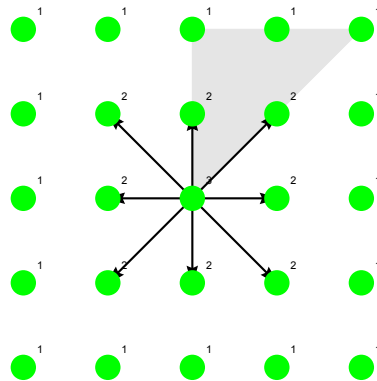


Figure 22.6: High Weight (4, 0)

High Weight $(0, 3)$:

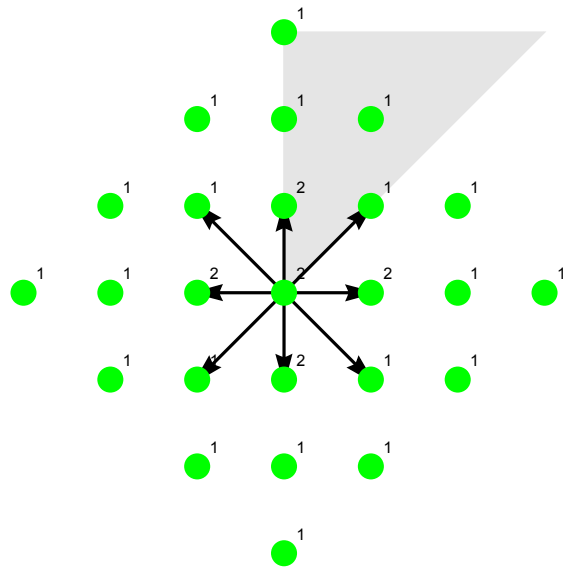


Figure 22.7: High Weight $(0, 3)$

Remark: Finally, two more random examples. Here is $(1, 0)$:

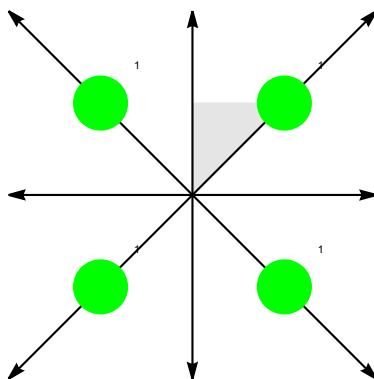


Figure 22.8: High Weight $(1, 0)$

High Weight $(1, 2)$:

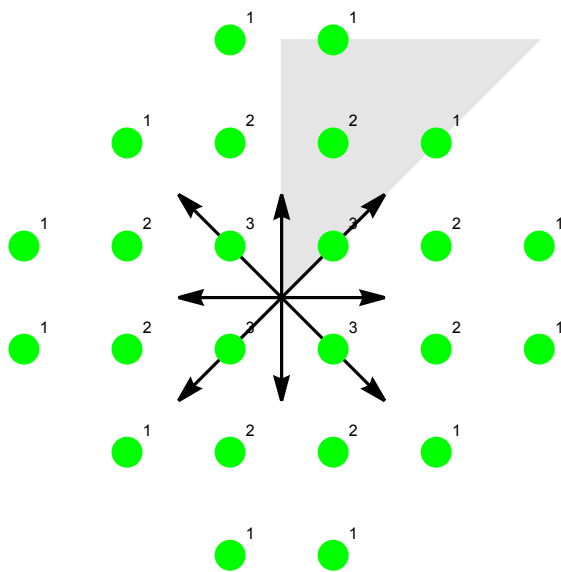


Figure 22.9: High Weight $(1, 2)$

Chapter 23

Characters for G_2

23.1 Introduction

In this short chapter, we revise our program to cover the last rank 2 group, G_2 , and calculate a few characters for representations of this group.

The root diagram is shown below. We can normalize so α has length 1 and thus $\alpha = (1, 0)$. Each angle between roots is thirty degrees, so the short root just right of the y -axis equals $(\cos 60, \sin 60) = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. Therefore $\beta = \lambda \left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$ where λ is chosen so the second coordinate is $\frac{\sqrt{3}}{2}$. This makes $\lambda = \sqrt{3}$ and thus $\beta = \left(-\frac{3}{2}, \frac{\sqrt{3}}{2}\right)$.

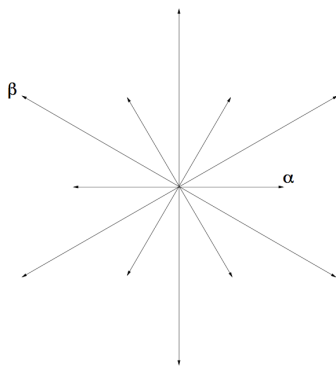
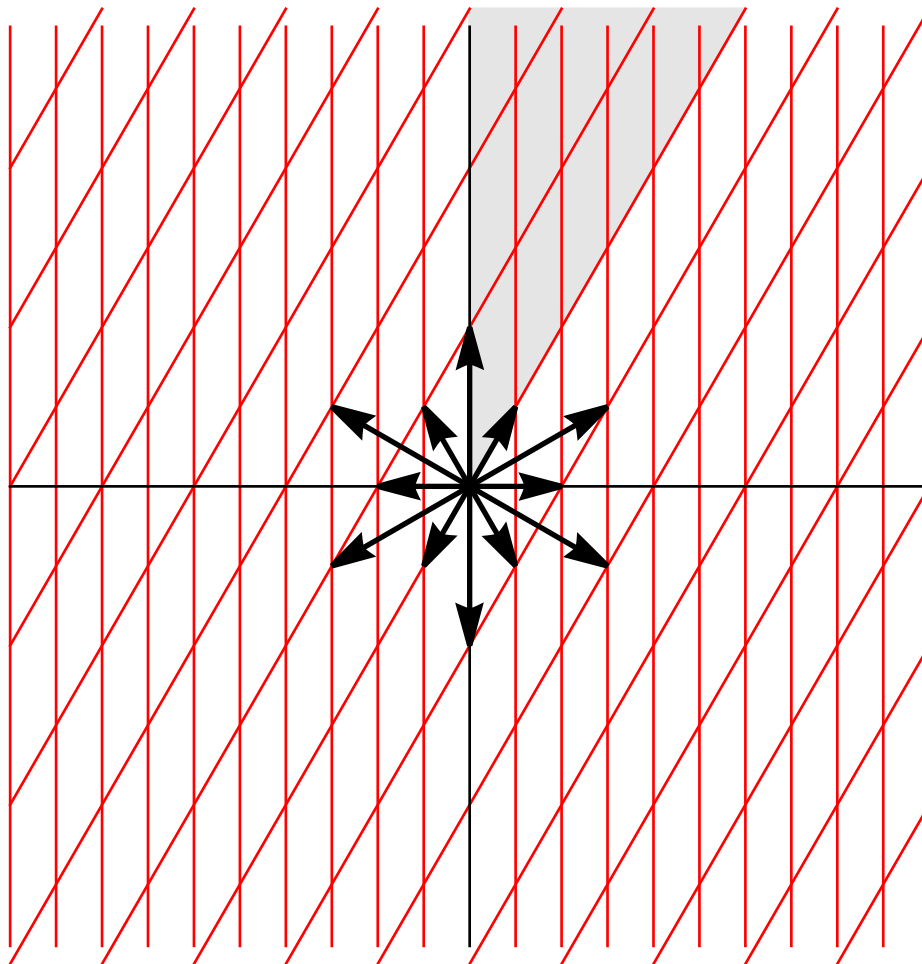


Figure 23.1: g_2

The corresponding coroots are $(2, 0)$ and $\frac{2\left(-\frac{3}{2}, \frac{\sqrt{3}}{2}\right)}{3} = \left(-1, \frac{1}{\sqrt{3}}\right)$. The dual lattice is drawn below.

Figure 23.2: Weight Lattice for G_2

Next we need to compute the Weyl group. Half of this group is generated by rotation counterclockwise by 60 degrees. This map sends α to $(1, 0)$ and $(1, 0)$ to $(-1, 1)$. Note that $\alpha = (2, -1)$. A brief calculation shows that the resulting matrix must be $\begin{pmatrix} -1 & -3 \\ 1 & 2 \end{pmatrix}$. The six powers of this matrix are then

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & -3 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} -2 & -3 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ -1 & -2 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ -1 & -1 \end{pmatrix}$$

Next we consider reflection across α . This map sends $\alpha = (2, -1)$ to itself and maps $(0, 1)$ to $(0, -1)$, where $(0, 1)$ is the long vertical root. A brief calculation shows that the matrix for this reflection is $\begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix}$. The six reflections in the Weyl group can be obtained by multiplying each rotation by this matrix:

$$\begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ -1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} -2 & -3 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} -1 & -3 \\ 0 & 1 \end{pmatrix}$$

Finally, in our code we need the vectors $-\alpha$ and $-\beta$. These are clearly $(-2, 1)$ and $(3, -2)$.

Now that these calculations are done, it is easy to modify the code. Three modifications are required, exactly as for $SO(5)$. First the Weyl group code now needs to read

```
(*Weyl Group*)
w1={{1,0},{0,1}};
w2={{-1, -3}, {1, 2}};
w3={{-2, -3}, {1, 1}};
w4={{-1, 0}, {0, -1}};
w5={{1, 3}, {-1, -2}};
w6={{2, 3}, {-1, -1}};

w7={{1, 0}, {-1, -1}};
w8={{2, 3}, {-1, -2}};
w9={{1, 3}, {0, -1}};
w10={{-1, 0}, {1, 1}};
w11={{-2, -3}, {1, 2}};
w12={{-1, -3}, {0, 1}};

w={w1,w2,w3,w4,w5,w6, w7, w8, w9, w10, w11, w12};
WeylOrder=12;
```

Next we must replace ListOuterRing[k, l] with the following code:

```
ListOuterRing[k_, l_] := Block[{alist, aweight}, alist = {};
  aweight = {k, l};
  While[((aweight[[1]] >= 0) && (aweight[[2]] >= 0)), alist = Insert[alist, aweight, 1];
  aweight = {aweight[[1]] - 2, aweight[[2]] + 1};];
  aweight = {k, l};
  aweight = {aweight[[1]] + 3, aweight[[2]] - 2};
  While[((aweight[[1]] >= 0) && (aweight[[2]] >= 0)), alist = Insert[alist, aweight, 1];
  aweight = {aweight[[1]] + 3, aweight[[2]] - 2};];
  Return[alist];];
```

Finally, we must replace the routine to draw weights with the following code:

```
sq3 = Sqrt[3];
c11 = 1/2;
s11 = sq3 / 2;
dy1 = (HighK/2) sq3 + (HighL) sq3;
dx1 = dy1 / sq3;
Gdot[k_, l_] := Disk[{k/2, (k/2 + 1) sq3}, .15];
Tdot[k_, l_, dim_] := Text[dim, {k/2 + .2, (k/2 + 1) sq3 + .2}];
Show[Graphics[{
  GrayLevel[0.9],
  Polygon[{{0, 0}, {0, dy1}, {dx1, dy1}}],
  AbsoluteThickness[2], RGBColor[0.0, 0.0, 0.0],
  Arrow[{{0, 0}, {1, 0}}],
  Arrow[{{0, 0}, {c11, s11}}],
  Arrow[{{0, 0}, {-c11, s11}}],
  Arrow[{{0, 0}, {-1, 0}}],
  Arrow[{{0, 0}, {-c11, -s11}}],
  Arrow[{{0, 0}, {c11, -s11}}],
  Arrow[{{0, 0}, {0, sq3}}],
  Arrow[{{0, 0}, {0, -sq3}}],
  Arrow[{{0, 0}, {sq3 s11, sq3 c11}}],
  Arrow[{{0, 0}, {-sq3 s11, sq3 c11}}],
  Arrow[{{0, 0}, {-sq3 s11, -sq3 c11}}],
  Arrow[{{0, 0}, {sq3 s11, -sq3 c11}}],

  RGBColor[0.0, 1.0, 0.0],
  (*Dot Commands*)
}],
{Gdot[k, l], Tdot[k, l, dim]}]
```

Remark: We end with a few examples. Even relatively small high weights lead to a large number of weights in this case.

High Weight $(1, 0)$:

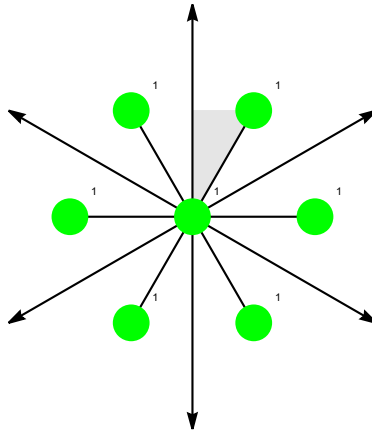


Figure 23.3: High Weight $(1, 0)$

High Weight $(0, 1)$:

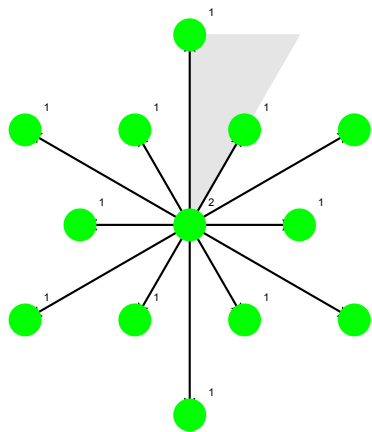


Figure 23.4: High Weight $(0, 1)$

High Weight (1, 1):

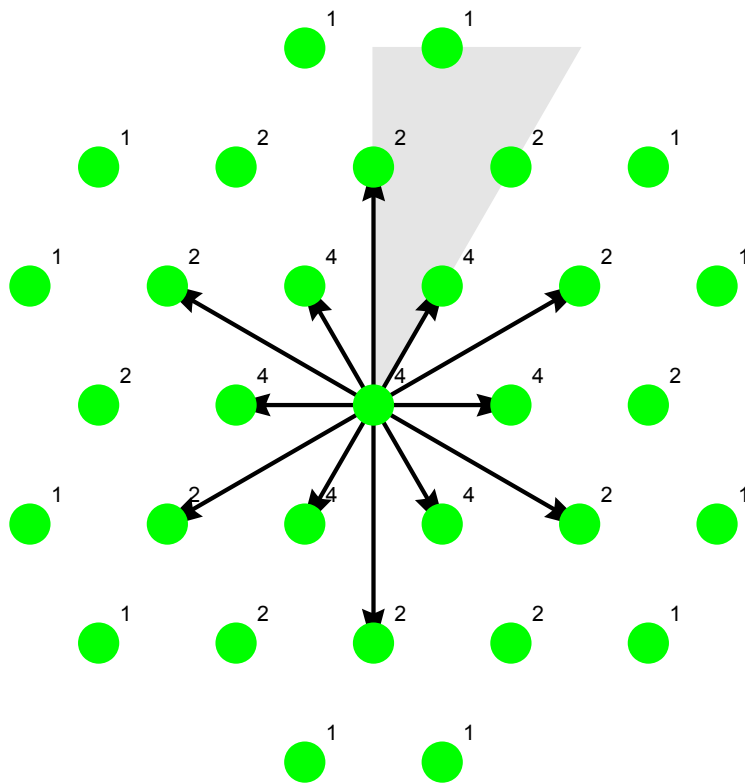


Figure 23.5: High Weight $(1, 1)$

High Weight (2, 4):

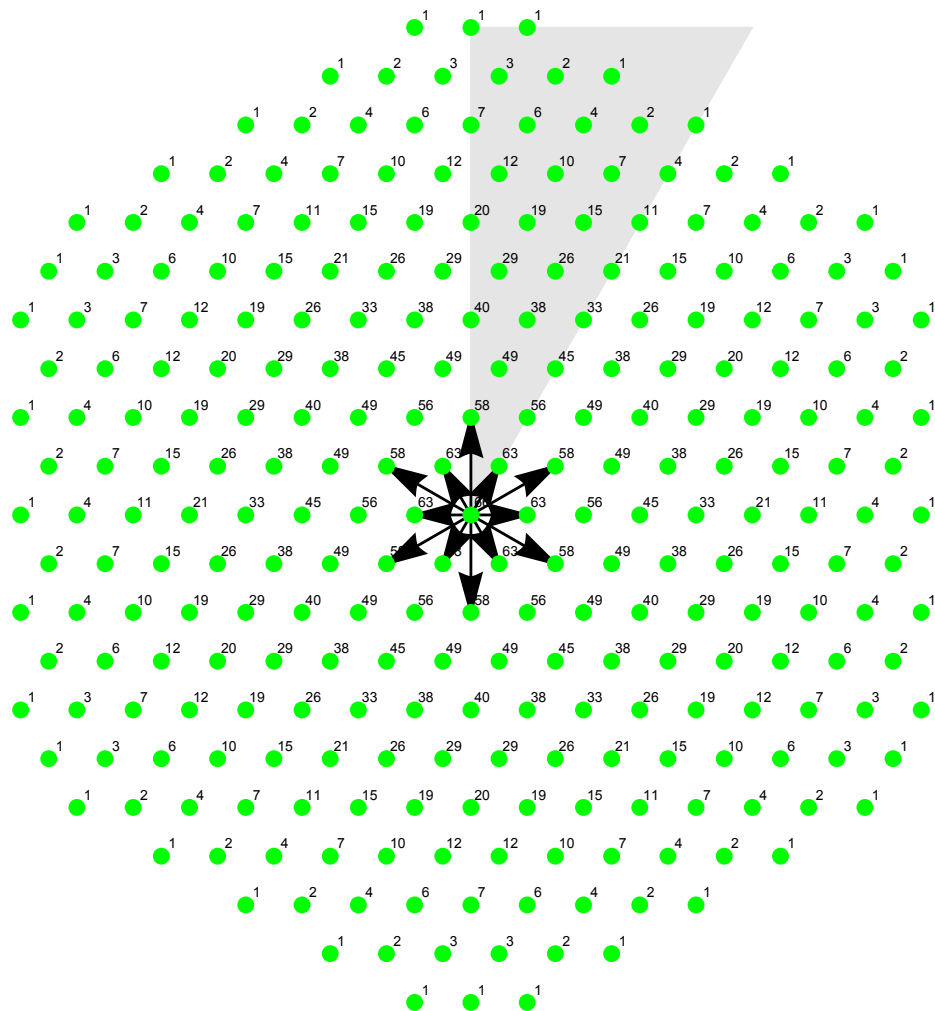


Figure 23.6: High Weight (2, 4)

23.2 Other Software

Around 1990, the computer algebra group at the Centre For Mathematics and Computer Science in Amsterdam wrote a program called LiE, which can compute characters given the highest weight of the representation, using Weyl's formula. The program is able to answer other questions about compact groups as well. This program is still available in 2020. The program runs on a server, so users can access the server and ask for a particular calculation. But the source code, written in C, is also available. It is easy to compile the code on a Macintosh. The web site <http://www-math.univ-poitiers.fr/~maavl/LiE/> provides a manual and further information, as well as the source code.

The online service is available at <http://www-math.univ-poitiers.fr/~maavl/LiE/form.html>.

For example, the server allows a user to ask for various items. I asked for a complete character table. The server asked for a group and I selected B2. The server asked for a highest weight and I selected $[2, 1]$. Then server then produced the following list:

$$\begin{aligned} &1X[2, 1] + 1X[3, -1] + 1X[0, 3] + 2X[1, 1] + 1X[-2, 5] + 2X[2, -1] + \\ &2X[-1, 3] + 1X[3, -3] + 3X[0, 1] + 1X[-3, 5] + 3X[1, -1] + 2X[-2, 3] + \\ &2X[2, -3] + 3X[-1, 1] + 1X[3, -5] + 3X[0, -1] + 1X[-3, 3] + 2X[1, -3] + \\ &2X[-2, 1] + 1X[2, -5] + 2X[-1, -1] + 1X[0, -3] + 1X[-3, 1] + 1X[-2, -1] \end{aligned}$$

Note that a number of problems were faced by the authors of this software that we skimmed above. In rank 2, the Weyl group is very small and can be coded by hand, but larger Weyl groups are much larger, so a systematic method must be invented to produce them. Our division algorithm to expand Weyl's quotient is notably ad-hoc; much more sophisticated methods must be employed in the general case.

The software is able to handle groups of rank smaller than about 10, and this is sufficient to process representations of the exceptional groups for small highest weight.

An advantage of this software is that it introduces the same basis for weights used here, so results are immediately meaningful. There is also software in Sage to compute characters, but that software uses a different coordinate basis and I found it less illuminating than LiE.

Chapter 24

The Symmetric Group

24.1 Introduction

After his papers on the Weyl Character Formula, Weyl turned to the matter of finding the actual irreducible representations described by these characters, at least for the groups $SU(n)$, $SO(n)$, and $Sp(n)$. These representations were determined in his book *The Classical Groups* published by Weyl in 1939. Unlike the earlier work, which depended on a beautiful mix of ideas from many branches of mathematics, this book was very algebraic.

The easiest case, and arguably the most important in physics, is $SU(n)$; we will describe its irreducible representations in the next chapter. The same ideas apply to the other groups, but with additional pesky details. These details are available in other books.

The group $SU(n)$ acts on a complex vector space V . Weyl's basic idea was to extend this representation to $V \otimes V \otimes \dots \otimes V$, and then decompose this representation into irreducible sub representations. We will discover that all irreducible representations of $SU(n)$ arise in this way.

The group of permutations $\mathcal{S}_n$ also acts on $V \otimes V \otimes \dots \otimes V$. The representations of $SU(n)$ and $\mathcal{S}_n$ commute with each other, since $SU(n)$ acts on each factor separately while preserving their order, and $\mathcal{S}_n$ permutes the factors but acts trivially on each one. A key result says that all operators commuting with the $\mathcal{S}_n$ come from $SU(n)$, and all operators commuting with the $SU(n)$ come from $\mathcal{S}_n$. This turns out to imply that the irreducible representations of the two groups are closely related, and ultimately the representations of $SU(n)$ arise from a close study of the representations of the finite group $\mathcal{S}_n$. This chapter is devoted to these representations.

24.2 Permutation Groups

Definition 48 A permutation on n objects is a one-to-one, onto map

$$\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$$

The composition of two permutations is again a permutation, and so the set of all permutations on n objects forms a group $\mathcal{S}_n$, called the symmetric group.

Remark: We can describe a permutation by writing two rows of numbers, with sources on the first row and images on the second row. For instance

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 1 & 3 & 7 & 4 & 2 & 5 \end{pmatrix}$$

Since $(\sigma \circ \tau)(k) = \sigma(\tau(k))$, permutations compose from right to left. Therefore

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 1 & 3 & 7 & 4 & 2 & 5 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 7 & 4 & 1 & 2 & 3 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 5 & 7 & 6 & 1 & 3 & 4 \end{pmatrix}$$

Be careful because a few authors multiply in the other order.

Each integer k generates a cycle $k, \sigma(k), \sigma^2(k), \dots$ until we return to the start. Any permutation can be written as a product of such cycles, as in the example below.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 1 & 3 & 7 & 4 & 2 & 5 \end{pmatrix} = (162)(3)(475)$$

This is called *cycle notation*; the order in which we write cycles does not matter, but it is customary to write longer cycles first and shorter ones later. Some authors omit cycles with only one element.

More generally, we can write products of cycles where the same number appears in more than one cycle. Then order definitely matters. This situation comes up when we compute products of permutations written in cycle notation, and of course we continue to work from right to left, so

$$(162)(3)(475) \circ (1634)(275) = (125)(3746)$$

The “cycle pattern” of a permutation is the unordered list of cycle lengths when the permutation is written in cycle notation. Thus the cycle pattern of $(1629)(3)(475)(8)$ is 4, 1, 3, 1, which is the same as 4, 3, 1, 1.

Theorem 123 Two permutations are conjugate in $\mathcal{S}_n$ if and only if they have the same cycle pattern.

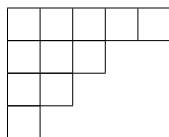
Proof: Let us compare σ and $\tau \circ \sigma \circ \tau^{-1}$. Suppose τ takes $1, 2, \dots, n$ to $k_1, k_2, \dots, k_n$. Then if σ takes i to j , $\tau \circ \sigma \circ \tau^{-1}$ takes k_i to k_j . Consequently if we know the cycle notation for σ , we can get the cycle notation for $\tau \circ \sigma \circ \tau^{-1}$ by just replacing each i with k_i . The cycle pattern remains the same. QED.

Remark: A cycle pattern for an element of $\mathcal{S}_n$ is determined by the list of cycle lengths, and thus by a list of positive integers which add up to n . Such a list is called a *partition* of n . An important chapter of number theory is devoted to determining properties of these integers.

The partitions of 3 are $3, 2 + 1, 1 + 1$, so $p_3 = 3$. The partitions of 4 are $4, 3 + 1, 2 + 2, 2 + 1 + 1, 1 + 1 + 1 + 1$, so $p_4 = 5$.

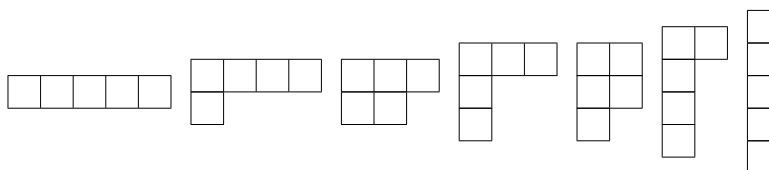
Remark: The order of $\mathcal{S}_n$ is $n!$ because we can send 1 to n numbers, and then send 2 to any of the $n - 1$ remaining numbers, etc. The number of irreducible representations of a finite group equals the number of conjugacy classes of the group. It follows that $\mathcal{S}_n$ has exactly p_n irreducible representations.

There is a graphical way to write these cycles; as lines of boxes. To draw the partition $5 + 3 + 2 + 1 = 11$, draw five boxes, and under them 3 boxes, and then 2 boxes and 1 box as shown below. A diagram of this sort is called a Young Diagram. Each line must start at the left side; lines grow shorter as we go down the diagram, but two lines of the same length are allowed.



Later on, we will put numbers in the boxes, and the Young approach will become more and more essential. But we might as well start with empty boxes right away.

Here are all possible diagrams when $n = 5$:



Let p_n be the number of partitions of n , and so the number of irreducible representations of $\mathcal{S}_n$. We will spend the rest of this section computing these numbers, a classical combinatorics problem going back to Euler. This computation is not used in later developments and can be skipped. We include it because of the beautiful mathematics involved and interesting history generated by the problem.

The first twenty-five values of p_n are

n	p_n		n	p_n		n	p_n		n	p_n
1	1		6	11		11	56		16	231
2	2		7	15		12	77		17	297
3	4		8	22		13	101		18	385
4	5		9	30		14	135		19	490
5	7		10	42		15	176		20	627

There is no closed formula for p_n , but there is a beautiful recursive way to compute them. Encode the p_n in a formal sum

$$1 + p_1x + p_2x^2 + p_3x^3 + p_4x^4 + p_5x^5 + \dots = 1 + x + 2x^2 + 3x^3 + 5x^4 + 7x^5 + \dots$$

This sum is formally

$$\prod_{k \geq 1} (1 + x^k + x^{2k} + x^{3k} + \dots)$$

We think of this product as follows. Suppose we want the first fifth terms of the product. Then terms x^n with $n > 50$ are irrelevant, so we can replace the first fifth terms by finite polynomials, and all other terms by 1. Thus the product makes sense.

An easy argument shows that the product will be our infinite series involving the p_n . When we multiply, we can get a power of x , say x^{50} , by picking a monomial from the first term, a monomial from the second term, etc., and finally a monomial from the 50th term, so these elements multiply to give x^{50} , and then adding over all the possible ways to do this. The monomials will be $x^{k_1}, x^{2k_2}, \dots, x^{50k_{50}}$, where $k_1 + 2k_2 + \dots + 50k_{50} = 50$. But such a choice of k_i is just a partition of 50 with k_1 1's, and k_2 2's, etc. So the number of ways to make the choice is p_{50} .

Euler noticed that $(1 + x^k + x^{2k} + x^{3k} + \dots) = \frac{1}{1-x^k}$ and therefore

$$1 + p_1x + p_2x^2 + p_3x^3 + p_4x^4 + p_5x^5 + \dots = \frac{1}{\prod_{k \geq 1} (1 - x^k)}$$

Now suppose that there is a nice formula for $\prod_{k \geq 1} (1 - x^k)$ so we can easily compute its expansion $1 + b_1x + b_2x^2 + \dots$. In that case,

$$(1 + p_1x + p_2x^2 + p_3x^3 + \dots) (1 + b_1x + b_2x^2 + b_3x^3 + \dots) = 1$$

and so each coefficient of the product except the initial one vanishes, or

$$p_n + p_{n-1}b_1 + p_{n-2}b_2 + \dots + b_n = 0$$

We could then easily compute the p_n recursively.

Euler then multiplied out the first 50 terms of $\prod(1 - x^k)$ and got

$$1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + x^{26} - \dots$$

There is a pattern here. After the initial terms, there are two negative terms and then two positive terms and then two negative terms and $\dots$ Moreover, the difference between the exponents of consecutive terms with the same sign is 1, 2, 3, 4, 5, $\dots$ and the difference between the exponents of consecutive terms with different signs is 3, 5, 7, 9, 11, $\dots$ The fact that this product has a sum with this pattern is called *Euler's Pentagonal Number Theorem*.

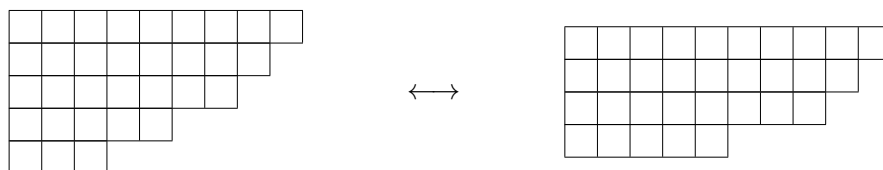
Euler tried to prove this theorem for many years. He first mentioned it in a letter to Daniel Bernoulli written in 1741. Euler writes “The other problem, to transform $(1 - x)(1 - x^2)(1 - x^3) \dots$ into $1 - x - x^2 + x^5 + x^7 - \dots$ follows easily by induction, if one multiplied many factors. The remainder of the series I do not see. This can be shown in a most pleasant investigation, together with tranquil pastime and the endurance of pertinacious labor, all three of which I lack.” He mentions the result several more times in letters to other mathematicians. Learning that d’Alembert wanted to leave mathematical research to regain his health, he wrote him “If in your spare time you should wish to do some research which does not require much effort, I will take the liberty to propose the expression $(1 - x)(1 - x^2)(1 - x^3) \dots$ etc., which upon expansion by multiplication gives the series $1 - x - x^2 + x^5 + x^7 - \dots$, which would seem very remarkable to me because of the law which we easily discover within it, but I do not see how his law may be deduced without induction of the proposed expression.” Eventually d’Alembert wrote back “regarding the series of which you have spoken, it is very peculiar, but I only see induction to show it. But no one is deeper and better versed on such matters than you.”

Finally after ten years, Euler found a proof. The proof is complicated. Later in the 1820’s the result came up naturally in the theory of elliptic functions. In 1881, the American mathematician Franklin gave a proof which involves no algebra at all. Hans Rademacher called this proof “the first major achievement of American mathematics.” We’ll give Franklin’s proof next.

The argument we used to get the p_n from $\prod_{k \geq 1} (1 + x^k + x^{2k} + x^{3k} + \dots)$ works similarly to give a combinatorial meaning to the coefficients of $\prod_{k \geq 1} (1 - x^k)$. This time only one power of x is involved in each term of the product, so we are talking about partitions of n in which each number occurs at most once. For instance, $5 = 5, 4 + 1, 3 + 2$. The minus sign in the terms of the product forces us to count partitions with an odd number of terms negatively and partitions with an even number of terms positively.

Let us concentrate on the fact that most of the time our product has zero coefficients. So most of the time, a number n has the same number of partitions with an even number of distinct summands as there are partitions with an odd number of distinct summands.

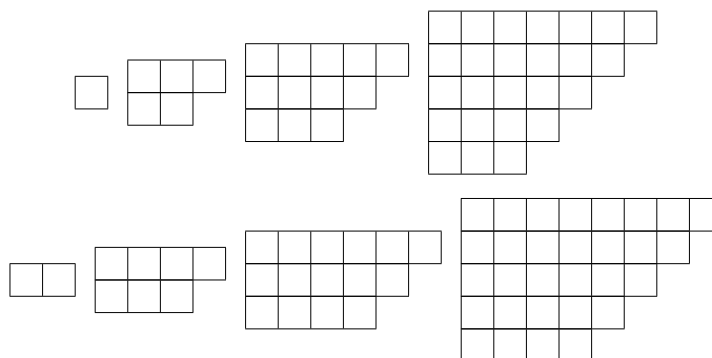
Franklin showed that this conclusion is true because there is a way to associate each partition with an even number of distinct summands with a unique corresponding partition with an odd number of distinct summands. This method is best illustrated with Young diagrams:



In this picture, the bottom row on the left has been moved to the ends of the first three rows of the diagram on the right, giving a new final diagonal.

Notice that in the left picture we can make this move but not its reverse. Indeed, if we move the final diagonal on the left to a new row on the bottom, we will have two rows with the same length, but our partitions have distinct summands. Similarly in the right picture we can move the diagonal down, but we cannot move the bottom row over because it is too long. A little thought shows that this principle will hold in general: given a diagram, only one operation is possible. If the number of elements in the final row is smaller than the number of elements in the final diagonal, we can move that row up but cannot move the diagonal down. If it is larger, we can move the diagonal down but cannot move the row up. So every partition with an odd number of distinct summands is associated with exactly one partition with an even number of distinct summands.

However, there are cases when neither move is possible; these cases occur when the final row and the final diagonal share an element. The reader can show that this happens exactly when our diagram belongs to one of two sequences shown below, and the reader can verify that neither move is possible in each of these cases.



The number of boxes on the top line is 1, 5, 12, 25, ... and the number of boxes on the bottom is 2, 7, 15, 30, ... and these are exactly the exponents where there are terms in the resulting product, terms with alternating signs because the number of rows increases by one in each of these sequences. QED.

Given this information, it is easy to write a program which will compute as many p_n as desired. The only caveat is that these numbers rapidly grow large, so it is important to use a program or library which can handle very large integers.

Here is a Mathematica program which does the task, for example:

```
FPentagonal[limit_] :=
Block[{N, f, P, Pinverse, k, index, n, i}, (* local variables *)
N = limit + 1; (* number of series coefficients *)
f[s_] := 0;
P = Array[f, N]; (* the p(n), initially filled with zeros *)
P[[1]] = 1; (* p(0) = p(1) = 1; arrays in Mathematica are one-based *) P[[2]] = 1;
Pinverse = Array[f, N]; (* inverse of p(n) series *)
k = 1; (* now fill in Pinverse using the Pentagonal Number Theorem *)
index = k (3 k - 1) / 2;
While[index <= N,
  Pinverse[[index]] = (-1)^k;
  index = k (3 k + 1) / 2;
  If[index <= N, Pinverse[[index]] = (-1)^k]; k = k + 1;
  index = k (3 k - 1) / 2;
];
For[n = 2, n < N, n = n + 1, (* compute p(n) as inverse of Pinverse *)
  P[[n]] = 0;
  For[i = 1, i < n, i = i + 1,
    P[[n]] = P[[n]] - P[[n - i]] Pinverse[[i]]
  ];
  P[[n]] = P[[n]] - Pinverse[[n]];
  Print[n, ": ", P[[n]]];
]
]
```

I don't know how many values of p_n were computed by Euler. In 1918, MacMahon in England computed the first 200 values of p_n . This table was extended to 600 by Gupta in 1935, and to 1000 by Gupta, Gwyther and Miller in 1958. I don't know if a computer was used for this final table.

The last value computed by MacMahon was $p_{200} = 3972999029388$. My Macbook Pro Portable Computer took less than five minutes to compute the first 10,000 numbers, and the number of irreducible representations of $S_{10,000}$ is exactly

$$p_{10000} = 36,167,251,325,636,293,988,820,471,890,$$

$$953,695,495,016,030,339,315,650,422,081,$$

868, 605, 887, 952, 568, 754, 066, 420, 592, 310,
556, 052, 906, 916, 435, 144

This is considerably more than the number of atoms in the universe, but in this case the items have been counted precisely, so that final digit is definitely 4 rather than 5.

I hope the reader saw the movie *The Man Who Knew Infinity*, starring Dev Patel as Ramanujan, with Jeremy Irons as G. H. Hardy, Toby Jones as Littlewood, and Kevin McNally as Major MacMahon. After MacMahon computed the first 200 values of p_n , Ramanujan pointed out that $p_4 = 5$ and after that every fifth entry in the table is divisible by 5. Also $p_5 = 7$ and after that, every seventh entry is divisible by 7. Finally $p_6 = 11$ and after that every eleventh entry is divisible by 11. Eventually Ramanujan was able to prove these facts; the proofs are not easy.

Ramanujan also had an estimate for the size of p_n :

$$p_n \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{2n}{3}}}$$

The movie shows MacMahon's rejection of this claim, and then the proof by Hardy and Ramanujan that it is a much better approximation than expected.

In 2006, Ken Ono proved that whenever N is an integer whose prime factorization does not contain 2 or 3, then there are integers a and b such that $p_{am+b} \equiv 0 \pmod{N}$. This generalizes Ramanujan's observation to all primes except 2 and 3; as far as I know, their status is still unknown. However, the new congruences are nowhere near as simple as Ramanujan's. For instance

$$\begin{aligned} p_{17303m+237} &\equiv 0 \pmod{13} \\ p_{40837937m+1122838} &\equiv 0 \pmod{17} \end{aligned}$$

24.3 The Sign of a Permutation

Definition 49 A transposition is a permutation which interchanges two integers and leaves all other integers fixed.

Examples $\sigma = (37)$ and $\tau = (12)$.

Theorem 124 Every permutation can be written as a product of transpositions.

Proof by induction on n : Suppose σ takes n to k . Then $(k \ n) \circ \sigma$ fixes n and thus is a permutation of $(1, 2, \dots, n-1)$. So it can be written as a product of transpositions, and σ is this product times the inverse of $(k \ n)$. This inverse is actually $(k \ n)$.

Theorem 125 Either all representations of a fixed permutation as a product of transitions have an even number of transpositions, or else all have an odd number of transpositions.

Proof: Suppose a permutation maps $(1, 2, \dots, n)$ to $(k_1, k_2, \dots, k_n)$. For each pair $i < j$, either k_i and k_j are in order or else they are out of order. Count the number of out of order pairs, and keep track of whether this count is even or odd. Before applying any permutation, all pairs are in order and the sum is even.

Now suppose we interchange k_i and k_j and leave all other numbers fixed. We claim the sum will change parity. If so, we are clearly done. Look closely at the picture below:

$$(k_1, \dots, k_i, \dots, k_s, \dots, k_j, \dots, k_n)$$

Consider pairs involving k_1 . The only pairs that will change will be (k_1, k_i) and (k_1, k_j) , but they will just switch roles, so the sum and parity will not change. This argument works with every k left of k_i . The same argument works for k_n and more generally for any k to the right of k_j .

Next look at k_s . The only pairs that matter are (k_i, k_s) and (k_s, k_j) . If k_s is smaller than both k_i and k_j , then one pair was originally out of order and one was not. After the switch, the parity of each pair changes, but it is still true that one pair is out of order and one is not. If k_s is greater than both k_i and k_j , the same argument applies.

If k_s is larger than k_i but smaller than k_j , then originally both pairs are in order, and after the switch they are both out of order. So the sum of out of order terms increases by two, but the parity does not change.

If k_s is smaller than k_i but larger than k_j , then originally both pairs are out of order, and after the switch they are both in order. So the sum of out of order terms decreases by two, but the parity does not change.

Finally, consider the term (k_i, k_j) . This is the only case where the parity changes when we switch. So multiplying by a transposition always changes the parity of the sum of out of order terms. QED.

Definition 50 The sign of a permutation σ , $\text{sgn}(\sigma)$, is 1 if it can be written as a product of an even number of transpositions, and -1 if it can be written as a product of an odd number of transpositions.

Theorem 126 $\text{sgn} : \mathcal{S}_n \rightarrow \{\pm 1\}$ is a group homomorphism.

Proof: Trivial.

Remark: The sgn is an important property of permutations, which comes up in all sorts of unexpected places, from the definition of orientation and the explanation of left vs right to the 15 puzzle. Recall this puzzle, with 15 plastic numbered pieces and one blank spot. We are given the puzzle with the pieces randomly placed, and wish to move the pieces until they form rows numbered 1, 2, 3, 4 and 5, 6, 7, 8 and 9, 10, 11, 12 and 13, 14, 15, blank. This movement will be a permutation, and thus have sign ± 1 . But each legal movement

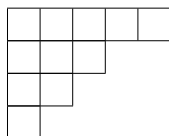
involves the blank, so we can keep track of the sign by just following the motion of the blank spot. Usually it starts in the bottom right and should be there at the end. This forces the permutation to have sign 1. So a puzzle starting in a random position with sign equal -1 cannot be solved.

Chapter 25

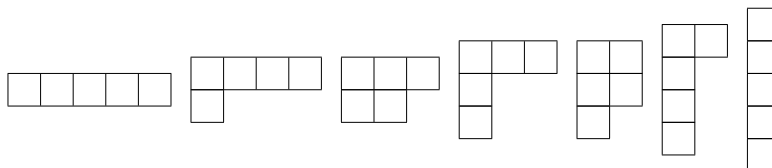
Irreducible Representations of $\mathcal{S}_n$

25.1 From Diagram to Representation

There is a graphical way to write a partition of n ; arrange n boxes on multiple lines as below. To draw the partition $5 + 3 + 2 + 1 = 11$, draw five boxes, and under them 3 boxes, and then 2 boxes and 1 box as shown. A diagram of this sort is called a Young Diagram. Each line must start at the left side; lines grow shorter as we go down the diagram, but two lines of the same length are allowed.



Here are all possible diagrams when $n = 5$:



In general there are p_n Young diagrams with n boxes, and so p_n irreducible representations of $\mathcal{S}_n$. The theorem that the number of conjugacy classes of a finite group equals the number of irreducible representations is proved by duality, not by associating a representation to each conjugacy class. So we should not expect to construct irreducible representations

from diagrams. But amazingly, there is such a construction in the special case of the symmetric group. In this chapter we will show how to directly construct an irreducible representation of $\mathcal{S}_n$ starting with a specific Young diagram. We begin with a few easy general results.

Definition 51 *Let G be a finite group and let K be a field. The group ring of G over K is the set of all formal linear combinations*

$$\sum k_i g_i$$

with $k_i \in K$ and $g_i \in G$. Such elements are added and multiplied in the obvious way, giving a (usually non-commutative) ring with unit $\mathcal{C}(G)$.

Remark: Suppose M is a left-module over the group ring. Then we can add elements of M and scalar multiply by elements of K , so M is a vector space over K . But in addition, we can scalar multiply by elements of G . If $g \in G$, let us write

$$g \cdot m = \rho(g)m$$

Then ρ is a representation of G on M . Conversely, if we are given ρ , we can define scalar multiplication by elements of G using the above formula, and extend linearly to obtain scalar multiplication by elements of $\mathcal{C}(G)$. So modules over the group ring are the same thing as representations of G .

There is an interesting special case of this. Suppose $c \in \mathcal{C}(G)$ is a fixed element. Then $\mathcal{C}(G)c$ is a subspace of the group ring, where we are deliberately multiplying on the right by c . As we will see, this set can be much smaller than all of $\mathcal{C}(G)$. Moreover, it is clearly invariant under multiplication from the left by scalars, and thus is a module over the group ring, or equivalently a representation of G .

We will show that all irreducible representations of the symmetric group can be obtained in this way. Let us begin by constructing two easy representations.

Suppose that $c = \sum_{g_i \in G} g_i$. Then $\mathcal{C}(G)c$ is one dimensional, because if we multiply on the left by a particular g , we obtain $\sum gg_i$ and the gg_i are just a rearrangement of the g_i . In particular, $\rho(g)c = c$, so the corresponding representation is the identity representation.

Suppose next G is a symmetric group and $c = \sum_{g_i \in G} \text{sgn}(g_i)g_i$. Then $\mathcal{C}(G)c$ is again one dimensional, because if we multiply on the left by a particular g , we obtain $\sum \text{sgn}(g_i)gg_i = \text{sgn}(g) \sum \text{sgn}(gg_i)gg_i$; this sum is just a rearrangement of c , so we obtain $\text{sgn}(g)c$. In particular $\rho(g)c = \text{sgn}(g)c$, so the corresponding representation of $\mathcal{S}_n$ is the sign representation.

Definition 52 Let λ be an arbitrary Young diagram with n boxes. Number the boxes as shown below.

$$\lambda = \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & & \\ \hline 7 & 8 & & \\ \hline 9 & & & \\ \hline \end{array}$$

Let A be the subgroup of $\mathcal{S}_n$ consisting of permutations which preserve the rows, and B be the subgroup of $\mathcal{S}_n$ of permutations which preserve the columns. Define

$$\begin{aligned} a_\lambda &= \sum_{\sigma \in A} \sigma \\ b_\lambda &= \sum_{\sigma \in B} \text{sgn}(\sigma) \sigma \\ c_\lambda &= a_\lambda b_\lambda \end{aligned}$$

Note that c_λ is an element of the group algebra $\mathcal{C}(\mathcal{S}_n)$.

Remark: We will prove that the resulting representation of $\mathcal{S}_n$ on $\mathcal{C}(\mathcal{S}_n)c_\lambda$ is irreducible. We will prove that distinct Young diagrams produce representations which are not isomorphic. Since the total number of Young diagrams is exactly the total number p_n of irreducible representations, it follows that the irreducible representations of $\mathcal{S}_n$ are exactly the above representations.

Remark: Our description leaves a lot of questions. It does not determine the dimensions of the representations. We will ultimately find a formula for these dimensions, but that will require substantial work. The description of the representations is very abstract, so it is not clear what the representations actually look like. We will find a little additional information, but not much.

Note that the Young diagram with only one row gives the identity representation, and the Young diagram with only one column gives the sign representation.

25.2 A Preliminary Observation

We are going to prove that the above representations are irreducible and give all irreducible representations of $\mathcal{S}_n$. We start with an amusing observation. A Young diagram with an arbitrary numbering of the boxes is called a *Young tableau*. The particular tableau obtained by numbering the first row $1, 2, \dots$, and continuing row by row is called the *canonical tableau*. When a permutation σ acts on a tableau, it acts on the numbers in the boxes, so if a box originally had k , then after the action it has $\sigma(k)$.

If we start with an arbitrary tableau, we would like to convert it to the canonical tableau by first permuting just columns, and then permuting just rows. Let's consider an example. Suppose we start with

9	5	7	4
6	3	2	
1			
8			

We begin by permuting the columns so the first row contains some permutation of 1, 2, 3, 4. In this case we can find permutations making the first row 1, 3, 2, 4. There are several ways to do this, but one gives

1	3	2	4
9	5	7	
6			
8			

We need a second row with 5, 6, 7 in some order, and we can do that by permuting the bottom three entries of the first column. This gives

1	3	2	4
6	5	7	
9			
8			

We need an 8 on the third row, and a permutation of the last two entries of the first column will do this, giving

1	3	2	4
6	5	7	
8			
9			

Now a row permutation in A will finish the job.

Remark: How could we get in trouble? Here's a bad starting position:

9	8	5	2
4	7	1	
6			
3			

The first column contains two numbers 3 and 4 that must end up in the first row, and no column permutation can make that happen.

Lemma 50 *Suppose T is a Young tableau and T_0 is the canonical tableau with the same Young diagram. Then T can be changed to T_0 by a series of column permutations, followed by a series of row permutations, if and only if no column of the diagram T contains two numbers which are on the same row of the canonical tableau.*

Proof: The condition is obviously necessary. If it holds, then no two entries on the top row of the canonical tableau can be in the same column of T by assumption, and the number of columns is the number of boxes on that top row. So we can permute the columns to make the top row contain the correct numbers in some order. By induction we can continue until all rows contain the correct numbers in some order. Now permute the rows to finish. QED.

Remark: Let us rewrite this as an assertion about the subgroups A and B associated with the canonical tableau T_0 . Let g be the permutation which maps the canonical tableau to T . Then the column operations on T each have the form gbg^{-1} where b is a column operation for T_0 , and thus the product of these operations has the same form. In the end, we obtain a tableau which only differs from T_0 in the order of its rows. So $gbg^{-1}T = aT_0$. Since $T = gT_0$, we have $gbg^{-1}gT_0 = aT_0$ and so $gbg^{-1}g = a$ or $gb = a$ or $g = ab^{-1}$. Thus we have proved

Lemma 51 *Suppose T is a Young tableau and T_0 is the canonical tableau with the same Young diagram. Let g be the permutation which maps T_0 to T . Then g can be written in the form ab for $a \in A$ and $b \in B$ if and only if no column of T contains two numbers which came from the same row of T_0 .*

25.3 Key Lemmas

If $a \in A$, then $aa_\lambda = a_\lambda$ because $a \sum_{\sigma \in A} \sigma = \sum (a\sigma)$, a rearrangement of the original sum. Similarly if $b \in B$, $b_\lambda b = \text{sgn}(b)b_\lambda$. Consequently $ac_\lambda b = \text{sgn}(b)c_\lambda$.

Lemma 52 *Conversely, if c is an element of the group algebra such that $acb = \text{sgn}(b)c$ whenever $a \in A$ and $b \in B$, then c is a multiple of c_λ .*

Proof: Notice first that $A \cap B = \{e\}$ and consequently an element $\sigma \in \mathcal{S}_n$ can be written in the form ab with $a \in A$ and $b \in B$ in at most one way.

Suppose $c = \sum n_g g$ satisfies the hypothesis of the lemma. It suffices to prove that $n_g = 0$ whenever g is not ab for $a \in A$ and $b \in B$. Indeed if c has this property, then $c = \sum_{a \in A, b \in B} c_{a,b} \text{sgn}(b)ab$ where the $c_{a,b}$ are numbers depending on a and b . As a runs over P , so does pa . As b runs over Q , so does bq . So our formula for c can be written $\sum_{a \in P, b \in Q} c_{pa,bq} \text{sgn}(bq)(pa)(bq)$. Rewriting slightly, we get $\text{sgn}(q)p \left(\sum_{a \in P, b \in Q} c_{pa,bq} \text{sgn}(b)ab \right) q$. But we are assuming that $c = \text{sgn}(q)pcq$, so $c_{pa,bq} = c_{a,b}$ for all p and q , and therefore the $c_{a,b}$ are constants independent of indices.

Suppose g is a permutation not of the form ab . Form the Young tableaux T whose boxes are numbered $g(1), g(2), g(3), \dots, g(n)$ by numbering row by row. Then the permutation g maps the canonical tableau to T . Since g is not of the form ab , a previous lemma says that some column of T has two numbers which are in the same row of the canonical tableau. Let

t be the permutation which fixes all numbers except these two. Then t is a row permutation of the canonical tableau, but a column permutation of T . This column permutation of T thus has the form gbg^{-1} where b is a column operation of T_0 . Note that b must also be a transposition. So $t = gbg^{-1}$ or $tgb^{-1} = g$. However $tc b^{-1} = \text{sgn}(b^{-1})c = -c$. Since g does not change sign under this operation, g cannot be one of the terms in c . QED.

Lemma 53 *Whenever $c \in \mathcal{C}(\mathcal{S}_n)$, $c_\lambda \cdot c \cdot c_\lambda$ is a multiple of c_λ . In particular $c_\lambda \cdot c_\lambda = n_\lambda c_\lambda$ for some complex n_λ .*

Proof: This follows immediately from lemma 3. Indeed if $a \in A$ and $b \in B$,

$$a(c_\lambda \cdot c \cdot c_\lambda)b = \text{sgn}(b)(c_\lambda \cdot c \cdot c_\lambda)$$

so the term in parentheses is a multiple of c_λ .

25.4 Irreducibility

Theorem 127 *If λ is a Young diagram, the representation of $\mathcal{S}_n$ on $\mathcal{C}c_\lambda$ is irreducible.*

Proof: Write $V_\lambda = \mathcal{C}c_\lambda$. We show first that this is not the zero space. Since the subgroups A and B satisfy $A \cap B = \{e\}$, every element $\sigma = ab$ in c_λ has a unique such expression. In particular, the identity occurs once with coefficient one in c_λ . So $c_\lambda \in V_\lambda$ is not zero.

If $W \subset V_\lambda$ is any subspace, then $c_\lambda W \subset c_\lambda V_\lambda = c_\lambda \mathcal{C}(\mathcal{S}_n)c_\lambda$ and by the previous lemma, each element of this space is a multiple of c_λ . So $c_\lambda W$ has dimension zero or one. If the dimension of this space is one, then $c_\lambda \in c_\lambda W$. Suppose in addition that W is invariant. Then $V_\lambda = \mathcal{C}c_\lambda \subset W$ and so $W = V_\lambda$.

The other possibility is that $c_\lambda W = 0$. In that case

$$W \cdot W \subset (\mathcal{C}(\mathcal{S}_n)c_\lambda)W = \mathcal{C}(\mathcal{S}_n)(c_\lambda W) = 0$$

We will show that this implies that $W = \{0\}$.

Since W is an invariant subspace of the representation space V_λ , it is an invariant subspace of the group algebra $\mathcal{C}$. We can find a complementary invariant subspace U by Maschke's theorem. Hence $\mathcal{C} = W \oplus U$ where both are invariant subspaces. The identity element can thus be written $e = w + u$. Multiplying both sides by an arbitrary $w_1 \in W$ gives $w_1 = w_1 \cdot w + w_1 \cdot u = 0 + w_1 \cdot u \in U$ and thus $W \subset U$. Since we have a direct sum, $W = \{0\}$. QED.

Corollary 12 *If λ is a Young diagram, $c_\lambda V_\lambda \neq \{0\}$.*

Proof: This follows from the proof of the previous result.

25.5 Replacing the Canonical Tableau with Another Tableau

We defined the Young symmetrizer associated with a canonical Young tableau. Actually we could have used any Young tableau. This would change the subgroups A and B , and thus the elements $a_\lambda, b_\lambda, c_\lambda$. For example, consider the Young diagram below. Its canonical tableau is T_0 and T is another tableau.

$$\lambda = \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \quad T_0 = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & & \\ \hline 7 & 8 & & \\ \hline 9 & & & \\ \hline \end{array} \quad T = \begin{array}{|c|c|c|c|} \hline 4 & 7 & 9 & 2 \\ \hline 1 & 3 & & \\ \hline 8 & 5 & & \\ \hline 6 & & & \\ \hline \end{array}$$

Note that $(1\ 9\ 7\ 5)$ is a column permutation in B for T_0 , but not in the B associated with T . On the other hand, $(1\ 4\ 6\ 8)$ is in the B associated to T .

Let g be the permutation mapping T_0 to T . In the above example, g maps 1 to 4, 2 to 7, and so on. As usual, let A and B denote the subgroups for T_0 . Then column permutations for T have the form gbg^{-1} where $b \in B$ and row permutations for T have the form gag^{-1} where $a \in A$. So the Young symmetrizer associated with T is

$$c_T = (ga_\lambda g^{-1})(gb_\lambda g^{-1}) = gc_\lambda g^{-1}$$

Thus

$$V_T = \mathcal{C}(\mathcal{S}_n)c_T = \mathcal{C}(\mathcal{S}_n)gc_\lambda g^{-1} = \left(g\mathcal{C}(\mathcal{S}_n)g^{-1}\right)gc_\lambda g^{-1} = g\left(\mathcal{C}(\mathcal{S}_n)c_\lambda\right)g^{-1} = gV_\lambda g^{-1}$$

We then get a commutative diagram

$$\begin{array}{ccc} V_\lambda & \xrightarrow{\sigma} & V_\lambda \\ gvg^{-1} \downarrow & & \downarrow gvg^{-1} \\ V_T & \xrightarrow{g\sigma g^{-1}} & V_T \end{array}$$

because $g(\sigma v)g^{-1} = (g\sigma g^{-1})(gvg^{-1})$. So the representation $\rho(\sigma)$ at the top has become the representation $g\rho(\sigma)g^{-1}$ at the bottom and the irreducible representations on V_λ and V_T are equivalent. In fact, we can decompose $\mathcal{C}(\mathcal{S}_n)$ into a sum of irreducible representation spaces, each of the form V_T for some Young tableau T . We get all irreducible representations of $\mathcal{S}_n$ in this way, and each irreducible representation occurs as many times in the decomposition as its dimension.

Definition 53 We order Young diagrams with the same number of boxes by writing $\lambda < \mu$ if λ has a shorter first row, or if their first rows have equal length and λ has a shorter second row, and so forth.

We will need one fact about this ordering.

Lemma 54 *Suppose λ and μ are Young tableau with the same number of boxes and $\lambda < \mu$. The symmetric group acts on both tableau. There is a transposition t in this group which is a column permutation of λ , but a row permutation of μ .*

Proof: If the first row of μ is longer than the first row of λ , then two elements of the first row of μ must be in the same column of λ and the result is clear. If these two rows have the same length, but two elements of the first row of μ are in the same column of λ , the result is also clear. If the top rows have the same length and each element of the top row of μ is in a different column of λ , we can apply column transformations to λ so λ and μ have the same first row up to a row permutation. Remove this row from both tableau and from now on restrict attention to permutations which fix all elements in this row. Repeat the argument for lower rows until a transposition is found.

25.6 More Lemmas

From now on, we use the canonical tableau to construct $a_\lambda, b_\lambda, c_\lambda$.

Lemma 55 *Let $\mathcal{C}$ be the group algebra of the symmetric group. Then*

- *If $\lambda < \mu$, then $b_\lambda \cdot \mathcal{C} \cdot a_\mu = 0$.*
- *If $\lambda < \mu$, then $c_\lambda \cdot \mathcal{C} \cdot c_\mu = 0$. In particular, $c_\lambda \cdot c_\mu = 0$.*
- *In fact, the conclusions of the previous item hold whenever $\lambda \neq \mu$.*

Proof: We first issue a warning. If λ and μ are different diagrams, they have different A and B subgroups, so we must make certain not to choose, say, $a \in A$ and then apply it to both c_λ and c_μ .

To prove the first item, it suffices to consider the case when $\mathcal{C}$ is replaced by a single $g \in \mathcal{S}_n$. Let T_0 be the canonical tableau for the Young diagram μ and let T be the corresponding diagram obtained by applying g to each entry of T_0 . If a_μ is the symmetrizer element for μ using the canonical tableau, then the corresponding element using T is $ga_\mu g^{-1}$. Moreover, $b_\lambda ga_\mu = 0$ if and only if $b_\lambda ga_\mu g^{-1} = 0$. So to prove the first item, we can ignore g and simply assume that the μ comes from an arbitrary tableau rather than the canonical tableau.

Since $\lambda < \mu$, there is a transposition t which is a column operation in λ , but a row operation in μ . Then

$$b_\lambda a_\mu = b_\lambda t^2 a_\mu = (b_\lambda t)(ta_\mu) = (-b_\lambda)(a_\mu) = -b_\lambda a_\mu$$

The second item then follows because if $c \in \mathcal{C}(\mathcal{S}_n)$, then

$$c_\lambda \cdot c \cdot c_\mu = a_\lambda \cdot b_\lambda \cdot c \cdot a_\mu \cdot b_\mu = a_\lambda (b_\lambda \cdot c \cdot a_\mu) b_\mu$$

To prove the third result, notice that the map $\sigma \rightarrow \sigma^{-1}$ from $\mathcal{S}_n$ to itself induces a map $\star : \mathcal{C} \rightarrow \mathcal{C}$ satisfying $c_1^\star \circ c_2^\star = (c_2 \circ c_1)^\star$. Notice also that $a_\lambda^\star = a_\lambda$ and $b_\lambda^\star = b_\lambda$. We will prove that $c_\lambda \mathcal{C} c_\mu = 0$ by showing that $\star$ of this element is zero. Indeed

$$(c_\lambda \mathcal{C} c_\mu)^\star = b_\mu^\star a_\mu^\star c_\lambda^\star b_\lambda^\star a_\lambda^\star = b_\mu (a_\mu c^\star b_\lambda) a_\lambda = 0$$

This time we do not claim that the item in parentheses is zero, but only that it is an element of $\mathcal{C}$, so this expression has the form $b_\mu c a_\lambda$ and vanishes by the first part of the lemma. QED

25.7 Different Diagrams Give Inequivalent Representations

Theorem 128 *Suppose λ and μ are distinct Young diagrams with the same number of boxes. By previous results, each generates an irreducible representation of $\mathcal{S}_n$. These representations are not equivalent.*

Proof: Finally, we prove V_λ and V_μ inequivalent. Without loss of generality, assume $\lambda < \mu$. Note that $c_\lambda V_\lambda \neq 0$ by the corollary of the theorem that V_λ is irreducible. But

$$c_\lambda V_\mu = c_\lambda \mathcal{C} c_\mu = 0$$

by the previous lemma.

On the other hand, if the representations were equivalent, then we could find a linear isomorphism $\varphi : V_\lambda \rightarrow V_\mu$ such that when $g \in G$ and $v \in V_\lambda$, $gv = \varphi^{-1}g\varphi(v)$. This result would extend to linear combinations of group elements and thus $c_\lambda v = \varphi^{-1}c_\lambda\varphi(v)$. But the right side is identically zero and the left side is not. QED.

Corollary 13 *All irreducible representations of $\mathcal{S}_n$ are given by Young diagrams and thus equivalent to a representation on*

$$V_\lambda = \mathcal{C} c_\lambda$$

Proof: The number of irreducible representations of $\mathcal{S}_n$ is p_n , the partition number of n . This is also the number of Young diagrams with n boxes, and thus the number of inequivalent irreducible representations V_λ .

25.8 A Formula for $c_\lambda \cdot c_\lambda$

If λ is a Young diagram, we earlier proved that $c_\lambda \cdot c_\lambda$ is a multiple of c_λ . We now prove

Theorem 129 *If λ is a Young diagram, $c_\lambda \cdot c_\lambda = n_\lambda c_\lambda$ where*

$$n_\lambda = \frac{n!}{\dim V_\lambda}$$

Proof: Consider the map $f(v) = v \cdot c_\lambda : \mathcal{C} \rightarrow \mathcal{C}$. The trace of this map is $n!$ because the coefficient of e in c_λ is 1 and this element leaves all basis vectors of $\mathcal{C}$ fixed, while every other group element in c_λ moves elements around.

Let $W \subset V_\lambda$ be the set of all elements $v \in V_\lambda$ such that $v \cdot c_\lambda = 0$. This is an invariant subspace of V_λ , so it is either $\{0\}$ or else V_λ . If $c_\lambda^2 = 0$, then $c_\lambda \in W$ so W is not the zero space and thus $W = V_\lambda$. The image of f would then belong to W , but $f = 0$ on W . So the trace of f would be zero. This contradiction shows that $c_\lambda^2 \neq 0$ and thus $c_\lambda \notin W$. So W cannot be V_λ and must instead be $\{0\}$.

Therefore $V_\lambda \cap \ker(f) = 0$. On the other hand, for every other μ , $V_\mu \subset \ker(f)$ because $\mathcal{C}c_\mu \cdot c_\lambda = 0$. So $\mathcal{C} = V_\lambda \oplus \ker(f)$. On V_λ , $f(a \cdot c_\lambda) = a \cdot c_\lambda^2 = n_\lambda \cdot a \cdot c_\lambda$ and thus it is constantly n_λ . So the trace of f is $n_\lambda \dim V_\lambda$, but also $n!$. Therefore, $n_\lambda = \frac{n!}{\dim V_\lambda}$.

25.9 Interchanging a_λ and b_λ

Theorem 130 *Let $\mathcal{C}$ be the group algebra for S_n . Then the left $\mathcal{C}$ modules $\mathcal{C}a_\lambda b_\lambda$ and $\mathcal{C}b_\lambda a_\lambda$ are isomorphic, and in particular define isomorphic representations of S_n . Thus the representation space V_λ can be defined as $\mathcal{C}c_\lambda$ where c_λ can be either $a_\lambda b_\lambda$ or $b_\lambda a_\lambda$.*

Proof: Define the maps below, where each map is right multiplication by the indicated element:

$$\mathcal{C}a_\lambda b_\lambda \xrightarrow{\frac{a_\lambda}{\sqrt{n_\lambda}}} \mathcal{C}b_\lambda a_\lambda$$

$$\mathcal{C}b_\lambda a_\lambda \xrightarrow{\frac{b_\lambda}{\sqrt{n_\lambda}}} \mathcal{C}a_\lambda b_\lambda$$

Notice that each map's image is in the indicated space. If $c \in \mathcal{C}$, the composition of these maps sends

$$ca_\lambda b_\lambda \rightarrow c \frac{a_\lambda b_\lambda a_\lambda b_\lambda}{n_\lambda} = ca_\lambda b_\lambda$$

It follows that the first map is one-to-one and the second map is onto. Applying $\star$ to the identity $(a_\lambda b_\lambda)(a_\lambda b_\lambda) = n_\lambda a_\lambda b_\lambda$ gives $(b_\lambda a_\lambda)(b_\lambda a_\lambda) = n_\lambda b_\lambda a_\lambda$. Therefore, composing in the other order is also the identity, and the second map is one-to-one and the first map is onto. In short, both maps are isomorphisms. QED.

Chapter 26

Specht Modules

26.1 Introduction

We would like to find the dimensions of the irreducible representations of the symmetric group. This is difficult using the previous construction. Given λ , we would have to construct c_λ , which has many terms, and then multiply it by all possible $g \in \mathcal{S}_n$, giving $n!$ expressions of the form gc_λ . Then we would have to perform row operations on these expressions to obtain a maximal linearly independent subset. Not pleasant for any g and not practical for all but very small n . Specht introduced a different but equivalent construction of this representation space with the important advantage that we can easily determine a basis for his space.

26.2 Specht Modules

Let λ be a Young diagram. Form the group algebra $\mathcal{C}(\mathcal{S}_n)$; elements of this space are linear combinations of permutations. If g is a permutation and T_0 is the standard tableau associated with λ , then $T = gT_0$ is another tableau associated with λ . Hence we can imagine that elements of the group algebra are just linear combinations of Young tableau.

For instance, when $\lambda = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}$, the group algebra contains linear combinations of

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array}$$

We call two such Young diagrams equivalent if they have the same numbers in the rows up to rearrangement of the items in that row. In the above list, the first two two diagrams are equivalent, as are the third and fourth, and the fifth and sixth. Each equivalence class

is determined by any of its elements, and we might as well select the element with all rows increasing. Let $\mathcal{D}$ be the set of all linear combinations of equivalence classes. Equivalently, $\mathcal{D}$ is the set of linear combinations of Young tableau with increasing rows. For instance, if

$\lambda = \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array}$ then elements of $\mathcal{D}$ are

$$a_1 \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} + a_2 \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + a_3 \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array}$$

We get a natural map $\mathcal{C} \rightarrow \mathcal{D}$ by mapping each Young tableau to the equivalent tableau with increasing rows. So

$$\begin{aligned} & b_1 \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} + b_2 \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} + b_3 \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + b_4 \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array} + b_5 \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} + b_6 \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} \rightarrow \\ & (b_1 + b_2) \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} + (b_3 + b_4) \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + (b_5 + b_6) \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \end{aligned}$$

If two tableau T_1 and T_2 are equivalent and we apply a permutation σ to both, the new tableau $\sigma(T_1)$ and $\sigma(T_2)$ are also equivalent. Indeed if we pick any row, say the top row, then T_1 and T_2 have the same numbers $k_1, \dots, k_s$ on this row up to reordering, so the numbers $\sigma(k_1), \dots, \sigma(k_s)$ are also the same up to reordering.

Therefore $\mathcal{S}_n$ acts on $\mathcal{D}$ and the map $\mathcal{C} \rightarrow \mathcal{D}$ commutes with the action of $\mathcal{S}_n$ on these spaces.

We also define a map $\mathcal{D} \rightarrow \mathcal{C}$ by sending each equivalence class of tableau to the sum of all tableau in the equivalence class. Thus when $\lambda = \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array}$ we sent both $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ and $\begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array}$ to $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ and now we send $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ to $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array}$. It is easy to see that this map also commutes with the action of $\mathcal{S}_n$ on the two spaces.

The composition $\mathcal{D} \rightarrow \mathcal{C} \rightarrow \mathcal{D}$ can be simply described, for if we take a tableau T with all rows ordered and map it to the sum of all equivalent tableau, and then map each element of this sum back to the original tableau T by reordering the rows, we get the original element multiplied by the number of elements in the equivalence class. If T_0 is the canonical tableau associated to λ , then $T = \sigma T_0$ and permutations of the rows have the form $\sigma \circ a \circ \sigma^{-1}$ where $a \in A$, the row group of T_0 . So the composition $\mathcal{D} \rightarrow \mathcal{C} \rightarrow \mathcal{D}$ is multiplication by $|A|$.

It is not so easy to describe the composition $\mathcal{C} \rightarrow \mathcal{D} \rightarrow \mathcal{C}$. It is certainly not multiplication by a constant since these spaces have different dimensions. We study this composition later.

Suppose that T is a given tableau and let B be the group of column permutations of T . This definitely depends on the specific T ; equivalent tableau will generate very different groups B . Define

$$c_T = \sum_{b \in B} \text{sgn}(b) b(T)$$

and define

$$e_T = \text{element induced by } c_T \text{ in } \mathcal{D}$$

These elements depend on the specific tableau T ; equivalent T produce very different c_T and e_T . For example, consider the equivalent tableau

$$T_1 = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} \text{ and } T_2 = \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} \quad c_{T_1} = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} \text{ and } c_{T_2} = \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array}$$

The first terms of these expressions are different but equivalent, but the second terms aren't even equivalent.

Let $W_\lambda \subset \mathcal{D}$ be the subspace generated by the e_T as T varies over all Young tableau with shape λ . Our goal is to prove the following theorem:

Theorem 131 *Suppose λ is a Young diagram.*

- *The space W_λ is invariant under $\mathcal{S}_n$.*
- *The resulting representation on W_λ is equivalent to the representation associated earlier to λ using Young symmetrizers.*
- *Consequently W_λ is irreducible, and distinct Young diagrams produce inequivalent representations.*
- *A Young tableau is called standard if the numbers in boxes increase along both rows and columns. The e_T for standard tableau T form a basis for W_λ .*
- *Consequently the dimension of the irreducible representation associated with λ equals the number of standard tableau with shape λ .*

Example: There are two standard tableau with shape $\begin{array}{|c|c|} \hline & \\ \hline & \\ \hline \end{array}$, namely $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$ and $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}$. So the associated irreducible representation of $\mathcal{S}_3$ has dimension two. In fact, $\mathcal{S}_3$ is the full symmetry group of an equilateral triangle, and the symmetries of such a triangle give this irreducible representation.

Let us give more details for this special case. There are six tableau associated to λ , and so we obtain the following c_T and e_T for these elements:

$$\begin{array}{lcl}
\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \\
\\
\begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} \\
\\
\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \\
\\
\begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} \\
\\
\begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} \\
\\
\begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} & c_T = \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} & e_T = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}
\end{array}$$

The first and last elements on the right differ in sign, the second and fourth differ in sign, and the third and fifth differ in sign. Therefore W_λ is generated by the first three terms on the right. But the first minus the third is the second. So the space is actually generated by the first and third terms, which are clearly independent. And these terms come from the two standard tableau on the extreme left.

Proof: The proof requires several steps and thus is spread over a number of pages.

Lemma 56 *Given $\sigma \in \mathcal{S}_n$ and a Young tableau T , $\sigma(c_T) = c_{\sigma T}$.*

Proof: Let T_0 be the canonical tableau so that $T = \sigma_1 T_0$. Then column operations in T have the form $\sigma_1 \circ b \circ \sigma_1^{-1}$ where b is a column operation in T_0 . So

$$\sigma c_T = \sum \text{sgn}(b) \sigma (\sigma_1 \circ b \circ \sigma_1^{-1}) \sigma_1 T_0 = \sum \text{sgn}(b) \sigma \sigma_1 b T_0$$

and

$$c_{\sigma T} = c_{\sigma \sigma_1 T_0} = \sum \text{sgn}(b) (\sigma \sigma_1) b (\sigma \sigma_1)^{-1} \sigma \sigma_1 T_0 = \sum \text{sgn}(b) (\sigma \sigma_1) b T_0$$

QED.

Corollary 14 *Given $\sigma \in \mathcal{S}_n$ and a Young tableau T , $\sigma(e_T) = e_{\sigma T}$. Hence W_λ is invariant under $\mathcal{S}_n$.*

Remark: Our next goal is to prove that the Specht representation on $\mathcal{W}_\lambda$ is equivalent to the earlier representation on $\mathcal{C}c_\lambda$. To show this, let us calculate the effect of the map $\mathcal{C} \rightarrow \mathcal{D} \rightarrow \mathcal{C}$ on an element of the form c_T . This is equivalent to considering the effect of

the map $\mathcal{D} \rightarrow \mathcal{C}$ on elements of the form e_T . Since these elements generate $\mathcal{W}_\lambda$, it is also equivalent to determining the image of $\mathcal{W}_\lambda$ under the map $\mathcal{W}_\lambda \subset \mathcal{D} \rightarrow \mathcal{C}$.

Let us start with a single Young tableau T . We can find a unique $\sigma \in \mathcal{S}_n$ such that $T = \sigma T_0$. Then c_T is a linear combination of Young tableaux obtained by applying column operations to T . Each such operation has the form $\sigma \circ b \circ \sigma^{-1}$ where b is a column operation of T_0 . So c_T is a sum of terms of the form $\text{sgn}(b) (\sigma \circ b \circ \sigma^{-1}) \sigma T_0 = \text{sgn}(b) \sigma b T_0$. The element $e_T \in \mathcal{D}$ is the same sum of Young diagrams, but now each is an equivalence class. If we map e_T back to $\mathcal{C}$, we map each of the Young equivalence classes to the sum of all elements in the class. The elements equivalent to $\sigma b T_0$ all have the form $\left((\sigma b) \circ a \circ (\sigma b)^{-1} \right) (\sigma b) T_0 = \sigma b a T_0$ where $a \in A$. So the final element we obtain from c_T in $\mathcal{C}$ is

$$\sum_{b \in B, a \in A} \text{sgn}(b) \sigma b a T_0 = \sigma b_\lambda a_\lambda T_0 = \sigma b_\lambda a_\lambda$$

Recall that the representations of $\mathcal{S}_n$ on $\mathcal{C}a_\lambda b_\lambda$ and $\mathcal{C}b_\lambda a_\lambda$ are equivalent. The end result of our calculation is that if we start with a permutation σ and the equivalent Young tableau $T = \sigma T_0$, the map $\mathcal{W}_\lambda \rightarrow \mathcal{C}$ sends e_T to $\sigma(b_\lambda a_\lambda)$. Since the e_T generate $\mathcal{W}_\lambda$ and the $\sigma(b_\lambda a_\lambda)$ generate the representation space determined by λ , this map between representation spaces is onto. On the other hand, it is one-to-one because the map

$$\mathcal{D} \rightarrow \mathcal{C} \rightarrow \mathcal{D}$$

is multiplication by a non-zero constant $|A|$. So we have proved the second and third part of our theorem.

Remark: Our claim about the dimension of the Specht module will be proved in later sections.

Remark: We have described the irreducible representations of the symmetric group in two different ways. The original description asked us to compute $c_\lambda = b_\lambda a_\lambda$. Then each permutation σ generates σc_λ and these elements generate the representation space. The complication is that these generators are not linearly independent.

The alternate approach thinks of permutations as Young tableau. The representation space contains linear combinations of these tableau, just as the group algebra contains linear combinations of permutations. But in the new representation space, we are allowed to permute elements on rows of Young diagrams without changing the final element. In both cases, the generators of the representation space come from permutations σ . But this time σ induces $T = \sigma T_0$, and the generator is e_T . As earlier, these generators are not linearly independent.

Our isomorphism between the two descriptions sends e_T for $T = \sigma T_0$ to σc_λ .

26.3 A Lower Bound for the Dimension of Representations

Definition 54 A Young tableau is called standard if the numbers in the tableau increase along each row and along each column.

Example 1: The standard tableau associated with  are

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array}$$

Example 2: There are five Young diagrams with four boxes; the standard tableau for these diagrams are listed below:

$$\begin{array}{cccc} \begin{array}{|c|c|c|c|} \hline & & & \\ \hline \end{array} & \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline \end{array} & & \\ \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline \end{array} & \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & & \\ \hline \end{array} & \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 3 & & \\ \hline \end{array} & \begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 2 & & \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline & \\ \hline \end{array} & \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array} & \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & 4 \\ \hline \end{array} & \\ \begin{array}{|c|c|} \hline & \\ \hline & \\ \hline & \\ \hline \end{array} & \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline 4 & \\ \hline \end{array} & \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline 4 & \\ \hline \end{array} & \begin{array}{|c|c|} \hline 1 & 4 \\ \hline 2 & \\ \hline 3 & \\ \hline \end{array} \\ \begin{array}{|c|} \hline \\ \hline \\ \hline \\ \hline \end{array} & \begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline 4 \\ \hline \end{array} & & \end{array}$$

Theorem 132 Let λ be a Young diagram. Each Young tableau associated with this diagram creates an element e_T of the Specht module. The elements created by standard Young tableau are linearly independent. Therefore the dimension of $\mathcal{W}_\lambda$ is at least the number of standard Young tableau with diagram λ .

Proof: Put a total order on the set of equivalence classes of Young tableau. Note that $\{T_1\} = \{T_2\}$ if every i is in the same row of T_1 and T_2 . Otherwise there is an i which is in different rows of the two diagrams. Find the largest such i . We say $\{T_1\} < \{T_2\}$ if this i is in a higher row of T_1 than its row in T_2 .

For example, notice that

$$\begin{array}{|c|c|c|} \hline 3 & 4 & 5 \\ \hline 1 & 2 & \\ \hline \end{array} < \begin{array}{|c|c|c|} \hline 2 & 4 & 5 \\ \hline 1 & 3 & \\ \hline \end{array} < \begin{array}{|c|c|c|} \hline 1 & 4 & 5 \\ \hline 2 & 3 & \\ \hline \end{array}$$

This relation clearly depends only on the equivalence classes of the tableau. Any two items can be compared, for start with the highest i and work down until you find an i in different rows. If no such i exists, the equivalence classes are equal.

Finally the relation is transitive, for if $\{T_1\} < \{T_2\}$ and $\{T_2\} < \{T_3\}$, then $\{T_1\} < \{T_3\}$; indeed start with the highest i and work down until i is in different rows of T_1 and T_2 , or of T_2 and T_3 . Then there are three possibilities: either i is in the same rows of T_1 and T_2 but higher in T_2 than T_3 , or else it is higher in T_1 than T_2 but in the same rows of T_2 and T_3 , or finally it is higher in T_1 than in T_2 and higher in T_2 than T_3 .

Suppose that T is a λ -tableau in which entries increase down all columns. Let e_T be the element in D_λ generated by this tableau. We claim

$$e_T = \{T\} \pm \{T_1\} \pm \dots \pm \{T_m\}$$

where $\{T_i\} < \{T\}$ for all i . Each T_i arises from T by a column permutation σ . If σ is not the identity, then there is an i such that $\sigma(i) \neq i$ and $\sigma(j) = j$ for $i < j$. Then i and $\sigma(i)$ are in the same column and all lower elements in that column are fixed by σ so that $\sigma(i)$ is higher than i . Therefore $\{T_i\} < \{T\}$.

Suppose that a linear combination $\sum \alpha_T e_T$ is zero in $\mathcal{D}$ where not all α are zero and each T is a standard tableau. Consider the set of equivalence classes $\{T\}$ with nonzero α_T and pick the maximal class among them using our ordering. Note that this class is unique because we have a total order and because two standard tableau cannot belong to the same equivalence class. Indeed if they did, then they would differ by row permutations, but each of these permutations would preserve order and thus be the identity. By the previous paragraph, the resulting linear combination of e_T contains a unique term $\alpha_T \{T\}$. But this is impossible because the various equivalence classes are linearly independent in $\mathcal{D}$. QED.

Chapter 27

The Robinson-Schensted Correspondence

If λ is a Young diagram, let f_λ be the number of standard associated Young tableau. We proved in the previous section that the dimension of the irreducible representation $V_\lambda = \mathcal{C}b_\lambda a_\lambda$ of $\mathcal{S}_n$ associated with λ is at least f_λ .

The goal of this chapter is to prove combinatorially that $\sum f_\lambda^2 = n!$. From representation theory, we know that for any finite group G with irreducible representations V_λ , we have $\sum (\dim V_\lambda)^2 = |G|$. Putting these results together, we obtain

Theorem 133 *If λ is a Young diagram, the e_T formed from standard Young tableau T form a basis for the Specht space. Consequently, the dimension of the irreducible representation associated with λ is exactly f_λ .*

27.1 History of the Robinson-Schensted Correspondence

We will prove that $\sum f_\lambda^2 = n!$ using a wonderful correspondence introduced by Schensted. Given a permutation of n letters σ , Schensted constructs two standard Young Tableau P and Q with identical diagrams. In this way, he obtains a map from the set of all permutations $\mathcal{S}_n$ to the set of all pairs of standard Young tableau with identical diagrams and n boxes. Moreover, he also constructs an inverse map going the other direction. So the number of permutations $n!$ equals the number of pairs of standard Young tableau $\sum_\lambda f_\lambda^2$.

This correspondence was introduced by Schensted in 1961. It later turned out that Robinson had invented a different procedure, now almost forgotten, that accomplished the same result in 1938. The procedure was then generalized by Knuth in 1970. The procedure is

now known as the Robinson-Schensted Correspondence or the Robinson-Schensted-Knuth Correspondence.

Incidentally, one of my sources writes

Schenstead is a weird man. He's a physicist, without a faculty position, who makes his money by inventing board games. In 1995, he changed his first name from Craige to Ea, because people kept forgetting the 'e' on the end of his name. Ea is the Babylonian name for the Sumerian god Enki. He then changed it again to Ea Ea.

This is neither here nor there, but the Robinson-Schensted Correspondence is a key component of combinatorial Young theory, with many beautiful consequences I won't have time to mention.

The Robinson-Schensted Correspondence comes up very early in volume 3 of *The Art of Computer Programming*. On pages 51 and 52, Knuth states and proves the Robinson-Schensted Correspondence by giving precise algorithms for going both ways in the correspondence.

Why is the Robinson-Schensted Correspondence described in a book about computer science? Because standard Young tableaux provide an important data structure to store and search for information. In applications, the data often dribbles in, rather than being provided all at once, and an algorithm is needed to insert new items into the data structure. That is essentially what Schensted's algorithm does. It constructs a Young diagram and fills it with data box by box, making sure that the data is always ordered along rows and along columns. Consequently it is easy to search for data, even though the data structure changes with time.

In some applications it is important to recover the data in the order that it was generated. In mathematical terms, we'd like to take the Young diagram and reconstruct the permutation which generated it. Schensted discovered that this can be done if additional information is saved during the construction of the original Young diagram. The additional data needed is merely the order in which the boxes of the Young diagram were created. An easy way to record this data is to create a second Young diagram with the same shape as the first one, this time filling each box with the number of the step which created it. This gives two standard Young tableaux with the same shape, called P and Q . We call P the "insertion tableau" because that is where the data is stored, and we call Q the "recording tableau" because that explains the order in which the data was stored. Given these two standard tableau, we will find a separate algorithm to go backward and reconstruct the permutation σ .

A geometrical way of looking at the Correspondence was found by Viennot. The web site

https://www.youtube.com/watch?v=K_qhmS07giQ&feature=youtu.be&t=585 has a lecture on Robinson-Schensted by Viennot himself, which very clearly explains the algorithm. This is the first of a series of wonderful combinatorial lectures on the site. I recommend this lecture, but with a warning, because the French write Young diagrams upside down.

The Robinson-Schensted Correspondence is a pleasant way to end the theory of representations of $\mathcal{S}_n$, since it fills an important gap in the theory and yet is proved by just describing two algorithms.

27.2 Proof of the Robinson-Schensted Correspondence

The Insertion Algorithm: We'll describe the insertion algorithm with words, and then with examples. Suppose we have a standard Young tableau and want to insert a new number k . In the existing tableau, the numbers increase along rows and along columns, but are not necessarily consecutive numbers, so the tableau might contain 4, 5, 7, 11, 13, and 17. The number k will be placed somewhere on the first row. If it is larger than any number currently in this first row, it will be placed in a new box at the end of the row. Otherwise we find the first number in this row larger than k . Call this number k_1 . Remove it from its box and place k in this box.

Now we have a new boxless number k_1 . We insert this number in the diagram exactly as we inserted k , except that we ignore the first row and place it somewhere in the second row. If it is greater than everything in the second row, we place it at the end of the row. Otherwise we find the earliest number on the second row which is greater than k_1 , call that number k_2 , remove that number from its box and replace it with k_1 . We place this k_2 somewhere in the third row using the same rule. Continue.

We'll use the example from Viennot's lecture to explain the correspondence. Consider the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 1 & 6 & 10 & 2 & 5 & 8 & 4 & 9 & 7 \end{pmatrix}$$

This permutation will generate the pair of standard Young tableau below:

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 7 & 9 \\ \hline 3 & 5 & 8 & & \\ \hline 6 & 10 & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & 9 \\ \hline 2 & 5 & 6 & & \\ \hline 8 & 10 & & & \\ \hline \end{array}$$

Our construction below suggests that the diagram already exists and we only need fill it with numbers. But in fact we are creating the diagram as the algorithm proceeds. So imagine that only the boxes with numbers exist in the pictures.

We map the permutation to the tableau as follows. The permutation begins with 3. Map this to the first possible place in P and put a 1 in this place in Q to indicate that we

created that box first.

$$P = \begin{array}{|c|c|c|c|c|} \hline 3 & & & & \\ \hline & & & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & & & & \\ \hline & & & & \\ \hline & & & & \\ \hline \end{array}$$

Now we want to insert the number 1. The number 3 in this row is larger than 1, so we replace the 3 by 1 in the top box. Now we have to place 3 somewhere on the second row. This row is empty, so we place it in the left-most position.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & & & & \\ \hline 3 & & & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & & & & \\ \hline 2 & & & & \\ \hline & & & & \\ \hline \end{array}$$

The next number is 6 and we can put it on the first row without any bumping.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 6 & & & \\ \hline 3 & & & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & & & \\ \hline 2 & & & & \\ \hline & & & & \\ \hline \end{array}$$

The next number is 10 and that can also be put on the first row without bumping.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 6 & 10 & & \\ \hline 3 & & & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & & \\ \hline 2 & & & & \\ \hline & & & & \\ \hline \end{array}$$

The next symbol is 2. The earliest number on the first row larger than 2 is 6, so we replace 6 with 2 and now must insert 6 somewhere in the second row. In fact it can go at the end of the second line and we get

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 10 & & \\ \hline 3 & 6 & & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & & \\ \hline 2 & 5 & & & \\ \hline & & & & \\ \hline \end{array}$$

The next number is 5. It replaces the 10 on the first row, and that 10 can go at the end of the second row.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 5 & & \\ \hline 3 & 6 & 10 & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & & \\ \hline 2 & 5 & 6 & & \\ \hline & & & & \\ \hline \end{array}$$

The next symbol is 8

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 5 & 8 & \\ \hline 3 & 6 & 10 & & \\ \hline & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & \\ \hline 2 & 5 & 6 & & \\ \hline & & & & \\ \hline \end{array}$$

The next symbol is 4. It replaces 5 on the top line. This 5 replaces 6 on the second line. This 6 falls to the third line.

$$P = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 4 & 8 \\ \hline 3 & 5 & 10 & \\ \hline 6 & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|} \hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 & 6 & \\ \hline 8 & & & \\ \hline \end{array}$$

The next symbol is 9. It ends the top line.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 8 & 9 \\ \hline 3 & 5 & 10 & & \\ \hline 6 & & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & 9 \\ \hline 2 & 5 & 6 & & \\ \hline 8 & & & & \\ \hline \end{array}$$

The final symbol is 7. It replaces 8 on the top line. This 8 replaces 10 on the second line. The 10 ends the third line.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 7 & 9 \\ \hline 3 & 5 & 8 & & \\ \hline 6 & 10 & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & 9 \\ \hline 2 & 5 & 6 & & \\ \hline 8 & 10 & & & \\ \hline \end{array}$$

There are a few questions about this algorithm, but we discuss them after giving the inverse algorithm.

The Inverse Algorithm:

Suppose we have a permutation σ , say

We construct the permutation from the diagrams P and Q as follows. Consult Q to find the last box added to the diagram. Find the same box in P and suppose it contains k . This k came from the largest element smaller than k on the previous line. Call this element k_1 . Replace k_1 by k and erase the contents of the box which originally contained k . This frees k_1 , which we will bump up the diagram.

Look at the line one up from the line now containing k . Find the largest element in this line that is smaller than k_1 . Call it k_2 . Replace k_2 by k_1 in its box, freeing k_2 to bump to a higher line.

Continue this process until an element is bumped out of the first line. This newly freed element is $\sigma(n)$ if the diagrams have n boxes.

Repeat the process for $n-1, n-2, \dots, 1$. Notice that the tableau P changes during the algorithm, but Q always remains the same.

Let us try this algorithm using the P and Q constructed earlier.

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 7 & 9 \\ \hline 3 & 5 & 8 & & \\ \hline 6 & 10 & & & \\ \hline \end{array} \quad Q = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 7 & 9 \\ \hline 2 & 5 & 6 & & \\ \hline 8 & 10 & & & \\ \hline \end{array}$$

We first compute $\sigma(10)$. The diagram Q shows that this step created the last element of the bottom line, which contains 10. We erase this 10. The largest element on the previous line smaller than 10 is 8. We replace 8 by 10 on this line. The largest element on the previous line smaller than 8 is 7. We replace 7 by 8 on this line. We conclude that $\sigma(10) = 7$ and replace P by

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 8 & 9 \\ \hline 3 & 5 & 10 & & \\ \hline 6 & & & & \\ \hline \end{array}$$

Next we compute $\sigma(9)$. We find 9 in Q and that box is the last box on the first line. So we erase 9 from this box, report that $\sigma(9) = 9$, and obtain a new P diagram

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 8 & \\ \hline 3 & 5 & 10 & & \\ \hline 6 & & & & \\ \hline \end{array}$$

Now we compute $\sigma(8)$. We find 8 in Q and discover that it is the first box on the bottom line. In P this box contains 6. So we erase 6 from the box and find the greatest element on the previous line smaller than 6. It is the box containing 5. We replace the 5 by 6, freeing up 5 to float upward. We find the greatest element on the top line smaller than 5. That box contains 4. We replace 4 by 5, freeing up 4. So $\sigma(8) = 4$. The new P is

$$P = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 5 & 8 & \\ \hline 3 & 6 & 10 & & \\ \hline & & & & \\ \hline \end{array}$$

At this point we glance back at our calculation of P and Q from σ and discover that we are reproducing the P 's computed there in reverse order. So the inverse algorithm is working as desired.

Proving that Robinson-Schensted Works in General

There are a couple of points that need to be checked. In the insertion algorithm we sometimes add a box to the right side of an existing row. We must check that this row never grows longer than the row above it, guaranteeing that we always get a genuine Young diagram.

We must also check that at each stage of the insertion algorithm, numbers in rows increase to the right and numbers in columns increase to the bottom.

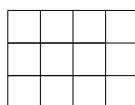
In the inverse algorithm, there are also things to check. We need to show that after erasing the contents of a box, the resulting non-empty boxes still form a Young diagram. And we need to show that after each stage, the new contents increase in both rows and columns.

Finally, we need to show that the two operations are inverse to each other.

One of these points for the inverse algorithm is very easy. We erase boxes in reverse order of the numbering of Q . Since Q is a standard Young tableau, the highest number is in a corner of the diagram. Erasing a corner always yields another Young tableau. We erase boxes in P in the same order, so P always remains legal.

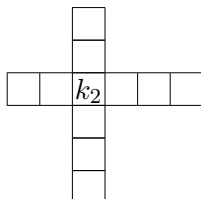
Most of the remaining tasks can be left to the reader, but just to be sure we will check the points which show that the insertion algorithm never runs into trouble.

First we show that adding boxes never creates a long row under a short row. Consider for example the following situation and imagine that the right box on the bottom row is about to be created. How could that occur?



We would be inserting k into the diagram. It couldn't add a box to the top line because then the process would stop. So it replaced a box containing k_1 on the top line, and that k_1 was pushed down to the second line. If the algorithm created a new box at the end of the second row then the process would stop. So it must have replaced a box containing k_2 on the second line with k_1 and this k_2 was pushed down to the third line, where it will now create a box at the end of the line. So k_2 is larger than any number on the third line. But k_2 was originally in the second line and the columns originally increased while going down. So this is impossible.

Finally we show that each step in the construction of P makes a tableau in which numbers increase along both rows and columns. The following picture will be useful for this argument. It shows the crucial step when k_2 in a row is replaced by k_1 dropping from a higher row. Only boxes in the row and column of k_2 are shown:



Since k_2 is the first number in the row larger than k_1 , and since no two numbers are equal, all numbers to the left of this box are smaller than k_1 . Moreover the numbers in the row were increasing before switching k_1 and k_2 , so all numbers in the row to the right of k_2 are larger than k_2 and thus certainly larger than k_1 . So if we replace k_2 with k_1 , the numbers in the row will still increase.

Similarly, the numbers in the column below k_2 are larger than k_2 , which is larger than k_1 , so that portion of the column will increase as we move down even if we replace k_2 by k_1 .

Finally, we will show that the numbers above k_2 must be smaller than k_1 and thus the whole column will increase after the switch. It clearly suffices to show that the number directly above k_2 is smaller than k_1 .

To see this, recall that k_1 used to be in the previous row until it was knocked out by an earlier step. Now k_1 is trying to find a place in the main row shown, knocking out k_2 .

Where was k_1 when it was in the previous row? If it was right of the column displayed above, then since numbers in its row increase as we go right, the number right above k_2 must have been smaller than k_1 , and that is what we are trying to show. So k_1 used to be directly above k_2 or further left. But if we go left of k_2 in its row, we know that the numbers are smaller than k_1 and if we go up, they get smaller still, so none of these numbers could have been k_1 . Therefore, k_1 must have been directly above k_2 . But it was replaced by something smaller when k_1 was knocked out to wander, so now something smaller than k_1 is directly above k_2 . QED.

Chapter 28

The Hook Formula

28.1 The Hook Formula and the Dimension of V_λ

Each box in a Young diagram forms a *hook* which extends to the right or down from the box. The hook length of the box is the number of boxes in this hook. Thus it is the number of boxes to the right or below the box, including the box itself. For the Young diagram on the left below, the hook lengths are indicated in the right half of the picture:

					7	6	4	2	1
					4	3	1		
					2	1			

Theorem 134 (Hook Formula) *The dimension of the representation space V_λ attached to a Young diagram with n boxes is*

$$\frac{n!}{\prod_{\text{boxes}} (\text{hook length})}$$

Example For the diagram displayed above, the dimension of V_λ will be

$$\frac{10!}{7 \cdot 6 \cdot 4 \cdot 2 \cdot 4 \cdot 3 \cdot 2} = 450$$

Remark: This formula has an interesting history. Classically, an entirely different formula was known for the dimension of V_λ . The hook formula was found in 1953 by Frame, Robinson, and Thrall. According to an article in Wikipedia quoting Sagan,

One Thursday in May of 1953, Robinson was visiting Frame at Michigan State University. Discussing the work of Staal (a student of Robinson), Frame was

led to conjecture the hook formula. At first Robinson could not believe that such a simple formula existed, but after trying some examples he became convinced, and together they proved the identity. On Saturday they went to the University of Michigan, where Frame presented their new result after a lecture by Robinson. This surprised Thrall, who was in the audience, because he had just proved the same result on the same day!

The original proofs of the Hook formula reduced it to the earlier known formula for $\dim V_\lambda$, but later Donald Knuth suggested a heuristic argument using probability theory.

Notice that a Young tableau is standard if and only if each box contains a number which is smallest in the numbers attached to its hook. There are $n!$ Young tableau of a given shape λ , and f_λ of these tableau are standard. So if we insert numbers at random into a diagram, the probability that we get a standard tableau is

$$\frac{f_\lambda}{n!}$$

On the other hand, when we are filling a given hook, the probability that the vertex will get the smallest number is

$$\frac{1}{\text{hook length}}$$

and therefore the probability that every vertex will get the smallest number is

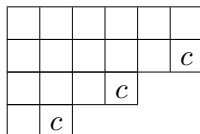
$$\prod_{\text{boxes}} \frac{1}{\text{hook length}}$$

So

$$\frac{f_\lambda}{n!} = \prod_{\text{boxes}} \frac{1}{\text{hook length}}$$

and the formula follows. Sadly, the events considered on the right side of this formula are not independent, so this is not a rigorous proof.

Proof of the Hook Formula: Later, in 1979, Greene, Nijenhuis, and Wilf found a rigorous probabilistic proof along these lines. Here is their proof. Given a Young diagram, a *corner* of the diagram is a box that is at the end of its row and simultaneously at the end of its column. For example, the corners of the following diagram are indicated with the letter “c”.



When numbers are put in the boxes to get a standard tableau, the 1 must be put in the top left position and the n must be put in one of the corners. So we get all possible standard

tableaux by removing a corner, filling the numbers $1, 2, \dots, n-1$ in the remaining boxes to obtain a standard tableaux, and then putting the number n in the box which was removed. It follows that

$$f_\lambda = \sum_{\text{corners } c} f_{\lambda,c}$$

where $f_{\lambda,c}$ is the number of standard tableaux in the shape λ after removing the corner c .

Now the goal is to show that the hook formula satisfies exactly the same equation. If so, we have an easy proof that the hook formula counts standard Young diagrams by induction on the number of boxes in the diagram.

Greene, Nijenhuis, and Wilf rewrite this formula as follows:

$$1 = \sum_{\text{corners } c} \frac{f_{\lambda,c}}{f_\lambda}$$

This formula begins to look like probability theory. To amplify this similarity, Green, Nijenhuis, and Wilf invent a game related to the formula.

Select a spot in the Young diagram λ at random. Think of this spot as the vertex of a hook and move to one of the other boxes in this hook at random, where each non-vertex move is equally likely. Continue moving in the same way. The game ends when we come to a corner. We are going to show that the probability that a game ends at a corner c is

$$\frac{f_{\lambda,c}}{f_\lambda}$$

where this time we calculate the f values using the hook formula rather than by counting standard tableaux. Since the game certainly ends at some corner, it will follow that

$$1 = \sum_{\text{corners } c} \frac{f_{\lambda,c}}{f_\lambda}$$

and the proof will be complete.

We begin by calculating $\frac{f_{\lambda,c}}{f_\lambda}$ assuming the f values are given by the hook formula. From now on we symbolize corners by their two coordinates (α, β) and we symbolize hook numbers by their two coordinates h_{ij} . Then

$$F = \frac{n!}{\prod h_{ij}}$$

$$F_{\alpha,\beta} = \frac{(n-1)!}{\prod_{(i,j) \neq (\alpha,\beta)} (h_{ij} \text{ or } h_{ij} - 1)}$$

and

$$\frac{F_{\alpha,\beta}}{F} = \frac{1}{n} \prod_{1 \leq i < \alpha} \frac{h_{i\beta}}{h_{i\beta} - 1} \prod_{1 \leq j < \beta} \frac{h_{\alpha j}}{h_{\alpha j} - 1} = \frac{1}{n} \prod_{1 \leq i < \alpha} \left(1 + \frac{1}{h_{i\beta} - 1}\right) \prod_{1 \leq j < \beta} \left(1 + \frac{1}{h_{\alpha j} - 1}\right)$$

Note that if $\alpha = 1$ there are no terms in the first product; in that case we take the product to be 1; similarly for the second product.

Now we consider our game more closely. A random action in the game starts at a box (a, b) and moves as follows:

$$(a, b) = (A_1, B_1) \rightarrow (A_2, B_2) \rightarrow \dots \rightarrow (A_m, B_m) = (\alpha, \beta)$$

Each individual move changes only one of the two coordinates. To better understand this feature, let us list the distinct first coordinates in order, forming a list

$$A : a = a_1, a_2, \dots, a_s = \alpha$$

and let us list the distinct second coordinates in order, forming a list

$$B : b = b_1, b_2, \dots, b_t = \beta$$

Let $P(A, B, a, b)$ be the probability that a trial starting at (a, b) has column list A , and row list B ; notice that such a trial ends at (α, β) . This symbol depends on (a, b) because we will later vary a and b , but we always deal with the same fixed corner (α, β) . We claim that

$$P(A, B, a, b) = \prod_{i \in A, i \neq \alpha} \frac{1}{h_{i\beta} - 1} \cdot \prod_{j \in B, j \neq \beta} \frac{1}{h_{\alpha j} - 1}$$

Assume this result is true. Then the probability that a game ends at the corner (α, β) is the sum over possible starting points (a, b) of the probability of starting at that point times the probability that a path from that point will end at (α, β) . Each such path has a list of columns A and a list of rows B where A is a sublist of $\{a, a+1, \dots, \alpha\}$ and B is a sublist of $\{b, b+1, \dots, \beta\}$. The total probability that an experiment starting at (a, b) gets to the corner (α, β) is the sum of all products $P(A, B, a, b)$ over sublist pairs A, B . So the probability that an experiment ends at the corner (α, β) is

$$\frac{1}{n} \sum_{(a,b) \in \lambda} \sum_{A, B} P(A, B, a, b)$$

and thus

$$\frac{1}{n} \sum_{(a,b) \in \lambda} \sum_{A, B} \prod_{i \in A, i \neq \alpha} \frac{1}{h_{i\beta} - 1} \prod_{j \in B, j \neq \beta} \frac{1}{h_{\alpha j} - 1}$$

A little thought shows that this is exactly the expression $\frac{F_{\alpha,\beta}}{F}$ studied at the start of the proof. For example, when the product

$$\prod_{1 \leq i < \alpha} \left(1 + \frac{1}{h_{i\beta} - 1} \right)$$

is expanded out, it will be a sum of products where each product is formed by taking either the 1 or the $\frac{1}{h_{i\beta}-1}$ from an individual term of the full product. Each of these subproducts has the form

$$\prod_{i \in A, i \neq \alpha} \frac{1}{h_{i\beta} - 1}$$

for a sublist A , and thus the full expansion is

$$\sum_A \prod_{i \in A, i \neq \alpha} \frac{1}{h_{i\beta} - 1}$$

Key Step of the Proof: The proof therefore reduces to showing that the following result is true.

$$P(A, B, a, b) = \prod_{i \in A, i \neq \alpha} \frac{1}{h_{i\beta} - 1} \cdot \prod_{j \in B, j \neq \beta} \frac{1}{h_{\alpha j} - 1}$$

Observe first that

$$P(A, B, a, b) = \frac{1}{h_{ab} - 1} \left[P(A - a_1, B, a_2, b_1) + P(A, B - b_1, a_1, b_2) \right]$$

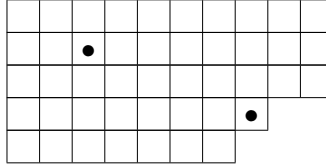
This equation says that any of the $h_{ab} - 1$ non-vertex boxes in the hook is equally likely for the first move, and that we either modify the first coordinate and leave the second coordinate alone, or we modify the second coordinate and leave the first coordinate alone. By induction $P(A - a_1, B, a_2, b_1)$ is our formula multiplied by $(h_{a_1\beta} - 1)$ to get rid of that starting term, and $P(A, B - b_1, a_1, b_2)$ is our formula multiplied by $(h_{\alpha b_1} - 1)$ to get rid of that starting term. Thus the formula involves proving that

$$\frac{(h_{a\beta} - 1) + (h_{\alpha b} - 1)}{h_{ab} - 1} = 1$$

or equivalently that

$$h_{ab} - 1 = (h_{a\beta} - 1) + (h_{\alpha b} - 1)$$

An example best illustrates why this is true. Roughly speaking, adding the two hooks on the bottom diagram gives the full hook from the dot in upper left.



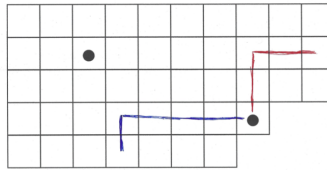


Figure 28.1: Sum of Two Hooks

28.2 Frightening Exercises

Volume 3 of Donald Knuth's *The Art of Computer Programming* is about Sorting and Searching. Very early in this book, Donald Knuth connects this subject with Young Diagrams and their combinatorial properties. He describes the Hook formula on page 60 and gives his heuristic proof there. He adds that no purely combinatorial proof of the Hook theorem was discovered until the work of Pak and Stoyanovskii in 1992. Exercise 39 asks the reader to reconstruct this work and prove the Hook theorem.

But the really frightening exercise may be exercise 38, which takes 15 lines of the book, sketches the argument of Greene, Nijenhuis and Wilf very briefly, and asks the reader to supply the details and use them to prove the Hook formula.

Chapter 29

Examples

29.1 $\mathcal{S}_3$

There are three Young diagrams with three boxes: $\begin{array}{|c|c|c|}\hline\Box & \Box & \Box \\ \hline\end{array}$, $\begin{array}{|c|c|}\hline\Box & \Box \\ \hline\Box & \\ \hline\end{array}$, $\begin{array}{|c|}\hline\Box \\ \hline\Box \\ \hline\Box \\ \hline\end{array}$. By the hook formula, the associated irreducible representations have dimensions 1, 2, 1. The first diagram gives the identity representation and the third gives the alternating representation. We are going to construct the two dimensional representation concretely.

The space $\mathcal{D}$ has dimension 3, with basis $\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \\ \hline\end{array}$, $\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \\ \hline\end{array}$, $\begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 & \\ \hline\end{array}$. In $\mathcal{D}$ these are representatives of equivalence classes, but the first two are also the standard Young tableau, producing a basis for the Specht module. It is convenient to change the sign of the second basis element and select the basis

$$e_1 = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \\ \hline\end{array} - \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 & \\ \hline\end{array} \quad e_2 = -\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \\ \hline\end{array} + \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 & \\ \hline\end{array}$$

The group $\mathcal{S}_3$ has a cyclic subgroup of order three containing the identity and $(1\ 2\ 3)$ and $(1\ 3\ 2)$, and then three transpositions $(1\ 2)$, $(1\ 3)$, $(2\ 3)$. A short calculation shows that $(1\ 2\ 3)$ maps

$$e_1 \rightarrow e_2 \quad e_2 \rightarrow -e_1 - e_2 \quad -e_1 - e_2 \rightarrow e_1$$

We easily find a basis in the plane that matches these results, and shows that the representation acting on the subgroup Z_3 generated by $(1\ 2\ 3)$ acts by rotations by 0, 120, and 240 degrees.

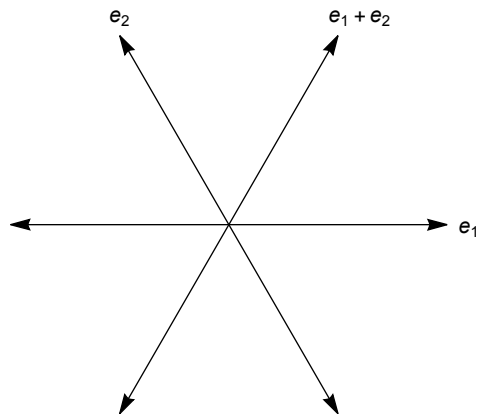


Figure 29.1: 2-dimensional representation

We must consider the action of the remaining transpositions. Consider the transposition $(1\ 3)$ and notice that it maps e_1 to $-e_1$. Indeed

$$\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \rightarrow \begin{array}{|c|c|} \hline 3 & 2 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} \sim \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array}$$

Also it maps $e_2 \rightarrow e_1 + e_2$ because

$$-\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \rightarrow -\begin{array}{|c|c|} \hline 3 & 1 \\ \hline 2 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline 2 & 1 \\ \hline 3 & \\ \hline \end{array} \sim -\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} = \left(\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \right) + \left(-\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 1 & \\ \hline \end{array} \right) = e_1 + e_2$$

So this transposition maps to reflection across the vertical dotted line below.

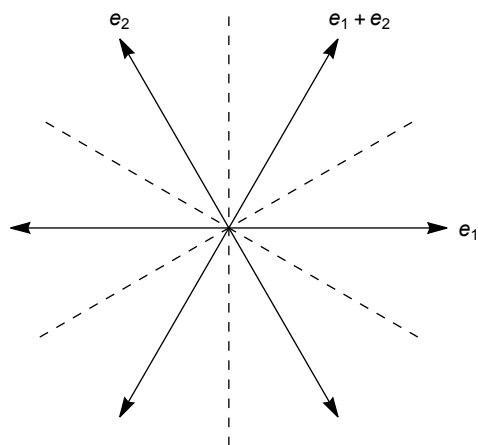


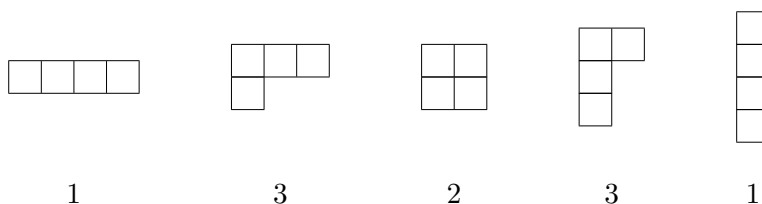
Figure 29.2: reflection lines

The other transpositions must map to reflections across the other dotted lines because $\mathcal{S}_3$ and its image in $GL(2, R)$ are both generated by the two rotations and the vertical reflection. We conclude that the Young diagram  gives the irreducible representation of $\mathcal{S}_3$ acting as the full group of symmetries of an equilateral triangle. But notice that the natural basis e_1, e_2 does not point to the vertices of this triangle. A fuller explanation will appear in the section on the tetrahedron.

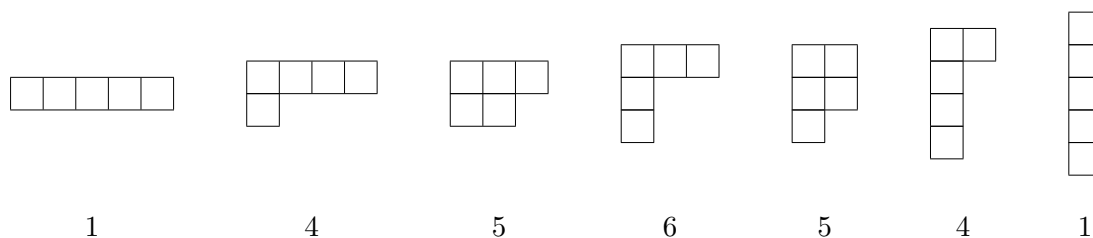
29.2 $\mathcal{S}_4$ and $\mathcal{S}_5$

These groups have orders 24 and 120, and $\mathcal{S}_5$ contains the first non-abelian simple group.

The Young diagrams for $\mathcal{S}_4$ and associated representation space dimensions are



The Young diagrams for $\mathcal{S}_5$ and associated representation space dimensions are



Notice that a Young diagram can be flipped over its diagonal to produce a second diagram, and in the above lists these two diagrams have the same dimension. There is a simple explanation for this symmetry. If $\sigma \rightarrow \rho(\sigma)$ is a representation, $\sigma \rightarrow \text{sgn}(\sigma)\rho(\sigma)$ is another representation. Repeating this operation twice returns to the original representation. Clearly the map preserves irreducibility.

Theorem 135 *If ρ is associated with a Young diagram, then $\text{sgn} \cdot \rho$ is associated with the Young diagram flipped over the diagonal.*

Proof: Let A be the group algebra. Define a map $\psi : A \rightarrow A$ by sending σ to $\text{sgn}(\sigma)\sigma$. This is an algebra isomorphism. Clearly $\psi(a_{P,\lambda}) = b_{P,\lambda}$ and $\psi(b_{Q,\lambda}) = a_{Q,\lambda}$ because ψ removes signs if they are present, and introduces signs if they are absent.

If we flip a Young diagram over the diagonal, the subgroups P and Q are reversed. Denote the corresponding subgroups of the flipped diagram by P_1, Q_1 . Then $\psi(a_P) = b_{Q_1}$ and $\psi(b_Q) = a_{P_1}$. Since $\psi(A) = A$, we have $\psi(Aa_Pb_Q) = Ab_{Q_1}a_{P_1}$.

If λ is the original Young diagram, $V_\lambda = Aa_Pb_Q$. If λ_1 is this diagram flipped over the diagonal, $V_{\lambda_1} = Aa_{P_1}b_{Q_1}$, but we earlier proved that this is isomorphic as a representation space to $Ab_{Q_1}a_{P_1}$. Hence ψ maps V_λ to V_{λ_1} . If $a = \sigma$ in A , the action of a on the first space is the same as the action of $\psi(a) = \text{sgn}(\sigma)\sigma$ on the second space. QED.

Remark: It remains to explain two representations of $\mathcal{S}_4$ and three representations of $\mathcal{S}_5$.

Consider first the representation for $\mathcal{S}_4$ associated with .

For a general group, there is an easy way to generate some irreducible representations: Find a normal subgroup H and an irreducible representation of G/H and map

$$G \rightarrow G/H \xrightarrow{\rho} GL(V)$$

This trick seldom works for $\mathcal{S}_n$ because the alternating group is simple, but it does work for $\mathcal{S}_4$ because the Klein subgroup $\mathcal{K} = \{e, (12)(34), (13)(24), (14)(23)\}$ is normal. Normality is easy because conjugation preserves the cycle structure and these are all possible products of two transpositions. To show it is a group, notice that in each element the two transpositions can be switched and the order in the transpositions can be switched. So it suffices to consider $(ab)(cd) \cdot (ac)(bd) = (ad)(bc)$. The group $\mathcal{S}_4/\mathcal{K}$ has 6 elements.

Notice that $(12)(123) = (23)$ and $(123)(12) = (13)$. These products are not equal in $\mathcal{S}_4/\mathcal{K}$ because $(23)(13)^{-1} = (23)(13) = (123)$ is not in $\mathcal{K}$. So the quotient group is not abelian, and the only non-abelian group of order 6 is $\mathcal{S}_3$. Since $\mathcal{S}_3$ has an irreducible 2-dimensional representation, we obtain an irreducible 2-dimensional representation of $\mathcal{S}_4$ and this must

be the representation associated with .

29.3 The Tetrahedron

We are about to describe an infinite number of irreducible representations of $\mathcal{S}_n$, one for each n . When $n = 3$, we will obtain the symmetries of an equilateral triangle, and for higher n we will obtain symmetries of tetrahedra of higher and higher dimensions. The basic idea is shown in the image at the top of the next page.

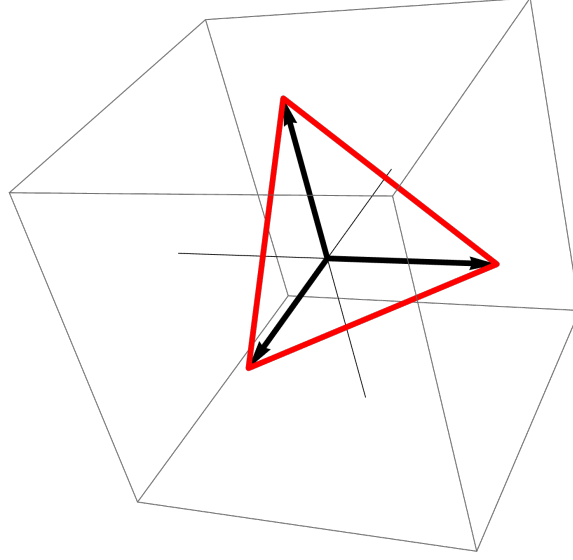


Figure 29.3: reflection lines

The symmetric group $\mathcal{S}_n$ acts on the basis vectors $e_1, e_2, \dots, e_n$ of $V = R^n$ or $V = C^n$ by permuting these vectors. Each permutation induces a linear transformation from V to V and we obtain a canonical representation of $\mathcal{S}_n$ on V . This representation is not irreducible; the subspace generated by $e_1 + e_2 + \dots + e_n$ is invariant, as is the subspace of all vectors $\sum c_i e_i$ with $\sum c_i = 0$. Thus $V = C \oplus W$, the sum of a one dimensional space and a space W of dimension $n - 1$. Both of these representations turn out to be irreducible.

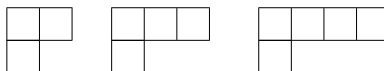
It is useful to form the hyperplane $P = \{v = \sum c_i e_i \mid \sum c_i = 1\}$. This hyperplane is not a subspace, of course. It contains all e_i . In the above picture it is the plane containing the red triangle. If we subtract $\frac{e_1 + \dots + e_n}{n}$ from each element of P , we obtain the subspace W . It follows that the plane P is invariant under the action of $\mathcal{S}_n$ and this representation is essentially the same as the representation on W .

The intersection of P with the main quadrant forms a generalized tetrahedron. It is a triangle when $n = 3$, an ordinary tetrahedron when $n = 4$, a 4-dimensional tetrahedron when $n = 5$, and so forth. The representation transformations of $\mathcal{S}_n$ are exactly the group of distance preserving symmetries of this general tetrahedron, whose vertices are the basis vectors e_i .

Note that in the case $n = 3$, the vertex $e_1 \in P$ is pushed back to $\frac{2}{3}e_1 - \frac{1}{3}e_2 - \frac{1}{3}e_3 \in W$, with similar formulas for the images of e_2 and e_3 . Two of these elements would have been better choices for the basis than the choice we picked in an earlier section.

We now show that the Young diagram which gives this tetrahedral representation for $\mathcal{S}_n$

has $n - 1$ boxes on the first row and 1 box on the second row. Consequently, this explains the following representations of $\mathcal{S}_3, \mathcal{S}_4$, and $\mathcal{S}_5$:



We will discuss the case $n = 5$ for concreteness, but everything we say generalizes. Notice that the dimension of $\mathcal{D}$ is $n = 5$ with basis

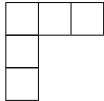
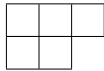
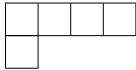
$$T_1 = \begin{array}{|c|c|c|c|c|} \hline 2 & 3 & 4 & 5 & \\ \hline 1 & & & & \\ \hline \end{array}, T_2 = \begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 5 & \\ \hline 2 & & & & \\ \hline \end{array}, T_3 = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 5 & \\ \hline 3 & & & & \\ \hline \end{array}, T_4 = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 3 & 5 & \\ \hline 4 & & & & \\ \hline \end{array}, T_5 = \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & \\ \hline 5 & & & & \\ \hline \end{array}$$

The last four of these are all the standard tableau, and provide basis vectors for the Specht module. For example, T_2 and T_3 yield

$$\begin{array}{|c|c|c|c|c|} \hline 1 & 3 & 4 & 5 & \\ \hline 2 & & & & \\ \hline \end{array} - \begin{array}{|c|c|c|c|c|} \hline 2 & 3 & 4 & 5 & \\ \hline 1 & & & & \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 4 & 5 & \\ \hline 3 & & & & \\ \hline \end{array} - \begin{array}{|c|c|c|c|c|} \hline 2 & 3 & 4 & 5 & \\ \hline 1 & & & & \\ \hline \end{array}$$

Notice that all four of these elements belong to the subspace $W \subset \mathcal{D}$. Since these four elements form a basis for the Specht module, and since the dimension of W is four, we conclude that the Specht module is exactly W . It follows that W is irreducible and corresponds to the Young diagram we are discussing. Thus that irreducible representation is essentially the representation on P and so the symmetries of the $n - 1$ dimensional tetrahedron. QED.

29.4 Remaining Cases for $\mathcal{S}_5$

There are two remaining Young diagrams in the case of $\mathcal{S}_5$:  and . These correspond to representations with dimensions 6 and 5. Recall that  corresponded to a hyperplane W acted on by symmetries of the 4-dimensional tetrahedron. It turns out that the first case above corresponds to $\Lambda^2(W)$. More generally, $\Lambda^k(W)$ is irreducible and has a Young diagram with n boxes, k in the first column, and the remaining $n - k$ boxes on the first row. Thus Young diagrams containing a single hook are understood.

The final representation, of dimension 5, is related to the dodecahedron and icosahedron, but the relationship is tricky. The alternating group A_5 is the group of rotational symmetries of these objects, but their full group of symmetries including reflections is not $\mathcal{S}_5$. Full details are available in other books.

29.5 Real and Complex Representations of $\mathcal{S}_n$

We have been applying results from representation theory which require a complex base field. But our pictures of representations, for instance those of the tetrahedra, are of real vector spaces. Luckily, the distinction does not matter for the symmetric groups. The representations we constructed from Young diagrams are defined and irreducible over either field. We proved that they are the only irreducible representations over C . It can then be shown that they are the only irreducible representations over R , by an argument we will summarize.

There is an elaborate theory of irreducible real representations, irreducible complex representations, and irreducible quaternionic representations of a finite group G . Each set can be divided into three disjoint subsets, called representations of *real type*, of *complex type*, and of *quaternionic type*. There is a one-to-one correspondence between the representations of real type over R , C , and H , given by starting with a real irreducible representation and noticing that the representation matrices are also complex matrices, and indeed also quaternionic matrices. There is also a one-to-one correspondence between the representations of quaternionic type over R , C , and H , given by starting with a quaternionic irreducible representation, and noticing that the quaternionic matrices can become complex matrices by replacing each q_{ij} with a 2×2 block of complex numbers, and then noticing that these complex matrices can become real matrices by replacing each c_{ij} with a 2×2 block of real numbers.

The representations of complex type over R , C , and H behave in a slightly more complicated fashion. Each real irreducible representation of complex type becomes a pair of complex irreducible representations of complex type, and this pair produces a single quaternion irreducible representation of complex type.

In the special case of $G = \mathcal{S}_n$, every irreducible complex representation comes from a Young diagram, and thus from a corresponding irreducible real representation given by the same Young diagram, so all complex irreducible representations are of real type. It follows that all irreducible real representations are also of real type, and thus in one-to-one correspondence with these complex irreducible representations. Therefore they are all associated with Young diagrams. QED.

Chapter 30

Induced Representations and The Alternating Group

Before returning to Lie Groups, we will find all irreducible representations of the alternating groups $\mathcal{A}_n$. We do this for several reasons. Since the $\mathcal{A}_5$ for $n \geq 5$ are one of the families of non-abelian finite simple groups, it is desirable to know their representations. The group $\mathcal{A}_5$ is the symmetry group of the icosahedron and dodecahedron and it is always pleasant to meet these objects again. Finally, $\mathcal{A}_n \subset \mathcal{S}_n$ is close to the entire group $\mathcal{S}_n$ and we'd like to know how their irreducible representations are related.

30.1 Induced Representations

Induced representations were defined early in the development of representation theory by Frobenius himself, who proved the key result known as the *Frobenius Reciprocity Theorem*. This result generalizes several important results in the theory, including the fact that $L^2(G)$ contains all irreducible representations of G .

All of these results generalize to compact Lie groups and we will sketch that generalization below. But the general case requires results from differential geometry and analysis which we have not covered. Thus the theory for finite groups will be completely rigorous, but a few details will be missing from the compact Lie case.

Suppose $H \subset G$ is a subgroup, not necessarily normal, and $\rho : H \rightarrow GL(V)$ is a representation of H . We would like to use this information to obtain a representation of G . But very easy cases show that it is often impossible to extend to $\rho : G \rightarrow GL(V)$.

Instead we use $\rho : H \rightarrow GL(V)$ to construct a representation of G on a different vector space. We first form $G \times V$, which we picture as a trivial vector bundle over G . The group G acts on this bundle and the base space by acting trivially on V , so $g(g_1 \times v) = gg_1 \times v$.

$$\begin{array}{ccc} G \times V & & \\ \downarrow & \pi & \\ G & & \end{array} \quad (30.1)$$

We introduce an equivalence relation on the top and bottom. Whenever $h \in H$, we say that on the top $(g, v) \sim (gh, \rho(h^{-1})v)$ and on the bottom $g \sim gh$. It is trivial to show that both are equivalence relations. Call the equivalence classes on the top E ; the equivalence classes on the bottom form G/H . Clearly there is an induced map

$$\begin{array}{ccc} E & & \\ \downarrow & \pi & \\ G/H & & \end{array} \quad (30.2)$$

This map commutes with the action of G on the top and bottom, so G still acts on E and G/H .

We claim that E is a vector bundle over G/H . This language just means that each $p \in G/H$ has an open neighborhood $\mathcal{U}$ for which there is a diffeomorphism $\pi^{-1}\mathcal{U} \rightarrow \mathcal{U} \times V$ preserving fibers and linear on each fibre. In the finite case, pick a representative for each coset of G/H . Then $\pi^{-1}(gH) = g \times V$ and in this way $E = G/H \times V$.

In the Lie group case, assume that H is a closed subgroup. It may not be possible to continuously pick representatives for each coset., but each $p \in G/H$ has an open neighborhood $\mathcal{U}$ over which a continuous map can be found $\eta : \mathcal{U} \rightarrow G$ such that $\pi \circ \eta$ is the identity, and the previous argument then shows that over $\mathcal{U}$, $E = \mathcal{U} \times V$.

The group G acts from the left on G/H . It also acts on $G \times V$ by acting on G ; this action commutes with the equivalence relation and thus G acts on E . Picture this as follows. On the base space, $p \rightarrow gp$ and in the fibre $E_p \rightarrow E_{gp}$ by a linear transformation.

Definition 55 *Let $H \subset G$ be a subgroup, and suppose $\rho : H \rightarrow GL(V)$ is a representation. The above construction gives a vector bundle*

$$\begin{array}{ccc} E & & \\ \downarrow & \pi & \\ G/H & & \end{array} \quad (30.3)$$

Let W be the space of global sections of this bundle, that is, maps $s : G/H \rightarrow E$ such that $\pi \circ s$ is the identity. The representation induced by ρ , $\text{Ind}(\rho)$, is the representation of G obtained by letting it act on these sections via $[\text{Ind}(\rho)_g s](p) = gs(g^{-1}p)$.

Remark: In the Lie case, this is an infinite dimensional representation. In that case we usually complete the space of sections to form a Hilbert Space. For the moment we restrict attention to the case of finite groups.

Theorem 136 (The Frobenius Reciprocity Theorem) *Suppose ρ is an irreducible representation of H . Let $\text{Ind}(\rho)$ be the induced representation of G . Let ψ be another irreducible representation of G . Then the number of times ψ occurs in the decomposition of $\text{Ind}(\rho)$ is equal to the number of times ρ occurs in the decomposition of ψ_H , where ψ_H is the restriction of ψ to H .*

Example: Suppose $H = \{e\} \subset G$ and let ρ be the identity representation. Then E is the trivial one-dimensional bundle over G , so $\text{Ind}(\rho) = L^2(G)$. The reciprocity theorem then says that the number of times ψ occurs in $L^2(G)$ is the number of times that $\psi_{\{e\}}$ contains the identity representation, but since $\psi_{\{e\}}$ is the identity matrix, this is just the dimension of ψ .

Remark: We prove the reciprocity theorem using character theory, and the main step is to compute the character of $\text{Ind}(\rho)$. The theorem immediately follows from the following two results:

Theorem 137 *The following equalities are correct, where σ is an H coset satisfying $g(\sigma) = \sigma$:*

$$\chi_{\text{Ind}(\rho)}(g) = \frac{1}{|H|} \sum_{\sigma, g_1 \in \sigma} \text{Tr}(\rho(g_1^{-1}gg_1))$$

$$< \chi_{\text{Ind}(\rho)}, \chi_\psi >_G = < \chi_\rho, \chi_{\psi_H} >_H$$

Proof: Every section in W is uniquely a sum of sections which are non-zero at only one point. If $\text{Ind}(\rho)_g$ acts on such a section, it produces a section which is non-zero at another point. The only cases that contribute to the trace are sections s such that s and gs are non-zero at the same point. Suppose this point is a coset σ . Then $s(\sigma) \neq 0$ and $s(g^{-1}\sigma) \neq 0$ and so $\sigma = g^{-1}\sigma$ or equivalently $\sigma = g\sigma$.

If $p = \sigma \in G/H$ and $g_1 \in \sigma$, then g_1 represents p . In that case every point in E_p has a unique representative of the form $g_1 \times v$. So once we have picked g_1 , we can identify E_p with V .

Notice that if $g_1 \in \sigma$ then $\sigma = g_1H = gg_1H$ and so $g_1^{-1}gg_1 \in H$. Say this expression equals h so $g_1^{-1}gg_1 = h$. So $gg_1 = g_1h$.

The representation $\text{Ind}(\rho)(g)$ maps E_p to E_p via $g_1 \times v \rightarrow gg_1 \times v$. Using the last equation in the previous paragraph, this is $g_1 \times v \rightarrow g_1h \times v$. But this last element is equivalent to $g_1 \times \rho(h)v$.

Summarizing, the induced action of g on sections maps a section which is non-zero at a coset σ to another section which is non-zero at the same σ if and only if $\sigma = g\sigma$. In this

case, E_σ can be identified with V and g maps this vector space to itself by $\rho(h)$ where $h = g_1^{-1}gg_1$.

The sections non-zero at a single point don't form a basis for W . To get a basis, we must pick a basis $e_1, \dots, e_k$ for V and consider the sections with values $g_1 \times e_1, \dots, g_1 \times e_k$ at g_1H . Using this basis, the map sending $g_1 \times v$ to $g_1 \times \rho(h)v$ has trace equal $\text{Tr}(\rho(h))$.

Putting all of this together, we find that every coset σ with $\sigma = g\sigma$ yields terms for the trace of $\text{Ind}(\rho)(g)$ and ultimately contributes $\text{Tr}(\rho(h))$ to the character evaluated at g , where $g_1 \in \sigma$ and $g_1^{-1}gg_1 = h$. The element g_1 is not unique. We can replace it by g_1h_1 whenever $h_1 \in H$ and get the same coset σ . Moreover, $(g_1h_1)^{-1}gg_1h_1 = h_1^{-1}hh_1$ and the trace of $\rho(h)$ equals the trace of $\rho(h_1^{-1}hh_1)$. So we can either pick one g_1 for each σ or else allow arbitrary $g_1 \in H$ and divide by $\frac{1}{|H|}$.

The follow aside will be useful in the proof that follows. The crucial equation above is $g_1H = gg_1H$. Suppose σ is an arbitrary H coset with the property $\sigma = g\sigma$. If g_1 is any element of σ , then $\sigma = g_1H = gg_1H$. So $g_1^{-1}gg_1 \in H$. Conversely, if $g_1^{-1}gg_1 \in H$, then g_1H contains gg_1 and also g_1 , so $\sigma = g_1H$ has the property $\sigma = g\sigma$.

Proof, continued: Now we prove the second equation. In the first equation below, we sum over $g \in G$, and then over cosets σ so $g\sigma = \sigma$ and finally over all $g_1 \in \sigma$.

$$< \chi_{\text{Ind}(\rho)}, \chi_\psi >_G = \frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{\sigma, g_1 \in \sigma} \text{Tr}(\rho(g_1^{-1}gg_1)) \overline{\chi_\psi(g)}$$

Since ρ is a representation of H , the expression $\text{Tr}(\rho(g))$ in the above formula only makes sense if $g \in H$. Extend the expression $\text{Tr}(\rho(g))$ to be zero if $g \notin G$. Then in the expression we can sum over *all* cosets σ rather than only those satisfying $\sigma = g\sigma$ because when $\sigma \neq g\sigma$, the expression $g_1^{-1}gg_1$ is not in H .

Notice finally that g_1 can be anything in G , and once g_1 is known, σ must be g_1H . So the final expression above equals

$$\frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{g_1 \in G} \text{Tr}(\rho(g_1^{-1}gg_1)) \overline{\chi_\psi(g)}$$

Since χ_ψ is defined for all g and is a class function, $\chi_\psi(g) = \chi_\psi(g_1^{-1}gg_1)$ and consequently our sum equals

$$\frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{g_1 \in G} \chi_\rho(g_1^{-1}gg_1) \overline{\chi_\psi(g_1^{-1}gg_1)}$$

As g runs over all elements of G , so does $g_1^{-1}gg_1$, so this expression is

$$\frac{1}{|G|} \sum_{g \in G} \frac{1}{|H|} \sum_{g_1 \in G} \chi_\rho(g) \overline{\chi_\psi(g)} = \frac{1}{|H|} \sum_{g \in G} \chi_\rho(g) \overline{\chi_\psi(g)} = \frac{1}{|H|} \sum_{g \in H} \chi_\rho(g) \overline{\chi_\psi(g)} = < \chi_\rho, \chi_{\psi_H} >_H$$

30.2 Subgroups $H \subset G$ of Index Two

Suppose H is a subgroup of a finite group G with index two. There are two important examples of this situation. The first is $\mathcal{A}_n \subset \mathcal{S}_n$ and the second is $Z_n \subset D_n$ where D_n is the dihedral group of all symmetries of a regular polygon and Z_n is the subgroup of all rotations of the polygon.

In general, we use induced representation theory to connect the irreducible representations of H to the irreducible representations of G . The relation is strong enough that if we know one of these sets, we can construct the other. In the next section, we use this result to find all irreducible representations of $\mathcal{A}_n$; that section is just a corollary of the results we prove here.

If $H \subset G$ has index 2, then H is necessarily normal. Indeed the left cosets are H and $G - H$. The right cosets are also H and $G - H$. If $h \in H$, then $hH = Hh$. If $g \in G - H$, gH is a left coset and it cannot be H , so $gH = G - H$. Similarly $Hg = G - H$.

Therefore G/H is a group, which must be Z_2 . This group has one non-trivial character, χ . If ψ is a representation of G , we get another, $\hat{\psi}$, by multiplying ψ by χ . Clearly ψ is irreducible if and only if $\hat{\psi}$ is irreducible. These representations may or may not be equivalent. Note that $\hat{\psi}(h) = \psi(h)$ if $h \in H$ and $\hat{\psi}(g) = -\psi(g)$ if $g \in (G - H)$.

If ρ is a representation of H and $\tau \in (G - H)$, we can get a second representation ρ^c of H using the formula $\rho^c(h) = \rho(\tau^{-1}h\tau)$. We say that ρ^c is a *conjugate* of ρ . At first glance we might suppose ρ^c and ρ are equivalent by writing $\rho(\tau^{-1}h\tau) = \rho(\tau)^{-1}\rho(h)\rho(\tau)$ so that $\rho(\tau)$ creates the equivalence, but notice that $\rho(\tau)$ isn't even defined. So ρ and ρ^c may or may not be equivalent. The equivalence class of ρ^c is uniquely determined, since any other element of $G - H$ equals τh_1 for some $h_1 \in H$ and then $\rho((\tau h_1)^{-1}h(\tau h_1)) = \rho(h_1)^{-1}\rho^c(h)\rho(h_1)$ is an equivalence between ρ^c defined using τh_1 and ρ^c defined using τ . Clearly ρ is irreducible if and only if ρ^c is irreducible.

In short, using the inclusion $H \subset G$, both the irreducible representations of H and the irreducible representations of G come in pairs, which may or may not be equivalent.

Remark: Now on to the main event. Suppose ρ is an irreducible representation of H . We will describe the induced representation on G . Select e as the representative of the coset H and select τ as the representative of the other coset, where $\tau \in (G - H)$. Then E over eH contains all $e \times v$ and E over τH contains all $\tau \times v$. Call these spaces V_1 and V_2 .

Suppose $h \in H$. Then h acts on V_1 by $e \times v \rightarrow h \times v \sim e \times \rho(h)v$. So H acts on V_1 via ρ .

Also $h \in H$ acts on V_2 by $\tau \times v \rightarrow h\tau \times v = \tau(\tau^{-1}h\tau) \times v$. Since H is normal, $\tau^{-1}h\tau \in H$ and the last element is equivalent to $\tau \times \rho(\tau^{-1}h\tau)v$. So the action of H on V_2 is given by $v \rightarrow \rho(\tau^{-1}h\tau)v$.

We conclude that if we start with a representation ρ of H , induce it to $\text{Ind}(\rho)$ of G , and restrict this representation back to H , we get $\rho \oplus \rho^c$.

To complete the picture of the induced representation for ρ , we must explain the action of τ on V_1 and V_2 . Note that τ maps $e \times v$ to $\tau \times v$. So it is the identity map from $V_1 \rightarrow V_2$. Similarly τ maps $\tau \times v \rightarrow \tau^2 \times v$. Note that $\tau^2 \in H$ because G/H is isomorphic to Z_2 , so $\tau^2 \times v$ is equivalent to $e \times \rho(\tau^2)v$. Thus we map $V_2 \rightarrow V_1$ by $v \rightarrow \rho(\tau^2)v$.

Lemma 57 *Let ψ be an irreducible representation of G . Then $\psi \sim \hat{\psi}$ if and only if*

$$\chi_\psi(g) = 0 \quad \text{whenever } g \notin H$$

Proof: Since $\psi(g) = \hat{\psi}(g)$ when $g \in H$, taking traces gives $\chi_\psi(g) = \chi_{\hat{\psi}}(g)$ whenever $g \in H$. Since $\psi(g) = -\hat{\psi}(g)$ when $g \notin H$, taking traces gives $\chi_\psi(g) = -\chi_{\hat{\psi}}(g)$ whenever $g \notin H$. Two irreducible representations are equivalent if and only if they have the same character, so the lemma follows immediately.

Lemma 58 *Let ψ be an irreducible representation of G . Then either ψ restricted to H is irreducible and ψ and $\hat{\psi}$ are not equivalent, or else ψ restricted to H is a sum of two irreducible representations and ψ and $\hat{\psi}$ are equivalent.*

Proof: Let χ be the character of ψ . Recall that the L^2 norm of χ is the number of irreducible components of ψ and in particular is 1 if ψ is irreducible. Therefore

$$\frac{1}{|G|} \sum_{g \in G} |\chi_\psi(g)|^2 = 1$$

and so

$$|G| = 2|H| = \sum_{h \in H} |\chi_\psi(h)|^2 + \sum_{g \in (G-H)} |\chi_\psi(g)|^2$$

Dividing by $|H|$ gives

$$2 = \frac{1}{|H|} \sum_{h \in H} |\chi_\psi(h)|^2 + \frac{1}{|H|} \sum_{g \in (G-H)} |\chi_\psi(g)|^2$$

The first term on the right is the L^2 -norm of ψ restricted to H . It follows that this term is either 1 or 2 and thus ρ is either irreducible or else breaks into a sum of two irreducible representations of H . Moreover the first term equals 1 just in case the second term is not zero and thus just in case ψ and $\hat{\psi}$ are not equivalent. QED.

Remark: The following theorem gives the key result of this section.

Theorem 138 *Let $H \subset G$ be subgroup of index 2 and let ψ be an irreducible representation of G . Then exactly one of the following is true:*

- *The representations ψ and $\hat{\psi}$ on G are not equivalent and ψ restricted to H is irreducible. If $\rho = \psi|_H$, then ρ and ρ^c are equivalent. If we restrict ψ to H and then induce this to G , we get a representation equivalent to $\psi \oplus \hat{\psi}$.*
- *The representations ψ and $\hat{\psi}$ of G are equivalent and ψ restricted to H is not irreducible. Restricting ψ to H gives a sum of two irreducible representations of H , ρ_1 and ρ_2 which are conjugate but not equivalent. If either ρ_1 or ρ_2 is induced back to G , the induced representation is equivalent to ψ .*

In this way, every irreducible representation of H can be formed by restricting an irreducible representation of G to H and then decomposing if necessary. Every irreducible representation of G can be formed by inducing an irreducible representation of H and then decomposing if necessary.

Proof: Suppose ψ is an irreducible representation of G , and $\psi|_H$ is irreducible. Let $\rho = \psi|_H$ and apply the Frobenius reciprocity theorem to ρ and ψ . Then ρ appears once in $\psi|_H$ and so ψ occurs once in $\text{Ind}(\rho)$. Moreover $\hat{\psi}$ is irreducible on G and its restriction to H is also ρ . By the Frobenius reciprocity theorem, $\hat{\psi}$ occurs once in $\text{Ind}(\rho)$. Since ψ and $\hat{\psi}$ are not equivalent and $\text{Ind}(\rho)$ equals the sum of their dimensions, we must have $\text{Ind}(\rho) = \psi \oplus \hat{\psi}$. Finally, our construction of $\text{Ind}(\rho)$ shows that the restriction of $\text{Ind}(\rho)$ to H is $\rho \oplus \rho^c$, but the restriction of $\psi \oplus \hat{\psi}$ is $\rho \oplus \rho$, so ρ and ρ^c are equivalent.

In the second case, ψ restricted to H is not irreducible, so it is the sum of two irreducible representations ρ_1 and ρ_2 of H , which might be equivalent. Apply the reciprocity theorem to ρ_1 , $\text{Ind}(\rho_1)$, and ψ . Then ρ_1 occurs once or twice in the restriction of ψ to H , so ψ occurs once or twice in $\text{Ind}(\rho_1)$. Since the restriction of ψ is a sum of both ρ_1 and ρ_2 , its dimension is larger than either of their dimensions and so ψ must occur once in $\text{Ind}(\rho_1)$. It follows that ρ_1 and ρ_2 must be inequivalent. It also follows that the restriction of $\text{Ind}(\rho_1)$ to H equals the restriction of ψ to H , and thus equals $\rho_1 \oplus \rho_2$. But we know that this restriction also equals $\rho_1 \oplus \rho_1^c$. So $\rho_2 = \rho_1^c$, and hence ρ_1 and ρ_1^c are not equivalent. QED.

30.3 Irreducible Representations of the Alternating Group

Theorem 139 *Consider irreducible representations ψ of S_n and their associated Young diagrams. Some Young diagrams remain unchanged when flipped over the diagonal, while others change; diagrams which change when flipped thus occur in pairs. In this situation, call the two related irreducible representations ψ_1 and ψ_2 .*

- *If the Young diagram changes when flipped, both ψ_1 and ψ_2 give the same representation when restricted to A_n and this representation is irreducible. If this representation*

of $\mathcal{A}_n$ is induced back up to $\mathcal{S}_n$, the induced representation splits into two irreducible representations, giving ψ_1 and ψ_2 .

- If the Young diagram remains the same when flipped, then when ψ is restricted to $\mathcal{A}_n$, it splits into a sum of two irreducible representations, each of dimension $\frac{\dim \psi}{2}$. These representations are not equivalent. If either is induced back up to $\mathcal{S}_n$, the induced representation is isomorphic to ψ .
- All irreducible representations of $\mathcal{A}_n$ arise in one of these two ways.

Example: The irreducible representations of $\mathcal{S}_n$ have dimensions 1, 4, 5, 6, 5, 4, 1. Each is paired with another except the representation of dimension 6. Thus the dimensions of the irreducible representations of $\mathcal{A}_5$ is 1, 4, 5, 3, 3.

Warning: In the case $n = 5$, Young diagrams are paired except for one diagram “in the middle of the list.” It is tempting to suppose that this is true in general, but a little experimentation finds n with multiple Young diagrams which remain the same when flipped.

Proof: Recall that if ψ is the representation of $\mathcal{S}_n$ for a Young diagram λ and we flip λ across the diagonal, then the new representation is ψ multiplied by the alternating character. Thus the two representations agree on $h \in \mathcal{A}_n$ and differ by a sign on $g \in \mathcal{S}_n - \mathcal{A}_n$. So these representations are ψ and $\hat{\psi}$ and they are not equivalent. This puts us in the first case of the main theorem in the previous section, so ψ restricted to $\mathcal{A}_n$ is irreducible. Moreover, if we induce this restriction back up to a representation of $\mathcal{S}_n$, we get a sum of ψ and $\hat{\psi}$. Thus we have a one-to-one correspondence between pairs of unequal Young diagrams, pairs of inequivalent irreducible representations on $\mathcal{S}_n$ whose restrictions to $\mathcal{A}_n$ are irreducible and agree, and irreducible representations on $\mathcal{S}$ with the property that when you Induce to a representation of G and then restrict to a representation of H you get a sum of two inequivalent irreducible representations.

The other situation occurs when we start with a Young diagram invariant under flipping. Then we get an irreducible representation ψ of $\mathcal{S}_n$ which is equivalent to $\hat{\psi}$. So we are in the second case of the previous theorem and ψ restricted to $\mathcal{A}_n$ is a sum of two inequivalent irreducible representations. The Induced representation of either of these irreducible representations is ψ . So we have a one-to-one correspondence between Young diagrams invariant under flipping, irreducible representations of $\mathcal{S}_n$ such that ψ and $\hat{\psi}$ are equivalent, and inequivalent pairs of irreducible representations of $\mathcal{A}_n$ that are conjugate. Inducing either representation in the pair back to G gives the original ψ . QED.

30.4 The Icosahedron and Dodecahedron

It is well known that $\mathcal{A}_5$ is the group of rotations of a dodecahedron. Thus we expect to find a three dimensional irreducible representation of $\mathcal{A}_5$. But curiously, there are two. On the other hand, it is surprising that $\mathcal{S}_n$ has no representation related to the dodecahedron, since it has no irreducible representations of dimension 3. All this will be explained.

Notice that $\mathcal{S}_5$ is *not* the full symmetry group of the dodecahedron. Indeed $-I$ is a symmetry of the dodecahedron which commutes with all rotations, so the full group of symmetries is $\mathcal{A}_5 \times \mathbb{Z}_2$. But no permutation in $\mathcal{S}_5$ commutes with everything.

It is possible to inscribe a cube in a dodecahedron. The cube has 12 edges and the dodecahedron has 12 faces; each face contains one edge of the cube. This edge passes through vertices separated by "two clicks". See the next illustration.

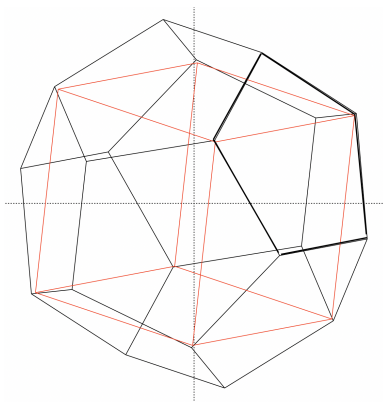


Figure 30.1: Cube In Dodecahedron

Since a face has five sides, there are five possible edges in each face. Once one edge of a cube is known, the remaining cube is unique. Therefore a cube can be inscribed in a dodecahedron in five different ways. Each rotation of the dodecahedron determines a permutation of these five objects.

In this way we get a map from the group of rotations of a dodecahedron to $\mathcal{S}_5$. There are 60 rotations of the cube and 120 elements of $\mathcal{S}_5$, so the image of the rotation group is a subgroup of $\mathcal{S}_5$ of order 60. We claim this image is $\mathcal{A}_5$. Since the image contains half of the elements of $\mathcal{S}_5$, it is a normal subgroup. The group $\mathcal{A}_5$ is also normal. If these two groups are not equal, their intersection will be a proper normal subgroup of $\mathcal{A}_5$, but $\mathcal{A}_5$ is simple.

In short, the group of rotations of the dodecahedron is isomorphic to $\mathcal{A}_5$. Each numbering of the five cubes leads to a specific isomorphism. If we think of these isomorphisms in

reverse as maps

$$\mathcal{A}_5 \rightarrow GL(R^3)$$

then each numbering of the cubes yields a representation of $\mathcal{A}_5$.

The group $\mathcal{S}_5$ consists of all permutations σ of $\{1, 2, 3, 4, 5\}$, and each of these gives a renumbering of the cubes in the dodecahedron. It is easy to see that if one numbering gives the representation $\rho(g)$, then renumbering by σ gives $\rho(\sigma^{-1}g\sigma)$.

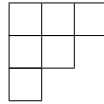
If $\sigma \in \mathcal{A}_5$, we can form the rotation $T = \rho(\sigma)$ and write the previous formula as

$$\rho(\sigma^{-1}g\sigma) = \rho(\sigma)^{-1}\rho(g)\rho(\sigma) = T^{-1}\rho(g)T$$

and consequently these representations are equivalent. So if we have a specific numbering and its associated representation, all even renumberings yield equivalent representations. Similarly all representations associated with odd renumberings are equivalent to each other.

Therefore if we fix a numbering of the cubes, we get two irreducible representations of $\mathcal{A}_5$ up to equivalence. One is the representation associated with the fixed numbering, $\rho(g)$. The other is $\rho(\tau^{-1}g\tau)$ for any odd permutation τ . We could, for instance, choose $\tau = (12)$.

This is in agreement with the general theory of representations of $\mathcal{A}_5$ from the previous section. According to that theory, the two dodecahedral representations of $\mathcal{A}_5$ come from the six dimensional irreducible representation of $\mathcal{S}_5$ associated with the Young diagram



When this representation is restricted to $\mathcal{A}_5$, it splits into two three dimensional irreducible representations which are conjugate but not equivalent.

Moreover, that theory says that the six dimensional irreducible representation of $\mathcal{S}_5$ is $Ind(\rho)$ where ρ is either of the inequivalent dodecahedral representations. Choosing $\tau = (12)$ as a representative of the nontrivial coset of G/T , we find that the representation space is $R^3 \oplus R^3$ where $\mathcal{A}_5$ acts in the first space as $\rho(h)$ and in the second as $\rho(\tau^{-1}h\tau)$. Since $\tau^2 = e$, the induced representation applied to τ sends $R^3 \rightarrow R^3$ in either direction by the identity map. The most general element of $G - H$ can be written as τh and thus acts on the two spaces by first applying $\rho(h) \oplus \rho^c(h)$ and then switching the components.

Another Visualization

A representation is more than just its image group in $GL(V)$; two representations can easily have the same image group and yet not be equivalent. There are two representations of $\mathcal{A}_5$ by rotations of the dodecahedron. Both of these representations have the same rotations, which multiply as their related permutations multiply, so we cannot distinguish these representations by just looking at the rotations. We'll describe a method which can tell them apart.

Fix one of the cubes and consider the rotational symmetries which map that cube to itself. There are 60 symmetries altogether and 5 cubes, so there are $\frac{60}{5} = 12$ such symmetries. On the other hand, a cube has 24 symmetries. Which symmetries come from symmetries of the dodecahedron?

A tetrahedron can be inscribed in a cube in exactly two ways. The elements of the group either preserve these tetrahedrons, or flip them. The subgroup of rotations that preserve the tetrahedrons has 12 elements.

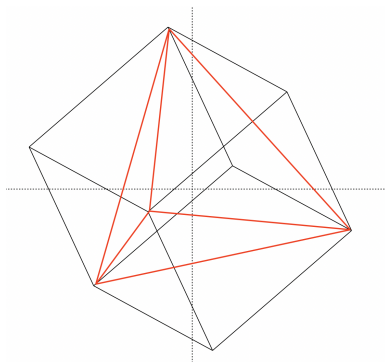


Figure 30.2: Tetrahedron in Cube

It is easy to check that only one subgroup of the symmetry group of a cube has 12 elements, and that is the group of rotations which fix the two tetrahedra as sets. So symmetries of our cubes coming from dodecahedral symmetries must preserve these tetrahedra.

Suppose we fix one cube and color its tetrahedra red and green. Then it is possible to color the tetrahedra in all other cubes red and green such that all permutations of these tetrahedra arising from the symmetries of the dodecahedron preserve these colors. We will let the reader give details after sketching the main idea. Suppose $\mathcal{C}_1$ and $\mathcal{C}_2$ are two cubes and we have dodecahedral symmetries s_1, s_2 mapping $\mathcal{C}_1$ to $\mathcal{C}_2$. Then $s_2^{-1}s_1$ maps $\mathcal{C}_1$ to itself and thus preserves tetrahedral colors. It follows that we can consistently color the tetrahedrons in $\mathcal{C}_2$ using either s_1 or s_2 and get the same result.

Completion: Using all this, we'll provide a way to visualize the six dimensional irreducible representation of $\mathcal{S}_5$. Draw two three dimensional pictures of dodecahedra. Form the five cubes in each, and the ten tetrahedra in each, and color half of the tetrahedra red and half green as before. Keep only the green tetrahedra in the left picture and only the red tetrahedra in the right picture.

If $h \in \mathcal{A}_5$, then let h act via $\rho(h)$ on the left side and by $\rho^c(h)$ on the right side. Thus h will permute the green tetrahedra on the left and will permute the red tetrahedra on the right, but these permutations will not be the same. Each permutation will be obtained by rotating the dodecahedra, but different rotations will be used on the left and right.

As for τ , it's action is easy to visualize. Just switch the tetrahedra so a green tetrahedra on the left is mapped to the red tetrahedra in the same cube on the right, and each red tetrahedron on the right is mapped to the corresponding green one on the left. In the general case of $(\tau h) \in (\mathcal{S}_5 - \mathcal{A}_5)$, first permute the red and green tetrahedra with h and then switch with τ . It follows that elements of the representation are completely determined by their permutations on the green and red tetrahedra.

Some readers might prefer to draw the green and red tetrahedra in the same picture. That is fine; the only trouble is that actions of h will use one rotation of the dodecahedron to rotate the green objects and a different rotation to rotate the red objects.

Chapter 31

Irreducible Representations of $SU(n)$

31.1 The Setup

In this chapter we find explicitly all irreducible representations of $SU(n)$. Our constructions will also work for $GL(n, C)$, $SL(n, C)$, $SL(n, R)$ and $U(n)$. We first sketch the central idea of the construction.

Let $V = C^n$ and notice that $SU(n)$ acts naturally on V . Consequently it also acts on $V \otimes V \otimes \dots \otimes V$ by

$$A(v_1 \otimes v_2 \otimes \dots \otimes v_k) = A(v_1) \otimes A(v_2) \otimes \dots \otimes A(v_k)$$

On the other hand, the symmetric group $\mathcal{S}_k$ acts on this same space via

$$\sigma(v_1 \otimes v_2 \otimes \dots \otimes v_k) = v_{\sigma^{-1}(1)} \otimes v_{\sigma^{-1}(2)} \otimes \dots \otimes v_{\sigma^{-1}(k)}$$

The actions of $SU(n)$ and $\mathcal{S}_k$ commute.

We can write $V \otimes V \otimes \dots \otimes V$ as a direct sum of irreducible representations of $\mathcal{S}_k$. These irreducible representations are associated with Young diagrams containing k boxes.

Suppose W is an irreducible, invariant subspace of $V \otimes \dots \otimes V$ under $\mathcal{S}_k$. We would like to use the commutativity of the actions of $SU(n)$ and $\mathcal{S}_k$ to argue that W is an invariant subspace under $SU(n)$, but the argument fails miserably. If $A \in SU(n)$, then $A(W)$ is another irreducible, invariant subspace giving the same irreducible representation of $\mathcal{S}_k$, since the action of A commutes with the action of $\mathcal{S}_k$. But this subspace need not equal W because this particular irreducible representation of $\mathcal{S}_k$ may occur more than once.

To fix this, we gather together all irreducible subspaces with equivalent irreducible representations. First we write $V \otimes \dots \otimes V$ as a sum of irreducible subspaces for S_k . We then sum all subspaces which provide the same irreducible representation of S_k up to equivalence. In this way we can write the tensor product as

$$W_1 \oplus W_2 \oplus \dots \oplus W_m$$

where each W_i is a sum of equivalent irreducible representations of S_k , and different i correspond to different irreducible representations.

It is then easy to prove that each $A \in SU(n)$ maps W_i back to itself. The action of $SU(n)$ on W_i can thus be written as a sum of irreducible representations of $SU(n)$. We will prove that they are all equivalent, and the irreducible representations associated with W_i and W_j for $i \neq j$ are not equivalent. Therefore the W_i , originally constructed using the decomposition of the action of S_n , could have just as easily been constructed using the action of $SU(n)$. Each W_i pairs up an irreducible representation of S_k with an associated irreducible representations of $SU(n)$.

Moreover, we will prove the following:

- the dimension of the irreducible representation of S_k , times the number of times it occurs, equals the dimension of W
- the dimension of the irreducible representation of $SU(n)$, times the number of times it occurs, equals the dimension of W
- the dimension of the irreducible representation of $SU(n)$ equals the number of times the irreducible representation of S_k occurs in W
- the dimension of the irreducible representation of S_k equals the number of times the irreducible representation of $SU(n)$ occurs in W

31.2 Example: The Case of Symmetric Tensors

The simplest irreducible representation of S_k is the identity representation. Each such representation is one-dimensional, generated by a nonzero element of $V \otimes \dots \otimes V$ left fixed by every permutation. For example, $v \otimes \dots \otimes v$ has this property. But there are other invariant elements. Suppose V has dimension two with basis x_1, x_2 . Then $x_1 \otimes x_1$ and $x_2 \otimes x_2$ are left fixed by all permutations, but so is $x_1 \otimes x_2 + x_2 \otimes x_1$.

The subspace of $V \otimes \dots \otimes V$ formed by tensors invariant under all elements of the permutation group S_k is called the space of symmetric tensors and denoted $S^k(V)$. It is easy to specify these elements concretely. Define a map $S : V \otimes \dots \otimes V \rightarrow V \otimes \dots \otimes V$ given

by

$$S(v_1 \otimes v_2 \otimes \dots \otimes v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} v_{\sigma^{-1}(1)} \otimes \dots \otimes v_{\sigma^{-1}(k)}$$

The symmetric tensors are exactly tensors in the image of this map. The $\frac{1}{k!}$ has been introduced to make $S^2 = S$, so S leaves symmetric tensors fixed.

We denote $S(v_1 \otimes v_2 \otimes \dots \otimes v_k) = v_1 v_2 \dots v_k$, noting that the symbols on the right commute. It is then easy to see that if $e_1, \dots, e_n$ is a basis for V , a basis for $S^k(V)$ consists of all $e_1^{k_1} e_2^{k_2} \dots e_n^{k_n}$ where $\sum k_i = k$.

From here it is easy to determine the dimension of $S^k(V)$. It is the number of choices of the k_i which sum to k . Suppose, for example, that $k = 8$ and $n = 3$. One choice of k_i is $2 + 5 + 1$. We can indicate this by writing down 8 + 2 spots and putting a 's in the spots for k_i and $_$ in the separator spots. So $2 + 5 + 1$ would be aa_aaaa_a . One possibility is $k_i = 0$, so the separate spots can appear in any of these positions. In general there are $n + k - 1$ spots and we must pick k spots for the a 's, but the order of the a 's does not matter. Thus the dimension of $S^k(V)$ is

$$\binom{n + k - 1}{k}$$

If we start with the identity representation of the symmetric group S_k , then the symmetric tensors generate representation spaces for this identity representation. So the associated space W is $S^k(V)$. Since the identity representation has dimension one, the related representation of $SU(n)$ acts on the full $S^k(V)$. This representation will be proved irreducible as a special case of theory.

Recall that much earlier we found all irreducible representations of $SU(2)$, and all of these representations had the form $S^k(V)$ for $V = C^2$ and varying k . So we are making progress. But there are other irreducible representations of $SU(n)$ for $n > 2$.

31.3 Other Examples

There is another one dimensional irreducible representation of S_k , the alternating representation. A non-zero vector generates this representation if it is left fixed by even permutations and its sign is changed by odd permutations. The set of all such vectors forms the space W . By our theory, $SU(n)$ acts irreducibly on this space, since the corresponding irreducible representation of S_k is one dimensional.

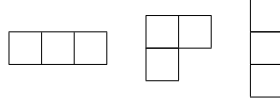
We define $\Lambda : V \otimes \dots \otimes V \rightarrow V \otimes \dots \otimes V$ by linear extension of

$$\Lambda(v_1 \otimes \dots \otimes v_k) = \frac{1}{k!} \sum_{\sigma} \text{sgn}(\sigma) v_{\sigma^{-1}(1)} \otimes \dots \otimes v_{\sigma^{-1}(k)}$$

Again $\Lambda^2 = \Lambda$ and the image of Λ is the set $\Lambda^k(V)$ of all skew-symmetric k -tensors. The image of $v_1 \otimes v_2 \otimes \dots \otimes v_k$ is denoted $v_1 \wedge v_2 \wedge \dots \wedge v_k$. If $e_1, \dots, e_n$ is a basis of V , a basis of $\Lambda^k(V)$ is formed by all $e_{i_1} \wedge e_{i_2} \wedge \dots \wedge e_{i_k}$ with $i_1 < i_2 < \dots < i_k$. Since the alternating representation of $\mathcal{S}_k$ is one-dimensional, the corresponding representation of $SU(n)$ is the full space $W = \Lambda^k(V)$ of alternating tensors. Notice that this space vanishes if $k > \dim V$. The dimension of $\Lambda^k(V)$ is $\binom{n}{k}$.

There are only two irreducible representations of $\mathcal{S}_2$, so $V \otimes V = S^2(V) \oplus \Lambda^2(V)$.

However, we get something new when there are three copies of V . This time there are three possible Young diagrams:



The first corresponds to $S^3(V)$ and the third corresponds to $\Lambda^3(V)$. The middle diagram produces a representation of $\mathcal{S}_3$ with dimension 2, so the corresponding irreducible representation of $SU(n)$ occurs twice in W . We'll determine this representation concretely in a later section, but we can already find its dimension. The dimension of $V \otimes V \otimes V$ is n^3 and the dimension of $S^3(V)$ is $\binom{n+2}{3}$ and the dimension of $\Lambda^3(V)$ is $\binom{n}{3}$, so the dimension of the missing representation of $SU(n)$ is half of

$$\begin{aligned} n^3 - \frac{(n+2)(n+1)n}{6} - \frac{n(n-1)(n-2)}{6} &= \frac{6n^3 - (n^3 + 3n^2 + 2n) - (n^3 - 3n^2 + 2n)}{6} \\ &= \frac{4n^3 - 4n}{6} = \frac{2n^3 - 2n}{3} \end{aligned}$$

and thus the missing irreducible representation of $SU(n)$ should have dimension $\frac{n^3 - n}{3} = \frac{(n-1)n(n+1)}{3}$. Notice that the numerator is indeed divisible by 3.

31.4 Irreducible Representations of $\mathcal{S}_k$ and the Structure of W

Fix a particular W_i . We will call this space W from now on so we can use i as an index of other things. This space is a direct sum of multiple copies of an irreducible representation of $\mathcal{S}_k$. Each such representation corresponds to a Young diagram which we will call λ . Number the boxes in this diagram in the standard way to obtain a Young tableau. For instance, we might have



Define

$$\begin{aligned} P_\lambda &= \{ \sigma \in \mathcal{S}_k \mid \sigma \text{ preserves the rows } \} \\ Q_\lambda &= \{ \sigma \in \mathcal{S}_k \mid \sigma \text{ preserves the columns } \} \\ a_\lambda &= \sum_{\sigma \in P} \sigma \\ b_\lambda &= \sum_{\sigma \in Q} \text{sgn}(\sigma) \sigma \end{aligned}$$

and let

$$c_\lambda = a_\lambda b_\lambda$$

This c_λ is an element of the group algebra $A(\mathcal{S}_k)$, called the Young symmetrizer. The subspace

$$A(\mathcal{S}_k)c_\lambda$$

is invariant under the action of the symmetric group and we proved that the symmetric group acts irreducibly on it. Distinct Young diagrams give inequivalent representations, and all irreducible representations have this form.

From now on, we shall just write A for the group algebra of $\mathcal{S}_k$. Notice carefully that usually there are distinct elements a_1 and a_2 of A such that $a_1 c_\lambda = a_2 c_\lambda$, so the representation space is not isomorphic to all of A .

When proving these results, we proved two useful facts about c_λ . First, $c_\lambda^2 = n_\lambda c_\lambda$ where $n_\lambda \neq 0$. And second, for any $a \in A$, $c_\lambda a c_\lambda$ is a complex multiple of c_λ .

These are the only results we need to obtain a full picture of W .

Notice that individual permutations σ are elements of the group algebra A , but also act on $V \times \dots \times V$ as maps from this space to itself. In the same way, a_λ, b_λ , and c_λ belong to the group algebra but also define maps from $V \otimes \dots \otimes V$ to itself. In the exposition which follows, we often denote the map induced by c_λ on w by writing $C_\lambda(w)$.

In particular, suppose we fix an element w in $V \otimes \dots \otimes V$. We can then map the group algebra A to $V \otimes \dots \otimes V$ by sending σ to σw . This map sends the representation space $A c_\lambda$ in the group algebra to a subspace of $V \otimes \dots \otimes V$ by $a c_\lambda \rightarrow a c_\lambda w = a C_\lambda(w)$. Both subspaces are invariant under the symmetric group and this map preserves that action, so the kernel of the map is an invariant subspace of our irreducible representation. Consequently

Lemma 59 *Either $C_\lambda(w) = 0$ or else the image of this map is a subspace of $V \otimes \dots \otimes V$ on which the symmetric group acts irreducibly with representation λ . In this second case, the image is contained in the W associated with λ .*

Lemma 60 *If $C_\lambda(w) \neq 0$, then C_λ maps every element of the representation space $AC_\lambda(w)$ inside $V \otimes \dots \otimes V$ to a complex multiple of $C_\lambda(w)$. Consequently, each of these irreducible representation spaces for the symmetric group inside $V \otimes \dots \otimes V$ is associated with an element $C_\lambda(w)$ which is unique up to complex scalars.*

Proof: This is an immediate consequence of the fact that $c_\lambda ac_\lambda \in A$ is a complex multiple of c_λ . QED.

Remark: We are *not* claiming that w is unique up to scalar. We are only claiming that $C_\lambda(w)$ is unique up to scalar. But we can say more. Since $c_\lambda^2 = n_\lambda c_\lambda$ for a nonzero n_λ ,

$$C_\lambda(w) = C_\lambda \left(\frac{C_\lambda(w)}{n_\lambda} \right)$$

and therefore we can choose w in the image of $C_\lambda : V \otimes \dots \otimes V \rightarrow V \otimes \dots \otimes V$

Theorem 140 *Every irreducible representation space in $V \otimes \dots \otimes V$ for a representation of $\mathcal{S}_k$ associated with the Young diagram λ has the form Aw where A is the group algebra of $\mathcal{S}_k$ and w is a nonzero element in the image of $C_k : V \otimes \dots \otimes V \rightarrow V \otimes \dots \otimes V$. This w is unique up to multiplication by a non-zero constant scalar, and every w occurs.*

.

Proof: If $w \neq 0$ is in the image of C_λ , then there is a w_1 such that $w = C_\lambda(w_1)$. Thus

$$Aw = AC_\lambda(w_1)$$

and so the symmetric group acts irreducibly on this space and the representation is associated with λ . Moreover, $C_\lambda(w_1) = w$ is uniquely determined up to a scalar.

Conversely, if we have an irreducible representation of $\mathcal{S}_k$ on a non-trivial subspace of $V \otimes \dots \otimes V$, then the group algebra element c_k must map to some non-zero $w \in V \otimes \dots \otimes V$. Since the representation space in the group algebra is Ac_λ , it is Aw in $V \otimes \dots \otimes V$. But c_λ maps to w , so c_λ^2 maps to $C_\lambda(w)$. Since $c_\lambda^2 = n_\lambda c_\lambda$, it follows that $n_\lambda c_\lambda$ maps to $C_\lambda(w)$ and so c_λ maps to $C_\lambda \left(\frac{w}{n_\lambda} \right)$. Putting these results together,

$$w = C_\lambda \left(\frac{w}{n_\lambda} \right)$$

and thus w is in the image of C_λ . QED.

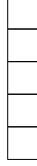
Remark: It follows that the image of $C_\lambda(V \otimes \dots \otimes V)$ is a subspace of the W associated with the representation λ of $\mathcal{S}_k$. If this irreducible representation is one dimensional, this subspace will turn out to be all of W . But otherwise, this subspace will turn out to be invariant under $SU(n)$ and provide an irreducible representation of this group. More details soon.

Notice that horizontal Young diagrams like



define elements $c_\lambda = \sum_\sigma \sigma$ and thus C_λ is our symmetrizer S up to the irrelevant factor $\frac{1}{k!}$, and the image of C_λ is the space of symmetric tensors $S^k(V)$.

Similarly vertical Young diagrams like



define elements $c_\lambda = \sum_\sigma \text{sgn}(\sigma)\sigma$ and thus C_λ is the map Λ up to the factor $\frac{1}{k!}$ and the image of C_λ is the space of alternating tensors $\Lambda^k(V)$.

31.5 $\text{Hom}_{\mathcal{S}_k}[V_1 \otimes \dots \otimes V_m, V_1 \otimes \dots \otimes V_m]$

Remark: We now determine exactly how $SU(n)$ acts on each W . Our analysis works for any of $G = GL(n, C), SL(n, C), GL(n, R), SL(n, R), U(n)$, or $SU(n)$, so we just call the group G in this section. If $A \in G$, we have

$$A(v_1 \otimes \dots \otimes v_k) = A(v_1) \otimes \dots \otimes A(v_k)$$

Such maps commute with the action of $\mathcal{S}_k$, and it is convenient to study the full vector space of all homomorphisms which similarly commute. This set is denoted

$$\text{Hom}_{\mathcal{S}_k}[V_1 \otimes \dots \otimes V_m, V_1 \otimes \dots \otimes V_m]$$

This is a finite dimensional vector space whose elements are linear transformations, but not necessarily isomorphisms or elements of G . However each $A \in G$ is such a map.

An arbitrary linear map

$$M : V \otimes \dots \otimes V \rightarrow V \otimes \dots \otimes V$$

is the same thing as an arbitrary linear map

$$M : W_1 \oplus \dots \oplus W_m \rightarrow W_1 \oplus \dots \oplus W_m$$

and such a map decomposes as a matrix of linear maps $M_{ij} : W_i \rightarrow W_j$. Clearly each of these maps commutes with the action of $\mathcal{S}_k$. Fix i and j ; then W_i is the sum of irreducible subspaces X_k , all equivalent, and W_j is the sum of irreducible subspaces Y_l , all equivalent, and so the map M_{ij} breaks up into a matrix of linear maps $X_k \rightarrow Y_l$, which again clearly

commute with the action of S_k . Schur's lemma shows that if $i \neq j$ then each piece is zero. So the only non-zero M_{ij} are $M_{ii} : W_i \rightarrow W_i$. We conclude that

$$\text{Hom}_{S_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V] = \bigoplus_i \text{Hom}_{S_k}[W_i, W_i]$$

and thus it suffices to study the W_i one by one.

Recall that the image of C_λ is a subspace of W . Pick a basis $t_1, \dots, t_s$ for this subspace. If A is the group algebra of S_k , then each At_i is an irreducible representation space isomorphic to Ac_λ inside A . If $M \in \text{Hom}_{S_k}[W_i, W_i]$, then M can be considered a matrix of maps M_{ij} sending At_i to At_j . Each of these maps is an intertwining operator between identical irreducible representation spaces, and thus by Schur's lemma a complex constant. Thus the matrix (M_{ij}) maps the space with basis $\{t_1, \dots, t_s\}$ to itself.

The representation space Ac_λ in the group algebra is generated by the elements σc_λ as σ runs over S_k . Find a subset $\sigma_1, \dots, \sigma_s$ giving a basis of the representation space. Since c_λ is in this space, we can require that σ_1 is the identity of the group.

Then clearly for each fixed j , the $\sigma_i t_j$ form a basis for the representation space At_j and thus the full set of $\sigma_i t_j$ form a basis for W .

Return to a typical $M \in \text{Hom}_{S_k}[W_i, W_i]$. Since this map commutes with the action of S_k and $M(t_j) = \sum_j M_{jk} t_k$, we have $M(\sigma_i t_j) = \sum_j M_{jk} \sigma_i t_k$. In short, not only does M leave the subspace with basis $t_1, \dots, t_s$ invariant, but it also leaves each of the subspaces with basis $\sigma_i t_1, \dots, \sigma_i t_s$ invariant, and the matrices expressing its action on each of these subspaces is exactly the same.

31.6 A Mental Picture of All of This

The previous section completes our picture of the two groups S_k and G both acting on a particular W . It took me weeks to develop this picture. So I'm going to repeat the story in this redundant section!

Splitting $V \otimes \dots \otimes V$ into irreducible subspaces under the symmetric group, we find that we can write

$$V \otimes \dots \otimes V = W_1 \oplus \dots \oplus W_m$$

where each W_i is associated with a particular irreducible representation. This representation is determined by a Young symmetrizer c_λ . So we can map c_λ to elements $t_1, t_2, \dots, t_s$ in W which generate independent copies of this representation summing to W . These t_i are certainly not unique, but that doesn't matter.

The abstract irreducible representation determined by c_λ is a subspace of the group algebra, and we can choose a basis $\sigma_i c_\lambda$ where the σ_i are particular elements of the symmetric group. The choice of these is not unique, but that doesn't matter. We insist that $\sigma_1 = e$.

If we fix j , the $\sigma_i t_j$ are a basis for the subspace of W determined by t_j . And as we vary both i and j , the $\sigma_i t_j$ form a basis for W .

It is useful to visualize the $\sigma_i t_j$ for fixed j as forming columns in W , where each column is an irreducible representation of the symmetric group.

We can also form rows in W by fixing i and varying j . One row contains all the t_j . Another row contains all the $\sigma_2 t_j$. Etc.

To understand the significance of these rows, consider

$$\text{Hom}_{\mathcal{S}_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$$

It is the set of all homomorphisms which commute with the action of $\mathcal{S}_k$, so in particular it contains all maps coming from the action of $GL(n, C)$ on $V \otimes \dots \otimes V$. Amazingly, this set of invariant homomorphisms has a very simple structure. First, each such homomorphism preserves the W_i and acts completely independently on W_i . So for instance the set contains homomorphisms which act on a particular W and are zero on all other W_j . Moreover, these invariant homomorphisms leave the rows invariant. So for instance they map the space generated by $t_1, \dots, t_s$ back on itself. There is no further restriction on individual rows, so *any* linear transformation from the space with basis $t_1, \dots, t_s$ to itself can occur. *However*, invariance under $\mathcal{S}_k$ requires that exactly the same transformation act on each of the parallel rows in W . So if the map from the space with basis $t_1, \dots, t_s$ to itself is given by a matrix M , then the map from the space with basis $\sigma_2 t_1, \dots, \sigma_2 t_s$ to itself must be given by exactly the same matrix M .

Imagine watching the action of such a homomorphism on the sum of the W_i . You see completely independent actions in the various W . In each particular W , the bottom row moves in an arbitrary way. But the higher rows all move in parallel. That's the picture!

In particular, the action of $GL(n, C)$ fits this picture. But be careful. This action has many more rules. First, all the W must move at once; these actions are no longer independent. Second, on the bottom row of a given W , we definitely do not see all non-singular homomorphisms; instead we see some complicated representation of $GL(n, C)$. If we look at the bottom rows of the various W , we find lots of different representations of $GL(n, C)$ acting simultaneously. Finally on a particular W , higher rows are acting in parallel with the action on the bottom rows. So the action of $GL(n, C)$ is *much, much* more complicated than the action of a general invariant homomorphism.

31.7 A Theorem of Schur and Weyl

The following result was first proved by Schur. Weyl recognized its importance and used it extensively in his work on the Classical Groups. Since $\text{Hom}_{\mathcal{S}_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$

is straightforward and easy, and the action of G is rich and complicated, the result at first sounds impossible. But it is true.

Theorem 141 (Schur-Weyl) *Every element of $\text{Hom}_{S_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ is a finite linear combination of maps $A \otimes A \otimes \dots \otimes A$ induced by $A \in GL(n, C)$.*

Proof: This result is surprising because if we write $V \otimes \dots \otimes V = W_1 \oplus \dots \oplus W_m$, then the map given by A on W_1 and zero on all other pieces commutes with S_k and certainly is not $A \otimes \dots \otimes A$.

There are many proofs of this result, most very algebraic. The following proof is given by Shlomo Sternberg in his book on Group Theory and Physics.

We first prove a supplemental lemma:

Lemma 61 *If V is a finite dimensional complex vector space, the space $S^k(V)$ of symmetric k -tensors is generated by tensors of the form $v \otimes v \otimes \dots \otimes v$.*

Remark: This looks abstract, but it is really a very concrete result. If $e_1, \dots, e_n$ is a basis of V , then the $e_1^{i_1} e_2^{i_2} \dots e_n^{i_n}$ with $\sum i_j = k$ form a basis for $S^k(V)$. Suppose $n = k = 2$. Then a basis for S^2 is $e_1^2, e_1 e_2, e_2^2$. The first and last have the form $v \otimes v$. The middle term is a linear combination of such terms because $(e_1 + e_2)^2 - e_1^2 - e_2^2 = 2e_1 e_2$. The lemma says that this works in general, but a brute force calculation would be messy.

Proof: Notice that $v \otimes \dots \otimes v$ is already in S^k and the S operator is the identity on such elements. Let $v = a_1 e_1 + \dots + a_n e_n$. Then

$$v \otimes \dots \otimes v = \left(\sum a_i e_i \right) \otimes \dots \otimes \left(\sum a_i e_i \right)$$

which equals

$$\sum_{k_1 + \dots + k_n = k} \frac{k_1! k_2! \dots k_n!}{k!} a_1^{k_1} \dots a_n^{k_n} e_1^{k_1} \dots e_n^{k_n}$$

Let $W \subset S^k$ be the subspace generated by all such elements. Since W is a subspace, it is closed under limits, and it follows that $(\partial/\partial a_1)^{k_1} \dots (\partial/\partial a_n)^{k_n}$ applied to this element is also in W . But this expression is a non-zero multiple of $e_1^{k_1} \dots e_n^{k_n}$. Indeed when the operator is applied to this particular monomial, it kills off all of the a_i coefficients, but when it is applied to some other monomial, some a_i are differentiated less often than needed, but others are differentiated more often than needed, killing off the term completely.

It follows that W contains a basis for the full $S^k(V)$ and thus equals $S^k(V)$. QED.

Proof of Schur-Weyl: We want to study elements of $\text{Hom}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ which commute with the action of the symmetric group. Ignore this action for a moment and consider the full set of homomorphisms. Recall that there is a canonical isomorphism

$V^* \otimes V \simeq \text{Hom}(V, V)$. Using this isomorphism, we have

$$\begin{aligned} \text{Hom}(V, V) \otimes \dots \otimes \text{Hom}(V, V) &\simeq (V^* \otimes V) \otimes \dots \otimes (V^* \otimes V) \\ &\simeq (V^* \otimes \dots \otimes V^*) \otimes (V \otimes \dots \otimes V) \\ &\simeq V \otimes \dots \otimes V)^* \otimes (V \otimes \dots \otimes V) \\ &\simeq \text{Hom}(V \otimes \dots \otimes V, V \otimes \dots \otimes V) \end{aligned}$$

It is easy to check that this isomorphism carries $A_1 \otimes \dots \otimes A_k$ to the map

$$v_1 \otimes \dots \otimes v_k \rightarrow A_1(v_1) \otimes \dots \otimes A_k(v_k)$$

Now $\text{Hom}_{S_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ is the set of homomorphisms which commute with the action of S_k on $V \otimes \dots \otimes V$. We can define an action of S_k on $\text{Hom}(V \otimes \dots \otimes V, V \otimes \dots \otimes V)$ by sending Y to sYs^{-1} . The set which interests us consists of all homomorphisms left fixed by this action. Indeed $sY = Ys$ if and only if $sYs^{-1} = Y$.

The group S_k also acts on $\text{Hom}(V, V) \otimes \dots \otimes \text{Hom}(V, V)$ by sending $A_1 \otimes \dots \otimes A_k$ to $A_{s^{-1}(1)} \otimes \dots \otimes A_{s^{-1}(k)}$. We claim that the above isomorphism sends one action to the other. Indeed

$$[s(A_1 \otimes \dots \otimes A_k)]v_1 \otimes \dots \otimes v_k = [A_{s^{-1}(1)} \otimes \dots \otimes A_{s^{-1}(k)}]v_1 \otimes \dots \otimes v_k = A_{s^{-1}(1)}v_1 \otimes \dots \otimes A_{s^{-1}(k)}v_k$$

and

$$\begin{aligned} s[(A_1 \otimes \dots \otimes A_k)(s^{-1}(v_1 \otimes \dots \otimes v_k))] &= s(A_1(v_{s(1)}) \otimes \dots \otimes A_k(v_{s(k)})) \\ &= A_{s^{-1}(1)}(v_{s^{-1}s(1)}) \otimes \dots \otimes A_{s^{-1}(k)}(v_{s^{-1}s(k)}) = A_{s^{-1}(1)}v_1 \otimes \dots \otimes A_{s^{-1}(k)}v_k \end{aligned}$$

It follows that $\text{Hom}_{S_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ is mapped by our isomorphism to the elements of $\text{Hom}(V, V) \otimes \dots \otimes \text{Hom}(V, V)$ invariant under all elements of S_k , that is, $S^k(\text{Hom}[V, V])$. By the preliminary lemma of this section, this space is spanned by all $A \otimes A \otimes \dots \otimes A$ for $A \in \text{Hom}(V, V)$. Let $W \subset S^k(\text{Hom}(V, V))$ be the subspace spanned by all $A \otimes A \otimes \dots \otimes A$ for A non-singular, that is, $A \in GL(n, C)$. Since this is a subspace of our vector space, it is closed under limits. But every matrix can be approximated arbitrarily closely by a non-singular matrix. Consequently, the space spanned by all $A \otimes A \otimes \dots \otimes A$ for $A \in GL(n, C)$ is the same as the space spanned by all such elements for arbitrary $A \in \text{Hom}(V, V)$. QED.

Remark: In the introductory section we studied the dual actions of $SU(n)$ and S_k on $V \otimes \dots \otimes V$. That entire section works without change if we replace $SU(n)$ with $GL(n, C)$. Thus we can write

$$V \otimes \dots \otimes V = W_1 \oplus \dots \oplus W_m$$

where each W_i is a sum of equivalent representation subspaces for a particular irreducible representation of the symmetric group $\mathcal{S}_k$, and also a sum of equivalent representations of $GL(n, C)$.

The theorem of Schur-Weyl will enable us to complete this picture when the group is $GL(n, C)$. We'll do that shortly. It turns out that the theorem of Schur-Weyl is also true for $SU(n)$, although the proof is a little harder. So the picture we are developing works in both cases.

Before finishing the story for $GL(n, C)$, let us look back at what has been done.

31.8 Consequences of the Schur-Weyl Theorem for $GL(n, C)$

Theorem 142 *Fix W and the basis $t_1, \dots, t_s$ for a subspace of this space. The representation of $GL(n, C)$ on this subspace is irreducible. Representations of $GL(n, C)$ obtained in this way from W_i and W_j with $i \neq j$ are inequivalent.*

Suppose the action of $GL(n, C)$ on the space with basis $t_1, \dots, t_s$ is reducible. Then there is a different basis $u_1, \dots, u_s$ for this space such that a non-trivial subspace with basis $u_1, \dots, u_t$ is invariant. So each matrix of the representation has a block of zeros in the lower left. Therefore every linear combination of such matrices has the same block of zeros. These linear combinations give all of $Hom_{\mathcal{S}_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ by the Schur-Weyl theorem. But we know that this set contains *all* linear homomorphisms of the space generated by $u_1, \dots, u_s$, since that is also the space generated by $t_1, \dots, t_s$. This contradiction proves the first half of the theorem.

If $i \neq j$ and the representation using the bottom row of W_i is equivalent to the representation using the bottom row of W_j , then there is a linear isomorphism S from the bottom row in W_i to the bottom row in W_j commuting with all $A \in GL(n, C)$. This map would then commute with all linear combinations of these A and thus with the entire $Hom_{\mathcal{S}_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$. But this set contains a homomorphism T which is the identity on W_i and zero on W_j , and so $S = S \circ T = T \circ S = 0$, a contradiction.

31.9 Schur-Weyl for $SL(n, C), GL(n, R), SL(n, R), U(n)$, and $SU(n)$

We now claim that the results in the first two sections remain true if we replace the group $GL(n, C)$ with any of $SL(n, C), GL(n, R), SL(n, R), U(n)$, or $SU(n)$. The proofs remain the same once we prove one key result.

Theorem 143 (Schur-Weyl) *Let G be $GL(n, C), SL(n, C), GL(n, R), SL(n, R), U(n)$, or $SU(n)$. Then every element of $Hom_{\mathcal{S}_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$ is a finite complex linear*

combination of maps $A \otimes \dots \otimes A$ induced by $A \in G$.

Proof: The result is clear for $SL(n, C)$, $GL(n, R)$ and $SL(n, R)$, since the complex linear span of elements of each of these groups equals the complex linear span of elements of $GL(n, C)$. The same proof can be given for $SU(n)$ once we know the result for $U(n)$. So we are reduced to proving this one case of the theorem. Our proof is taken from a paper by Guillaume Aubrun.

Let $A \in GL(n, C)$. It suffices to prove that $A \otimes \dots \otimes A$ can be written as a limit of terms that are linear combinations of elements $U \otimes \dots \otimes U$, where U is unitary. Indeed, the linear span of terms of the form $U \otimes \dots \otimes U$ for unitary U would then contain all $A \otimes \dots \otimes A$, and then by the Schur-Weyl theorem proved earlier, would be all of $Hom_{S_k}[V \otimes \dots \otimes V, V \otimes \dots \otimes V]$.

By a result in section 2.4, we can write $A = UH$ where $U \in U(n)$ and H is Hermitian with positive real eigenvalues. Every Hermitian matrix can be diagonalized via some orthonormal basis, so we can find a unitary matrix U_1 such that $U_1 H U_1^{-1}$ is diagonal with positive real terms on the diagonal. It follows that $A = U_1 H U_1^{-1} U = U_1 H U_2$ with U_i diagonal. Let us write $A = U_1 H(s_1, \dots, s_n) U_2$ where the middle term stands for a diagonal matrix with real diagonal entries $s_1, \dots, s_n$. Multiplying A by a positive real multiple, we can assume that all s_i satisfy $-1 < s_i < 1$. In the end, we will prove that this positive multiple has the required property, and then of course A itself also does.

Notice that if we replace each s_i with a complex z_i of absolute value one, we obtain a unitary matrix.

Select a basis $e_1, \dots, e_n$ for V . We obtain an induced basis $e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_k}$ for $V \otimes \dots \otimes V$. With respect to this basis, $A \otimes \dots \otimes A$ is represented by a giant matrix whose entries are polynomials in $s_1, \dots, s_n$. Call this matrix $A_{s_1, \dots, s_n}^{\otimes k}$.

Let γ be a counterclockwise unit circle in the complex plane and recall that

$$\frac{1}{2\pi i} \int_{\gamma} \frac{z^k}{z-s} dz = s^k$$

Integrate each term of this matrix to obtain

$$\left(\frac{1}{2\pi i} \right)^n \int_{\gamma} \dots \int_{\gamma} A_{z_1, \dots, z_n}^{\otimes k} \frac{1}{z_1 - s_1} \dots \frac{1}{z_n - s_n} dz_1 \dots dz_n = A_{s_1, \dots, s_n}^{\otimes k} = A \otimes \dots \otimes A$$

For each fixed $z_1, \dots, z_n$, the expression in the integral sign is a unitary matrix of the form $U \otimes \dots \otimes U$, multiplied by a fixed complex number $\frac{1}{z_1 - s_1} \dots \frac{1}{z_n - s_n}$. If we replace the integral with a Riemann sum, the sum is a linear combination of maps of the form $U \otimes \dots \otimes U$ for $U \in U(n)$. It follows that $A \otimes \dots \otimes A$ is a limit of such combinations. QED.

31.10 A Basis for the Image of C_λ

The tensor space $C_\lambda(V \otimes \dots \otimes V)$ is a complicated object because the subgroups P and Q used in its definition do not commute. Consequently, tensors in this space aren't necessarily symmetric in their rows and skew in their columns. In fact, if σ is a permutation of the rows, then $c_\lambda \sigma = c_\lambda$, and if σ is a permutation of the columns, then $\sigma c_\lambda = \text{sgn}(\sigma) c_\lambda$. It follows that if $w = C_\lambda(v)$ and σ is a permutation of the rows, then $w = C_\lambda(\sigma v)$. This can be used to greatly restrict the set of v 's we use to construct the image of C_λ . Similarly, if σ is a permutation of the columns, then $\sigma(C_\lambda(v)) = \text{sgn}(\sigma) C_\lambda(v)$, so tensors in the image of C_λ are automatically skew in their columns.

Using the easiest possible case, we show how these facts can be used to construct a basis for the image of C_λ . The case we study is the subspace of $V \otimes V \otimes V$ corresponding to the Young diagram



Let $e_1, e_2, \dots, e_n$ be a basis of V . Then the $e_{i_1} \otimes e_{i_2} \otimes e_{i_3}$ form a basis for $V \otimes V \otimes V$, and so the $C_\lambda(e_{i_1} \otimes e_{i_2} \otimes e_{i_3})$ generate the image of C_λ . We will prove that a basis is given by the set $\mathcal{B} = \{ C_\lambda(e_{i_1} \otimes e_{i_2} \otimes e_{i_3}) \mid i_1 \leq i_2 \text{ and } i_1 < i_3 \}$.

It is tempting to give a glib proof of this. Since $C_\lambda(v) = C_\lambda(\sigma v)$ whenever σ permutes rows, $e_{i_1} \otimes e_{i_2} \otimes e_{i_3}$ and $e_{i_2} \otimes e_{i_1} \otimes e_{i_3}$ have the same image, so we may as well assume that $i_1 \leq i_2$. That much is correct. But the corresponding argument for $i_1 < i_3$ fails because $C_\lambda(\sigma v) = \sigma C_\lambda(v)$ is false for column permutations.

However, it is encouraging that the result predicts the correct dimension for the space. Earlier we discovered that this space must have dimension $\frac{(n-1)n(n+1)}{3}$. Suppose $i_1 = k$. Then i_2 can have any of the $n - k + 1$ values $k, k + 1, \dots, n$ and i_3 can have any of the $n - k$ values $k + 1, k + 2, \dots, n$. Thus the total number of elements is

$$\begin{aligned} \sum_{k=1}^n (n - k + 1)(n - k) &= \sum_{k=1}^n (n - k)^2 + \sum_{k=1}^n (n - k) = \sum_{k=0}^{n-1} k^2 + \sum_{k=0}^{n-1} k = \\ &= \frac{(n-1)n(2(n-1)+1)}{6} + \frac{(n-1)n}{2} = \\ &= \frac{n}{6} \left((n-1)(2n-1) + 3n-3 \right) = \frac{n}{6} (2n^2 - 2) = \frac{n(n-1)(n+1)}{3} \end{aligned}$$

Since the dimension is correct, it suffices to prove that the images listed generate the image of C_λ . To do this, we order basis vectors of $V \otimes \dots \otimes V$ lexicographically. When comparing $e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_k}$ and $e_{j_1} \otimes e_{j_2} \otimes \dots \otimes e_{j_k}$, we say the first of these is smaller if $i_1 < j_1$, or else if $i_1 = j_1$ and $i_2 < j_2$, or else ...

Suppose we have an element in the image of C_λ and write this image as a linear combination of the above basis vectors of $V \otimes V \otimes V$. Order these basis vectors. We will prove that by subtracting an appropriate multiple of one of the vectors in $\mathcal{B}$, we can obtain an element of the image of C_λ whose smallest term is larger than before. Continuing the process, we eventually generate the element as a linear combination of images of vectors in $\mathcal{B}$.

Let us denote the element $b = e_{i_1} \otimes e_{i_2} \otimes e_{i_3}$ by

$$\begin{array}{|c|c|} \hline i_1 & i_2 \\ \hline i_3 & \\ \hline \end{array}$$

Then $C_\lambda(b)$ is

$$\begin{array}{|c|c|} \hline i_1 & i_2 \\ \hline i_3 & \\ \hline \end{array} + \begin{array}{|c|c|} \hline i_2 & i_1 \\ \hline i_3 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline i_3 & i_2 \\ \hline i_1 & \\ \hline \end{array} - \begin{array}{|c|c|} \hline i_3 & i_1 \\ \hline i_2 & \\ \hline \end{array}$$

Notice that, as predicted, this tensor is skew in the columns, but not symmetric in the rows. If $b \in \mathcal{B}$, then $i_1 \leq i_2$ and $i_1 < i_3$ and it follows that the first element in $C_\lambda(b)$ is smaller than all remaining elements unless $i_1 = i_2$, and in that case the first two elements are equal. So if we have a sum of terms in the image of C_λ and b is the smallest term in this expression, then subtracting an appropriate scalar multiple of $C_\lambda(b)$ will remove this smallest term entirely. As explained at the end of the previous paragraph, this completes the proof that images of elements in $\mathcal{B}$ generate the image of C_λ .

Remark: In this special case, a dimension argument shows that the elements we obtained must be linearly independent in the image of C_λ . But our method also proves this directly. Each basis vector induces an expression with a different smallest term, so a dependence relation will imply that the coefficient of the basis vector with the minimal smallest term must vanish, and then the next one must vanish, etc.

Remark: The special case we have just proved is true in general, and with the same proof. The proof is somewhat messy to write down, but no unexpected surprises arise in the process. Here is the theorem.

Theorem 144 *Let λ be a Young diagram. Fill the boxes in this diagram with basis vectors e_{i_j} of V , where repetitions are allowed and not all basis vectors need occur. Let each such filled diagram represent an element of $V \otimes \dots \otimes V$ by just reading the boxes from left to right, and then from top to bottom, and forming the tensor product of the terms. Let $\mathcal{B}$ be the set of such filled diagrams such that across the rows the basis vectors remain the same or increase, and down the columns the basis vectors strictly increase. Apply C_λ to each tensor in $\mathcal{B}$. The resulting tensors form a basis of the image of C_λ .*

Remark Added Later: Actually the proof is quite easy using the following observations. Suppose a term contains certain basis vectors e_{i_j} in Young boxes. Any other term in the Young symmetrization of this term will contain the same elements permuted in some way. The lexicographically smallest permutation of the terms using only row permutations and

column permutations in any order is the one selected by our rules for $\mathcal{B}$. To finish the argument, we must prove that when we symmetrize this element of $\mathcal{B}$, all terms with this lexicographically smallest form have positive coefficients rather than negative coefficients, so this term does not cancel out under Young symmetrization. It is relatively easy to prove this lemma. QED.

Remark: Notice that as a consequence we can compute the dimension of the representation associated with a Young diagram λ . The calculation will be straightforward but messy. Later we'll state a beautiful formula for the result.

Remark: Notice that $\mathcal{B}$ is empty exactly when the number of rows in the Young diagram is larger than n , the dimension of V . This proves

Theorem 145 *An irreducible representation of S_k appears in the decomposition of*

$$V \otimes \dots \otimes V$$

if and only if the number of rows in the Young diagram is at most the dimension of V . Thus an irreducible representation of $G = GL(n, C), SL(n, C), U(n), SU(n)$ appears in the decomposition if and only if the number of rows in the Young diagram is at most n .

31.11 A Calculation Given Without Proof

One goal of these notes is to prove all assertions made. This section contains one exception, which we will not use in the remaining notes. If laziness can be overcome, a proof might be provided at a later date.

Theorem 146 *Consider the irreducible representation of G associated with a Young diagram λ . Let $\text{Hook}(\lambda)$ be the positive integer given by the Hook formula, described in the chapter on representations of the symmetric group. Fill each box of the Young diagram with an integer in the following way: put $n = \dim V$ in the upper left box, and fill the first column with the integers $n, n-1, n-2, \dots$. Then fill each row with increasing integers; the first row has $n, n+1, n+2, \dots$, the second row has $n-1, n, n+1, \dots$ and so forth. The dimension of the associated irreducible representation of G is*

$$\frac{\prod(\text{numbers in boxes})}{\text{Hook}(\lambda)}$$

Example: Consider the representation associated with



The hook length of the top left corner is 3, and all other boxes have hook length 1, so $Hook(\lambda) = 3$. When we fill the boxes as indicated, we obtain

n	$n + 1$
$n - 1$	

The dimension of the representation is thus

$$\frac{(n-1)n(n+1)}{3}$$

Incidentally, when $n = 3$ and we are dealing with $SU(3)$, this number is 8, and our representation is the representation used by the physicists for the 8-fold way.

Hint of Possible Proof: Notice that

$$\frac{\prod(\text{numbers in boxes})}{Hook(\lambda)} = \frac{\prod(\text{numbers in boxes})}{n!} \cdot \frac{n!}{Hook(\lambda)}$$

The second element of this product is the dimension of the irreducible representation of $\mathcal{S}_k$ associated with λ . The first element of the product does not involve the Hook formula. So a natural approach proving the result would be to find a nice combinatorial object on which $\mathcal{S}_k$ acts and break up the combinatorial object into subsets on which $\mathcal{S}_k$ acts simply transitively and finally count the number of such subsets.

31.12 Highest Weights

We already know that the irreducible representations of $SU(n)$ are completely determined by their highest weights. In this section, we compute the highest weights of the representations obtained from Young diagrams. It will turn out that every possible highest weight occurs in this way, and therefore every irreducible representation of $SU(n)$ is obtained from a Young diagram by the method described earlier.

The maximal torus of $SU(n)$ is the set of all diagonal matrices

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

where each $|\lambda_i| = 1$ and the product of the λ_i is also 1. This element acts on $e_{i_1} \otimes \dots \otimes e_{i_n}$ by multiplying it by $\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_n^{m_n}$, where the basis element contains m_1 1's, $\dots$, and m_n n 's. This coefficient does not depend on the order of the terms, so either

$$C_\lambda(e_{i_1} \otimes \dots \otimes e_{i_n})$$

is zero, or else it is a weight vector with weight $\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_n^{m_n}$. Since these elements generate the full representation space, this is a list of all possible weights.

A particularly interesting weight vector is obtained by filling the Young diagram with constant rows and columns which increase by one:

e_1	e_1	e_1	e_1	e_1
e_2	e_2	e_2		
e_3	e_3	e_3		
e_4				

Notice that elements of this form are part of the basis obtained earlier, and thus certainly non-zero. They have weights

$$\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_n^{m_n}$$

where $m_1 \geq m_2 \geq \dots \geq m_n \geq 0$. The m_i here are the lengths of the rows of the Young diagram. These will turn out to be the highest weights. Before showing this, we need to clarify one issue.

Since we are dealing with $SU(n)$, the product of the λ_i is 1, and thus $\lambda_n = \lambda_1^{-1} \dots \lambda_{n-1}^{-1}$. Thus the weight assigned to our element ought to be written

$$\lambda_1^{m_1-m_n} \lambda_2^{m_2-m_n} \dots \lambda_{n-1}^{m_{n-1}-m_n}$$

Notice that these coefficients still decrease and are still greater than or equal to zero. However, we can now get the same highest weight in several different ways, since

$$(m_1, m_2, \dots, m_{n-1}, 0)$$

$$(m_1 + 1, m_2 + 1, \dots, m_{n-1} + 1, 1)$$

and etc. all give the same highest weight.

It will ultimately turn out that the various representations listed above are different on $U(n)$ but become identical when restricted to $SU(n)$. We will fix this by restricting to Young diagrams with at most $n - 1$ rows when dealing with $SU(n)$. The reader may find this slightly confusing because when we decomposed the representation on $V \otimes V \otimes \dots \otimes V$ into representations given by Young diagrams, we proved that the representations attached to different W_i are distinct. However, in the case currently being discussed, the different representations come from taking tensor products of different numbers of V and then decomposing.

Here is our ultimate theorem:

Theorem 147 *There is a one-to-one correspondence between irreducible representations of $SU(n)$ and Young diagrams with at most $n - 1$ rows. Given a diagram with m_1 boxes on the first row, m_2 boxes on the second row, etc., the corresponding highest weight is*

$$\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_{n-1}^{m_{n-1}}$$

Note that

$$m_1 \geq m_2 \geq \dots \geq m_{n-1}$$

Proof: Applying C_λ to the term

e_1	e_1	e_1	e_1	e_1
e_2	e_2	e_2		
e_3	e_3	e_3		
e_4				

gives a non-zero element by our result giving a basis for the image of C_λ . This image element is a weight vector with the indicated weight. We want to prove that it is a highest weight, that is, that adding a positive root to this weight is never another weight. A calculation at the end of the discussion then shows that these highest weights give all possible highest weights, and we will be done.

Our first job is to find the roots in this language. If G is any Lie group, G acts on itself by conjugation, and this induces an action of G on its Lie algebra $\mathcal{G}$ by conjugation. The roots of a compact G are the weights of this conjugate action extended to $\mathcal{G} \otimes C$. This extension is important because we only deal with representations on complex vector spaces.

The Lie algebra of $SU(n)$ is the set of all complex matrices A with zero trace such that $-\bar{A}^T = A$. Conjugation leaves the diagonal elements of these matrices fixed. The remaining elements of the algebra $su(n)$ are generated by complex matrices which are zero except in the ij -th and ji -th entries, for $i < j$, where $a_{ji} = -\bar{a}_{ij}$. A brief calculation shows that conjugation multiplies the ij -element by $\lambda_i \lambda_j^{-1}$ and multiplies the ji -element by $\lambda_j \lambda_i^{-1}$. The new element is again in the Lie algebra because $|\lambda_i| = 1$ and so $\bar{\lambda}_i = \lambda_i^{-1}$:

$$-\overline{\lambda_i \lambda_j^{-1} a_{ij}} = -\lambda_i^{-1} \lambda_j \bar{a}_{ij} = \lambda_j \lambda_i^{-1} a_{ji}$$

Note that this a_{ij} matrix is not a root vector, because conjugation multiplies the top-right term by one factor and the bottom-left term by the inverse of that factor. But this is what we should expect, because in our general theory of compact Lie groups, the roots corresponded to three-dimensional subalgebras isomorphic to $su(2)$. That general theory implies that the root vectors will be in the complexification $\mathcal{G} \otimes C$. Indeed, the matrix with a 1 in the ij spot and zeros everywhere else is a root vector with root $\lambda_i \lambda_j^{-1}$ and the matrix with a 1 in the ji spot and zeros everywhere else is a root vector with root $\lambda_j \lambda_i^{-1}$. Both matrices are in the complexification of $su(n)$.

Looking back at the chapter on irreducible representations of compact groups, we find that our notation is considerably different than the notation in that chapter. So we translate. In that chapter, we considered the universal cover of the maximal torus and identified it with the Lie algebra $\mathcal{T}$ of the torus. We identified weights with linear maps $\mathcal{T} \rightarrow \mathcal{R}$ via the formula $\varphi(t)e_w = e^{2\pi i w(t)}e_w$. So let us write

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = \begin{pmatrix} e^{2\pi i t_1} & & & \\ & e^{2\pi i t_2} & & \\ & & \ddots & \\ & & & e^{2\pi i t_n} \end{pmatrix}$$

where now the Lie algebra of the torus has coordinates $(t_1, t_2, \dots, t_n)$ and $\sum t_i = 0$. Then

$$\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_n^{m_n} = e^{2\pi i (m_1 t_1 + m_2 t_2 + \dots + m_n t_n)}$$

and so in our previous notation the weight is the map

$$(t_1, \dots, t_n) \rightarrow m_1 t_1 + \dots + m_n t_n$$

Since $\sum t_i = 0$, we can also write this

$$(t_1, \dots, t_{n-1}) \rightarrow (m_1 - m_n)t_1 + \dots + (m_{n-1} - m_n)t_{n-1}$$

which agrees with an earlier observation at the start of this section.

A similar remark applies to the roots. We have

$$\lambda_i \lambda_j^{-1} = e^{2\pi i (t_i - t_j)}$$

and so in our previous notation the corresponding root is

$$(t_1, \dots, t_n) \rightarrow (t_i - t_j)$$

Recall that we then separate the roots into positive and negative roots, and the positive roots were

$$\alpha_{ij} = t_i - t_j \quad \text{for} \quad i < j$$

We are now in a position to prove that $m_1 t_1 + m_2 t_2 + \dots + m_{n-1} t_{n-1}$ is a highest weight. Suppose we add $t_i - t_j$ to this element. The resulting weight vector could be written uniquely in terms of the basis vectors we earlier obtained from the representation space. In particular, the lexicographically earliest term in the weight vector would have to come from the basis vector associated with m_1 copies of e_1 , etc., and in particular $m_i + 1$ copies of e_i . But there is no such basis vector, because all m_1 copies of e_1 would have to be in the first row, all m_2 copies of e_2 would be in the second row, etc., and finally the i th row

would be filled with m_i copies of e_i but there is a left over e_i . This would have to go in the next row, but that is impossible because the entries in the columns strictly increase.

Remark: We are almost, but not quite, done. In the general theory, we associated with each highest weight a series of non-negative integers, one attached to each simple root. These integers completely classify irreducible representations. At this point we know all highest weights, but we don't know how they correspond to these integers. Recall that the simple roots, in the case of $SU(n)$, are $\alpha_1 = t_1 - t_2, \alpha_2 = t_2 - t_3, \dots, \alpha_{n-1} = t_{n-1} - t_n$.

If w is a highest weight, we attach to each α_i the integer

$$\frac{2w(t_{\alpha_i})}{\langle t_{\alpha_i}, t_{\alpha_i} \rangle}$$

Here $\langle \cdot, \cdot \rangle$ is an invariant metric on $su(n)$, which is unique up to a positive scalar, and t_{α_i} is the element of the torus dual to the root. This integer is independent of the positive scalar factor. We proved that $-tr(XY)$ is an invariant metric on $su(n)$. Since elements of $su(n)$ satisfy $A = -\bar{A}^T$, diagonal entries of the matrices are purely imaginary and an element $(t_1, \dots, t_n) \in \mathcal{T}$ is given by the matrix

$$\begin{pmatrix} it_1 & & & \\ & it_2 & & \\ & & \ddots & \\ & & & it_n \end{pmatrix}$$

But then the invariant inner product of $(t_1, t_2, \dots, t_n)$ and $(u_1, u_2, \dots, u_n)$ is $\sum t_i u_i$. So our invariant metric is the ordinary Euclidean metric.

But then the t_i form an orthonormal basis and the dual vector has the coordinates of the original vector. In particular, if $\alpha_i(t) = t_i - t_{i+1}$, then $t_{\alpha_i} = e_i - e_{i+1}$. Notice that $\langle \alpha_i, \alpha_i \rangle = \langle e_i - e_{i+1}, e_i - e_{i+1} \rangle = 2$ and consequently

$$w(t_{\alpha_i}) = (m_1 t_1 + \dots + m_n t_n)(e_i - e_{i+1}) = m_i - m_{i+1}$$

Theorem 148 *The representation associated with a Young diagram whose first row has m_1 boxes, second row has m_2 boxes, and so forth, until the n th row has $m_n = 0$ boxes, has a highest weight $\lambda_1^{m_1} \lambda_2^{m_2} \dots \lambda_{n-1}^{m_{n-1}}$, and corresponds to the assignment to each point in the Dynkin diagram below of integers $m_1 - m_2, m_2 - m_3, \dots, m_{n-1} - m_n$. Conversely, given $n - 1$ non-negative integers, we can uniquely find the related integers $m_1, \dots, m_{n-1}$.*

$$\mathbf{A}_n \quad \bullet \text{---} \bullet \text{---} \bullet \quad \dots \quad \bullet \text{---} \bullet \quad n \geq 1$$

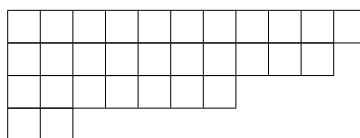
Remark: Note that $SU(n)$ for $n \geq 2$ corresponds to type A_{n-1} for $n - 1 \geq 1$. The identity representation attaches 0 to every node.

31.13 Examples of Irreducible Representations of $SU(n)$

The easy way to start with a labeled Dynkin diagram with labels $(L_1, L_2, \dots, L_{n-1})$ and obtain the corresponding Young diagram is to begin from the right side of the labels and work left, while starting at the bottom of the Young diagram and working up. There will be L_{n-1} boxes on the bottom row. Then L_{n-2} determines how many *additional* boxes will be on the next higher row, etc.

Another way to think of this is that the sum of the labels gives the number of boxes on the top row, the sum of all labels except the first gives the number of boxes on the second row, and so forth.

So $(1, 3, 5, 2)$ gives the Young diagram



In particular, this allows us to read off the representations corresponding to a single non-zero label. So

$$(1, 0, 0, 0, 0) \leftrightarrow \begin{array}{|c|} \hline \square \\ \hline \end{array} \leftrightarrow \Lambda^1(V)$$

$$(0, 1, 0, 0, 0) \leftrightarrow \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} \leftrightarrow \Lambda^2(V)$$

$$(0, 0, 1, 0, 0) \leftrightarrow \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \leftrightarrow \Lambda^3(V)$$

$$(0, 0, 0, 1, 0) \leftrightarrow \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \leftrightarrow \Lambda^4(V)$$

$$(0, 0, 0, 0, 1) \leftrightarrow \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \leftrightarrow \Lambda^5(V)$$

In general, the dot on the right of the Dynkin diagram for $SU(n)$ gives $\Lambda^{n-1}(V)$. As explained earlier, we could continue to the final $\Lambda^n(V)$, but that would be the one dimensional determinant representation, and it is trivial for $SU(n)$. In our scheme, the trivial representation is obtained by assigning zero to every node, and we don't want to list it more than once.

Notice that $(5, 0, 0, 0, 0)$ corresponds to $\square\square\square\square\square$ and to $S^5(V)$, and in general $(m, 0, 0, 0, 0)$ corresponds to a Young diagram with a single row of length m , and thus to the space of symmetric tensors $S^m(V)$.

31.14 Weyl's Unitary Trick

The previous sections complete our discovery of a large number of irreducible representations for each of the groups $GL(n, C)$, $SL(n, C)$, $GL(n, R)$, $SL(n, R)$, $U(n)$, $SU(n)$. The first two groups are complex analytic manifolds and their representations are complex analytic. The remaining groups are real manifolds and their representations are representations of real Lie groups, but still acting on C^n by complex linear maps.

When we talk about representations of these groups, we always speak of complex analytic representations of the first two groups, and ordinary Lie representations of the remaining groups.

Recall that Lie theory reduces questions about representations of Lie groups to questions about representations of their Lie algebras. In particular, Hermann Weyl used this idea to prove

Theorem 149 (Weyl's Unitary Trick) *If the Lie algebra of a connected Lie group is semisimple, then every representation of the group is equivalent to a direct sum of irreducible representations.*

Proof: Let $\rho : G \rightarrow GL(n, C)$ be such a representation. It induces a corresponding algebra representation $\rho^* : g \rightarrow gl(n, C)$. This is linear over the reals, but it induces a corresponding map linear over the complex numbers: $\rho^* \otimes C : g \otimes C \rightarrow gl(n, C)$. A deep theorem in the classification of complex Lie algebras states that whenever g is a real semisimple algebra, there is a second real semisimple algebra with negative-definite Killing form, g_1 , so $g \otimes C$ and $g_1 \otimes C$ are isomorphic. Thus $\rho \otimes C : g_1 \otimes C \rightarrow gl(n, C)$ exists, and restriction to real scalars gives $g_1 \rightarrow gl(n, C)$. We proved that if g_1 is a semisimple algebra with negative-definite Killing form, then g_1 is the Lie algebra of a compact Lie group. A difficult theorem states that the universal cover of this compact Lie group is compact. In this way, the original representation has induced a representation of a simply-connected compact Lie group. We proved that all such representations can be written as direct sums of irreducible representations. Tracing the process back through groups and algebras then decomposes the original representation into a direct sum of irreducible representations.

Remark: We don't need the two difficult steps in this proof because we only apply this argument to the groups on our list, where we can directly check all the assertions. Notice

that the Lie algebras of the groups on our list are

$$\begin{array}{ll}
 GL(n, C) & C \oplus sl(n, C) \\
 SL(n, C) & sl(n, C) \\
 GL(n, R) & R \oplus sl(n, R) \\
 SL(n, R) & sl(n, R) \\
 U(n) & R \oplus su(n) \\
 SU(n) & su(n)
 \end{array}$$

Moreover, $sl(n, R) \otimes C$ and $su(n) \otimes C$ are isomorphic to $sl(n, C)$.

Consequently, we know the complete story of representation theory for $SL(n, C)$, $SL(n, R)$, and $SU(n)$. In each of these cases, every representation can be written uniquely up to order as a sum of irreducible representations, and these irreducible representations are described by exactly the same data that we developed for $SU(n)$.

Since $U(n)$ is a compact group, it also satisfies the theorem that every representation can be written uniquely up to order as a sum of irreducible representations. The Lie algebra of $U(n)$ is a sum of two ideals, and each of these ideals corresponds to a closed subgroup of the main group. The left term corresponds to the set of constant diagonal matrices λI where $|\lambda| = 1$, and the right side corresponds to $SU(n)$. Notice that the group of all λI is just the circle group S^1 .

If ρ is a representation of $U(n)$ on a vector space V , it restricts to a representation of S^1 . This representation can be written as a sum of irreducible representations of S^1 , which are all one dimensional. Indeed, there is one such representation for each integer k , namely

$$\lambda \rightarrow \lambda^k$$

Collecting equivalent representations together, we can write $V = W_1 \oplus \dots \oplus W_k$, a trick we used earlier. Because the groups S^1 and $SU(n)$ commute, each element T of $\rho(SU(n))$ is an intertwining operator. Moreover, each such element acting on $W_1 \oplus \dots \oplus W_k$ can be thought of as a matrix T_{ij} mapping $W_i \rightarrow W_j$. By the usual argument from Schur's lemma, these maps are zero when $i \neq j$, so each W_i is invariant under the action of $\rho(SU(n))$ and consequently under the full action of $U(n)$. In particular, if ρ is an irreducible representation of $U(n)$, there is only a single W and the diagonal matrices λI act on the representation space by multiplication by λ^k for some fixed integer k .

The irreducible representations of $SU(n)$ are known, and associated with Young diagrams. Each of these representation spaces is actually invariant under the full $U(n)$ using the same Young procedure. It follows that every irreducible representation of $U(n)$ equals a representation given by a Young diagram with at most $n - 1$ rows, except that if $A \in U(n)$ does not have determinant 1, then A equals $(\lambda I)B$ for some B , and then the action of A equals the action of B multiplied by λ^k .

Notice that $\det A = \lambda^n$ and so another way to state this result is that ρ is determined by giving a Young diagram with at most $n - 1$ rows and an extra integer k . The representation is the standard representation of $U(n)$ associated with this Young diagram, multiplied by $(\det A)^{k/n}$.

However, there is a problem because $\det A$ can be any complex number of absolute value one, and taking the n th root is multiple valued, and it is not possible to select continuously one of these choices. So the determinant expression only makes sense when k is a multiple of n . We may as well just write $(\det A)^k$.

Our final conclusion is

Theorem 150 *Every irreducible representation of $U(n)$ is given by a Young diagram with at most $n - 1$ rows, together with an integer k . The representation is the standard representation of $U(n)$ determined by this Young diagram, multiplied by $(\det A)^k$.*

31.15 Representations of $GL(n, R)$ and $GL(n, C)$

Most of our results for $U(n)$ ultimately work for $GL(n, R)$ and $GL(n, C)$. Recall again that we are studying arbitrary continuous representations of $GL(n, R)$, which by Lie theory are actually C^∞ , but holomorphic representations of $GL(n, C)$. We will make one restriction. The group $GL(n, R)$ has two connected components, determined by the sign of $\det A$, and we will only deal with the connected subgroup $GL^+(n, R)$.

However, there is a difficulty. It is not true that every representation of $GL(n, R)$ can be written as a sum of irreducible representations. For example, it is easy to check that the following is a representation of $GL^+(n, R)$:

$$A \rightarrow \begin{pmatrix} 1 & \log(\det A) \\ 0 & 1 \end{pmatrix}$$

The only invariant subspace of this representation is the x -axis.

From now on, suppose ρ is a representation of G , where G is either $GL(n, R)$ or $GL(n, C)$. Each G has two closed subgroups which commute with each other and generate the full group. The first is the one-dimensional group of diagonal matrices λI . The second is the group of matrices with determinant one, which we will call SL ; it is either $SL(n, R)$ or $SL(n, C)$. We can restrict ρ to either of these subgroups. Its restriction to SL is fully reducible as a sum of irreducible representations. The restriction of ρ to λI is more problematic. Let us assume that it also decomposes as a sum of irreducible representations. This is the basic assumption we make in this section, and under this assumption we will prove that ρ is a sum of irreducible representations and determine exactly what they look like.

The group λI is isomorphic to the group of all positive real numbers λ when $G = GL^+(n, R)$, and isomorphic to the group of all nonzero complex numbers λ when $G = GL(n, C)$. In both cases the irreducible representations are one-dimensional and have the form

$$\lambda \rightarrow \lambda^c$$

for some fixed constant c . In the real case there are no restrictions on c because $\lambda^c = e^{c \ln(\lambda)}$. In the complex case, λ^c is usually multiple-valued and it is impossible to continuously pick one of these values, unless c is an integer. So in that case c is an arbitrary integer.

Our previous analysis then works as before. We can gather together equivalent irreducible representations of λI and write $V = W_1 \oplus \dots \oplus W_k$. We know that when we restrict ρ to the subgroup SL , it breaks into a sum of irreducible representations. Each element of one of these representations acts as an intertwining operator for the representation on λI , so as before, each irreducible representation of the special group is in exactly one W_i . It follows that the representation of λI is constant on each of the direct summands for SL . The entire representation space is the sum of these irreducible spaces for SL , so ρ is a sum of irreducible representations. We can extend the representations of SL to all of G since Young diagrams work for both the full group and the subgroup SL . So the irreducible representations of G are representations given by a Young diagram, multiplied by λ^c for an appropriate c .

This result is not quite satisfactory because it asks us to start with an arbitrary $A \in G$ and extract λI from it. Notice that $\det A = \lambda^n$. So in the real case we can write $\lambda^c = \lambda^{c/n}$ and the new c/n is just another constant. Clearly the choice of λ is unique for each A in this real case. Consequently we have proved the following theorem:

Theorem 151 *Let ρ be a representation of $GL(n, R)$ and suppose the restriction of ρ to the diagonal subgroup λI decomposes into a sum of irreducible representations. Then ρ itself decomposes into a sum of irreducible representations of $GL(n, R)$. Moreover, each of these irreducible representations has the form*

$$(\det A)^c \rho_1(A)$$

where c is some fixed real constant and ρ_1 is an irreducible representation formed by a Young diagram. Any c and ρ_1 can occur, and they are uniquely determined by the irreducible representation.

The case $G = GL(n, C)$ is slightly trickier. There are two problems. The first is that λI may belong to SL even if $\lambda \neq 1$; indeed this happens whenever $\lambda^n = 1$. In this case the value of $\rho(\lambda I)$ is completely determined already by the restriction of ρ to SL .

Given A , we want to extract λ . As before, $\lambda^n = \det A$ and so $\lambda^c = (\det A)^{c/n}$. But now there are multiple choices for this λ . We want to make one choice, continuously over all of

$GL(n, C)$. This is only possible if c/n is an integer. We already knew that c had to be an integer, but now that integer must be a multiple of n . Luckily, this removes the ambiguity in the previous paragraph.

Here is the final theorem:

Theorem 152 *Let ρ be a representation of $GL(n, C)$ and suppose the restriction of ρ to the diagonal subgroup λI decomposes into a sum of irreducible representations. Then ρ itself decomposes into a sum of irreducible representations of $GL(n, C)$. Moreover, each of these irreducible representations has the form*

$$(\det A)^k \rho_1(A)$$

where k is some fixed integer and ρ_1 is an irreducible representation formed by a Young diagram. Any k and ρ_1 can occur, and they are uniquely determined by the irreducible representation.

31.16 That Pesky Assumption

Our treatments of $GL(n, R)$ and $GL(n, C)$ required assuming that the representation ρ on λI is a sum of irreducible representations. To understand this assumption more clearly, we determine all representations of R^+ and C^* , that is, the positive real numbers and the non-zero complex numbers. By Lie theory, such representations are determined by the induced Lie algebra representation ρ^* . Since the groups are one dimensional, ρ^* is determined by its value at the identity of the group, which is an arbitrary complex matrix. By changing the basis, we can put this matrix in Jordan canonical form. Thus assume it is a sum of smaller blocks along the diagonal, each of the form

$$\begin{pmatrix} c & 1 & & \\ & c & 1 & \\ & & c & 1 \\ & & & c \end{pmatrix}$$

The pesky assumption is that all of these blocks are 1×1 .

Call one of these blocks T . The representation itself sends e^t to e^{tT} . Note that $\lambda = e^t$.

A brief calculation gives

$$e^{tT} = \begin{pmatrix} e^{ct} & te^{ct} & \frac{t^2}{2!}e^{ct} & \frac{t^3}{3!}e^{ct} \\ & e^{ct} & te^{ct} & \frac{t^2}{2!}e^{ct} \\ & & e^{ct} & te^{ct} \\ & & & e^{ct} \end{pmatrix}$$

But recall that $e^t = \lambda$ and so $e^{ct} = \lambda^t$ and $t = \ln \lambda$.

In the complex case, choosing a single continuous value from the multiple values of $\ln \lambda$ will be impossible. The conclusion is that Jordan boxes of size greater than one do not induce representations of λI . In short, the pesky assumption is automatically true. So

Theorem 153 *In the complex case, every representation of λI can be written as a sum of irreducible representations, and therefore every representation of $GL(n, C)$ can be written as a sum of irreducible representations, each of the form*

$$(\det A)^k \rho_1$$

where k is an integer and ρ_1 is a representation of $GL(n, C)$ associated with a Young diagram.

Remark: We do not get off this easy in the case $GL^+(n, R)$ because we already know that the diagonal group has representations which do not decompose into irreducible representations. However, the previous formula shows that in this case, the representation of λI contains terms of the form $\ln(\det A)$ to various powers. These logarithmic terms do not occur if our pesky assumption is true.

Definition 56 *A representation ρ of $G = GL^+(n, R)$ is said to be algebraic if whenever $A \in G$, the coefficients of the matrix $\rho(A)$ are rational functions of the coefficients of A .*

Remark: Clearly $(\det A)^k$ and representations associated with Young diagrams have this form.

Theorem 154 *Every algebraic representation of $GL^+(n, R)$ can be decomposed into a sum of irreducible algebraic representations. The algebraic irreducible representations have the form*

$$(\det A)^k \rho_1$$

where k is an integer and ρ_1 is a representation of $GL^+(n, R)$ associated with a Young diagram.