Groups and Representation Theory

Brian C. Hall
University of Notre Dame
Department of Mathematics
Computing Center and Mathematics Building
Notre Dame, IN 46556 U.S.A.
bhall@nd.edu

November, 1999

Contents

Chapter 1
GROUPS

1.1 Definition of a Group, and Basic Properties

Definition 1 *A* ***group*** *is a set G, together with a map of $G \times G$ into G (denoted $g_1 * g_2$) with the following properties:*

First, associativity: for all $g_1, g_2 \in G$,

$$g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3. \tag{1.1}$$

Second, there exists an element e in G such that for all $g \in G$,

$$g * e = e * g = g. \tag{1.2}$$

and such that for all $g \in G$, there exists $h \in G$ with

$$g * h = h * g = e. \tag{1.3}$$

If $g * h = h * g$ for all $g, h \in G$, then the group is said to be ***commutative*** *(or* ***abelian****).*

Remarks

The element e is (as we shall see momentarily) unique, and is called the **identity element** of the group, or simply the **identity**. Part of the definition of a group is that multiplying a group element g by the identity on *either the right or the left* must give back g.

The map of $G \times G$ into G is called the **product operation** for the group. Part of the definition of a group G is that the product operation map $G \times G$ into G, i.e., that the product of two elements of G be again an element of G. This property is refered to as **closure**.

Given a group element g, a group element h such that $g * h = h * g = e$ is called an **inverse** of g. We shall see momentarily that each group element has a *unique* inverse.

Given a set and an operation, there are four things that must be checked to show that this is a group: *closure*, *associativity*, existence of an *identity*, and existence of *inverses*.

Proposition 2 (Uniqueness of the Identity) *Let G be a group, and let $e, f \in G$ be such that for all $g \in G$*

$$e * g = g * e = g$$
$$f * g = g * f = g.$$

Then $e = f$.

Proof

Since e is an identity, we have

$$e * f = f.$$

On the other hand, since f is an identity, we have

$$e * f = e.$$

Thus $e = e * f = f$. $\square$

Proposition 3 (Uniqueness of Inverses) *Let G be a group, e the (unique) identity of G, and g, h, k arbitrary elements of G. Suppose that*

$$g * h = h * g = e$$
$$g * k = k * g = e.$$

Then $h = k$.

Proof

We know that $g * h = g * k$ $(= e)$. Multiplying on the left by h gives

$$h * (g * h) = h * (g * k).$$

By associativity, this gives

$$(h * g) * h = (h * g) * k,$$

and so

$$e * h = e * k$$
$$h = k.$$

This is what we wanted to prove. $\square$

Proposition 4 *Let G be a group, e the identity element of G, and g an arbitrary element of G. Suppose $h \in G$ satisfies either $h * g = e$ or $g * h = e$. Then h is the (unique) inverse of g.*

Proof

To show that h is the inverse of g, we must show *both* that $h * g = e$ and $g * h = e$. Suppose we know, say, that $h * g = e$. Then our goal is to show that this implies that $g * h = e$.

Since $h * g = e$,

$$g * (h * g) = g * e = g.$$

By associativity, we have

$$(g * h) * g = g.$$

Now, by the definition of a group, g has an inverse. Let k be that inverse. (Of course, in the end, we will conclude that $k = h$, but we cannot assume that now.) Multiplying on the right by k and using associativity again gives

$$\begin{aligned} ((g * h) * g) * k &= g * k = e \\ (g * h) * (g * k) &= e \\ (g * h) * e &= e \\ g * h &= e. \end{aligned}$$

A similar argument shows that if $g * h = e$, then $h * g = e$. □

Remark

Note that in order to show that $h * g = e$ implies $g * h = e$, we used the fact that g has an inverse, since it is an element of a group. In more general contexts (that is, in some system which is not a group), one may have $h * g = e$ but not $g * h = e$. (See Exercise 10.)

Notation

For any group element g, its unique inverse will be denoted g^{-1}.

Proposition 5 (Properties of Inverses) *Let G be a group, e its identity, and g, h arbitrary elements of G. Then*

$$\begin{aligned} \left(g^{-1}\right)^{-1} &= g \\ (gh)^{-1} &= h^{-1}g^{-1} \\ e^{-1} &= e. \end{aligned}$$

Proof

Exercise.

1.2 Some Examples of Groups

Notation

From now on, we will denote the product of two group elements g_1 and g_2 simply by $g_1 g_2$, instead of the more cumbersome $g_1 * g_2$. Moreover, since we have associativity, we will write simply $g_1 g_2 g_3$ in place of $(g_1 g_2) g_3$ or $g_1(g_2 g_3)$.

The trivial group

The set with one element, e, is a group, with the group operation being defined as $ee = e$. This group is commutative.

Associativity is automatic, since both sides of (1.1) must be equal to e. Of course, e itself is the identity, and is its own inverse. Commutativity is also automatic.

The integers

The set $\mathbb{Z}$ of integers forms a group with the product operation being addition. This group is commutative.

First, we check *closure*, namely, that addition maps $\mathbb{Z} \times \mathbb{Z}$ into $\mathbb{Z}$, i.e., that the sum of two integers is an integer. Since this is obvious, it remains only to check *associativity*, *identity*, and *inverses*. Addition is associative; zero is the additive identity (i.e., $0 + n = n + 0 = n$, for all $n \in \mathbb{Z}$); each integer n has an additive inverse, namely, $-n$. Since addition is commutative, $\mathbb{Z}$ is a commutative group.

The reals and $\mathbb{R}^n$

The set $\mathbb{R}$ of real numbers also forms a group under the operation of addition. This group is commutative. Similarly, the n-dimensional Euclidean space $\mathbb{R}^n$ forms a group under the operation of vector addition. This group is also commutative.

The verification is the same as for the integers.

Non-zero real numbers under multiplication

The set of non-zero real numbers forms a group with respect to the operation of multiplication. This group is commutative.

Again we check closure: the product of two non-zero real numbers is a non-zero real number. Multiplication is associative; one is the multiplicative identity; each non-zero real number x has a multiplicative inverse, namely, $\frac{1}{x}$. Since multiplication of real numbers is commutative, this is a commutative group.

This group is denoted $\mathbb{R}^*$.

Non-zero complex numbers under multiplication

The set of non-zero complex numbers forms a group with respect to the operation of complex multiplication. This group is commutative.

This group in denoted $\mathbb{C}^*$.

Complex numbers of absolute value one under multiplication

The set of complex numbers with absolute value one (i.e., of the form $e^{i\theta}$) forms a group under complex multiplication. This group is commutative.

This group is the unit circle, denoted S^1.

Invertible matrices

For each positive integer n, the set of all $n \times n$ invertible matrices with real entries forms a group with respect to the operation of matrix multiplication. This group in non-commutative, for $n \geq 2$.

We check closure: the product of two invertible matrices is invertible, since $(AB)^{-1} = B^{-1}A^{-1}$. Matrix multiplication is associative; the identity matrix (with ones down the diagonal, and zeros elsewhere) is the identity element; by definition, an invertible matrix has an inverse. Simple examples show that the group is non-commutative, except in the trivial case $n = 1$. (See Exercise 7.)

This group is called the **general linear group** (over the reals), and is denoted $\mathsf{GL}(n;\mathbb{R})$.

Symmetric group (permutation group)

The set of one-to-one, onto maps of the set $\{1, 2, \cdots n\}$ to itself forms a group under the operation of composition. This group is non-commutative for $n \geq 3$.

We check closure: the composition of two one-to-one, onto maps is again one-to-one and onto. Composition of functions is associative; the identity map (which sends 1 to 1, 2 to 2, etc.) is the identity element; a one-to-one, onto map has an inverse. Simple examples show that the group is non-commutative, as long as n is at least 3. (See Exercise 9.)

This group is called the **symmetric group**, and is denoted S_n. A one-to-one, onto map of $\{1, 2, \cdots n\}$ is a permutation, and so S_n is also called the **permutation group**. The group S_n has $n!$ elements.

Integers mod n

The set $\{0, 1, \cdots n - 1\}$ forms a group under the operation of addition **mod** n. This group is commutative.

Explicitly, the group operation is the following. Consider $a, b \in \{0, 1 \cdots n - 1\}$. If $a + b < n$, then $a + b$ **mod** $n = a + b$, if $a + b \geq n$, then $a + b$ **mod** $n = a + b - n$.

(Since a and b are less than n, $a+b-n$ is less than n; thus we have closure.) To show associativity, note that both $(a+b \textbf{ mod } n)+c \textbf{ mod } n$ and $a+(b+c \textbf{ mod } n) \textbf{ mod } n$ are equal to $a+b+c$, minus some multiple of n, and hence differ by a multiple of n. But since both are in the set $\{0,1,\cdots n-1\}$, the only possible multiple on n is zero. Zero is still the identity for addition **mod** n. The inverse of an element $a \in \{0,1,\cdots n-1\}$ is $n-a$. (Exercise: check that $n-a$ is in $\{0,1,\cdots n-1\}$, and that $a+(n-a) \textbf{ mod } n = 0$.) The group is commutative because ordinary addition is commutative.

This group is refered to as "$\mathbb{Z}$ **mod** n," and is denoted $\mathbb{Z}_n$.

1.3 Subgroups, the Center, and Direct Products

Definition 6 *A* ***subgroup*** *of a group G is a subset H of G with the following properties:*

1. *The identity is an element of H.*
2. *If $h \in H$, then $h^{-1} \in H$.*
3. *If $h_1, h_2 \in H$, then $h_1 h_2 \in H$.*

Remark

The conditions on H guarantee that H is a group, with the same product operation as G (but restricted to H). Closure is assured by (3), associativity follows from associativity in G, and the existence of an identity and of inverses is assured by (1) and (2).

Examples

Every group G has at least two subgroups: G itself, and the one-element subgroup $\{e\}$. (If G itself is the trivial group, then these two subgroups coincide.) These are called the **trivial subgroups** of G.

The set of even integers is a subgroup of $\mathbb{Z}$: zero is even, the negative of an even integer is even, and the sum of two even integers is even.

The set H of $n \times n$ real matrices with determinant one is a subgroup of $\mathsf{GL}(n;\mathbb{R})$. The set H is a sub*set* of $\mathsf{GL}(n;\mathbb{R})$ because any matrix with determinant one is invertible. The identity matrix has determinant one, so 1 is satisfied. The determinant of the inverse is the reciprocal of the determinant, so 2 is satisfied; and the determinant of a product is the product of the determinants, so 3 is satisfied. This group is called the **special linear group** (over the reals), and is denoted $\mathsf{SL}(n;\mathbb{R})$.

Additional examples, as well as some non-examples, are given in Exercise 1.

Definition 7 *The* ***center*** *of a group G is the set of all $g \in G$ such that $gh = hg$ for all $h \in G$.*

Exercise

Show that the center of any group G is a subgroup G.

Definition 8 *Let G and H be groups, and consider the Cartesian product of G and H, i.e., the set of ordered pairs (g, h) with $g \in G, h \in H$. Define a product operation on this set as follows:*

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2).$$

This operation makes the Cartesian product of G and H into a group, called the ***direct product*** *of G and H and denoted $G \times H$.*

Remark

It is a simple matter to check that this operation truly makes $G \times H$ into a group. For example, the identity element of $G \times H$ is the pair (e_1, e_2), where e_1 is the identity for G, and e_2 is the identity for H.

1.4 Homomorphisms and Isomorphisms

Definition 9 *Let G and H be groups. A map $\phi : G \to H$ is called a* ***homomorphism*** *if $\phi(g_1 g_2) = \phi(g_1)\phi(g_2)$ for all $g_1, g_2 \in G$. If in addition, ϕ is one-to-one and onto, then ϕ is called an* ***isomorphism****. An isomorphism of a group with itself is called an* ***automorphism****.*

Proposition 10 *Let G and H be groups, e_1 the identity element of G, and e_2 the identity element of H. If $\phi : G \to H$ is a homomorphism, then $\phi(e_1) = e_2$, and $\phi(g^{-1}) = \phi(g)^{-1}$ for all $g \in G$.*

Proof

Let g be any element of G. Then $\phi(g) = \phi(ge_1) = \phi(g)\phi(e_1)$. Multiplying on the left by $\phi(g)^{-1}$ gives $e_2 = \phi(e_1)$. Now consider $\phi(g^{-1})$. Since $\phi(e_1) = e_2$, we have $e_2 = \phi(e_1) = \phi(gg^{-1}) = \phi(g)\phi(g^{-1})$. In light of Prop. 4, we conclude that $\phi(g^{-1})$ is the inverse of $\phi(g)$. $\square$

Definition 11 *Let G and H be groups, $\phi : G \to H$ a homomorphism, and e_2 the identity element of H. The* ***kernel*** *of ϕ is the the set of all $g \in G$ for which $\phi(g) = e_2$.*

Proposition 12 *Let G and H be groups, and $\phi : G \to H$ a homomorphism. Then the kernel of ϕ is a subgroup of G.*

Proof

Easy. $\square$

Examples

Given any two groups G and H, we have the trivial homomorphism from G to H: $\phi(g) = e$ for all $g \in G$. The kernel of this homomorphism is all of G.

In any group G, the identity map ($id(g) = g$) is an automorphism of G, whose kernel is just $\{e\}$.

Let $G = H = \mathbb{Z}$, and define $\phi(n) = 2n$. This is a homomorphism of $\mathbb{Z}$ to itself, but not an automorphism. The kernel of this homomorphism is just $\{0\}$.

The determinant is a homomorphism of $\mathsf{GL}(n, \mathbb{R})$ to $\mathbb{R}^*$. The kernel of this map is $\mathsf{SL}\,(n, \mathbb{R})$.

Additional examples are given in Exercises 11 and 6.

Remark

If there exists an isomorphism from G to H, then G and H are said to be **isomorphic**, and this relationship is denoted $G \cong H$. (See Exercise 3.) Two groups which are isomorphic should be thought of as being (for all practical purposes) the same group.

1.5 Exercises

Recall the definitions of the groups $\mathsf{GL}(n; \mathbb{R})$, S_n, $\mathbb{R}^*$, and $\mathbb{Z}_n$ from Sect. 1.2, and the definition of the group $\mathsf{SL}(n; \mathbb{R})$ from Sect. 1.3.

1. In (a)-(f), you are given a group G and a subset H of G. In each case, determine whether H is a subgroup of G.

   (a) $G = \mathbb{Z},\ H = \{\text{odd integers}\}$

   (b) $G = \mathbb{Z},\ H = \{\text{multiples of } 3\}$

   (c) $G = \mathsf{GL}(n; \mathbb{R}),\ H = \{A \in \mathsf{GL}(n; \mathbb{R})\,|\det A \text{ is an integer}\}$

   (d) $G = \mathsf{SL}(n; \mathbb{R}),\ H = \{A \in \mathsf{SL}(n; \mathbb{R})\,|\text{all the entries of } A \text{ are integers}\}$

   Hint: recall Kramer's rule for finding the inverse of a matrix.

   (e) $G = \mathsf{GL}(n; \mathbb{R}),\ H = \{A \in \mathsf{GL}(n; \mathbb{R})\,|\text{all of the entries of } A \text{ are rational}\}$

   (f) $G = \mathbb{Z}_9,\ H = \{0, 2, 4, 6, 8\}$

2. Verify the properties of inverses in Prop. 5.

3. Let G and H be groups. Suppose there exists an isomorphism ϕ from G to H. Show that there exists an isomorphism from H to G.

4. Show that the set of positive real numbers is a subgroup of $\mathbb{R}^*$. Show that this group is isomorphic to the group $\mathbb{R}$.

5. Show that the set of automorphisms of any group G is itself a group, under the operation of composition. This group is the **automorphism group** of G, $Aut(G)$.

6. Given any group G, and any element g in G, define $\phi_g : G \to G$ by $\phi_g(h) = ghg^{-1}$. Show that ϕ_g is an automorphism of G. Show that the map $g \to \phi_g$ is a homomorphism of G into $Aut(G)$, and that the kernel of this map is the center of G.

   Note: An automorphism which can be expressed as ϕ_g for some $g \in G$ is called an **inner automorphism**; any automorphism of G which is not equal to any ϕ_g is called an **outer automorphism**.

7. Give an example of two 2×2 invertible real matrices which do not commute. (This shows that $\mathsf{GL}(2, \mathbf{R})$ is not commutative.)

8. Show that in any group G, the center of G is a subgroup.

9. An element σ of the permutation group S_n can be written in two-row form,

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix}$$

   where σ_i denotes $\sigma(i)$. Thus

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

   is the element of S_3 which sends 1 to 2, 2 to 3, and 3 to 1. When multiplying (i.e., composing) two permutations, one performs the one on the right first, and then the one on the left. (This is the usual convention for composing functions.)

   Compute

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

   and

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

   Conclude that S_3 is not commutative.

10. Consider the set $\mathbb{N}= \{0, 1, 2, \cdots\}$ of natural numbers, and the set $\mathcal{F}$ of *all* functions of $\mathbb{N}$ to itself. Composition of functions defines a map of $\mathcal{F} \times \mathcal{F}$ into $\mathcal{F}$, which is associative. The identity $(id(n) = n)$ has the property that $id \circ f = f \circ id = f$, for all f in $\mathcal{F}$. However, since we do not restrict to funtions

which are one-to-one and onto, not every element of $\mathcal{F}$ has an inverse. Thus $\mathcal{F}$ is not a group.

Give an example of two functions f, g in $\mathcal{F}$ such that $f \circ g = id$, but $g \circ f \neq id$. (Compare with Prop. 4.)

11. Consider the groups $\mathbb{Z}$ and $\mathbb{Z}_n$. For each a in $\mathbb{Z}$, define a **mod** n to be the unique element b of $\{0, 1, \cdots n-1\}$ such that a can be written as $a = kn + b$, with k an integer. Show that the map $a \to a$ **mod** n is a homomorphism of $\mathbb{Z}$ into $\mathbb{Z}_n$.

12. Let G be a group, and H a subgroup of G. H is called a **normal subgroup** of G if given any $g \in G$, and $h \in H$, ghg^{-1} is in H.

    Show that any subgroup of a commutative group is normal. Show that in any group G, the trivial subgroups G and $\{e\}$ are normal. Show that the center of any group is a normal subgroup. Show that if ϕ is a homomorphism from G to H, then the kernel of ϕ is a normal subgroup of G.

    Show that $\mathsf{SL}(n; \mathbb{R})$ is a normal subgroup of $\mathsf{GL}(n; \mathbb{R})$.

    Note: a group G with no normal subgroups other than G and $\{e\}$ is called **simple**.

Chapter 2
MATRIX LIE GROUPS

2.1 Definition of a Matrix Lie Group

Recall that the **general linear group** over the reals, denoted $\mathsf{GL}(n;\mathbb{R})$, is the group of all $n\times n$ invertible matrices with real entries. We may similarly define $\mathsf{GL}(n;\mathbb{C})$ to be the group of all $n\times n$ invertible matrices with complex entries. Of course, $\mathsf{GL}(n;\mathbb{R})$ is contained in $\mathsf{GL}(n;\mathbb{C})$.

Definition 13 *Let A_n be a sequence of complex matrices. We say that A_n* ***converges*** *to a matrix A if each entry of A_n converges to the corresponding entry of A, i.e., if $(A_n)_{ij}$ converges to A_{ij} for all $1\leq i,j\leq n$.*

Definition 14 *A* ***matrix Lie group*** *is any subgroup H of $\mathsf{GL}(n;\mathbb{C})$ with the following property: if A_n is any sequence of matrices in H, and A_n converges to some matrix A, then either $A\in H$, or A is not invertible.*

Remarks

The condition on H amounts to saying that H is a closed subset of $\mathsf{GL}(n;\mathbb{C})$. (This is not the same as saying that H is closed in the space of all matrices.) Thus Definition 14 is equivalent to saying that a matrix Lie group is a **closed subgroup** of $\mathsf{GL}(n;\mathbb{C})$.

The condition that H be a *closed* subgroup, as opposed to merely a subgroup, should be regarded as a technicality, in that most of the *interesting* subgroups of $\mathsf{GL}(n;\mathbb{C})$ have this property. (Almost all of the matrix Lie groups H we will consider have the stronger property that if A_n is any sequence of matrices in H, and A_n converges to some matrix A, then $A\in H$.)

There is a topological structure on the set of $n\times n$ complex matrices which goes with the above notion of convergence. This topological structure is defined by identifying the space of $n\times n$ matrices with $\mathbb{C}^{n^2}$ in the obvious way and using the usual topological structure on $\mathbb{C}^{n^2}$.

Counterexamples

An example of a subgroup of $\mathsf{GL}(n;\mathbb{C})$ which is not closed (and hence is not a matrix Lie group) is the set of all $n\times n$ invertible matrices all of whose enties are real and rational. This is in fact a subgroup of $\mathsf{GL}(n;\mathbb{C})$, but not a closed subgroup. That is, one can (easily) have a sequence of invertible matrices with rational entries converging to an invertible matix with some irrational entries. (In fact, *every* real invertible matrix is the limit of some sequence of invertible matrices with rational entries.)

Another example of a group of matrices which is not a matrix Lie group is the following subgroup of $\mathsf{GL}(2,\mathbb{C})$. Let a be an irrational real number, and let

$$H = \left\{ \left(\begin{array}{cc} e^{it} & 0 \\ 0 & e^{ita} \end{array} \right) | t \in \mathbb{R} \right\}$$

Clearly, H is a subgroup of $\mathsf{GL}(2,\mathbb{C})$. Because a is irrational, the matrix $-I$ is not in H, since to make e^{it} equal to -1, we must take t to be an odd integer multiple of π, in which case ta cannot be an odd integer multiple of π. On the other hand, by taking $t = (2n+1)\pi$ for a suitably chosen integer n, we can make ta arbitrarily *close* to an odd integer multiple of π. (It is left to the reader to verify this.) Hence we can find a sequence of matrices in H which converges to $-I$, and so H is not a matrix Lie group. See Exercise 1.

2.2 Examples of Matrix Lie Groups

Mastering the subject of Lie groups involves not only learning the general theory, but also familiarizing oneself with examples. In this section, we introduce some of the most important examples of (matrix) Lie groups.

The general linear groups $\mathsf{GL}(n;\mathbb{R})$ and $\mathsf{GL}(n;\mathbb{C})$

The general linear groups (over $\mathbb{R}$ or $\mathbb{C}$) are themselves matrix Lie groups. Of course, $\mathsf{GL}(n;\mathbb{C})$ is a subgroup of itself. Furthermore, if A_n is a sequence of matrices in $\mathsf{GL}(n;\mathbb{C})$ and A_n converges to A, then by the definition of $\mathsf{GL}(n;\mathbb{C})$, either A is in $\mathsf{GL}(n;\mathbb{C})$, or A is not invertible.

Moreover, $\mathsf{GL}(n;\mathbb{R})$ is a subgroup of $\mathsf{GL}(n;\mathbb{C})$, and if $A_n \in \mathsf{GL}(n;\mathbb{R})$, and A_n converges to A, then the entries of A are real. Thus either A is not invertible, or $A \in \mathsf{GL}(n;\mathbb{R})$.

The special linear groups $\mathsf{SL}(n;\mathbb{R})$ and $\mathsf{SL}(n;\mathbb{C})$

The **special linear group** (over $\mathbb{R}$ or $\mathbb{C}$) is the group of $n \times n$ invertible matrices (with real or complex entries) having determinant one. Both of these are subgroups of $\mathsf{GL}(n;\mathbb{C})$, as noted in Chapter 1. Furthermore, if A_n is a sequence of matrices with determinant one, and A_n converges to A, then A also has determinant one, because the determinant is a continuous function. Thus $\mathsf{SL}(n;\mathbb{R})$ and $\mathsf{SL}(n;\mathbb{C})$ are matrix Lie groups.

The orthogonal and special orthogonal groups, $\mathsf{O}(n)$ and $\mathbf{SO}(n)$

An $n \times n$ real matrix A is said to be **orthogonal** if the column vectors that make up A are orthonormal, that is, if

$$\sum_{i=1}^{n} A_{ij} A_{ik} = \delta_{jk}$$

Equivalently, A is orthogonal if it preserves the inner product, namely, if $\langle x, y\rangle = \langle Ax, Ay\rangle$ for all vectors x, y in $\mathbb{R}^n$. (Angled brackets denote the usual inner product on $\mathbb{R}^n$, $\langle x, y\rangle = \sum_i x_i y_i$.) Still another equivalent definition is that A is orthogonal if $A^{tr}A = I$, i.e., if $A^{tr} = A^{-1}$. (A^{tr} is the transpose of A, $(A^{tr})_{ij} = A_{ji}$.) See Exercise 2.

Since $\det A^{tr} = \det A$, we see that if A is orthogonal, then $\det(A^{tr}A) = (\det A)^2 = \det I = 1$. Hence $\det A = \pm 1$, for all orthogonal matrices A.

This formula tells us, in particular, that every orthogonal matrix must be invertible. But if A is an orthogonal matrix, then

$$\left\langle A^{-1}x, A^{-1}y\right\rangle = \left\langle A\left(A^{-1}x\right), A\left(A^{-1}x\right)\right\rangle = \langle x, y\rangle$$

Thus the inverse of an orthogonal matrix is orthogonal. Furthermore, the product of two orthogonal matrices is orthogonal, since if A and B both preserve inner products, then so does AB. Thus the set of orthogonal matrices forms a group.

The set of all $n \times n$ real orthogonal matrices is the **orthogonal group** $\mathsf{O}(n)$, and is a subgroup of $\mathsf{GL}(n; \mathbb{C})$. The limit of a sequence of orthogonal matices is orthogonal, because the relation $A^{tr}A = I$ is preserved under limits. Thus $\mathsf{O}(n)$ is a matrix Lie group.

The set of $n \times n$ orthogonal matrices with determinant one is the **special orthogonal group** $\mathsf{SO}(n)$. Clearly this is a subgroup of $\mathsf{O}(n)$, and hence of $\mathsf{GL}(n; \mathbb{C})$. Moreover, both orthogonality and the property of having determinant one are preserved under limits, and so $\mathsf{SO}(n)$ is a matrix Lie group. Since elements of $\mathsf{O}(n)$ already have determinant ± 1, $\mathsf{SO}(n)$ is "half" of $\mathsf{O}(n)$.

Geometrically, elements of $\mathsf{O}(n)$ are either rotations, or combinations of rotations and reflections. The elements of $\mathsf{SO}(n)$ are just the rotations.

See also Exercise 6.

The unitary and special unitary groups, $\mathsf{U}(n)$ and $\mathsf{SU}(n)$

An $n \times n$ complex matrix A is said to be **unitary** if the column vectors of A are orthonormal, that is, if

$$\sum_{i=1}^{n} \overline{A_{ij}} A_{ik} = \delta_{jk}$$

Equivalently, A is unitary if it preserves the inner product, namely, if $\langle x, y\rangle = \langle Ax, Ay\rangle$ for all vectors x, y in $\mathbb{C}^n$. (Angled brackets here denote the inner product on $\mathbb{C}^n$, $\langle x, y\rangle = \sum_i \overline{x_i} y_i$. We will adopt the convention of putting the complex conjugate on the left.) Still another equivalent definition is that A is unitary if $A^*A = I$, i.e., if $A^* = A^{-1}$. (A^* is the adjoint of A, $(A^*)_{ij} = \overline{A_{ji}}$.) See Exercise 3.

Since $\det A^* = \overline{\det A}$, we see that if A is unitary, then $\det(A^*A) = |\det A|^2 = \det I = 1$. Hence $|\det A| = 1$, for all unitary matrices A.

This in particular shows that every unitary matrix is invertible. The same argument as for the orthogonal group shows that the set of unitary matrices forms a group.

The set of all $n \times n$ unitary matrices is the **unitary group** $\mathsf{U}(n)$, and is a subgroup of $\mathsf{GL}(n; \mathbb{C})$. The limit of unitary matrices is unitary, so $\mathsf{U}(n)$ is a matrix Lie group. The set of unitary matrices with determinant one is the **special unitary group** $\mathsf{SU}(n)$. It is easy to check that $\mathsf{SU}(n)$ is a matrix Lie group. Note that a unitary matrix can have determinant $e^{i\theta}$ for any θ, and so $\mathsf{SU}(n)$ is a smaller subset of $\mathsf{U}(n)$ than $\mathsf{SO}(n)$ is of $\mathsf{O}(n)$. (Specifically, $\mathsf{SO}(n)$ has the same dimension as $\mathsf{O}(n)$, whereas $\mathsf{SU}(n)$ has dimension one less than that of $\mathsf{U}(n)$.)

See also Exercise 8.

The complex orthogonal groups, $\mathsf{O}(n; \mathbb{C})$ and $\mathsf{SO}(n; \mathbb{C})$

Consider the bilinear form $(\ )$ on $\mathbf{C}^n$ defined by $(x, y) = \sum x_i y_i$. This form is not an inner product, because of the lack of a complex conjugate in the definition. The set of all $n \times n$ complex matrices A which preserve this form, (i.e., such that $(Ax, Ay) = (x, y)$ for all $x, y \in \mathbf{C}^n$) is the **complex orthogonal group** $\mathsf{O}(n; \mathbb{C})$, and is a subgroup of $\mathsf{GL}(n; \mathbb{C})$. (The proof is the same as for $\mathsf{O}(n)$.) An $n \times n$ complex matrix A is in $\mathsf{O}(n; \mathbb{C})$ if and only if $A^{tr}A = I$. It is easy to show that $\mathsf{O}(n; \mathbb{C})$ is a matrix Lie group, and that $\det A = \pm 1$, for all A in $\mathsf{O}(n; \mathbb{C})$. Note that $\mathsf{O}(n; \mathbb{C})$ is *not* the same as the unitary group $\mathsf{U}(n)$. The group $\mathsf{SO}(n; \mathbb{C})$ is defined to be the set of all A in $\mathsf{O}(n; \mathbb{C})$ with $\det A = 1$. Then $\mathsf{SO}(n; \mathbb{C})$ is also a matrix Lie group.

The generalized orthogonal and Lorentz groups

Let n and k be positive integers, and consider $\mathbb{R}^{n+k}$. Define a symmetric bilinear form $[\ ]_{n+k}$ on $\mathbb{R}^{n+k}$ by the formula

$$[x, y]_{n,k} = x_1 y_1 + \cdots + x_n y_n - x_{n+1} y_{n+1} \cdots - y_{n+k} x_{n+k} \tag{2.1}$$

The set of $(n+k) \times (n+k)$ real matrices A which preserve this form (i.e., such that $[Ax, Ay]_{n,k} = [x, y]_{n,k}$ for all $x, y \in \mathbb{R}^{n+k}$) is the **generalized orthogonal group** $\mathsf{O}(n; k)$, and it is a subgroup of $\mathsf{GL}(n + k; \mathbb{R})$ (Ex. 4). Since $\mathsf{O}(n; k)$ and $\mathsf{O}(k; n)$ are essentially the same group, we restrict our attention to the case $n \geq k$. It is not hard to check that $\mathsf{O}(n; k)$ is a matrix Lie group.

If A is an $(n+k)\times(n+k)$ real matrix, let $A^{(i)}$ denote the i^{th} column vector of A, that is

$$A^{(i)} = \begin{pmatrix} A_{1,i} \\ \vdots \\ A_{n+k,i} \end{pmatrix}$$

Then A is in $\mathsf{O}(n;k)$ if and only if the following conditions are satisfied:

$$\begin{array}{rcll} \left[A^{(i)}, A^{(j)}\right]_{n,k} & = & 0 & i \neq j \\ \left[A^{(i)}, A^{(i)}\right]_{n,k} & = & 1 & 1 \leq i \leq n \\ \left[A^{(i)}, A^{(i)}\right]_{n,k} & = & -1 & n+1 \leq i \leq n+k \end{array} \tag{2.2}$$

Let g denote the $(n+k)\times(n+k)$ diagonal matrix with ones in the first n diagonal entries, and minus ones in the last k diagonal entries. Then A is in $\mathsf{O}(n;k)$ if and only if $A^{tr}gA = g$ (Ex. 4). Taking the determinant of this equation gives $(\det A)^2 \det g = \det g$, or $(\det A)^2 = 1$. Thus for any A in $\mathsf{O}(n;k)$, $\det A = \pm 1$.

The group $\mathsf{SO}(n;k)$ is defined to be the set of matrices in $\mathsf{O}(n;k)$ with $\det A = 1$. This is a subgroup of $\mathsf{GL}(n+k;\mathbb{R})$, and is a matrix Lie group.

Of particular interest in physics is the **Lorentz group** $\mathsf{O}(3;1)$. (Sometimes the phrase Lorentz group is used more generally to refer to the group $\mathsf{O}(n;1)$ for any $n \geq 1$.) See also Exercise 7.

The symplectic groups $\mathsf{Sp}(n;\mathbb{R})$, $\mathsf{Sp}(n;\mathbb{C})$, and $\mathsf{Sp}(n)$

The special and general linear groups, the orthogonal and unitary groups, and the symplectic groups (which will be defined momentarily) make up the **classical groups**. Of the classical groups, the symplectic groups have the most confusing definition, partly because there are three sets of them ($\mathsf{Sp}(n;\mathbb{R})$, $\mathsf{Sp}(n;\mathbb{C})$, and $\mathsf{Sp}(n)$), and partly because they involve skew-symmetric bilinear forms rather than the more familiar symmetric bilinear forms. To further confuse matters, the notation for referring to these groups is not consistent from author to author.

Consider the skew-symmetric bilinear form B on $\mathbb{R}^{2n}$ defined as follows:

$$B[x,y] = \sum_{i=1}^{n} x_i y_{n+i} - x_{n+i} y_i \tag{2.3}$$

The set of all $2n\times 2n$ matrices A which preserve B (i.e., such that $B[Ax,Ay] = B[x,y]$ for all $x,y \in \mathbb{R}^{2n}$) is the **real symplectic group** $\mathsf{Sp}(n;\mathbb{R})$, and it is a subgroup of $\mathsf{GL}(2n;\mathbb{R})$. It is not difficult to check that this is a matrix Lie group (Exercise 5). This group arises naturally in the study of classical mechanics. If J is the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

then $B[x,y] = \langle x, Jy\rangle$, and it is possible to check that a $2n \times 2n$ real matrix A is in $\mathsf{Sp}(n;\mathbb{R})$ if and only if $A^{tr}JA = J$. (See Exercise 5.) Taking the determinant of this identity gives $(\det A)^2 \det J = \det J$, or $(\det A)^2 = 1$. This shows that $\det A = \pm 1$, for all $A \in \mathsf{Sp}(n;\mathbb{R})$. In fact, $\det A = 1$ for all $A \in \mathsf{Sp}(n;\mathbb{R})$, although this is not obvious.

One can define a bilinear form on $\mathbb{C}^n$ by the same formula (2.3). (This form is bilinear, not hermitian, and involves no complex conjugates.) The set of $2n \times 2n$ complex matrices which preserve this form is the **complex symplectic group** $\mathsf{Sp}(n;\mathbb{C})$. A $2n \times 2n$ complex matrix A is in $\mathsf{Sp}(n;\mathbb{C})$ if and only if $A^{tr}JA = J$. (Note: this condition involves A^{tr}, *not* A^*.) This relation shows that $\det A = \pm 1$, for all $A \in \mathsf{Sp}(n;\mathbb{C})$. In fact $\det A = 1$, for all $A \in \mathsf{Sp}(n;\mathbb{C})$.

Finally, we have the **compact symplectic group** $\mathsf{Sp}(n)$ defined as

$$\mathsf{Sp}(n) = \mathsf{Sp}(n;\mathbb{C}) \cap \mathsf{U}(2n).$$

See also Exercise 9. For more information and a proof of the fact that $\det A = 1$, for all $A \in \mathsf{Sp}(n;\mathbb{C})$, see Miller, Sect. 9.4. What we call $\mathsf{Sp}(n;\mathbb{C})$ Miller calls $\mathsf{Sp}(n)$, and what we call $\mathsf{Sp}(n)$, Miller calls $\mathsf{USp}(n)$.

The Heisenberg group H

The set of all 3×3 real matrices A of the form

$$A = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \tag{2.4}$$

where a, b, and c are arbitrary real numbers, is the **Heisenberg group**. It is easy to check that the product of two matrices of the form (2.4) is again of that form, and clearly the identity matrix is of the form (2.4). Furthermore, direct computation shows that if A is as in (2.4), then

$$A^{-1} = \begin{pmatrix} 1 & -a & ac-b \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{pmatrix}$$

Thus H is a subgroup of $\mathsf{GL}(3;\mathbb{R})$. Clearly the limit of matrices of the form (2.4) is again of that form, and so H is a matrix Lie group.

It is not evident at the moment why this group should be called the Heisenberg group. We shall see later that the Lie algebra of H gives a realization of the Heisenberg commutation relations of quantum mechanics. (See especially Chapter 5, Exercise 10.)

See also Exercise 10.

The groups $\mathbb{R}^*$, $\mathbb{C}^*$, S^1, $\mathbb{R}$, and $\mathbb{R}^n$

Several important groups which are not naturally groups of matrices can (and will in these notes) be thought of as such.

The group $\mathbb{R}^*$ of non-zero real numbers under multiplication is isomorphic to $\mathsf{GL}(1,\mathbb{R})$. Thus we will regard $\mathbb{R}^*$ as a matrix Lie group. Similarly, the group $\mathbb{C}^*$ of non-zero complex numbers under multiplication is isomorphic to $\mathsf{GL}(1;\mathbb{C})$, and the group S^1 of complex numbers with absolute value one is isomorphic to $\mathsf{U}(1)$.

The group $\mathbb{R}$ under addition is isomorphic to $\mathsf{GL}(1;\mathbb{R})^+$ (1×1 real matrices with positive determinant) via the map $x \to [e^x]$. The group $\mathbb{R}^n$ (with vector addition) is isomorphic to the group of diagonal real matrices with positive diagonal entries, via the map

$$(x_1,\cdots,x_n) \to \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_n} \end{pmatrix}.$$

The Euclidean and Poincare groups

The Euclidean group $\mathsf{E}(n)$ is by definition the group of all one-to-one, onto, distance-preserving maps of $\mathbb{R}^n$ to itself, that is, maps $f:\mathbb{R}^n \to \mathbb{R}^n$ such that $d\left(f\left(x\right), f\left(y\right)\right) = d\left(x,y\right)$ for all $x,y\in\mathbb{R}^n$. Here d is the usual distance on $\mathbb{R}^n$, $d\left(x,y\right) = \left|x-y\right|$. Note that we don't assume *anything* about the structure of f besides the above properties. In particular, f need not be linear. The orthogonal group $\mathsf{O}(n)$ is a subgroup of $\mathsf{E}(n)$, and is the group of all *linear* distance-preserving maps of $\mathbb{R}^n$ to itself. The set of translations of $\mathbb{R}^n$ (i.e., the set of maps of the form $T_x(y) = x+y$) is also a subgroup of $\mathsf{E}(n)$.

Proposition 15 *Every element T of $\mathsf{E}(n)$ can be written uniquely as an orthogonal linear transformation followed by a translation, that is, in the form*

$$T = T_x R$$

with $x\in\mathbb{R}^n$, and $R\in\mathsf{O}(n)$.

We will not prove this here. The key step is to prove that every one-to-one, onto, distance-preserving map of $\mathbb{R}^n$ to itself which fixes the origin must be linear.

Following Miller, we will write an element $T = T_x R$ of $\mathsf{E}(n)$ as a pair $\{x,R\}$. Note that for $y\in\mathbb{R}^n$,

$$\{x,R\}\,y = Ry + x$$

and that

$$\{x_1,R_1\}\{x_2,R_2\}y = R_1(R_2 y + x_2) + x_1 = R_1R_2y + (x_1 + R_1x_2)$$

Thus the product operation for $\mathsf{E}(n)$ is the following:

$$\{x_1, R_1\}\{x_2, R_2\} = \{x_1 + R_1 x_2, R_1 R_2\} \tag{2.5}$$

The inverse of an element of $\mathsf{E}(n)$ is given by

$$\{x, R\}^{-1} = \{-R^{-1}x, R^{-1}\}$$

Now, as already noted, $\mathsf{E}(n)$ is not a subgroup of $\mathsf{GL}(n;\mathbb{R})$, since translations are not linear maps. However, $\mathsf{E}(n)$ is isomorphic to a subgroup of $\mathsf{GL}(n+1;\mathbb{R})$, via the map which associates to $\{x, R\} \in \mathsf{E}(n)$ the following matrix

$$\begin{pmatrix} & & & x_1 \\ & R & & \vdots \\ & & & x_n \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{2.6}$$

This map is clearly one-to-one, and it is a simple computation to show that it is a homomorphism. Thus $\mathsf{E}(n)$ is isomorphic to the group of all matrices of the form (2.6) (with $R \in \mathsf{O}(n)$). The limit of things of the form (2.6) is again of that form, and so we have expressed the Euclidean group $\mathsf{E}(n)$ as a matrix Lie group.

We similarly define the Poincaré group $\mathsf{P}(n;1)$ to be the group of all transformations of $\mathbb{R}^{n+1}$ of the form

$$T = T_x A$$

with $x \in \mathbb{R}^{n+1}$, $A \in \mathsf{O}(n;1)$. This is the group of affine transformations of $\mathbb{R}^{n+1}$ which preserve the Lorentz "distance" $d_L(x,y) = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 - (x_{n+1} - y_{n+1})^2$. (An affine transformation is one of the form $x \to Ax + b$, where A is a linear transformation and b is constant.) The group product is the obvious analog of the product (2.5) for the Euclidean group.

The Poincaré group $\mathsf{P}(n;1)$ is isomorphic to the group of $(n+2) \times (n+2)$ matrices of the form

$$\begin{pmatrix} & & & x_1 \\ & A & & \vdots \\ & & & x_{n+1} \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{2.7}$$

with $A \in \mathsf{O}(n;1)$. The set of matrices of the form (2.7) is a matrix Lie group.

2.3 Compactness

Definition 16 *A matrix Lie group G is said to be **compact** if the following two conditions are satisfied:*

1. *If A_n is any sequece of matrices in G, and A_n converges to a matrix A, then A is in G.*

2. *There exists a constant C such that for all $A \in G$, $|A_{ij}| \leq C$ for all $1 \leq i, j \leq n$.*

Remarks

This is not the usual topological definition of compactness. However, the set of all $n \times n$ complex matrices can be thought of as $\mathbb{C}^{n^2}$. The above definition says that G is compact if it is a *closed, bounded* subset of $\mathbb{C}^{n^2}$. It is a standard theorem from elementary analysis that a subset of $\mathbb{C}^m$ is compact (in the usual sense that every open cover has a finite subcover) if and only if it is closed and bounded.

All of our examples of matrix Lie groups except $\mathsf{GL}(n;\mathbb{R})$ and $\mathsf{GL}(n;\mathbb{C})$ have property (1). Thus it is the boundedness condition (2) that is most important.

The property of compactness has very important implications. For example, if G is compact, then every irreducible unitary representation of G is finite-dimensional.

Examples of compact groups

The groups $\mathsf{O}(n)$ and $\mathsf{SO}(n)$ are compact. Property (1) is satisfied because the limit of orthogonal matrices is orthogonal and the limit of matrices with determinant one has determinant one. Property (2) is satisfied because if A is orthogonal, then the column vectors of A have norm one, and hence $|A_{ij}| \leq 1$, for all $1 \leq i, j \leq n$. A similar argument shows that $\mathsf{U}(n)$, $\mathsf{SU}(n)$, and $\mathsf{Sp}(n)$ are compact. (This includes the unit circle, $S^1 \cong \mathsf{U}(1)$.)

Examples of non-compact groups

All of the other examples given of matrix Lie groups are non-compact. $\mathsf{GL}(n;\mathbb{R})$ and $\mathsf{GL}(n;\mathbb{C})$ violate property (1), since a limit of invertible matrices may be non-invertible. $\mathsf{SL}\,(n;\mathbb{R})$ and $\mathsf{SL}\,(n;\mathbb{C})$ violate (2), except in the trivial case $n = 1$, since

$$A_n = \begin{pmatrix} n & & & & \\ & \frac{1}{n} & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

has determinant one, no matter how big n is.

The following groups also violate (2), and hence are non-compact: $\mathsf{O}(n;\mathbb{C})$ and $\mathsf{SO}(n;\mathbb{C})$; $\mathsf{O}(n;k)$ and $\mathsf{SO}(n;k)$ $(n \geq 1,\ k \geq 1)$; the Heisenberg group H; $\mathsf{Sp}\,(n;\mathbb{R})$ and $\mathsf{Sp}\,(n;\mathbb{C})$; $\mathsf{E}(n)$ and $\mathsf{P}(n;1)$; $\mathbb{R}$ and $\mathbb{R}^n$; $\mathbb{R}^*$ and $\mathbb{C}^*$. It is left to the reader to provide examples to show that this is the case.

2.4 Connectedness

Definition 17 *A matrix Lie group G is said to be* **connected** *if given any two matrices A and B in G, there exists a continuous path $A(t)$, $a \leq t \leq b$, lying in G with $A(a) = A$, and $A(b) = B$.*

Remarks

This property is what is called **path-connected** in topology, which is not (in general) the same as connected. However, it is a fact (not particulary obvious at the moment) that a matrix Lie group is connected if and only if it is path-connected. So in a slight abuse of terminology we shall continue to refer to the above property as connectedness. (See Section 2.7.)

A matrix Lie group G which is not connected can be decomposed (uniquely) as a union of several pieces, called **components**, such that two elements of the same component can be joined by a continuous path, but two elements of different components cannot.

Proposition 18 *If G is a matrix Lie group, then the component of G containing the identity is a subgroup of G.*

Proof

Saying that A and B are both in the component containing the identity means that there exist continuous paths $A(t)$ and $B(t)$ with $A(0) = B(0) = I$, $A(1) = A$, and $B(1) = B$. But then $A(t)B(t)$ is a continuous path starting at I and ending at AB. Thus the product of two elements of the identity component is again in the identity component. Furthermore, $A(t)^{-1}$ is a continuous path starting at I and ending at A^{-1}, and so the inverse of any element of the identity component is again in the identity component. Thus the identity component is a subgroup. □

Proposition 19 *The group $\mathsf{GL}(n;\mathbb{C})$ is connected for all $n \geq 1$.*

Proof

Consider first the case $n = 1$. A 1×1 invertible complex matrix A is of the form $A = [\lambda]$ with $\lambda \in \mathbb{C}^*$, the set of non-zero complex numbers. But given any two non-zero complex numbers, we can easily find a continuous path which connects them and does not pass through zero.

For the case $n \geq 1$, we use the Jordan canonical form. Every $n \times n$ complex matrix A can be written as

$$A = CBC^{-1}$$

where B is the Jordan canonical form. The only property of B we will need is that B is upper-triangular:

$$B = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

If A is invertible, then all the λ_i's must be non-zero, since $\det A = \det B = \lambda_1 \cdots \lambda_n$.

Let $B(t)$ be obtained by multiplying the part of B above the diagonal by $(1-t)$, for $0 \leq t \leq 1$, and let $A(t) = CB(t)C^{-1}$. Then $A(t)$ is a continuous path which starts at A and ends at CDC^{-1}, where D is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

This path lies in $\mathsf{GL}(n;\mathbb{C})$ since $\det A(t) = \lambda_1 \cdots \lambda_n$ for all t.

But now, as in the case $n = 1$, we can define $\lambda_i(t)$ which connects each λ_i to 1 in $\mathbf{C}^*$, as t goes from 1 to 2. Then we can define

$$A(t) = C \begin{pmatrix} \lambda_1(t) & & 0 \\ & \ddots & \\ 0 & & \lambda_n(t) \end{pmatrix} C^{-1}$$

This is a continuous path which starts at CDC^{-1} when $t = 1$, and ends at I $(= CIC^{-1})$ when $t = 2$. Since the $\lambda_i(t)$'s are always non-zero, $A(t)$ lies in $\mathsf{GL}(n;\mathbb{C})$.

We see, then, that every matrix A in $\mathsf{GL}(n;\mathbb{C})$ can be connected to the identity by a continuous path lying in $\mathsf{GL}(n;\mathbb{C})$. Thus if A and B are two matrices in $\mathsf{GL}(n;\mathbb{C})$, they can be connected by connecting each of them to the identity. □

Proposition 20 *The group* $\mathsf{SL}\,(n;\mathbb{C})$ *is connected for all* $n \geq 1$.

Proof

The proof is almost the same as for $\mathsf{GL}(n;\mathbb{C})$, except that we must be careful to preserve the condition $\det A = 1$. Let A be an arbitrary element of $\mathsf{SL}\,(n;\mathbb{C})$. The case $n = 1$ is trivial, so we assume $n \geq 2$. We can define $A(t)$ as above for $0 \leq t \leq 1$, with $A(0) = A$, and $A(1) = CDC^{-1}$, since $\det A(t) = \det A = 1$. Now define $\lambda_i(t)$ as before for $1 \leq i \leq n-1$, and define $\lambda_n(t)$ to be $[\lambda_1(t) \cdots \lambda_{n-1}(t)]^{-1}$. (Note that since $\lambda_1 \cdots \lambda_n = 1$, $\lambda_n(0) = \lambda_n$.) This allows us to connect A to the identity while staying within $\mathsf{SL}\,(n;\mathbb{C})$. □

Proposition 21 *The groups* $\mathsf{U}(n)$ *and* $\mathsf{SU}(n)$ *are connected, for all* $n \geq 1$.

Proof

By a standard result of linear algebra, every unitary matrix has an orthonormal basis of eigenvectors, with eigenvalues of the form $e^{i\theta}$. It follows that every unitary matrix U can be written as

$$U = U_1 \begin{pmatrix} e^{i\theta_1} & & 0 \\ & \ddots & \\ 0 & & e^{i\theta_n} \end{pmatrix} U_1^{-1} \tag{2.8}$$

with U_1 unitary and $\theta_i \in \mathbb{R}$. Conversely, as is easily checked, every matrix of the form (2.8) is unitary. Now define

$$U(t) = U_1 \begin{pmatrix} e^{i(1-t)\theta_1} & & 0 \\ & \ddots & \\ 0 & & e^{i(1-t)\theta_n} \end{pmatrix} U_1^{-1}$$

As t ranges from 0 to 1, this defines a continuous path in $\mathsf{U}(n)$ joining U to I. This shows that $\mathsf{U}(n)$ is connected.

A slight modification of this argument, as in the proof of Proposition 20, shows that $\mathsf{SU}(n)$ is connected. $\square$

Proposition 22 *The group $\mathsf{GL}(n;\mathbb{R})$ is not connected, but has two components. These are $\mathsf{GL}(n;\mathbb{R})^+$, the set of $n \times n$ real matrices with positive determinant, and $\mathsf{GL}(n;\mathbb{R})^-$, the set of $n \times n$ real matrices with negative determinant.*

Proof

$\mathsf{GL}(n;\mathbb{R})$ cannot be connected, for if $\det A > 0$ and $\det B < 0$, then any continuous path connecting A to B would have to include a matrix with determinant zero, and hence pass outside of $\mathsf{GL}(n;\mathbb{R})$.

The proof that $\mathsf{GL}(n;\mathbb{R})^+$ is connected is given in Exercise 14. Once $\mathsf{GL}(n;\mathbb{R})^+$ is known to be connected, it is not difficult to see that $\mathsf{GL}(n;\mathbb{R})^-$ is also connected. For let C be any matrix with negative determinant, and take A, B in $\mathsf{GL}(n;\mathbb{R})^-$. Then $C^{-1}A$ and $C^{-1}B$ are in $\mathsf{GL}(n;\mathbb{R})^+$, and can be joined by a continuous path $D(t)$ in $\mathsf{GL}(n;\mathbb{R})^+$. But then $CD(t)$ is a continuous path joining A and B in $\mathsf{GL}(n;\mathbb{R})^-$. $\square$

The following table lists some matrix Lie groups, indicates whether or not the group is connected, and gives the number of components.

Group	Connected?	Components
$\mathsf{GL}(n;\mathbb{C})$	yes	1
$\mathsf{SL}\,(n;\mathbb{C})$	yes	1
$\mathsf{GL}(n;\mathbb{R})$	no	2
$\mathsf{SL}\,(n;\mathbb{R})$	yes	1
$\mathsf{O}(n)$	no	2
$\mathsf{SO}(n)$	yes	1
$\mathsf{U}(n)$	yes	1
$\mathsf{SU}(n)$	yes	1
$\mathsf{O}(n;1)$	no	4
$\mathsf{SO}(n;1)$	no	2
Heisenberg	yes	1
$\mathsf{E}\,(n)$	no	2
$\mathsf{P}(n;1)$	no	4

Proofs of some of these results are given in Exercises 7, 11, 13, and 14. (The connectedness of the Heisenberg group is immediate.)

2.5 Simple-connectedness

Definition 23 *A connected matrix Lie group G is said to be **simply connected** if every loop in G can be shrunk continuously to a point in G.*

More precisely, G is simply connected if given any continuous path $A(t)$, $0 \leq t \leq 1$, lying in G with $A(0) = A(1)$, there exists a continuous function $A(s,t)$, $0 \leq s,t \leq 1$, taking values in G with the following properties: 1) $A(s,0) = A(s,1)$ for all s, 2) $A(0,t) = A(t)$, and 3) $A(1,t) = A(1,0)$ for all t.

Remarks

You should think of $A(t)$ as a loop, and $A(s,t)$ as a parameterized family of loops which shrinks $A(t)$ to a point. Condition 1) says that for each value of the parameter s, we have a loop; condition 2) says that when $s = 0$ the loop is the specified loop $A(t)$; and condition 3) says that when $s = 1$ our loop is a point.

It is customary to speak of simple-connectedness only for connected matrix Lie groups, even though the definition makes sense for disconnected groups.

Proposition 24 *The group* $\mathsf{SU}(2)$ *is simply connected.*

Proof

Exercise 8 shows that $\mathsf{SU}(2)$ may be thought of (topologically) as the three-dimensional sphere S^3 sitting inside $\mathbb{R}^4$. It is well-known that S^3 is simply connected. $\square$

The condition of simple-connectedness is extremely important. One of our most important theorems will be that if G is simply connected, then there is a natural one-to-one correspondence between the representations of G and the representations of its Lie algebra.

Without proof, we give the following table.

Group	**Simply connected?**
$\mathsf{GL}(n;\mathbb{C})$	no
$\mathsf{SL}(n;\mathbb{C})$	yes
$\mathsf{GL}(n;\mathbb{R})$	no
$\mathsf{SL}(n;\mathbb{R})$	no
$\mathsf{SO}(n)$	no
$\mathsf{U}(n)$	no
$\mathsf{SU}(n)$	yes
$\mathsf{SO}(1;1)$	yes
$\mathsf{SO}(n;1)$ $(n \geq 2)$	no
Heisenberg	yes

2.6 Homomorphisms and Isomorphisms

Definition 25 *Let G and H be matrix Lie groups. A map ϕ from G to H is called a* ***Lie group homomorphism*** *if 1) ϕ is a group homomorphism and 2) ϕ is continuous. If in addition, ϕ is one-to-one and onto, and the inverse map ϕ^{-1} is continuous, then ϕ is called a* ***Lie group isomorphism****.*

Remarks

The condition that ϕ be continuous should be regarded as a technicality, in that it is very difficult to give an example of a group homomorphism between two matrix Lie groups which is not continuous. In fact, if $G = \mathbb{R}$ and $H = \mathbb{C}^*$, then any group homorphism from G to H which is even measurable (a very weak condition) must be continuous. (See W. Rudin, *Real and Complex Analysis,* Chap. 9, Ex. 17.)

If G and H are matrix Lie groups, and there exists a Lie group isomorphism from G to H, then G and H are said to be **isomorphic**, and we write $G \cong H$. Two matrix Lie groups which are isomorphic should be thought of as being essentially the same group. (Note that by definition, the inverse of Lie group isomorphism is continuous, and so also a Lie group isomorphism.)

Example: $\mathsf{SU}(2)$ and $\mathsf{SO}(3)$

A very important topic for us will be the relationship between the groups $\mathsf{SU}(2)$ and $\mathsf{SO}(3)$. This example is designed to show that $\mathsf{SU}(2)$ and $\mathsf{SO}(3)$ are almost (but not quite!) isomorphic. Specifically, there exists a Lie group homomorphism ϕ which maps $\mathsf{SU}(2)$ onto $\mathsf{SO}(3)$, and which is *two*-to-one. (See Miller 7.1 and Bröcker, Chap. I, 6.18.)

Consider the space V of all 2×2 complex matrices which are self-adjoint and have trace zero. This is a three-dimensional *real* vector space with the following basis

$$A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad A_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}; \quad A_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

We may define an inner product on V by the formula

$$\langle A, B \rangle = \frac{1}{2}\mathrm{trace}(AB)$$

(Exercise: check that this is an inner product.)

Direct computation shows that $\{A_1, A_2, A_3\}$ is an orthonormal basis for V. Having chosen an orthonormal basis for V, we can indentify V with $\mathbb{R}^3$.

Now, if U is an element of $\mathsf{SU}(2)$, and A is an element of V, then it is easy to see that UAU^{-1} is in V. Thus for each $U \in \mathsf{SU}(2)$, we can define a linear map ϕ_U of V to itself by the formula

$$\phi_U(A) = UAU^{-1}$$

(This definition would work for $U \in \mathsf{U}(2)$, but we choose to restrict our attention to $\mathsf{SU}(2)$.) Moreover, given $U \in \mathsf{SU}(2)$, and $A, B \in V$, note that

$$\langle \phi_U(A), \phi_U(B) \rangle = \frac{1}{2}\mathrm{trace}(UAU^{-1}UBU^{-1}) = \frac{1}{2}\mathrm{trace}(AB) = \langle A, B \rangle$$

Thus ϕ_U is an orthogonal transformation of $V \cong \mathbb{R}^3$, which we can think of as an element of $\mathsf{O}(3)$.

We see, then, that the map $U \to \phi_U$ is a map of $\mathsf{SU}(2)$ into $\mathsf{O}(3)$. It is very easy to check that this map is a homomorphism (i.e., $\phi_{U_1 U_2} = \phi_{U_1}\phi_{U_2}$), and that it is continuous. Thus $U \to \phi_U$ is a Lie group homomorphism of $\mathsf{SU}(2)$ into $\mathsf{O}(3)$.

Recall that every element of $\mathsf{O}(3)$ has determinant ± 1. Since $\mathsf{SU}(2)$ is connected (Exercise 8), and the map $U \to \phi_U$ is continuous, ϕ_U must actually map into $\mathsf{SO}(3)$. Thus $U \to \phi_U$ is a Lie group homomorphism of $\mathsf{SU}(2)$ into $\mathsf{SO}(3)$.

The map $U \to \phi_U$ is not one-to-one, since for any $U \in \mathsf{SU}(2)$, $\phi_U = \phi_{-U}$. (Observe that if U is in $\mathsf{SU}(2)$, then so is $-U$.) It is possible to show that ϕ_U is a two-to-one map of $\mathsf{SU}(2)$ onto $\mathsf{SO}(3)$. (See Miller.)

2.7 Lie Groups

A Lie group is something which is simultaneously a group and a differentiable manifold (see Definition 26). As the terminology suggests, every matrix Lie group is a Lie group, although this requires proof (Theorem 27). I have decided to restrict attention to matrix Lie groups, except in emergencies, for three reasons. First, this makes the course accessible to students who are not familiar with the theory of differentiable manifolds. Second, this makes the definition of the Lie algebra and of the exponential mapping far more comprehensible. Third, all of the important examples of Lie groups are (or can easily be represented as) matrix Lie groups.

Alas, there is a price to pay for this simplification. Certain important topics (notably, the universal cover) are considerably complicated by restricting to the matrix case. Nevertheless, I feel that the advantages outweigh the disadvantages in an introductory course such as this.

Definition 26 *A **Lie group** is a differentiable manifold G which is also a group, and such that the group product*

$$G \times G \to G$$

and the inverse map $g \to g^{-1}$ are differentiable.

For the reader who is not familiar with the notion of a differentiable manifold, here is a brief recap. (I will consider only manifolds embedded in some $\mathbb{R}^n$, which is a harmless assumption.) A subset M of $\mathbf{R}^n$ is called a **k-dimensional differentiable manifold** if given any $m_0 \in M$, there exists a smooth (non-linear) coordinate system $(x^1, \cdots x^n)$ defined in a neighborhood U of m_0 such that

$$M \cap U = \left\{ m \in U \left| x^{k+1}(m) = c_1, \cdots, x^n(m) = c_{n-k} \right. \right\}$$

This says that locally, after a suitable change of variables, M looks like the k-dimensional hyperplane in $\mathbb{R}^n$ obtained by setting all but the first k coordinates equal to constants.

For example, $S^1 \subset \mathbb{R}^2$ is a one-dimensional differentiable manifold because in the usual polar coordinates (θ, r), S^1 is the set $r = 1$. Of course, polar coordinates are not globally defined, because θ is undefined at the origin, and because θ is not "single-valued." But given any point m_0 in S^1, we can define polar coordinates in a neighborhood U of m_0, and then $S^1 \cap U$ will be the set $r = 1$.

Note that while we assume that our differentiable manifolds are embedded in some $\mathbb{R}^n$ (a harmless assumption), we are *not* saying that a Lie group has to be embedden in $\mathbb{R}^{n^2}$, or that the group operation has to have anything to do with matrix multiplication. A Lie group is simply a subset G of some $\mathbb{R}^n$ which is a differentiable manifold, together with *any* map from $G \times G$ into G which makes G into a group (and such that the group operations are smooth). It is remarkable that almost (but not quite!) every Lie group is isomorphic to a matrix Lie group.

Note also that it is far from obvious that a matrix Lie group must be a Lie group, since our definition of a matrix Lie group G does not say anything about G being a manifold. It is not too difficult to verify that all of our examples of matrix Lie groups are Lie groups, but in fact we have the following result which makes such verifications unnecessary:

Theorem 27 *Every matrix Lie group is a Lie group.*

Discussion

Let us consider first the group $\mathsf{GL}(n;\mathbb{R})$. The space of all $n \times n$ real matrices can be thought of as $\mathbb{R}^{n^2}$. Since $\mathsf{GL}(n;\mathbb{R})$ is the set of all matrices A with $\det A \neq 0$, $\mathsf{GL}(n;\mathbb{R})$ is an open subset of $\mathbb{R}^{n^2}$. (That is, given an invertible matrix A, there is a neigborhood U of A such that every matrix $B \in U$ is also invertible.) Thus $\mathsf{GL}(n;\mathbb{R})$ is an n^2-dimensional smooth manifold. Furthermore, the matrix product AB is clearly a smooth (even polynomial) function of the entries of A and B, and (in light of Kramer's rule) A^{-1} is a smooth function of the entries of A. Thus $\mathsf{GL}(n;\mathbb{R})$ is a Lie group.

Similarly, if we think of the space of $n \times n$ complex matrices as $\mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$, then the same argument shows that $\mathsf{GL}(n;\mathbb{C})$ is a Lie group.

Thus, to prove that every matrix Lie group is a Lie group, it suffices to show that a closed subgroup of a Lie group is a Lie group. This is proved in Bröcker and tom Dieck, Chapter I, Theorem 3.11. The proof is not too difficult, but it requires the exponential mapping, which we have not yet introduced. (See Chapter 3.)

It is customary to call a map ϕ between two Lie groups a Lie group homomorphism if ϕ is a group homomorphism and ϕ is *smooth*, whereas we have (in Definition 25) required only that ϕ be continuous. However, the following Proposition shows that our definition is equivalent to the more standard one.

Proposition 28 *Let G and H be Lie groups, and ϕ a group homomorphism from G to H. Then if ϕ is continuous it is also smooth.*

Thus group homomorphisms from G to H come in only two varieties: the very bad ones (discontinuous), and the very good ones (smooth). There simply aren't any intermediate ones. (See, for example, Exercise 16.) For proof, see Bröcker and tom Dieck, Chapter I, Proposition 3.12.

Remark

In light of Theorem 27, every matrix Lie group is a (smooth) manifold. As such, a matrix Lie group is automatically locally path connected. It follows that a matrix Lie group is path connected if and only if it is connected. (See Remarks following Definition 17.)

2.8 Exercises

1. Let a be an irrational real number. Show that the set of numbers of the form $e^{2\pi i n a}$, $n \in \mathbb{Z}$, is dense in S^1. Now let G be the following subgroup of $\mathsf{GL}(2;\mathbb{C})$:

$$G = \left\{ \begin{pmatrix} e^{it} & 0 \\ 0 & e^{iat} \end{pmatrix} | t \in \mathbb{R} \right\}$$

Show that

$$\overline{G} = \left\{ \begin{pmatrix} e^{it} & 0 \\ 0 & e^{is} \end{pmatrix} | t, s \in \mathbb{R} \right\},$$

where $\overline{G}$ denotes the closure of the set G inside the space of 2×2 matrices.

Note: The group $\overline{G}$ can be thought of as the torus $S^1 \times S^1$, which in turn can be thought of as $[0, 2\pi] \times [0, 2\pi]$, with the ends of the intervals identified. The set $G \subset [0, 2\pi] \times [0, 2\pi]$ is called an **irrational line**. Draw a picture of this set and you should see why G is dense in $[0, 2\pi] \times [0, 2\pi]$.

2. *Orthogonal groups*. Let $\langle\ \rangle$ denote the standard inner product on $\mathbb{R}^n$, $\langle x, y \rangle = \sum_i x_i y_i$. Show that a matrix A preserves inner products if and only if the column vectors of A are orthonormal.

   Show that for any $n \times n$ real matrix B,

   $$\langle Bx, y \rangle = \left\langle x, B^{tr} y \right\rangle$$

   where $\left(B^{tr}\right)_{ij} = B_{ji}$. Using this fact, show that a matrix A preserves inner products if and only if $A^{tr} A = I$.

   Note: a similar analysis applies to the complex orthogonal groups $\mathsf{O}(n; \mathbb{C})$ and $\mathsf{SO}(n; \mathbb{C})$.

3. *Unitary groups*. Let $\langle\ \rangle$ denote the standard inner product on $\mathbb{C}^n$, $\langle x, y \rangle = \sum_i \overline{x_i} y_i$. Following Exercise 2, show that $A^* A = I$ if and only if $\langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{C}^n$. $((A^*)_{ij} = \overline{A_{ji}}.)$

4. *Generalized orthogonal groups.* Let $[x, y]_{n,k}$ be the symmetric bilinear form on $\mathbb{R}^{n+k}$ defined in (2.1). Let g be the $(n+k) \times (n+k)$ diagonal matrix with first n diagonal entries equal to one, and last k diagonal entries equal to minus one:

   $$g = \begin{pmatrix} I_n & 0 \\ 0 & -I_k \end{pmatrix}$$

   Show that for all $x, y \in \mathbb{R}^{n+k}$,

   $$[x, y]_{n,k} = \langle x, gy \rangle$$

   Show that a $(n+k) \times (n+k)$ real matrix A is in $\mathsf{O}(n; k)$ if and only if $A^{tr} g A = g$. Show that $\mathsf{O}(n; k)$ and $\mathsf{SO}(n; k)$ are subgroups of $\mathsf{GL}(n+k; \mathbb{R})$, and are matrix Lie groups.

5. *Symplectic groups.* Let $B[x, y]$ be the skew-symmetric bilinear form on $\mathbb{R}^{2n}$ given by $B[x, y] = \sum_{i=1}^n x_i y_{n+i} - x_{n+i} y_i$. Let J be the $2n \times 2n$ matrix

   $$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

Show that for all $x, y \in \mathbb{R}^{2n}$

$$B[x, y] = \langle x, Jy \rangle$$

Show that a $2n \times 2n$ matrix A is in $\mathsf{Sp}(n; \mathbb{R})$ if and only if $A^{tr} J A = J$. Show that $\mathsf{Sp}(n; \mathbb{R})$ is a subgroup of $\mathsf{GL}(2n; \mathbb{R})$, and a matrix Lie group.

Note: a similar analysis applies to $\mathsf{Sp}(n; \mathbb{C})$.

6. *The groups* $\mathsf{O}(2)$ *and* $\mathsf{SO}(2)$. Show that the matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

is in $\mathsf{SO}(2)$, and that

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}$$

Show that every element A of $\mathsf{O}(2)$ is of one of the two forms

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{2.9}$$

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \tag{2.10}$$

(If A is of the first form, then $\det A = 1$; if A is of the second form, then $\det A = -1$.)

Hint: Recall that for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to be in $\mathsf{O}(2)$, the column vectors $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ must be unit vectors, and must be orthogonal.

7. *The groups* $\mathsf{O}(1;1)$ *and* $\mathsf{SO}(1;1)$. Show that

$$A = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}$$

is in $\mathsf{SO}(1;1)$, and that

$$\begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \begin{pmatrix} \cosh s & \sinh s \\ \sinh s & \cosh s \end{pmatrix} = \begin{pmatrix} \cosh(t+s) & \sinh(t+s) \\ \sinh(t+s) & \cosh(t+s) \end{pmatrix}$$

Show that every element of $\mathsf{O}(1;1)$ can be written in one of the four forms

$$\begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \tag{2.11}$$

$$\begin{pmatrix} -\cosh t & \sinh t \\ \sinh t & -\cosh t \end{pmatrix} \tag{2.12}$$

$$\begin{pmatrix} \cosh t & -\sinh t \\ \sinh t & -\cosh t \end{pmatrix} \tag{2.13}$$

$$\begin{pmatrix} -\cosh t & -\sinh t \\ \sinh t & \cosh t \end{pmatrix} \tag{2.14}$$

(Since $\cosh t$ is always positive, there is no overlap among the four cases. Matrices of the first two forms have determinant one; matrices of the last two forms have determinant minus one.)

Hint: For $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to be in $\mathsf{O}(1;1)$, we must have $a^2 - c^2 = 1$, $b^2 - d^2 = -1$, and $ab - cd = 0$. The set of points (a, c) in the plane with $a^2 - c^2 = 1$ (i.e., $a = \pm\sqrt{1+c^2}$) is a hyperbola.

8. *The group* $\mathsf{SU}(2)$. Show that if α, β are arbitrary complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$, then the matrix

$$A = \begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix} \tag{2.15}$$

is in $\mathsf{SU}(2)$. Show that every $A \in \mathsf{SU}(2)$ can be expressed in the form (2.15) for a unique pair (α, β) satisfying $|\alpha|^2 + |\beta|^2 = 1$. (Thus $\mathsf{SU}(2)$ can be thought of as the three-dimensional sphere S^3 sitting inside $\mathbf{C}^2 = \mathbb{R}^4$. In particular, this shows that $\mathsf{SU}(2)$ is connected and simply connected.)

9. *The groups* $\mathsf{Sp}(1;\mathbb{R})$, $\mathsf{Sp}(1;\mathbf{C})$, *and* $\mathsf{Sp}(1)$. Show that $\mathsf{Sp}(1;\mathbb{R}) = \mathsf{SL}(2;\mathbb{R})$, $\mathsf{Sp}(1;\mathbf{C}) = \mathsf{SL}(2;\mathbf{C})$, and $\mathsf{Sp}(1) = \mathsf{SU}(2)$.

10. *The Heisenberg group.* Determine the center $Z(H)$ of the Heisenberg group H. Show that the quotient group $H/Z(H)$ is abelian.

11. *Connectedness of* $\mathsf{SO}(n)$. Show that $\mathsf{SO}(n)$ is connected, following the outline below.

For the case $n = 1$, there is not much to show, since a 1×1 matrix with determinant one must be $[1]$. Assume, then, that $n \geq 2$. Let e_1 denote the vector

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

in $\mathbb{R}^n$. Given any unit vector $v \in \mathbb{R}^n$, show that there exists a continuous path $R(t)$ in $\mathsf{SO}(n)$ with $R(0) = I$ and $R(1)v = e_1$. (Thus any unit vector can be "continuously rotated" to e_1.)

Now show that any element R of $\mathsf{SO}(n)$ can be connected to an element of $\mathsf{SO}(n-1)$, and proceed by induction.

12. *The polar decomposition of* $\mathsf{SL}(n;\mathbb{R})$. Show that every element A of $\mathsf{SL}(n;\mathbb{R})$ can be written uniquely in the form $A = RH$, where R is in $\mathsf{SO}(n)$, and H is a symmetric, positive-definite matrix with determinant one. (That is, $H^{tr} = H$, and $\langle x, Hx\rangle \geq 0$ for all $x \in \mathbb{R}^n$).

    Hint: If A *could* be written in this form, then we would have

    $$A^{tr}A = H^{tr}R^{tr}RH = HR^{-1}RH = H^2$$

    Thus H would have to be the unique positive-definite symmetric square root of $A^{tr}A$.

    Note: A similar argument gives polar decompositions for $\mathsf{GL}(n;\mathbb{R})$, $\mathsf{SL}(n;\mathbb{C})$, and $\mathsf{GL}(n;\mathbb{C})$. For example, every element A of $\mathsf{SL}(n;\mathbb{C})$ can be written uniquely as $A = UH$, with U in $\mathsf{SU}(n)$, and H a self-adjoint positive-definite matrix with determinant one.

13. *The connectedness of* $\mathsf{SL}(n;\mathbb{R})$. Using the polar decomposition of $\mathsf{SL}(n;\mathbb{R})$ (Ex. 12) and the connectedness of $\mathsf{SO}(n)$ (Ex. 11), show that $\mathsf{SL}(n;\mathbb{R})$ is connected.

    Hint: Recall that if H is a real, symmetric matrix, then there exists a *real* orthogonal matrix R_1 such that $H = R_1 D R_1^{-1}$, where D is diagonal.

14. *The connectedness of* $\mathsf{GL}(n;\mathbb{R})^+$. Show that $\mathsf{GL}(n;\mathbb{R})^+$ is connected.

15. Show that the set of translations is a normal subgroup of the Euclidean group, and also of the Poincaré group. Show that $(\mathsf{E}(n)/\text{translations}) \cong \mathsf{O}(n)$.

16. *Harder.* Show that every Lie group homomorphism ϕ from $\mathbb{R}$ to S^1 is of the form $\phi(x) = e^{iax}$ for some $a \in \mathbb{R}$. In particular, every such homomorphism is smooth.

Chapter 3
LIE ALGEBRAS AND THE EXPONENTIAL MAPPING

3.1 The Matrix Exponential

The exponential of a matrix plays a crucial role in the theory of Lie groups. The exponential enters into the definition of the Lie algebra of a matrix Lie group (Section 3.5 below), and is the mechanism for passing information from the Lie algebra to the Lie group. Since many computations are done much more easily at the level of the Lie algebra, the exponential is indispensible.

Let X be an $n \times n$ real or complex matrix. We wish to define the exponential of X, e^X or $\exp X$, by the usual power series

$$e^X = \sum_{m=0}^{\infty} \frac{X^m}{m!}. \tag{3.1}$$

We will follow the convention of using letters such as X and Y for the variable in the matrix exponential.

Proposition 29 *For any $n \times n$ real or complex matrix X, the series (3.1) converges. The matrix exponential e^X is a continuous function of X.*

Before proving this, let us review some elementary analysis. Recall that the norm of a vector x in $\mathbb{C}^n$ is defined to be

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum |x_i|^2}.$$

This norm satisfies the triangle inequality

$$\|x + y\| \leq \|x\| + \|y\|.$$

The norm of a matrix A is defined to be

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Equivalently, $\|A\|$ is the smallest number λ such that $\|Ax\| \leq \lambda \|x\|$ for all $x \in \mathbf{C}^n$.

It is not hard to see that for any $n \times n$ matrix A, $\|A\|$ is finite. Furthermore, it is easy to see that for any matrices A, B

$$\|AB\| \leq \|A\| \, \|B\| \tag{3.2}$$
$$\|A + B\| \leq \|A\| + \|B\| . \tag{3.3}$$

It is also easy to see that a sequence of matrices A_m converges to a matrix A if and only if $\|A_m - A\| \to 0$. (Compare this with Definition 13 of Chapter 2.)

A sequence of matrices A_m is said to be a **Cauchy sequence** if $\|A_m - A_l\| \to 0$ as $m, l \to \infty$. Thinking of the space of matrices as $\mathbb{R}^{n^2}$ or $\mathbb{C}^{n^2}$, and using a standard result from analysis, we have the following:

Proposition 30 *If A_m is a sequence of $n \times n$ real or complex matrices, and A_m is a Cauchy sequence, then there exists a unique matrix A such that A_m converges to A.*

That is, every Cauchy sequence converges.

Now, consider an infinite series whose terms are matrices:

$$A_0 + A_1 + A_2 + \cdots . \tag{3.4}$$

If

$$\sum_{m=0}^{\infty} \|A_m\| < \infty$$

then the series (3.4) is said to **converge absolutely**. If a series converges absolutely, then it is not hard to show that the partial sums of the series form a Cauchy sequence, and hence by Proposition 30, the series converges. That is, any series which converges absolutely also converges. (The converse is not true; a series of matrices can converge without converging absolutely.)

Proof

In light of (3.2), we see that

$$\|X^m\| \leq \|X\|^m ,$$

and hence

$$\sum_{m=0}^{\infty} \left\| \frac{X^m}{m!} \right\| \leq \sum_{m=0}^{\infty} \frac{\|X\|^m}{m!} = e^{\|X\|} < \infty.$$

Thus the series (3.1) converges absolutely, and so it converges.

To show continuity, note that since X^m is a continuous function of X, the partial sums of (3.1) are continuous. But it is easy to see that (3.1) converges uniformly on each set of the form $\{\|X\| \leq R\}$, and so the sum is again continuous. $\square$

Proposition 31 *Let X, Y be arbitrary $n \times n$ matrices. Then*

1. $e^0 = I$.
2. e^X *is invertible, and* $\left(e^X\right)^{-1} = e^{-X}$.
3. $e^{(\alpha+\beta)X} = e^{\alpha X} e^{\beta X}$ *for all real or complex numbers* α, β.
4. *If* $XY = YX$, *then* $e^{X+Y} = e^X e^Y = e^Y e^X$.
5. *If* C *is invertible, then* $e^{CXC^{-1}} = Ce^X C^{-1}$.
6. $\left\| e^X \right\| \leq e^{\|X\|}$.

Remark

It is *not* true in general that $e^{X+Y} = e^X e^Y$, although by 4) it is true if X and Y commute. This is a crucial point, which we will consider in detail later. (See the Lie product formula in Section 3.4 and the Baker-Campbell-Hausdorff formula in Chapter 4.)

Proof

Point 1) is obvious. Points 2) and 3) are special cases of point 4). To verify point 4), we simply multiply power series term by term. (It is left to the reader to verify that this is legal.) Thus

$$e^X e^Y = \left(I + X + \frac{X^2}{2!} + \cdots \right) \left(I + Y + \frac{Y^2}{2!} + \cdots \right).$$

Multiplying this out and collecting terms where the power of X plus the power of Y equals m, we get

$$e^X e^Y = \sum_{m=0}^{\infty} \sum_{k=0}^{m} \frac{X^k}{k!} \frac{Y^{m-k}}{(m-k)!} = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} X^k Y^{m-k}. \tag{3.5}$$

Now because (and *only* because) X and Y commute,

$$(X+Y)^n = \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} X^k Y^{m-k},$$

and so (3.5) becomes

$$e^X e^Y = \sum_{m=0}^{\infty} \frac{1}{m!} (X+Y)^m = e^{X+Y}.$$

To prove 5), simply note that

$$\left(CXC^{-1}\right)^m = CX^mC^{-1}$$

and so the two sides of 5) are the same term by term.

Point 6) is evident from the proof of Proposition 29. □

Proposition 32 *Let X be a $n \times n$ complex matrix, and view the space of all $n \times n$ complex matrices as $\mathbb{C}^{n^2}$. Then e^{tX} is a smooth curve in $\mathbb{C}^{n^2}$, and*

$$\frac{d}{dt}e^{tX} = Xe^{tX} = e^{tX}X.$$

In particular,

$$\left.\frac{d}{dt}\right|_{t=0} e^{tX} = X.$$

Proof

Differentiate the power series for e^{tX} term-by-term. (You might worry whether this is valid, but you shouldn't. For each i, j, $\left(e^{tX}\right)_{ij}$ is given by a convergent power series in t, and it is a standard theorem that you can differentiate power series term-by-term.) □

3.2 Computing the Exponential of a Matrix

Case 1: X is diagonalizable

Suppose that X is a $n \times n$ real or complex matrix, and that X is diagonalizable over $\mathbb{C}$, that is, that there exists an invertible complex matrix C such that $X = CDC^{-1}$, with

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Observe that e^D is the diagonal matrix with eigenvalues $e^{\lambda_1}, \cdots, e^{\lambda_n}$, and so in light of Proposition 31, we have

$$e^X = C\begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix}C^{-1}.$$

Thus if you can explicitly diagonalize X, you can explicitly compute e^X. Note that if X is real, then although C may be complex and the λ_i's may be complex, e^X must come out to be real, since each term in the series (3.1) is real.

For example, take

$$X = \begin{pmatrix} 0 & -a \\ a & 0 \end{pmatrix}.$$

Then the eigenvectors of X are $\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\begin{pmatrix} i \\ 1 \end{pmatrix}$, with eigenvalues $-ia$ and ia, respectively. Thus the invertible matrix

$$C = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

maps the basis vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to the eigenvectors of X, and so (check) $C^{-1}XC$ is a diagonal matrix D. Thus $X = CDC^{-1}$:

$$\begin{aligned} e^X &= \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} e^{-ia} & 0 \\ 0 & e^{ia} \end{pmatrix} \begin{pmatrix} 1/2 & -i/2 \\ -i/2 & 1/2 \end{pmatrix} \\ &= \begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix}. \end{aligned}$$

Note that explicitly if X (and hence a) is real, then e^X is real.

Case 2: X is nilpotent

An $n \times n$ matrix X is said to be **nilpotent** if $X^m = 0$ for some positive integer m. Of course, if $X^m = 0$, then $X^l = 0$ for all $l > m$. In this case the series (3.1) which defines e^X terminates after the first m terms, and so can be computed explicitly.

For example, compute e^{tX}, where

$$X = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that

$$X^2 = \begin{pmatrix} 0 & 0 & ac \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and that $X^3 = 0$. Thus

$$e^{tX} = \begin{pmatrix} 1 & ta & tb + \frac{1}{2}t^2 ac \\ 0 & 1 & tc \\ 0 & 0 & 1 \end{pmatrix}.$$

Case 3: X arbitrary

A general matrix X may be neither nilpotent nor diagonalizable. However, it follows from the Jordan canonical form that X can be written (Exercise 2) in the form $X = S + N$ with S diagonalizable, N nilpotent, and $SN = NS$. (See Exercise 2.) Then, since N and S commute,

$$e^X = e^{S+N} = e^S e^N$$

and e^S and e^N can be computed as above.

For example, take

$$X = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}.$$

Then

$$X = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}.$$

The two terms clearly commute (since the first one is a multiple of the identity), and so

$$e^X = \begin{pmatrix} e^a & 0 \\ 0 & e^a \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^a & e^a b \\ 0 & e^a \end{pmatrix}.$$

3.3 The Matrix Logarithm

We wish to define a matrix logarithm, which should be an inverse function to the matrix exponential. Defining a logarithm for matrices should be at least as difficult as defining a logarithm for complex numbers, and so we cannot hope to define the matrix logarithm for all matrices, or even for all invertible matrices. We will content ourselves with defining the logarithm in a neighborhood of the identity matrix.

The simplest way to define the matrix logarithm is by a power series. We recall the situation for complex numbers:

Lemma 33 *The function*

$$\log z = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(z-1)^m}{m}$$

is defined and analytic in a circle of radius one about $z = 1$.

For all z with $|z - 1| < 1$,

$$e^{\log z} = z.$$

For all u with $|u| < \log 2$, $|e^u - 1| < 1$ and

$$\log e^u = u.$$

Proof

The usual logarithm for real, positive numbers satisfies

$$\frac{d}{dx}\log(1-x) = \frac{-1}{1-x} = -\left(1 + x + x^2 + \cdots\right)$$

for $|x| < 1$. Integrating term-by-term and noting that $\log 1 = 0$ gives

$$\log(1-x) = -\left(x + \tfrac{x^2}{2} + \tfrac{x^3}{3} + \cdots\right).$$

Taking $z = 1 - x$ (so that $x = 1 - z$), we have

$$\begin{aligned}\log z &= -\left((1-z) + \tfrac{(1-z)^2}{2} + \tfrac{(1-z)^3}{3} + \cdots\right) \\ &= \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(z-1)^m}{m}.\end{aligned}$$

This series has radius of convergence one, and defines a complex analytic function on the set $\{|z-1| < 1\}$, which coincides with the usual logarithm for real z in the interval $(0,2)$. Now, $\exp(\log z) = z$ for $z \in (0,2)$, and by analyticity this identity continues to hold on the whole set $\{|z-1| < 1\}$.

On the other hand, if $|u| < \log 2$, then

$$|e^u - 1| = \left|u + \tfrac{u^2}{2!} + \cdots\right| \leq |u| + \frac{|u|^2}{2!} + \cdots$$

so that

$$|e^u - 1| \leq e^{|u|} - 1 < 1.$$

Thus $\log(\exp u)$ makes sense for all such u. Since $\log(\exp u) = u$ for real u with $|u| < \log 2$, it follows by analyticity that $\log(\exp u) = u$ for all complex numbers with $|u| < \log 2$. $\square$

Theorem 34 *The function*

$$\log A = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(A-I)^m}{m} \tag{3.6}$$

is defined and continuous on the set of all $n \times n$ complex matrices A with $\|A - I\| < 1$, and $\log A$ *is real if A is real.*

For all A with $\|A - I\| < 1$,

$$e^{\log A} = A.$$

For all X with $\|X\| < \log 2$, $\left\|e^X - 1\right\| < 1$ and

$$\log e^X = X.$$

Proof

It is easy to see that the series (3.6) converges absolutely whenever $\|A - I\| < 1$. The proof of continuity is essentially the same as for the exponential. If A is real, then every term in the series (3.6) is real, and so $\log A$ is real.

We will now show that $\exp(\log A) = A$ for all A with $\|A - I\| < 1$. We do this by considering two cases.

Case 1: A is diagonalizable.

Suppose that $A = CDC^{-1}$, with D diagonal. Then $A - I = CDC^{-1} - I = C(D - I)C^{-1}$. It follows that $(A - I)^m$ is of the form

$$(A-I)^m = C\begin{pmatrix} (z_1-1)^m & & 0 \\ & \ddots & \\ 0 & & (z_n-1)^m \end{pmatrix} C^{-1},$$

where $z_1, \cdots, z_n$ are the eigenvalues of A.

Now, if $\|A - I\| < 1$, then certainly $|z_i - 1| < 1$ for $i = 1, \cdots, n$. (Think about it.) Thus

$$\sum_{m=1}^{\infty} (-1)^{m+1} \frac{(A-I)^m}{m} = C\begin{pmatrix} \log z_1 & & 0 \\ & \ddots & \\ 0 & & \log z_n \end{pmatrix} C^{-1}$$

and so by the Lemma

$$e^{\log A} = C\begin{pmatrix} e^{\log z_1} & & 0 \\ & \ddots & \\ 0 & & e^{\log z_n} \end{pmatrix} C^{-1} = A.$$

Case 2: A is not diagonalizable.

If A is not diagonalizable, then, using the Jordan canonical form, it is not difficult to construct a sequence A_m of diagonalizable matrices with $A_m \to A$. (See Exercise 4.) If $\|A - I\| < 1$, then $\|A_m - I\| < 1$ for all sufficiently large m. By Case 1, $\exp(\log A_m) = A_m$, and so by the continuity of exp and log, $\exp(\log A) = A$.

Thus we have shown that $\exp(\log A) = A$ for all A with $\|A - I\| < 1$. Now, the same argument as in the complex case shows that if $\|X\| < \log 2$, then $\left\|e^X - I\right\| < 1$. But then the same two-case argument as above shows that $\log(\exp X) = X$ for all such X. $\square$

Proposition 35 *There exists a constant c such that for all $n \times n$ matrices B with $\|B\| < \frac{1}{2}$*

$$\|\log(I + B) - B\| \leq c\,\|B\|^2.$$

Proof

Note that

$$\log(I+B)-B=\sum_{m=2}^{\infty}(-1)^m\frac{B^m}{m}=B^2\sum_{m=2}^{\infty}(-1)^m\frac{B^{m-2}}{m}$$

so that

$$\|\log(I+B)-B\|\leq\|B\|^2\sum_{m=2}^{\infty}\frac{\left(\frac{1}{2}\right)^m}{m}.$$

This is what we want. □

Proposition 36 *Let X be any $n\times n$ complex matrix, and let C_m be a sequence of matrices such that $\|C_m\|\leq\frac{\text{const.}}{m^2}$. Then*

$$\lim_{m\to\infty}\left[I+\frac{X}{m}+C_m\right]^m=e^X.$$

Proof

The expression inside the brackets is clearly tending to I as $m\to\infty$, and so is in the domain of the logarithm for all sufficiently large m. Now

$$\log\left(I+\frac{X}{m}+C_m\right)=\frac{X}{m}+C_m+E_m$$

where E_m is an error term which, by Proposition 35 satisfies $\|E_m\|\leq c\left\|\frac{X}{m}+C_m\right\|^2\leq\frac{\text{const.}}{m^2}$. But then

$$I+\frac{X}{m}+C_m=\exp\left(\frac{X}{m}+C_m+E_m\right),$$

and so

$$\left[I+\frac{X}{m}+C_m\right]^m=\exp\left(X+mC_m+mE_m\right).$$

Since both C_m and E_m are of order $\frac{1}{m^2}$, we obtain the desired result by letting $m\to\infty$ and using the continuity of the exponential. □

3.4 Further Properties of the Matrix Exponential

In this section we give three additional results involving the exponential of a matrix, which will be important in our study of Lie algebras.

Theorem 37 (Lie Product Formula) *Let X and Y be $n\times n$ complex matrices. Then*

$$e^{X+Y}=\lim_{m\to\infty}\left(e^{\frac{X}{m}}e^{\frac{Y}{m}}\right)^m.$$

Remark

This theorem has a big brother, called the Trotter product formula, which gives the same result in the case where X and Y are suitable unbounded operators on an infinite-dimensional Hilbert space. The Trotter formula is described, for example, in M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I, VIII.8.

Proof

Using the power series for the exponential and multiplying, we get

$$e^{\frac{X}{m}}e^{\frac{Y}{m}} = I + \frac{X}{m} + \frac{Y}{m} + C_m,$$

where (check!) $\|C_m\| \leq \frac{const.}{m^2}$. Since $e^{\frac{X}{m}}e^{\frac{Y}{m}} \to I$ as $m \to \infty$, $e^{\frac{X}{m}}e^{\frac{Y}{m}}$ is in the domain of the logarithm for all sufficiently large m. But

$$\begin{aligned} \log\left(e^{\frac{X}{m}}e^{\frac{Y}{m}}\right) &= \log\left(I + \frac{X}{m} + \frac{Y}{m} + C_m\right) \\ &= \frac{X}{m} + \frac{Y}{m} + C_m + E_m \end{aligned}$$

where by Proposition 35 $\|C_m\| \leq const. \left\|\frac{X}{m} + \frac{Y}{m} + C_m\right\|^2 \leq \frac{const.}{m^2}$. Exponentiating the logarithm gives

$$e^{\frac{X}{m}}e^{\frac{Y}{m}} = \exp\left(\frac{X}{m} + \frac{Y}{m} + C_m + E_m\right)$$

and

$$\left(e^{\frac{X}{m}}e^{\frac{Y}{m}}\right)^m = \exp\left(X + Y + mC_m + mE_m\right).$$

Since both C_m and E_m are of order $\frac{1}{m^2}$, we have (using the continuity of the exponential)

$$\lim_{m\to\infty}\left(e^{\frac{X}{m}}e^{\frac{Y}{m}}\right)^m = \exp\left(X + Y\right)$$

which is the Lie product formula. □

Theorem 38 *Let X be an $n \times n$ real or complex matrix. Then*

$$\det\left(e^X\right) = e^{\mathrm{trace}(X)}.$$

Proof

There are three cases, as in Section 3.2.

Case 1: A is diagonalizable. Suppose there is a complex invertible matrix C such that

$$X = C \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} C^{-1}.$$

Then

$$e^X = C \begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix} C^{-1}.$$

Thus $\operatorname{trace}(X) = \sum \lambda_i$, and $\det(e^X) = \prod e^{\lambda_i} = e^{\sum \lambda_i}$. (Recall that $\operatorname{trace}(CDC^{-1}) = \operatorname{trace}(D)$.)

Case 2: X is nilpotent. If X is nilpotent, then it cannot have any non-zero eigenvalues (check!), and so all the roots of the characteristic polynomial must be zero. Thus the Jordan canonical form of X will be strictly upper triangular. That is, X can be written as

$$X = C \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} C^{-1}.$$

In that case (it is easy to see) e^X will be upper triangular, with ones on the diagonal:

$$e^X = C \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix} C^{-1}.$$

Thus if X is nilpotent, $\operatorname{trace}(X) = 0$, and $\det(e^X) = 1$.

Case 3: X arbitrary. As pointed out in Section 3.2, every matrix X can be written as the sum of two commuting matrices S and N, with S diagonalizable (over $\mathbb{C}$) and N nilpotent. Since S and N commute, $e^X = e^S e^N$. So by the two previous cases

$$\det\left(e^X\right) = \det\left(e^S\right)\det\left(e^N\right) = e^{\operatorname{trace}(S)} e^{\operatorname{trace}(N)} = e^{\operatorname{trace}(X)},$$

which is what we want. □

Definition 39 *A function $A : \mathbb{R} \to \mathsf{GL}(n;\mathbb{C})$ is called a **one-parameter group** if*

1. *A is continuous,*

2. *$A(0) = I$,*

3. *$A(t+s) = A(t)A(s)$ for all $t, s \in \mathbb{R}$.*

Theorem 40 (One-parameter Subgroups) *If A is a one-parameter group in $\mathsf{GL}(n;\mathbb{C})$, then there exists a unique $n \times n$ complex matrix X such that*

$$A(t) = e^{tX}.$$

Remark

By taking $n = 1$, and noting that $\mathsf{GL}(1;\mathbb{C}) \cong \mathbb{C}^*$, this Theorem provides an alternative method of solving Exercise 16 in Chapter 2.

Proof

The uniqueness is immediate, since if there is such an X, then $X = \frac{d}{dt}\big|_{t=0} A(t)$. So we need only worry about existence.

The first step is to show that $A(t)$ must be smooth. This follows from Proposition 28 in Chapter 2 (which we did not prove), but we give a self-contained proof. Let $f(s)$ be a smooth real-valued function supported in a small neighborhood of zero, with $f(s) \geq 0$ and $\int f(s)ds = 1$. Now look at

$$B(t) = \int A(t+s)f(s)\, ds. \tag{3.7}$$

Making the change-of-variable $u = t + s$ gives

$$B(t) = \int A(u)f(u-t)\, du.$$

It follows that $B(t)$ is differentiable, since derivatives in the t variable go onto f, which is smooth.

On the other hand, if we use the identity $A(t+s) = A(t)A(s)$ in (3.7), we have

$$B(t) = A(t) \int A(s)f(s)\, ds.$$

Now, the conditions on the function f, together with the continuity of A, guarantee that $\int A(s)f(s)\, ds$ is close to $A(0) = I$, and hence is invertible. Thus we may write

$$A(t) = B(t) \left(\int A(s)f(s)ds \right)^{-1}. \tag{3.8}$$

Since $B(t)$ is smooth and $\int A(s)f(s)ds$ is just a constant matrix, this shows that $A(t)$ is smooth.

Now that $A(t)$ is known to be differentiable, we may define

$$X = \left.\frac{d}{dt}\right|_{t=0} A(t).$$

Our goal is to show that $A(t) = e^{tX}$. Since $A(t)$ is smooth, a standard calculus result (extended trivially to handle matrix-valued functions) says

$$\|A(t) - (I + tX)\| \leq \text{const.}t^2.$$

It follows that for each fixed t,

$$A\left(\tfrac{t}{m}\right) = I + \tfrac{t}{m}X + O\left(\tfrac{1}{m^2}\right).$$

Then, since A is a one-parameter group

$$A(t) = \left[A\left(\tfrac{t}{m}\right)\right]^m = \left[I + \tfrac{t}{m}X + O\left(\tfrac{1}{m^2}\right)\right]^m.$$

Letting $m \to \infty$ and using Proposition 36 from Section 3.3 shows that $A(t) = e^{tX}$. $\square$

3.5 The Lie Algebra of a Matrix Lie Group

The Lie algebra is an indespensible tool in studying matrix Lie groups. On the one hand, Lie algebras are simpler than matrix Lie groups, because (as we will see) the Lie algebra is a linear space. Thus we can understand much about Lie algebras just by doing linear algebra. On the other hand, the Lie algebra of a matrix Lie group contains much information about that group. (See for example, Proposition 51 in Section 3.7, and the Baker-Campbell-Hausdorff Formula (Chapter 4).) Thus many questions about matrix Lie groups can be answered by considering a similar but easier problem for the Lie algebra.

Definition 41 *Let G be a matrix Lie group. Then the* ***Lie algebra*** *of G, denoted $\mathfrak{g}$, is the set of all matrices X such that e^{tX} is in G for all real numbers t.*

Remarks

Note that even if G is a subgroup of $\mathsf{GL}(n;\mathbb{C})$ we do *not* require that e^{tX} be in G for all complex t, but only for all *real* t. Also, it is definitely not enough to have just e^X in G. That is, it is easy to give an example of an X and a G such that $e^X \in G$ but $e^{tX} \notin G$ for some values of t. Such an X is not in the Lie algebra of G.

It is customary to use lower case Gothic (Fraktur) characters such as $\mathfrak{g}$ and $\mathfrak{h}$ to refer to Lie algebras.

Physicists' Convention

Physicists are accustomed to considering the map $X \to e^{iX}$ instead of $X \to e^{X}$. Thus a physicist would think of the Lie algebra of G as the set of all matrices X such that $e^{itX} \in G$ for all real t. In the physics literature, the Lie algebra is frequently refered to as the space of "infinitesimal group elements." See Bröcker and tom Dieck, Chapter I, 2.21. The physics literature does not always distinguish clearly between a matrix Lie group and its Lie algebra.

Before examining general properties of the Lie algebra, let us compute the Lie algebras of the matrix Lie groups introduced in the previous chapter.

The general linear groups

If X is any $n \times n$ complex matrix, then by Proposition 31, e^{tX} is invertible. Thus the Lie algebra of $\mathsf{GL}(n;\mathbb{C})$ is the space of all $n \times n$ complex matrices. This Lie algebra is denoted $\mathsf{gl}(n;\mathbb{C})$.

If X is any $n \times n$ real matrix, then e^{tX} will be invertible and real. On the other hand, if e^{tX} is real for all real t, then $X = \left.\frac{d}{dt}\right|_{t=0} e^{tX}$ will also be real. Thus the Lie algebra of $\mathsf{GL}(n;\mathbb{R})$ is the space of all $n \times n$ real matrices, denoted $\mathsf{gl}(n;\mathbb{R})$.

Note that the preceding argument shows that if G is a subgroup of $\mathsf{GL}(n;\mathbb{R})$, then the Lie algebra of G must consist entirely of real matrices. We will use this fact when appropriate in what follows.

The special linear groups

Recall Theorem 38: $\det\left(e^{X}\right) = e^{\mathrm{trace} X}$. Thus if $\mathrm{trace}X = 0$, then $\det\left(e^{tX}\right) = 1$ for all real t. On the other hand, if X is any $n \times n$ matrix such that $\det\left(e^{tX}\right) = 1$ for all t, then $e^{(t)(\mathrm{trace}X)} = 1$ for all t. This means that $(t)(\mathrm{trace}X)$ is an integer multiple of $2\pi i$ for all t, which is only possible if $\mathrm{trace}X = 0$. Thus the Lie algebra of $\mathsf{SL}\,(n;\mathbb{C})$ is the space of all $n \times n$ complex matrices with trace zero, denoted $\mathsf{sl}(n;\mathbb{C})$.

Similarly, the Lie algebra of $\mathsf{SL}\,(n;\mathbb{R})$ is the space of all $n \times n$ real matrices with trace zero, denoted $\mathsf{sl}\,(n;\mathbb{R})$.

The unitary groups

Recall that a matrix U is unitary if and only if $U^* = U^{-1}$. Thus e^{tX} is unitary if and only if

$$\left(e^{tX}\right)^* = \left(e^{tX}\right)^{-1} = e^{-tX}. \tag{3.9}$$

But by taking adjoints term-by-term, we see that $\left(e^{tX}\right)^* = e^{tX^*}$, and so (3.9) becomes

$$e^{tX^*} = e^{-tX}. \tag{3.10}$$

Clearly, a sufficient condition for (3.10) to hold is that $X^* = -X$. On the other hand, if (3.10) holds for all t, then by differentiating at $t = 0$, we see that $X^* = -X$ is necessary.

Thus the Lie algebra of $\mathsf{U}(n)$ is the space of all $n \times n$ complex matrices X such that $X^* = -X$, denoted $\mathsf{u}(n)$.

By combining the two previous computations, we see that the Lie algebra of $\mathsf{SU}(n)$ is the space of all $n \times n$ complex matrices X such that $X^* = -X$ and $\operatorname{trace} X = 0$, denoted $\mathsf{su}(n)$.

The orthogonal groups

The identity component of $\mathsf{O}(n)$ is just $\mathsf{SO}(n)$. Since (Proposition 42) the exponential of a matrix in the Lie algebra is automatically in the identity component, the Lie algebra of $\mathsf{O}(n)$ is the same as the Lie algebra of $\mathsf{SO}(n)$.

Now, an $n \times n$ real matrix R is orthogonal if and only if $R^{tr} = R^{-1}$. So, given an $n \times n$ real matrix X, e^{tX} is orthogonal if and only if $(e^{tX})^{tr} = (e^{tX})^{-1}$, or

$$e^{tX^{tr}} = e^{-tX}. \tag{3.11}$$

Clearly, a sufficient condition for this to hold is that $X^{tr} = -X$. If (3.11) holds for all t, then by differentiating at $t = 0$, we must have $X^{tr} = -X$.

Thus the Lie algebra of $\mathsf{O}(n)$, as well as the Lie algebra of $\mathsf{SO}(n)$, is the space of all $n \times n$ real matrices X with $X^{tr} = -X$, denoted $\mathsf{so}(n)$. Note that the condition $X^{tr} = -X$ forces the diagonal entries of X to be zero, and so explicitly the trace of X is zero.

The same argument shows that the Lie algebra of $\mathsf{SO}(n;\mathbb{C})$ is the space of $n \times n$ complex matrices satisfying $X^{tr} = -X$, denoted $\mathsf{so}(n;\mathbb{C})$. This is not the same as $\mathsf{su}(n)$.

The generalized orthogonal groups

A matrix A is in $\mathsf{O}(n;k)$ if and only if $A^{tr}gA = g$, where g is the $(n+k) \times (n+k)$ diagonal matrix with the first n diagonal entries equal to one, and the last k diagonal entries equal to minus one. This condition is equivalent to the condition $g^{-1}A^{tr}g = A^{-1}$, or, since explicitly $g^{-1} = g$, $gA^{tr}g = A^{-1}$. Now, if X is an $(n+k) \times (n+k)$ real matrix, then e^{tX} is in $\mathsf{O}(n;k)$ if and only if

$$ge^{tX^{tr}}g = e^{tgX^{tr}g} = e^{-tX}.$$

This condition holds for all real t if and only if $gX^{tr}g = -X$. Thus the Lie algebra of $\mathsf{O}(n;k)$, which is the same as the Lie algebra of $\mathsf{SO}(n;k)$, consists of all $(n+k)\times(n+k)$ real matrices X with $gX^{tr}g = -X$. This Lie algebra is denoted $\mathsf{so}(n;k)$.

(In general, the group $\mathsf{SO}(n;k)$ will not be connected, in contrast to the group $\mathsf{SO}(n)$. The identity component of $\mathsf{SO}(n;k)$, which is also the identity component of

$\mathsf{O}(n;k)$, is denoted $\mathsf{SO}(n;k)_I$. The Lie algebra of $\mathsf{SO}(n;k)_I$ is the same as the Lie algebra of $\mathsf{SO}(n;k)$.)

The symplectic groups

These are denoted $\mathsf{sp}(n;\mathbb{R})$, $\mathsf{sp}(n;\mathbb{C})$, and $\mathsf{sp}(n)$. The calculation of these Lie algebras is similar to that of the generalized orthogonal groups, and I will just record the result here. Let J be the matrix in the definition of the symplectic groups. Then $\mathsf{sp}(n;\mathbb{R})$ is the space of $2n \times 2n$ real matrices X such that $JX^{tr}J = X$, $\mathsf{sp}(n;\mathbb{C})$ is the space of $2n \times 2n$ complex matrices satisfying the same condition, and $\mathsf{sp}(n) = \mathsf{sp}(n;\mathbb{C}) \cap \mathsf{u}(2n)$.

The Heisenberg group

Recall the the Heisenberg group H is the group of all 3×3 real matrices A of the form

$$A = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \tag{3.12}$$

Recall also that in Section 3.2, Case 2, we computed the exponential of a matrix of the form

$$X = \begin{pmatrix} 0 & \alpha & \beta \\ 0 & 0 & \gamma \\ 0 & 0 & 0 \end{pmatrix} \tag{3.13}$$

and saw that e^X was in H. On the other hand, if X is any matrix such that e^{tX} is of the form (3.12), then all of the entries of $X = \left.\frac{d}{dt}\right|_{t=0} e^{tX}$ which are on or below the diagonal must be zero, so that X is of form (3.13).

Thus the Lie algebra of the Heisenberg group is the space of all 3×3 real matrices which are strictly upper triangular.

The Euclidean and Poincaré groups

Recall that the Euclidean group $\mathsf{E}(n)$ is (or can be thought of as) the group of $(n+1) \times (n+1)$ real matrices of the form

$$\begin{pmatrix} & & & x_1 \\ & R & & \vdots \\ & & & x_n \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

with $R \in \mathsf{O}(n)$. Now if X is an $(n+1)\times(n+1)$ real matrix such that e^{tX} is in $\mathsf{E}(n)$ for all t, then $X = \left.\frac{d}{dt}\right|_{t=0} e^{tX}$ must be zero along the bottom row:

$$X = \begin{pmatrix} & & y_1 \\ & Y & \vdots \\ & & y_n \\ 0 & \cdots & 0 \end{pmatrix} \tag{3.14}$$

Our goal, then, is to determine which matrices of the form (3.14) are actually in the Lie algebra of the Euclidean group. A simple computation shows that for $n \geq 1$

$$\begin{pmatrix} & & y_1 \\ & Y & \vdots \\ & & y_n \\ 0 & \cdots & 0 \end{pmatrix}^n = \begin{pmatrix} & Y^n & Y^{n-1}y \\ & & \\ 0 & \cdots & 0 \end{pmatrix},$$

where y is the column vector with entries $y_1, \cdots, y_n$. It follows that if X is as in (3.14), then e^{tX} is of the form

$$e^{tX} = \begin{pmatrix} & & & * \\ & e^{tY} & & \vdots \\ & & & * \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Now, we have already established that e^{tY} is in $\mathsf{O}(n)$ for all t if and only if $Y^{tr} = -Y$. Thus we see that the Lie algebra of $\mathsf{E}(n)$ is the space of all $(n+1)\times(n+1)$ real matrices of the form (3.14) with Y satisfying $Y^{tr} = -Y$.

A similar argument shows that the Lie algebra of $\mathsf{P}(n;1)$ is the space of all $(n+2)\times(n+2)$ real matrices of the form

$$\begin{pmatrix} & & y_1 \\ & Y & \vdots \\ & & y_{n+1} \\ 0 & \cdots & 0 \end{pmatrix}$$

with $Y \in \mathsf{so}(n;1)$.

3.6 Properties of the Lie Algebra

We will now establish various basic properties of the Lie algebra of a matrix Lie group. The reader is invited to verify by direct calculation that these general properties hold for the examples computed in the previous section.

Proposition 42 *Let G be a matrix Lie group, and X an element of its Lie algebra. Then e^X is an element of the identity component of G.*

Proof

By definition of the Lie algebra, e^{tX} lies in G for all real t. But as t varies from 0 to 1, e^{tX} is a continuous path connecting the identity to e^{X}. $\square$

Proposition 43 *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$. Let X be an element of $\mathfrak{g}$, and A an element of G. Then AXA^{-1} is in $\mathfrak{g}$.*

Proof

This is immediate, since by Proposition 31,

$$e^{t(AXA^{-1})} = Ae^{tX}A^{-1},$$

and $Ae^{tX}A^{-1} \in G$. $\square$

Theorem 44 *Let G be a matrix Lie group, $\mathfrak{g}$ its Lie algebra, and X, Y elements of $\mathfrak{g}$. Then*

1. *$sX \in \mathfrak{g}$ for all real numbers s,*
2. *$X + Y \in \mathfrak{g}$,*
3. *$XY - YX \in \mathfrak{g}$.*

Remark

If you are following the physics convention for the definition of the Lie algebra, then condition 3 should be replaced with the condition $-i\,(XY - YX) \in \mathfrak{g}$.

Proof

Point 1 is immediate, since $e^{t(sX)} = e^{(ts)X}$, which must be in G if X is in $\mathfrak{g}$. Point 2 is easy to verify *if* X and Y commute, since then $e^{t(X+Y)} = e^{tX}e^{tY}$. If X and Y do not commute, this argument does not work. However, the Lie product formula says that

$$e^{t(X+Y)} = \lim_{m\to\infty} \left(e^{tX/m}e^{tY/m}\right)^m.$$

Because X and Y are in the Lie algebra, $e^{tX/m}$ and $e^{tY/m}$ are in G, as is $\left(e^{tX/m}e^{tY/m}\right)^m$, since G is a group. But now because G is a matrix Lie group, the limit of things in G must be again in G, provided that the limit is invertible. Since $e^{t(X+Y)}$ is automatically invertible, we conclude that it must be in G. This shows that $X + Y$ is in $\mathfrak{g}$.

Now for point 3. Recall (Proposition 32) that $\left.\frac{d}{dt}\right|_{t=0} e^{tX} = X$. It follows that $\left.\frac{d}{dt}\right|_{t=0} e^{tX}Y = XY$, and hence by the product rule (Exercise 1)

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=0} \left(e^{tX}Ye^{-tX}\right) &= (XY)e^0 + (e^0Y)(-X) \\ &= XY - YX. \end{aligned}$$

But now, by Proposition 43, $e^{tX}Ye^{-tX}$ is in $\mathfrak{g}$ for all t. Since we have (by points 1 and 2) established that $\mathfrak{g}$ is a real vector space, it follows that the derivative of any smooth curve lying in $\mathfrak{g}$ must be again in $\mathfrak{g}$. Thus $XY - YX$ is in $\mathfrak{g}$. $\square$

Definition 45 *Given two $n \times n$ matrices A and B, the* ***bracket*** *(or* ***commutator****) of A and B is defined to be simply*

$$[A, B] = AB - BA.$$

According to Theorem 44, the Lie algebra of any matrix Lie group is closed under brackets.

The following very important theorem tells us that a Lie group homomorphism between two Lie groups gives rise in a natural way to a map between the corresponding Lie algebras. In particular, this will tell us that two isomorphic Lie groups have "the same" Lie algebras. (That is, the Lie algebras are isomorphic in the sense of Section 3.8.) See Exercise 6.

Theorem 46 *Let G and H be matrix Lie groups, with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively. Suppose that $\phi : G \to H$ be a Lie group homomorphism. Then there exists a unique real linear map $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ such that*

$$\phi(e^X) = e^{\widetilde{\phi}(X)}$$

for all $X \in \mathfrak{g}$. The map $\widetilde{\phi}$ has following additional properties

1. *$\widetilde{\phi}\left(AXA^{-1}\right) = \phi(A)\widetilde{\phi}(X)\phi(A)^{-1}$, for all $X \in \mathfrak{g}$, $A \in G$.*

2. *$\widetilde{\phi}([X, Y]) = \left[\widetilde{\phi}(X), \widetilde{\phi}(Y)\right]$, for all $X, Y \in \mathfrak{g}$.*

3. *$\widetilde{\phi}(X) = \left.\frac{d}{dt}\right|_{t=0} \phi(e^{tX})$, for all $X \in \mathfrak{g}$.*

If G, H, and K are matrix Lie groups and $\phi : H \to K$ and $\psi : G \to H$ are Lie group homomorphisms, then

$$\widetilde{\phi \circ \psi} = \widetilde{\phi} \circ \widetilde{\psi}.$$

Remarks

In practice, given a Lie group homomorphism ϕ, the way one goes about computing $\widetilde{\phi}$ is by using Property 3. Of course, since $\widetilde{\phi}$ is (real) linear, it suffices to compute $\widetilde{\phi}$ on a basis for $\mathfrak{g}$. In the language of differentiable manifolds, Property 3 says that $\widetilde{\phi}$ is the derivative (or differential) of ϕ at the identity, which is the standard definition of $\widetilde{\phi}$. (See also Exercise 19.)

A linear map with propery (2) is called a **Lie algebra homomorphism**. (See Section 3.8.) This theorem says that every Lie group homomorphism gives rise to a Lie algebra homomorphism. We will see eventually that the converse is true *under certain circumstances*. Specifically, suppose that G and H are Lie groups, and $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism. If G is *connected and simply connected*, then there exists a unique Lie group homomorphism $\phi : G \to H$ such that ϕ and $\widetilde{\phi}$ are related as in Theorem 46.

Proof of the Theorem

The proof is similar to the proof of Theorem 44. Since ϕ is a continuous group homomorphism, $\phi(e^{tX})$ will be a one-parameter subgroup of H, for each $X \in \mathfrak{g}$. Thus by Theorem 40, there is a unique Z such that

$$\phi\left(e^{tX}\right) = e^{tZ} \tag{3.15}$$

for all $t \in \mathbb{R}$. This Z must lie in $\mathfrak{h}$ since $e^{tZ} = \phi\left(e^{tX}\right) \in H$.

We now define $\widetilde{\phi}(X) = Z$, and check in several steps that $\widetilde{\phi}$ has the required properties.

Step 1: $\phi(e^X) = e^{\widetilde{\phi}(X)}$.
This follows from (3.15) and our definition of $\widetilde{\phi}$, by putting $t = 1$.

Step 2: $\widetilde{\phi}(sX) = s\widetilde{\phi}(X)$ *for all* $s \in \mathbb{R}$.
This is immediate, since if $\phi(e^{tX}) = e^{tZ}$, then $\phi(e^{tsX}) = e^{tsZ}$.

Step 3: $\widetilde{\phi}(X + Y) = \widetilde{\phi}(X) + \widetilde{\phi}(Y)$.
By Steps 1 and 2,

$$e^{t\widetilde{\phi}(X+Y)} = e^{\widetilde{\phi}[t(X+Y)]} = \phi\left(e^{t(X+Y)}\right).$$

By the Lie product formula, and the fact that ϕ is a continuous homomorphism:

$$\begin{aligned} &= \phi\left(\lim_{m\to\infty}\left(e^{tX/m}e^{tY/m}\right)^m\right) \\ &= \lim_{m\to\infty}\left(\phi\left(e^{tX/m}\right)\phi(e^{tY/m})\right)^m. \end{aligned}$$

But then we have

$$e^{t\widetilde{\phi}(X+Y)} = \lim_{m\to\infty}\left(e^{t\widetilde{\phi}(X)/m}e^{t\widetilde{\phi}(Y)/m}\right)^m = e^{t\left(\widetilde{\phi}(X)+\widetilde{\phi}(Y)\right)}.$$

Differentiating this result at $t = 0$ gives the desired result.

Step 4: $\widetilde{\phi}\,(AXA^{-1}) = \phi(A)\widetilde{\phi}(X)\phi(A)^{-1}$.
By Steps 1 and 2,

$$\exp t\widetilde{\phi}(AXA^{-1}) = \exp\widetilde{\phi}(tAXA^{-1}) = \phi\left(\exp tAXA^{-1}\right).$$

Using a property of the exponential and Step 1, this becomes

$$\begin{aligned}\exp t\widetilde{\phi}(AXA^{-1}) &= \phi\left(Ae^{tX}A^{-1}\right) = \phi(A)\phi(e^{tX})\phi(A)^{-1}\\ &= \phi(A)e^{t\widetilde{\phi}(X)}\phi(A)^{-1}.\end{aligned}$$

Differentiating this at $t = 0$ gives the desired result.

Step 5: $\widetilde{\phi}([X,Y]) = \left[\widetilde{\phi}(X),\widetilde{\phi}(Y)\right]$.
Recall from the proof of Theorem 44 that

$$[X,Y] = \left.\tfrac{d}{dt}\right|_{t=0} e^{tX}Ye^{-tX}.$$

Hence

$$\widetilde{\phi}\,([X,Y]) = \widetilde{\phi}\left(\left.\tfrac{d}{dt}\right|_{t=0} e^{tX}Ye^{-tX}\right) = \left.\tfrac{d}{dt}\right|_{t=0}\widetilde{\phi}\left(e^{tX}Ye^{-tX}\right)$$

where we have used the fact that a derivative commutes with a linear tranformation.
But then by Step 4,

$$\begin{aligned}\widetilde{\phi}\,([X,Y]) &= \left.\tfrac{d}{dt}\right|_{t=0}\phi(e^{tX})\widetilde{\phi}(Y)\phi(e^{-tX})\\ &= \left.\tfrac{d}{dt}\right|_{t=0} e^{t\widetilde{\phi}(X)}\widetilde{\phi}(Y)e^{-t\widetilde{\phi}(X)}\\ &= \left[\widetilde{\phi}(X),\widetilde{\phi}(Y)\right].\end{aligned}$$

Step 6: $\widetilde{\phi}(X) = \left.\tfrac{d}{dt}\right|_{t=0}\phi(e^{tX})$.
This follows from (3.15) and our definition of $\widetilde{\phi}$.

Step 7: $\widetilde{\phi}$ *is the unique real-linear map such that* $\phi(e^X) = e^{\widetilde{\phi}(X)}$.
Suppose that ψ is another such map. Then

$$e^{t\psi(X)} = e^{\psi(tX)} = \phi(e^{tX})$$

so that

$$\psi(X) = \tfrac{d}{dt}\Big|_{t=0} \phi(e^{tX}).$$

Thus by Step 6, ψ coincides with $\widetilde{\phi}$.

Step 8: $\widetilde{\phi \circ \psi} = \widetilde{\phi} \circ \widetilde{\psi}$.
For any $X \in \mathfrak{g}$,

$$\phi \circ \psi\left(e^{tX}\right) = \phi\left(\psi\left(e^{tX}\right)\right) = \phi\left(e^{t\widetilde{\psi}(X)}\right) = e^{t\widetilde{\phi}(\widetilde{\psi}(X))}.$$

Thus $\widetilde{\phi \circ \psi}(X) = \widetilde{\phi} \circ \widetilde{\psi}(X)$. $\square$

Definition 47 (The Adjoint Mapping) *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$. Then for each $A \in G$, define a linear map $\mathrm{Ad}A : \mathfrak{g} \to \mathfrak{g}$ by the formula*

$$\mathrm{Ad}A(X) = AXA^{-1}.$$

We will let Ad denote the map $A \to \mathrm{Ad}A$.

Proposition 48 *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$. Then for each $A \in G$, $\mathrm{Ad}A$ is an invertible linear tranformation of $\mathfrak{g}$ with inverse $\mathrm{Ad}A^{-1}$, and $\mathrm{Ad} : G \to \mathsf{GL}(\mathfrak{g})$ is a group homomorphism.*

Proof

Easy. Note that Proposition 43 guarantees that $\mathrm{Ad}A(X)$ is actually in $\mathfrak{g}$ for all $X \in \mathfrak{g}$. $\square$

Since $\mathfrak{g}$ is a real vector space with some dimension k, $\mathsf{GL}(\mathfrak{g})$ is essentially the same as $\mathsf{GL}(k;\mathbb{R})$. Thus we will regard $\mathsf{GL}(\mathfrak{g})$ as a matrix Lie group. It is easy to show that $\mathrm{Ad} : G \to \mathsf{GL}(\mathfrak{g})$ is continuous, and so is a Lie group homomorphism. By Theorem 46, there is an associated real linear map $\widetilde{\mathrm{Ad}}$ from the Lie algebra of G to the Lie algebra of $\mathsf{GL}(\mathfrak{g})$, i.e., from $\mathfrak{g}$ to $\mathsf{gl}(\mathfrak{g})$, with the property that

$$e^{\widetilde{\mathrm{Ad}}X} = \mathrm{Ad}\left(e^X\right).$$

Proposition 49 *Let G be a matrix Lie group, let $\mathfrak{g}$ its Lie algebra, and let $\mathrm{Ad} : G \to \mathsf{GL}(\mathfrak{g})$ be the Lie group homomorphism defined above. Let $\widetilde{\mathrm{Ad}} : \mathfrak{g} \to gl(\mathfrak{g})$ be the associated Lie algebra map. Then for all $X, Y \in \mathfrak{g}$*

$$\widetilde{\mathrm{Ad}}X(Y) = [X, Y].$$

Proof

Recall that by Theorem 46, $\widetilde{\mathrm{Ad}}$ can be computed as follows:

$$\widetilde{\mathrm{Ad}}X = \left.\tfrac{d}{dt}\right|_{t=0} \mathrm{Ad}(e^{tX}).$$

Thus

$$\begin{aligned}\widetilde{\mathrm{Ad}}X(Y) &= \left.\tfrac{d}{dt}\right|_{t=0} \mathrm{Ad}(e^{tX})(Y) = \left.\tfrac{d}{dt}\right|_{t=0} e^{tX}Ye^{-tX} \\ &= [X,Y]\end{aligned}$$

which is what we wanted to prove. See also Exercise 13. □

3.7 The Exponential Mapping

Definition 50 *If G is a matrix Lie group with Lie algebra $\mathfrak{g}$, then the **exponential mapping** for G is the map*

$$\exp : \mathfrak{g} \to G.$$

In general the exponential mapping is neither one-to-one nor onto. Nevertheless, it provides an crucial mechanism for passing information between the group and the Lie algebra. The following result says that the exponential mapping is *locally* one-to-one and onto, a result that will be essential later.

Proposition 51 *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Then there exist a neighborhood U of zero in $\mathfrak{g}$ and a neighborhood V of I in G such that the exponential mapping takes U homeomorphically onto V.*

Proof

We follow the proof of Theorem I.3.11 in Bröcker and tom Dieck. In view of what we have proved about the matrix logarithm, we know this result for the case of $\mathsf{GL}(n;\mathbb{C})$. To prove the general case, we consider a matrix Lie group $G < \mathsf{GL}(n;\mathbb{C})$, with Lie algebra $\mathfrak{g}$.

Lemma 52 *Suppose g_n are elements of G, and that $g_n \to I$. Let $Y_n = \log g_n$, which is defined for all sufficiently large n. Suppose $Y_n/\left\|Y_n\right\| \to Y \in \mathsf{gl}(n;\mathbb{C})$. Then $Y \in \mathfrak{g}$.*

Proof of Lemma

To show that $Y \in \mathfrak{g}$, we must show that $\exp tY \in G$ for all $t \in \mathbb{R}$. As $n \to \infty$, $(t/\left\|Y_n\right\|)Y_n \to tY$. Note that since $g_n \to I$, $Y_n \to 0$, and so $\left\|Y_n\right\| \to 0$. Thus we can find integers m_n such that $(m_n\left\|Y_n\right\|) \to t$. Then $\exp(m_nY_n) =$

$\exp[(m_n\|Y_n\|)(Y_n/\|Y_n\|)] \to \exp(tY)$. But $\exp(m_nY_n) = \exp(Y_n)^{m_n} = (g_n)^{m_n} \in G$, and G is closed, so $\exp(tY) \in G$. □

We think of $\mathsf{gl}(n;\mathbb{C})$ as $\mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$. Then $\mathfrak{g}$ is a subspace of $\mathbb{R}^{2n^2}$. Let D denote the orthogonal complement of $\mathfrak{g}$ with respect to the usual inner product on $\mathbb{R}^{2n^2}$. Consider the map $\Phi : \mathfrak{g} \oplus D \to \mathsf{GL}(n;\mathbb{C})$ given by

$$\Phi(X,Y) = e^X e^Y.$$

Of course, we can identify $\mathfrak{g} \oplus D$ with $\mathbb{R}^{2n^2}$. Moreover, $\mathsf{GL}(n;\mathbb{C})$ is an open subset of $\mathsf{gl}(n;\mathbb{C}) \cong \mathbb{R}^{2n^2}$. Thus we can regard Φ as a map from $\mathbb{R}^{2n^2}$ to itself.

Now, using the properties of the matrix exponential, we see that

$$\left.\frac{d}{dt}\right|_{t=0} \Phi(tX,0) = X$$
$$\left.\frac{d}{dt}\right|_{t=0} \Phi(0,tY) = Y.$$

This shows that the derivative of Φ at the point $0 \in \mathbb{R}^{2n^2}$ is the identity. (Recall that the derivative at a point of a function from $\mathbb{R}^{2n^2}$ to itself is a linear map of $\mathbb{R}^{2n^2}$ to itself, in this case the identity map.) In particular, the derivative of Φ at 0 is invertible. Thus the inverse function theorem says that Φ has a continuous local inverse, defined in a neigborhood of I.

Now let U be any neighborhood of zero in $\mathfrak{g}$. I want to show that $\exp(U)$ contains a neighborhood of I in G. Suppose not.Then we can find a sequence $g_n \in G$ with $g_n \to I$ such that no g_n is in $\exp(U)$. Since Φ is locally invertible, we can write g_n (for large n) uniquely as $g_n = \exp(X_n)\exp(Y_n)$, with $X_n \in \mathfrak{g}$ and $Y_n \in D$. Since $g_n \to I$ and Φ^{-1} is continuous, X_n and Y_n tend to zero. Thus (for large n), $X_n \in U$. So we must have (for large n) $Y_n \neq 0$, otherwise g_n would be in $\exp(U)$.

Let $\widetilde{g}_n = \exp(Y_n) = \exp(-X_n)\,g_n$. Note that $\widetilde{g}_n \in G$ and $\widetilde{g}_n \to I$. Since the unit ball in D is compact, we can choose a subsequence of $\{Y_n\}$ (still called $\{Y_n\}$) so that $Y_n/\|Y_n\|$ converges to some $Y \in D$, with $\|Y\| = 1$. But then by the Lemma, $Y \in \mathfrak{g}$! This is a contradiction, because D is the orthogonal complement of $\mathfrak{g}$.

So for every neighborhood U of zero in $\mathfrak{g}$, $\exp(U)$ contains a neighborhood of the identity in G. If we make U small enough, then the exponential will be one-to-one on $\overline{U}$. (The existence of the matrix logarithm implies that the exponential is one-to-one near zero.) Let log denote the inverse map, defined on $\exp\left(\overline{U}\right)$. Since $\overline{U}$ is compact, and exp is one-to-one and continuous on $\overline{U}$, log will be continuous. (This is a standard topological result.) So take V to be a neighborhood of I contained in $\exp\left(\overline{U}\right)$, and let $U' = \exp^{-1}(V) \cap U$. Then U' is open and the exponential takes U' homeomorphically onto V. □

Definition 53 *If U and V are as in Proposition 51, then the inverse map* $\exp^{-1} : V \to \mathfrak{g}$ *is called the logarithm for G.*

Corollary 54 *If G is a connected matrix Lie group, then every element A of G can be written in the form*

$$A = e^{X_1} e^{X_2} \cdots e^{X_n} \tag{3.16}$$

for some $X_1, X_2, \cdots X_n$ in $\mathfrak{g}$.

Proof

Recall that for us, saying G is connected means that G is path-connected. This certainly means that G is connected in the usual topological sense, namely, the only non-empty subset of G that is both open and closed is G itself. So let E denote the set of all $A \in G$ that can be written in the form (3.16). In light of the Proposition, E contains a neighborhood V of the identity. In particular, E is non-empty.

We first claim that E is open. To see this, consider $A \in E$. Then look at the set of matrices of the form AB, with $B \in V$. This will be a neighborhood of A. But every such B can be written as $B = e^X$ and A can be written as $A = e^{X_1} e^{X_2} \cdots e^{X_n}$, so $AB = e^{X_1} e^{X_2} \cdots e^{X_n} e^X$.

Now we claim that E is closed (in G). Suppose $A \in G$, and there is a sequence $A_n \in E$ with $A_n \to A$. Then $AA_n^{-1} \to I$. Thus we can choose some n_0 such that $AA_{n_0}^{-1} \in V$. Then $AA_{n_0}^{-1} = e^X$ and $A = A_{n_0} e^X$. But by assumption, $A_{n_0} = e^{X_1} e^{X_2} \cdots e^{X_n}$, so $A = e^{X_1} e^{X_2} \cdots e^{X_n} e^X$. Thus $A \in E$, and E is closed.

Thus E is both open and closed, so $E = G$. $\square$

3.8 Lie Algebras

Definition 55 *A **finite-dimensional real or complex Lie algebra** is a finite-dimensional real or complex vector space $\mathfrak{g}$, together with a map $[\ ]$ from $\mathfrak{g} \times \mathfrak{g}$ into $\mathfrak{g}$, with the following properties:*

1. *$[\ ]$ is bilinear.*
2. *$[X, Y] = -[Y, X]$ for all $X, Y \in \mathfrak{g}$.*
3. *$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ for all $X, Y, Z \in \mathfrak{g}$.*

Remarks

Condition 3 is called the **Jacobi identity**. Note also that Condition 2 implies that $[X, X] = 0$ for all $X \in \mathfrak{g}$. The same three conditions define a Lie algebra over an arbitrary field $\mathbf{F}$, except that if $\mathbf{F}$ has characteristic two, then one should add the condition $[X, X] = 0$, which doesn't follow from skew-symmetry in characteristic two. We will deal only with finite-dimensional Lie algebras, and will from now on interpret "Lie algebra" as "finite-dimensional Lie algebra."

A Lie algebra is in fact an algebra in the usual sense, but the product operation [ ] for this algebra is neither commutative nor associative. The Jacobi identity should be thought of as a substitute for associativity.

Proposition 56 *The space* $\mathsf{gl}(n;\mathbb{R})$ *of all* $n \times n$ *real matrices is a real Lie algebra with respect to the bracket operation* $[A,B] = AB - BA$. *The space* $\mathsf{gl}(n;\mathbb{C})$ *of all* $n \times n$ *complex matrices is a complex Lie algebra with respect to the analogous bracket operation.*

Let V *is a finite-dimensional real or complex vector space, and let* $\mathsf{gl}(V)$ *denote the space of linear maps of* V *into itself. Then* $\mathsf{gl}(V)$ *becomes a real or complex Lie algebra with the bracket operation* $[A,B] = AB - BA$.

Proof

The only non-trivial point is the Jacobi identity. The only way to prove this is to write everything out and see, and this is best left to the reader. Note that each triple bracket generates four terms, for a total of twelve. Each of the six orderings of $\{X, Y, Z\}$ occurs twice, once with a plus sign and once with a minus sign. $\square$

Definition 57 *A* ***subalgebra*** *of a real or complex Lie algebra* $\mathfrak{g}$ *is a subspace* $\mathfrak{h}$ *of* $\mathfrak{g}$ *such that* $[H_1, H_2] \in \mathfrak{h}$ *for all* $H_1, H_2 \in \mathfrak{h}$. *If* $\mathfrak{g}$ *is a complex Lie algebra, and* $\mathfrak{h}$ *is a real subspace of* $\mathfrak{g}$ *which is closed under brackets, then* $\mathfrak{h}$ *is said to be a* ***real subalgebra*** *of* $\mathfrak{g}$.

If $\mathfrak{g}$ *and* $\mathfrak{h}$ *are Lie algebras, then a linear map* $\phi : \mathfrak{g} \to \mathfrak{h}$ *is called a* ***Lie algebra homomorphism*** *if* $\phi([X,Y]) = [\phi(X), \phi(Y)]$ *for all* $X, Y \in \mathfrak{g}$. *If in addition* ϕ *is one-to-one and onto, then* ϕ *is called a* ***Lie algebra isomorphism***. *A Lie algebra isomorphism of a Lie algebra with itself is called a* ***Lie algebra automorphism***.

Remarks

A subalgebra of a Lie algebra is again a Lie algebra. A real subalgebra of a complex Lie algebra is a real Lie algebra. The inverse of a Lie algebra isomorphism is again a Lie algebra isomorphism.

Proposition 58 *The Lie algebra* $\mathfrak{g}$ *of a matrix Lie group* G *is a real Lie algebra.*

Proof

By Theorem 44, $\mathfrak{g}$ is a real subalgebra of $\mathsf{gl}(n;\mathbb{C})$ complex matrices, and is thus a real Lie algebra. $\square$

Theorem 59 (Ado) *Every finite-dimensional real Lie algebra is isomorphic to a subalgebra of* $\mathsf{gl}(n;\mathbb{R})$. *Every finite-dimensional complex Lie algebra is isomorphic to a (complex) subalgebra of* $\mathsf{gl}(n;\mathbb{C})$.

This remarkable theorem is proved in Varadarajan. The proof is well beyond the scope of this course (which is after all a course on Lie *groups*), and requires a deep understanding of the structure of complex Lie algebras. The theorem tells us that every Lie algebra is (isomorphic to) a Lie algebra of matrices. (This is in contrast to the situation for Lie groups, where most but not all Lie groups are matrix Lie groups.)

Definition 60 *Let $\mathfrak{g}$ be a Lie algebra. For $X \in \mathfrak{g}$, define a linear map* $\mathrm{ad}X : \mathfrak{g} \to \mathfrak{g}$ *by*

$$\mathrm{ad}X(Y) = [X, Y].$$

*Thus "*ad*" (i.e., the map $X \to \mathrm{ad}X$) can be viewed as a linear map from $\mathfrak{g}$ into $\mathsf{gl}(\mathfrak{g})$, where $\mathsf{gl}(\mathfrak{g})$ denotes the space of linear operators from $\mathfrak{g}$ to $\mathfrak{g}$.*

Remark

Since $\mathrm{ad}X(Y)$ is just $[X, Y]$, it might seem foolish to introduce the additional "ad" notation. However, thinking of $[X, Y]$ as a linear map in Y for each fixed X, gives a somewhat different perspective. In any case, the "ad" notation is extremely useful in some situations. For example, instead of writing

$$[X, [X, [X, [X, Y]]]]$$

we can now write

$$(\mathrm{ad}X)^4 (Y).$$

This kind of notation will be essential in Section 4.1.

Proposition 61 *If $\mathfrak{g}$ is a Lie algebra, then*

$$\mathrm{ad}[X, Y] = \mathrm{ad}X\mathrm{ad}Y - \mathrm{ad}Y\mathrm{ad}X = [\mathrm{ad}X, \mathrm{ad}Y].$$

That is, ad: $\mathfrak{g} \to \mathsf{gl}(\mathfrak{g})$ is a Lie algebra homomorphism.

Proof

Observe that

$$\mathrm{ad}[X, Y](Z) = [[X, Y], Z]$$

whereas

$$[\mathrm{ad}X, \mathrm{ad}Y](Z) = [X, [Y, Z]] - [Y, [X, Z]].$$

So we require that

$$[[X,Y],Z] = [X,[Y,Z]] - [Y,[X,Z]]$$

or equivalently

$$0 = [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]]$$

which is exactly the Jacobi identity. $\square$

Recall that for any $X \in \mathfrak{g}$, and any $A \in G$, we define

$$\mathrm{Ad}A(X) = AXA^{-1}$$

and that Ad: $G \to \mathsf{GL}(\mathfrak{g})$ is a Lie group homomorphism. We showed (Proposition 49) that the associated Lie algebra homomorphism $\widetilde{\mathrm{Ad}} : \mathfrak{g} \to \mathsf{gl}(\mathfrak{g})$ is given by

$$\widetilde{\mathrm{Ad}}X(Y) = [X,Y].$$

In our new notation, we may say

$$\widetilde{\mathrm{Ad}} = \mathrm{ad}$$

By the defining property of $\widetilde{\mathrm{Ad}}$, we have the following identity: For all $X \in \mathfrak{g}$,

$$\mathrm{Ad}(e^X) = e^{\mathrm{ad}X}. \tag{3.17}$$

Note that both sides of (3.17) are linear operators on the Lie algebra $\mathfrak{g}$. This is an important relation, which can also be verified directly, by expanding out both sides. (See Exercise 13.)

Structure Constants

Let $\mathfrak{g}$ be a finite-dimensional real or complex Lie algebra, and let $X_1, \cdots, X_n$ be a basis for $\mathfrak{g}$ (as a vector space). Then for each i, j, $[X_i, X_j]$ can be written uniquely in the form

$$[X_i, X_j] = \sum_{k=1}^{n} c_{ijk} X_k.$$

The constants c_{ijk} are called the **structure constants** of $\mathfrak{g}$ (with respect to the chosen basis). Clearly, the structure constants determine the bracket operation on $\mathfrak{g}$. In some of the literature, the structure constants play an important role, although we will not have occasion to use them in this course. (In the physics literature, the structure constants are defined as $[X_i, X_j] = \sqrt{-1}\sum_k c_{ijk} X_k$, reflecting the factor of $\sqrt{-1}$ difference between the physics definition of the Lie algebra and our own.)

The structure constants satisfy the following two conditions,

$$c_{ijk} + c_{jik} = 0$$
$$\sum_m (c_{ijm}c_{mkl} + c_{jkm}c_{mil} + c_{kim}c_{mjl}) = 0$$

for all i, j, k, l. The first of these conditions comes from the skew-symmetry of the bracket, and the second comes from the Jacobi identity. (The reader is invited to verify these conditions for himself.)

3.9 The Complexification of a Real Lie Algebra

Definition 62 *If V is a finite-dimensional real vector space, then the* ***complexification*** *of V, denoted $V_{\mathbb{C}}$, is the space of formal linear combinations*

$$v_1 + iv_2$$

with $v_1, v_2 \in V$. This becomes a real vector space in the obvious way, and becomes a complex vector space if we define

$$i(v_1 + iv_2) = -v_2 + iv_1.$$

Remarks

We could more pedantically define $V_{\mathbb{C}}$ to be the space of ordered pairs (v_1, v_2), but this is notationally cumbersome. It is straightforward to verify that the above definition really makes $V_{\mathbb{C}}$ into a complex vector space. We will regard V as a real subspace of $V_{\mathbb{C}}$ in the obvious way.

Proposition 63 *Let $\mathfrak{g}$ be a finite-dimensional real Lie algebra, and $\mathfrak{g}_{\mathbb{C}}$ its complexification (as a real vector space). Then the bracket operation on $\mathfrak{g}$ has a unique extension to $\mathfrak{g}_{\mathbb{C}}$ which makes $\mathfrak{g}_{\mathbb{C}}$ into a complex Lie algebra. The complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ is called the* ***complexification*** *of the real Lie algebra $\mathfrak{g}$.*

Proof

The uniqueness of the extension is obvious, since if the bracket operation on $\mathfrak{g}_{\mathbb{C}}$ is to be bilinear, then it must be given by

$$[X_1 + iX_2, Y_1 + iY_2] = ([X_1, Y_1] - [X_2, Y_2]) + i([X_1, Y_2] + [X_2, Y_1]). \tag{3.18}$$

To show exisitence, we must now check that (3.18) is really bilinear and skew-symmetric, and that it satisfies the Jacobi identity. It is clear that (3.18) is *real* bilinear, and skew-symmetric. The skew-symmetry means that if (3.18) is complex linear in the first factor, it is also complex linear in the second factor. Thus we need only show that

$$[i(X_1+iX_2), Y_1+iY_2] = i\,[X_1+iX_2, Y_1+iY_2]\,. \tag{3.19}$$

Well, the left side of (3.19) is

$$[-X_2+iX_1, Y_1+iY_2] = (-[X_2,Y_1]-[X_1,Y_2]) + i\,([X_1,Y_1]-[X_2,Y_2])$$

whereas the right side of (3.19) is

$$\begin{aligned} & i\,\{([X_1,Y_1]-[X_2,Y_2]) + i\,([X_2,Y_1]+[X_1,Y_2])\} \\ & = (-[X_2,Y_1]-[X_1,Y_2]) + i\,([X_1,Y_1]-[X_2,Y_2])\,, \end{aligned}$$

and indeed these are equal.

It remains to check the Jacobi identity. Of course, the Jacobi identity holds if X, Y, and Z are in $\mathfrak{g}$. But now observe that the expression on the left side of the Jacobi identity is (complex!) linear in X for fixed Y and Z. It follows that the Jacobi identity holds if X is in $\mathfrak{g}_{\mathbb{C}}$, and Y, Z in $\mathfrak{g}$. The same argument then shows that we can extend to Y in $\mathfrak{g}_{\mathbb{C}}$, and then to Z in $\mathfrak{g}_{\mathbb{C}}$. Thus the Jacobi identity holds in $\mathfrak{g}_{\mathbb{C}}$. $\square$

Proposition 64 *The Lie algebras* $\mathsf{gl}(n;\mathbb{C})$, $\mathsf{sl}(n;\mathbb{C})$, $\mathsf{so}(n;\mathbb{C})$, *and* $\mathsf{sp}(n;\mathbb{C})$ *are complex Lie algebras, as is the Lie algebra of the complex Heisenberg group. In addition, we have the following isomorphisms of complex Lie algebras*

$$\begin{array}{ccc} \mathsf{gl}\,(n;\mathbb{R})_{\mathbb{C}} & \cong & \mathsf{gl}(n;\mathbb{C}) \\ \mathsf{u}(n)_{\mathbb{C}} & \cong & \mathsf{gl}(n;\mathbb{C}) \\ \mathsf{sl}\,(n;\mathbb{R})_{\mathbb{C}} & \cong & \mathsf{sl}(n;\mathbb{C}) \\ \mathsf{so}(n)_{\mathbb{C}} & \cong & \mathsf{so}(n;\mathbb{C}) \\ \mathsf{sp}(n;\mathbb{R})_{\mathbb{C}} & \cong & \mathsf{sp}(n;\mathbb{C}) \\ \mathsf{sp}(n)_{\mathbb{C}} & \cong & \mathsf{sp}(n;\mathbb{C}). \end{array}$$

Proof

From the computations in the previous section we see easily that the specified Lie algebras are in fact complex subalgebras of $\mathsf{gl}(n;\mathbb{C})$, and hence are complex Lie algebras.

Now, $\mathsf{gl}(n;\mathbb{C})$ is the space of all $n\times n$ complex matrices, whereas $\mathsf{gl}\,(n;\mathbb{R})$ is the space of all $n\times n$ real matrices. Clearly, then, every $X\in\mathsf{gl}\,(n;\mathbb{C})$ can be written uniquely in the form X_1+iX_2, with $X_1, X_2 \in \mathsf{gl}\,(n;\mathbb{R})$. This gives us a complex vector space isomorphism of $\mathsf{gl}\,(n;\mathbb{R})_{\mathbb{C}}$ with $\mathsf{gl}(n;\mathbb{C})$, and it is a triviality to check that this is a Lie algebra isomorphism.

On the other hand, $\mathsf{u}(n)$ is the space of all $n \times n$ complex skew-self-adjoint matrices. But if X is any $n \times n$ complex matrix, then

$$\begin{aligned} X &= \frac{X - X^*}{2} + \frac{X + X^*}{2} \\ &= \frac{X - X^*}{2} + i\frac{(-iX) - (-iX)^*}{2}. \end{aligned}$$

Thus X can be written as a skew matrix plus i times a skew matrix, and it is easy to see that this decomposition is unique. Thus every X in $\mathsf{gl}(n;\mathbb{C})$ can be written uniquely as $X_1 + iX_2$, with X_1 and X_2 in $\mathsf{u}(n)$. It follows that $\mathsf{u}(n)_{\mathbb{C}} \cong \mathsf{gl}(n;\mathbb{C})$.

The verification of the remaining isomorphisms is similar, and is left as an exercise to the reader. □

Remarks

Note that $\mathsf{u}(n)_{\mathbb{C}} \cong \mathsf{gl}\,(n;\mathbb{R})_{\mathbb{C}} \cong \mathsf{gl}(n;\mathbb{C})$. However, $\mathsf{u}(n)$ is *not* isomorphic to $\mathsf{gl}\,(n;\mathbb{R})$, except when $n = 1$. The real Lie algebras $\mathsf{u}(n)$ and $\mathsf{gl}\,(n;\mathbb{R})$ are called **real forms** of the complex Lie algebra $\mathsf{gl}(n;\mathbb{C})$. A given complex Lie algebra may have several non-isomorphic real forms. See Exercise 11.

Physicists do not always clearly distinguish between a matrix Lie group and its (real) Lie algebra, or between a real Lie algebra and its complexification. Thus, for example, some references in the physics literature to $\mathsf{SU}(2)$ actually refer to the complexified Lie algebra, $\mathsf{sl}(2;\mathbb{C})$.

3.10 Exercises

1. *The product rule.* Recall that a matrix-valued function $A(t)$ is smooth if each $A_{ij}(t)$ is smooth. The derivative of such a function is defined as

$$\left(\frac{dA}{dt}\right)_{ij} = \frac{dA_{ij}}{dt}$$

or equivalently,

$$\frac{d}{dt}A(t) = \lim_{h\to 0}\frac{A(t+h) - A(t)}{h}.$$

Let $A(t)$ and $B(t)$ be two such functions. Prove that $A(t)B(t)$ is again smooth, and that

$$\frac{d}{dt}\left[A(t)B(t)\right] = \frac{dA}{dt}B(t) + A(t)\frac{dB}{dt}.$$

2. Using the Jordan canonical form, show that every $n \times n$ matrix A can be written as $A = S + N$, with S diagonalizable (over $\mathbb{C}$), N nilpotent, and $SN = NS$. Recall that the Jordan canonical form is block diagonal, with each block of the form
$$\begin{pmatrix} \lambda & & * \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}.$$

3. Let X and Y be $n \times n$ matrices. Show that there exists a constant C such that
$$\left\| e^{(X+Y)/m} - e^{X/m} e^{Y/m} \right\| \leq \frac{C}{m^2}$$
for all integers $m \geq 1$.

4. Using the Jordan canonical form, show that every $n \times n$ complex matrix A is the limit of a sequence of diagonalizable matrices.

   Hint: If the characteristic polynomial of A has n distinct roots, then A is diagonalizable.

5. Give an example of a matrix Lie group G and a matrix X such that $e^X \in G$, but $X \notin \mathfrak{g}$.

6. Show that two isomorphic matrix Lie groups have isomorphic Lie algebras.

7. *The Lie algebra* $\mathsf{so}(3;1)$. Write out explicitly the general form of a 4×4 real matrix in $\mathsf{so}(3;1)$.

8. Verify directly that Proposition 43 and Theorem 44 hold for the Lie algebra of $\mathsf{SU}(n)$.

9. *The Lie algebra* $\mathsf{su}(2)$. Show that the following matrices form a basis for the real Lie algebra $\mathsf{su}(2)$:
$$E_1 = \tfrac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad E_2 = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad E_3 = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$
   Compute $[E_1, E_2]$, $[E_2, E_3]$, and $[E_3, E_1]$. Show that there is an invertible linear map $\phi : \mathsf{su}(2) \to \mathbb{R}^3$ such that $\phi([X,Y]) = \phi(X) \times \phi(Y)$ for all $X, Y \in \mathsf{su}(2)$, where $\times$ denotes the cross-product on $\mathbb{R}^3$.

10. *The Lie algebras* $\mathsf{su}(2)$ *and* $\mathsf{so}(3)$. Show that the real Lie algebras $\mathsf{su}(2)$ and $\mathsf{so}(3)$ are isomorphic.

    Note: Nevertheless, the corresponding *groups* $\mathsf{SU}(2)$ and $\mathsf{SO}(3)$ are not isomorphic. (Although $\mathsf{SO}(3)$ is isomorphic to $\mathsf{SU}(2)/\{I, -I\}$.)

11. *The Lie algebras* $\mathsf{su}(2)$ *and* $\mathsf{sl}(2;\mathbb{R})$. Show that $\mathsf{su}(2)$ and $\mathsf{sl}(2;\mathbb{R})$ are not isomorphic Lie algebras, even though $\mathsf{su}(2)_{\mathbb{C}} \cong \mathsf{sl}(2;\mathbb{R})_{\mathbb{C}}$.

    Hint: Using Exercise 9, show that $\mathsf{su}(2)$ has no two-dimensional subalgebras.

12. Let G be a matrix Lie group, and $\mathfrak{g}$ its Lie algebra. For each $A \in G$, show that $\mathrm{Ad}A$ is a Lie algebra automorphism of $\mathfrak{g}$.

13. Ad *and* ad. Let X and Y be matrices. Show by induction that

$$(\mathrm{ad}X)^n\,(Y) = \sum_{k=0}^{n} \binom{n}{k} X^k Y (-X)^{n-k}.$$

    Now show by direct computation that

$$e^{\mathrm{ad}X}(Y) = \mathrm{Ad}(e^X)Y = e^X Y e^{-X}.$$

    You may assume that it is legal to multiply power series term-by-term. (This result was obtained indirectly in Equation 3.17.)

    Hint: Recall that Pascal's Triangle gives a relationship between things of the form $\binom{n+1}{k}$ and things of the form $\binom{n}{k}$.

14. *The complexification of a real Lie algebra.* Let $\mathfrak{g}$ be a real Lie algebra, $\mathfrak{g}_{\mathbb{C}}$ its complexification, and $\mathfrak{h}$ an arbitrary complex Lie algebra. Show that every real Lie algebra homomorphism of $\mathfrak{g}$ into $\mathfrak{h}$ extends uniquely to a complex Lie algebra homomorphism of $\mathfrak{g}_{\mathbb{C}}$ into $\mathfrak{h}$. (This is the **universal property** of the complexification of a real Lie algebra. This property can be used as an alternative definition of the complexification.)

15. *The exponential mapping for* $\mathsf{SL}\,(2;\mathbb{R})$. Show that the image of the exponential mapping for $\mathsf{SL}\,(2;\mathbb{R})$ consists of precisely those matrices $A \in \mathsf{SL}\,(2;\mathbb{R})$ such that $\mathrm{trace}\,(A) > -2$, together with the matrix $-I$ (which has trace -2). You will need to consider the possibilities for the eigenvalues of a matrix in the Lie algebra $\mathsf{sl}\,(2;\mathbb{R})$ and in the group $\mathsf{SL}\,(2;\mathbb{R})$. In the Lie algebra, show that the eigenvalues are of the form $(\lambda, -\lambda)$ or $(i\lambda, -i\lambda)$ with λ real. In the group, show that the eigenvalues are of the form $(\alpha, 1/a)$ or $(-a, -1/a)$ with a real and positive, or else of the form $\left(e^{i\theta}, e^{-i\theta}\right)$, with θ real. The case of a repeated eigenvalue ($(0,0)$ in the Lie algebra and $(1,1)$ or $(-1,-1)$ in the group) will have to be treated separately.

    Show that the image of the exponential mapping is not dense in $\mathsf{SL}\,(2;\mathbb{R})$.

16. Using Exercise 4, show that the exponential mapping for $\mathsf{GL}(n;\mathbb{C})$ maps onto a dense subset of $\mathsf{GL}(n;\mathbb{C})$.

17. *The exponential mapping for the Heisenberg group.* Show that the exponential mapping from the Lie algebra of the Heisenberg group to the Heisenberg group is one-to-one and onto.

18. *The exponential mapping for* $\mathsf{U}(n)$. Show that the exponential mapping from $\mathsf{u}(n)$ to $\mathsf{U}(n)$ is onto, but not one-to-one. (Note that this shows that $\mathsf{U}(n)$ is connected.)

    Hint: Every unitary matrix has an orthonormal basis of eigenvectors.

19. Let G be a matrix Lie group, and $\mathfrak{g}$ its Lie algebra. Let $A(t)$ be a smooth curve lying in G, with $A(0) = I$. Let $X = \left.\frac{d}{dt}\right|_{t=0} A(t)$. Show that $X \in \mathfrak{g}$.

    Hint: Use Proposition 36.

    Note: This shows that the Lie algebra $\mathfrak{g}$ coincides with what would be called the **tangent space at the identity** in the language of differentiable manifolds.

20. Consider the space $\mathsf{gl}(n;\mathbb{C})$ of all $n \times n$ complex matrices. As usual, for $X \in \mathsf{gl}(n;\mathbb{C})$, define $\mathrm{ad}X : \mathsf{gl}(n;\mathbb{C}) \to \mathsf{gl}(n;\mathbb{C})$ by $\mathrm{ad}X(Y) = [X, Y]$. Suppose that X is a diagonalizable matrix. Show, then, that $\mathrm{ad}X$ is diagonalizable as an operator on $\mathsf{gl}(n;\mathbb{C})$.

    Hint: Consider first the case where X is actually diagonal.

    Note: The problem of diagonalizing $\mathrm{ad}X$ is an important one that we will encounter again in Chapter 6, when we consider semisimple Lie algebras.

Chapter 4
THE BAKER-CAMPBELL-HAUSDORFF FORMULA

4.1 The Baker-Campbell-Hausdorff Formula for the Heisenberg Group

A crucial result of Chapter 5 will be the following: Let G and H be matrix Lie groups, with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, and suppose that G is connected and simply connected. Then if $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism, there exists a unique Lie group homomorphism $\phi : G \to H$ such that ϕ and $\widetilde{\phi}$ are related as in Theorem 46. This result is extremely important because it implies that if G is connected and simply connected, then there is a natural one-to-one correspondence between the representations of G and the representations of its Lie algebra $\mathfrak{g}$ (as explained in Chapter 5). In practice, it is much easier to determine the representations of the Lie algebra than to determine directly the representations of the corresponding group.

This result (relating Lie algebra homomorphisms and Lie group homomorphisms) is deep. The "modern" proof (e.g., Varadarajan, Theorem 2.7.5) makes use of the Frobenius theorem, which is both hard to understand and hard to prove (Varadarajan, Section 1.3). Our proof will instead use the Baker-Campbell-Hausdorff formula, which is more easily stated and more easily motivated than the Frobenius theorem, but still deep.

The idea is the following. The desired group homomorphism $\phi : G \to H$ must satisfy

$$\phi\left(e^X\right) = e^{\widetilde{\phi}(X)}. \tag{4.1}$$

We would like, then, to *define* ϕ by this relation. This approach has two serious difficulties. First, a given element of G may not be expressible as e^X, and even if it is, the X may not be unique. Second, it is very far from clear why the ϕ in (4.1) (even to the extent it is well-defined) should be a group homomorphism.

It is the second issue which the Baker-Campbell-Hausdorff formula addresses. (The first issue will be addressed in the next chapter; it is there that the simple connectedness of G comes into play.) Specifically, (one form of) the Baker-Campbell-Hausdorff formula says that if X and Y are sufficiently small, then

$$\log(e^X e^Y) = X + Y + \tfrac{1}{2}[X,Y] + \tfrac{1}{12}[X,[X,Y]] - \tfrac{1}{12}[Y,[X,Y]] + \cdots. \tag{4.2}$$

It is not supposed to be evident at the moment what "$\cdots$" refers to. The only important point is that all of the terms in (4.2) are given in terms of X and Y, brackets of X and Y, brackets of brackets involving X and Y, etc. Then because $\widetilde{\phi}$ is a Lie algebra homomorphism,

$$\begin{aligned}\widetilde{\phi}\left(\log\left(e^{X}e^{Y}\right)\right) &= \widetilde{\phi}(X)+\widetilde{\phi}(Y)+\tfrac{1}{2}[\widetilde{\phi}(X),\widetilde{\phi}(Y)] \\ &\quad +\tfrac{1}{12}[\widetilde{\phi}(X),[\widetilde{\phi}(X),\widetilde{\phi}(Y)]]-\tfrac{1}{12}[\widetilde{\phi}(Y),[\widetilde{\phi}(X),\widetilde{\phi}(Y)]]+\cdots \\ &= \log\left(e^{\widetilde{\phi}(X)}e^{\widetilde{\phi}(Y)}\right) \end{aligned} \tag{4.3}$$

The relation (4.3) is extremely significant. For of course

$$e^{X}e^{Y}=e^{\log(e^{X}e^{Y})}$$

and so by (4.1),

$$\phi\left(e^{X}e^{Y}\right)=e^{\widetilde{\phi}(\log(e^{X}e^{Y}))}.$$

Thus (4.3) tells us that

$$\phi\left(e^{X}e^{Y}\right)=e^{\log\left(e^{\widetilde{\phi}(X)}e^{\widetilde{\phi}(Y)}\right)}=e^{\widetilde{\phi}(X)}e^{\widetilde{\phi}(Y)}=\phi(e^{X})\phi(e^{Y}).$$

Thus, the Baker-Campbell-Hausdorff formula shows that on elements of the form e^{X}, with X small, ϕ is a group homomorphism. (See Corollary 68 below.)

The Baker-Campbell-Hausdorff formula shows that all the information about the group product, at least near the identity, is "encoded" in the Lie algebra. Thus if $\widetilde{\phi}$ is a Lie algebra homomorphism (which by definition preserves the Lie algebra structure), and if we define ϕ near the identity by (4.1), then we can expect ϕ to preserve the group structure, i.e., to be a group homomorphism.

In this section we will look at how all of this works out in the very special case of the Heisenberg group. In the next section we will consider the general situation.

Theorem 65 *Suppose X and Y are $n\times n$ complex matrices, and that X and Y commute with their commutator. That is, suppose that*

$$[X,[X,Y]]=[Y,[X,Y]]=0.$$

Then

$$e^{X}e^{Y}=e^{X+Y+\frac{1}{2}[X,Y]}.$$

Remark

This is the special case of (4.2) in which the series terminates after the $[X,Y]$ term.

Proof

Let X and Y be as in the statement of the theorem. We will prove that in fact

$$e^{tX}e^{tY} = \exp\left(tX + tY + \frac{t^2}{2}[X,Y]\right),$$

which reduces to the desired result in the case $t = 1$. Since by assumption $[X,Y]$ commutes with everything in sight, the above relation is equivalent to

$$e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]} = e^{t(X+Y)}. \tag{4.4}$$

Let us call the left side of (4.4) $A(t)$ and the right side $B(t)$. Our strategy will be to show that $A(t)$ and $B(t)$ satisfy the same differential equation, with the same initial conditions. We can see right away that

$$\frac{dB}{dt} = B(t)(X+Y).$$

On the other hand, differentiating $A(t)$ by means of the product rule gives

$$\frac{dA}{dt} = e^{tX}Xe^{tY}e^{-\frac{t^2}{2}[X,Y]} + e^{tX}e^{tY}Ye^{-\frac{t^2}{2}[X,Y]} + e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}(-t[X,Y]). \tag{4.5}$$

(You can verify that the last term on the right is correct by differentiating term-by-term.)

Now, since X and Y commute with $[X,Y]$, they also commute with $e^{-\frac{t^2}{2}[X,Y]}$. Thus the second term on the right in (4.5) can be rewritten as

$$e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}Y.$$

The first term on the right in (4.5) is more complicated, since X does not necessarily commute with e^{tY}. However,

$$\begin{aligned} Xe^{tY} &= e^{tY}e^{-tY}Xe^{tY} \\ &= e^{tY}\operatorname{Ad}\left(e^{-tY}\right)(X) \\ &= e^{tY}e^{-t\operatorname{ad}Y}(X). \end{aligned}$$

But since $[X,[X,Y]] = 0$,

$$e^{-t\operatorname{ad}Y}(X) = X - t[Y,X] = X + t[X,Y]$$

with all higher terms being zero. Using the fact that everything commutes with $e^{-\frac{t^2}{2}[X,Y]}$ gives

$$e^{tX}Xe^{tY}e^{-\frac{t^2}{2}[X,Y]} = e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}(X + t[X,Y])$$

Making these substitutions into (4.5) gives

$$\begin{aligned}\frac{dA}{dt} &= e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}\left(X + t\left[X,Y\right]\right) + e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}Y + e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}\left(-t\left[X,Y\right]\right)\\ &= e^{tX}e^{tY}e^{-\frac{t^2}{2}[X,Y]}\left(X+Y\right)\\ &= A\left(t\right)\left(X+Y\right).\end{aligned}$$

Thus $A(t)$ and $B(t)$ satisfy the same differential equation. Moreover, $A(0) = B(0) = I$. Thus by standard uniqueness results for ordinary differential equations, $A(t) = B(t)$ for all t. □

Theorem 66 *Let H denote the Heisenberg group, and $\mathfrak{h}$ its Lie algebra. Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$, and let $\widetilde{\phi} : \mathfrak{h} \to \mathfrak{g}$ be a Lie algebra homomorphism. Then there exists a unique Lie group homomorphism $\phi : H \to G$ such that*

$$\phi\left(e^X\right) = e^{\widetilde{\phi}(X)}$$

for all $X \in \mathfrak{h}$.

Proof

Recall that the Heisenberg group has the very special property that its exponential mapping is one-to-one and onto. Let "log" denote the inverse of this map. Define $\phi : H \to G$ by the formula

$$\phi\left(A\right) = e^{\widetilde{\phi}(\log A)}.$$

We will show that ϕ is a Lie group homomorphism.

If X and Y are in the Lie algebra of the Heisenberg group (3×3 strictly upper-triangular matrices), then $[X, Y]$ is of the form

$$\begin{pmatrix} 0 & 0 & a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

such a matrix commutes with both X and Y. That is, X and Y commute with their commutator. Since $\widetilde{\phi}$ is a Lie algebra homomorphism, $\widetilde{\phi}(X)$ and $\widetilde{\phi}(Y)$ will also commute with their commutator:

$$\begin{aligned}\left[\widetilde{\phi}\left(X\right), \left[\widetilde{\phi}\left(X\right), \widetilde{\phi}\left(Y\right)\right]\right] &= \widetilde{\phi}\left(\left[X,\left[X,Y\right]\right]\right) = 0\\ \left[\widetilde{\phi}\left(Y\right), \left[\widetilde{\phi}\left(X\right), \widetilde{\phi}\left(Y\right)\right]\right] &= \widetilde{\phi}\left(\left[Y,\left[X,Y\right]\right]\right) = 0.\end{aligned}$$

We want to show that ϕ is a homomorphism, i.e., that $\phi(AB) = \phi(A)\phi(B)$. Well, A can be written as e^X for a unique $X \in \mathfrak{h}$ and B can be written as e^Y for a unique $Y \in \mathfrak{h}$. Thus by Theorem 65

$$\phi(AB) = \phi\left(e^X e^Y\right) = \phi\left(e^{X+Y+\frac{1}{2}[X,Y]}\right).$$

Using the definition of ϕ and the fact that $\widetilde{\phi}$ is a Lie algebra homomorphism:

$$\phi(AB) = \exp\left(\widetilde{\phi}(X) + \widetilde{\phi}(Y) + \frac{1}{2}\left[\widetilde{\phi}(X), \widetilde{\phi}(Y)\right]\right).$$

Finally, using Theorem 65 again we have

$$\phi(AB) = e^{\widetilde{\phi}(X)} e^{\widetilde{\phi}(Y)} = \phi(A)\phi(B).$$

Thus ϕ is a group homomorphism. It is easy to check that ϕ is continuous (by checking that log, exp, and $\widetilde{\phi}$ are all continuous), and so ϕ is a Lie group homomorphism. Moreover, ϕ by definition has the right relationship to $\widetilde{\phi}$. Furthermore, since the exponential mapping is one-to-one and onto, there can be at most one ϕ with $\phi\left(e^X\right) = e^{\widetilde{\phi}(X)}$. So we have uniqueness. $\square$

4.2 The General Baker-Campbell-Hausdorff Formula

The importance of the Baker-Campbell-Hausdorff formula lies not in the details of the formula, but in the fact that there is one, and the fact that it gives $\log(e^X e^Y)$ in terms of brackets of X and Y, brackets of brackets, etc. This tells us something very important, namely that (at least for elements of the form e^X, X small) the group product for a matrix Lie group G is *completely expressible in terms of the Lie algebra*. (This is because $\log\left(e^X e^Y\right)$, and hence also $e^X e^Y$ itself, can be computed in Lie-algebraic terms by (4.2).)

We will actually state and prove an integral form of the Baker-Campbell-Hausdorff formula, rather than the series form (4.2). However, the integral form is sufficient to obtain the desired result (4.3). (See Corollary 68.) The series form of the Baker-Campbell-Hausdorff formula is stated precisely and proved in Varadarajan, Sec. 2.15.

Consider the function

$$g(z) = \frac{\log z}{1 - \frac{1}{z}}.$$

This function is defined and analytic in the disk $\{|z-1| < 1\}$, and thus for z in this set, $g(z)$ can be expressed as

$$g(z) = \sum_{m=0}^{\infty} a_m (z-1)^m.$$

This series has radius of convergence one.

Now suppose V is a finite-dimensional complex vector space. Choose an arbitrary basis for V, so that V can be identified with $\mathbb{C}^n$ and thus the norm of a linear operator on V can be defined. Then for any operator A on V with $\|A - I\| < 1$, we can define

$$g(A) = \sum_{m=0}^{\infty} a_m (A-1)^m.$$

We are now ready to state the integral form of the Baker-Campbell-Hausdorff formula.

Theorem 67 (Baker-Campbell-Hausdorff) *For all $n \times n$ complex matrices X and Y with $\|X\|$ and $\|Y\|$ sufficiently small,*

$$\log\left(e^X e^Y\right) = X + \int_0^1 g(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y})(Y)\,dt. \tag{4.6}$$

Corollary 68 *Let G be a matrix Lie group and $\mathfrak{g}$ its Lie algebra. Suppose that $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{gl}(n; \mathbf{C})$ is a Lie algebra homomorphism. Then for all sufficiently small X, Y in $\mathfrak{g}$, $\log\left(e^X e^Y\right)$ is in $\mathfrak{g}$, and*

$$\widetilde{\phi}\left[\log\left(e^X e^Y\right)\right] = \log\left(e^{\widetilde{\phi}(X)} e^{\widetilde{\phi}(Y)}\right). \tag{4.7}$$

Remarks

Note that $e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}$, and hence also $g(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y})$, is a linear operator on the space $\mathfrak{gl}(n; \mathbb{C})$ of all $n \times n$ complex matrices. In (4.6), this operator is being applied to the matrix Y. The fact that X and Y are assumed small guarantees that $e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}$ is close to the identity operator on $\mathfrak{gl}(n; \mathbb{C})$ for all $0 \leq t \leq 1$. This ensures that $g(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y})$ is well defined.

If X and Y commute, then we expect to have $\log\left(e^X e^Y\right) = \log(e^{X+Y}) = X+Y$. Exercise 3 asks you to verify that the Baker-Campbell-Hausdorff formula indeed gives $X + Y$ in that case.

Formula (4.6) is admittedly horrible-looking. However, we are interested not in the details of the formula, but in the fact that it expresses $\log\left(e^X e^Y\right)$ (and hence $e^X e^Y$) in terms of the Lie-algebraic quantities $\mathrm{ad}X$ and $\mathrm{ad}Y$.

The goal of the Baker-Campbell-Hausdorff theorem is to compute $\log\left(e^X e^Y\right)$. You may well ask, "Why don't we simply expand both exponentials and the logarithm in power series and multiply everything out?" Well, you can do this, and if you do it for the first several terms you will get the same answer as B-C-H. However, there is a serious problem with this approach, namely: How do you know that the terms in such an expansion are expressible in terms of commutators? Consider for example the quadratic term. It is clear that this will be a linear combination of X^2, Y^2, XY, and

YX. But to be expressible in terms of commutators it must actually be a constant times $(XY - YX)$. Of course, for the quadratic term you can just multiply it out and see, and indeed you get $\frac{1}{2}(XY - YX) = \frac{1}{2}[X, Y]$. But it is far from clear how to prove that a similar result occurs for all the higher terms. See Exercise 4.

Proof of the Corollary

We begin by proving that the corollary follows from the integral form of the Baker-Campbell-Hausdorff formula. The proof is conceptually similar to the reasoning in Equation (4.3). Note that if X and Y lie in some Lie algebra $\mathfrak{g}$ then $\mathrm{ad}X$ and $\mathrm{ad}Y$ will preserve $\mathfrak{g}$, and so also will $g(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y})(Y)$. Thus whenever formula (4.6) holds, $\log\left(e^X e^Y\right)$ will lie in $\mathfrak{g}$. It remains only to verify (4.7). The idea is that if $\widetilde{\phi}$ is Lie algebra homomorphism, then it will take a big horrible looking expression involving 'ad' and X and Y, and turn it into the same expression with X and Y replaced by $\widetilde{\phi}(X)$ and $\widetilde{\phi}(Y)$.

More precisely, since $\widetilde{\phi}$ is a Lie algebra homomorphism,

$$\widetilde{\phi}[Y, X] = [\widetilde{\phi}(Y), \widetilde{\phi}(X)]$$

or

$$\widetilde{\phi}\left(\mathrm{ad}Y(X)\right) = \mathrm{ad}\widetilde{\phi}(Y)\left(\widetilde{\phi}(X)\right).$$

More generally,

$$\widetilde{\phi}\left((\mathrm{ad}Y)^n(X)\right) = \left(\mathrm{ad}\widetilde{\phi}(Y)\right)^n\left(\widetilde{\phi}(X)\right).$$

This being the case,

$$\begin{aligned}
\widetilde{\phi}\left(e^{\mathrm{ad}Y}(X)\right) &= \sum_{m=0}^{\infty}\frac{t^m}{m!}\widetilde{\phi}\left((\mathrm{ad}Y)^n(X)\right)\\
&= \sum_{m=0}^{\infty}\frac{t^m}{m!}\left(\mathrm{ad}\widetilde{\phi}(Y)\right)^n\left(\widetilde{\phi}(X)\right)\\
&= e^{t\mathrm{ad}\widetilde{\phi}(Y)}\left(\widetilde{\phi}(X)\right).
\end{aligned}$$

Similarly,

$$\widetilde{\phi}\left(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y}(X)\right) = e^{\mathrm{ad}\widetilde{\phi}(X)}e^{t\mathrm{ad}\widetilde{\phi}(Y)}\left(\widetilde{\phi}(X)\right).$$

Assume now that X and Y are small enough that B-C-H applies to X and Y, and to $\widetilde{\phi}(X)$ and $\widetilde{\phi}(Y)$. Then, using the linearity of the integral and reasoning similar to the above, we have:

$$\begin{aligned}\widetilde{\phi}\left(\log\left(e^{X}e^{Y}\right)\right) &= \widetilde{\phi}(X) + \int_0^1 \sum_{m=0}^{\infty} a_m \widetilde{\phi}\left[\left(e^{\operatorname{ad}X}e^{t\operatorname{ad}Y} - I\right)^n (X)\right] dt \\ &= \widetilde{\phi}(X) + \int_0^1 \sum_{m=0}^{\infty} a_m \left(e^{\operatorname{ad}\widetilde{\phi}(X)}e^{t\operatorname{ad}\widetilde{\phi}(Y)} - I\right)^n (\widetilde{\phi}(X))\, dt \\ &= \log\left(e^{\widetilde{\phi}(X)}e^{\widetilde{\phi}(Y)}\right).\end{aligned}$$

This is what we wanted to show. □

Before coming to the proof Baker-Campbell-Hausdorff formula itself, we will obtain a result concerning derivatives of the exponential mapping. This result is valuable in its own right, and will play a central role in our proof of the Baker-Campbell-Hausdorff formula.

Observe that if X and Y commute, then

$$e^{X+tY} = e^X e^{tY}$$

and so

$$\left.\tfrac{d}{dt}\right|_{t=0} e^{X+tY} = e^X \left.\tfrac{d}{dt}\right|_{t=0} e^{tY} = e^X Y.$$

In general, X and Y do not commute, and

$$\left.\tfrac{d}{dt}\right|_{t=0} e^{X+tY} \neq e^X Y.$$

This, as it turns out, is an important point. In particular, note that in the language of multivariate calculus

$$\left.\tfrac{d}{dt}\right|_{t=0} e^{X+tY} = \left\{\begin{array}{l}\text{directional derivative of } exp \text{ at } X, \\ \text{in the direction of } Y\end{array}\right. . \qquad (4.8)$$

Thus computing the left side of (4.8) is the same as computing all of the directional derivatives of the (matrix-valued) function *exp*. We expect the directional derivative to be a linear function of Y, for each fixed X.

Now, the function

$$\frac{1-e^{-z}}{z} = \frac{1-(1-z+\frac{z^2}{2!}-\cdots)}{z}$$

is an entire analytic function of z, even at $z=0$, and is given by the power series

$$\frac{1-e^{-z}}{z} = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{z^{n-1}}{n!} = 1 - \frac{z}{2!} + \frac{z^2}{3!} - \cdots.$$

This series (which has infinite radius of convergence), make sense when z is repaced by a linear operator A on some finite-dimensional vector space.

Theorem 69 (Derivative of Exponential) *Let X and Y be $n\times n$ complex matrices. Then*

$$\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} = e^{X}\left\{\frac{I-e^{-\mathrm{ad}X}}{\mathrm{ad}X}(Y)\right\}$$
$$= e^{X}\left\{Y-\frac{[X,Y]}{2!}+\frac{[X,[X,Y]]}{3!}-\cdots\right\}. \tag{4.9}$$

More generally, if $X(t)$ is a smooth matrix-valued funtion, then

$$\left.\frac{d}{dt}\right|_{t=0} e^{X(t)} = e^{X(0)}\left\{\frac{I-e^{-\mathrm{ad}X(0)}}{\mathrm{ad}X(0)}\left(\left.\tfrac{dX}{dt}\right|_{t=0}\right)\right\}. \tag{4.10}$$

Remarks

Note that the directional derivative in (4.9) is indeed linear in Y for each fixed X. Note also that (4.9) is just a special case of (4.10), by taking $X(t)=X+tY$, and evaluating at $t=0$.

Furthermore, observe that if X and Y commute, then only the first term in the series (4.9) survives. In that case, we obtain $\left.\frac{d}{dt}\right|_{t=0} e^{X+tY}=e^{X}Y$ as expected.

Proof of the Theorem

It is possible to prove this Theorem by expanding everything in a power series and differentiating term-by-term; we will not take that approach. We will prove only form (4.9) of the derivative formula, but the form (4.10) follows by the chain rule.

Let us use the Lie product formula, and let us assume for the moment that it is legal to interchange limit and derivative. (We will consider this issue at the end.) Then we have

$$e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} = e^{-X}\lim_{n\to\infty}\left.\frac{d}{dt}\right|_{t=0}\left(e^{X/n}e^{tY/n}\right)^{n}.$$

We now apply the product rule (generalized to n factors) to obtain

$$e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} = e^{-X}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left[\left(e^{X/n}e^{tY/n}\right)^{n-k-1}\left(e^{X/n}e^{tY/n}Y/n\right)\left(e^{X/n}e^{tY/n}\right)^{k}\right]_{t=0}$$
$$= e^{-X}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left(e^{X/n}\right)^{n-k-1}\left(e^{X/n}Y/n\right)\left(e^{X/n}\right)^{k}$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left(e^{X/n}\right)^{-k}Y\left(e^{X/n}\right)^{k}.$$

But

$$\begin{aligned}\left(e^{X/n}\right)^{-k} Y \left(e^{X/n}\right)^{k} &= \left[\mathrm{Ad}\left(e^{-X/n}\right)\right]^{k}(Y)\\ &= \left(e^{-\mathrm{ad}X/n}\right)^{k}(Y)\end{aligned}$$

(where we have used the relationship between Ad and ad). So we have

$$e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\left(e^{-\mathrm{ad}X/n}\right)^{k}(Y). \tag{4.11}$$

Observe now that $\sum_{k=0}^{n-1}\left(e^{-\mathrm{ad}X/n}\right)^{k}$ is a geometric series. Let us now reason for a moment at the purely formal level. Using the usual formula for geometric series, we get

$$\begin{aligned}e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} &= \lim_{n\to\infty}\frac{1}{n}\frac{I-\left(e^{-\mathrm{ad}X/n}\right)^{n}}{I-e^{-\mathrm{ad}X/n}}(Y)\\ &= \lim_{n\to\infty}\frac{I-e^{-\mathrm{ad}X}}{n\left[I-\left(I-\frac{\mathrm{ad}X}{n}+\frac{(\mathrm{ad}X)^2}{n^2 2!}-\cdots\right)\right]}(Y)\\ &= \lim_{n\to\infty}\frac{I-e^{-\mathrm{ad}X}}{\mathrm{ad}X-\frac{(\mathrm{ad}X)^2}{n2!}+\cdots}(Y)\\ &= \frac{I-e^{-\mathrm{ad}X}}{\mathrm{ad}X}(Y).\end{aligned}$$

This is what we wanted to show!

Does this argument make sense at any rigorous level? In fact it does. As usual, let us consider first the diagonalizable case. That is, assume that $\mathrm{ad}X$ is diagonalizable as an operator on $\mathsf{gl}(n;\mathbb{C})$, and assume that Y is an eigenvector for $\mathrm{ad}X$. This means that $\mathrm{ad}X(Y) = [X,Y] = \lambda Y$, for some $\lambda \in \mathbb{C}$. Now, there are two cases, $\lambda = 0$ and $\lambda \neq 0$. The $\lambda = 0$ case corresponds to the case in which X and Y commute, and we have already observed that the Theorem holds trivially in that case.

The interesting case, then, is the case $\lambda \neq 0$. Note that $(\mathrm{ad}X)^n(Y) = \lambda^n Y$, and so

$$\left(e^{-\mathrm{ad}X/n}\right)^{k}(Y) = \left(e^{-\lambda/n}\right)^{k}(Y).$$

Thus the geometric series in (4.11) becomes an ordinary complex-valued series, with ratio $e^{-\lambda/n}$. Since $\lambda \neq 0$, this ratio will be different from one for all sufficiently large n. Thus we get

$$e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} = \left(\lim_{n\to\infty}\frac{1}{n}\frac{I-\left(e^{-\lambda/n}\right)^{n}}{I-e^{-\lambda/n}}\right)Y.$$

There is now no trouble in taking the limit as we did formally above to get

$$\begin{aligned} e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY} &= \frac{1-e^{-\lambda}}{\lambda}Y \\ &= \frac{I-e^{-\mathrm{ad}X}}{\mathrm{ad}X}(Y). \end{aligned}$$

We see then that the Theorem holds in the case that $\mathrm{ad}X$ is diagonalizable and Y is an eigenvector of $\mathrm{ad}X$. If $\mathrm{ad}X$ is diagonalizable but Y is not an eigenvector, then Y is a linear combination of eigenvectors and applying the above computation to each of those eigenvectors gives the desired result.

We need, then, to consider the case where $\mathrm{ad}X$ is not diagonalizable. But (Exercise 20), if X is a diagonalizable matrix, then $\mathrm{ad}X$ will be diagonalizable as an operator on $\mathsf{gl}(n;\mathbb{C})$. Since, as we have already observed, every matrix is the limit of diagonalizable matrices, we are essentially done. For it is easy to see by differentiating the power series term-by-term that $e^{-X}\left.\frac{d}{dt}\right|_{t=0} e^{X+tY}$ exists and varies continuously with X. Thus once we have the Theorem for all diagonalizable X we have it for all X by passing to the limit.

The only unresolved issue, then, is the interchange of limit and derivative which we performed at the very beginning of the argument. I do not want to spell this out in detail, but let us see what would be involved in justifying this. A standard theorem in elementary analysis says that if $f_n(t) \to f(t)$ pointwise, and in addition df_n/dt converges uniformly to some function $g(t)$, then $f(t)$ is differentiable and $df/dt = g(t)$. (E.g., Theorem 7.17 in W. Rudin's *Principles of Mathematical Analysis.*) The key requirement is that the *derivatives* converge uniformly. Uniform convergence of the f_n's themselves is definitely not sufficient.

In our case, $f_n(t) = e^{-X}\left(e^{X/n}e^{tY/n}\right)^n$. The Lie product formula says that this converges pointwise to $e^{-X}e^{X+tY}$. We need, then, to show that

$$\frac{d}{dt}e^{-X}\left(e^{X/n}e^{tY/n}\right)^n$$

converges uniformly to some $g(t)$, say on the interval $-1 \leq t \leq 1$. This computation is similar to what we did above, with relatively minor modifications to account for the fact that we do not take $t = 0$ and to make sure the convergence is uniform. This part of the proof is left as an exercise to the reader. $\square$

Proof of the Baker-Campbell-Hausdorff Formula

We now turn to the proof of the Baker-Campbell-Hausdorff formula itself. Our argument follows Miller, Sec. 5.1, with minor differences of convention. (Warning: Miller's "Ad" is what we call "ad.") Define

$$Z(t) = \log\left(e^X e^{tY}\right)$$

If X and Y are sufficiently small, then $Z(t)$ is defined for $0 \leq t \leq 1$. It is left as an exercise to verify that $Z(t)$ is smooth. Our goal is to compute $Z(1)$.

By definition

$$e^{Z(t)} = e^X e^{tY}$$

so that

$$e^{-Z(t)} \frac{d}{dt} e^{Z(t)} = \left(e^X e^{tY}\right)^{-1} e^X e^{tY} Y = Y.$$

On the other hand, by Theorem 69,

$$e^{-Z(t)} \frac{d}{dt} e^{Z(t)} = \left\{ \frac{I - e^{-\mathrm{ad}Z(t)}}{\mathrm{ad}Z(t)} \right\} \left(\frac{dZ}{dt} \right).$$

Hence

$$\left\{ \frac{I - e^{-\mathrm{ad}Z(t)}}{\mathrm{ad}Z(t)} \right\} \left(\frac{dZ}{dt} \right) = Y.$$

If X and Y are small enough, then $Z(t)$ will also be small, so that $\left(I - e^{-\mathrm{ad}Z(t)}\right)/\mathrm{ad}Z(t)$ will be close to the identity and thus invertible. So

$$\frac{dZ}{dt} = \left\{ \frac{I - e^{-\mathrm{ad}Z(t)}}{\mathrm{ad}Z(t)} \right\}^{-1} (Y). \tag{4.12}$$

Recall that $e^{Z(t)} = e^X e^{tY}$. Applying the homomorphism 'Ad' gives

$$\mathrm{Ad}\left(e^{Z(t)}\right) = \mathrm{Ad}\left(e^X\right) \mathrm{Ad}\left(e^{tY}\right).$$

By the relationship (3.17) between 'Ad' and 'ad,' this becomes

$$e^{\mathrm{ad}Z(t)} = e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}$$

or

$$\mathrm{ad}Z(t) = \log\left(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}\right).$$

Plugging this into (4.12) gives

$$\frac{dZ}{dt} = \left\{ \frac{I - \left(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}\right)^{-1}}{\log\left(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}\right)} \right\}^{-1} (Y). \tag{4.13}$$

But now observe that

$$g(z) = \left\{ \frac{1 - z^{-1}}{\log z} \right\}^{-1}$$

so, formally, (4.13) is the same as

$$\frac{dZ}{dt} = g\left(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y}\right)(Y). \tag{4.14}$$

Reasoning as in the proof of Theorem 69 shows easily that this formal argument is actually correct.

Now we are essentially done, for if we note that $Z(0) = X$ and integrate (4.14), we get

$$Z(1) = X + \int_0^1 g(e^{\mathrm{ad}X} e^{t\mathrm{ad}Y})(Y)\,dt$$

which is the Baker-Campbell-Hausdorff formula. □

4.3 The Series Form of the Baker-Campbell-Hausdorff Formula

Let us see how to get the first few terms of the series form of B-C-H from the integral form. Recall the function

$$\begin{aligned} g(z) &= \frac{z\log z}{z-1} \\ &= \frac{[1+(z-1)]\left[(z-1) - \frac{(z-1)^2}{2} + \frac{(z-1)^3}{3}\cdots\right]}{(z-1)} \\ &= [1+(z-1)]\left[1 - \frac{z-1}{2} + \frac{(z-1)^2}{3}\right]. \end{aligned}$$

Multiplying this out and combining terms gives

$$g(z) = 1 + \frac{1}{2}(z-1) - \frac{1}{6}(z-1)^2 + \cdots.$$

The closed-form expression for g is

$$g(z) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n(n+1)}(z-1)^n.$$

Meanwhile

$$\begin{aligned} e^{\mathrm{ad}X} e^{t\mathrm{ad}Y} - I &= \left(I + \mathrm{ad}X + \frac{(\mathrm{ad}X)^2}{2} + \cdots\right)\left(I + t\mathrm{ad}Y + \frac{t^2(\mathrm{ad}Y)^2}{2} + \cdots\right) - I \\ &= \mathrm{ad}X + t\mathrm{ad}Y + t\mathrm{ad}X\mathrm{ad}Y + \frac{(\mathrm{ad}X)^2}{2} + \frac{t^2(\mathrm{ad}Y)^2}{2} + \cdots. \end{aligned}$$

The crucial observation here is that $e^{\mathrm{ad}X}e^{t\mathrm{ad}Y} - I$ has no zero-order term, just first-order and higher in $\mathrm{ad}X/\mathrm{ad}Y$. Thus $\left(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y} - I\right)^n$ will contribute only terms of degree n or higher in $\mathrm{ad}X/\mathrm{ad}Y$.

We have, then, up to degree two in $\mathrm{ad}X/\mathrm{ad}Y$

$$\begin{aligned} g\left(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y}\right) &= I + \frac{1}{2}\left[\mathrm{ad}X + t\mathrm{ad}Y + t\mathrm{ad}X\mathrm{ad}Y + \frac{(\mathrm{ad}X)^2}{2} + \frac{t^2(\mathrm{ad}Y)^2}{2} + \cdots\right] \\ &\quad - \frac{1}{6}\left[\mathrm{ad}X + t\mathrm{ad}Y + \cdots\right]^2 \\ &= I + \frac{1}{2}\mathrm{ad}X + \frac{t}{2}\mathrm{ad}Y + \frac{t}{2}\mathrm{ad}X\mathrm{ad}Y + \frac{(\mathrm{ad}X)^2}{4} + \frac{t^2(\mathrm{ad}Y)^2}{4} \\ &\quad - \frac{1}{6}\left[(\mathrm{ad}X)^2 + t^2(\mathrm{ad}Y)^2 + t\mathrm{ad}X\mathrm{ad}Y + t\mathrm{ad}Y\mathrm{ad}X\right] \\ &\quad + \text{ higher-order terms.} \end{aligned}$$

We now to apply $g\left(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y}\right)$ to Y and integrate. So (neglecting higher-order terms) by B-C-H, and noting that any term with adY acting first is zero:

$$\begin{aligned} &\log\left(e^X e^Y\right) \\ &= X + \int_0^1 \left[Y + \frac{1}{2}[X,Y] + \frac{1}{4}[X,[X,Y]] - \frac{1}{6}[X,[X,Y]] - \frac{t}{6}[Y,[X,Y]]\right] dt \\ &= X + Y + \frac{1}{2}[X,Y] + \left(\frac{1}{4} - \frac{1}{6}\right)[X,[X,Y]] - \frac{1}{6}\int_0^1 t\,dt\,[Y,[X,Y]]. \end{aligned}$$

Thus if we do the algebra we end up with

$$\begin{aligned} \log\left(e^X e^Y\right) &= X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] \\ &\quad + \text{ higher order terms.} \end{aligned}$$

This is the expression in (4.2).

4.4 Subgroups and Subalgebras

Suppose that G is a matrix Lie group, H another matrix Lie group, and suppose that $H \subset G$. Then certainly the Lie algebra $\mathfrak{h}$ of H will be a subalgebra of the Lie algebra $\mathfrak{g}$ of G. Does this go the other way around? That is given a Lie group G with Lie algebra $\mathfrak{g}$, and a subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, is there a matrix Lie group H whose Lie algebra is $\mathfrak{h}$?

In the case of the Hiesenberg group, the answer is yes. This is easily seen using the fact that the exponential mapping is one-to-one and onto, together with the special form of the Baker-Campbell-Hausdorff formula. (See Exercise 6.)

Unfortunately, the answer in general is no. For example, let $G = \mathsf{GL}(2;\mathbb{C})$ and let

$$\mathfrak{h} = \left\{ \left(\begin{array}{cc} it & 0 \\ 0 & ita \end{array} \right) \middle| \, t \in \mathbb{R} \right\},$$

where a is irrational. If there is going to be a matrix Lie group H with Lie algebra $\mathfrak{h}$, then H would contain the set

$$H_0 = \left\{ \left(\begin{array}{cc} e^{it} & 0 \\ 0 & e^{ita} \end{array} \right) \middle| \, t \in \mathbb{R} \right\}.$$

To be a matrix Lie group, H would have to be closed in $\mathsf{GL}(2;\mathbb{C})$, and so it would contain the closure of H_0, which (see) is the set

$$H_1 = \left\{ \left(\begin{array}{cc} e^{it} & 0 \\ 0 & e^{is} \end{array} \right) \middle| \, s, t \in \mathbb{R} \right\}.$$

But then the Lie algebra of H would have to contain the Lie algebra of H_1, which is two-dimensional!

Fortunately, all is not lost. We can still get a subgroup H for each subalgebra $\mathfrak{h}$, if we weaken the condition that H be a matrix Lie group. In the above example, the subgroup we want is H_0, despite the fact that H_0 is not a matrix Lie group.

Definition 70 *If H is any subgroup of $\mathsf{GL}(n;\mathbb{C})$, define the Lie algebra $\mathfrak{h}$ of H to be the set of all matrices X such that*

$$e^{tX} \in H$$

for all real t.

Definition 71 *If G is a matrix Lie group with Lie algebra $\mathfrak{g}$, then H is a **connected Lie subgroup** of G if*

i) H is a subgroup of G

ii) H is connected

iii) the Lie algebra $\mathfrak{h}$ of H is a subspace of $\mathfrak{g}$

iv) Every element of H can be written in the form $e^{X_1}e^{X_2}\cdots e^{X_n}$, with $X_1, \cdots, X_n \in \mathfrak{h}$.

Theorem 72 *If G is a matrix Lie group with Lie algebra $\mathfrak{g}$, and H is a connected Lie subgroup of G, then the Lie algebra $\mathfrak{h}$ of H is a subalgebra of $\mathfrak{g}$.*

Proof

Since by definition $\mathfrak{h}$ is a subspace of $\mathfrak{g}$, it remains only to show that $\mathfrak{h}$ is closed under brackets. So assume $X, Y \in \mathfrak{h}$. Then e^{tX} and e^{sY} are in H, and so (since H is a subgroup) is the element

$$e^{tX}e^{sY}e^{-tX} = \exp\left[s\left(e^{tX}Ye^{-tX}\right)\right].$$

This shows that $e^{tX}Ye^{-tX}$ is in $\mathfrak{h}$ for all t. But $\mathfrak{h}$ is a subspace of $\mathfrak{g}$, which is necessarily a closed subset of $\mathfrak{g}$. Thus

$$[X, Y] = \left.\frac{d}{dt}\right|_{t=0} e^{tX}Ye^{-tX} = \lim_{h \to 0} \frac{\left(e^{hX}Ye^{-hX} - Y\right)}{h}$$

is in $\mathfrak{h}$. (This argument is precisely the one we used to show that the Lie algebra of a matrix Lie group is a closed under brackets, once we had established that it is a subspace.) □

We are now ready to state the main theorem of this section, which is our second major application of the Baker-Campbell-Hausdorff formula.

Theorem 73 *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Let $\mathfrak{h}$ be a Lie subalgebra of $\mathfrak{g}$. Then there exists a unique connected Lie subgroup H of G such that the Lie algebra of H is $\mathfrak{h}$.*

Remark

Given a matrix Lie group G and a subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, the associated connected Lie subgroup H *might* be a matrix Lie group. This will happen precisely if H is a closed subset of G. There are various conditions under which you can prove that H is closed. For example, if $G = \mathsf{GL}(n; \mathbb{C})$, and $\mathfrak{h}$ is semisimple, then H is automatically closed, and hence a matrix Lie group. (See Helgason, Chapter II, Exercises and Further Results, D.)

Proof

Not written at this time. □

4.5 Exercises

1. The **center** of a Lie algebra $\mathfrak{g}$ is defined to be the set of all $X \in \mathfrak{g}$ such that $[X, Y] = 0$ for all $Y \in \mathfrak{g}$. Now consider the Heisenberg group

$$H = \left\{ \left(\begin{array}{ccc} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{array} \right) | a, b, c \in \mathbb{R} \right\}$$

with Lie algebra

$$\mathfrak{h}=\left\{\left(\begin{array}{ccc} 0 & \alpha & \beta \\ 0 & 0 & \gamma \\ 0 & 0 & 0 \end{array}\right) |\alpha,\beta,\gamma\in\mathbb{R}\right\}.$$

Determine the center $Z(\mathfrak{h})$ of $\mathfrak{h}$. For any $X, Y \in \mathfrak{h}$, show that $[X,Y] \in Z(\mathfrak{h})$. This implies, in particular that both X and Y commute with their commutator $[X,Y]$.

Show by direct computation that for any $X, Y \in \mathfrak{h}$,

$$e^X e^Y = e^{X+Y+\frac{1}{2}[X,Y]}. \tag{4.15}$$

2. Let X be a $n \times n$ complex matrix. Show that

$$\frac{I-e^{-X}}{X}$$

is invertible if and only if X has no eigenvalue of the form $\lambda = 2\pi i n$, with n an non-zero integer.

Hint: When is $\left(1-e^{-z}\right)/z$ equal to zero?

Remark: This exercise, combined with the formula in Theorem 69, gives the following result (in the language of differentiable manifolds): The exponential mapping $\exp : \mathfrak{g} \to G$ is a local diffeomorphism near $X \in \mathfrak{g}$ if and only $\mathrm{ad}X$ has no eigenvalue of the form $\lambda = 2\pi i n$, with n a non-zero integer.

3. Verify that the right side of the Baker-Campbell-Hausdorff formula (4.6) reduces to $X+Y$ in the case that X and Y commute.

Hint: Compute first $e^{\mathrm{ad}X}e^{t\mathrm{ad}Y}(Y)$ and $\left(e^{\mathrm{ad}X}e^{t\mathrm{ad}Y}-I\right)(Y)$.

4. Compute $\log\left(e^X e^Y\right)$ through third order in X/Y by using the power series for the exponential and the logarithm. Show that you get the same answer as the Baker-Campbell-Hausdorff formula.

5. Using the techniques in Section 4.3, compute the series form of the Baker-Campbell-Hausdorff formula up through fourth-order brackets. (We have already computed up through third-order brackets.)

6. Let $\mathfrak{a}$ be a subalgebra of the Lie algebra of the Heisenberg group. Show that $\exp(\mathfrak{a})$ is a connected Lie subgroup of the Heisenberg group. Show that in fact $\exp(\mathfrak{a})$ is a matrix Lie group.

7. Show that every connected Lie subgroup of $\mathsf{SU}(2)$ is closed. Show that this is not the case for $\mathsf{SU}(3)$.

Chapter 5
BASIC REPRESENTATION THEORY

5.1 Representations

Definition 74 *Let G be a matrix Lie group. Then a* ***finite-dimensional complex representation*** *of G is a Lie group homomorphism*

$$\Pi : G \to \mathsf{GL}(n; \mathbb{C})$$

($n \geq 1$) or more generally a Lie group homomorphism

$$\Pi : G \to \mathsf{GL}(V)$$

where V is a finite-dimensional complex vector space (with $dim(V) \geq 1$). A ***finite-dimensional real representation*** *of G is a Lie group homomorphism Π of G into $\mathsf{GL}(n; \mathbb{R})$ or into $\mathsf{GL}(V)$, where V is a finite-dimensional real vector space.*

If $\mathfrak{g}$ is a real or complex Lie algebra, then a ***finite-dimensional complex representation*** *of $\mathfrak{g}$ is a Lie algebra homomorphism π of $\mathfrak{g}$ into $\mathsf{gl}(n; \mathbb{C})$ or into $gl(V)$, where V is a finite-dimensional complex vector space. If $\mathfrak{g}$ is a real Lie algebra, then a* ***finite-dimensional real representation*** *of $\mathfrak{g}$ is a Lie algebra homomorphism π of $\mathfrak{g}$ into $\mathsf{gl}(n; \mathbb{R})$ or into $\mathsf{gl}(V)$.*

If Π or π is a one-to-one homomorphism, then the representation is called ***faithful****.*

Remarks

You should think of a reprentation as a (linear) **action** of a group or Lie algebra on a vector space. (Since, say, to every $g \in G$ there is associated an operator $\Pi(g)$, which acts on the vector space V.) In fact, we will use terminology such as, "Let Π be a representation of G acting on the space V." Even if $\mathfrak{g}$ is a real Lie algebra, we will consider mainly complex representations of $\mathfrak{g}$. After making a few more definitions, we will discuss the question of why one should be interested in studying representations.

Definition 75 *Let Π be a finite-dimensional real or complex representation of a matrix Lie group G, acting on a space V. A subspace W of V is called* ***invariant*** *if $\Pi(A)w \in W$ for all $w \in W$ and all $A \in G$. An invariant subspace W is called*

non-trivial *if* $W \neq \{0\}$ *and* $W \neq V$. *A representation with no non-trivial invariant subspaces is called* ***irreducible****.*

The terms ***invariant****,* ***non-trivial****, and* ***irreducible*** *are defined analogously for representations of Lie algebras.*

Definition 76 *Let* G *be a matrix Lie group, let* Π *be a representation of* G *acting on the space* V, *and let* Σ *be a representation of* G *acting on the space* W. *A linear map* $\phi : V \to W$ *is called a* ***morphism*** *(or* ***intertwining map****) of representations if*

$$\phi(\Pi(A)v) = \Sigma(A)\phi(v)$$

for all $A \in G$ *and all* $v \in V$. *The analogous property defines morphisms of representations of a Lie algebra.*

If ϕ *is a morphism of representations, and in addition* ϕ *is invertible, then* ϕ *is said to be an* ***isomorphism*** *of representations. If there exists an isomorphism between* V *and* W, *then the representations are said to be* ***isomorphic*** *(or* ***equivalent****).*

Remarks

Two isomorphic representations should be regarded as being "the same" representation. A typical problem in representation theory is to determine, up to isomorphism, all the irreducible representations of a particular group or Lie algebra. In Section 5.4 we will determine all the finite-dimensional complex irreducible representations of the Lie algebra $\mathsf{su}(2)$.

Proposition 77 *Let* G *be a matrix Lie group with Lie algebra* $\mathfrak{g}$, *and let* Π *be a (finite-dimensional real or complex) representation of* G, *acting on the space* V. *Then there is a unique representation* π *of* $\mathfrak{g}$ *acting on the same space such that*

$$\Pi(e^X) = e^{\pi(X)}$$

for all $X \in \mathfrak{g}$. *The representation* π *can be computed as*

$$\pi(X) = \left.\frac{d}{dt}\right|_{t=0} \Pi\left(e^{tX}\right)$$

and satisfies

$$\pi\left(AXA^{-1}\right) = \Pi(A)\pi(X)\Pi(A)^{-1}$$

for all $X \in \mathfrak{g}$ *and all* $A \in G$.

Proof

Theorem 46 in Chapter 3 states that for each Lie group homomorphism $\phi : G \to H$ there is an associated Lie algebra homomorphism $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$. Take $H = \mathsf{GL}(V)$ and $\phi = \Pi$. Since the Lie algebra of $\mathsf{GL}(V)$ is $\mathsf{gl}(V)$ (since the exponential of any operator is invertible), the associated Lie algebra homomorphism $\widetilde{\phi} = \pi$ maps from $\mathfrak{g}$ to $\mathsf{gl}(V)$, and so constitutes a representation of $\mathfrak{g}$.

The properties of π follow from the properties of $\widetilde{\phi}$ given in Theorem 6. □

Proposition 78 *Let $\mathfrak{g}$ be a real Lie algebra, and $\mathfrak{g}_{\mathbb{C}}$ its complexification. Then every finite-dimensional complex representation π of $\mathfrak{g}$ has a unique extension to a (complex-linear) representation of $\mathfrak{g}_{\mathbb{C}}$, also denoted π. The representation of $\mathfrak{g}_{\mathbb{C}}$ satisfies*

$$\pi(X + iY) = \pi(X) + i\pi(Y)$$

for all $X \in \mathfrak{g}$.

Proof

This follows from Exercise 14 of Chapter 3.

Definition 79 *Let G be a matrix Lie group, let $\mathcal{H}$ be a Hilbert space, and let $U(\mathcal{H})$ denote the group of unitary operators on $\mathcal{H}$. Then a homomorphism $\Pi : G \to U(\mathcal{H})$ is called a **unitary representation** of G if Π satisfies the following continuity condition: If $A_n, A \in G$ and $A_n \to A$, then*

$$\Pi(A_n)v \to \Pi(A)v$$

*for all $v \in \mathcal{H}$. A unitary representation with no non-trivial closed invariant subspaces is called **irreducible**.*

Remarks

This continuity condition is called **strong continuity**. One could require the even stronger condition that $\|\Pi(A_n) - \Pi(A)\| \to 0$, but this turns out to be too stringent a requirement. In practice, any homomorphism of G into $U(\mathcal{H})$ you can write down explicitly will be strongly continuous.

One could try to define some analog of unitary representations for Lie algebras, but there are serious technical difficulties associated with getting the "right" definition.

5.2 Why Study Representations?

If a representation Π is a faithful representation of a matrix Lie group G, then $\{\Pi(A)\,|A \in G\}$ is a group of matrices which is isomorphic to the original group G. Thus Π allows us to *represent* G as a group of matrices. This is the motivation for the term representation. (Of course, we still call Π a representation even if it is not faithful.)

Despite the origin of the term, the point of representation theory is *not* (at least in this course) to represent a group as a group of matrices. After all, all of our groups are already matrix groups! While it might seem redundant to study representations of a group which is already represented as a group of matrices, this is precisely what we are going to do.

The reason for this is that a representation can be thought of (as we have already noted) as an action of our group on some vector space. Such actions (representations) arise naturally in many branches of both mathematics and physics, and it is important to understand them.

A typical example would be a differential equation in three-dimensional space which has rotational symmetry. If the equation has rotational symmetry, then the space of solutions will be invariant under rotations. Thus the space of solutions will constitute a representation of the rotation group $\mathsf{SO}(3)$. If you know what all of the representations of $\mathsf{SO}(3)$ are, this can help immensely in narrowing down what the space of solutions can be. (As we will see, $\mathsf{SO}(3)$ has lots of other representations besides the obvious one in which $\mathsf{SO}(3)$ acts on $\mathbb{R}^3$.)

In fact, one of the chief applications of representation theory is to exploit symmetry. If a system has symmetry, then the set of symmetries will form a group, and understanding the representations of the symmetry group allows you to use that symmetry to simplify the problem.

In addition, studying the representations of a group G (or of a Lie algebra $\mathfrak{g}$) can give information about the group (or Lie algebra) itself. For example, if G is a *finite* group, then associated to G is something called the **group algebra**. The structure of this group algebra can be described very nicely in terms of the irreducible representations of G.

In this course, we will be interested primarily in computing the finite-dimensional irreducible complex representations of matrix Lie groups. As we shall see, this problem can be reduced almost completely to the problem of computing the finite-dimensional irreducible complex representations of the associated Lie algebra. In this chapter, we will discuss the theory at an elementary level, and will consider in detail the example of $\mathsf{SO}(3)$ and $\mathsf{SU}(2)$. In Chapter 6, we will study the representations of $\mathsf{SU}(3)$, which is substantially more involved than that of $\mathsf{SU}(2)$, and give an overview of the representation theory of a very important class of Lie groups, namely, the semisimple ones.

5.3 Examples of Representations

The Standard Representation

A matrix Lie group G is by definition a subset of some $\mathsf{GL}(n;\mathbb{R})$ or $\mathsf{GL}(n;\mathbb{C})$. The inclusion map of G into $\mathsf{GL}(n)$ (i.e., $\Pi(A) = A$) is a representation of G, called the **standard representation** of G. Thus for example the standard representation of $\mathsf{SO}(3)$ is the one in which $\mathsf{SO}(3)$ acts in the usual way on $\mathbb{R}^3$. If G is a subgroup of $\mathsf{GL}(n;\mathbb{R})$ or $\mathsf{GL}(n;\mathbb{C})$, then its Lie algebra $\mathfrak{g}$ will be a subalgebra of $\mathsf{gl}(n;\mathbb{R})$ or $\mathsf{gl}(n;\mathbb{C})$. The inclusion of $\mathfrak{g}$ into $\mathsf{gl}(n;\mathbb{R})$ or $\mathsf{gl}(n;\mathbb{C})$ is a representation of $\mathfrak{g}$, called the **standard representation**.

The Trivial Representation

Consider the one-dimensional complex vector space $\mathbb{C}$. Given any matrix Lie group G, we can define the **trivial representation** of G, $\Pi : G \to \mathsf{GL}(1;\mathbb{C})$, by the formula

$$\Pi(A) = I$$

for all $A \in G$. Of course, this is an irreducible representation, since $\mathbb{C}$ has no non-trivial subspaces, let alone non-trivial invariant subspaces. If $\mathfrak{g}$ is a Lie algebra, we can also define the **trivial representation** of $\mathfrak{g}$, $\pi : \mathfrak{g} \to \mathsf{gl}(1;\mathbb{C})$, by

$$\pi(X) = 0$$

for all $X \in \mathfrak{g}$. This is an irreducible representation.

The Adjoint Representation

Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. We have already defined the adjoint mapping

$$\mathrm{Ad} : G \to \mathsf{GL}(\mathfrak{g})$$

by the formula

$$\mathsf{Ad}A(X) = AXA^{-1}.$$

Recall that Ad is a Lie group homomorphism. Since Ad is a Lie group homomorphism into a group of invertible operators, we see that in fact Ad is a representation of G, acting on the space $\mathfrak{g}$. Thus we can now give Ad its proper name, the **adjoint representation** of G. The adjoint representation is a real representation of G.

Similarly, if $\mathfrak{g}$ is a Lie algebra, we have

$$\mathrm{ad} : \mathfrak{g} \to \mathsf{gl}(\mathfrak{g})$$

defined by the formula

$$\mathsf{ad}X(Y) = [X, Y].$$

We know that ad is a Lie algebra homomorphism (Chapter 3, Proposition 61), and is therefore a representation of $\mathfrak{g}$, called the **adjoint representation**. In the case that $\mathfrak{g}$ is the Lie algebra of some matrix Lie group G, we have already established (Chapter 3, Proposition 49 and Exercise 13) that Ad and ad are related as in Proposition 77.

Note that in the case of $\mathsf{SO}(3)$ the standard representation and the adjoint representation are both three dimensional real representations. In fact these two representations are equivalent (Exercise 4).

Some Representations of $\mathsf{SU}(2)$

Consider the space V_m of homogeneous polynomials in two complex variables with total degree m ($m \geq 0$). That is, V_m is the space of functions of the form

$$f(z_1, z_2) = a_0 z_1^m + a_1 z_1^{m-1} z_2 + a_2 z_1^{m-2} z_2^2 \cdots + a_m z_2^m \tag{5.1}$$

with $z_1, z_2 \in \mathbb{C}$ and the a_i's arbitrary complex constants. The space V_m is an $(m+1)$-dimensional complex vector space.

Now by definition an element U of $\mathsf{SU}(2)$ is a linear transformation of $\mathbb{C}^2$. Let z denote the pair $z = (z_1, z_2)$ in $\mathbb{C}^2$. Then we may define a linear transformation $\Pi_m(U)$ on the space V_m by the formula

$$[\Pi_m(U)f](z) = f(U^{-1}z). \tag{5.2}$$

Explicitly, if f is as in (5.1), then

$$[\Pi_m(U)f](z_1, z_2) = \sum_{k=0}^{m} a_k \left(U_{11}^{-1} z_1 + U_{12}^{-1} z_2\right)^{m-k} \left(U_{21}^{-1} z_1 + U_{22}^{-1} z_2\right)^k.$$

By expanding out the right side of this formula we see that $\Pi_m(U)f$ is again a homogeneous polynomial of degree m. Thus $\Pi_m(U)$ actually maps V_m into V_m.

Now, compute

$$\begin{aligned}\Pi_m(U_1)[\Pi_m(U_2)f](z) &= [\Pi_m(U_2)f](U_1^{-1}z) = f\left(U_2^{-1}U_1^{-1}z\right) \\ &= \Pi_m(U_1U_2)f(z).\end{aligned}$$

Thus Π_m is a (finite-dimensional complex) representation of $\mathsf{SU}(2)$. (It is very easy to do the above computation incorrectly.) The inverse in definition (5.2) is necessary in order to make Π_m a representation. It turns out that each of the representations Π_m of $\mathsf{SU}(2)$ is irreducible, and that every finite-dimensional irreducible representation of $\mathsf{SU}(2)$ is equivalent to one (and only one) of the Π_m's. (Of course, no two of the Π_m's are equivalent, since they don't even have the same dimension.)

Let us now compute the corresponding Lie algebra representation π_m. According to Proposition 77, π_m can be computed as

$$\pi_m(X) = \left.\frac{d}{dt}\right|_{t=0} \Pi_m\left(e^{tX}\right).$$

So

$$\left(\pi_m(X)f\right)(z) = \left.\frac{d}{dt}\right|_{t=0} f\left(e^{-tX}z\right).$$

Now let $z(t)$ be the curve in $\mathbb{C}^2$ defined as $z(t) = e^{-tX}z$, so that $z(0) = z$. Of course, $z(t)$ can be written as $z(t) = (z_1(t), z_2(t))$, with $z_i(t) \in \mathbb{C}$. By the chain rule,

$$\pi_m(X)f = \frac{\partial f}{\partial z_1}\left.\frac{dz_1}{dt}\right|_{t=0} + \frac{\partial f}{\partial z_2}\left.\frac{dz_2}{dt}\right|_{t=0}.$$

But $dz/dt|_{t=0} = -Xz$, so we obtain the following formula for $\pi_m(X)$

$$\pi_m(X)f = -\frac{\partial f}{\partial z_1}\left(X_{11}z_1 + X_{12}z_2\right) - \frac{\partial f}{\partial z_2}\left(X_{21}z_1 + X_{22}z_2\right). \tag{5.3}$$

Now, according to Proposition 78, every finite-dimensional complex representation of the Lie algebra $\mathsf{su}(2)$ extends uniquely to a complex-linear representation of the complexification of $\mathsf{su}(2)$. But the complexification of $\mathsf{su}(2)$ is (isomorphic to) $\mathsf{sl}(2;\mathbb{C})$ (Chapter 3, Proposition 64). To see that this is so, note that $\mathsf{sl}(2;\mathbb{C})$ is the space of all 2×2 complex matrices with trace zero. But if X is in $\mathsf{sl}(2;\mathbb{C})$, then

$$X = \frac{X - X^*}{2} + \frac{X + X^*}{2} = \frac{X - X^*}{2} + i\frac{X + X^*}{2i}$$

where both $(X - X^*)/2$ and $(X + X^*)/2i$ are in $\mathsf{su}(2)$. (Check!) It is easy to see that this decomposition is unique, so that every $X \in \mathsf{sl}(2;\mathbb{C})$ can be written uniquely as $X = X_1 + iY_1$ with $X_1, Y_1 \in \mathsf{su}(2)$. Thus $\mathsf{sl}(2;\mathbb{C})$ is isomomorphic as a vector space to $\mathsf{su}(2)_{\mathbf{C}}$. But this is in fact an isomorphism of Lie algebras, since in both cases

$$[X_1 + iY_1, X_2 + iY_2] = [X_1, X_2] - [Y_1, Y_2] + i\left([X_1, Y_2] + [X_2, Y_1]\right).$$

(See Exercise 5.)

So, the representation π_m of $\mathsf{su}(2)$ given by (5.3) extends to a representation of $\mathsf{sl}(2;\mathbb{C})$, which we will also call π_m. I assert that in fact formula (5.3), still holds for $X \in \mathsf{sl}(2;\mathbb{C})$. Why is this? Well, (5.3) is undoubtedly (complex) linear, and it agrees with the original π_m for $X \in \mathsf{su}(2)$. But there is only one complex linear extension of π_m from $\mathsf{su}(2)$ to $\mathsf{sl}(2;\mathbb{C})$, so this must be it!

So, for example, consider the element

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

in the Lie algebra $\mathsf{sl}(2;\mathbb{C})$. Applying formula (5.3) gives

$$(\pi_m(H)f)(z) = -\frac{\partial f}{\partial z_1}z_1 + \frac{\partial f}{\partial z_2}z_2.$$

Thus we see that

$$\pi_m(H) = -z_1\frac{\partial}{\partial z_1} + z_2\frac{\partial}{\partial z_2}. \tag{5.4}$$

Applying $\pi_m(H)$ to a basis element $z_1^k z_2^{m-k}$ we get

$$\pi_m(H)z_1^k z_2^{m-k} = -kz_1^k z_2^{m-k} + (m-k)z_1^k z_2^{m-k} = (m-2k)z_1^k z_2^{m-k}.$$

Thus $z_1^k z_2^{m-k}$ is an eigenvector for $\pi_m(H)$ with eigenvalue $(m-2k)$. In particular, $\pi_m(H)$ is diagonalizable.

Let X and Y be the elements

$$X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \quad Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

in $\mathsf{sl}(2;\mathbb{C})$. Then (5.3) tells us that

$$\pi_m(X) = -z_2\tfrac{\partial}{\partial z_1}; \quad \pi_m(Y) = -z_1\tfrac{\partial}{\partial z_2}$$

so that

$$\begin{aligned} \pi_m(X)z_1^k z_2^{m-k} &= -kz_1^{k-1}z_2^{m-k+1} \\ \pi_m(Y)z_1^k z_2^{m-k} &= (k-m)z_1^{k+1}z_2^{m-k-1}. \end{aligned} \tag{5.5}$$

Proposition 80 *The representation π_m is an irreducible representation of* $\mathsf{sl}(2;\mathbb{C})$.

Proof

It suffices to show that every non-zero invariant subspace of V_m is in fact equal to V_m. So let W be such a space. Since W is assumed non-zero, there is at least one non-zero element w in W. Then w can be written uniquely in the form

$$w = a_0 z_1^m + a_1 z_1^{m-1}z_2 + a_2 z_1^{m-2}z_2^2 \cdots + a_m z_2^m$$

with at least one of the a_k's non-zero. Let k_0 be the largest value of k for which $a_k \neq 0$, and consider

$$\pi_m(X)^{k_0}w.$$

Since (by (5.5)) each application of $\pi_m(X)$ lowers the power of z_1 by 1, $\pi_m(X)^{k_0}$ will kill all the terms in w whose power of z_1 is less than k_0, that is, all except the $a_{k_0} z_1^{k_0} z_2^{m-k_0}$ term. On the other hand, we compute easily that

$$\pi_m(X)^{k_0}\left(a_{k_0} z_1^{k_0} z_2^{m-k_0}\right) = k_0!(-1)^{k_0} a_{k_0} z_2^m.$$

We see, then, that $\pi_m(X)^{k_0} w$ is a *non-zero* multiple of z_2^m. Since W is assumed invariant, W must contain this multiple of z_2^m, and so also z_2^m itself.

But now it follows from (5.5) that $\pi_m(Y)^k z_2^m$ is a *non-zero* multiple of $z_1^k z_2^{m-k}$. Therefore W must also contain $z_1^k z_2^{m-k}$ for all $0 \leq k \leq m$. Since these elements form a basis for V_m, we see that in fact $W = V_m$, as desired. □

Two Unitary Representations of $\mathsf{SO}(3)$

Let $\mathcal{H} = L^2(\mathbb{R}^3, dx)$. For each $R \in \mathsf{SO}(3)$, define an operator $\Pi_1(R)$ on $\mathcal{H}$ by the formula

$$[\Pi_1(R)f](x) = f\left(R^{-1}x\right).$$

Since Lebesgue measure dx is rotationally invariant, $\Pi_1(R)$ is a unitary operator for each $R \in \mathsf{SO}(3)$. The calculation of the previous subsection shows that the map $R \to \Pi_1(R)$ is a homomorphism of $\mathsf{SO}(3)$ into $U(\mathcal{H})$. This map is strongly continuous, and hence constitutes a unitary representation of $\mathsf{SO}(3)$.

Similarly, we may consider the unit sphere $S^2 \subset \mathbb{R}^3$, with the usual surface measure Ω. Of course, any $R \in \mathsf{SO}(3)$ maps S^2 into S^2. For each R we can define $\Pi_2(R)$ acting on $L^2(S^2, d\Omega)$ by

$$[\Pi_2(R)f](x) = f\left(R^{-1}x\right).$$

Then Π_2 is a unitary representation of $\mathsf{SO}(3)$.

Neither of the unitary representations Π_1 and Π_2 is irreducible. In the case of Π_2, $L^2(S^2, d\Omega)$ has a very nice decomposition as the orthogonal direct sum of finite-dimensional invariant subspaces. This decomposition is the theory of "spherical harmonics," which are well known in the physics (and mathematics) literature.

A Unitary Representation of the Reals

Let $\mathcal{H} = L^2(\mathbb{R}, dx)$. For each $a \in \mathbb{R}$, define $T_a : \mathcal{H} \to \mathcal{H}$ by

$$(T_a f)(x) = f(x - a).$$

Clearly T_a is a unitary operator for each $a \in \mathbb{R}$, and clearly $T_a T_b = T_{a+b}$. The map $a \to T_a$ is strongly continuous, so T is a unitary representation of $\mathbb{R}$. This representation is not irreducible. The theory of the Fourier transform allows you to determine all the closed, invariant subspaces of $\mathcal{H}$ (W. Rudin, *Real and Complex Analysis*, Theorem 9.17).

The Unitary Representations of the Real Heisenberg Group

Consider the Heisenberg group

$$H = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} |a, b, c \in \mathbb{R} \right\}.$$

Now consider a real, non-zero constant, which for reasons of historical convention we will call $\hbar$("aitch-bar"). Now for each $\hbar \in \mathbb{R}\backslash\{0\}$, define a unitary operator $\Pi_\hbar$ on $L^2(\mathbb{R}, dx)$ by

$$\Pi_\hbar \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} f = e^{-i\hbar b} e^{i\hbar cx} f(x-a). \tag{5.6}$$

It is clear that the right side of (5.6) has the same norm as f, so $\Pi_\hbar$ is indeed unitary.

Now compute

$$\begin{aligned} &\Pi_\hbar \begin{pmatrix} 1 & \widetilde{a} & \widetilde{b} \\ 0 & 1 & \widetilde{c} \\ 0 & 0 & 1 \end{pmatrix} \Pi_\hbar \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} f \\ &= e^{-i\hbar\widetilde{b}} e^{i\hbar\widetilde{c}x} e^{-i\hbar b} e^{i\hbar c(x-\widetilde{a})} f(x - \widetilde{a} - a) \\ &= e^{-i\hbar(\widetilde{b}+b+c\widetilde{a})} e^{i\hbar(\widetilde{c}+c)x} f\left(x - (\widetilde{a}+a)\right). \end{aligned}$$

This shows that the map $A \to \Pi_\hbar(A)$ is a homomorphism of the Heisenberg group into $U\left(L^2(\mathbb{R})\right)$. This map is strongly continuous, and so $\Pi_\hbar$ is a unitary representation of H.

Note that a typical unitary operator $\Pi_\hbar(A)$ consists of first translating f, then multiplying f by the function $e^{i\hbar cx}$, and then multiplying f by the constant $e^{-i\hbar b}$. Multiplying f by the function $e^{i\hbar cx}$ has the effect of translating the Fourier transform of f, or in physical language, "translating f in momentum space." Now, if U_1 is an ordinary translation and U_2 is a translation of the Fourier transform (i.e., $U_2 =$ multiplication by some $e^{i\hbar cx}$), then U_1 and U_2 will not commute, but $U_1 U_2 U_1^{-1} U_2^{-1}$ will be simply multiplication by a constant of absolute value one. Thus $\{\Pi_\hbar(A)\,|A \in H\}$ is the group of operators on $L^2(\mathbb{R})$ generated by ordinary translations and translations in Fourier space. It is this representation of the Heisenberg group which motivates its name. (See also Exercise 10.)

It follows fairly easily from standard Fourier transform theory (e.g., W. Rudin, *Real and Complex Analysis*, Theorem 9.17) that for each $\hbar \in \mathbb{R}\backslash\{0\}$ the representation $\Pi_\hbar$ is irreducible. Furthermore, these are (up to equivalence) almost all of the irreducible unitary representations of H. The only remaining ones are the one-dimensional representations $\Pi_{\alpha,\beta}$

$$\Pi_{\alpha,\beta}\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} = e^{i(\alpha a+\beta c)}I$$

with $\alpha, \beta \in \mathbb{R}$. (The $\Pi_{\alpha,\beta}$'s are the irreducible unitary representations in which the center of H acts trivially.) The fact that $\Pi_{\hbar}$'s and the $\Pi_{\alpha,\beta}$'s are all of the (strongly continuous) irreducible unitary representations of H is closely related to the celebrated Stone-Von Neumann theorem in mathematical physics. See, for example, M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. 3, Theorem XI.84. See also Exercise 11.

5.4 The Irreducible Representations of $\mathsf{su}(2)$

In this section we will compute (up to equivalence) all the finite-dimensional irreducible complex representations of the Lie algebra $\mathsf{su}(2)$. This computation is important for several reasons. In the first place, $\mathsf{su}(2) \cong \mathsf{so}(3)$, and the representations of $\mathsf{so}(3)$ are of physical significance. (The computation we will do here is found in every standard textbook on quantum mechanics, under the heading "angular momentum.") In the second place, the representation theory of $\mathsf{su}(2)$ is an illuminating example of how one uses commutation relations to determine the representations of a Lie algebra. In the third place, in determining the representations of general semisimple Lie algebras (Chapter 6), we will explicitly use the representation theory of $\mathsf{su}(2)$.

Now, every finite-dimensional complex representation π of $\mathsf{su}(2)$ extends by Prop. 78 to a complex-linear representation (also called π) of the complexification of $\mathsf{su}(2)$, namely $\mathsf{sl}(2;\mathbb{C})$.

Proposition 81 *Let π be a complex representation of $\mathsf{su}(2)$, extended to a complex-linear representation of $\mathsf{sl}(2;\mathbb{C})$. Then π is irreducible as a representation of $\mathsf{su}(2)$ if and only if it is irreducible as a representation of $\mathsf{sl}(2;\mathbb{C})$.*

Proof

Let us make sure we are clear about what this means. Suppose that π is a complex representation of the (real) Lie algebra $\mathsf{su}(2)$, acting on the complex space V. Then saying that π is irreducible means that there is no non-trivial invariant *complex* subspace $W \subset V$. That is, even though $\mathsf{su}(2)$ is a real Lie algebra, when considering complex representations we are interested only in complex invariant subspaces.

Now, suppose that π is irreducible as a representation of $\mathsf{su}(2)$. If W is a (complex) subspace of V which is invariant under $\mathsf{sl}(2;\mathbb{C})$, then certainly W is invariant under $\mathsf{su}(2) \subset \mathsf{sl}(2;\mathbb{C})$. Therefore $W = \{0\}$ or $W = V$. Thus π is irreducible as a representation of $\mathsf{sl}(2;\mathbb{C})$.

On the other hand, suppose that π is irreducible as a representation of $\mathsf{sl}(2;\mathbb{C})$, and suppose that W is a (complex) subspace of V which is invariant under $\mathsf{su}(2)$. Then W will also be invariant under $\pi(X+iY)=\pi(X)+i\pi(Y)$, for all $X,Y\in\mathsf{su}(2)$. Since every element of $\mathsf{sl}(2;\mathbb{C})$ can be written as $X+iY$, we conclude that in fact W is invariant under $\mathsf{sl}(2;\mathbb{C})$. Thus $W=\{0\}$ or $W=V$, so π is irreducible as a representation of $\mathsf{su}(2)$. $\square$

We see, then that studying the irreducible representations of $\mathsf{su}(2)$ is equivalent to studying the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$. Passing to the complexified Lie algebra makes our computations easier.

We will use the following basis for $\mathsf{sl}(2;\mathbb{C})$:

$$H=\begin{pmatrix}1&0\\0&-1\end{pmatrix};\quad X=\begin{pmatrix}0&1\\0&0\end{pmatrix};\quad Y=\begin{pmatrix}0&0\\1&0\end{pmatrix}$$

which have the commutation relations

$$\begin{aligned}[][H,X]&=2X\\ [H,Y]&=-2Y\\ [X,Y]&=H\end{aligned}.$$

If V is a (finite-dimensional complex) vector space, and A,B, and C are operators on V satisfying

$$\begin{aligned}[][A,B]&=2B\\ [A,C]&=-2C\\ [B,C]&=A\end{aligned}$$

then because of the skew-symmetry and bilinearity of brackets, the linear map $\pi:\mathsf{sl}(2;\mathbb{C})\to\mathsf{gl}(V)$ satisfying

$$\pi(H)=A;\quad \pi(X)=B;\quad \pi(Y)=C$$

will be a representation of $\mathsf{sl}(2;\mathbb{C})$.

Theorem 82 *For each integer $m\geq 0$, there is an irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ with dimension $m+1$. Any two irreducible representations of $\mathsf{sl}(2;\mathbb{C})$ with the same dimension are equivalent. If π is an irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ with dimension $m+1$, then π is equivalent to the representation π_m described in Section 5.3.*

Proof

Let π be an irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ acting on a (finite-dimensional complex) space V. Our strategy is to diagonalize the operator $\pi(H)$. Of course, *a priori*, we don't know that $\pi(H)$ is diagonalizable. However, because we are working over the (algebraically closed) field of complex numbers, $\pi(H)$ must have at least one eigenvector.

The following lemma is the key to the entire proof.

Lemma 83 *Let u be an eigenvector of $\pi(H)$ with eigenvalue $\alpha \in \mathbb{C}$. Then*

$$\pi(H)\pi(X)u = (\alpha + 2)\pi(X)u.$$

Thus either $\pi(X)u = 0$, or else $\pi(X)u$ is an eigenvector for $\pi(H)$ with eigenvalue $\alpha + 2$. Similarly,

$$\pi(H)\pi(Y)u = (\alpha - 2)\pi(Y)u$$

so that either $\pi(Y)u = 0$, or else $\pi(Y)u$ is an eigenvector for $\pi(H)$ with eigenvalue $\alpha - 2$.

Remark

We call $\pi(X)$ the "raising operator," because it has the effect of raising the eigenvalue of $\pi(H)$ by 2, and we call $\pi(Y)$ the "lowering operator."

Proof of the Lemma

We know that $[\pi(H), \pi(X)] = \pi([H, X]) = 2\pi(X)$. Thus

$$\pi(H)\pi(X) - \pi(X)\pi(H) = 2\pi(X)$$

or

$$\pi(H)\pi(X) = \pi(X)\pi(H) + 2\pi(X).$$

Thus

$$\begin{aligned} \pi(H)\pi(X)u &= \pi(X)\pi(H)u + 2\pi(X)u \\ &= \pi(X)(\alpha u) + 2\pi(X)u \\ &= (\alpha + 2)\pi(X)u. \end{aligned}$$

Similarly, $[\pi(H), \pi(Y)] = -2\pi(Y)$, and so

$$\pi(H)\pi(Y) = \pi(Y)\pi(H) - 2\pi(Y)$$

so that

$$\begin{aligned} \pi(H)\pi(Y)u &= \pi(Y)\pi(H)u - 2\pi(Y)u \\ &= \pi(Y)(\alpha u) - 2\pi(Y)u \\ &= (\alpha - 2)\pi(Y)u. \end{aligned}$$

This is what we wanted to show. $\square$

As we have observed, $\pi(H)$ must have at least one eigenvector u ($u \neq 0$), with some eigenvalue $\alpha \in \mathbb{C}$. By the lemma,

$$\pi(H)\pi(X)u = (\alpha + 2)\pi(X)u$$

and more generally

$$\pi(H)\pi(X)^n u = (\alpha + 2n)\pi(X)^n u.$$

This means that either $\pi(X)^n u = 0$, or else $\pi(X)^n u$ is an eigenvector for $\pi(H)$ with eigenvalue $(\alpha + 2n)$.

Now, an operator on a finite-dimensional space can have only finitely many distinct eigenvalues. Thus the $\pi(X)^n u$'s cannot all be different from zero. Thus there is some $N \geq 0$ such that

$$\pi(X)^N u \neq 0$$

but

$$\pi(X)^{N+1} u = 0.$$

Define $u_0 = \pi(X)^N u$ and $\lambda = \alpha + 2N$. Then

$$\pi(H)u_0 = \lambda u_0 \tag{5.7}$$
$$\pi(X)u_0 = 0 \tag{5.8}$$

Then define

$$u_k = \pi(Y)^k u_0$$

for $k \geq 0$. By the second part of the lemma, we have

$$\pi(H)u_k = (\lambda - 2k)\, u_k. \tag{5.9}$$

Since, again, $\pi(H)$ can have only finitely many eigenvalues, the u_k's cannot all be non-zero.

Lemma 84 *With the above notation,*

$$\begin{aligned} \pi(X)u_k &= [k\lambda - k(k-1)]\, u_{k-1} \quad (k > 0) \\ \pi\,(X)\,u_0 &= 0. \end{aligned}$$

Proof of the lemma

We proceed by induction on k. In the case $k = 1$ we note that $u_1 = \pi(Y)u_0$. Using the commutation relation $[\pi(X), \pi(Y)] = \pi(H)$ we have

$$\pi(X)u_1 = \pi(X)\pi(Y)u_0 = (\pi(Y)\pi(X) + \pi(H))\, u_0.$$

But $\pi(X)u_0 = 0$, so we get

$$\pi(X)u_1 = \lambda u_0$$

which is the lemma in the case $k = 1$.

Now, by definition $u_{k+1} = \pi(Y)u_k$. Using (5.9) and induction we have

$$\begin{aligned}
\pi(X)u_{k+1} &= \pi(X)\pi(Y)u_k \\
&= (\pi(Y)\pi(X) + \pi(H))\, u_k \\
&= \pi(Y)\,[k\lambda - k(k-1)]\, u_{k-1} + (\lambda - 2k)u_k \\
&= [k\lambda - k(k-1) + (\lambda - 2k)]\, u_k.
\end{aligned}$$

Simplifying the last expression give the Lemma. □

Since $\pi(H)$ can have only finitely many eigenvalues, the u_k's cannot all be non-zero. There must therefore be an integer $m \geq 0$ such that

$$u_k = \pi(Y)^k u_0 \neq 0$$

for all $k \leq m$, but

$$u_{m+1} = \pi(Y)^{m+1} u_0 = 0.$$

Now if $u_{m+1} = 0$, then certainly $\pi(X)u_{m+1} = 0$. Then by Lemma 84,

$$0 = \pi(X)u_{m+1} = [(m+1)\lambda - m(m+1)]\, u_m = (m+1)(\lambda - m)u_m.$$

But $u_m \neq 0$, and $m + 1 \neq 0$ (since $m \geq 0$). Thus in order to have $(m+1)(\lambda - m)u_m$ equal to zero, we must have $\lambda = m$.

We have made considerable progress. Given a finite-dimensional irreducible representation π of $\mathsf{sl}(2;\mathbb{C})$, acting on a space V, there exists an integer $m \geq 0$ and non-zero vectors $u_0, \cdots u_m$ such that (putting λ equal to m)

$$\begin{aligned}
&\pi(H)u_k = (m - 2k)u_k \\
&\pi(Y)u_k = u_{k+1} \quad (k < m) \\
&\pi(Y)u_m = 0 \\
&\pi(X)u_k = [km - k(k-1)]\, u_{k-1} \quad (k > 0) \\
&\pi(X)u_0 = 0
\end{aligned} \tag{5.10}$$

The vectors $u_0, \cdots u_m$ must be linearly independent, since they are eigenvectors of $\pi(H)$ with distinct eigenvalues. Moreover, the $(m+1)$-dimensional span of $u_0, \cdots u_m$ is explicitly invariant under $\pi(H)$, $\pi(X)$, and $\pi(Y)$, and hence under $\pi(Z)$ for all $Z \in \mathsf{sl}(2;\mathbb{C})$. Since π is irreducible, this space must be all of V.

We have now shown that every irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ is of the form (5.10). It remains to show that everything of the form (5.10) is a representation, and that it is irreducible. That is, if we *define* $\pi(H)$, $\pi(X)$, and $\pi(Y)$ by (5.10) (where the u_k's are basis elements for some $(m+1)$-dimensional vector space), then we want to show that they have the right commutation relations to form a representation of $\mathsf{sl}(2;\mathbb{C})$, and that this representation is irreducible.

The computation of the commutation relations of $\pi(H)$, $\pi(X)$, and $\pi(Y)$ is straightforward, and is left as an exercise. Note that when dealing with $\pi(Y)$, you should treat separately the vectors u_k, $k < m$, and u_m. Irreducibility is also easy to check, by imitating the proof of Proposition 80. (See Exercise 6.)

We have now shown that there is an irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ in each dimension $m+1$, by explicitly writing down how H, X, and Y should act (Equation 5.10) in a basis. But we have shown more than this. We also have shown that any $(m+1)$-dimensional irreducible representation of $\mathsf{sl}(2;\mathbb{C})$ must be of the form (5.10). It follows that any two irreducible representations of $\mathsf{sl}(2;\mathbb{C})$ of dimension $(m+1)$ must be equivalent. For if π_1 and π_2 are two irreducible representations of dimension $(m+1)$, acting on spaces V_1 and V_2, then V_1 has a basis $u_0, \cdots u_m$ as in (5.10) and V_2 has a similar basis $\widetilde{u}_0, \cdots \widetilde{u}_m$. But then the map $\phi : V_1 \to V_2$ which sends u_k to $\widetilde{u}_k$ will be an isomorphism of representations. (Think about it.)

In particular, the $(m+1)$-dimensional representation π_m described in Section 5.3 must be equivalent to (5.10).This can be seen explicitly by introducing the following basis for V_m:

$$u_k = [\pi_m(Y)]^k \left(z_2^m\right) = (-1)^k \frac{m!}{(m-k)!} z_1^k z_2^{m-k} \qquad (k \leq m).$$

Then by definition $\pi_m(Y)u_k = u_{k+1}$ $(k < m)$, and it is clear that $\pi_m(Y)u_m = 0$. It is easy to see that $\pi_m(H)u_k = (m-2k)u_k$. The only thing left to check is the behavior of $\pi_m(X)$. But direct computation shows that

$$\pi_m(X)u_k = k(m-k+1)u_{k-1} = [km - k(k-1)]\, u_{k-1}.$$

as required.

This completes the proof of Theorem 82. □

5.5 Direct Sums of Representations and Complete Reducibility

One way of generating representations is to take some representations you know and combine them in some fashion. We will consider two methods of generating new representations from old ones, namely direct sums and tensor products of representations. In this section we consider direct sums; in the next section we look at tensor products. (There is one other standard construction of this sort, namely the dual of a representation. See Exercise 14.)

Definition 85 *Let G be a matrix Lie group, and let $\Pi_1, \Pi_2, \cdots \Pi_n$ be representations of G acting on vector spaces $V_1, V_2, \cdots V_n$. Then the **direct sum** of $\Pi_1, \Pi_2, \cdots \Pi_n$ is a representation $\Pi_1 \oplus \cdots \oplus \Pi_n$ of G acting on the space $V_1 \oplus \cdots \oplus V_n$, defined by*

$$[\Pi_1 \oplus \cdots \oplus \Pi_n(A)]\,(v_1, \cdots v_n) = (\Pi_1(A)v_1, \cdots, \Pi_n(A)v_n)$$

for all $A \in G$.

*Similarly, if $\mathfrak{g}$ is a Lie algebra, and $\pi_1, \pi_2, \cdots \pi_n$ are representations of $\mathfrak{g}$ acting on $V_1, V_2, \cdots V_n$, then we define the **direct sum** of $\pi_1, \pi_2, \cdots \pi_n$, acting on $V_1 \oplus \cdots \oplus V_n$ by*

$$[\pi_1 \oplus \cdots \oplus \pi_n(X)]\,(v_1, \cdots v_n) = (\pi_1(X)v_1, \cdots, \pi_n(X)v_n)$$

for all $X \in \mathfrak{g}$.

Remark

It is trivial to check that, say, $\Pi_1 \oplus \cdots \oplus \Pi_n$ is really a representation of G.

Definition 86 *A finite-dimensional representation of a group or Lie algebra, acting on a space V, is said to be **completely reducible** if the following property is satisfied: Given an invariant subspace $W \subset V$, and a second invariant subspace $U \subset W \subset V$, there exists a third invariant subspace $\widetilde{U} \subset W$ such that $U \cap \widetilde{U} = \{0\}$ and $U + \widetilde{U} = W$.*

The following Proposition shows that complete reducibility is a nice property for a representation to have.

Proposition 87 *A finite-dimensional completely reducible representation of a group or Lie algebra is equivalent to a direct sum of (one or more) irreducible representations.*

Proof

The proof is by induction on the dimension of the space V. If $\dim V = 1$, then automatically the representation is irreducible, since then V is has no non-trivial subspaces, let alone non-trivial invariant subspaces. Thus V is a direct sum of irreducible representations, with just one summand, namely V itself.

Suppose, then, that the Proposition holds for all representations with dimension strictly less than n, and that $\dim V = n$. If V is irreducible, then again we have a direct sum with only one summand, and we are done. If V is not irreducible, then there exists a non-trivial invariant subspace $U \subset V$. Taking $W = V$ in the definition of complete reducibility, we see that there is another invariant subspace $\widetilde{U}$ with $U \cap \widetilde{U} = \{0\}$ and $U + \widetilde{U} = V$. That is, $V \cong U \oplus \widetilde{U}$ as a vector space.

But since U and $\widetilde{U}$ are invariant, they can be viewed as representations in their own right. (That is, the *action* of our group or Lie algebra on U or $\widetilde{U}$ is a representation.) It is easy to see that in fact V is isomorphic to $U \oplus \widetilde{U}$ as a representation. Furthermore, it is easy to see that both U and $\widetilde{U}$ are completely reducible representations, since every invariant subspace W of, say, U is also an invariant subspace of V. But since U is non-trivial (i.e., $U \neq \{0\}$ and $U \neq V$), we have $\dim U < \dim V$ and $\dim \widetilde{U} < \dim V$. Thus by induction $U \cong U_1 \oplus \cdots U_n$ (as representations), with the U_i's irreducible, and $\widetilde{U} \cong \widetilde{U}_1 \oplus \cdots \widetilde{U}_m$, with the $\widetilde{U}_i$'s irreducible, so that $V \cong U_1 \oplus \cdots U_n \oplus \widetilde{U}_1 \oplus \cdots \widetilde{U}_m$. $\square$

Remark

Certain groups and Lie algebras have the property that *every* (finite-dimensional) representation is completely reducible. This is a very nice property, because it implies (by the above Proposition) that every representation is equivalent to a direct sum of irreducible representations. (And, as it turns out, this decomposition is essentially unique.) Thus for such groups and Lie algebras, if you know (up to equivalence) what all the irreducible representations are, then you know (up to equivalence) what *all* the representations are.

Unfortunately, not every representation is irreducible. For example, the standard representation of the Heisenberg group is not completely reducible. (See Exercise 8.)

Proposition 88 *Let G be a matrix Lie group. Let Π be a finite-dimensional unitary representation of G, acting on a finite-dimensional real or complex Hilbert space V. Then Π is completely reducible.*

Proof

So, we are assuming that our space V is equipped with an inner product, and that $\Pi(A)$ is unitary for each $A \in G$. Suppose that $W \subset V$ is invariant, and that $U \subset W \subset V$ is also invariant. Define

$$\widetilde{U} = U^{\perp} \cap W.$$

Then of course $\widetilde{U} \cap U = \{0\}$, and standard Hilbert space theory implies that $\widetilde{U} + U = W$.

It remains only to show that $\widetilde{U}$ is invariant. So suppose that $v \in U^{\perp} \cap W$. Since W is assumed invariant, $\Pi(A)v$ will be in W for any $A \in G$. We need to show that $\Pi(A)v$ is perpendicular to U. Well, since $\Pi(A^{-1})$ is unitary, then for any $u \in U$

$$\langle u, \Pi(A)v \rangle = \left\langle \Pi(A^{-1})u, \Pi(A^{-1})\Pi(A)v \right\rangle = \left\langle \Pi(A^{-1})u, v \right\rangle .$$

But U is assumed invariant, and so $\Pi(A^{-1})u \in U$. But then since $v \in U^{\perp}$, $\langle \Pi(A^{-1})u, v \rangle = 0$. This means that

$$\langle u, \Pi(A)v \rangle = 0$$

for all $u \in U$, i.e., $\Pi(A)v \in U^{\perp}$.

Thus $\widetilde{U}$ is invariant, and we are done. □

Proposition 89 *If G is a finite group, then every finite-dimensional real or complex representation of G is completely reducible.*

Proof

Suppose that Π is a representation of G, acting on a space V. Choose an arbitrary inner product $\langle\ \rangle$ on V. Then define a new inner product $\langle\ \rangle_G$ on V by

$$\langle v_1, v_2 \rangle_G = \sum_{g \in G} \langle \Pi(g)v_1, \Pi(g)v_2 \rangle .$$

It is very easy to check that indeed $\langle\ \rangle_G$ is an inner product. Furthermore, if $h \in G$, then

$$\begin{aligned} \langle \Pi(h)v_1, \Pi(h)v_2 \rangle_G &= \sum_{g \in G} \langle \Pi(g)\Pi(h)v_1, \Pi(g)\Pi(h)v_2 \rangle \\ &= \sum_{g \in G} \langle \Pi(gh)v_1, \Pi(gh)v_2 \rangle . \end{aligned}$$

But as g ranges over G, so does gh. Thus in fact

$$\langle \Pi(h)v_1, \Pi(h)v_2 \rangle_G = \langle v_1, v_2 \rangle_G .$$

That is, Π is a unitary representation with respect to the inner product $\langle\ \rangle_G$. Thus Π is completely reducible by Proposition 88. □

There is a variant of the above argument which can be used to prove the following result:

Proposition 90 *If G is a compact matrix Lie group, then every finite-dimensional real or complex representation of G is completely reducible.*

Proof

This proof requires the notion of Haar measure. A **left Haar measure** on a matrix Lie group G is a non-zero measure μ on the Borel σ-algebra in G with the following two properties: 1) it is locally finite, that is, every point in G has a neighborhood with finite measure, and 2) it is left-translation invariant. Left-translation invariance means that $\mu(gE) = \mu(E)$ for all $g \in G$ and for all Borel sets $E \subset G$, where

$$gE = \{ge \,|e \in E\}.$$

It is a fact which we cannot prove here that every matrix Lie group has a left Haar measure, and that this measure is unique up to multiplication by a constant. (One can analogously define right Haar measure, and a similar theorem holds for it. Left Haar measure and right Haar measure may or may not coincide; a group for which they do is called **unimodular**.)

Now, the key fact for our purpose is that left Haar measure is finite if and only if the group G is compact. So if Π is a finite-dimensional representation of a compact group G acting on a space V, then let $\langle\ \rangle$ be an arbitrary inner product on V, and define a new inner product $\langle\ \rangle_G$ on V by

$$\langle v_1, v_2\rangle_G = \int_G \langle \Pi(g)v_1, \Pi(g)v_2\rangle \; d\mu(g),$$

where μ is left Haar measure. Again, it is easy to check that $\langle\ \rangle_G$ is an inner product. Furthermore, if $h \in G$, then by the left-invariance of μ

$$\begin{aligned}\langle \Pi(h)v_1, \Pi(h)v_2\rangle_G &= \int_G \langle \Pi(g)\Pi(h)v_1, \Pi(g)\Pi(h)v_2\rangle \; d\mu(g) \\ &= \int_G \langle \Pi(gh)v_1, \Pi(gh)v_2\rangle \; d\mu(g) \\ &= \langle v_1, v_2\rangle_G .\end{aligned}$$

So Π is a unitary representation with respect to $\langle\ \rangle_G$, and thus completely reducible. Note that $\langle\ \rangle_G$ is well-defined only because μ is finite. $\square$

5.6 Tensor Products of Representations

Let U and V be finite-dimensional real or complex vector spaces. We wish to define the **tensor product** of U and V, which is will be a new vector space $U \otimes V$ "built" out of U and V. We will discuss the idea of this first, and then give the precise definition.

We wish to consider a formal "product" of an element u of U with an element v of V, denoted $u \otimes v$. The *space* $U \otimes V$ is then the space of linear combinations of such products, i.e., the space of elements of the form

$$a_1 u_1 \otimes v_1 + a_2 u_2 \otimes v_2 + \cdots + a_n u_n \otimes v_n. \tag{5.11}$$

Of course, if "$\otimes$" is to be interpreted as a product, then it should be bilinear. That is, we should have

$$\begin{aligned} (u_1 + au_2) \otimes v &= u_1 \otimes v + au_2 \otimes v \\ u \otimes (v_1 + av_2) &= u \otimes v_1 + au \otimes v_2. \end{aligned}$$

We do not assume that the product is commutative. (In fact, the product in the other order, $v \otimes u$, is in a different space, namely, $V \otimes U$.)

Now, if $e_1, e_2, \cdots, e_n$ is a basis for U and $f_1, f_2, \cdots, f_m$ is a basis for V, then using bilinearity it is easy to see that any element of the form (5.11) can be written as a linear combination of the elements $e_i \otimes f_j$. In fact, it seems reasonable to expect that $\{e_i \otimes f_j \,|0 \leq i \leq n, 0 \leq j \leq m\}$ should be a basis for the space $U \otimes V$. This in fact turns out to be the case.

Definition 91 *If U and V are finite-dimensional real or complex vector spaces, then a* ***tensor product*** *of U with V is a vector space W, together with a bilinear map $\phi : U \times V \to W$ with the following property: If ψ is any bilinear map of $U \times V$ into a vector space X, then there exists a unique linear map $\widetilde{\psi}$ of W into X such that the following diagram commutes:*

$$\begin{array}{ccc} U \times V & \overset{\phi}{\to} & W \\ \psi \searrow & & \swarrow \widetilde{\psi} \\ & X & \end{array} .$$

Remark

Note that the *bilinear* map ψ from $U \times V$ into X turns into the *linear* map $\widetilde{\psi}$ of W into X. This is one of the points of tensor products: bilinear maps on $U \times V$ turn into linear maps on W.

Theorem 92 *If U and V are any finite-dimensional real or complex vector spaces, then a tensor product (W, ϕ) exists. Furthermore, (W, ϕ) is unique up to canonical isomorphism. That is, if (W_1, ϕ_1) and (W_2, ϕ_2) are two tensor products, then there exists a unique vector space isomorphism $\Phi : W_1 \to W_2$ such that the following diagram commutes*

$$\begin{array}{ccc} U \times V & \overset{\phi_1}{\to} & W_1 \\ \phi_2 \searrow & & \swarrow \Phi \\ & W_2 & \end{array} .$$

Suppose that (W, ϕ) is a tensor product, and that $e_1, e_2, \cdots, e_n$ is a basis for U and $f_1, f_2, \cdots, f_m$ is a basis for V. Then $\{\phi(e_i, f_j) \,|0 \leq i \leq n, 0 \leq j \leq m\}$ is a basis for W.

Proof

Exercise 12.

Notation

Since the tensor product of U and V is essentially unique, we will let $U \otimes V$ denote an arbitrary tensor product space, and we will write $u \otimes v$ instead of $\phi(u, v)$. In this notation, the Theorem says that $\{e_i \otimes f_j \,|0 \leq i \leq n, 0 \leq j \leq m\}$ is a basis for $U \otimes V$, as expected. Note in particular that

$$\dim (U \otimes V) = (\dim U)\,(\dim V)$$

(not $\dim U + \dim V$).

Discussion

The defining property of $U \otimes V$ is called the **universal property** of tensor products. While it may seem that we are taking a simple idea and making it confusing, in fact there is a point to this universal property. Suppose we want to define a linear map T from $U \otimes V$ into some other space. The most sensible way to define this is to define T on elements of the form $u \otimes v$. (You might try defining it on a basis, but this forces you to worry about whether things depend on the choice of basis.) Now, every element of $U \otimes V$ is a linear combination of things of the form $u \otimes v$. However, this representation is far from unique. (Since, say, if $u = u_1 + u_2$, then you can rewrite $u \otimes v$ as $u_1 \otimes v + u_2 \otimes v$.)

Thus if you try to define T by what it does to elements of the form $u \otimes v$, you have to worry about whether T is well-defined. This is where the universal property comes in. Suppose that $\psi(u, v)$ is some bilinear expression in u, v. Then the universal property says precisely that there is a unique linear map T $(= \widetilde{\psi})$ such that

$$T(u \otimes v) = \psi(u, v).$$

(Think about it and make sure that you see that this is really what the universal property says.)

The conclusion is this: You can define a linear map T on $U \otimes V$ by defining it on elements of the form $u \otimes v$, and this will be well-defined, *provided* that $T(u \otimes v)$ is bilinear in (u, v). The following Proposition shows how to make use of this idea.

Proposition 93 *Let U and V be finite-dimensional real or complex vector spaces. Let $A : U \to U$ and $B : V \to V$ be linear operators. Then there exists a unique linear operator from $U \otimes V$ to $U \otimes V$, denoted $A \otimes B$, such that*

$$A \otimes B(u \otimes v) = (Au) \otimes (Bv)$$

for all $u \in U$, $v \in V$.

If A_1, A_2 *are linear operators on* U *and* B_1, B_2 *are linear operators on* V, *then*

$$(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1 A_2) \otimes (B_1 B_2).$$

Proof

Define a map ψ from $U \times V$ into $U \otimes V$ by

$$\psi(u, v) = (Au) \otimes (Bv).$$

Since A and B are linear, and since $\otimes$ is bilinear, ψ will be a bilinear map of $U \times V$ into $U \otimes V$. But then the universal property says that there is an associated linear map $\widetilde{\psi} : U \otimes V \to U \otimes V$ such that

$$\widetilde{\psi}(u \otimes v) = \psi(u, v) = (Au) \otimes (Bv).$$

Then $\widetilde{\psi}$ is the desired map $A \otimes B$.

Now, if A_1, A_2 are operators on U and B_1, B_2 are operators on V, then compute that

$$\begin{aligned}(A_1 \otimes B_1)(A_2 \otimes B_2)(u \otimes v) &= (A_1 \otimes B_1)(A_2 u \otimes B_2 v) \\ &= A_1 A_2 u \otimes B_1 B_2 v.\end{aligned}$$

This shows that $(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1 A_2) \otimes (B_1 B_2)$ are equal on elements of the form $u \otimes v$. Since every element of $U \otimes V$ can be written as a linear combination of things of the form $u \otimes v$ (in fact of $e_i \otimes f_j$), $(A_1 \otimes B_1)(A_2 \otimes B_2)$ and $(A_1 A_2) \otimes (B_1 B_2)$ must be equal on the whole space. $\square$

We are now ready to define tensor products of representations. There are two different approaches to this, both of which are important. The first approach starts with a representation of a group G acting on a space V and a representation of another group H acting on a space U, and produces a representation of the product group $G \times H$ acting on the space $U \otimes V$. The second approach starts with two different representations of the same group G, acting on spaces U and V, and produces a representation of G acting on $U \otimes V$. Both of these approaches can be adapted to apply to Lie algebras.

Definition 94 *Let* G *and* H *be matrix Lie groups. Let* Π_1 *be a representation of* G *acting on a space* U *and let* Π_2 *be a representation of* H *acting on a space* V. *The the* ***tensor product*** *of* Π_1 *and* Π_2 *is a representation* $\Pi_1 \otimes \Pi_2$ *of* $G \times H$ *acting on* $U \otimes V$ *defined by*

$$\Pi_1 \otimes \Pi_2(A, B) = \Pi_1(A) \otimes \Pi_2(B)$$

for all $A \in G$ *and* $B \in H$.

Remarks

Using the above Proposition, it is very easy to check that indeed $\Pi_1 \otimes \Pi_2$ is a representation of $G \times H$.

Now, if G and H are matrix Lie groups, that is, G is a closed subgroup of $\mathsf{GL}(n;\mathbb{C})$ and H is a closed subgroup of $\mathsf{GL}(m;\mathbb{C})$, then $G \times H$ can be regarded in an obvious way as a closed subgroup of $\mathsf{GL}(n+m;\mathbb{C})$. Thus the direct product of matrix Lie groups can be regarded as a matrix Lie group. It is easy to check that the Lie algebra of $G \times H$ is isomorphic to the direct sum of the Lie algebra of G and the Lie algebra of H. See Exercise 13.

In light of Proposition 77, the representation $\Pi_1 \otimes \Pi_2$ of $G \times H$ gives rise to a representation of the Lie algebra of $G \times H$, namely $\mathfrak{g} \oplus \mathfrak{h}$. The following Proposition shows that this representation of $\mathfrak{g} \oplus \mathfrak{h}$ is not what you might expect at first.

Proposition 95 *Let G and H be matrix Lie groups, let Π_1, Π_2 be representations of G, H respectively, and consider the representation $\Pi_1 \otimes \Pi_2$ of $G \times H$. Let $\pi_1 \otimes \pi_2$ denote the associated representation of the Lie algebra of $G \times H$, namely $\mathfrak{g} \oplus \mathfrak{h}$. Then for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$*

$$\pi_1 \otimes \pi_2(X, Y) = \pi_1(X) \otimes I + I \otimes \pi_2(Y).$$

Proof

Suppose that $u(t)$ is a smooth curve in U and $v(t)$ is a smooth curve in V. Then we verify the product rule in the usual way:

$$\begin{aligned}
&\lim_{h\to 0} \frac{u(t+h) \otimes v(t+h) - u(t) \otimes v(t)}{h} \\
&= \lim_{h\to 0} \frac{u(t+h) \otimes v(t+h) - u(t+h) \otimes v(t)}{h} + \frac{u(t+h) \otimes v(t) - u(t) \otimes v(t)}{h} \\
&= \lim_{h\to 0} \left[u(t+h) \otimes \frac{(v(t+h) - v(t))}{h} \right] + \lim_{h\to 0} \left[\frac{(u(t+h) - u(t))}{h} \otimes v(t) \right].
\end{aligned}$$

Thus

$$\frac{d}{dt}\left(u(t) \otimes v(t)\right) = \frac{du}{dt} \otimes v(t) + u(t) \otimes \frac{dv}{dt}.$$

This being the case, we can compute $\pi_1 \otimes \pi_2(X, Y)$:

$$\begin{aligned}
\pi_1 \otimes \pi_2(X, Y)(u \otimes v) &= \left.\frac{d}{dt}\right|_{t=0} \Pi_1 \otimes \Pi_2(e^{tX}, e^{tY})(u \otimes v) \\
&= \left.\frac{d}{dt}\right|_{t=0} \Pi_1(e^{tX})u \otimes \Pi_2(e^{tY})v \\
&= \left(\left.\frac{d}{dt}\right|_{t=0} \Pi_1(e^{tX})u \right) \otimes v + u \otimes \left(\left.\frac{d}{dt}\right|_{t=0} \Pi_2(e^{tY})v \right).
\end{aligned}$$

This shows that $\pi_1 \otimes \pi_2(X, Y) = \pi_1(X) \otimes I + I \otimes \pi_2(Y)$ on elements of the form $u \otimes v$, and therefore on the whole space $U \otimes V$. $\square$

Definition 96 *Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras, and let π_1 and π_2 be representations of $\mathfrak{g}$ and $\mathfrak{h}$, acting on spaces U and V. Then the* ***tensor product*** *of π_1 and π_2, denoted $\pi_1 \otimes \pi_2$, is a representation of $\mathfrak{g} \oplus \mathfrak{h}$ acting on $U \otimes V$, given by*

$$\pi_1 \otimes \pi_2(X, Y) = \pi_1(X) \otimes I + I \otimes \pi_2(Y)$$

for all $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$.

Remarks

It is easy to check that this indeed defines a representation of $\mathfrak{g} \oplus \mathfrak{h}$. Note that if we defined $\pi_1 \otimes \pi_2(X, Y) = \pi_1(X) \otimes \pi_2(Y)$, this would *not* be a represntation of $\mathfrak{g} \oplus \mathfrak{h}$, for this is not even a linear map. (E.g., we would then have $\pi_1 \otimes \pi_2(2X, 2Y) = 4\pi_1 \otimes \pi_2(X, Y)$!) Note also that the above definition applies even if π_1 and π_2 do not come from a representation of any matrix Lie group.

Definition 97 *Let G be a matrix Lie group, and let Π_1 and Π_2 be representations of G, acting on spaces V_1 and V_2. Then the* ***tensor product*** *of Π_1 and Π_2 is a representation of G acting on $V_1 \otimes V_2$ defined by*

$$\Pi_1 \otimes \Pi_2(A) = \Pi_1(A) \otimes \Pi_2(A)$$

for all $A \in G$.

Proposition 98 *With the above notation, the associated representation of the Lie algebra $\mathfrak{g}$ satisfies*

$$\pi_1 \otimes \pi_2(X) = \pi_1(X) \otimes I + I \otimes \pi_2(X)$$

for all $X \in \mathfrak{g}$.

Proof

Using the product rule,

$$\begin{aligned} \pi_1 \otimes \pi_2(X)\left(u \otimes v\right) &= \left.\frac{d}{dt}\right|_{t=0} \Pi_1\left(e^{tX}\right) u \otimes \Pi_2\left(e^{tX}\right) v \\ &= \pi_1\left(X\right) u \otimes v + v \otimes \pi_2\left(X\right) u. \end{aligned}$$

This is what we wanted to show. $\square$

Definition 99 *If $\mathfrak{g}$ is a Lie algebra, and π_1 and π_2 are representations of $\mathfrak{g}$ acting on spaces V_1 and V_2, then the* ***tensor product*** *of π_1 and π_2 is a representation of $\mathfrak{g}$ acting on the space $V_1 \otimes V_2$ defined by*

$$\pi_1 \otimes \pi_2(X) = \pi_1(X) \otimes I + I \otimes \pi_2(X)$$

for all $X \in \mathfrak{g}$.

Remarks

It is easy to check that $\Pi_1 \otimes \Pi_2$ and $\pi_1 \otimes \pi_2$ are actually representations of G and $\mathfrak{g}$, respectively. There is some ambiguity in the notation, say, $\Pi_1 \otimes \Pi_2$. For even if Π_1 and Π_2 are both representations of the same group G, we could still regard $\Pi_1 \otimes \Pi_2$ as a representation of $G \times G$, by taking $H = G$ in definition 94. We will rely on context to make clear whether we are thinking of $\Pi_1 \otimes \Pi_2$ as a representation of $G \times G$ or as representation of G.

Suppose Π_1 and Π_2 are *irreducible* representations of a group G. If we regard $\Pi_1 \otimes \Pi_2$ as a representation of G, it may no longer be irreducible. If it is not irreducible, one can attempt to decompose it as a direct sum of irreducible representations. This process is called **Clebsch-Gordan** theory. In the case of $\mathsf{SU}(2)$, this theory is relatively simple. (In the physics literature, the problem of analyzing tensor products of representations of $\mathsf{SU}(2)$ is called "addition of angular momentum.") See Exercise 15.

5.7 Schur's Lemma

Let Π and Σ be representations of a matrix Lie group G, acting on spaces V and W. Recall that a **morphism** of representations is a linear map $\phi : V \to W$ with the property that

$$\phi\left(\Pi(A)v\right) = \Sigma(A)\left(\phi(v)\right)$$

for all $v \in V$ and all $A \in G$. Schur's Lemma is an extremely important result which tells us about morphisms of irreducible representations. Part of Schur's Lemma applies to both real and complex representations, but part of it applies only to complex representations.

It is desirable to be able to state Schur's lemma simultaneously for groups and Lie algebras. In order to do so, we need to indulge in a common abuse of notation. If, say, Π is a representation of G acting on a space V, we will refer to V as the representation, without explicit reference to Π.

Theorem 100 (Schur's Lemma)

1. *Let V and W be irreducible real or complex representations of a group or Lie algebra, and let $\phi : V \to W$ be a morphism. Then either $\phi = 0$ or ϕ is an isomorphism.*

2. *Let V be an irreducible complex representation of a group or Lie algebra, and let $\phi : V \to V$ be a morphism of V with itself. Then $\phi = \lambda I$, for some $\lambda \in \mathbb{C}$.*

3. *Let V and W be irreducible complex representations of a group or Lie algebra, and let $\phi_1, \phi_2 : V \to W$ be non-zero morphisms. Then $\phi_1 = \lambda\phi_2$, for some $\lambda \in \mathbb{C}$.*

Corollary 101 *Let* Π *be an irreducible complex representation of a matrix Lie group* G*. If* A *is in the center of* G*, then* $\Pi(A) = \lambda I$*. Similarly, if* π *is an irreducible complex representation of a Lie algebra* $\mathfrak{g}$*, and if* X *is in the center of* $\mathfrak{g}$ *(i.e.,* $[X, Y] = 0$ *for all* $Y \in \mathfrak{g}$*), then* $\pi(X) = \lambda I$*.*

Corollary 102 *An irreducible complex representation of a commutative group or Lie algebra is one-dimensionsal.*

Proof of Corollary 1

We prove the group case; the proof of the Lie algebra case is the same. If A is in the center of G, then for all $B \in G$,

$$\Pi(A)\Pi(B) = \Pi(AB) = \Pi(BA) = \Pi(B)\Pi(A).$$

But this says exactly that $\Pi(A)$ is a morphism of Π with itself. So by Point 2 of Schur's lemma, $\Pi(A)$ is a multiple of the identity. □

Proof of Corollary 2

Again, we prove only the group case. If G is commutative, then the center of G is all of G, so by the previous corollary $\Pi(A)$ is a multiple of the identity for each $A \in G$. But this means that *every* subspace of V is invariant! Thus the only way that V can fail to have a non-trivial invariant subspace is for it not to have any non-trivial subspaces. This means that V must be one-dimensional. (Recall that we do not allow V to be zero-dimensional.) □

Proof of Schur's Lemma

As usual, we will prove just the group case; the proof of the Lie algebra case requires only the obvious notational changes.

Proof of 1. Saying that ϕ is a morphism means $\phi(\Pi(A)v) = \Sigma(A)\,(\phi(v))$ for all $v \in V$ and all $A \in G$. Now suppose that $v \in \ker(\phi)$. Then

$$\phi(\Pi(A)v) = \Sigma(A)\phi(v) = 0.$$

This shows that $\ker\phi$ is an invariant subspace of V. Since V is irreducible, we must have $\ker\phi = 0$ or $\ker\phi = V$. Thus ϕ is either one-to-one or zero.

Suppose ϕ is one-to-one. Then the image of ϕ is a non-zero subspace of W. On the other hand, the image of ϕ is invariant, for if $w \in W$ is of the form $\phi(v)$ for some $v \in V$, then

$$\Sigma(A)w = \Sigma(A)\phi(v) = \phi(\Pi(A)v).$$

Since W is irreducible and image(V) is non-zero and invariant, we must have image$(V) = W$. Thus ϕ is either zero or one-to-one and onto.

Proof of 2. Suppose now that V is an irreducible *complex* representation, and that $\phi : V \to V$ is a morphism of V to itself. This means that $\phi\Pi(A) = \Pi(A)\phi$ for all $A \in G$, i.e., that ϕ commutes with all of the $\Pi(A)$'s. Now, since we are over an algebraically complete field, ϕ must have at least one eigenvalue $\lambda \in \mathbb{C}$. Let U denote the eigenspace for ϕ associated to the eigenvalue λ, and let $u \in U$. Then for each $A \in G$

$$\phi\left(\Pi(A)u\right) = \Pi(A)\phi(v) = \lambda\Pi(A)u.$$

Thus applying $\Pi(A)$ to an eigenvector of ϕ with eigenvalue λ yields another eigenvector of ϕ with eigenvalue λ. That is, U is invariant.

Since λ is an eigenvalue, $U \neq 0$, and so we must have $U = V$. But this means that $\phi(v) = \lambda v$ for all $v \in V$, i.e., that $\phi = \lambda I$.

Proof of 3. If $\phi_2 \neq 0$, then by (1) ϕ_2 is an isomorphism. Now look at $\phi_1 \circ \phi_2^{-1}$. As is easily checked, the composition of two morphisms is a morphism, so $\phi_1 \circ \phi_2^{-1}$ is a morphism of W with itself. Thus by (2), $\phi_1 \circ \phi_2^{-1} = \lambda I$, whence $\phi_1 = \lambda\phi_2$. $\square$

5.8 Group Versus Lie Algebra Representations

We know from Chapter 3 (Theorem 46) that every Lie group homomorphism gives rise to a Lie algebra homomorphism. In particular, this shows (Proposition 77) that every representation of a matrix Lie group gives rise to a representation of the associated Lie algebra. The goal of this section is to investigate the reverse process. That is, given a representation of the Lie algebra, under what circumstances is there an associated representation of the Lie group?

The climax of this section is Theorem 105, which states that if G is a *connected* and *simply connected* matrix Lie group with Lie algebra $\mathfrak{g}$, and if π is a representation of $\mathfrak{g}$, then there is a unique representation Π of G such that Π and π are related as in Proposition 77. Our proof of this theorem will make use of the Baker-Campbell-Hausdorff formula from Chapter 4. Before turning to this general theorem, we will examine two special cases, namely $\mathsf{SO}(3)$ and $\mathsf{SU}(2)$, for which we can work things out by hand. See Bröcker and tom Dieck, Chapter II, Section 5.

We have shown (Theorem 82) that every irreducible complex representation of $\mathsf{su}(2)$ is equivalent to one of the representations π_m described in Section 5.3. (Recall that the irreducible complex representations of $\mathsf{su}(2)$ are in one-to-one correspondence with the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$.) Each of the representations π_m of $\mathsf{su}(2)$ was constructed from the corresponding representation Π_m of the group $\mathsf{SU}(2)$. Thus we see, by brute force computation, that every irreducible complex representation of $\mathsf{su}(2)$ actually comes from a representation of the group $\mathsf{SU}(2)$! This is consistent with the fact that $\mathsf{SU}(2)$ is simply connected (Chapter 2, Prop. 24).

Let us now consider the situation for $\mathsf{SO}(3)$. (Which is not simply connected.) We know from Exercise 10 of Chapter 3 that the Lie algebras $\mathsf{su}(2)$ and $\mathsf{so}(3)$ are isomorphic. In particular, if we take the basis

$$E_1 = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad E_2 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad E_3 = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

for $\mathsf{su}(2)$ and the basis

$$F_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad F_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad F_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

then direct computation shows that $[E_1, E_2] = E_3$, $[E_2, E_3] = E_1$, $[E_3, E_1] = E_2$, and similarly with the E's replaced by the F's. Thus the map $\phi : \mathsf{so}(3) \to \mathsf{su}(2)$ which takes F_i to E_i will be a Lie algebra isomorphism.

Since $\mathsf{su}(2)$ and $\mathsf{so}(3)$ are isomorphic Lie algebras, they must have "the same" representations. Specifically, if π is a representation of $\mathsf{su}(2)$, then $\pi \circ \phi$ will be a representation of $\mathsf{so}(3)$, and every representation of $\mathsf{so}(3)$ is of this form. In particular, the irreducible representations of $\mathsf{so}(3)$ are precisely of the form $\sigma_m = \pi_m \circ \phi$. We wish to determine, for a particular m, whether there is a representation Σ_m of the group $\mathsf{SO}(3)$ such that σ_m and Σ_m are related as in Proposition 77.

Proposition 103 *Let $\sigma_m = \pi_m \circ \phi$ be the irreducible complex representations of the Lie algebra $\mathsf{so}(3)$ ($m \geq 0$). If m is even, then there is a representation Σ_m of the group $\mathsf{SO}(3)$ such that σ_m and Σ_m are related as in Proposition 77. If m is odd, then there is no such representation of $\mathsf{SO}(3)$.*

Remarks

Note that the condition that m be even is equivalent to the condition that $\dim V_m = m+1$ be odd. Thus it is the odd-dimensional representations of the Lie algebra $\mathsf{so}(3)$ which come from group representations.

In the physics literature, the representations of $\mathsf{su}(2)/\mathsf{so}(3)$ are labeled by the parameter $l = m/2$. In terms of this notation, a representation of $\mathsf{so}(3)$ comes from a representation of $\mathsf{SO}(3)$ if and only if l is an integer. The representations with l an integer are called "integer spin"; the others are called "half-integer spin."

Proof

Case 1: m odd. In this case, we want to prove that there is no representation Σ_m such that σ_m and Σ_m are related as in Proposition 77. (We have already considered the case $m = 1$ in Exercise 7.) Suppose, to the contrary, that there is such a Σ_m. Then Proposition 77 says that

$$\Sigma_m(e^X) = e^{\sigma_m(X)}$$

for all $X \in \mathsf{so}(3)$. In particular, take $X = 2\pi F_1$. Then, computing as in Chapter 3, Section 3.2 we see that

$$e^{2\pi F_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi & -\sin 2\pi \\ 0 & \sin 2\pi & \cos 2\pi \end{pmatrix} = I.$$

Thus on the one hand $\Sigma_m\left(e^{2\pi F_1}\right) = \Sigma_m(I) = I$, while on the other hand $\Sigma_m\left(e^{2\pi F_1}\right) = e^{2\pi\sigma_m(F_1)}$.

Let us compute $e^{2\pi\sigma_m(F_1)}$. By definition, $\sigma_m(F_1) = \pi_m(\phi(F_1)) = \pi_m(E_1)$. But, $E_1 = \frac{i}{2}H$, where as usual

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We know that there is a basis $u_0, u_1, \cdots, u_m$ for V_m such that u_k is an eigenvector for $\pi_m(H)$ with eigenvalue $m - 2k$. This means that u_k is also an eigenvector for $\sigma_m(F_1) = \frac{i}{2}\pi_m(H)$, with eigenvalue $\frac{i}{2}(m - 2k)$. Thus in the basis $\{u_k\}$ we have

$$\sigma_m(F_1) = \begin{pmatrix} \frac{i}{2}m & & & \\ & \frac{i}{2}(m-2) & & \\ & & \ddots & \\ & & & \frac{i}{2}(-m) \end{pmatrix}.$$

But we are assuming the m is odd! This means that $m - 2k$ is an odd integer. Thus $e^{2\pi\frac{i}{2}(m-2k)} = -1$, and in the basis $\{u_k\}$

$$e^{2\pi\sigma_m(F_1)} = \begin{pmatrix} e^{2\pi\frac{i}{2}m} & & & \\ & e^{2\pi\frac{i}{2}(m-2)} & & \\ & & \ddots & \\ & & & e^{2\pi\frac{i}{2}(-m)} \end{pmatrix} = -I.$$

Thus on the one hand, $\Sigma_m\left(e^{2\pi F_1}\right) = \Sigma_m(I) = I$, while on the other hand $\Sigma_m\left(e^{2\pi F_1}\right) = e^{2\pi\sigma_m(F_1)} = -I$. This is a contradiction, so there can be no such group representation Σ_m.

Case 2: m is even. We will use the following:

Lemma 104 *There exists a Lie group homomorphism $\Phi : \mathsf{SU}(2) \to \mathsf{SO}(3)$ such that*

1. *Φ maps $\mathsf{SU}(2)$ onto $\mathsf{SO}(3)$*
2. $\ker\Phi = \{I, -I\}$
3. *The associated Lie algebra homomorphism $\widetilde{\Phi} : \mathsf{su}(2) \to \mathsf{so}(3)$ is an isomorphism which takes E_i to F_i. That is, $\widetilde{\Phi} = \phi^{-1}$.*

Proof of the Lemma

Exercise 17. □

Now consider the representations Π_m of $\mathsf{SU}(2)$. I claim that if m is even, then $\Pi_m(-I) = I$. To see this, note that

$$e^{2\pi E_1} = \exp\begin{pmatrix} \pi i & 0 \\ 0 & -\pi i \end{pmatrix} = -I.$$

Thus $\Pi_m(-I) = \Pi_m(e^{2\pi E_1}) = e^{\pi_m(2\pi E_1)}$. But as in Case 1,

$$e^{\pi_m(2\pi E_1)} = \begin{pmatrix} e^{2\pi \frac{i}{2} m} & & & \\ & e^{2\pi\frac{i}{2}(m-2)} & & \\ & & \ddots & \\ & & & e^{2\pi \frac{i}{2}(-m)} \end{pmatrix}.$$

Only, this time, m is even, and so $\frac{i}{2}(m-2k)$ is an integer, so that $\Pi_m(-I) = e^{\pi_m(2\pi E_1)} = I$.

Since $\Pi_m(-I) = I$, $\Pi_m(-U) = \Pi_m(U)$ for all $U \in \mathsf{SU}(2)$. According to Lemma 104, for each $R \in \mathsf{SO}(3)$, there is a unique pair of elements $\{U, -U\}$ such that $\Phi(U) = \Phi(-U) = R$. Since $\Pi_m(U) = \Pi_m(-U)$, it makes sense to define

$$\Sigma_m(R) = \Pi_m(U).$$

It is easy to see that Σ_m is a Lie group homomorphism (hence, a representation). By construction, we have

$$\Pi_m = \Sigma_m \circ \Phi. \tag{5.12}$$

Now, if $\widetilde{\Sigma}_m$ denotes the Lie algebra representation associated to Σ_m, then it follows from (5.12) that

$$\pi_m = \widetilde{\Sigma}_m \circ \widetilde{\Phi}.$$

But the Lie algebra homomorphism $\widetilde{\Phi}$ takes E_i to F_i, that is, $\widetilde{\Phi} = \phi^{-1}$. So $\pi_m = \widetilde{\Sigma}_m \circ \phi^{-1}$, or $\widetilde{\Sigma}_m = \pi_m \circ \phi$. Thus $\widetilde{\Sigma}_m = \sigma_m$, which is what we want to show. □

It is now time to state the main theorem.

Theorem 105 *1. Let G, H be a matrix Lie groups, let $\phi_1, \phi_2 : G \to H$ be Lie group homomorphisms, and let $\widetilde{\phi}_1, \widetilde{\phi}_2 : \mathfrak{g} \to \mathfrak{h}$ be the associated Lie algebra homomorphisms. If G is connected and $\widetilde{\phi}_1 = \widetilde{\phi}_2$, then $\phi_1 = \phi_2$.*

2. Let G, H be a matrix Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$. Let $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. If G is connected and simply connected, then there exists a unique Lie group homomorphism $\phi : G \to H$ such that ϕ and $\widetilde{\phi}$ are related as in Theorem 46 of Chapter 3.

This has the following corollaries.

Corollary 106 *Suppose G and H are connected, simply connected matrix Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$. If $\mathfrak{g} \cong \mathfrak{h}$ then $G \cong H$.*

Proof

Let $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra isomorphism. By Theorem 105, there exists an associated Lie group homomorphism $\phi : G \to H$. Since $\widetilde{\phi}^{-1} : \mathfrak{h} \to \mathfrak{g}$ is also a Lie algebra homomorphism, there is a corresponding Lie group homomorphism $\psi : H \to G$. We want to show that ϕ and ψ are inverses of each other.

Well, $\widetilde{\phi \circ \psi} = \widetilde{\phi} \circ \widetilde{\psi} = I_{\mathfrak{h}}$, so by the Point 1 of the Theorem, $\phi \circ \psi = I_H$. Similarly, $\psi \circ \phi = I_G$. □

Corollary 107
1. *Let G be a connected matrix Lie group, let Π_1 and Π_2 be representations of G, and let π_1 and π_2 be the associated Lie algebra representations. If π_1 and π_2 are equivalent, then Π_1 and Π_2 are equivalent.*
2. *Let G be connected and simply connected. If π is a representation of $\mathfrak{g}$, then there exists a representation Π of G, acting on the same space, such that Π and π are related as in Proposition 77.*

Proof of the Corollary

For (1), let Π_1 act on V and Π_2 on W. We assume that the associated Lie algebra representations are equivalent, i.e., that there exists an invertible linear map $\phi : V \to W$ such that

$$\phi\left(\pi_1(X)v\right) = \pi_2(X)\phi(v)$$

for all $X \in \mathfrak{g}$ and all $v \in V$. This is the same as saying that $\phi\pi_1(X) = \pi_2(X)\phi$, or equivalently that $\phi\pi_1(X)\phi^{-1} = \pi_2(X)$ (for all $X \in \mathfrak{g}$).

Now define a map $\Sigma_2 : G \to \mathsf{GL}(W)$ by the formula

$$\Sigma_2(A) = \phi\Pi_1(A)\phi^{-1}.$$

It is trivial to check that Σ_2 is a homomorphism. Furthermore, differentiation shows that the associated Lie algebra homomorphism is

$$\sigma_2(X) = \phi\pi_1(X)\phi^{-1} = \pi_2(X)$$

for all X. Then by (1) in the Theorem, we must also have $\Sigma_2 = \Pi_2$, i.e.,

$$\phi\Pi_1(A)\phi^{-1} = \Pi_2(A)$$

for all $A \in G$. But this shows that Π_1 and Π_2 are equivalent.

Point (2) of the Corollary follows immediately from Point (2) of the Theorem, by taking $H = \mathsf{GL}(V)$. □

Proof of the Theorem

Step 1: Verify Point (1) of the Theorem.

Since G is connected, Corollary 54 of Chapter 3 tells us that every element A of G is a finite product of the form $A = \exp X_1 \exp X_2 \cdots \exp X_n$, with $X_i \in \mathfrak{g}$. But then if $\widetilde{\phi}_1 = \widetilde{\phi}_2$, we have

$$\phi_1\left(e^{X_1}\cdots e^{X_n}\right) = e^{\widetilde{\phi}_1(X_1)}\cdots e^{\widetilde{\phi}_1(X_n)} = e^{\widetilde{\phi}_2(X_1)}\cdots e^{\widetilde{\phi}_2(X_n)} = \phi_2\left(e^{X_1}\cdots e^{X_n}\right).$$

So we now need only prove Point (2).

Step 2: Define ϕ in a neighborhood of the identity.

Proposition 51 of Chapter 3 says that the exponential mapping for G has a local inverse which maps a neighborhood V of the identity into the Lie algebra $\mathfrak{g}$. On this neighborhood V we can *define* $\phi : V \to H$ by

$$\phi(A) = \exp\left\{\widetilde{\phi}(\log A)\right\}.$$

That is

$$\phi = \exp\circ\widetilde{\phi}\circ\log.$$

(Note that if there is to be a homomorphism ϕ as in Theorem 46 of Chapter 3, then on V, ϕ must be $\exp\circ\widetilde{\phi}\circ\log$.)

It follows from Corollary 68 to the Baker-Campbell-Hausdorff formula that this ϕ is a "local homomorphism." That is, if A and B are in V, and *if* AB happens to be in V as well, then $\phi(AB) = \phi(A)\phi(B)$. (See the discussion at the beginning of Chapter 4.)

Step 3: Define ϕ along a path.

Recall that when we say G is connected, we really mean that G is path-connected. Thus for any $A \in G$, there exists a path $A(t) \in G$ with $A(0) = I$ and $A(1) = A$. A compactness argument shows that there exists numbers $0 = t_0 < t_1 < t_2 \cdots < t_n = 1$ such that

$$A(s)A(t_i)^{-1} \in V \tag{5.13}$$

for all s between t_i and t_{i+1}.

In particular, for $i = 0$, we have $A(s) \in V$ for $0 \le s \le t_1$. Thus we can define $\phi(A(s))$ by Step 2 for $s \in [0, t_1]$. Now, for $s \in [t_1, t_2]$ we have by (**??**) $A(s)A(t_1)^{-1} \in V$. Moving the $A(t_1)$ to the other side, this means that for $s \in [t_1, t_2]$ we can write

$$A(s) = \left[A(s)A(t_1)^{-1}\right] A(t_1).$$

with $A(s)A(t_1)^{-1} \in V$. If ϕ is to be a homomorphism, we must have

$$\phi(A(s)) = \phi\left(\left[A(s)A(t_1)^{-1}\right] A(t_1)\right) = \phi\left(A(s)A(t_1)^{-1}\right) \phi(A(t_1)). \tag{5.14}$$

But $\phi(A(t_1))$ has already been defined, and we can define $\phi(A(s)A(t_1)^{-1})$ by Step 2. In this way we can use (5.14) to define $\phi(A(s))$ for $s \in [t_1, t_2]$.

Proceeding on in the same way, we can define $\phi(A(s))$ successively on each interval $[t_i, t_{i+1}]$ until eventually we have defined $\phi(A(s))$ on the whole time interval $[0, 1]$. This in particular serves to define $\phi(A(1)) = \phi(A)$.

Step 4: Prove independence of path.

In Step 3, we "defined" $\phi(A)$ by defining ϕ along a path joining the identity to A. For this to make sense as a definition of $\phi(A)$ we have to prove that the answer is independent of the choice of path, and also, for a particular path, independent of the choice of partition $(t_0, t_1, \cdots t_n)$.

To establish independence of partition, we first show that passing from a particular partition to a refinement of that partition doesn't change the answer. (A refinement of a partition is one which contains all the points of the original partition, plus some other ones.) This is proved by means of the Baker-Campbell-Hausdorff formula. For example, suppose we insert an extra partition point s between t_0 and t_1. Under the old partition we have

$$\phi(A(t_1)) = \exp \circ \widetilde{\phi} \circ \log(A(t_1)). \tag{5.15}$$

Under the new partition we write

$$A(t_1) = \left[A(t_1)A(s)^{-1}\right] A(s)$$

so that

$$\phi(A(t_1)) = \exp \circ \widetilde{\phi} \circ \log\left(A(t_1)A(s)^{-1}\right) \exp \circ \widetilde{\phi} \circ \log(A(s)). \tag{5.16}$$

But (as noted in Step 2), Corollary 68 of the Baker-Campbell-Hausdorff formula (Chapter 4, Section 4.2) implies that for A and B sufficiently near the identity

$$\exp \circ \widetilde{\phi} \circ \log(AB) = \left[\exp \circ \widetilde{\phi} \circ \log(A)\right] \left[\exp \circ \widetilde{\phi} \circ \log(B)\right].$$

Thus the right sides of (5.15) and (5.16) are equal. Once we know that passing to a refinement doesn't change the answer, we have independence of partition. For any two partitions of $[0, 1]$ have a common refinement, namely, the union of the two.

Once we know independence of partition, we need to prove independence of path. It is at this point that we use the fact that G is simply connected. In particular, because of simple connectedness, any two paths $A_1(t)$ and $A_2(t)$ joining the identity to A will be homotopic with endpoints fixed. (This is a standard topological fact.) Using this, we want to prove that Step 3 gives the same answer for A_1 and A_2.

Our strategy is to deform A_1 into A_2 in a series of steps, where during each step we only change the path in a small time interval $(t, t+\epsilon)$, keeping everything fixed on $[0,t]$ and on $[t+\epsilon, 1]$. Since we have independence of partition, we can take t and $t+\epsilon$ to be partion points. Since the time interval is small, we can assume there are no partition points between t and $t+\epsilon$. Then we have

$$\phi\left(A(t+\epsilon)\right) = \phi\left(A(t+\epsilon)A(t)^{-1}\right)\phi\left(A(t)\right)$$

where $\phi\left(A(t+\epsilon)A(t)^{-1}\right)$ is defined as in Step 2.

But notice that our value for $\phi\left(A(t+\epsilon)\right)$ depends only on $A(t)$ and $A(t+\epsilon)$, not on how we get from $A(t)$ to $A(t+\epsilon)$! Thus the value $\phi\left(A(t+\epsilon)\right)$ doesn't change as we deform the path. But if $\phi\left(A(t+\epsilon)\right)$ doesn't change as we deform the path, neither does $\phi\left(A(1)\right)$, since the path isn't changing on $[t+\epsilon, 1]$.

Since A_1 and A_2 are homotopic with endpoints fixed, it is possible (by a standard topological argument) to deform A_1 into A_2 in a series of small steps as above.

Step 5: Prove that ϕ is a homomorphism, and is properly related to $\widetilde{\phi}$.

Now that we have independence of path (and partition), we can give a simpler description of how to compute ϕ. Given any group element A, A can be written in the form

$$A = C_n C_{n-1} \cdots C_1$$

with each C_i in V. (This follows from the (path-)connectedness of G.) We can then choose a path $A(t)$ which starts at the identity, then goes to C_1, then to C_2C_1, and so on to $C_nC_{n-1}\cdots C_1 = A$. We can choose a partition so that $A(t_i) = C_iC_{i-1}\cdots C_1$. By the way we have defined things

$$\phi(A) = \phi\left(A(1)A(t_{n-1})^{-1}\right)\phi\left(A(t_{n-1})A(t_{n-2})^{-1}\right)\cdots\phi\left(A(t_1)A(0)\right).$$

But

$$A(t_i)A(t_{i-1})^{-1} = \left(C_iC_{i-1}\cdots C_1\right)\left(C_{i-1}\cdots C_1\right)^{-1} = C_i$$

so

$$\phi(A) = \phi(C_n)\phi(C_{n-1})\cdots\phi(C_1).$$

Now suppose that A and B are two elements of G and we wish to compute $\phi(AB)$. Well, write

$$\begin{aligned} A &= C_nC_{n-1}\cdots C_1 \\ B &= D_nD_{n-1}\cdots D_1. \end{aligned}$$

Then

$$\begin{aligned}\phi(AB) &= \phi(C_n C_{n-1} \cdots C_1 D_n D_{n-1} \cdots D_1) \\ &= [\phi(C_n) \cdots \phi(C_1)]\,[\phi(D_n) \cdots \phi(D_1)] \\ &= \phi(A)\phi(B).\end{aligned}$$

We see then that ϕ is a homomorphism. It remains only to verify that ϕ has the proper relationship to $\widetilde{\phi}$. But since ϕ is defined near the identity to be $\phi = \exp \circ \widetilde{\phi} \circ \log$, we see that

$$\left.\frac{d}{dt}\right|_{t=0} \phi\left(e^{tX}\right) = \left.\frac{d}{dt}\right|_{t=0} e^{t\widetilde{\phi}(X)} = \widetilde{\phi}(X).$$

Thus $\widetilde{\phi}$ is the Lie algebra homomorphism associated to the Lie group homomorphism ϕ.

This completes the proof of Theorem 105. □

5.9 Covering Groups

It is at this point that we pay the price for our decision to consider only *matrix* Lie groups. For the universal covering group of a matrix Lie group (defined below) is always a Lie group, but *not* always a matrix Lie group. For example, the universal covering group of $\mathsf{SL}(n;\mathbb{R})$ $(n \geq 2)$ is a Lie group, but not a matrix Lie group. (See Exercise 20.)

The notion of a universal cover allows us to determine, in the case of a non-simply connected group, which representations of the Lie algebra correspond to representations of the group. See Theorem 113 below.

Definition 108 *Let G be a connected matrix Lie group. A* ***universal covering group*** *of G (or just* ***universal cover****) is a connected, simply connected Lie group $\widetilde{G}$, together with a Lie group homomorphism $\phi : \widetilde{G} \to G$ (called the* ***projection map****) with the following properties:*

1. *ϕ maps $\widetilde{G}$ onto G.*
2. *There is a neighborhood U of I in $\widetilde{G}$ which maps homeomorphically under ϕ onto a neighborhood V of I in G.*

Proposition 109 *If G is any connected matrix Lie group, then a universal covering group $\widetilde{G}$ of G exists and is unique up to canonical isomorphism.*

We will not prove this theorem, but the idea of proof is as follows. We assume that G is a matrix Lie group, hence a Lie group (that is, a manifold). As a manifold, G has a topological universal cover $\widetilde{G}$ which is a connected, simply connected manifold. The universal cover comes with a "projection map" $\phi : \widetilde{G} \to G$ which is a local homeomorphism. Now, since G is not only a manifold but also a group, $\widetilde{G}$ also becomes a group, and the projection map ϕ becomes a homomorphism.

Proposition 110 *Let G be a connected matrix Lie group, $\widetilde{G}$ its universal cover, and ϕ the projection map from $\widetilde{G}$ to G. Suppose that $\widetilde{G}$ is a matrix Lie group with Lie algebra $\widetilde{\mathfrak{g}}$. Then the associated Lie algebra map*

$$\widetilde{\phi} : \widetilde{\mathfrak{g}} \to \mathfrak{g}$$

is an isomorphism.

Remark

In light of this Proposition, we often say that G and $\widetilde{G}$ have the same Lie algebra.

The above Proposition is true even if $\widetilde{G}$ is not a matrix Lie group. But to make sense out of the Proposition in that case, we need the definition of the Lie algebra of a general Lie group, which we have not defined.

Proof

Exercise 18. □

Examples

The universal cover of S^1 is $\mathbb{R}$, and the projection map is the map $x \to e^{ix}$. The universal cover of $\mathsf{SO}(3)$ is $\mathsf{SU}(2)$, and the projection map is the homomorphism described in Lemma 104.

More generally, we can consider $\mathsf{SO}(n)$ for $n \geq 3$. As it turns out, for $n \geq 3$ the universal cover of $\mathsf{SO}(n)$ is a double cover. (That is, the projection map ϕ is two-to-one.) The universal cover of $\mathsf{SO}(n)$ is called $\mathsf{Spin}(n)$, and may be constructed as a certain group of invertible elements in the **Clifford algebra** over $\mathbb{R}^n$. See Bröcker and tom Dieck, Chapter I, Section 6, especially Propositions I.6.17 and I.6.19. In particular, $\mathsf{Spin}(n)$ *is* a matrix Lie group.

The case $n = 4$ is quite special. It turns out that the universal cover of $\mathsf{SO}(4)$ (i.e., $\mathsf{Spin}(4)$) is isomorphic to $\mathsf{SU}(2) \times \mathsf{SU}(2)$. This is best seen by regarding $\mathbb{R}^4$ as the quaternion algebra.

Theorem 111 *Let G be a matrix Lie group, and suppose that $\widetilde{G}$ is also a matrix Lie group. Identify the Lie algebra of $\widetilde{G}$ with the Lie algebra $\mathfrak{g}$ of G as in Proposition 110. Suppose that H is a matrix Lie group with Lie algebra $\mathfrak{h}$, and that $\widetilde{\phi} : \mathfrak{g} \to \mathfrak{h}$ is a homomorphism. Then there exists a unique Lie group homomorphism $\phi : \widetilde{G} \to H$ such that ϕ and $\widetilde{\phi}$ are related as in Theorem 46 of Chapter 3.*

Proof

$\widetilde{G}$ is simply connected. □

Corollary 112 *Let G and $\widetilde{G}$ be as in Theorem 111, and let π be a representation of $\mathfrak{g}$. Then there exists a unique representation $\widetilde{\Pi}$ of $\widetilde{G}$ such that*

$$\pi(X) = \left.\frac{d}{dt}\right|_{t=0} \widetilde{\Pi}\left(e^{tX}\right)$$

for all $X \in \mathfrak{g}$.

Theorem 113 *Let G and $\widetilde{G}$ be as in Theorem 111, and let $\phi : \widetilde{G} \to G$. Now let π be a representation of $\mathfrak{g}$, and $\widetilde{\Pi}$ the associated representation of $\widetilde{G}$, as in the Corollary. Then there exists a representation Π of G corresponding to π if and only if*

$$\ker \widetilde{\Pi} \supset \ker \phi.$$

Proof

Exercise 19. □

5.10 Exercises

1. Let G be a matrix Lie group, and $\mathfrak{g}$ its Lie algebra. Let Π_1 and Π_2 be representations of G, and let π_1 and π_2 be the associated representations of $\mathfrak{g}$ (Proposition 77). Show that if Π_1 and Π_2 are equivalent representations of G, then π_1 and π_2 are equivalent representations of $\mathfrak{g}$. Show that if G is connected, and if π_1 and π_2 are equivalent representations of $\mathfrak{g}$, then Π_1 and Π_2 are equivalent representations of G.

   Hint: Use Corollary 54 of Chapter 3.

2. Let G be a connected matrix Lie group with Lie algebra $\mathfrak{g}$. Let Π be a representation of G acting on a space V, and let π be the associated Lie algebra representation. Show that a subspace $W \subset V$ is invariant for Π if and only if it is invariant for π. Show that Π is irreducible if and only if π is irreducible.

3. Suppose that Π is a finite-dimensional unitary representation of a matrix Lie group G. (That is, V is a finite-dimensional Hilbert space, and Π is a continuous homomorphism of G into $U(V)$.) Let π be the associated representation of the Lie algebra $\mathfrak{g}$. Show that for each $X \in \mathfrak{g}$, $\pi(X)^* = -\pi(X)$.

4. Show explicitly that the adjoint representation and the standard representation are equivalent representations of the Lie algebra $\mathsf{so}(3)$. Show that the adjoint and standard representations of the group $\mathsf{SO}(3)$ are equivalent.

5. Consider the elements E_1, E_2, and E_3 in $\mathsf{su}(2)$ defined in Exercise 9 of Chapter 3. These elements form a basis for the real vector space $\mathsf{su}(2)$. Show directly that E_1, E_2, and E_3 form a basis for the complex vector space $\mathsf{sl}(2;\mathbb{C})$.

6. Define a vector space with basis $u_0, u_1 \cdots u_m$. Now define operators $\pi(H)$, $\pi(X)$, and $\pi(Y)$ by formula (5.10). Verify by direct computation that the operators defined by (5.10) satisfy the commutation relations $[\pi(H), \pi(X)] = 2\pi(X)$, $[\pi(H), \pi(Y)] = -2\pi(Y)$, and $[\pi(X), \pi(Y)] = \pi(H)$. (Thus $\pi(H)$, $\pi(X)$, and $\pi(Y)$ define a representation of $\mathsf{sl}(2;\mathbb{C})$.) Show that this representation is irreducible.

   Hint: It suffices to show, for example, that $[\pi(H), \pi(X)] = 2\pi(X)$ on each basis element. When dealing with $\pi(Y)$, don't forget to treat separately the case of u_k, $k < m$, and the case of u_m.

7. We can define a two-dimensional representation of $\mathsf{so}(3)$ as follows:

$$\pi\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix};$$

$$\pi\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix};$$

$$\pi\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

   (You may assume that this actually gives a representation.) Show that there is no group representation Π of $\mathsf{SO}(3)$ such that Π and π are related as in Proposition 77.

   Hint: If $X \in \mathsf{so}(3)$ is such that $e^X = I$, and Π is any representation of $\mathsf{SO}(3)$, then $\Pi(e^X) = \Pi(I) = I$.

   Remark: In the physics literature, this non-representation of $\mathsf{SO}(3)$ is called "spin $\frac{1}{2}$."

8. Consider the standard representation of the Heisenberg group, acting on $\mathbb{C}^3$. Determine all subspaces of $\mathbb{C}^3$ which are invariant under the action of the Heisenberg group. Is this representation completely reducible?

9. Give an example of a representation of the commutative group $\mathbb{R}$ which is not completely reducible.

10. Consider the unitary representations $\Pi_\hbar$ of the real Heisenberg group. Assume that there is some sort of associated representation $\pi_\hbar$ of the Lie algebra, which should be given by

$$\pi_\hbar(X)f = \left.\frac{d}{dt}\right|_{t=0} \Pi_\hbar\left(e^{tX}\right) f$$

(We have not proved any theorem of this sort for infinite-dimensional unitary representations.)

Computing in a purely formal manner (that is, ignoring all technical issues) compute

$$\pi_\hbar\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \pi_\hbar\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}; \quad \pi_\hbar\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Verify (still formally) that these operators have the right commutation relations to generate a representation of the Lie algebra of the real Heisenberg group. (That is, verify that on this basis, $\pi_\hbar[X,Y] = [\pi_\hbar(X), \pi_\hbar(Y)]$.)

Why is this computation not rigorous?

11. Consider the Heisenberg group over the field $\mathbb{Z}_p$ of integers mod p, with p prime, namely

$$H_p = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} |a, b, c \in \mathbb{Z}_p \right\}.$$

This is a subgroup of the group $\mathsf{GL}(3; \mathbb{Z}_p)$, and has p^3 elements.

Let V_p denote the space of complex-valued functions on $\mathbb{Z}_p$, which is a p-dimensional complex vector space. For each non-zero $n \in \mathbb{Z}_p$, define a representation of H_p by the formula

$$(\Pi_n f)(x) = e^{-i2\pi nb/p} e^{i2\pi ncx/p} f(x-a) \;\; x \in \mathbb{Z}_p.$$

(These representations are analogous to the unitary representations of the real Heisenberg group, with the quantity $2\pi n/p$ playing the role of $\hbar$.)

a) Show that for each n, Π_n is actually a representation of H_p, and that it is irreducible.

b) Determine (up to equivalence) all the one-dimensional representations of H_p.

c) Show that every irreducible representation of H_p is either one-dimensional or equivalent to one of the Π_n's.

12. Prove Theorem 92.

Hints: For existence, choose bases $\{e_i\}$ and $\{f_j\}$ for U and V. Then define a space W which has as a basis $\{w_{ij} \,|0 \le i \le n, 0 \le j \le m\}$. Define $\phi(e_i, f_j) = w_{ij}$ and extend by bilinearity. For uniqueness, use the universal property.

13. Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras, and consider the vector space $\mathfrak{g} \oplus \mathfrak{h}$. Show that the following operation makes $\mathfrak{g} \oplus \mathfrak{h}$ into a Lie algebra

$$[(X_1, Y_1), (X_2, Y_2)] = ([X_1, X_2], [Y_1, Y_2]) .$$

Now let G and H be matrix Lie groups, with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$. Show that $G \times H$ can be regarded as a matrix Lie group in an obvious way, and that the Lie algebra of $G \times H$ is isomorphic to $\mathfrak{g} \oplus \mathfrak{h}$.

14. Suppose that π is a representation of a Lie algebra $\mathfrak{g}$ acting on a finite-dimensional vector space V. Let V^* denote as usual the dual space of V, that is, the space of linear functionals on V. If A is a linear operator on V, let A^{tr} denote the dual or transpose operator on V^* ,

$$\left(A^{tr}\phi\right)(v) = \phi\,(Av)$$

for $\phi \in V^*$, $v \in V$. Define a representation π^* of $\mathfrak{g}$ on V^* by the formula

$$\pi^*(X) = -\pi\left(X^{tr}\right).$$

a) Show that π^* is really a representation of $\mathfrak{g}$.

b) Show that $(\pi^*)^*$ is isomorphic to π.

c) Show that π^* is irreducible if and only if π is.

d) What is the analogous construction of the dual representation for representations of groups?

15. Recall the spaces V_m introduced in Section 5.3, viewed as representations of the Lie algebra $\mathsf{sl}(2;\mathbb{C})$. In particular, consider the space V_1 (which has dimension 2).

a) Regard $V_1 \otimes V_1$ as a representation of $\mathsf{sl}(2;\mathbb{C})$, as in Definition 99. Show that this representation is not irreducible.

b) Now view $V_1 \otimes V_1$ as a representation of $\mathsf{sl}(2;\mathbb{C}) \oplus \mathsf{sl}(2;\mathbb{C})$, as in Definition 96. Show that this representation is irreducible.

c) More generally, show that $V_m \otimes V_n$ is irreducible as a representation of $\mathsf{sl}(2;\mathbb{C}) \oplus \mathsf{sl}(2;\mathbb{C})$, but reducible (except if one of n or m is zero) as a representation of $\mathsf{sl}(2;\mathbb{C})$.

16. Show explicitly that $\exp : \mathsf{so}(3) \to \mathsf{SO}(3)$ is onto.

Hint: Using the fact that $\mathsf{SO}(3) \subset \mathsf{SU}(3)$, show that the eigenvalues of $R \in \mathsf{SO}(3)$ must be of one of the three following forms: $(1, 1, 1)$, $(1, -1, -1)$, or

$(1, e^{i\theta}, e^{-i\theta})$. In particular, R must have an eigenvalue equal to one. Now show that in a suitable orthonormal basis, R is of the form

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}.$$

17. *Proof of Lemma 104.*

    Let $\{E_1, E_2, E_3\}$ be the usual basis for $\mathsf{su}(2)$, and $\{F_1, F_2, F_3\}$ be the basis for $\mathsf{so}(3)$ introduced in Section 5.8. Identify $\mathsf{su}(2)$ with $\mathbb{R}^3$ by identifying the basis $\{E_1, E_2, E_3\}$ with the standard basis for $\mathbb{R}^3$. Consider $\mathrm{ad}E_1$, $\mathrm{ad}E_2$, and $\mathrm{ad}E_3$ as operators on $\mathsf{su}(2)$, hence on $\mathbb{R}^3$. Show that $\mathrm{ad}E_i = F_i$, for $i = 1, 2, 3$. In particular, *ad* is a Lie algebra isomorphism of $\mathsf{su}(2)$ onto $\mathsf{so}(3)$.

    Now consider $\mathrm{Ad} : \mathsf{SU}(2) \to \mathsf{GL}(\mathsf{SU}(2)) = \mathsf{GL}(3; \mathbb{R})$. Show that the image of *Ad* is precisely $\mathsf{SO}(3)$. Show that the kernel of *Ad* is $\{I, -I\}$.

    Show that $\mathrm{Ad} : \mathsf{SU}(2) \to \mathsf{SO}(3)$ is the homomorphism Φ required by Lemma 104.

18. *Proof of Proposition 110.*

    Suppose that G and $\widetilde{G}$ are matrix Lie groups. Suppose that $\phi : \widetilde{G} \to G$ is a Lie group homomorphism such that ϕ maps some neighborhood U of I in $\widetilde{G}$ homeomorphically onto a neighborhood V of I in G. Prove that the associated Lie algebra map $\widetilde{\phi} : \widetilde{\mathfrak{g}} \to \mathfrak{g}$ is an isomorphism.

    Hints: Suppose that $\widetilde{\phi}$ were not one-to-one. Show, then, that there exists a sequence of points A_n in $\widetilde{G}$ with $A_n \neq I$, $A_n \to I$ and $\phi(A_n) = I$, giving a contradiction.

    To show that $\widetilde{\phi}$ is onto, use Step 1 of the proof of Theorem 105 to show that on a sufficiently small neighborhood of zero in $\widetilde{\mathfrak{g}}$,

    $$\widetilde{\phi} = \log \circ \phi \circ \exp .$$

    Use this to show that the image of $\widetilde{\phi}$ contains a neighborhood of zero in $\mathfrak{g}$. Now use linearity to show that the image of $\widetilde{\phi}$ is all of $\mathfrak{g}$.

19. *Proof of Theorem 113.*

    First suppose that $\ker \widetilde{\Pi} \supset \ker \phi$. Then construct Π as in the proof of Proposition 103.

    Now suppose that there is a representation Π of G for which the associated Lie algebra representation is π. We want to show, then, that $\ker \widetilde{\Pi} \supset \ker \phi$. Well, define a new representation Σ of $\widetilde{G}$ by

    $$\Sigma = \Pi \circ \phi.$$

Show that the associated Lie algebra homomorphism σ is equal to π, so that, by Point (1) of Theorem 105, $\widetilde{\Pi} = \Sigma$. What can you say about the kernel of Σ?

20. Fix an integer $n \geq 2$.

    a) Show that every (finite-dimensional complex) representation of the Lie algebra $\mathsf{sl}(n;\mathbb{R})$ gives rise to a representation of the group $\mathsf{SL}(n;\mathbb{R})$, even though $\mathsf{SL}(n;\mathbb{R})$ is not simply connected. (You may use the fact that $\mathsf{SL}(n;\mathbb{C})$ is simply connected.)

    b) Show that the universal cover of $\mathsf{SL}(n;\mathbb{R})$ is not isomorphic to any matrix Lie group. (You may use the fact that $\mathsf{SL}(n;\mathbb{R})$ is not simply connected.)

21. Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$, let $\mathfrak{h}$ be a subalgebra of $\mathfrak{g}$, and let H be the unique connected Lie subgroup of G with Lie algebra $\mathfrak{h}$. Suppose that there exists a compact simply connected matrix Lie group K such that the Lie algebra of K is isomorphic to $\mathfrak{h}$. Show that H is closed. Is H necessarily isomorphic to K?

Chapter 6
THE REPRESENTATIONS OF $\mathsf{SU}(3)$, AND BEYOND

6.1 Preliminaries

There is a theory of the representations of semisimple groups/Lie algebras which includes as a special case the representation theory of $\mathsf{SU}(3)$. However, I feel that it is worthwhile to examine the case of $\mathsf{SU}(3)$ separately. I feel this way partly because $\mathsf{SU}(3)$ is an important group in physics, but chiefly because the general semisimple theory is difficult to digest. Considering a non-trivial example makes it much clearer what is going on. In fact, all of the elements of the general theory are present already in the case of $\mathsf{SU}(3)$, so we do not lose too much by considering at first just this case.

The main result of this chapter is Theorem 1, which states that an irreducible finite-dimensional representation of $\mathsf{SU}(3)$ can be classified in terms of its "highest weight." This is analogous to labeling the irreducible representations V_m of $\mathsf{SU}(2)/\mathsf{sl}(2;\mathbb{C})$ by the highest eigenvalue of $\pi_m(H)$. (The highest eigenvalue of $\pi_m(H)$ in V_m is precisely m.) We will then discuss, without proofs, what the corresponding results are for general semisimple Lie algebras.

The group $\mathsf{SU}(3)$ is connected and simply connected (Bröcker and tom Dieck), so by Corollary 1 of Chapter 5, the finite-dimensional representations of $\mathsf{SU}(3)$ are in one-to-one correspondence with the finite-dimensional representations of the Lie algebra $\mathsf{su}(3)$. Meanwhile, the complex representations of $\mathsf{su}(3)$ are in one-to-one correspondence with the complex-linear representations of the complexified Lie algebra $\mathsf{su}(3)_{\mathbb{C}}$. But $\mathsf{su}(3)_{\mathbb{C}} \cong \mathsf{sl}(3;\mathbb{C})$, as is easily verified. Moreover, since $\mathsf{SU}(3)$ is connected, it follows that a subspace $W \subset V$ is invariant under the action of $\mathsf{SU}(3)$ if and only if it is invariant under the action of $\mathsf{sl}(3;\mathbb{C})$. Thus we have the following:

Proposition 114 *There is a one-to-one correspondence between the finite-dimensional complex representations* Π *of* $\mathsf{SU}(3)$ *and the finite-dimensional complex-linear representations* π *of* $\mathsf{sl}(3;\mathbb{C})$. *This correspondence is determined by the property that*

$$\Pi\left(e^X\right) = e^{\pi(X)}$$

for all $X \in \mathsf{su}(3) \subset \mathsf{sl}(3;\mathbb{C})$.

The representation Π *is irreducible if and only the representation* π *is irreducible. Moreover, a subspace* $W \subset V$ *is invariant for* Π *if and only if it is invariant for* π.

Since $\mathsf{SU}(3)$ is compact, Proposition 90 of Chapter 5 tells us that all the finite-dimensional representations of $\mathsf{SU}(3)$ are completely reducible. The above proposition then implies that all the finite-dimensional representations of $\mathsf{sl}(3;\mathbb{C})$ are completely reducible.

Moreover, we can apply the same reasoning to the group $\mathsf{SU}(2)$, its Lie algebra $\mathsf{su}(2)$, and its complexified Lie algebra $\mathsf{sl}(2;\mathbb{C})$. Since $\mathsf{SU}(2)$ is simply connected, there is a one-to-one correspondence between the complex representations of $\mathsf{SU}(2)$ and the representations of the complexified Lie algebra $\mathsf{sl}(2;\mathbb{C})$. Since $\mathsf{SU}(2)$ is compact, all of the representations of $\mathsf{SU}(2)$–and therefore also of $\mathsf{sl}(2;\mathbb{C})$–are completely reducible. Thus we have establised the following.

Proposition 115 *Every finite-dimensional (complex-linear) representation of* $\mathsf{sl}(2;\mathbb{C})$ *or* $\mathsf{sl}(3;\mathbb{C})$ *is completely reducible. In particular, every finite-dimensional representation of* $\mathsf{sl}(2;\mathbb{C})$ *or* $\mathsf{sl}(3;\mathbb{C})$ *decomposes as a direct sum of irreducible invariant subspaces.*

We will use the following basis for $\mathsf{sl}(3;\mathbb{C})$:

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad H_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad X_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$Y_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad Y_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad Y_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note that the span of $\{H_1, X_1, Y_1\}$ is a subalgebra of $\mathsf{sl}(3;\mathbb{C})$ which is isomorphic to $\mathsf{sl}(2;\mathbb{C})$, by ignoring the third row and the third column. Similary, the span of $\{H_2, X_2, Y_2\}$ is a subalgebra isomorphic to $\mathsf{sl}(2;\mathbb{C})$, by ignoring the first row and first column. Thus we have the following commutation relations

$$\begin{array}{rclrcl} [H_1, X_1] & = & 2X_1 & [H_2, X_2] & = & 2X_2 \\ [H_1, Y_1] & = & -2Y_1 & [H_2, Y_2] & = & -2Y_2 \\ [X_1, Y_1] & = & H_1 & [X_2, Y_2] & = & H_2. \end{array}$$

We now list all of the commutation relations among the basis elements which involve at least one of H_1 and H_2. (This includes some repetitions of the commutation relations above.)

$$\begin{array}{rclcrcl}
[H_1,H_2] & = & 0 & & & & \\
\\
[H_1,X_1] & = & 2X_1 & & [H_1,Y_1] & = & -2Y_1 \\
[H_2,X_1] & = & -X_1 & & [H_2,Y_1] & = & Y_1 \\
\\
[H_1,X_2] & = & -X_2 & & [H_1,Y_2] & = & Y_2 \\
[H_2,X_2] & = & 2X_2 & & [H_2,Y_2] & = & -2Y_2 \\
\\
[H_1,X_3] & = & X_3 & & [H_1,Y_3] & = & -Y_3 \\
[H_2,X_3] & = & X_3 & & [H_2,Y_3] & = & -Y_3
\end{array} \tag{6.1}$$

We now list all of the remaining commutation relations.

$$\begin{array}{rcl}
[X_1,Y_1] & = & H_1 \\
[X_2,Y_2] & = & H_2 \\
[X_3,Y_3] & = & H_1+H_2
\end{array}$$

$$\begin{array}{rclcrcl}
[X_1,X_2] & = & X_3 & & [Y_1,Y_2] & = & -Y_3 \\
[X_1,Y_2] & = & 0 & & [X_2,Y_1] & = & 0 \\
\\
[X_1,X_3] & = & 0 & & [Y_1,Y_3] & = & 0 \\
[X_2,X_3] & = & 0 & & [Y_2,Y_3] & = & 0 \\
\\
[X_2,Y_3] & = & Y_1 & & [X_3,Y_2] & = & X_1 \\
[X_1,Y_3] & = & -Y_2 & & [X_3,Y_1] & = & -X_2
\end{array}$$

Note that there is a kind of symmetry between the X_i's and the Y_i's. If a relation in the first column involves an X_i and/or a Y_j, the corresponding relation in the second column will involve a Y_i and/or an X_j. (E.g., we have the relation $[H_1,X_2]=-X_2$ in the first column, and the relation $[H_2,Y_2]=Y_2$ in the second column.) See Exercise 1.

All of the analysis we will do for the representations of $\mathsf{sl}(3;\mathbb{C})$ will be in terms of the above basis. From now on, all representations of $\mathsf{sl}(3;\mathbb{C})$ will be assumed to be finite-dimensional and complex-linear.

6.2 Weights and Roots

Our basic strategy in classifying the representations of $\mathsf{sl}(3;\mathbb{C})$ is to simultaneously diagonalize $\pi(H_1)$ and $\pi(H_2)$. Since H_1 and H_2 commute, $\pi(H_1)$ and $\pi(H_2)$ will also commute, and so there is at least a chance that $\pi(H_1)$ and $\pi(H_2)$ can be simultaneously diagonalized.

Definition 116 *If (π, V) is a representation of $\mathsf{sl}(3;\mathbb{C})$, then an ordered pair $\mu = (\mu_1, \mu_2) \in \mathbb{C}^2$ is called a* ***weight*** *for π if there exists $v \neq 0$ in V such that*

$$\begin{aligned} \pi(H_1)v &= \mu_1 v \\ \pi(H_2)v &= \mu_2 v. \end{aligned} \tag{6.2}$$

The vector v is called a ***weight vector*** *corresponding to the weight μ. If $\mu = (\mu_1, \mu_2)$ is a weight, then the space of all vectors v satisfying (6.2) is the* ***weight space*** *corresponding to the weight μ.*

Thus a weight is simply a pair of simultaneous eigenvalues for $\pi(H_1)$ and $\pi(H_2)$.

Proposition 117 *Every representation of $\mathsf{sl}(3;\mathbb{C})$ has at least one weight.*

Proof

Since we are working over the complex numbers, $\pi(H_1)$ has at least one eigenvalue μ_1. Let $W \subset V$ be the eigenspace for $\pi(H_1)$ with eigenvalue μ_1. I assert that W is invariant under $\pi(H_2)$. To see this consider $w \in W$, and compute

$$\begin{aligned} \pi(H_1)\left(\pi(H_2)w\right) &= \pi(H_2)\pi(H_1)w \\ &= \pi(H_2)\left(\mu_1 w\right) = \mu_1 \pi(H_2)w. \end{aligned}$$

This shows that $\pi(H_2)w$ is either zero or an eigenvector for $\pi(H_1)$ with eigenvalue μ_1; thus W is invariant.

Thus $\pi(H_2)$ can be viewed as an operator on W. Again, since we are over $\mathbb{C}$, the restriction of $\pi(H_2)$ to W must have at least one eigenvector w with eigenvalue μ_2. But then w is a simultaneous eigenvector for $\pi(H_1)$ and $\pi(H_2)$ with eigenvalues μ_1 and μ_2. □

Now, every representation π of $\mathsf{sl}(3;\mathbb{C})$ can be viewed, by restriction, as a representation of the subalgebra $\{H_1, X_1, Y_1\} \cong \mathsf{sl}(2;\mathbb{C})$. Note that, even if π is irreducible as a representation of $\mathsf{sl}(3;\mathbb{C})$, there is no reason to expect that it will still be irreducible as a representation of the subalgebra $\{H_1, X_1, Y_1\}$. Nevertheless, π restricted to $\{H_1, X_1, Y_1\}$ must be *some* finite-dimensional representation of $\mathsf{sl}(2;\mathbb{C})$. The same reasoning applies to the restriction of π to the subalgebra $\{H_2, X_2, Y_2\}$, which is also isomorphic to $\mathsf{sl}(2;\mathbb{C})$.

Proposition 118 *Let (π, V) be any finite-dimensional complex-linear representation of $\mathsf{sl}(2;\mathbb{C}) = \{H, X, Y\}$. Then all the eigenvalues of $\pi(H)$ are integers.*

Proof

By Proposition 115, V decomposes as a direct sum of irreducible invariant subspaces V_i. Each V_i must be one of the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$, which we have classified. In particular, in each V_i, $\pi(H)$ can be diagonalized, and the eigenvalues of $\pi(H)$ are integers. Thus $\pi(H)$ can be diagonalized on the whole space V, and all of the eigenvalues are integers. □

Corollary 119 *If π is a representation of $\mathsf{sl}(3;\mathbb{C})$, then all of the weights of π are of the form*

$$\mu = (m_1, m_2)$$

with m_1 and m_2 integers.

Proof

Apply Proposition 118 to the restriction of π to $\{H_1, X_1, Y_1\}$, and to the restriction of π to $\{H_2, X_2, Y_2\}$. □

Our strategy now is to begin with one simultaneous eigenvector for $\pi(H_1)$ and $\pi(H_2)$, and then to apply $\pi(X_i)$ or $\pi(Y_i)$, and see what the effect is. The following definition is relevant in this context. (See Lemma 121 below.)

Definition 120 *An ordered pair $\alpha = (\alpha_1, \alpha_2) \in \mathbb{C}^2$ is called a* ***root*** *if*

1. *α_1 and α_2 are not both zero, and*
2. *there exists $Z \in \mathsf{sl}(3;\mathbb{C})$ such that*

$$\begin{aligned}[H_1, Z] &= \alpha_1 Z \\ [H_2, Z] &= \alpha_2 Z.\end{aligned}$$

The element Z is called a ***root vector*** *corresponding to the root α.*

That is, a root is a non-zero weight for the adjoint representation. The commutation relations (6.1) tell us what the roots for $\mathsf{sl}(3;\mathbb{C})$ are. There are six roots.

$$\begin{array}{cc} \alpha & \mathbf{Z} \\ (2,-1) & X_1 \\ (-1,2) & X_2 \\ (1,1) & X_3 \\ (-2,1) & Y_1 \\ (1,-2) & Y_2 \\ (-1,-1) & Y_3 \end{array} \tag{6.3}$$

It is convenient to single out the two roots corresponding to X_1 and X_2 and give them special names:

$$\begin{aligned}\alpha^{(1)} &= (2,-1)\\ \alpha^{(2)} &= (-1,2).\end{aligned} \tag{6.4}$$

The roots $\alpha^{(1)}$ and $\alpha^{(2)}$ are called the **simple roots**. They have the property that all of the roots can be expressed as linear combinations of $\alpha^{(1)}$ and $\alpha^{(2)}$ with *integer* coefficients, and these coefficients are either all greater than or equal to zero or all less than or equal to zero. This is verified by direct computation:

$$\begin{array}{ccc}(2,-1) & = & \alpha^{(1)}\\ (-1,2) & = & \alpha^{(2)}\\ (1,1) & = & \alpha^{(1)}+\alpha^{(2)}\\ (-2,1) & = & -\alpha^{(1)}\\ (1,-2) & = & -\alpha^{(2)}\\ (-1,-1) & = & -\alpha^{(1)}-\alpha^{(2)}.\end{array}$$

The significance of the roots for the representation theory of $\mathsf{sl}(3;\mathbb{C})$ is contained in the following Lemma. Although its proof is very easy, this Lemma plays a crucial role in the classification of the representations of $\mathsf{sl}(3;\mathbb{C})$. Note that this Lemma is the analog of Lemma 83 of Chapter 5, which was the key to the classification of the representations of $\mathsf{sl}(2;\mathbb{C})$.

Lemma 121 *Let $\alpha = (\alpha_1,\alpha_2)$ be a root, and $Z_\alpha \neq 0$ a corresponding root vector in $\mathsf{sl}(3;\mathbf{C})$. Let π be a representation of $\mathsf{sl}(3;\mathbb{C})$, $\mu = (m_1,m_2)$ a weight for π, and $v \neq 0$ a corresponding weight vector. Then*

$$\begin{aligned}\pi(H_1)\pi(Z_\alpha)v &= (m_1+\alpha_1)\pi(Z_\alpha)v\\ \pi(H_2)\pi(Z_\alpha)v &= (m_2+\alpha_2)\pi(Z_\alpha)v.\end{aligned}$$

Thus either $\pi(Z_\alpha)v = 0$ or else $\pi(Z_\alpha)v$ is a new weight vector with weight

$$\mu+\alpha = (m_1+\alpha_1, m_2+\alpha_2).$$

Proof

The definition of a root tells us that we have the commutation relation $[H_1 Z_\alpha] = \alpha_1 Z_\alpha$. Thus

$$\begin{aligned}\pi(H_1)\pi(Z_\alpha)v &= (\pi(Z_\alpha)\pi(H_1) + \alpha_1\pi(Z_a))\,v\\ &= \pi(Z_\alpha)(m_1 v) + \alpha_1\pi(Z_\alpha)v\\ &= (m_1+\alpha_1)\pi(Z_\alpha)v.\end{aligned}$$

A similar argument allows us to compute $\pi(H_2)\pi(Z_\alpha)v$. $\square$

6.3 Highest Weights and the Classification Theorem

We see then that if we have a representation with a weight $\mu = (m_1, m_2)$, then by applying the root vectors $X_1, X_2, X_3, Y_1, Y_2, Y_3$ we can get some new weights of the form $\mu + \alpha$, where α is the root. Of course, some of the weight vectors may simply give zero. In fact, since our representation is finite-dimensional, there can be only finitely many weights, so we must get zero quite often. By analogy to the classification of the representations of $\mathsf{sl}(2;\mathbf{C})$, we would like to single out in each representation a "highest" weight, and then work from there. The following definition gives the "right" notion of highest.

Definition 122 *Let $\alpha^{(1)} = (2,-1)$ and $\alpha^{(2)} = (-1,2)$ be the roots introduced in (6.4). Let μ_1 and μ_2 be two weights. Then μ_1 is **higher** than μ_2 (or equivalently, μ_2 is **lower** than μ_1) if $\mu_1 - \mu_2$ can be written in the form*

$$\mu_1 - \mu_2 = a\alpha^{(1)} + b\alpha^{(2)}$$

with $a \geq 0$ and $b \geq 0$. This relationship is written as $\mu_1 \succeq \mu_2$ or $\mu_2 \preceq \mu_1$.

*If π is a representation of $\mathsf{sl}(3;\mathbb{C})$, then a weight μ_0 for π is said to be a **highest weight** if for all weights μ of π, $\mu \preceq \mu_0$.*

Note that the relation of "higher" is only a *partial* ordering. That is, one can easily have μ_1 and μ_2 such that μ_1 is neither higher nor lower than μ_2. For example, $\alpha^{(1)} - \alpha^{(2)}$ is neither higher nor lower than 0. This in particular means that a finite set of weights need not have a highest element. (E.g., the set $\left\{0, \alpha^{(1)} - \alpha^{(2)}\right\}$ has no highest element.)

We are now ready to state the main theorem regarding the irreducible representations of $\mathsf{sl}(3;\mathbb{C})$.

Theorem 123

1. *Every irreducible representation π of $\mathsf{sl}(3;\mathbb{C})$ is the direct sum of its weight spaces. That is, $\pi(H_1)$ and $\pi(H_2)$ are simultaneously diagonalizable.*

2. *Every irreducible representation of $\mathsf{sl}(3;\mathbb{C})$ has a unique highest weight μ_0, and two equivalent irreducible representations have the same highest weight.*

3. *Two irreducible representations of $\mathsf{sl}(3;\mathbb{C})$ with the same highest weight are equivalent.*

4. *If π is an irreducible representation of $\mathsf{sl}(3;\mathbb{C})$, then the highest weight μ_0 of π is of the form*

$$\mu_0 = (m_1, m_2)$$

with m_1 and m_2 non-negative integers.

5. *Conversely, if m_1 and m_2 are non-negative integers, then there exists a unique irreducible representation π of $\mathsf{sl}(3;\mathbb{C})$ with highest weight $\mu_0 = (m_1, m_2)$.*

Note the parallels between this result and the classification of the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$: In each irreducible representation of $\mathsf{sl}(2;\mathbb{C})$, $\pi(H)$ is diagonalizable, and there is a largest eigenvalue of $\pi(H)$. Two irreducible representations of $\mathsf{sl}(2;\mathbb{C})$ with the same largest eigenvalue are equivalent. The highest eigenvalue is always a non-negative integer, and conversely, for every non-negative integer m, there is an irreducible representation with highest eigenvalue m.

However, note that in the classification of the representations of $\mathsf{sl}(3;\mathbb{C})$ the notion of "highest" does not mean what we might have thought it should mean. For example, the weight $(1,1)$ is higher than the weights $(-1,2)$ and $(2,-1)$. (In fact, $(1,1)$ is the highest weight for the adjoint representation, which is irreducible.)

It is possible to obtain much more information about the irreducible representations besides the highest weight. For example, we have the following formula for the dimension of the representation with highest weight (m_1, m_2).

Theorem 124 *The dimension of the irreducible representation with highest weight (m_1, m_2) is*

$$\frac{1}{2}(m_1+1)(m_2+1)(m_1+m_2+2).$$

We will not prove this formula. It is a consequence of the "Weyl character formula." See Humphreys, Section 24.3. Humphreys refers to $\mathsf{sl}(3;\mathbb{C})$ as A_2.

6.4 Proof of the Classification Theorem

It will take us some time to prove Theorem 1. The proof will consist of a series of Propositions.

Proposition 125 *In every irreducible representation (π, V) of $\mathsf{sl}(3;\mathbb{C})$, $\pi(H_1)$ and $\pi(H_2)$ can be simultaneously diagonalized. That is, V is the direct sum of its weight spaces.*

Proof

Let W be the direct sum of the weight spaces in V. Equivalently, W is the space of all vectors $w \in V$ such that w can be written as a linear combination of simultaneous eigenvectors for $\pi(H_1)$ and $\pi(H_2)$. Since (Proposition 117) π always has at least one weight, $W \neq \{0\}$.

On the other hand, Lemma 121 tells us that if Z_α is a root vector corresponding to the root α, then $\pi(Z_\alpha)$ maps the weight space corresponding to μ into the weight space corresponding to $\mu + \alpha$. Thus W is invariant under the action of all of the root vectors, namely, under the action X_1, X_2, X_3, Y_1, Y_2, and Y_3. Since W is certainly invariant under the action of H_1 and H_2, W is invariant. Thus by irreducibility, $W = V$. $\square$

Definition 126 *A representation* (π, V) *of* $\mathsf{sl}(3;\mathbb{C})$ *is said to be a* ***highest weight cyclic representation with weight*** $\mu_0 = (m_1, m_2)$ *if there exists* $v \neq 0$ *in* V *such that*

1. *v is a weight vector with weight μ_0.*
2. *$\pi(X_1)v = \pi(X_2)v = 0$.*
3. *The smallest invariant subspace of V containing v is all of V.*

The vector v is called a ***cyclic vector*** *for π.*

Proposition 127 *Let* (π, V) *be a highest weight cyclic representation of* $\mathsf{sl}(3;\mathbb{C})$ *with weight* μ_0. *Then*

1. *π has highest weight μ_0.*
2. *The weight space corresponding to the highest weight μ_0 is one-dimensional.*

Proof

Let v be as in the definition. Consider the subspace W of V spanned by elements of the form

$$w = \pi(Y_{i_1})\pi(Y_{i_2})\cdots\pi(Y_{i_n})v \tag{6.5}$$

with each $i_l = 1, 2$, and $n \geq 0$. (If $n = 0$, it is understood that w in (6.5) is equal to v.) I assert that W is invariant. To see this, it suffices to check that W is invariant under each of the basis elements.

By definition, W is invariant under $\pi(Y_1)$ and $\pi(Y_2)$. It is thus also invariant under $\pi(Y_3) = -[\pi(Y_1), \pi(Y_2)]$.

Now, Lemma 121 tells us that applying a root vector $Z_\alpha \in \mathsf{sl}(3;\mathbb{C})$ to a weight vector v with weight μ gives either zero, or else a new weight vector with weight $\mu+\alpha$. Now, by assumption, v is a weight vector with weight μ_0. Furthermore, Y_1 and Y_2 are root vectors with roots $-\alpha^{(1)} = (-2, 1)$ and $-\alpha^{(2)} = (1, -2)$, respectively. (See Equation (6.3).) Thus each application of $\pi(Y_1)$ or $\pi(Y_2)$ subtracts $\alpha^{(1)}$ or $\alpha^{(2)}$ from the weight. In particular, each non-zero element of the form (6.5) is a simultaneous eigenvector for $\pi(H_1)$ and $\pi(H_2)$. Thus W is invariant under $\pi(H_1)$ and $\pi(H_2)$.

To show that W is invariant under $\pi(X_1)$ and $\pi(X_2)$, we argue by induction on n. For $n = 0$, we have $\pi(X_1)v = \pi(X_2)v = 0 \in W$. Now consider applying $\pi(X_1)$

or $\pi(X_2)$ to a vector of the form (6.5). Recall the commutation relations involving an X_1 or X_2 and a Y_1 or Y_2:

$$\begin{array}{rclcrcl} [X_1, Y_1] & = & H_1 & \quad & [X_1, Y_2] & = & 0 \\ [X_2, Y_1] & = & 0 & \quad & [X_2, Y_2] & = & H_2. \end{array}$$

Thus (for i and j equal to 1 or 2) $\pi(X_i)\pi(Y_j) = \pi(Y_j)\pi(X_i) + \pi(H_{ij})$, where H_{ij} is either H_1 or H_2 or zero. Hence (for i equal to 1 or 2)

$$\begin{aligned} &\pi(X_i)\pi(Y_{i_1})\pi(Y_{i_2})\cdots\pi(Y_{i_n})v \\ &= \pi(Y_{i_1})\pi(X_i)\pi(Y_{i_2})\cdots\pi(Y_{i_n})v + \pi(H_{ij})\pi(Y_{i_2})\cdots\pi(Y_{i_n})v. \end{aligned}$$

But $\pi(X_i)\pi(Y_{i_2})\cdots\pi(Y_{i_n})v$ is in W by induction, and $\pi(H_{ij})\pi(Y_{i_2})\cdots\pi(Y_{i_n})v$ is in W since W is invariant under $\pi(H_1)$ and $\pi(H_2)$.

Finally, W is invariant under $\pi(X_3)$ since $\pi(X_3) = [\pi(X_1), \pi(X_2)]$. Thus W is invariant. Since by definition W contains v, we must have $W = V$.

Since Y_1 is a root vector with root $-\alpha^{(1)}$ and Y_2 is a root vector with root $-\alpha^{(2)}$, Lemma 121 tells us that each element of the form (6.5) is either zero or a weight vector with weight $\mu_0 - \alpha^{(i_1)} - \cdots - \alpha^{(i_n)}$. Thus $V = W$ is spanned by v together with weight vectors with weights lower than μ_0. Thus μ_0 is the highest weight for V.

Furthermore,every element of W can be written as a multiple of v plus a linear combination of weight vectors with weights lower than μ_0. Thus the weight space corresponding to μ_0 is spanned by v; that is, the weight space corresponding to μ_0 is one-dimensional. □

Proposition 128 *Every irreducible representation of* $\mathsf{sl}(3;\mathbb{C})$ *is a highest weight cylic representation, with a unique highest weight* μ_0.

Proof

Uniqueness is immediate, since by the previous Proposition, μ_0 is the highest weight, and two distinct weights cannot both be highest.

We have already shown that every irreducible representation is the direct sum of its weight spaces. Since the representation is finite-dimensional, there can be only finitely many weights. It follows that there must exist a weight μ_0 such that there is no weight $\mu \neq \mu_0$ with $\mu \succeq \mu_0$. This says that there is no weight higher than μ_0 (which is *not* the same as saying the μ_0 is highest). But if there is no weight higher than μ_0, then for any non-zero weight vector v with weight μ_0, we must have

$$\pi(X_1)v = \pi(X_2)v = 0.$$

(For otherwise, say, $\pi(X_1)v$ will be a weight vector with weight $\mu_0 + \alpha^{(1)} \succ \mu_0$.)

Since π is assumed irreducible, the smallest invariant subspace containing v must be the whole space; therefore the representation is highest weight cyclic. □

Proposition 129 *Every highest weight cyclic representation of* $\mathsf{sl}(3;\mathbb{C})$ *is irreducible.*

Proof

Let (π, V) be a highest weight cyclic representation with highest weight μ_0 and cyclic vector v. By complete reducibility (Proposition 115), V decomposes as a direct sum of irreducible representations

$$V \cong \bigoplus_i V_i. \tag{6.6}$$

By Proposition 125, each of the V_i's is the direct sum of its weight spaces. Thus since the weight μ_0 occurs in V, it must occur in some V_i. On the other hand, Proposition 127 says that the weight space corresponding to μ_0 is one-dimensional, that is, v is (up to a constant) the *only* vector in V with weight μ_0. Thus V_i must contain v. But then that V_i is an invariant subspace containg v, so $V_i = V$. Thus there is only one term in the sum (6.6), and V is irreducible. □

Proposition 130 *Two irreducible representations of* $\mathsf{sl}(3;\mathbb{C})$ *with the same highest weight are equivalent.*

Proof

We now know that a representation is irreducible if and only if it is highest weight cyclic. Suppose that (π, V) and (σ, W) are two such representations with the same highest weight μ_0. Let v and w be the cyclic vectors for V and W, respectively. Now consider the representation $V \oplus W$, and let U be smallest invariant subspace of $V \oplus W$ which contains the vector (v, w).

By definition, U is a highest weight cyclic representation, therefore irreducible by Proposition. 129. Consider the two "projection" maps $P_1 : V \oplus W \to V$, $P_1(v, w) = v$ and $P_2 : V \oplus W \to W$, $P_1(v, w) = w$. It is easy to check that P_1 and P_2 are morphisms of representations. Therefore the restrictions of P_1 and P_2 to $U \subset V \oplus W$ will also be morphisms.

Clearly neither $P_1|_U$ nor $P_2|_U$ is the zero map (since both are non-zero on (v, w)). Moreover, U, V, and W are all irreducible. Therefore, by Schur's Lemma, $P_1|_U$ is an isomorphism of U with V, and $P_2|_U$ is an isomorphism of U with W. Thus $V \cong U \cong W$. □

Proposition 131 *If* π *is an irreducible representation of* $\mathsf{sl}(3;\mathbb{C})$*, then the highest weight of* π *is of the form*

$$\mu = (m_1, m_2)$$

with m_1 *and* m_2 *non-negative integers.*

Proof

We already know that *all* of the weights of π are of the form (m_1, m_2), with m_1 and m_2 integers. We must show that if $\mu_0 = (m_1, m_2)$ is the highest weight, then m_1 and m_2 are both non-negative. For this, we again use what we know about the representations of $\mathsf{sl}(2;\mathbb{C})$. The following result can be obtained from the proof of the classification of the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$.

> Let (π, V) be any finite-dimensional representation of $\mathsf{sl}(2;\mathbb{C})$. Let v be an eigenvector for $\pi(H)$ with eigenvalue λ. If $\pi(X)v = 0$, then λ is a non-negative integer.

Now, if π is an irreducible representation of $\mathsf{sl}(3;\mathbf{C})$ with highest weight $\mu_0 = (m_1, m_2)$, and if $v \neq 0$ is a weight vector with weight μ_0, then we must have $\pi(X_1)v = \pi(X_2)v = 0$. (Otherwise, μ_0 wouldn't be highest.) Thus applying the above result to the restrictions of π to $\{H_1, X_1, Y_1\}$ and to $\{H_2, X_2, Y_2\}$ shows that m_1 and m_2 must be non-negative. $\square$

Proposition 132 *If m_1 and m_2 are non-negative integers, then there exists an irreducible representation of $\mathsf{sl}(3;\mathbb{C})$ with highest weight $\mu = (m_1, m_2)$.*

Proof

Note that the trivial representation is an irreducible representation with highest weight $(0,0)$. So we need only construct representations with at least one of m_1 and m_2 positive.

First, we construct two irreducible representations with highest weights $(1,0)$ and $(0,1)$. (These are the so-called **fundamental representations**.) The standard representation of $\mathsf{sl}(3;\mathbb{C})$ is an irreducible representation with highest weight $(1,0)$, as is easily checked. To construct an irreducible representation with weight $(0,1)$ we modify the standard representation. Specifically, we define

$$\pi(Z) = -Z^{tr} \tag{6.7}$$

for all $Z \in \mathsf{sl}(3;\mathbb{C})$. Using the fact that $(AB)^{tr} = B^{tr}A^{tr}$, it is easy to check that

$$-[Z_1, Z_2]^{tr} = \left[-Z_1^{tr}, -Z_2^{tr}\right]$$

so that π is really a representation. (This is isomorphic to the dual of the standard representation, as defined in Exercise 14 of Chapter 5.) It is easy to see that π is an irreducible representation with highest weight $(0,1)$.

Let (π_1, V_1) denote $\mathbb{C}^3$ acted on by the standard representation, and let v_1 denote a weight vector corresponding to the highest weight $(1,0)$. (So, $v_1 = (1,0,0)$.) Let (π_2, V_2) denote $\mathbb{C}^3$ acted on by the representation (6.7), and let v_2 denote a weight vector for the highest weight $(0,1)$. (So, $v_2 = (0,0,1)$.) Now consider the representation

$$V_1 \otimes V_1 \cdots \otimes V_1 \otimes V_2 \otimes V_2 \cdots V_2$$

where V_1 occurs m_1 times, and V_2 occurs m_2 times. Note that the action of $\mathsf{sl}(3;\mathbb{C})$ on this space is

$$\begin{aligned} Z \to\ & (\pi_1(Z) \otimes I \cdots \otimes I) \\ & + (I \otimes \pi_1(Z) \otimes I \cdots \otimes I) + \cdots + (I \otimes \cdots I \otimes \pi_2(Z)) \,. \end{aligned} \tag{6.8}$$

Let π_{m_1,m_2} denote this representation.

Consider the vector

$$v_{m_1,m_2} = v_1 \otimes v_1 \cdots \otimes v_1 \otimes v_2 \otimes v_2 \cdots \otimes v_2.$$

Then applying (6.8) shows that

$$\begin{aligned} \pi_{m_1,m_2}(H_1)v_{m_1,m_2} &= m_1 v_{m_1,m_2} \\ \pi_{m_1,m_2}(H_2)v_{m_1,m_2} &= m_2 v_{m_1,m_2} \\ \pi_{m_1,m_2}(X_1)v_{m_1,m_2} &= 0 \\ \pi_{m_1,m_2}(X_2)v_{m_1,m_2} &= 0. \end{aligned} \tag{6.9}$$

Now, the representation π_{m_1,m_2} is *not* irreducible (unless $(m_1,m_2) = (1,0)$ or $(0,1)$). However, if we let W denote the smallest invariant subspace containing the vector v_{m_1,m_2}, then in light of (6.9), W will be highest weight cyclic with highest weight (m_1,m_2). Therefore by Proposition 129, W is irreducible with highest weight (m_1,m_2).

Thus W is the representation we want. $\square$

We have now completed the proof of Theorem 1.

6.5 An Example: Highest Weight $(1,1)$

To obtain the irreducible representation with highest weight $(1,1)$ we are supposed to take the tensor product of the irreducible representations with highest weights $(1,0)$ and $(0,1)$, and then extract a certain invariant subspace. Let us establish some notation for the representations $(1,0)$ and $(0,1)$. In the standard representation, the weight vectors for

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad H_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix};$$

are the standard basis elements for $\mathbb{C}^3$, namely, e_1, e_2, and e_3. The corresponding weights are $(1,0)$, $(-1,1)$, and $(0,-1)$. The highest weight is $(1,0)$.

Recall that

$$Y_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad Y_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Thus

$$\begin{array}{rclrcl} Y_1(e_1) & = & e_2 & Y_2(e_1) & = & 0 \\ Y_1(e_2) & = & 0 & Y_2(e_2) & = & e_3 \\ Y_1(e_3) & = & 0 & Y_2(e_3) & = & 0. \end{array} \tag{6.10}$$

Now, the representation with highest weight $(0,1)$ is the representation $\pi(Z) = -Z^{tr}$, for $Z \in \mathsf{sl}(3;\mathbb{C})$. Let us define

$$\overline{Z} = -Z^{tr}$$

for all $Z \in \mathsf{sl}(3;\mathbb{C})$. Thus $\pi(Z) = \overline{Z}$. Note that

$$\overline{H_1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \overline{H_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The weight vectors are again e_1, e_2, and e_3, with weights $(-1,0)$, $(1,-1)$, and $(0,1)$. The highest weight is $(0,1)$.

Define new basis elements

$$\begin{array}{rcl} f_1 & = & e_3 \\ f_2 & = & -e_2 \\ f_3 & = & e_1. \end{array}$$

Then since

$$\overline{Y_1} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \overline{Y_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix};$$

we have

$$\begin{array}{rclrcl} \overline{Y_1}(f_1) & = & 0 & \overline{Y_2}(f_1) & = & f_2 \\ \overline{Y_1}(f_2) & = & f_3 & \overline{Y_2}(f_2) & = & 0 \\ \overline{Y_1}(f_3) & = & 0 & \overline{Y_2}(f_3) & = & 0. \end{array} \tag{6.11}$$

Note that the highest weight vector is $f_1 = e_3$.

So, to obtain an irreducible representation with highest weight $(1,1)$ we are supposed to take the tensor product of the representations with highest weights $(1,0)$ and $(0,1)$, and then take the smallest invariant subspace containing the vector $e_1 \otimes f_1$. In light of the proof of Proposition 127, this smallest invariant subspace is obtained by starting with $e_1 \otimes f_1$ and applying all possible combinations of Y_1 and Y_2.

Recall that if π_1 and π_2 are two representations of the Lie algebra $\mathsf{sl}(3;\mathbb{C})$, then

$$\begin{aligned}
(\pi_1 \otimes \pi_2)(Y_1) &= \pi_1(Y_1) \otimes I + I \otimes \pi_2(Y_1) \\
(\pi_1 \otimes \pi_2)(Y_2) &= \pi_1(Y_2) \otimes I + I \otimes \pi_2(Y_2).
\end{aligned}$$

In our case we want $\pi_1(Y_i) = Y_i$ and $\pi_2(Y_i) = \overline{Y_i}$. Thus

$$\begin{aligned}
(\pi_1 \otimes \pi_2)(Y_1) &= Y_1 \otimes I + I \otimes \overline{Y_1} \\
(\pi_1 \otimes \pi_2)(Y_2) &= Y_2 \otimes I + I \otimes \overline{Y_2}.
\end{aligned}$$

The actions of Y_i and $\overline{Y_i}$ are described in (6.10) and (6.11).

Note that $\pi_1 \otimes \pi_2$ is *not* an irreducible representation. The representation $\pi_1 \otimes \pi_2$ has dimension 9, whereas the smallest invariant subspace containing $e_1 \otimes f_1$ has, as it turns out, dimension 8.

So, it remains only to begin with $e_1 \otimes f_1$, apply Y_1 and Y_2 repeatedly until we get zero, and then figure out what dependence relations exist among the vectors we get. These computations are done on a supplementary page. Note that the weight $(0,0)$ has multiplicity two. This is because, starting with $e_1 \otimes f_1$, applying Y_1 and then Y_2 gives something different than applying Y_2 and then Y_1.

6.6 The Weyl Group

The set of weights of an arbitrary irreducible representation of $\mathsf{sl}(3;\mathbb{C})$ has a certain symmetry associated to it. This symmetry is in terms of something called the "Weyl group." (My treatment of the Weyl group follows Bröcker and tom Dieck, Chap. IV, 1.3.) We consider the following subgroup of $\mathsf{SU}(3)$:

$$W = \left\{ \begin{array}{ccc}
w_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; & w_1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; & w_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \\
w_3 = -\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}; & w_4 = -\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}; & w_5 = -\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}
\end{array} \right\}.$$

These are simply the matrices which permute the standard basis elements of $\mathbb{C}^3$, with an adjustment of overall sign when necessary to make the determinant equal one.

Now, for any $A \in \mathsf{SU}(3)$, we have the associated map $\mathrm{Ad}A : \mathsf{su}(3) \to \mathsf{su}(3)$, where

$$\mathrm{Ad}A(X) = AXA^{-1}.$$

Now, since each element of $\mathsf{sl}(3;\mathbb{C})$ is of the form $Z = X + iY$ with $X, Y \in \mathsf{su}(3)$, it follows that $\mathsf{sl}(3;\mathbb{C})$ is invariant under the map $Z \to AZA^{-1}$. That is, we can think of $\mathrm{Ad}A$ as a map of $\mathsf{sl}(3;\mathbb{C})$ to itself.

The reason for selecting the above group is the following: If $w \in W$, then $\mathrm{Ad}w(H_1)$ and $\mathrm{Ad}w(H_2)$ are linear combinations of H_1 and H_2. That is, each $\mathrm{Ad}w$ preserves the space spanned by H_1 and H_2. (There are other elements of $\mathsf{SU}(3)$ with this property, notably, the diagonal elements. However, these actually commute with H_1 and H_2. Thus the adjoint action of these elements on the span of H_1 and H_2 is trivial and therefore uninteresting. See Exercise 3.)

Now, for each $w \in W$ and each irreducible representation π of $\mathsf{sl}(3;\mathbb{C})$, let's define a new representation π_w by the formula

$$\pi_w(X) = \pi\left(\mathrm{Ad}w^{-1}(X)\right) = \pi(w^{-1}Xw).$$

Since $\mathrm{Ad}w^{-1}$ is a Lie algebra automorphism, π_w will in fact be a representation of $\mathsf{sl}(3;\mathbb{C})$.

Recall that since $\mathsf{SU}(3)$ is simply connected, then for each representation π of $\mathsf{sl}(3;\mathbb{C})$ there is an associated representation Π of $\mathsf{SU}(3)$ (acting on the same space) such that

$$\Pi\left(e^X\right) = e^{\pi(X)}$$

for all $X \in \mathsf{su}(3) \subset \mathsf{sl}(3;\mathbb{C})$. The representation Π has the property that

$$\pi(AXA^{-1}) = \Pi(A)\pi(X)\Pi(A)^{-1} \tag{6.12}$$

for all $X \in \mathsf{su}(3)$. Again since every element of $\mathsf{sl}(3;\mathbb{C})$ is of the form $X + iY$ with $X, Y \in \mathsf{su}(3)$, it follows that (6.12) holds also for $X \in \mathsf{sl}(3;\mathbb{C})$.

In particular, taking $A = w^{-1} \in W$ we have

$$\pi_w(X) = \pi(w^{-1}Xw) = \Pi(w)^{-1}\pi(X)\Pi(w) \tag{6.13}$$

for all $X \in \mathsf{sl}(3;\mathbb{C})$.

Proposition 133 *For each representation π of $\mathsf{sl}(3;\mathbb{C})$ and for each $w \in W$, the representation π_w is equivalent to the representation π.*

Proof

We need a map $\phi : V \to V$ with the property that

$$\phi\left(\pi_w(X)v\right) = \pi(X)\phi(v)$$

for all $v \in V$. This is the same as saying that $\phi\pi_w(X) = \pi(X)\phi$, or equivalently that $\pi_w(X) = \phi^{-1}\pi(X)\phi$. But in light of (6.13), we can take $\phi = \Pi(w)$. $\square$

Remark

Although π and π_w are equivalent, they are not equal. That is, in general $\pi(X) \neq \pi_w(X)$. You should think of π and π_w as differing by a change of basis on V, where the change-of-basis matrix is $\Pi(w)$. Two representations that differ just by a change of basis are automatically equivalent.

Corollary 134 *Let π be a representation of $\mathsf{sl}(3;\mathbb{C})$ and $w \in W$. Then a pair $\mu = (m_1, m_2)$ is a weight for π if and only if it is a weight for π_w. The multiplicity of μ as a weight of π is the same as the multiplicity of μ as a weight for π_w.*

Proof

Equivalent representations must have the same weights and the same multiplicities. □

Let us now compute explicitly the action of $\mathrm{Ad}w^{-1}$ on the span of H_1 and H_2, for each $w \in W$. This is a straightforward computation.

$$\begin{array}{rclcrcl}
w_0^{-1}H_1w_0 & = & H_1 & & w_3^{-1}H_1w_3 & = & -H_1 \\
w_0^{-1}H_2w_0 & = & H_2 & & w_3^{-1}H_2w_3 & = & H_1+H_2 \\
 & & & & & & \\
w_1^{-1}H_1w_1 & = & -H_1-H_2 & & w_4^{-1}H_1w_4 & = & -H_2 \\
w_1^{-1}H_2w_1 & = & H_1 & & w_4^{-1}H_2w_4 & = & -H_1 \\
 & & & & & & \\
w_2^{-1}H_1w_2 & = & H_2 & & w_5^{-1}H_1w_5 & = & H_1+H_2 \\
w_2^{-1}H_2w_2 & = & -H_1-H_2 & & w_5^{-1}H_2w_5 & = & -H_2.
\end{array} \tag{6.14}$$

We can now see the significance of the Weyl group. Let π be a representation of $\mathsf{sl}(3;\mathbb{C})$, $\mu = (m_1, m_2)$ a weight, and $v \neq 0$ a weight vector with weight μ. Then, for example,

$$\begin{aligned}
\pi_{w_1}(H_1)v &= \pi(w_1^{-1}H_1w_1)v = \pi(-H_1 - H_2)v = (-m_1 - m_2)v \\
\pi_{w_1}(H_2)v &= \pi(w_1^{-1}H_2w_1)v = \pi(H_1)v = m_1v.
\end{aligned}$$

Thus v is a weight vector for π_w with weight $(-m_1 - m_2, m_1)$. But by Corollary 134, the weights of π and of π_w are the same!

> **Conclusion**: If $\mu = (m_1, m_2)$ is a weight for π, so is $(-m_1 - m_2, m_1)$. The multiplicities of (m_1, m_2) and $(-m_1 - m_2, m_1)$ are the same.

Of course, a similar argument applies to each of the other elements of the Weyl group. Specifically, if μ is a weight for some representation π, and w is an element of W, then there will be some new weight which must also be a weight of π. We will denote this new weight $w \cdot \mu$. For example, if $\mu = (m_1, m_2)$, then

$w_1 \cdot \mu = (-m_1 - m_2, m_1)$. (We define $w \cdot \mu$ so that if v is a weight vector for π with weight μ, then v will be a weight for π_w with weight $w \cdot \mu$.) From (6.14) we can read off what $w \cdot \mu$ is for each w.

$$\begin{array}{rclrcl} w_0 \cdot (m_1, m_2) &=& (m_1, m_2) & w_3 \cdot (m_1, m_2) &=& (-m_1, m_1 + m_2) \\ w_1 \cdot (m_1, m_2) &=& (-m_1 - m_2, m_1) & w_4 \cdot (m_1, m_2) &=& (-m_2, -m_1) \\ w_2 \cdot (m_1, m_2) &=& (m_2, -m_1 - m_2) & w_5 \cdot (m_1, m_2) &=& (m_1 + m_2, -m_2) \end{array} \tag{6.15}$$

It is straightforward to check that

$$w_i \cdot (w_j \cdot \mu) = (w_i w_j) \cdot \mu. \tag{6.16}$$

We have now proved the following.

Theorem 135 *If $\mu = (m_1, m_2)$ is a weight and w is an element of the Weyl group, let $w \cdot \mu$ be defined by (6.15). If π is a finite-dimensional representation of $\mathsf{sl}(3; \mathbb{C})$, then μ is a weight for π if and only if $w \cdot \mu$ is a weight for π. The multiplicity of μ is the same as the multiplicity of $w \cdot \mu$.*

If we think of the weights $\mu = (m_1, m_2)$ as sitting inside $\mathbb{R}^2$, then we can think of (6.15) as a finite group of linear transformations of $\mathbb{R}^2$. (The fact that this is a *group* of transformations follows form (6.16).) Since this is a *finite* group of transformations, it is possible to choose an inner product on $\mathbb{R}^2$ such that the action of W is orthogonal. (As in the proof of Proposition 89 in Chapter 5.) In fact, there is (up to a constant) exactly one such inner product. In this inner product, the action (6.15) of the Weyl group is generated by a 120° rotation and a reflection about the y-axis. Equivalently, the Weyl group is the symmetry group of an equilateral triangle centered at the origin with one vertex on the y-axis.

6.7 Complex Semisimple Lie Algebras

This section gives a brief synopsis of the structure theory and representation theory of complex semisimple Lie algebras. The moral of the story is that all such Lie algebras look and feel a lot like $\mathsf{sl}(3; \mathbb{C})$. This section will not contain any proofs.

If $\mathfrak{g}$ is a Lie algebra, a subspace $I \subset \mathfrak{g}$ is said to be an **ideal** if $[X, Y] \in I$ for all $X \in \mathfrak{g}$ and all $Y \in I$. A Lie algebra $\mathfrak{g}$ is a said to be **simple** if $\dim \mathfrak{g} \geq 2$ and $\mathfrak{g}$ has no ideals other than $\{0\}$ and $\mathfrak{g}$. A Lie algebra $\mathfrak{g}$ is said to be **semisimple** if $\mathfrak{g}$ can be written as the direct sum of simple Lie algebras.

In this section we consider semisimple Lie algebras over the complex numbers. Examples of complex semisimple Lie algebras include $\mathsf{sl}(n; \mathbb{C})$, $\mathsf{so}(n; \mathbb{C})$ $(n \geq 3)$, and $\mathsf{sp}(n; \mathbb{C})$. All of these are actually simple, except for $\mathsf{so}(4; \mathbb{C})$ which is isomorphic to $\mathsf{sl}(2; \mathbb{C}) \oplus \mathsf{sl}(2; \mathbb{C})$.

Definition 136 *Let $\mathfrak{g}$ be a complex semisimple Lie algebra. A subspace $\mathfrak{h}$ of $\mathfrak{g}$ is said to be a **Cartán subalgebra** if*

1. *$\mathfrak{h}$ is abelian. That is, $[H_1, H_2] = 0$ for all $H_1, H_2 \in \mathfrak{h}$.*

2. *$\mathfrak{h}$ is maximal abelian. That is, if $X \in \mathfrak{g}$ satisfies $[H, X] = 0$ for all $H \in \mathfrak{h}$, then $X \in \mathfrak{h}$.*

3. *For all $H \in \mathfrak{h}$, $\mathrm{ad}H : \mathfrak{g} \to \mathfrak{g}$ is diagonalizable.*

Since all the H's commute, so do the $\mathrm{ad}H$'s. (I.e., $[\mathrm{ad}H_1, \mathrm{ad}H_2] = \mathrm{ad}\,[H_1, H_2] = 0$.) By assumption, each $\mathrm{ad}H$ is diagonalizable, and they commute, therefore the $\mathrm{ad}H$'s are simultaneously diagonalizable. (Using a standard linear algebra fact.) Let $\mathfrak{h}^*$ denote the dual of $\mathfrak{h}$, namely, the space of linear functionals on $\mathfrak{h}$.

Definition 137 *If $\mathfrak{g}$ is a complex semisimple Lie algebra and $\mathfrak{h}$ a Cartán subalgebra, then an element α of $\mathfrak{h}^*$ is said to be a* ***root*** *(for $\mathfrak{g}$ with respect to $\mathfrak{h}$) if α is non-zero and there exists $Z \neq 0$ in $\mathfrak{g}$ such that*

$$[H, Z] = \alpha(H)Z \tag{6.17}$$

for all $H \in \mathfrak{h}$. (Thus a root is a non-zero set of simultaneous eigenvalues for the $\mathrm{ad}H$'s.)

The vector Z is called a ***root vector*** *corresponding to the root α, and the space of all $Z \in \mathfrak{g}$ satisfying (6.17) is the* ***root space*** *corresponding to α. This space is denoted $\mathfrak{g}^\alpha$.*

The set of all roots will be denoted Δ.

Note that if $\mathfrak{g} = \mathsf{sl}\,(3; \mathbb{C})$, then one Cartán subalgebra is the space spanned by H_1 and H_2. The roots (with respect to this Cartán subalgebra) have been calculated in (6.3).

Theorem 138 *If $\mathfrak{g}$ is a complex semisimple Lie algebra, then a Cartán subalgebra $\mathfrak{h}$ exists. If $\mathfrak{h}_1$ and $\mathfrak{h}_2$ are two Cartán subalgebras, then there is an automorphism of $\mathfrak{g}$ which takes $\mathfrak{h}_1$ to $\mathfrak{h}_2$. In particular, any two Cartán subalgebras have the same dimension.*

From now on, $\mathfrak{g}$ will denote a complex semisimple Lie algebra, and $\mathfrak{h}$ a fixed Cartán subalgebra in $\mathfrak{g}$.

Definition 139 *The* ***rank*** *of a complex semisimple Lie algebra is the dimension of a Cartán subalgebra.*

For example, the rank of $\mathsf{sl}\,(n; \mathbb{C})$ is $n - 1$. One Cartán subalgebra in $\mathsf{sl}\,(n; \mathbb{C})$ is the space of diagonal matrices with trace zero. (Note that in the case $n = 3$ the space of diagonal matrices with trace zero is precisely the span of H_1 and H_2.) Both $\mathsf{so}(2n; \mathbb{C})$ and $\mathsf{so}(2n + 1; \mathbb{C})$ have rank n.

Definition 140 *Let (π, V) be a finite-dimensional, complex-linear representation of $\mathfrak{g}$. Then $\mu \in \mathfrak{h}^*$ is called a* ***weight*** *for π if there exists $v \neq 0$ in V such that*

$$\pi(H)v = \mu(H)v$$

for all $H \in \mathfrak{h}$. The vector v is called a ***weight vector*** *for the weight μ.*

Note that the roots are precisely the non-zero weights for the adjoint representation.

Lemma 141 *Let α be a root and Z a corresponding root vector. Let μ be a weight for a representation π and v a corresponding weight vector. Then either $\pi(Z)v = 0$ or else $\pi(Z)v$ is a weight vector with weight $\mu + \alpha$.*

Proof

Same as for $\mathsf{sl}(3; \mathbf{C})$. $\square$

Definition 142 *A set of roots $\{\alpha_1, \cdots \alpha_l\}$ is called a* ***simple system*** *(or* ***basis****) if*

1. *$\{\alpha_1, \cdots \alpha_l\}$ is a vector space basis for $\mathfrak{h}^*$.*
2. *Every root $\alpha \in \Delta$ can be written in the form*

$$\alpha = n_1\alpha_1 + n_2\alpha_2 + \cdots + n_l\alpha_l$$

with each n_i an integer, and such that the n_i's are either all non-negative or all non-positive.

A root α is said to be ***positive*** *(with respect to the given simple system) if the n_i's are non-negative; otherwise α is* ***negative****.*

If $\mathfrak{g} = \mathsf{sl}(3; \mathbb{C})$ and $\mathfrak{h} = \{H_1, H_2\}$, then one simple system of roots is $\left\{\alpha^{(1)}, \alpha^{(2)}\right\} = \{(2, -1), (-1, 2)\}$ (with the corresponding root vectors being X_1 and X_2). The positive roots are $\{(2, -1), (-1, 2), (1, 1)\}$. The negative roots are $\{(-2, 1), (1, -2), (-1, -1)\}$.

Definition 143 *Let $\{\alpha_1, \cdots \alpha_l\}$ be a simple system of roots and let μ_1 and μ_2 be two weights. Then μ_1 is* ***higher*** *than μ_2 (or μ_2 is* ***lower*** *than μ_1) if $\mu_1 - \mu_2$ can be written as*

$$\mu_1 - \mu_2 = a_1\alpha_1 + a_2\alpha_2 + \cdots + a_l\alpha_l$$

with $a_i \geq 0$. This relation is denoted $\mu_1 \succeq \mu_2$ or $\mu_2 \preceq \mu_1$.

A weight μ_0 for a representation π is ***highest*** *if all the weights μ of π satisfy $\mu \preceq \mu_0$.*

The following deep theorem captures much of the structure theory of semisimple Lie algebras.

Theorem 144 *Let $\mathfrak{g}$ be a complex semisimple Lie algebra, $\mathfrak{h}$ a Cartán subalgebra, and Δ the set of roots. Then*

1. *For each root $\alpha \in \Delta$, the corresponding root space $\mathfrak{g}^{\alpha}$ is one-dimensional.*
2. *If α is a root, then so is $-\alpha$.*
3. *A simple system of roots $\{\alpha_1, \cdots \alpha_l\}$ exists.*

We now need to identify the correct set of weights to be highest weights of irreducible representations.

Theorem 145 *Let $\{\alpha_1, \cdots \alpha_l\}$ denote a simple system of roots, X_i an element of the root space $\mathfrak{g}^{\alpha_i}$ and Y_i an element of the root space $\mathfrak{g}^{-\alpha_i}$. Define*

$$H_i = [X_i, Y_i].$$

Then it is possible to choose X_i and Y_i such that

1. *Each H_i is non-zero and contained in $\mathfrak{h}$.*
2. *The span of $\{H_i, X_i, Y_i\}$ is a subalgebra of $\mathfrak{g}$ isomorphic (in the obvious way) to $\mathsf{sl}(2;\mathbb{C})$.*
3. *The set $\{H_1, \cdots H_l\}$ is a basis for $\mathfrak{h}$.*

Note that (in most cases) the set of all H_i's, X_i's, and Y_i's $(i = 1, 2, \cdots l)$ do *not* span $\mathfrak{g}$. In the case $\mathfrak{g} = \mathsf{sl}(3;\mathbb{C})$, $l = 2$, and the span of $H_1, X_1, Y_1, H_2, X_2, Y_2$ represents only six of the eight dimensions of $\mathsf{sl}(3;\mathbb{C})$. Nevertheless the subalgebras $\{H_i, X_i, Y_i\}$ play an important role.

We are now ready to state the main theorem.

Theorem 146 *Let $\mathfrak{g}$ be a complex semisimple Lie algebra, $\mathfrak{h}$ a Cartán subalgebra, and $\{\alpha_1, \cdots \alpha_l\}$ a simple system of roots. Let $\{H_1, \cdots H_l\}$ be as in Theorem 145. Then*

1. *In each irreducible representation π of $\mathfrak{g}$, the $\pi(H)$'s are simultaneously diagonalizable.*
2. *Each irreducible representation of $\mathfrak{g}$ has a unique highest weight.*
3. *Two irreducible representations of $\mathfrak{g}$ with the same highest weight are equivalent.*

4. *If μ_0 is the highest weight of an irreducible representation of $\mathfrak{g}$, then for $i = 1, 2, \cdots l$, $\mu_0(H_i)$ is a non-negative integer.*

5. *Conversely, if $\mu_0 \in \mathfrak{h}^*$ is such that $\mu_0(H_i)$ is a non-negative integer for all $i = 1, 2, \cdots l$, then there is an irreducible representation of $\mathfrak{g}$ with highest weight μ_0.*

The weights μ_0 as in 4) and 5) are called **dominant integral weights**.

6.8 Exercises

1. Show that for any pair of $n \times n$ matrices X and Y,

$$\left[X^{tr}, Y^{tr}\right] = -\left[X, Y\right]^{tr}.$$

Using this fact and the fact that $X_i^{tr} = Y_i$ for $i = 1, 2, 3$, explain the symmetry between X's and Y's in the commutation relations for $\mathsf{sl}\,(3; \mathbb{C})$. For example, show that the relation $[Y_1, Y_2] = -Y_3$ can be obtained from the relation $[X_1, X_2] = X_3$ by taking transposes. Show that the relation $[H_1, Y_2] = Y_2$ follows from the relation $[H_1, X_2] = -X_2$.

2. Recall the definition of the dual π^* of a representation π from Exercise 14 of Chapter 5. Consider this for the case of representations of $\mathsf{sl}\,(3; \mathbb{C})$.

a) Show that the weights of π^* are the negatives of the weights of π.

b) Show that if π is the irreducible representation of $\mathsf{sl}\,(3; \mathbb{C})$ with highest weight (m_1, m_2) then π^* is the irreducible representation with highest weight (m_2, m_1).

Hint: If you identify V and V^* by choosing a basis for V, then A^{tr} is just the usual matrix transpose.

3. Let $\mathfrak{h}$ denote the subspace of $\mathsf{sl}\,(3; \mathbb{C})$ spanned by H_1 and H_2. Let G denote the group of all matrices $A \in \mathsf{SU}(3)$ such that $\mathrm{Ad}A$ preserves $\mathfrak{h}$. Now let G_0 denote the group of all matrices $A \in \mathsf{SU}(3)$ such that $\mathrm{Ad}A$ is the identity on $\mathfrak{h}$, i.e., such that $\mathrm{Ad}A(H_1) = H_1$ and $\mathrm{Ad}A(H_2) = H_2$. Show that G_0 is a normal subgroup of G. Compute G and G_0. Show that G/G_0 is isomorphic to the Weyl group W.

4. a) Verify Theorems 144 and 145 explicitly for the case $\mathfrak{g} = \mathsf{sl}\,(n; \mathbb{C})$.

b) Consider the task of trying to prove Theorem 146 for the case of $\mathsf{sl}\,(n; \mathbb{C})$. Now that you have done (a), what part of the proof goes through the same way as for $\mathsf{sl}\,(3; \mathbb{C})$? At what points in the proof of the corresponding theorem for $\mathsf{sl}\,(3; \mathbb{C})$ did we use special properties of $\mathsf{sl}\,(3; \mathbb{C})$?

Hint: Most of it is the same, but there is one critical point which we do something which does not generalize to $\mathsf{sl}\,(n; \mathbb{C})$.

Chapter 7
FINAL EXAMINATION

1. Let G be a connected matrix Lie group, and let $\mathrm{Ad} : G \to \mathsf{GL}(\mathfrak{g})$ be the adjoint representation of G. Show that

$$\ker(\mathrm{Ad}) = Z(G)$$

where $Z(G)$ denotes the center of G. If $G = \mathsf{O}(2)$, compute $\ker(\mathrm{Ad})$ and $Z(G)$ and show that they are not equal.

Hint: You should use the fact that if G is connected, then every $A \in G$ can be written in the form $A = e^{X_1}e^{X_2}\cdots e^{X_n}$, with $X_i \in \mathfrak{g}$.

2. Let G be a finite, commutative group. Show that the number of equivalence classes of irreducible complex representations of G is equal to the number of elements in G.

Hint: Use the fact that every finite, commutative group is a product of cyclic groups.

3. a) Show that if $R \in \mathsf{O}(2)$, and $\det R = -1$, then R has two real, orthogonal eigenvectors with eigenvalues 1 and -1.

b) Let R be in $\mathsf{O}(n)$. Show that there exists a subspace W of $\mathbb{R}^n$ which is invariant under both R and R^{-1}, and such that $\dim W = 1$ or 2. Show that $W^\perp$ (the orthogonal complement of W) is also invariant under R and R^{-1}. Show that the restrictions of R and R^{-1} to W and to $W^\perp$ are orthogonal. (That is, show that these restrictions preserve inner products.)

c) Let R be in $\mathsf{O}(n)$. Show that $\mathbb{R}^n$ can be written as the orthogonal direct sum of subspaces W_i such that

(a) 1) Each W_i is invariant under R and R^{-1},

(b) 2) Each W_i has dimension 1 or 2, and

(c) 3) If $\dim W_i = 2$, then the restriction of R to W_i has determinant one.

d) Show that the exponential mapping for $\mathsf{SO}(n)$ is onto. Make sure you use the fact that the elements of $\mathsf{SO}(n)$ have determinant one.

Note: This provides an alternative proof that the group $\mathsf{SO}(n)$ is connected.

4. Determine, up to equivalence, all of the finite-dimensional, irreducible (complex-linear) representations of the Lie algebra $\mathsf{sl}(2;\mathbb{C}) \oplus \mathsf{sl}(2;\mathbb{C})$. Can your answer be expressed in terms of a sort of "highest weight"?

   Hint: Imitate the proof of the classification of the irreducible representations of $\mathsf{sl}(2;\mathbb{C})$.

5. Consider the irreducible representation (π, V) of $\mathsf{sl}\,(3;\mathbb{C})$ with highest weight $(0,2)$. Following the procedure in Chapter 6, Section 5, determine

   1) The dimension of V.

   2) All of the weights of π.

   3) The multiplicity of each of the weights. (That is, the dimension of the corresponding weight spaces.)

Chapter 8
BIBLIOGRAPHY

1. **Theodor Bröcker and Tammo tom Dieck**, *Representations of Compact Lie Groups.* Springer-Verlag, 1985.

   A good reference for basic facts on compact groups and their representations, including characters and orthogonality relations. Analyzes representations from a more analytic and less algebraic viewpoint than other authors.

2. **William Fulton and Joe Harris**, *Representation theory. A First Course.* Graduate Texts in Mathematics, 129. Readings in Mathematics, Springer-Verlag, 1991. Has lots of examples. Written from an algebraic point of view.

3. **Sigurdur Helgason**, *Differential Geometry, Lie Groups, and Symmetric Spaces.* Academic Press, 1978.

   A good reference for a lot of things. Includes structure theory of semisimple groups.

4. **James E. Humphreys**, *Introduction to Lie Algebras and Representation Theory.* Springer-Verlag, 1972.

   A standard reference for the Lie algebra side of things (no Lie groups).

5. **N. Jacobson**, *Lie Algebras.* Interscience Tracts No. 10, John Wiley and Sons, 1962.

   Another good reference for Lie algebras.

6. **Anthony W. Knapp**, *Lie groups: beyond an introduction.* Birkhauser, 1996. Good complement to Helgason on such matters as structure theory of Lie groups. As title suggests, not the place to start, but a good reference.

7. **W. Miller**, *Symmetry Groups and Their Applications.* Academic Press.

   Oriented toward applications to physics. Includes theory of finite groups.

8. **Jean-Pierre Serre**, *Complex Semisimple Lie Algebras.* Springer-Verlag, 1987.

   A very concise summary of structure theory and representation theory of semisimple Lie algebras.

9. **Jean-Pierre Serre**, *Linear Representations of Finite Groups.* Springer-Verlag.

An introduction to both complex and modular representations of finite groups.

10. **Barry Simon**, *Representations of finite and compact Lie groups*, American Mathematical Society, 1996. Covers much of the same material as Bröcker and tom Dieck, but from a more analytical perspective.

11. **Frank W. Warner**, *Foundations of Differentiable Manifolds and Lie Groups.* Springer-Verlag, 1983.

Key word in the title is *foundations.* Gives a modern treatment of differentiable manifolds, and then proves some important, non-trivial theorems about Lie groups, including the relationship between subgroups and subalgebras, and the relationship between representations of the Lie algebra and of the Lie group.

12. **V.S. Varadarajan**, *Lie Groups, Lie Algebras, and Their Representations.* Springer-Verlag, 1974.

A comprehensive treatment of both Lie groups and Lie algebras.