Introduction to Lie groups, Lie algebras and their representations
Lecture Notes 2017*

Lecturer:
Prof. Alexander Gorodnik

Material marked with a (*) is nonexaminable.

Contents

*Lecture Notes by Jonathan Robbins

1 Matrix Lie Groups

1.1 Some notation and terminology

Let $v, w \in \mathbb{C}^n$. The hermitian inner product of v and w, denoted $\langle v, w \rangle$, is given by

$$\langle v, w \rangle = \sum_{j=1}^{n} v_j^* w_j.$$

Recall that

$$\langle v, \lambda w \rangle = \lambda \langle v, w \rangle, \quad \langle \lambda v, w \rangle = \lambda^* \langle v, w \rangle, \quad \langle v, w \rangle = \langle w, v \rangle^*,$$

for $\lambda \in \mathbb{C}$. The norm of $v \in \mathbb{C}^n$, denoted $||v||$, is given by

$$||v|| = \langle v, v \rangle^{1/2}$$

Let $\mathbb{C}^{n \times n}$ denote the space of complex $n \times n$ matrices. Let I_n denote the $n \times n$ identity matrix. Given $A \in \mathbb{C}^{n \times n}$, we denote its hermitian conjugate by

$$A^\dagger = A^{*T}, \text{ i.e. } A^\dagger_{jk} = A^*_{kj}.$$

If we identify $\mathbb{C}^{n \times n}$ with $\mathbb{C}^{n^2}$, then the hermitian inner product on $\mathbb{C}^{n \times n}$ may be written as

$$\begin{aligned}\langle A, B \rangle = \sum_{j,k=1}^{n} A^*_{jk} B_{jk} = \sum_{jk=1}^{n} A^\dagger_{kj} B_{jk} = \sum_{k=1}^{n} \left(A^\dagger B\right)_{kk} \\ = \mathrm{Tr}(A^\dagger B).\end{aligned}$$

The norm of $A \in \mathbb{C}^{n \times n}$ is then given by

$$||A|| = \langle A, A \rangle^{1/2}.$$

This norm on $\mathbb{C}^{n \times n}$ is called the Frobenius norm, in contrast to other, different definitions of the norm that you may have seen, e.g. $||A|| = \max_{v \in \mathbb{C}^n, ||v||=1} ||A \cdot v||$, or $||A|| = \max_{j,k} |A_{jk}|$. The Frobenius norm satisfies the usual properties of a norm, e.g. $||\lambda A|| = |\lambda| \, ||A||$ for $\lambda \in \mathbb{C}$ and $||A + B|| \le ||A|| + ||B||$ (triangle inequality) in addition to the following:

$$||A \cdot v|| \le ||A|| \, ||v||,$$

$$||AB|| \le ||A|| \, ||B||,$$

$$||A^\dagger|| = ||A||,$$

for all $A, B \in \mathbb{C}^{n \times n}$ and $v \in \mathbb{C}^n$.

Given $A \in \mathbb{C}^{n \times n}$, we define the open δ-ball around A, denoted $B_\delta(A)$, to be

$$B_\delta(A) = \{B \in \mathbb{C}^{n \times n} \mid ||A - B|| < \delta\}.$$

We say that $W \subset \mathbb{C}^{n \times n}$ is <u>open</u> if $\forall A \in W$, $\exists \delta > 0$ such that $B_\delta(A) \subset W$.

We will also have occasion to consider the subspace of real $n \times n$ matrices, denoted $\mathbb{R}^{n \times n}$.

1.2 Matrix groups.

Definition 1.1 (Matrix group)**.** $G \subset \mathbb{C}^{n \times n}$ is a **matrix group** if G is a group under matrix multiplication, i.e.

$$I_n \in G, \quad A \in G \implies A^{-1} \in G, \quad A, B \in G \implies AB \in G.$$

(Note that matrix multiplication is associative.)

Example 1.2 (Matrix groups)**.**

a) $GL(n, \mathbb{C})$, the general complex linear group, is the set of invertible complex $n \times n$ matrices. That is,

$$GL(n, \mathbb{C}) = \{A \in \mathbb{C}^{n \times n} \mid \det A \neq 0\}.$$

Similarly, $GL(n, \mathbb{R})$ denotes the general real linear group of invertible real $n \times n$ matrices.

b) $SL(n, \mathbb{C})$, the special complex linear group, is the set of invertible complex $n \times n$ matrices with determinant equal to one. That is,

$$SL(n, \mathbb{C}) = \{A \in \mathbb{C}^{n\times n} \mid \det A = 1\}.$$

Similarly, $SL(n, \mathbb{R})$ denotes the special real linear group of invertible real $n \times n$ matrices with determinant equal to one.

c) $G = \{I_n, -I_n\}$ is a matrix group consisting of just two elements.

d) $SO(n)$, the special orthogonal group, is the set of real orthogonal $n \times n$ matrices with determinant equal to one. That is,

$$SO(n) = \{A \in \mathbb{R}^{n\times n} \mid A^T = A^{-1} \text{ and } \det A = 1\}.$$

1.3 Matrix Lie groups.

Informally, a matrix Lie group is a matrix group whose members are smoothly parameterised by some number of real coordinates. The number of coordinates in the parameterisation is the dimension of the matrix Lie group. We proceed to formalise this idea.

Definition 1.3 (Open relative to an enclosing set)**.** Let $W \subset X \subset \mathbb{C}^{n\times n}$. W is **open with respect to X** if for all $A \in W$, there exists $\delta > 0$ such that

$$B_\delta(A) \cap X \subset W.$$

Example 1.4 (Open with respect to enclosing set)**.** The notion of a set being open with respect to an enclosing set is not restricted to subsets of matrices. To illustrate it is easier to take an example from within $\mathbb{R}^3$. Let $X \subset \mathbb{R}^3$ be the xy-plane, i.e.

$$X = \{\mathbf{r} = (x, y, z) \mid z = 0\}.$$

Let $W \subset X$ be the open unit disk about the origin in the xy-plane, i.e.

$$W = \{\mathbf{r} = (x, y, z) \mid z = 0,\ x^2 + y^2 < 1\}.$$

W is not open as a subset of $\mathbb{R}^3$, but it is open with respect to X, that is, as a subset of the xy-plane.

There is a related notion of being closed with respect to an enclosing set.

Definition 1.5 (Closed relative to an enclosing set)**.** Let $W \subset X \subset \mathbb{C}^{n\times n}$. W is **closed with respect to X** if for all sequences $A_m \in W$ with A_m converging to a limit $A \in \mathbb{C}^{n\times n}$, then if $A \in X$, we must have $A \in W$.

Example 1.6 (Closed with respect to enclosing set)**.** The notion of a set being closed with respect to an enclosing set is also not restricted to subsets of matrices, so we will illustrate with an example in $\mathbb{R}^3$. Let $X \subset \mathbb{R}^3$ be the open unit disk about the origin in the xy-plane, i.e.

$$X = \{\mathbf{r} = (x, y, z) \mid z = 0,\ x^2 + y^2 < 1\}.$$

Let W be the open unit interval about the origin along the x-axis, i.e.

$$W = \{\mathbf{r} = (x, y, z) \mid y = z = 0,\ x^2 < 1\}.$$

W is not a closed subset of $\mathbb{R}^3$; the sequence $w_m = (1 - 1/m, 0, 0)$ converges to $(1, 0, 0)$, which does not belong to W. However, if a sequence in W converges to a point that belongs to X, then that point necessarily belongs to W. Hence W is closed with respect to X.

(*) The notion of being open or closed with respect to an enclosing set has a more general formulation in the context of general topological spaces. A topological space is a set S together with a family $\mathcal{F}$ of subsets designated as open and satisfying certain properties: The empty set and S itself should be open, the union of an arbitrary number of open subsets should be an open subset, and the intersection of a finite number of open subsets should be an open set. A subset X of a topological space S may be regarded as a topological space in its own right. The family of open subsets of X is given by the intersection of the open subsets of S with X itself. This topology is called the *induced topology* on X; it is induced by the topology on S.

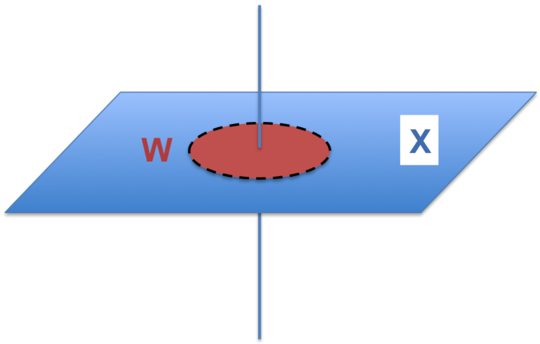


Figure 1: W, the open unit disk about the origin in the xy-plane, is not an open subset of $\mathbb{R}^3$, but it is an open subset of X, xy-plane.

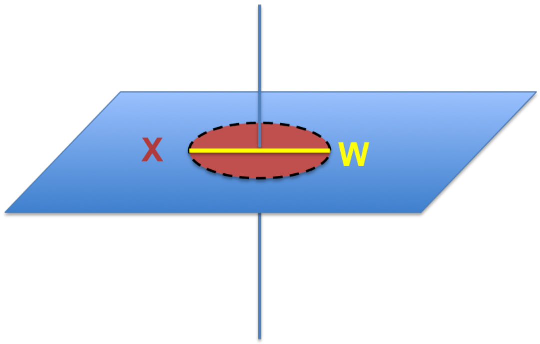


Figure 2: W, the open unit interval about the origin on the x-axis, is not a closed subset of $\mathbb{R}^3$, but it is a closed subset of X, the unit disk about the origin in the xy-plane.

Open and closed with respect to an enclosing set.

Definition 1.7 (Matrix Lie group)**.** A **matrix Lie group of dimension** d is a matrix group $G \subset \mathbb{C}^{n\times n}$ along with the following structure:

- an open subset $P \subset \mathbb{R}^d$ containing 0,
- a subset $V_I \subset G$ containing the identity I_n which is open with respect to G,
- a map

$$\Phi : P \to V_I; x \mapsto \Phi(x) \in G$$

satisfying the following properties:

a) Φ is 1-1 and onto,

b) $\Phi(0) = I_n$,

c) Φ is smooth, i.e. $\mathrm{Re}\,\Phi_{jk}(x)$, $\mathrm{Im}\,\Phi_{jk}(x)$ are smooth functions of x for all $1 \leq j,k \leq n$,

d) the d matrices

$$\xi_\alpha := \frac{\partial \Phi}{\partial x^\alpha}(0), \quad \alpha = 1,\dots,d$$

are linearly independent over $\mathbb{R}$. That is, if $\sum_{\alpha=1}^d c_\alpha \xi_\alpha = 0$ for $c_\alpha \in \mathbb{R}$, then $c_\alpha = 0$ for all $1 \leq \alpha \leq d$.

The set P is called the parameter domain. The map Φ provides a smooth parameterisation of a neighbourhood of the identity open with respect to G by points in the parameter domain. By convention, the identity is parameterised by 0.

Example 1.8 (Matrix Lie groups)**.**

a) $GL(n, C)$ is a matrix Lie group of dimension $2n^2$ (elements depend on n^2 complex parameters, hence $2n^2$ real parameters). We may take the parameter domain to be given by

$$P \subset \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} = \{(X,Y)\,|\,||X||, ||Y|| < \tfrac{1}{4}\}.$$

We take the parameterisation to be given by

$$\Phi(X,Y) = I_n + X + iY.$$

It is clear that $\Phi(X,Y)$ is nonsingular, since

$$||\Phi(X,Y)\cdot v|| = ||(I_n + X + iY)\cdot v|| \geq ||I_n v|| - ||X\cdot v|| - ||Y \cdot v|| \geq (1 - ||X|| - ||Y||)\,||v|| > \tfrac{1}{2}||v||,$$

where we have used properties of the matrix norm including the triangle inequality. Thus, $\Phi(X,Y) \in GL(n,\mathbb{C})$. It is also clear that $\Phi(0) = I_n$ and that Φ is 1-1, i.e. $\Phi(X_1,Y_1) = \Phi(X_2,Y_2)$ implies that $X_1 = X_2$ and $Y_1 = Y_2$. We take the neighbourhood V_I to be given by $\Phi(P)$, i.e. the image of P under Φ, so that $\Phi : P \to V_I$ is automatically onto.

Finally, let us compute

$$\xi_{(jk)} := \frac{\partial \Phi}{\partial X_{jk}}(0), \quad \eta_{jk} := \frac{\partial \Phi}{\partial Y_{jk}}(0)$$

(for this example, the notation $\xi_{(jk)}$ and $\eta_{(jk)}$ is more convenient than the generic notation ξ_α). Let $B(j,k) \in \mathbb{C}^{n\times n}$ denote the $n \times n$ matrix with a single nonzero element, namely the (j,k)th element, which is equal to 1. Thus,

$$\frac{\partial X}{\partial X_{jk}} = \frac{\partial Y}{\partial Y_{jk}} = B(j,k).$$

It follows that

$$\xi_{(jk)} = B(j,k), \quad \eta_{(jk)} = iB(j,k).$$

Clearly, the $B(j,k)$'s are linearly independent of each other. For fixed j,k, the two matrices $B(j,k)$ and $iB(j,k)$ are clearly not linearly independent over $\mathbb{C}$; indeed, if we let $M_1 = B(j,k)$ and $M_2 = iB(j,k)$, then obviously $M_1 + iM_2 = 0$. However, M_1 and M_2 are linearly independent over $\mathbb{R}$; for all real coefficients c_1 and c_2, $c_1 M_1 + c_2 M_2$ vanishes if and only if $c_1 = c_2 = 0$.

b) $G = \{I_n, -I_n\}$ is trivially a $d = 0$-dimensional matrix Lie group. The parameter domain P may be taken to consist of 0 only, and V_I to consist of I_n only.

c) $SO(n)$. One way to demonstrate the existence of a parameterisation for $SO(n)$ as well as other matrix groups characterised by a set of equations among the matrix elements, is via the Implicit Function Theorem. However, this approach does not yield an explicit parameterisation. A nice explicit parameterisation for $SO(n)$ is provided by the Cayley transform. Let

$$\mathbb{R}_-^{n\times n} = \{A \in \mathbb{R}^{n\times n} \mid A^T = -A\}$$

denote the space of antisymmetric matrices. $\mathbb{R}_-^{n\times n}$ is a real vector space of dimension $n(n-1)/2$, and therefore may be identified with $\mathbb{R}^d$ for $d = n(n-1)/2$ (an antisymmetric matrix is determined by its elements above the main diagonal, and there are $n(n-1)/2$ of these). Since the eigenvalues of an antisymmetric matrix are pure imaginary[1], it follows that $(I_n \pm A)$ is invertible.

We define a map Φ as follows:

$$\Phi(A) = (I_n - A)(I_n + A)^{-1}.$$

Φ is called the Cayley transform. Let us show that $\Phi(A) \in SO(n)$. We have that

$$\begin{aligned}\Phi(A)^T\Phi(A) = \left((I_n - A)(I_n + A)^{-1}\right)^T (I_n - A)(I_n + A)^{-1} &= (I_n + A^T)^{-1}(I_n - A^T)(I_n - A)(I_n + A)^{-1} \\ &= (I_n - A)(I_n + A)^{-1}(I_n - A)^{-1}(I_n + A).\end{aligned}$$

Since the matrices $I_n + A$ and $I_n - A$ commute (easily checked), it follows that

$$\Phi(A)^T\Phi(A) = (I_n + A^T)^{-1}(I_n - A)(I_n - A^T)(I_n + A)^{-1} = I_n,$$

so that $\Phi(A)^{-1} = \Phi(A)^T$. Also,

$$\det \Phi(A) = \det\left((I_n - A)(I_n + A)^{-1}\right) = \frac{\det(I_n - A)}{\det(I_n + A)} = 1,$$

since $\det(I_n - A) = \det(I_n - A)^T = \det(I_n + A)$.

The Cayley transform is self-inversive. That is, $\Phi(\Phi(A)) = A$, as we now verify:

$$\Phi(\Phi(A)) = (I_n - \Phi(A))(I_n + \Phi(A))^{-1} = \left(I_n - (I_n - A)(I_n + A)^{-1}\right)\left((I_n + (I_n - A)(I_n + A)^{-1}\right)^{-1}$$

$$= ((I_n + A) - (I_n - A))(I_n + A)^{-1}\left(((I_n + A) + (I_n - A))(I_n + A)^{-1}\right)^{-1} = 2A(I_n + A)^{-1}(I_n + A)I_n/2 = A.$$

Thus, we may let V_I be the subset of $SO(n)$ whose elements R have no eigenvalue equal to -1, i.e. $\det(I_n + R) \neq 0$. Then Φ is a smooth, 1-1 map from $\mathbb{R}_-^{n\times n}$ onto V_I.

[1] This follows from the fact that the eigenvalues of a hermitian matrix are real, and the fact that if A is antisymmetric, then iA is hermitian.

Finally, we should check that the matrices

$$\xi_{(jk)} := \frac{\partial \Phi}{\partial A_{jk}}(0), \quad 1 \le j < k \le n,$$

are linearly independent over $\mathbb{R}$. We note that

$$\frac{\partial A}{\partial A_{jk}} = B(j,k) - B(k,j),$$

where $B(j,k)$ is introduced in a) above. Also, we have the following expression for the derivative of $M^{-1}(t)$ with respect to t in terms of $\dot{M}(t) := dM/dt(t)$:

$$\frac{dM^{-1}}{dt} = -M^{-1}\dot{M}M^{-1},$$

which follows from differentiating the relation $M(t)M^{-1}(t) = I_n$. It follows that

$$\frac{\partial (I+A)^{-1}}{\partial A_{jk}} = -(I+A)^{-1}(B(j,k) - B(k,j))(I+A)^{-1},$$

so that

$$\left.\frac{\partial (I+A)^{-1}}{\partial A_{jk}}\right|_{A=0} = B(k,j) - B(j,k).$$

Therefore,

$$\xi_{jk} = 2(B(k,j) - B(j,k)).$$

It is then clear that

$$\sum_{k=1}^{n}\sum_{j=1}^{k-1} c^{jk}\xi_{jk} = 0 \iff c^{jk} = 0, \text{ for all } 1 \le j < k \le n.$$

It would be awkward always to have to produce a parameterisation of a matrix group in order to establish that it is a matrix Lie group. The following basic result gives an independent topological characterisation of matrix Lie groups.

Theorem 1.9 (Characterising property of matrix Lie groups)**.** Let $G \subset \mathbb{C}^{n\times n}$ be a matrix group. If G is closed with respect to $GL(n,\mathbb{C})$, then G is a matrix Lie group.

(*) Proof. See Problem 2.3 on Problem Sheet 1. □

Note that G need not be closed in $\mathbb{C}^{n\times n}$. For example, $GL(n,\mathbb{C})$ is not closed in $\mathbb{C}^{n\times n}$, since the matrices $A_m = m^{-1}I_n \in GL(n,\mathbb{C})$ converge to $0 \notin GL(n,\mathbb{C})$.

Proposition 1.10 (Converse of Theorem 1.9.)**.** If $G \subset \mathbb{C}^{n\times n}$ is a matrix Lie group, then G is closed with respect to $GL(n,\mathbb{C})$.

(*) Proof. Suppose that $A_m \in G$ converges to A and that $\det A \neq 0$. We must show that $A \in G$.

1) Since G is a matrix Lie group, there exists a smooth 1-1 map Φ from $P \subset \mathbb{R}^d$, an open subset of $\mathbb{R}^d$ containing 0, onto $V_I \subset G$, a subset of G open with respect to $GL(n,\mathbb{C})$ with $I_n \in G$, such that $\Phi(0) = I_n$.

2) Without loss of generality, we may assume that P is bounded. If not, we can define a new parameter domain, $P' = \Psi^{-1}(P)$, where Ψ is the map from the unit ball about the origin in $\mathbb{R}^d$ onto all of $\mathbb{R}^d$ given by

$$\Psi : B_1(0) \to \mathbb{R}^d; \quad x \mapsto \frac{x}{1-||x||}.$$

Ψ is invertible with smooth inverse Ψ^{-1} given by

$$\Psi^{-1} : \mathbb{R}^d \to B_1(0); \quad y \mapsto \frac{y}{1+||y||}.$$

A new parameter map Φ' defined on the bounded domain P' can be taken as $\Phi \circ \Psi$.

3) Since V_I is open with respect to $GL(n,\mathbb{C})$, there exists $\epsilon > 0$ such that $B_{2\epsilon}(I_n) \cap G \subset V_I$. It follows that $\overline{B}_\epsilon(I_n) \cap G$ is contained in V_I, where

$$\overline{B}_\epsilon(I_n) = \left\{ M \in \mathbb{C}^{n\times n} \,|\, ||M - I_n|| \leq \epsilon \right\}.$$

Let $Q = \Phi^{-1}(\overline{B}_\epsilon(I_n)) \subset P$. Since Φ is continuous and $\overline{B}_\epsilon(I_n)$ is closed, it follows that Q is closed. Since P is bounded, so is Q.

4) First, let us suppose that A_m converges to A and that $||A - I_n|| < \epsilon$. Then letting $x_m := \Phi^{-1}(A_m)$, we must have $x_m \in Q$ for m large enough. Since Q is closed and bounded, it is compact (Heine-Borel theorem). Therefore, x_m has a subsequence which converges to $x \in Q \subset P$ (Bolzano-Weierstrass theorem). Since Φ is continuous, it follows that $\Phi(x) = A$. Therefore, $A \in G$, as required.

5) It remains to consider the case where the limit A of the sequence A_m does not belong to $\overline{B}_\epsilon(I_n)$. By assumption, A is invertible. We note that A_m^{-1} converges to A^{-1}. Let $K = ||A^{-1}||$. We may choose M sufficiently large so that i) $||A_M^{-1}|| < 2K$ and ii) for all $m \geq M$, $||A_m - A_M|| < \epsilon/(2K)$.

6) Let

$$\tilde{A}_m = A_M^{-1} A_m.$$

We note that $\tilde{A}_m$ converges to $\tilde{A} := A_M^{-1} A$ We note as well that $\tilde{A}_m \in G$, since $\tilde{A}_m$ is a product of elements of G. Moreover, for $m > M$,

$$||\tilde{A}_m - I_n|| = ||A_M^{-1}(A_m - A_M)|| \leq ||A_M^{-1}||\,||(A_m - A_M)|| \leq 2K\frac{\epsilon}{2K} = \epsilon.$$

By the argument in 4. applied to $\tilde{A}_m$, it follows that $\tilde{A} \in G$. But $\tilde{A} = A_M^{-1} A$, and since $A_M \in G$, it follows that $A \in G$, as required.

□

Example 1.11 (Matrix groups not closed with respect to $GL(n,\mathbb{C})$).

a) The set of complex $n \times n$ matrices with rational elements (that is, the real and imaginary parts of the matrix elements are rational) and nonzero determinant, denoted $GL(n,\mathbb{Q})$, is a group (the inverse of a matrix with rational entries is rational, as is the product of two such matrices). Clearly $GL(n,\mathbb{Q})$ is not closed with respect to $GL(n,\mathbb{C})$; sequences of rational matrices may converge to an invertible matrix with irrational entries.

b) Let

$$G = \left\{ A(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{\pi i t} \end{pmatrix} \Big| \, t \in \mathbb{R} \right\}.$$

It is easy to check that G is matrix group.

We note that

$$J_{+-} := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \notin G, \quad J_{-+} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \notin G, \quad -I_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \notin G,$$

since $\exp(\pi i t) = \pm 1$ implies that t is an integer N, but e^{iN} is not equal to ± 1 for any integer N apart from $N = 0$, in which case $A(t) = A(0) = I_2$.

(*) However, as we now argue, there is a sequence of integers N_m such that e^{iN_m} converges to one of J_{+-}, J_{-+} or $-I_2$. ∓ 1. Indeed, by the Dirichlet Approximation Theorem (see below), we can find an increasing sequence of integers N_m and P_m such that N_m/P_m is an increasingly good approximation to π; specifically,

$$\left| \pi - \frac{N_m}{P_m} \right| < \frac{1}{P_m^2}$$

Then

$$N_m = P_m \pi + r_m,$$

where the remainder r_m satisfies $|r_m| < 1/P_m$.

Without loss of generality, we may assume that N_m and P_m have no common factors. In particular, they cannot both be even. Therefore, the sequence $\{(N_m, P_m)\}$ must contain at least one of the following three types of infinite subsequences: i) N_m odd, P_m even, ii) N_m even, P_m odd, iii) P_m, N_m both odd. In Case i), $e^{i\pi N_m} = -1$, while

$$\lim_{m\to\infty} e^{iN_m} = \lim_{m\to\infty} e^{iP_m\pi} e^{ir_m} = \lim_{m\to\infty} e^{ir_m} = 1,$$

so that

$$\lim_{m\to\infty} A(N_m) = J_{+-}.$$

In Case ii), $e^{i\pi N_m} = 1$, while

$$\lim_{m\to\infty} e^{iN_m} = \lim_{m\to\infty} e^{iP_m\pi} e^{ir_m} = -\lim_{m\to\infty} e^{ir_m} = -1,$$

so that

$$\lim_{m\to\infty} A(N_m) = J_{-+}.$$

In Case iii), $e^{i\pi N_m} = -1$, while

$$\lim_{m\to\infty} e^{iN_m} = \lim_{m\to\infty} e^{iP_m\pi} e^{ir_m} = -\lim_{m\to\infty} e^{ir_m} = -1,$$

so that

$$\lim_{m\to\infty} A(N_m) = -I_2.$$

Thus, at least one of J_{+-}, J_{-+} or $-I_2$ is a limit point of G. Thus, G is not closed in $GL(2, \mathbb{C})$, and therefore is not a matrix Lie group.

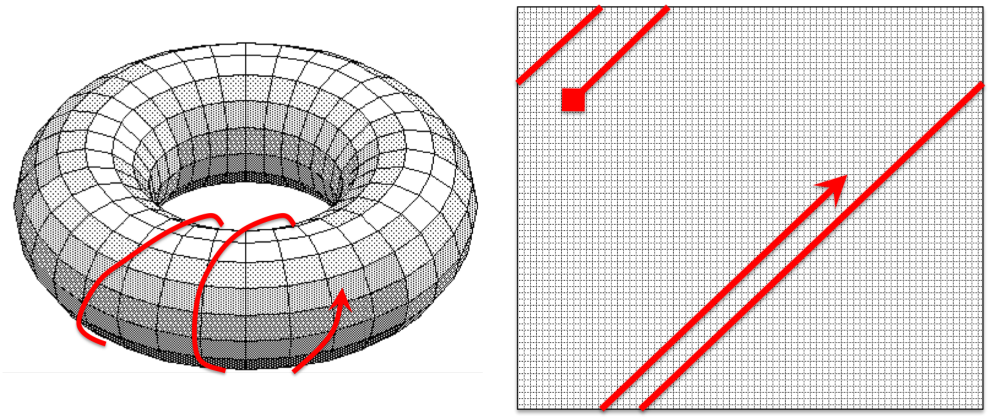


Figure 3: The group G as a subgroup of the group T^2 of 2×2 complex diagonal matrices with diagonal elements $\exp(i\theta_1)$ and $\exp(i\theta_2)$ on the unit circle complex. Elements of T^2 are parameterised by a pair of angles (θ_1, θ_2) defined modulo 2π. Thus, T^2 may be identified with the two-torus (a), or with the square with sides identified (b). G corresponds to elements of the form $(\theta_1, \theta_2) = (t, \pi t)$ for $-\infty < t < \infty$. G defines a line on the torus, part of which is shown above, which is dense (it passes arbitrarily close to every point).

Dirichlet Approximation Theorem (*): We may write

$$m\pi = q_m + r_m,$$

where q_m is a positive integer and r_m, the remainder, satisfies $0 \le r_m < 1$. For all m, there exists at least one pair of positive integers n, n' with $1 \le n' < n \le m+1$ such that $0 < |r_n - r_{n'}| < 1/m$ (this is the pigeon hole principle: if you have $m+1$ numbers distributed between 0 and 1, then at least two of them must be as close as $1/m$). Therefore,

$$0 < |r_n - r_{n'}| = \left|(n - n')\pi - (q_n - q_{n'})\right| < \frac{1}{m}.$$

Letting $P_m = n - n'$ and $N_m = q_n - q_{n'}$, and noting that $P_m \leq m$, we get that

$$0 < \left| \pi - \frac{N_m}{P_m} \right| < \frac{1}{mP_m} \leq \frac{1}{P_m^2}.$$

We note that P_m must go to infinity as m goes to infinity. (There are a finite number of rational numbers with denominator less than D, and as π is irrational, the distance of the closest of these to π will be greater than $1/m$ for $m > D$).

2 The Lie algebra of a matrix Lie group

Let $G \subset \mathbb{C}^{n\times n}$ be a matrix Lie group. By a *smooth curve through the identity*, we mean a nonempty open interval $(-T, T)$ and a smooth map

$$A : (-T, T) \to G; \quad t \mapsto A(t) \in G$$

such that $A(0) = I_n$. Smooth means that $A_{jk}(t)$ is a smooth (infinitely differentiable) function of t for all $1 \leq j, k \leq n$.

Definition 2.1 (Matrix Lie algebra)**.** The **Lie algebra** of a matrix Lie group G, denoted $\mathfrak{g}$, is the subset of $\mathbb{C}^{n\times n}$ given by matrices $\dot{A}(0)$, where $A(t)$ is a smooth curve through identity. That is, $a \in \mathfrak{g}$ if and only if $a = \dot{A}(0)$ for some smooth curve $A(t)$ through the identity.

Proposition 2.2 (Lie algebra as vector space)**.** $\mathfrak{g}$ is a vector space over $\mathbb{R}$.

Proof.

i) Let $a_1, a_2 \in \mathfrak{g}$. We want to show that $a_1 + a_2 \in \mathfrak{g}$. Let $a_1 = \dot{A}_1(0)$ and $a_2 = \dot{A}_2(0)$ for two smooth curves $A_1(t)$ and $A_2(t)$ through the identity. Let $A(t) := A_1(t)A_2(t)$. Then $A(t)$ is a smooth curve through the identity; as a product of smooth curves, it is smooth, and $A(0) = A_1(0)A_2(0) = I_n$. From the product rule (which applies to derivatives of products of matrices).

$$\dot{A}(0) = \dot{A}_1(0)A_2(0) + A_1(0)\dot{A}_2(0) = a_1 + a_2,$$

so $a_1 + a_2 \in \mathfrak{g}$.

ii) Let $a \in \mathfrak{g}$ and $\lambda \in \mathbb{R}$. We want to show that $\lambda a \in \mathfrak{g}$. Let $a = \dot{A}(0)$ for a smooth curve $A(t)$ through the identity. Let $\tilde{A}(t) := A(\lambda t)$. Then $\tilde{A}(t)$ is a smooth curve through the identity, and

$$\dot{\tilde{A}}(0) = \lambda \dot{A}(0) = \lambda a,$$

so that $a \in \mathfrak{g}$.

□

Proposition 2.3 (Characterisation of the Lie algebra in terms of Lie group parameterisation)**.** Let $\mathfrak{g}$ be the Lie algebra of a matrix Lie group $G \subset \mathbb{C}^{n\times n}$. In terms of a parameterisation $\Phi : P \to V_I$ of G, $\mathfrak{g}$ is given by the real span of the matrices ξ_α given by

$$\xi_\alpha := \frac{\partial \Phi}{\partial x^\alpha}(0).$$

That is,

$$\mathfrak{g} = \left\{ \sum_{\alpha=1}^{d} x^\alpha \xi_\alpha \,|\, x^\alpha \in \mathbb{R} \right\}.$$

As a vector space, $\mathfrak{g}$ is d dimensional.

Proof. Let $A(t)$ be a smooth curve through the identity. Without loss of generality, we may assume that the interval of definition, $(-T, T)$, is sufficiently small so that $A(t) \in V_I$ for all $t \in (-T, T)$. Let $x(t) = \Phi^{-1}(A(t))$. Clearly $x(0) = 0$. As shown below, $x(t)$ is smooth. Therefore, since $A(t) = \Phi(x(t))$, it follows from the Chain Rule that

$$a = \dot{A}(0) = \frac{d}{dt}\Phi(x(t))\bigg|_{t=0} = \sum_{\alpha=1}^{d} \dot{x}^\alpha(0)\xi_\alpha.$$

Therefore, a may be expressed as a real linear combination of the ξ_α's, so that

$$\mathfrak{g} \subset \left\{ \sum_{\alpha=1}^{d} x^\alpha \xi_\alpha \,|\, x^\alpha \in \mathbb{R} \right\}.$$

Conversely, ξ_α belongs to $\mathfrak{g}$, since $\xi_\alpha = \dot{A}_\alpha(0)$, where $A_\alpha(t) = \Phi(te_{(\alpha)})$, and $e_{(\alpha)}$ is the unit vector in the α direction. Since $\mathfrak{g}$ is a vector space (Proposition 2.2), it follows that

$$\mathfrak{g} \supset \left\{ \sum_{\alpha=1}^{d} x^\alpha \xi_\alpha \,|\, x^\alpha \in \mathbb{R} \right\}.$$

Thus,

$$\mathfrak{g} = \left\{ \sum_{\alpha=1}^{d} x^\alpha \xi_\alpha \,|\, x^\alpha \in \mathbb{R} \right\},$$

as required. Since the ξ_α's are linearly independent, it follows that $\dim \mathfrak{g} = d$.

(*) Finally, we show that $x(t)$ is smooth. Note that we cannot argue on the basis that Φ^{-1} is smooth, since Φ^{-1} is defined on G, a subset of $\mathbb{C}^{n\times n}$, and we haven't defined what it means for such a function to be smooth (it does make sense to say that Φ^{-1} is continuous, however). We have that

$$\Phi(x(t)) = A(t),$$

so that

$$\Phi(x(t+h)) - \Phi(x(t)) = A(t+h) - A(t).$$

By the Mean Value Theorem for maps,

$$A(t+h) - A(t) = M(t,h)h,$$

where

$$M(t,h) := \int_0^1 \frac{dA}{dt}(t+\tau h)\, d\tau.$$

Clearly $M(t,h)$ is continuous in t and h and approaches $\dot{A}(t)$ as h goes to 0. Similarly,

$$\Phi(x(t+h)) - \Phi(x(t)) = \int_0^1 \frac{d}{d\tau}\Phi(\tau x(t+h) + (1-\tau)x(t))\, d\tau = \sum_{\alpha=1}^{d} (x^\alpha(t+h) - x^\alpha(t))\eta_\alpha(t,h),$$

where

$$\eta_\alpha(t,h) := \int_0^1 \frac{\partial \Phi}{\partial x^\alpha}(\tau x(t+h)) + (1-\tau)x(t))\, d\tau.$$

$\eta_\alpha(t,h)$ is continuous in t and h and approaches $\partial\Phi/\partial x^\alpha(x(t))$ as h approaches 0. Therefore,

$$\sum_{\alpha=1}^{d} (x^\alpha(t+h) - x^\alpha(t))\eta_\alpha(t,h) = M(t,h)h.$$

For t and h sufficiently small, the $\eta_\alpha(t,h)$'s are linearly independent (since they are close to the ξ_α's, which are linearly independent). Therefore, the relation above may be continuously inverted to obtain

$$x^\alpha(t+h) - x^\alpha(t) = \chi^\alpha(t,h)h,$$

where $\chi^\alpha(t,h)$ is continuous and bounded in h and t. It follows that

$$\lim_{h\to 0} \frac{x^\alpha(t+h) - x^\alpha(t)}{h}$$

exists and is continuous in t, so that $x(t)$ is continuously differentiable. The argument may be repeated to show that $x(t)$ is smooth, but in fact only continuous differentiability is needed.

□

Proposition 2.4 (Properties of matrix exponential)**.**

In what follows, $a \in \mathbb{C}^{n\times n}$ and $s, t \in \mathbb{R}$

1)
$$e^0 = I_n$$

2)
$$e^{(s+t)a} = e^{sa}e^{ta}$$

Note that in general, $e^{a+b} \neq e^a e^b$.

3) If $A \in B_{1/2}(I_n)$, there exists a unique $a \in B_{\log 2}(0)$ such that $e^a = A$. Indeed,

$$a = \log A = \log(I - (I - A)) = -\sum_{j=1}^{\infty} \frac{(I - A)}{j}.$$

4)
$$\lim_{m\to\infty} \left(I + \frac{a}{m}\right)^m = e^a.$$

Indeed, we have the following stronger version: If $r_m \in \mathbb{C}^{n\times n}$ and $||r_m|| < k/m^2$ for some $k > 0$, then

$$\lim_{m\to\infty} \left(I + \frac{a}{m} + r_m\right)^m = e^a.$$

5)
$$\det e^a = e^{\operatorname{Tr} a}$$

Proof. 1) and 2) are elementary. 3) – 5) are dealt with in Problems 1.2 – 1.4 respectively. □

Proposition 2.5 (Exponential map)**.** If $\alpha \in \mathfrak{g}$, then $e^{t\alpha} \in G$ for all $t \in \mathbb{R}$.

Proof. By assumption, $\alpha = \dot{A}(0)$, where $A(t)$ is a smooth curve through the identity. Let

$$A_m := A(t/m)^m.$$

Then $A_m \in G$ (since $A(t/m) \in G$ and G is closed under multiplication).

Since $A(t)$ is smooth, it has a second-order Taylor polynomial,

$$A\left(\frac{t}{m}\right) = A(0) + \dot{A}((0)t/m + R_2(t/m) = I + t\alpha/m + R_2(t/m),$$

where, for $|t| < T$, we have that the remainder term satisfies

$$||R_2(t/m)|| \leq C(t/m)^2,$$

for some $C > 0$. It follows that

$$A_m = (I + t\alpha/m + R_2(t/m))^m.$$

By Property 4 of the matrix exponential from Proposition 2.4,

$$\lim_{m\to\infty} A_m = \lim_{m\to\infty} (I + t\alpha/m + R_2(t/m))^m = e^{t\alpha}.$$

Note that, by Property 5 of the matrix exponential from Proposition 2.4,

$$\det e^{t\alpha} = e^{t \operatorname{tr} \alpha} \neq 0.$$

Therefore, $e^{t\alpha} \in GL(n, \mathbb{C})$. Since G is closed in $GL(n, \mathbb{C})$ and A_m converges to $e^{ta} \in GL(n, \mathbb{C})$, it follows that $e^{t\alpha} \in G$, as required. □

Definition 2.6 (Exponential map)**.** The matrix exponential restricted to the Lie algebra $\mathfrak{g}$ of a matrix Lie group $G \subset \mathbb{C}^{n\times n}$ is called the **exponential map**,

$$\exp : \mathfrak{g} \to G; \quad \alpha \mapsto e^{\alpha}.$$

The following shows that the exponential map provides a parameterisation of a neighbourhood of the identity satisfying all requirements in our definition of a matrix Lie group.

Theorem 2.7 (Exponential map as parameterisation)**.** Let G be a d-dimensional matrix Lie group with Lie algebra $\mathfrak{g}$. Let ξ_μ, $1 \le \mu \le d$, be a basis for $\mathfrak{g}$. Then for some $\delta > 0$, the map

$$\widehat{\Phi} : B_\delta(0) \subset \mathbb{R}^d \to V_I; \quad x \mapsto \exp(x^\mu \xi_\mu),$$

where $V_I := \widehat{\Phi}(B_\delta(0))$ is a neighbourhood of the identity, is smooth, 1-1 and onto, with $\widehat{\Phi}(0) = I$. Moreover, the d matrices

$$\frac{\partial \widehat{\Phi}}{\partial x^\mu}(0), \quad 1 \le \mu \le d,$$

are linearly independent (indeed, they are just the ξ_μ's).

Proof. See Problem 2.2. □

In the preceding, we assumed that G is a matrix Lie group. With some additional work, the exponential map leads to a proof of Theorem 1.9, namely that a matrix group that is closed with respect to $GL(n, \mathbb{C})$ is a matrix Lie group – see Problem 2.3. The idea is to define a putative Lie algebra $\tilde{\mathfrak{g}}$ as the set of matrices α for which $\exp(t\alpha) \in G$ for all t. One then shows that either the exponential map on $\tilde{\mathfrak{g}}$ provides a good parameterisation of G or else G is not closed with respect to $GL(n, \mathbb{C})$.

Next, we establish the basic properties of the Lie algebra.

Proposition 2.8 (Adjoint action)**.** If $B \in G$ and $\alpha \in \mathfrak{g}$, then $B\alpha B^{-1} \in \mathfrak{g}$.

Proof. Let $\alpha = \dot{A}(0)$, where $A(t)$ is a smooth curve in G with $A(0) = I$. Let $\tilde{A}(t) = BA(t)B^{-1}$. Then $\tilde{A}(t)$ is a smooth curve in G with $\tilde{A}(0) = I$. We have that

$$\dot{\tilde{A}}(0) = B\alpha B^{-1},$$

so that $B\alpha B^{-1} \in \mathfrak{g}$. □

Notation. Let Ad_B denote the map

$$\mathrm{Ad}_B : \mathfrak{g} \to \mathfrak{g}; \alpha \mapsto B\alpha B^{-1}.$$

Ad is called the **Adjoint action** of G on $\mathfrak{g}$.

It is easy to verify that Ad_B is linear and that

$$\mathrm{Ad}_B \, \mathrm{Ad}_C = \mathrm{Ad}_{BC} \, .$$

(We'll use this relation in Section 3.)

Given $\alpha, \beta \in \mathfrak{g}$, we define their **Lie bracket**, denoted $[\alpha, \beta]$, by

$$[\alpha, \beta] = \alpha\beta - \beta\alpha.$$

Proposition 2.9 (Lie bracket)**.** If $\alpha, \beta \in \mathfrak{g}$, then $[\alpha, \beta] \in \mathfrak{g}$.

Proof. Let

$$\beta(s) = \mathrm{Ad}_{\exp(s\alpha)} \, \beta = e^{s\alpha} \beta e^{-s\alpha}.$$

Note that $\beta(s)$ is a smooth curve in $\mathfrak{g}$ (not in G, of course). Its derivative also lies in $\mathfrak{g}$. (The tangent to a curve in a vector space may itself be regarded as an element of the vector space. This is seen explicitly if one introduces a basis; $\beta(s) = \beta^\mu(s)\xi_\mu$ implies that $\beta'(s) = \beta^{\mu\prime}(s)\xi_\mu$). We have that

$$\beta'(s) = \alpha\beta(s) - \beta(s)\alpha$$

lies in $\mathfrak{g}$. In particular

$$\beta'(0) = \alpha\beta - \beta\alpha \in \mathfrak{g}.$$

□

Proposition 2.10 (Properties of the Lie bracket)**.**

(i) Linearity

$$[\alpha, c_1\beta_+ c_2\beta_2] = c_1[\alpha, \beta_1] + c_2[\alpha, \beta_2], \text{ for all } \alpha, \beta_1, \beta_2 \in \mathfrak{g} \text{ and } c_1, c_2 \in \mathbb{R}.$$

(ii) Antisymmetry

$$[\alpha, \beta] = -[\beta, \alpha].$$

(iii) Jacobi identity

$$[\alpha, [\beta, \gamma]] = [[\alpha, \beta], \gamma] + [\beta, [\alpha, \gamma]].$$

Proof. Properties (i) and (ii) are immediate. The Jacobi identity follows from straightforward calculation, e.g.

$$[\alpha, [\beta, \gamma]] = [\alpha, \beta\gamma - \gamma\beta] = \alpha\beta\gamma - \beta\gamma\alpha + \alpha\gamma\beta - \gamma\beta\alpha.$$

Expanding the other terms similarly, you can verify (iii). □

The Jacobi identity will be important in what follows.

A (real or complex) vector space $\mathfrak{g}$ satisfying the preceding properties is called a (real or complex) **Lie algebra**. Matrix Lie groups lead to real Lie algebras, as we have seen, but Lie algebras can be defined and studied in their own right. (Lie algebras do not uniquely determine corresponding Lie groups).

Notation. Given $\alpha \in \mathfrak{g}$, let ad_α define the map

$$\mathrm{ad}_\alpha : \mathfrak{g} \to \mathfrak{g}; \quad \beta \mapsto [\alpha, \beta].$$

ad is called the **adjoint** action of $\mathfrak{g}$ on $\mathfrak{g}$.

Example 2.11. [$U(n)$, unitary group in n dimensions.]

A matrix $U \in C^{n\times n}$ is **unitary** if $U^\dagger U = I$. Let $U(n)$ denote the set of unitary matrices in $\mathbb{C}^{n\times n}$.

- $U(n)$ is a matrix group.

  Let U_1, U_2 be unitary. The $U_1U_2(U_1U_2)^\dagger = U_1U_2U_2^\dagger U_1^\dagger = I$, so $U(n)$ is closed under multiplication. Clearly I is unitary. If U is unitary, then since $U^{-1} = U^\dagger$, it follows that $U^{-1}U^{-1\dagger} = U^{-1}U = I$, so that U^{-1} is unitary.

- $U(n)$ is a matrix Lie group.

  It suffices to show that $U(n)$ is closed. Suppose $U_m \in U(n)$, and U_m converges to V. Then

  $$V^\dagger V = \lim_{m\to\infty} U_m^\dagger U_m = I,$$

  so that $V \in U(n)$.

- Lie algebra of $U(n)$.

  It is conventional to denote the Lie algebra of $U(n)$ by $u(n)$. Suppose that $U(t) \in U(n)$ is a smooth curve in $U(n)$ with $U(0) = I$. Then

  $$U^\dagger(t)U(t) = I.$$

  Differentiating with respect to t, we get

  $$\dot{U}^\dagger(t)U(t) + U^\dagger(t)\dot{U}(t) = 0.$$

  Setting $t = 0$ and letting $\alpha = \dot{U}(0)$, we have that

  $$\alpha^\dagger + \alpha = 0.$$

  A matrix $a \in \mathbb{C}^{n\times n}$ is **hermitian** if $a^\dagger = a$, and **antihermitian** if $a^\dagger = -a$. We let $\mathbb{C}^{n\times n}_\pm$ denote the set of hermitian (+) and antihermitian (-) matrices in $\mathbb{C}^{n\times n}$.

  Thus, we have shown that $u(n)$ consists of antihermitian matrices, so that $u(n) \subset \mathbb{C}^{n\times n}_-$.

Next, we show that the converse is true, i.e. $\mathbb{C}_-^{n\times n} \subset u(n)$. Let α be an antihermitian matrix, and let $U(t) = e^{t\alpha}$. Then

$$U^\dagger(t) = e^{t\alpha^\dagger} = e^{-t\alpha},$$

so that

$$U^\dagger(t)\, U(t) = e^{-t\alpha} e^{t\alpha} = I.$$

Therefore, $U(t)$ is a smooth curve in $U(n)$. Clearly $U(0) = I$, Therefore, $\dot{U}(0)$ belongs to $u(n)$. But $\dot{U}(0) = \alpha$.

- Dimension of $U(n)$

  The dimension of $U(n)$ is the vector space dimension of its Lie algebra, the (real) vector space of antihermitian matrices. An antihermitian matrix is parameterised by its n diagonal elements, which are necessarily imaginary, and its $n(n-1)/2$ components above the diagonal, which can be complex (elements below the diagonal are complex conjugates of the ones above). Therefore, the number of independent parameters is $n + 2n(n-1)/2 = n^2$, so that $d = n^2$.

- Adjoint action.

  Let $U \in U(n)$ and $\alpha \in u(n)$. Then

  $$\mathrm{Ad}_U\, \alpha = U^\dagger \alpha U.$$

  It is easy to verify that $\mathrm{Ad}_U\, \alpha$ is antihermitian, since

  $$\left(U^\dagger \alpha U\right)^\dagger = U^\dagger \alpha^\dagger U = -U^\dagger \alpha U.$$

- Lie bracket.

  Let $\alpha, \beta \in u(n)$. It is easy to verify that $[\alpha, \beta] \in u(n)$, i.e. that $[\alpha, \beta]$ is antihermitian. Indeed,

  $$[\alpha, \beta]^\dagger = (\alpha\beta - \beta\alpha)^\dagger = (\beta^\dagger\alpha^\dagger - \alpha^\dagger\beta^\dagger) = (\beta\alpha - \alpha\beta) = -[\alpha, \beta].$$

- $\exp : u(n) \to U(n)$ is onto.

  You can show that the matrix exponential maps $u(n)$ onto all of $U(n)$. In other words, every unitary matrix has an antihermitian logarithm. See Problem 2.4 (d).

3 The Baker-Campbell-Hausdorff Theorem

So far, we started with a matrix Lie group, G, and defined its Lie algebra, $\mathfrak{g}$, in terms of G. We then derived properties of $\mathfrak{g}$ from properties of G. It turns out that this construction can be reversed in part. That is, starting with a Lie algebra $\mathfrak{g}$, you can reconstruct a neighbourhood of the identity of a corresponding Lie group, G. Note, however, that a Lie algebra does not uniquely determine a Lie group; many Lie groups have the same Lie algebra.

(Side remark: In particular, the Lie algebra does not determine the topology of a corresponding Lie group. Here is a very simple example (if you're not familiar with the tensor product of matrices, just ignore): the tensor product of a matrix Lie group with a finite group of matrices produces a new matrix Lie group with the same Lie algebra as the original, but which by construction is not path-connected. More interesting examples arise from quotients of matrix Lie groups by finite normal subgroups. The quotients aren't necessarily matrix Lie groups themselves, but in some cases they are; we'll see some specific examples later. With this remark I'm anticipating some things to come, so if it doesn't make sense, don't worry.)

Theorem 3.1 (Baker-Campbell-Hausdorff Theorem.)**.**

Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Then there exists $\delta > 0$ such that if $\alpha, \beta \in \mathfrak{g}$ and $||\alpha||, ||\beta|| < \delta$, then there exists a unique $\gamma \in \mathfrak{g}$ with $||\gamma|| < 1$ such that

$$e^\alpha e^\beta = e^\gamma.$$

Moreover, γ belongs to the Lie algebra generated by α and β, and is given by the following universal formula:

$$\gamma = \beta + \int_0^1 g\left(e^{t\,\mathrm{ad}_\alpha} e^{\mathrm{ad}_\beta}\right) \alpha\, dt,$$

where

$$g(z) = \frac{\log z}{z-1}.$$

This formula may be understood as follows: We have that

$$g(z) = g(1-(1-z)) = \sum_{j=0}^{\infty} \frac{(1-z)^j}{j+1},$$

which converges for $|z| < 1$. Therefore,

$$g\left(e^{t\,\mathrm{ad}_\alpha} e^{\mathrm{ad}_\beta}\right) = \sum_{j=0}^{\infty} \frac{1}{j+1} \left(e^{\mathrm{ad}_\alpha} e^{\mathrm{ad}_\beta} - I\right)^j.$$

Regarding α and β as small, this may be expanded in a formal power series in ad_α and ad_β, and the t-integral evaluated. For example, evaluating through terms involving up to two Lie brackets, we obtain

$$\gamma = \alpha + \beta + \tfrac{1}{2}[\alpha,\beta] + \tfrac{1}{12}\left([\alpha,[\alpha,\beta]] + [\beta,[\beta,\alpha]]\right) + O(4).$$

See Problem 3.1.

First, let us derive a useful formula in its own right. Recall that for fixed $\alpha \in \mathbb{C}^{n\times n}$, the differential equation

$$\dot{A}(t) = \alpha A(t), \quad A(0) = I,$$

has the unique solution $A(t) = e^{t\alpha}$; indeed, this is one way to define the matrix exponential. Allowing α to depend on t, there are two ways we might generalise this problem. First, we might consider the differential equation

$$\dot{A}(t) = \alpha(t) A(t), \quad A(0) = I.$$

This leads to the Volterra kernel or, in physics language, the time-ordered product or Born/Feynman-diagram series. The second problem, which is the one we will consider here, is how to calculate $d/dt\, e^{\alpha(t)}$.

Proposition 3.2 (Derivative of t-dependent exponential)**.** Let $\alpha(t)$ be a smooth curve in $\mathbb{C}^{n\times n}$. Then

$$\frac{d}{dt} e^{\alpha(t)} = \left(\int_0^1 e^{s\alpha(t)} \dot{\alpha}(t) e^{-s\alpha(t)}\, ds\right) e^{\alpha(t)}.$$

Note that if $\alpha(t) = t\alpha$, we recover the elementary case above.

Proof. We will work with formal power series. Convergence isn't hard to establish, but we won't do this here. We have that

$$\frac{d}{dt} e^{\alpha(t)} = \frac{d}{dt} \sum_{j=0}^{\infty} \frac{\alpha^j(t)}{j!}.$$

Differentiating the jth term, we get

$$\frac{d}{dt} \alpha^j(t) = \sum_{k=0}^{j-1} \alpha(t)^k \dot{\alpha}(t) \alpha^{j-k-1};$$

note that $\alpha(t)$ and $\dot{\alpha}(t)$ do not necessarily commute. From now on, to make the expressions more compact, we'll omit the argument t. Combining the preceding expansions, we get that

$$\frac{d}{dt} e^{\alpha} = \sum_{j=0}^{\infty} \sum_{k=0}^{j-1} \frac{\alpha^k \dot{\alpha} \alpha^{j-k-1}}{j!} = \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \frac{\alpha^k \dot{\alpha} \alpha^{j-k-1}}{j!}.$$

Change the summation variable, letting $l = j-(k+1)$ to obtain

$$\frac{d}{dt} e^{\alpha} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{\alpha^k\, \dot{\alpha}\, \alpha^l}{(k+l+1)!}.$$

We make use of the integral representation

$$\frac{k!\, l!}{(k+l+1)!} = \int_0^1 s^k (1-s)^l\, ds$$

(side remark: for k, l non-integral, this is the integral representation of the β-function). Substituting into the power series, we obtain

$$\frac{d}{dt}e^{\alpha} = \int_0^1 \left(\sum_{k=0}^{\infty} \frac{(s\alpha)^k}{k!}\right) \dot{\alpha} \left(\sum_{l=0}^{\infty} \frac{((1-s)\alpha)^l}{l!}\right) ds = \int_0^1 e^{s\alpha}\,\dot{\alpha}\,e^{(1-s)\alpha}\,ds,$$

which yields the required result. $\square$

The following establishes a connection between the Adjoint and adjoint actions via the matrix exponential.

Proposition 3.3 (Adjoint and adjoint action.)**.** Let $\alpha, \beta \in \mathfrak{g}$. Then

$$e^{\alpha}\beta e^{-\alpha} = e^{\mathrm{ad}_{\alpha}}\beta := \sum_{j=0}^{\infty} \frac{(\mathrm{ad}_{\alpha})^j}{j!}\beta,$$

where $\mathrm{ad}_{\alpha}^2\,\beta = \mathrm{ad}_{\alpha}(\mathrm{ad}_{\alpha}\,\beta) = [\alpha,[\alpha,\beta]]$, etc. Equivalently,

$$\mathrm{Ad}_{e^{\alpha}} = e^{\mathrm{ad}_{\alpha}},$$

or

$$\mathrm{Ad} \circ \exp = \exp \circ \,\mathrm{ad}\,.$$

Proof. We introduce a parameter t, and use the uniqueness of solutions to ODE's. Let

$$\beta_1(t) = e^{t\alpha}\beta e^{-t\alpha},$$
$$\beta_2(t) = e^{t\,\mathrm{ad}_{\alpha}}\beta.$$

We have that

$$\dot{\beta}_1 = [\alpha, \beta_1(t)], \quad \beta_1(0) = \beta,$$

while

$$\dot{\beta}_2 = \mathrm{ad}_{\alpha}\,\beta_2(t) = [\alpha, \beta_2(t)], \quad \beta_2(0) = \beta.$$

Thus $\beta_1(t)$ and $\beta_2(t)$ satisfy the same ODE with the same initial condition. It follows that $\beta_1(t) = \beta_2(t)$. $\square$

Proof of Theorem 3.1. Let

$$C(t) = e^{t\alpha}e^{\beta}.$$

By Property 2 of Proposition2.4, for $||\alpha||$, $||\beta||$ and $|t|$ sufficiently small, $C(t)$ has a unique logarithm near 0, which we denote by $\gamma(t)$:

$$\gamma(t) = \log C(t).$$

Let us determine the differential equation satisfied by $\gamma(t)$. From Proposition 3.2, we have that

$$\frac{d}{dt}e^{\gamma(t)} = \left(\int_0^1 e^{s\gamma(t)}\dot{\gamma}(t)e^{-s\gamma(t)}\,ds\right)e^{\gamma(t)}.$$

From Proposition 3.3, this may be written as

$$\frac{d}{dt}e^{\gamma(t)} = \left(\int_0^1 e^{s\,\mathrm{ad}_{\gamma(t)}}\dot{\gamma}(t)\,ds\right)e^{\gamma(t)}$$

(note that $\mathrm{ad}_{s\gamma} = s\,\mathrm{ad}_{\gamma}$ due to the linearity of the Lie bracket). Noting that

$$\int_0^1 e^{sx}\,ds = \frac{e^x - 1}{x} := F(x),$$

we may write this compactly as

$$\frac{d}{dt}e^{\gamma(t)} = \left(F(\mathrm{ad}_{\gamma(t)})\dot{\gamma}(t)\right)e^{\gamma(t)}$$

From the definition of $\gamma(t)$,

$$\frac{d}{dt}e^{\gamma(t)} = \dot{C}(t) = \alpha C(t) = \alpha e^{\gamma(t)}.$$

Equating these two expressions, we get that

$$F(\mathrm{ad}_{\gamma(t)})\dot{\gamma}(t) = \alpha.$$

Letting

$$G(x) = \frac{1}{F(x)} = \frac{x}{e^x - 1},$$

we obtain, on applying $G(\mathrm{ad}_\gamma(t))$ to both sides of the preceding, that

$$\dot{\gamma}(t) = G(\mathrm{ad}_{\gamma(t)})\alpha.$$

We would like to express the right-hand side in terms of ad_α and ad_β. For this we will use the relation between ad, Ad and exp given in Proposition 3.3. Define $g(z)$ by $G(x) = g(e^x)$. Equivalently, we have that

$$g(z) = G(\log z) = \frac{\log z}{z - 1}.$$

Then

$$G(\mathrm{ad}_{\gamma(t)}) = g\left(e^{\mathrm{ad}_{\gamma(t)}}\right) = g(\mathrm{Ad}_{e^{\gamma(t)}}),$$

from Proposition 3.3. But

$$\mathrm{Ad}_{e^{\gamma(t)}} = \mathrm{Ad}_{C(t)} = \mathrm{Ad}_{e^{t\alpha}e^{\beta}} = \mathrm{Ad}_{e^{t\alpha}}\,\mathrm{Ad}_{e^{\beta}} = e^{t\,\mathrm{ad}_\alpha}e^{\mathrm{ad}_\beta},$$

again using Proposition 3.3. Then

$$G(\mathrm{ad}_{\gamma(t)}) = g\left(e^{t\,\mathrm{ad}_\alpha}e^{\mathrm{ad}_\beta}\right).$$

Substituting above, we get that

$$\dot{\gamma}(t) = g\left(e^{t\,\mathrm{ad}_\alpha}e^{\mathrm{ad}_\beta}\right)\alpha.$$

Now we may integrate with respect to t, noting that $\gamma(0) = \beta$ and $\gamma(1) = \gamma$. We obtain

$$\gamma = \beta + \int_0^1 g\left(e^{t\,\mathrm{ad}_\alpha}e^{\mathrm{ad}_\beta}\right)\alpha\,dt.$$

□

4 SU(2) and SO(3)

We will study a particular matrix Lie group, SU(2), in greater detail. In addition to its importance in applications (it is closely related to rotations in $\mathbb{R}^3$), it provides useful, simple but representative examples of general results, and is a building block for the general theory, particularly the classification of simple Lie algebras.

4.1 Parameterisation of SU(2)

SU(2) is the group of 2×2 unitary matrices with determinant equal to 1. It is a subgroup of the group $U(2)$ of 2×2 unitary matrices. Note that the determinant of a unitary matrix necessarily has modulus equal to 1.

Let $u \in \mathrm{SU}(2)$. We may write

$$u = \begin{pmatrix} w & x \\ y & z \end{pmatrix}, \quad w, x, y, z \in \mathbb{C}.$$

The fact that u is unitary means that $u^\dagger = u^{-1}$. We have that

$$u^\dagger = \begin{pmatrix} w^* & y^* \\ x^* & z^* \end{pmatrix},$$

$$u^{-1} = \frac{1}{\det u}\begin{pmatrix} z & -x \\ -y & w \end{pmatrix} = \begin{pmatrix} z & -x \\ -y & w \end{pmatrix},$$

where we have used the fact that $\det u = 1$. It follows that u is of the form

$$u = \begin{pmatrix} w & x \\ -x^* & w^* \end{pmatrix}, \quad w, x \in \mathbb{C}.$$

Moreover, the fact that $\det u = 1$ implies that

$$|w|^2 + |x|^2 = 1.$$

4.2 Real parameterisation. Pauli matrices.

Write w and z in terms of their real and imaginary components:

$$w = a_0 + ia_3,$$
$$x = a_2 + ia_1.$$

Then

$$u = \begin{pmatrix} a_0 + ia_3 & a_2 + ia_1 \\ -a_2 + ia_1 & a_0 - ia_3 \end{pmatrix}, \quad a_0, a_1, a_2, a_3 \in \mathbb{R},$$

$$= a_0 I_2 + ia_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + ia_2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + ia_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Here, I_2 denotes the 2×2 identity matrix, to be distinguished from the 3×3 identity matrix I_3, which will appear later on. We define the matrices

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

and use the notation $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. These matrices are called the **Pauli matrices**. We note that the Pauli matrices are hermitian and traceless;

$$\sigma_j^\dagger = \sigma_j, \quad \operatorname{tr} \sigma_j = 0.$$

We also write $\mathbf{a} = (a_1, a_2, a_3) \in \mathbb{R}^3$, and $a = (a_0, \mathbf{a}) \in \mathbb{R}^4$. Then we may write

$$u = a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma}, \quad ||a|| = 1. \tag{2}$$

4.3 Topology of SU(2). Relation to S^3

From the paramaterisation (2), it follows that SU(2) is in 1-1 correspondence with S^3, the unit sphere in $\mathbb{R}^4$. Moreover, the map $a \in S^3 \mapsto a_0 I + i\mathbf{a} \cdot \boldsymbol{\sigma} \in \mathbb{C}^{2\times 2}$ is smooth. Thus, we may regard SU(2) as S^3 endowed with its usual topology inherited from $\mathbb{R}^4$ as well as a group structure.

With the identification of S^3 and SU(2) in view, certain topological properties of SU(2) become apparent. First, as S^3 is a closed and bounded subset of $\mathbb{R}^4$, it is *compact*; every open cover has a finite subcover, and every sequence has a convergent subsequence. Next, SU(2) is (path) *connected*; that is, any two elements u_1 and u_2 in SU(2) can be joined by a continuous path in SU(2). Finally, SU(2) is *simply connected*; that is, if $u(t)$, $0 < t < 1$, is a closed, continuous curve in SU(2), so that $u(0) = u(1) = u_*$, then $u(t)$ can be continuously deformed into the constant curve $u(t) = u_*$ while keeping its endpoints fixed at u_* throughout. More explicitly, there exists an SU(2)-valued function $U(t,s)$ continuous in s, t with $0 \le s, t \le 1$, such that $U(t, 0) = u(t)$, $U(0, s) = U(1, s) = u_*$, $U(t, 1) = u_*$. These facts can be established analytically; see Problem Sheet 2.

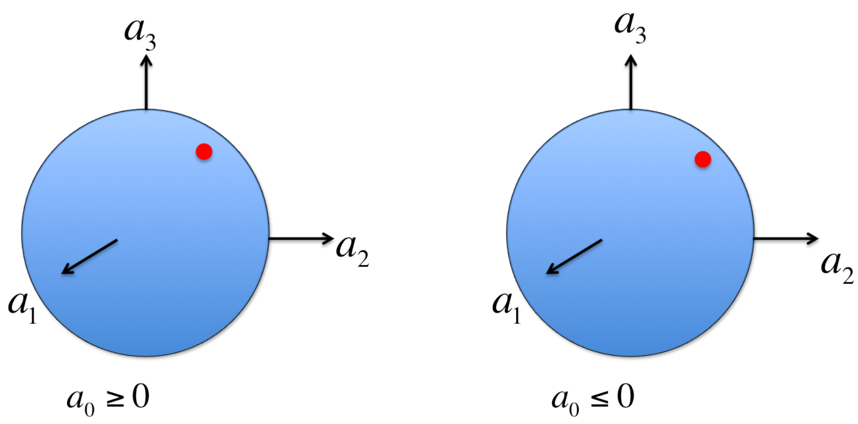


Figure 4: S^3 may be regarded as the union of two copies of the closed unit ball $||\mathbf{a}|| \le 1$ in $\mathbb{R}^3$ with corresponding points on the surfaces of the two balls identified. The ball on the left corresponds to $a_0 = \sqrt{1-\mathbf{a}\cdot\mathbf{a}}$, and the ball on the right to $a_0 = -\sqrt{1-\mathbf{a}\cdot\mathbf{a}}$. The surfaces of the two balls correspond to $||\mathbf{a}|| = 1$ and $a_0 = 0$ – hence the identification of corresponding points. The centres of the two balls correspond to $a_0 = 1, \mathbf{a} = 0$ (i.e., the identity I_2 in SU(2)) and $a_0 = -1, \mathbf{a} = 0$ ($-I_2$ in SU(2)).

4.4 Product in SU(2). Relation to Quarternions.

It is straightforward to verify the following:

$$\sigma_1\sigma_2 = -\sigma_2\sigma_1 = i\sigma_3, \quad \sigma_2\sigma_3 = -\sigma_3\sigma_2 = i\sigma_1, \quad \sigma_3\sigma_1 = -\sigma_1\sigma_3 = i\sigma_2,$$
$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I_2.$$

These can be summarised more concisely as

$$\sigma_j\sigma_k = \delta_{jk}I_2 + i\epsilon_{jkl}\sigma_l,$$

where the sum $\sum_{l=1}^3$ is implied in the last term, or

$$(\mathbf{a}\cdot\boldsymbol{\sigma})(\mathbf{b}\cdot\boldsymbol{\sigma}) = \mathbf{a}\cdot\mathbf{b}I_2 + i(\mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}. \tag{3}$$

These expressions allow us to write down the product of matrices in SU(2) in terms of the parameterisation (2). Let $u = a_0 I_2 + i\mathbf{a}\cdot\boldsymbol{\sigma}$ and $v = b_0 I_2 + i\mathbf{b}\cdot\boldsymbol{\sigma}$. Then

$$uv = (a_0 b_0 - \mathbf{a}\cdot\mathbf{b})I_2 + i(a_0\mathbf{b} + b_0\mathbf{a} - \mathbf{a}\times\mathbf{b})\cdot\boldsymbol{\sigma}. \tag{4}$$

*In the product formula (4), the fact that $u, v \in \mathrm{SU}(2)$ means that $||a||^2 = ||b||^2 = 1$. But the formula remains valid if the constraint on the norm of a and b is lifted. We are led to introduce the following, larger set of matrices: Let

$$q(x) = x_0 I_2 + i\mathbf{x}\cdot\boldsymbol{\sigma}, \quad x = (x_0, \mathbf{x}) \in \mathbb{R}^4,$$

and let

$$\mathbb{H} = \{q(x)\,|\,x \in \mathbb{R}^4\}.$$

$\mathbb{H}$ is in fact a parameterisation (or representation) of the quarternions in $\mathbb{C}^{2\times 2}$. Let us record some facts about quarternions, all of which are easily derived from (1):

(i) Product: $q(x)q(y) = q(z)$, where $z = (x\cdot y, x_0\mathbf{y} + y_0\mathbf{x} - \mathbf{x}\times\mathbf{y})$

(ii) Conjugate: $q^\dagger(x) = q(x_0, -\mathbf{x})$

(iii) Inner product: $\langle q(x), q(y)\rangle = \frac{1}{2}\,\mathrm{tr}\left(q(x)^\dagger q(y)\right) = x\cdot y := x_0 y_0 + \mathbf{x}\cdot\mathbf{y}$ (thus, inner product on $\mathbb{H}$ coincides with inner product on $\mathbb{R}^4$)

(iv) Norm: $||q(x)||^2 = \left\langle q^\dagger(x), q(x)\right\rangle$

(v) Norm is multiplicative: $||q(x)q(y)|| = ||q(x)||\,||q(y)||$

(vi) Inverse: For $x \neq 0$, $q(x)^{-1} = q(x)^\dagger/||q(x)||$

Note that the conjugate, inner product and norm are just as for $\mathbb{C}^{2\times 2}$ (apart from the factor of $\frac{1}{2}$ in the definition of the inner product).

SU(2) may be identified with the group of *unit quarternions*, i.e. quarternions with unit norm.

The quarternions are, additionally, an *associative division algebra*; they form an associative algebra (commutative and associative addition operation, associative product, distributive law), and every nonzero element has a multiplicative inverse). Up to isomorphism, there are just three real associative division algebras: the real numbers, the complex numbers, and the quarternions. Note that in the first two cases, too, the set of elements of unit norm form a multiplicative group (in the case of the $\mathbb{R}$, $\{\pm\}$; in the case of $\mathbb{C}$, $\{e^{i\theta}\}$).

4.5 Lie algebra su(2)

In Example 2.11, we found that $u(2)$, the Lie algebra of the group $U(2)$ of 2×2 unitary matrices, is given by $\mathbb{C}^{2\times 2}_{-}$, the space of 2×2 antihermitian matrices. As SU(2) is subgroup of $U(2)$, its Lie algebra, su(2), must be a subspace of $\mathbb{C}^{2\times 2}_{-}$. Let us determine the constraint on elements of su(2) that follows from the unit-determinant condition.

Let $u(t) \in \mathrm{SU}(2)$ be a smooth curve in SU(2) with $u(0) = I_2$. Since $\det u(t) = 1$, we have that

$$\frac{d}{dt}\det u(t) = 0.$$

From the general formula for the derivative of the determinant, we have that

$$\frac{d}{dt}\det u(t) = \mathrm{tr}(u^{-1}(t)\dot{u}(t))\det u(t).$$

Evaluating at $t = 0$, we get that

$$0 = \mathrm{tr}\,\dot{u}(0).$$

Thus, elements of su(2), in addition to being antihermitian, must also be traceless.

Let $\mathbb{C}^{2\times 2}_{-0}$ denote the space of 2×2 traceless antihermitian matrices. It is easily seen that $\mathbb{C}^{2\times 2}_{-0}$ consists of linear combinations of the Pauli matrices with imaginary coefficients:

$$\mathbb{C}^{2\times 2}_{-0} = \{i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}\,|\,\boldsymbol{\alpha}\in\mathbb{R}^3\}.$$

Therefore, we have shown that

$$\mathrm{su}(2) \subset \mathbb{C}^{2\times 2}_{-0}.$$

It is easy to see that the inclusion goes the other way, i.e.

$$\mathrm{su}(2) \supset \mathbb{C}^{2\times 2}_{-0}.$$

Indeed, let $u(t) = \exp(it\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})$. We saw in Example 2.11 that $u(t)$ is unitary for all real t. We also have that

$$\frac{d}{dt}\det u = \mathrm{tr}\left(u^{-1}\dot{u}\right)\det u = \mathrm{tr}\left(u^{-1}ui\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}\right)\det u = i\,\mathrm{tr}(\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})\det u = 0.$$

Thus, $\det u(t)$ is constant. Since $\det u(0) = 1$, it follows that $\det u(t) = 1$. Thus, $u(t) \in \mathrm{su}(2)$. Since $\dot{u}(0) = i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}$, it follows that all 2×2 traceless antihermitian matrices belong to su(2).

We conclude that

$$\mathrm{su}(2) = \mathbb{C}^{2\times 2}_{-0} = \{i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}\,|\,\boldsymbol{\alpha}\in\mathbb{R}^3\}.$$

The Lie bracket is easily worked out from (3). We have that

$$[i\sigma_j, i\sigma_k] = -2i\epsilon_{jkl}\sigma_l,$$

or

$$[i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}, i\boldsymbol{\beta}\cdot\boldsymbol{\sigma}] = -2i(\boldsymbol{\alpha}\times\boldsymbol{\beta})\cdot\boldsymbol{\sigma}.$$

The Jacobi identity is equivalent to the identity

$$\boldsymbol{\alpha} \times (\boldsymbol{\beta} \times \boldsymbol{\gamma}) = (\boldsymbol{\alpha} \times \boldsymbol{\beta}) \times \boldsymbol{\gamma} + \boldsymbol{\beta} \times (\boldsymbol{\alpha} \times \boldsymbol{\gamma}).$$

4.6 Exponential map

The exponential map $\exp : \mathrm{su}(2) \to \mathrm{SU}(2)$ can be evaluated explicitly – see Problem Sheet 2. We have that

$$e^{i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}} = \cos\alpha I_2 + i \sin\alpha\, \hat{\boldsymbol{\alpha}} \cdot \boldsymbol{\sigma},$$

where $\alpha = ||\boldsymbol{\alpha}||$ and $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}/||\boldsymbol{\alpha}||$.

4.7 Inner product on su(2).

As a subspace of $\mathbb{C}^{2\times 2}$, su(2) is endowed with an inner product. The definition in Section 1.1 is to take the inner product of u, v to be $\mathrm{tr}(u^\dagger v)$. Here, it will be convenient to introduce a factor of 1/2; we define

$$\langle u, v\rangle = \frac{1}{2}\,\mathrm{tr}(u^\dagger v).$$

Note that with this definition, the identity has norm 1. Also, the matrices $i\boldsymbol{\sigma}$ form an orthonormal basis for su(2);

$$\big\langle i\sigma_j, i\sigma_k\big\rangle = \tfrac{1}{2}\,\mathrm{tr}((i\sigma_j)^\dagger i\sigma_k) = \tfrac{1}{2}\,\mathrm{tr}(\sigma_j\sigma_k) = \tfrac{1}{2}\,\mathrm{tr}(\delta_{jk} I + i\epsilon_{jkl}\sigma_l) = \delta_{jk}.$$

Therefore,

$$\langle i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}, i\boldsymbol{\beta}\cdot\boldsymbol{\sigma}\rangle = \boldsymbol{\alpha}\cdot\boldsymbol{\beta}.$$

4.8 Adjoint action. Relation to rotations.

The Adjoint action is defined in the usual way:

$$\mathrm{Ad}_u\, i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma} = u(i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})u^\dagger.$$

As the Adjoint action is a linear map on su(2), we may write that

$$Ad_u\, i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma} = i(R(u)\boldsymbol{\alpha})\cdot\boldsymbol{\sigma},$$

where $R(u) : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map on $\mathbb{R}^3$, i.e. a 3×3 real matrix (since we have introduced a basis). We want to determine the properties of $R(u)$ and derive explicit formulas for it.

Proposition 4.1. *$R(u)$ is an orthogonal matrix.*

Proof. We have that

$$(R(u)\boldsymbol{\alpha})\cdot(R(u)\boldsymbol{\beta}) = \Big\langle u(i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})u^\dagger, u(i\boldsymbol{\beta}\cdot\boldsymbol{\sigma})u^\dagger\Big\rangle = \tfrac{1}{2}\,\mathrm{tr}\left((u(i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})u^\dagger)^\dagger(u(i\boldsymbol{\beta}\cdot\boldsymbol{\sigma})u^\dagger)\right)$$
$$= \tfrac{1}{2}\,\mathrm{tr}(u(i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})^\dagger u^\dagger u(i\boldsymbol{\beta}\cdot\boldsymbol{\sigma})u^\dagger) = \tfrac{1}{2}\,\mathrm{tr}\left((i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma})^\dagger i\boldsymbol{\beta}\cdot\boldsymbol{\sigma}\right) = \langle i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}, i\boldsymbol{\beta}\cdot\boldsymbol{\sigma}\rangle = \boldsymbol{\alpha}\cdot\boldsymbol{\beta}.$$

□

Proposition 4.2.

$$\det R(u) = 1.$$

Proof. An orthogonal matrix necessarily has determinant equal to 1 or −1. Also, since $R(I_2) = I_3$, we have that $\det R(I_2) = 1$. Take $u_* \in \mathrm{SU}(2)$. Since SU(2) is connected, there exists a continuous curve $u(t)$ in SU(2) with $u(0) = I_2$ and $u(1) = u_*$. $\det(R(u(t))$ is continuous, and therefore constant (since it is either 1 or −1), and therefore equal to 1 (since it equals 1 at $t = 0$). Therefore, $\det R(u_*) = 1$ (and u_* is arbitrary). □

We may derive an explicit formula for $R(u)$. Indeed, letting $\hat{e}_j$ denote the jth unit vector in $\mathbb{R}^3$ (e.g., $\hat{e}_1 = (1,0,0)^T$, $\hat{e}_2 = (0,1,0)^T$, etc) we have that

$$R_{jk}(u) = \hat{e}_j \cdot R(u)\,\hat{e}_k = \langle i\sigma_j, (i(R(u)\,\hat{e}_k) \cdot \boldsymbol{\sigma} \rangle = \tfrac{1}{2}\operatorname{tr}\left((i\sigma_j)^\dagger (u\, i\sigma_k u^\dagger)\right).$$

Writing

$$u = a_0 I_2 + i\mathbf{a}\cdot\boldsymbol{\sigma},$$

it is straightforward to obtain (see Problem Sheet 2)

$$R_{jk}(u) = (a_0^2 - \mathbf{a}\cdot\mathbf{a})\delta_{jk} + 2a_j a_k + 2a_0 a_r \epsilon_{rjk}. \tag{5}$$

Equivalently,

$$R(u)\,\mathbf{r} = (a_0^2 - \mathbf{a}\cdot\mathbf{a})\,\mathbf{r} + 2(\mathbf{a}\cdot\mathbf{r})\,\mathbf{a} - 2a_0\mathbf{a}\times\mathbf{r}. \tag{6}$$

Recall the general fact from Section 2 (which is easily verified) that $\mathrm{Ad}_u\,\mathrm{Ad}_v = \mathrm{Ad}_{uv}$. This implies that

$$R(u)R(v) = R(uv).$$

Let SO(3) denote the set of 3×3 real orthogonal matrices with determinant equal to 1. It is easily checked that SO(3) is a matrix Lie group (it forms a group and is closed in $\mathbb{R}^{3\times 3}$). We may summarise the considerations so far with the observation that the map

$$R : \mathrm{SU}(2) \to \mathrm{SO}(3);\, u \mapsto R(u)$$

is a homomorphism of matrix Lie groups.

Next, we show that the map R is onto. We will make use of the fact that the action of any element of SO(3) is described by a rotation about some axis through the origin. Let $\mathcal{R}(\hat{\mathbf{n}}, \theta)$ denote the rotation about $\hat{\mathbf{n}}$ by an angle θ, where $\hat{\mathbf{n}}$ is a unit vector in $\mathbb{R}^3$. The convention is that the rotation is anticlockwise with respect to $\hat{\mathbf{n}}$. We have that

$$\mathcal{R}(\hat{\mathbf{n}}, \theta)\,\hat{\mathbf{n}} = \hat{\mathbf{n}},$$

while if $\mathbf{v}$ is perpendicular to $\hat{\mathbf{n}}$, we have that

$$\mathcal{R}(\hat{\mathbf{n}}, \theta)\mathbf{v} = \cos\theta\mathbf{v} + \sin\theta\hat{\mathbf{n}}\times\mathbf{v}.$$

Proposition 4.3. Let

$$u = \cos\frac{\theta}{2} I_2 - i\sin\frac{\theta}{2}\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}.$$

Then

$$R(u) = \mathcal{R}(\hat{\mathbf{n}}, \theta).$$

Proof. From (6), noting that

$$a_0 = \cos\frac{\theta}{2}, \quad \mathbf{a} = -\sin\frac{\theta}{2}\hat{\mathbf{n}},$$

we get that

$$R(u)\hat{\mathbf{n}} = \left(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\right)\hat{\mathbf{n}} + 2\sin^2\frac{\theta}{2}\hat{\mathbf{n}} = \hat{\mathbf{n}},$$

and for $\mathbf{v}$ perpendicular to $\hat{\mathbf{n}}$,

$$R(u)\,\mathbf{v} = \cos\theta\,\mathbf{v} + \sin\theta\,\hat{\mathbf{n}}\times\mathbf{v}.$$

□

Another way to establish Proposition 4.3 is through Proposition 3.3 – see Problem Sheet 2.

Proposition 4.3 implies that the map $u \mapsto R(u)$ from SU(2) to SO(3) is onto, since every rotation can be realised with appropriate choice of u.

Proposition 4.4.

$$\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{I_2, -I_2\}.$$

Proof. We have already shown that the map $u \mapsto R(u)$ is a surjective homomorphism of SU(2). Therefore, SO(3) is isomorphic to the quotient of SU(2) by the kernel of R. We need to determine for which u we have $R(u) = I_3$. From (5), with $u = a_0 I_2 + i\mathbf{a}\cdot\boldsymbol{\sigma}$, this holds if and only if $a_0^2 = 1$, $\mathbf{a} = 0$, i.e. $a_0 = \pm 1$, $\mathbf{a} = 0$, or $u = \pm I_2$. □

Clearly, $u(-a) = -u(a)$. Therefore, under the map R, both $u(a)$ and $u(-a)$ are mapped to the same element of SO(3).

4.9 *Hamilton's theory of turns

The relation between SU(2) and SO(3) provides a geometric construction for the product of two rotations, due to Hamilton. Given rotations $\mathcal{R}(\hat{\mathbf{n}}_1, \theta_1)$ and $\mathcal{R}(\hat{\mathbf{n}}_2, \theta_2)$, let $\hat{\mathbf{m}}_0$ denote a unit vector that lies on the intersection of the two great circles normal to $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$. Let $\hat{\mathbf{m}}_1$ denote the vector obtained by rotating $\hat{\mathbf{m}}_0$ by an angle $-\theta_1/2$ about $\hat{\mathbf{n}}_1$. Similarly, let $\hat{\mathbf{m}}_2$ denote the vector obtained by rotating $\hat{\mathbf{m}}_0$ by an angle $\theta_2/2$ about $\hat{\mathbf{n}}_2$. Let $\hat{\mathbf{n}}$ denote the unit vector normal to $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$, and let $\theta/2$ denote the angle by which $\hat{\mathbf{m}}_1$ is rotated about $\hat{\mathbf{n}}$ to get to $\hat{\mathbf{m}}_2$. Then

$$\mathcal{R}(\hat{\mathbf{n}}_1, \theta_1)\mathcal{R}(\hat{\mathbf{n}}_2, \theta_2) = \mathcal{R}(\hat{\mathbf{n}}, \theta).$$

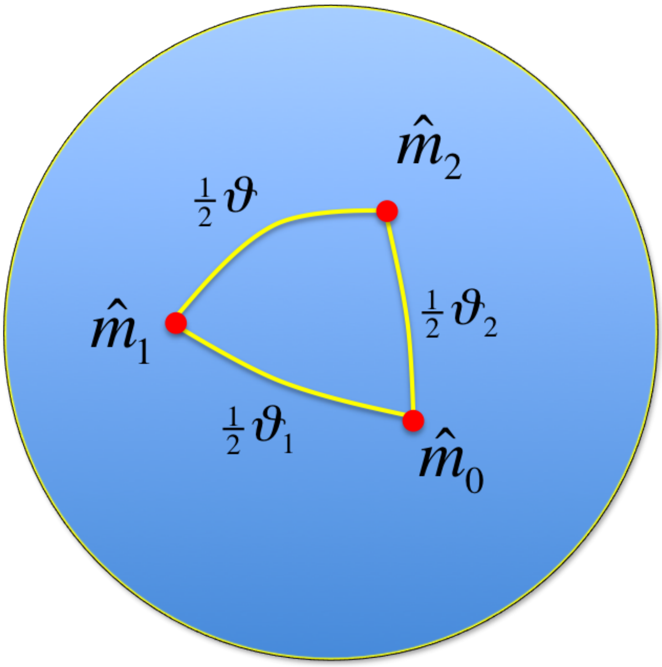


Figure 5:

4.10 *Topology of SO(3)

From the relationship between SU(2) and SO(3), we can deduce certain facts about the topology of SO(3) using what we know about the topology of SU(2) $\cong S^3$.

First, we argue that SO(3) is path connected. Let $\mathcal{R}_0, \mathcal{R}_1 \in$ SO(3), and let $u_0, u_1 \in$ SU(2) be such that $R(u_0) = \mathcal{R}_1$ and $R(u_1) = \mathcal{R}_2$. Since SU(2) is connected, there exists a continuous path $u(t)$ in SU(2) with $u(0) = u_0$ and $u(1) = 1$. Then $\mathcal{R}(t) = R(u(t))$ is a continuous path in SO(3) from $\mathcal{R}_0$ to $\mathcal{R}_1$.

From the fact that SU(2) is simply connected, one can show that SO(3) = SU(2)/$\{\pm I_2\}$ is doubly connected; the fundamental group of SO(3) is isomorphic to $\{\pm I_2\} \cong \mathbb{Z}_2$. We won't give details; the idea is that closed loops in SO(3) can be distinguished according to whether their pre-images in SU(2) are closed or not (if not, the endpoints of their pre-images in SU(2) differ by a sign). Under concatenation, the two classes of closed loops in SO(3) with endpoints at I_3, say, obey the group law of $\mathbb{Z}_2$.

This is the basis for the "belt trick".

5 Haar measure

5.1 Motivation – invariant measure on $\mathbb{R}$

Let f be a function on $\mathbb{R}$, which may be either real- or complex-valued. We define the weighted integral of f by

$$\langle f\rangle := \int_{-\infty}^{\infty} f(x)\rho(x)\,dx,$$

where the real weight function $\rho(x)$ is nonnegative (we assume throughout that f and ρ are such that the integral converges). Given $a \in \mathbb{R}$, we define

$$f_a(x) := f(x-a),$$

which is the function obtained by translating the argument of f by a. We want to choose the weight function $\rho(x)$ so that $\langle f\rangle = \langle f_a\rangle$ for all f and for all a. That is,

$$\int_{-\infty}^{\infty} f(x)\rho(x)\,dx = \int_{-\infty}^{\infty} f(x-a)\rho(x)\,dx. \tag{7}$$

It's easy to determine the condition on ρ implied by this requirement, but we'll go through it in detail to prepare the ground for the analogous argument for functions defined on a general group (the real numbers $\mathbb{R}$ under addition constitute a group, of course, so the present case is a particular example). On the right-hand side of (7), we make the change of variables $y = x - a$. We get that

$$\int_{-\infty}^{\infty} f(x-a)\rho(x)\,dx = \int_{-\infty}^{\infty} f(y)\rho(y+a)\,dy.$$

On the left-hand side, we make the trivial change of variable $y = x$, just in order to facilitate comparison with the left-hand side:

$$\int_{-\infty}^{\infty} f(x)\rho(x)\,dx = \int_{-\infty}^{\infty} f(y)\rho(y)\,dy.$$

Equating the two expressions, we get that

$$\int_{-\infty}^{\infty} f(y)\rho(y)\,dy = \int_{-\infty}^{\infty} f(y)\rho(y+a)\,dy,$$

or

$$\int_{-\infty}^{\infty} f(y)(\rho(y+a)-\rho(y))\,dy = 0.$$

As this relation must hold for all f, we conclude that

$$\rho(y+a) = \rho(y),$$

which in turn must hold for all a. Therefore, ρ must be constant, i.e. translation-invariant. Equivalently,

$$\rho(y) = \rho(0) \text{ for all } y.$$

5.2 Invariant measure on matrix Lie group

We want to consider integration of a (real- or complex-valued) function f defined on a matrix Lie group $G \subset \mathrm{GL}(n,\mathbb{C})$. To make sense of integration over G, we will make use of a parameterisation. Suppose G is d dimensional. Then there exists an open set $P \subset \mathbb{R}^d$ containing the origin and a smooth 1-1 map $\Phi : P \to G$ with $\Phi(0) = I_n$. For simplicity, we will assume that Φ is onto; that is, G is parameterised by a single domain P. This is not generally the case. In general, $\Phi(P)$ covers only part of G, and a number of overlapping parameterisations are required to cover all of G. However, even with this simplifying assumption, we can capture the main idea underlying invariant measures on a matrix Lie group. The general case can in principle be dealt with using the similar arguments. A more natural approach is to regard G as a differentiable manifold. While this is beyond the scope of our treatment, we give a brief sketch in Section 5.8.

We define

$$\langle f\rangle := \int_P f(\Phi(x))\rho(x)dx,$$

where dx denotes the d-dimensional volume element, and $\rho(x)$ is a nonnegative weight. Given $A \in G$, we define functions $f_{L,A}$ and $f_{R,A}$ on G by

$$f_{L,A}(X) = f(A^{-1}X), \quad f_{R,A}(X) = f(XA^{-1}),$$

where $X \in G$. $f_{L,A}$ is obtained from f by multiplying its argument on the left by A^{-1}, and $f_{R,A}$ is obtained from f by multiplying its argument on the right by A^{-1}. Note that if G is not abelian, then $f_{L,A}$ and $f_{R,A}$ are in general different. We want to determine $\rho_L(x)$ so that, for $\rho = \rho_L$,

$$\langle f \rangle = \langle f_{L,A} \rangle \tag{8}$$

for all functions f and all $A \in G$ (we assume the integral converges). Similarly, we want to determine $\rho_R(x)$ so that, for $\rho = \rho_R$,

$$\langle f \rangle = \langle f_{R,A} \rangle \tag{9}$$

for all functions f and all $A \in G$. $\rho_L(x)$ and $\rho_R(x)$ are called left-invariant and right-invariant Haar measures (with respect to the parameterisation given by Φ). We will obtain formulas for them and determine under what circumstances they coincide.

5.3 Group multiplication in terms of parameters

As we will be working with the parameterisation (in order to define integration), we want to express group multiplication in terms of parameters. Given $A \in G$, we define a map $L_A : P \to P$ by

$$L_A(x) = \Phi^{-1}(A\Phi(x)), \text{ or } \Phi(L_A(x)) = A\Phi(x).$$

That is, L_A is the map on the parameter domain that describes multiplication on the left by A. Given x, $L_A(x)$ is obtained by taking the associated group element $\Phi(x)$, multiplying on the left by A, and then determining the parameter values which correspond to $A\Phi(x)$. It can be shown that $L_A(x)$ is smooth, 1-1 and onto (given our assumption that $\Phi(P) = G$).

We have that

$$L_{AB}(x) = L_A(L_B(x)), \tag{10}$$

since

$$L_{AB}(x) = \Phi^{-1}(AB\Phi(x)) = \Phi^{-1}(A\Phi(L_B(x))) = L_A(L_B(x)).$$

Letting $B = A^{-1}$, we conclude that

$$(L_A)^{-1} = L_{A^{-1}},$$

so that L_A is invertible with smooth inverse. Therefore, L_A is a diffeomorphism.

Let us differentiate the composition law (10). The Chain Rule gives

$$L'_{AB}(x) = L'_A(L_B(x))L'_B(x), \tag{11}$$

where $L'_A(x) \in \mathbb{R}^{d\times d}$ denotes the Jacobian, i.e.

$$[L'_A(x)]_{jk} = \frac{\partial L^j_A}{\partial x^k}(x).$$

Similarly, we can express multiplication on the right in terms of parameters. Given $A \in G$, we define a map $R_A : P \to P$ by

$$R_A(x) = \Phi^{-1}(\Phi(x)A), \text{ or } \Phi(R_A(x)) = \Phi(x)A.$$

That is, R_A is the map on the parameter domain that describes multiplication on the right by A.

The composition law for right multiplication is given by

$$R_{AB}(x) = R_B(R_A(x)), \tag{12}$$

so that the maps R_A and R_B come out in the opposite order to how they appear in the product AB. It's easy to see why. For left multiplication, to obtain ABX from X, first we multiply on the left by B and then by A. For right multiplication, to obtain XAB from X, first we multiply on the right by A and then by B. To verify explicitly,

$$R_{AB}(x) = \Phi^{-1}(\Phi(x)AB) = \Phi^{-1}(\Phi(R_A(x))B) = R_B(R_A(x)).$$

Letting $B = A^{-1}$, we conclude that

$$(R_A)^{-1} = R_{A^{-1}},$$

so that R_A is invertible with smooth inverse. Therefore, R_A is a diffeomorphism.

Differentiating the composition law (12), we get

$$R'_{AB}(x) = R'_B(R_A(x))R'_A(x). \tag{13}$$

5.4 Calculation of ρ_L and ρ_R

First, we consider the left-invariant weight ρ_L. We look for a nonnegative function $\rho_L(x)$ on P such that (8) holds for all (suitable) functions f on G and for all $A \in G$. In terms of the left-translation map introduced above, we have that

$$f_{L,A}(\Phi(x)) = f(A^{-1}\Phi(x)) = f(\Phi(L_{A^{-1}}(x))).$$

Therefore, the condition (8) becomes

$$\int_P f(\Phi(x))\rho_L(x)\,dx = \int_P f(\Phi(L_{A^{-1}}(x)))\rho_L(x)\,dx, \tag{14}$$

On the right-hand side of (14), we make the change of variables

$$y = L_{A^{-1}}(x),$$

so that

$$x = L_A(y), \quad dx = \left|\det L'_A(y)\right|\,dy,$$

where the factor $\left|\det L'_A(y)\right|$ accounts for the volume element according to the change-of-variables formula for multidimensional integrals. The right-hand side becomes

$$\int_P f(\Phi(y))\rho_L(L_A(y))\left|\det L'_A(y)\right|\,dy.$$

On the left-hand side, to facilitate comparison, we make the trivial change of variables $y = x$ to obtain

$$\int_P f(\Phi(y))\rho_L(y)\,dy.$$

Then (14) becomes

$$\int_P f(\Phi(y))\rho_L(y)\,dy = \int_P f(\Phi(y))\rho_L(L_A(y))\left|\det L'_A(y)\right|\,dy.$$

In order for this to hold for all f and for all $A \in G$, we must have

$$\rho_L(L_A(y)) = \frac{\rho_L(y)}{\left|\det L'_A(y)\right|}, \tag{15}$$

which must hold for all $y \in P$ and for all $A \in G$.

Let us see what we can learn from (15) in the special case that $y = 0$. Recalling our convention $\Phi(0) = I_n$, we note that $\Phi(L_A(0)) = A\Phi(0) = A$. Introducing $x \in P$ by the relation $\Phi(x) = A$ (in other words, x denotes the parameter values associated with A), we may write that

$$\rho_L(L_A(0)) = \rho_L(x),$$

so that (15) becomes

$$\boxed{\rho_L(x) = \frac{\rho_L(0)}{\left|\det L'_{\Phi(x)}(0)\right|}.} \tag{16}$$

This gives an explicit formula for $\rho_L(x)$. (Note that ρ_L is determined up to an arbitrary multiplicative constant, which is reflected in the fact that $\rho_L(x)$ is proportional to $\rho_L(0)$.)

It remains to check that $\rho_L(x)$ as given by (16) satisfies the condition (15) for all $y \in P$. Using (16), we get for the right-hand side of (15) that

$$\frac{\rho_L(y)}{\left|\det L'_A(y)\right|} = \frac{\rho_L(0)}{\left|\det L'_A(y)\right| \left|\det L'_{\Phi(y)}(0)\right|}.$$

From the Chain Rule (11) and the fact that $y = L_{\Phi(y)}(0)$, we get that

$$L'_A(y)L'_{\Phi(y)}(0) = L'_A(L_{\Phi(y)}(0))L'_{\Phi(y)}(0) = L'_{A\Phi(y)}(0),$$

so that

$$\left|\det L'_A(y)\right| \left|\det L'_{\Phi(y)}(0)\right| = \left|\det L'_{A\Phi(y)(0)}\right|.$$

Therefore, the right-hand side of (15) may be written as

$$\frac{\rho_L(0)}{\left|\det L'_{A\Phi(y)}(0)\right|}.$$

On the left-hand side of (15), we have

$$\rho_L(L_A(y)) = \rho_L(L_{A\Phi(y)}(0)) = \frac{\rho_L(0)}{\left|\det L'_{A\Phi(y)}(0)\right|},$$

where we have used the formula (16). This coincides with the right-hand side, so that (15) is satisfied for all $y \in P$ and for all $A \in G$.

The formula (16) has a simple geometric interpretation, as illustrated in Figure 6.

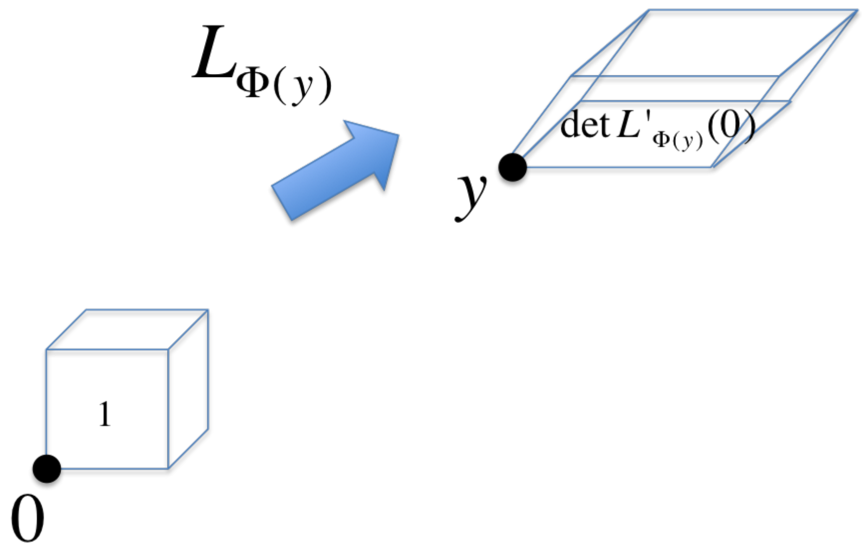


Figure 6: Left-invariant weight. An (infinitesimal) volume element based on the origin in the parameter domain P is mapped under left multiplication by $\Phi(y)$ into a volume element at y. The volume of the mapped element changes by a factor of $\left|\det L'_{\Phi(y)}(0)\right|$. In the invariant integral, however, we want both volume elements weighted equally. Hence, we compensate the change of volume by taking the weight $\rho_L(y)$ to be $\rho_L(0)/\left|\det L'_{\Phi(y)}(0)\right|$.

The derivation of the right-invariant weight proceeds similarly. The only small difference to note is in the order of factors in the composition rule, $R_{AB} = R_B \circ R_A$. We briefly summarise the calculation. We look for a nonnegative function $\rho_R(x)$ on P such that (9) holds for all (suitable) functions f on G and for all $A \in G$. In terms of the right-translation map introduced above, we have that

$$f_{R,A}(\Phi(x)) = f(\Phi(x)A^{-1}) = f(\Phi(R_{A^{-1}}(x))).$$

Therefore, the condition (9) becomes

$$\int_P f(\Phi(x))\rho_R(x)\,dx = \int_P f(\Phi(R_{A^{-1}}(x)))\rho_R(x)\,dx. \tag{17}$$

On the right-hand side of (14), we make the change of variables

$$y = R_{A^{-1}}(x),$$

so that

$$x = R_A(y), \quad dx = \left|\det R'_A(y)\right|\,dy.$$

The right-hand side becomes

$$\int_P f(\Phi(y))\rho_R(R_A(y))\left|\det R'_A(y)\right|\,dy.$$

Then (17) becomes

$$\int_P f(\Phi(y))\rho_R(y)\,dy = \int_P f(\Phi(y))\rho_R(R_A(y))\left|\det R'_A(y)\right|\,dy.$$

We obtain the condition

$$\rho_R(R_A(y)) = \frac{\rho_R(y)}{\left|\det R'_A(y)\right|}, \tag{18}$$

which must hold for all $y \in P$ and for all $A \in G$. For $y = 0$, and $x = \Phi^{-1}(A)$, (18) yields the explicit formula

$$\boxed{\rho_R(x) = \frac{\rho_R(0)}{\left|\det R'_{\Phi(x)}(0)\right|}.} \tag{19}$$

Let us check that $\rho_R(x)$ as given by (19) satisfies the condition (18) for all $y \in P$. Using (19), we get for the right-hand side of (18) that

$$\frac{\rho_R(y)}{\left|\det R'_A(y)\right|} = \frac{\rho_R(0)}{\left|\det R'_A(y)\right|\,\left|\det R'_{\Phi(y)}(0)\right|}.$$

From the Chain Rule (13) and the fact that $y = R_{\Phi(y)}(0)$, we get that

$$R'_A(y)R'_{\Phi(y)}(0) = R'_A(R_{\Phi(y)}(0))R'_{\Phi(y)}(0) = R'_{\Phi(y)A}(0),$$

so that

$$\left|\det R'_A(y)\right|\,\left|\det R'_{\Phi(y)}(0)\right| = \left|\det R'_{\Phi(y)A}(0)\right|.$$

Therefore, the right-hand side of (18) may be written as

$$\frac{\rho_L(0)}{\left|\det R'_{\Phi(y)A}(0)\right|}.$$

On the left-hand side of (18), we have

$$\rho_R(R_A(y)) = \rho_R(R_{\Phi(y)A}(0)) = \frac{\rho_L(0)}{\left|\det R'_{\Phi(y)A}(0)\right|},$$

where we have used the formula (19). This coincides with the right-hand side, so that (18) is satisfied for all $y \in P$ and for all $A \in G$.

5.5 Example: Affine group over $\mathbb{R}$

We take the affine group over $\mathbb{R}$, denoted $\mathrm{Aff}(\mathbb{R})$, to be the subgroup of $\mathrm{GL}(2,\mathbb{R})$ consisting of matrices of the form

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}, \text{ where } a > 0.$$

It is easy to check that $\mathrm{Aff}(\mathbb{R})$ is a matrix Lie group. We take as parameterisation

$$\Phi : \mathbb{R}^2 = \{(\lambda, b)\} \to \mathrm{Aff}(\mathbb{R}); (\lambda, b) \mapsto \Phi(\lambda, b) = \begin{pmatrix} e^\lambda & b \\ 0 & 1 \end{pmatrix}.$$

In this case, G *is* covered by the parameterisation, i.e. $\Phi(\mathbb{R}^2) = \mathrm{Aff}(\mathbb{R})$. The nomenclature derives from the fact that

$$\begin{pmatrix} e^\lambda & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} e^\lambda x + b \\ 1 \end{pmatrix},$$

so that these matrices generate the affine transformations $x \mapsto e^\lambda x + b$ on $\mathbb{R}$.

It is easy to work out that

$$\Phi(\lambda_1, b_1)\Phi(\lambda_2, b_2) = \Phi(\lambda_1 + \lambda_2, e^{\lambda_1} b_2 + b_1).$$

From this relation we can read off the left- and right-translation maps, as follows:

$$\begin{aligned} L_{\Phi(\lambda,b)}(x, y) &= (\lambda + x, b + e^\lambda y), \\ R_{\Phi(\lambda,b)}(x, y) &= (x + \lambda, y + e^x b). \end{aligned}$$

The Jacobians of the left- and right-translation maps are given by

$$\begin{aligned} L'_{\Phi(\lambda,b)}(x, y) &= \frac{\partial(\lambda + x, b + e^\lambda y)}{\partial(x, y)} = \begin{pmatrix} 1 & 0 \\ 0 & e^\lambda \end{pmatrix}, \\ R'_{\Phi(\lambda,b)}(x, y) &= \frac{\partial(x + \lambda, y + e^x b)}{\partial(x, y)} = \begin{pmatrix} 1 & 0 \\ e^x b & 1 \end{pmatrix}. \end{aligned}$$

Their determinants are given by

$$\begin{aligned} \det L'_{\Phi(\lambda,b)}(x, y) &= e^\lambda, \\ \det R'_{\Phi(\lambda,b)}(x, y) &= 1, \end{aligned}$$

both of which are independent of x and y. Thus, from (16) and (19), the left-invariant and right-invariant weights are given by

$$\begin{aligned} \rho_L(\lambda, b) &= e^{-\lambda} \rho_L(0, 0), \\ \rho_R(\lambda, b) &= \rho_R(0, 0), \end{aligned}$$

where $\rho_L(0, 0)$, $\rho_R(0, 0)$ are arbitrary constants. Thus, the left- and right-invariant weights for the affine group do not coincide. Figure 7 below gives a geometrical argument.

5.6 Bi-invariant Haar measure and the modular function

From (16) and (19), we have that

$$\frac{\rho_R(x)}{\rho_L(x)} = C \left| \frac{\det L'_{\Phi(x)}(0)}{\det R'_{\Phi(x)}(0)} \right|, \tag{20}$$

where $C = \rho_R(0)/\rho_L(0)$. The ratio of the left- and right-invariant Haar measures turns out to be related to the Adjoint action.

As discussed in Section 2, the Adjoint action on a matrix Lie group is the linear map on its Lie algebra $\mathfrak{g}$ given by $a \mapsto \mathrm{Ad}_A\, a = AaA^{-1}$. As the Lie algebra $\mathfrak{g}$ is a d-dimensional real vector space, by

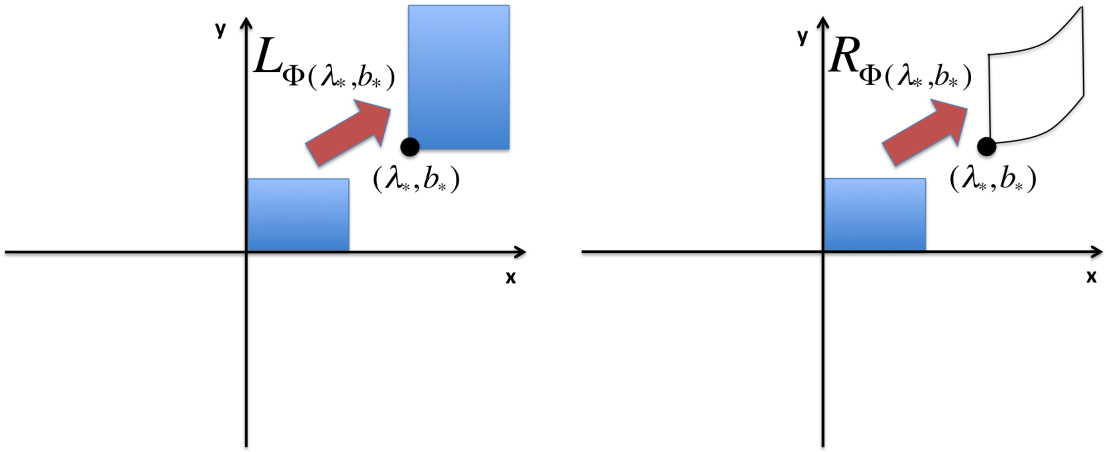


Figure 7: (a) Left multiplication by $\Phi(\lambda_*, b_*)$ is applied to a rectangular area element based at the origin. The image is a rectangular area element, translated by (λ_*, b_*) and scaled in the vertical direction by e^{λ^*}. Thus, the area of the image is changed by a factor of e^{λ_*}. (b) Right multiplication by $\Phi(\lambda_*, b_*)$ is applied to a rectangular area element based at the origin. The image is translated by (λ_*, b_*) and sheared, but its area is unchanged.

.

introducing a basis $\xi_1, \ldots, \xi_d$ for $\mathfrak{g}$, we may represent Ad_A by a matrix $M_A \in \mathbb{R}^{d\times d}$, as follows: Letting $\mathrm{Ad}_A(\sum_{j=1}^d c^j \xi_j)$ be given by $\sum_{k=1}^d d^k \xi_k$, we have that

$$d = M_A \cdot c.$$

Under the change of basis

$$\xi_j \mapsto \eta_j = \sum_{k=1} S_{kj}\xi_k,$$

where $S \in \mathbb{R}^{d\times d}$ is nonsingular, the matrix M_A is replaced by

$$N_A = S^{-1} M_A S.$$

However, we see that the determinants of N_A and M_A are the same. Therefore, the determinant of the Adjoint map $\det M_A$ is a basis independent, and may be regarded as an intrinsic function on a matrix Lie group, called the modular function.

Definition 5.1 (Modular function)**.** *The modular function $\Delta : G \to \mathbb{R}$ is the function on G given by*

$$\Delta(A) := \det M_A,$$

where M_A is a matrix representation of the Adjoint map $\mathrm{Ad}_A : \mathfrak{g} \to \mathfrak{g}$ with respect to a basis on $\mathfrak{g}$.

Let us record some properties of the modular function.

Proposition 5.1 (Properties of the modular function)**.** *Let Δ be the modular function on a d-dimensional matrix Lie group $G \subset \mathrm{GL}(n, \mathbb{C})$.*

(i) $\Delta(I_n) = 1$

(ii) $\Delta(AB) = \Delta(A)\Delta(B)$, *for all* $A, B \in G$

(iii) $\Delta(A) \neq 0$, *for all* $A \in G$

(iv) If G is connected, then Δ is nonnegative.

Proof.

(i) Ad_{I_n} is the identity map in $\mathfrak{g}$, so that $M_A = I_d$, and $\Delta(I_n) = \det I_d = 1$.

(ii) By Proposition 2.8, $\mathrm{Ad}_{AB} = \mathrm{Ad}_A \mathrm{Ad}_B$. It follows that $M_{AB} = M_A M_B$, so that

$$\Delta(AB) = \det M_{AB} = \det M_A M_B = \det M_A \det M_B = \Delta(A)\Delta(B).$$

(iii) From the previous two results, since $\Delta(A)\Delta(A^{-1}) = 1$, it follows that $\Delta(A)$ cannot vanish.

(iv) Suppose that G is connected. Given $A_* \in G$, and suppose that $\Delta(A_*) < 0$. Let $A(t)$, $0 \leq t \leq 1$ be a continuous curve joining I_n to A_*. Since $\Delta(A(0)) = 1$, $\Delta(A(1)) < 0$ and $\Delta(A(t))$ is continuous in t, it follows that $\Delta(A(t)$ must vanish for some $A(t) \in G$, contradicting the preceding result.

□

The parameterisation Φ provides an explicit formula for $\Delta(A)$.

Proposition 5.2.

$$\Delta(A) = \frac{\det L'_A(0)}{\det R'_A(0)}.$$

Proof. The parameterisation Φ provides a basis for $\mathfrak{g}$, namely

$$\xi_j = \frac{\partial \Phi}{\partial x^j}(0) = \left.\frac{d}{dt}\right|_{t=0} \Phi(te_{(j)}),$$

where $e_{(j)} \in \mathbb{R}^d$ is the unit vector in the jth direction. Then for $x \in \mathbb{R}^d$, $M_A \cdot x$ is defined by

$$\sum_{j=1}^{d} (M_A \cdot x)_j \xi_j := \left.\frac{d}{dt}\right|_{t=0} A\Phi(tx)A^{-1}.$$

But

$$A\Phi(tx)A^{-1} = \Phi(R_{A^{-1}}(L_A(tx))).$$

By the Chain Rule,

$$\left.\frac{d}{dt}\right|_{t=0} A\Phi(tx)A^{-1} = \Phi'(R_{A^{-1}}(L_A(0)))R'_{A^{-1}}(L_A(0))L'_A(0) \cdot x.$$

Since $R_A(0) = L_A(0) = \Phi^{-1}(A)$, we have that $R_{A^{-1}}(L_A(0)) = 0$ and

$$R'_{A^{-1}}(L_A(0)) = R'_{A^{-1}}(R_A(0)).$$

By the Chain Rule applied to the identity $R_{A^{-1}}(R_A(z)) = z$, we get that

$$R'_{A^{-1}}(R_A(0))R'_A(0) = I_d,$$

or

$$R'_{A^{-1}}(R_A(0)) = \left(R'_A(0)\right)^{-1}.$$

Substituting above, we get that

$$\left.\frac{d}{dt}\right|_{t=0} A\Phi(tx)A^{-1} = \sum_j \left(\left(R'_A(0)\right)^{-1} L'_A(0) \cdot x\right)_j \xi_j,$$

so that

$$M_A = \left(R'_A(0)\right)^{-1} L'_A(0).$$

Taking determinants, we get that

$$\det \mathrm{Ad}_A = \det M_A = \frac{\det L'_A(0)}{\det R'_A(0)},$$

as required. □

Proposition 5.3.

$$\frac{\rho_R(\Phi^{-1}(A))}{\rho_L(\Phi^{-1}(A))} = C|\Delta(A)|.$$

Proof. This follows from (20) and Proposition 5.2. □

Definition 5.1 (Bi-invariant Haar measure)**.** We say that a matrix Lie group G has a bi-invariant Haar measure if there exists a weight function $\rho(x)$ that satisfies both the conditions for left- and right-invariance, i.e. (15) and (18).

Definition 5.2 (Unimodular group)**.** A matrix Lie group G is **unimodular** if

$$\Delta(A) = 1.$$

Proposition 5.4. *A connected matrix Lie group G has a biinvariant Haar measure if and only if G is unimodular.*

Proof. If G is unimodular, it's clear from Proposition 5.3 that ρ_L/ρ_R is constant, so that both ρ_L and ρ_R are left- and right-invariant. Conversely, if G has a bi-invariant Haar measure, then ρ_R/ρ_L is constant From Proposition 5.3, this implies that $|\Delta(A)|$ is constant. Since $\Delta(I_n) = 1$ and G is connected, this implies that $\Delta(A) = 1$. □

An important class of unimodular groups are compact connected groups. A matrix Lie group $G \subset \mathrm{GL}(n, \mathbb{C})$ is *compact* if it is a closed and bounded subset of $\mathbb{C}^{n\times n}$

Proposition 5.5. *A compact connected matrix Lie group is unimodular.*

Proof. The argument is based on the fact that a continuous function on a compact space is bounded.

Since G is connected, it follows from Proposition 5.1(iv) that the modular function Δ is nonnegative. Suppose that G is not unimodular. Then there exists $A \in G$ such that $\det \mathrm{Ad}_A \neq 1$. From Proposition 5.1(ii), $\Delta(A) = 1/\Delta(A^{-1})$. Therefore, by replacing A with A^{-1} if necessary, we may assume that $\Delta(A) > 1$. It follows that Δ is unbounded on G, since by Proposition 5.1(ii), $\Delta(A^n) = \Delta(A)^n$. Since Δ is continuous, it follows that G cannot be compact. □

5.7 Haar measure on SU(2)

SU(2) is compact, so that we expect it to have a bi-invariant measure.

We take P to be the open unit ball in $\mathbb{R}^3$, i.e.

$$P = \{\mathbf{b} \in \mathbb{R}^3 \,|\, b < 1\}.$$

We define $\Phi : P \to \mathrm{SU}(2)$ by

$$\Phi(\mathbf{b}) = aI_2 + i\mathbf{b}\cdot\boldsymbol{\sigma}, \quad a = (1-b^2)^{1/2}.$$

In this case, $\Phi(P)$ does not cover all of SU(2), but rather half of it, since $SU(2)$ consists of all elements of the form $aI_2 + i\mathbf{b}\cdot\boldsymbol{\sigma}$ with $a = \pm(1-b^2)^{1/2}$. To compute the Haar measure, it turns out to be sufficient to consider the contribution where a is positive; we'll return to this point in the end.

Multiplication on SU(2) is given by

$$\Phi(\mathbf{b}_1)\Phi(\mathbf{b}_2) = \Phi(\mathbf{c}),$$

where

$$\mathbf{c} = a_1\mathbf{b}_2 + a_2\mathbf{b}_1 - \mathbf{b}_1 \times \mathbf{b}_2.$$

It follows that the left- and right-multiplication maps are given by

$$L_{\Phi(\mathbf{b})}(\mathbf{y}) = a\mathbf{y} + x\mathbf{b} - \mathbf{b}\times\mathbf{y},$$
$$R_{\Phi(\mathbf{b})}(\mathbf{y}) = x\mathbf{b} + a\mathbf{y} - \mathbf{y}\times\mathbf{b},$$

where

$$x = (1-y^2)^{1/2}.$$

Their Jacobians are given by

$$L'_{\Phi(\mathbf{b})}(\mathbf{y}) = \begin{pmatrix} a - \frac{y_1}{x}b_1 & -\frac{y_2}{x}b_1 + b_3 & -\frac{y_3}{x}b_1 - b_2 \\ -\frac{y_1}{x}b_2 - b_3 & a - \frac{y_2}{x}b_2 & -\frac{y_3}{x}b_2 + b_1 \\ -\frac{y_1}{x}b_3 + b_2 & -\frac{y_2}{x}b_3 - b_1 & a - \frac{y_3}{x}b_3 \end{pmatrix}$$
$$R'_{\Phi(\mathbf{b})}(\mathbf{y}) = \begin{pmatrix} a - \frac{y_1}{x}b_1 & -\frac{y_2}{x}b_1 - b_3 & -\frac{y_3}{x}b_1 + b_2 \\ -\frac{y_1}{x}b_2 + b_3 & a - \frac{y_2}{x}b_2 & -\frac{y_3}{x}b_2 - b_1 \\ -\frac{y_1}{x}b_3 - b_2 & -\frac{y_2}{x}b_3 + b_1 & a - \frac{y_3}{x}b_3 \end{pmatrix}$$

This expressions look rather formidable, but they simplify when evaluated at $\mathbf{y} = 0$ (in which case $x = 1$) to yield

$$L'_{\Phi(\mathbf{b})}(0) = \begin{pmatrix} a & b_3 & -b_2 \\ -b_3 & a & b_1 \\ b_2 & -b_1 & a \end{pmatrix},$$
$$R'_{\Phi(\mathbf{b})}(0) = \begin{pmatrix} a & -b_3 & b_2 \\ b_3 & a & -b_1 \\ -b_2 & b_1 & a \end{pmatrix},$$

Taking determinants, we get

$$\det L'_{\Phi(\mathbf{b})}(0) = a^3 + ab^2 = a(a^2 + b^2) = a,$$
$$\det R'_{\Phi(\mathbf{b})}(0) = a^3 + ab^2 = a(a^2 + b^2) = a.$$

Thus, the left- and right-invariant weights are given by

$$\rho_L(\mathbf{b}) = \frac{\rho_L(0)}{\sqrt{1-b^2}},$$
$$\rho_R(\mathbf{b}) = \frac{\rho_R(0)}{\sqrt{1-b^2}}. \tag{21}$$

These differ by a constant factor, as expected.

The expressions (21) hold also for the "other half" of SU(2), i.e. elements of the form $aI_2 + i\mathbf{b}\cdot\boldsymbol{\sigma}$ with $a = -(1-b^2)^{1/2}$. This can be "seen" from (15) and (18) after taking $A = -I_2$. Note that multiplication by $-I_2$ on either the left or the right sends $aI_2 + i\mathbf{b}\cdot\boldsymbol{\sigma}$ to $-aI_2 + -i\mathbf{b}\cdot\boldsymbol{\sigma}$, i.e. from one half of SU(2) to the other. As $L_{-I_2}(\mathbf{b}) = R_{-I_2}(\mathbf{b}) = -\mathbf{b}$, we have that $\det L'_{I_2}(\mathbf{b}) = \det R'_{I_2}(\mathbf{b}) = -1$, and $|\det L'_{-I_2}(\mathbf{b})| = |\det R'_{-I_2}(\mathbf{b})| = 1$. Therefore, (21) holds throughout SU(2). Warning: this argument is really a fudge, as L_{-I_2} and R_{-I_2} are not well defined (cf the definitions: the domain of Φ^{-1} is the "positive half", P, of SU(2)). A full and proper treatment of Haar measure on Lie groups is best carried out within the framework of differentiable manifolds – see below. In fact, Haar measure generalises to locally compact topological groups, which do not necessarily have a differentiable structure, but this is definitely beyond our scope.

5.8 *Intrinsic definition of Haar measure

This will be a brief account without definitions – it's intended to give an impression, and for you to follow up independently if you are interested. We regard G as a d-dimensional differentiable manifold, and assume that the multiplication and inversion operations are smooth. Given $g \in G$, we define diffeomorphisms $L_g : G \to G; x \mapsto gx$ and $R_g : G \to G; x \mapsto xg$, which describe left and right multiplication. Let ρ_e denote a d-form on the tangent space of G at the identity, T_eG, which may be identified with the Lie algebra $\mathfrak{g}$. We define d-forms ρ_L and ρ_R on G by the formulas

$$\rho_L(g) = L^*_{g^{-1},e}\rho_e, \quad \rho_R(g) = R^*_{g^{-1},e}\rho_e,$$

where L^*_g and R^*_g denote the pullbacks by L_g and R_g. Then one can show that

$$L^*_g\rho_L = \rho_L, \quad R^*_g\rho_R = \rho_R,$$

so that ρ_L and ρ_R are everywhere nonvanishing volume forms on G invariant respectively under left and right multiplication.

5.9 Haar Measure on SO(n)

We compute the Haar measure on SO(n) in terms of the Cayley parameterisation given in Example 1.8 c). The calculation is adapted from a problem on the 2016 examination paper.

The Cayley transform on $n \times n$ real matrices X is given by

$$\Phi(X) = (I - X)(I + X)^{-1},$$

where X is restricted to have no eigenvalue equal to -1. Let $\mathbb{R}_{-}^{n\times n}$ denote the space of antisymmetric matrices. As shown in Example 1.8 c), $\Phi : \mathbb{R}_{-}^{n\times n} \to \mathrm{SO}(n)$ is a good parameterisation of $\mathrm{SO}(n)$, and $\Phi(\Phi(X)) = X$. We compute the left-invariant Haar $\rho_L(X)$ (it's also right invariant, since $\mathrm{SO}(n)$ is compact) according to the following plan:

- First, we get an explicit formula for L_R, regarded as a map $L_R : \mathbb{R}_{-}^{n\times n} \to \mathbb{R}_{-}^{n\times n}$. L_R is defined implicitly by the equation $\Phi(L_R(Y)) = R\,\Phi(Y)$. Applying Φ to both sides, we get

$$L_R(Y) = \Phi(R\,\Phi(Y)).$$

- Since the Haar measure ρ_L is expressed in terms of the determinant of the Jacobian $L'_{\Phi(X)}(0)$, we compute $L_{\Phi(X)}(\epsilon Y)$ through first order in ϵ for $X, Y \in \mathbb{R}_{-}^{n\times n}$. The coefficient of ϵ in the expansion of $L_{\Phi(X)}(\epsilon Y)$ is then $L'_{\Phi(X)}(0) \cdot Y$, ie the directional derivative of $L_L\Phi(X)(W)$ at $W = 0$ along Y. We have that

$$L_{\Phi(X)}(\epsilon Y) = \Phi(\Phi(X)\Phi(\epsilon Y)).$$

Noting that, for any $n \times n$ matrix W, we have that $(I + \epsilon W)^{-1} = I - \epsilon W + O(\epsilon^2)$, it follows that

$$\Phi(\epsilon Y) = (I - \epsilon Y)(I + \epsilon Y)^{-1} = (I - \epsilon Y)(I - \epsilon Y + O(\epsilon^2)) = I - 2\epsilon Y + O(\epsilon^2).$$

Let

$$Z = \Phi(X), \text{ so that } \Phi(Z) = \Phi(\Phi(X)) = X.$$

Then

$$L_{\Phi(X)}(\epsilon Y) = \Phi(Z(I - 2\epsilon Y + O(\epsilon^2)).$$

From the definition of Φ, it follows that

$$L_{\Phi(X)}(\epsilon Y) = (I - Z + 2\epsilon ZY)(I + Z - 2\epsilon ZY)^{-1} + O(\epsilon^2).$$

But

$$(I + Z - 2\epsilon ZY) = (I - 2\epsilon ZY(I + Z)^{-1})(I + Z),$$

so that

$$(I + Z - 2\epsilon ZY)^{-1} = (I + Z)^{-1}(I + 2\epsilon ZY(I + Z)^{-1}) + O(\epsilon^2).$$

Neglecting terms of $O(\epsilon^2)$, we get that

$$L_{\Phi(X)}(\epsilon Y) = (I - Z + 2\epsilon ZY)(I + Z)^{-1}(I + 2\epsilon ZY(I + Z)^{-1}) = \Phi(Z) + 2\epsilon(I + \Phi(Z))ZY(I + Z)^{-1}).$$

Recalling that $\Phi(Z) = X$ and that $Z = (I - X)(I + X)^{-1}$, we get, neglecting $O(\epsilon^2)$, that

$$L_{\Phi(X)}(\epsilon Y) = X + 2\epsilon(I + X)(I - X)(I + X)^{-1}Y(I + Z)^{-1}) = X + 2\epsilon(I - X)Y(I + Z)^{-1}).$$

Also,

$$I + Z = I + (I - X)(1 + X)^{-1} = (I + X + I - X)(1 + X)^{-1} = 2(1 + X)^{-1},$$

so that

$$(I + Z)^{-1} = \tfrac{1}{2}(1 + X).$$

Finally, we obtain

$$L_{\Phi(X)}(\epsilon Y) = X + \epsilon(I - X)Y1 + X) + O(\epsilon^2).$$

It follows that

$$\frac{d}{dt}L_{\Phi(X)}(tY)\bigg|_{t=0} = L'_{\Phi(X)}(0) \cdot Y = (I - X)Y(I + X).$$

- We need to compute the determinant of $L'_{\Phi(X)}(0)$. We do this by finding the eigenvectors and eigenvalues of $L'_{\Phi(X)}(0)$, regarded as a linear map $\mathcal{A}$ on $\mathbb{R}^{n\times n}_{-} \to \mathbb{R}^{n\times n}_{-} \to$ given by

$$Y \mapsto \mathcal{A}\cdot Y := (I-X)Y(I+X)$$

(it's less writing to use $\mathcal{A}$ rather than $L'_{\Phi(X)}(0)$). As the eigenvalues of $\mathcal{A}$ may be complex (in fact, they turn out to be imaginary), it will be convenient to extend $\mathcal{A}$ to act on the complexification of $\mathbb{R}^{n\times n}_{-}$, namely the vector space $\mathbb{C}^{n\times n}_{-}$ of complex antisymmetric matrices.

To find eigenvectors of $\mathcal{A}$ (note: these eigenvectors are in fact $n\times n$ antisymmetric matrices), we note that since X is real antisymmetric, it is diagonalisable with imaginary eigenvalues $i\omega_j$ Let $v_{(j)} \in \mathbb{C}^n$ denote the corresponding eigenvectors. Let

$$Y_{(jk)} = v_{(j)}v^T_{(k)} - v_{(k)}v^T_{(j)}, \quad \text{i.e., } [Y_{(jk)}]_{ab} = v_{(j)a}v_{(k)b} - v_{(k)a}v_{(j)b},$$

so that $Y_{(jk)} \in \mathbb{C}^{n\times n}_{-}$. Then

$$\begin{aligned}\mathcal{A}\cdot Y_{(jk)} = (I-X)\left(v_{(j)}v^T_{(k)} - v_{(k)}v^T_{(j)}\right)(I+X) &= (1-i\omega_j)v_{(j)}v^T_{(k)}(1-i\omega_k) - (1-i\omega_k)v_{(j)}v^T_{(k)}(1-i\omega_j)\\ &= (1-i\omega_j)(1-i\omega_k)Y_{(jk)}.\end{aligned}$$

It follows that $Y_{(jk)}$ is an eigenvector of $\mathcal{A}$ with eigenvalue $(1-i\omega_j)(1-i\omega_k)$. As the $v_{(j)}$'s form a basis for $\mathbb{C}^n$, the $Y_{(jk)}$'s, with $j<k$, form a basis for $\mathbb{C}^{n\times n}_{-}$, and exhaust the eigenvectors of $\mathcal{A}$.

- $\det L'_{\Phi(X)}(0)$, or $\det\mathcal{A}$, is given by the product of the eigenvalues of $\mathcal{A}$. Thus,

$$\det L'_{\Phi(X)}(0) = \prod_{j<k}(1-i\omega_j)(1-i\omega_k).$$

Evidently, each factor $(1-i\omega_j)$ appears $(n-1)$ times in the product on the right-hand side. (There are $n(n-1)/2$ terms in the product, and each term contains two factors, so the number of factors is $n(n-1)$. The product is necessarily symmetric in the factors $(1-i\omega_1)$ through $(1-i\omega_n)$.) It follows that

$$\prod_{j<k}(1-i\omega_j)(1-i\omega_k) = \left(\prod_{j=1}^n(1-i\omega_j)\right)^{n-1}.$$

But the $(1-i\omega_j)$'s are just the eigenvalues of $I-X$, so that

$$\left(\prod_{j=1}^n(1-i\omega_j)\right)^{n-1} = (\det(I-iX))^{n-1}.$$

It follows that

$$\rho_L(X) = \frac{C}{\det L'_{\Phi(X)}(0)} = C(\det(I-iX))^{1-n},$$

where C is a constant independent of X.

6 Representations: Basic properties

6.1 Definition of representation

Let V be a complex vector space, $L(V)$ the space of linear maps on V, and $GL(V)\subset L(V)$ the group of invertible linear maps on V.

For example, we may take $V=\mathbb{C}^n$, in which case $L(V)=\mathbb{C}^{n\times n}$ and $GL(V)=GL(n,\mathbb{C})$. It is useful to keep this concrete realisation in mind, in which elements of $L(V)$ are $n\times n$ complex matrices and elements of $GL(V)$ are invertible complex matrices. However, it is also useful to allow for a more general point of view. There are examples where V isn't naturally identified with $\mathbb{C}^n$. For example, the set of homogeneous polynomials of degree n in three complex variables z_1, z_2, z_3 consists of linear combinations with complex coefficients of the monomials $z_1^{m_1}z_2^{m_2}z_3^{m_3}$ with $m_1+m_2+m_3=n$. These form a complex vector space of dimension $(n+1)(n+2)/2$, but there is no canonical identification with $\mathbb{C}^{(n+1)(n+2)/2}$.

Definition 6.1 (Group representations). Let G be a group (we do not yet restrict to matrix groups, nor matrix Lie groups), and let V be a complex vector space. A **complex representation** (Γ, V) of G, or **rep'n** for short, is a group homomorphism from G to $GL(V)$. That is, we have a map

$$\Gamma : G \to GL(V); g \mapsto \Gamma(g)$$

which satisfies

$$\Gamma(g)\Gamma(h) = \Gamma(gh).$$

V is often called the **carrier space** of the representation, $\dim V$ is the **dimension** of the representation. If V is a real vector space, (Γ, V) is said to be a **real representation**.

Example 6.2 (Examples of representations).

a) Natural representation. Let $G \subset \mathbb{C}^{n\times n}$ be a matrix group. The map $A \in G \mapsto \Gamma(A) := A$ yields a representation of G on $\mathbb{C}^n$. That is, a matrix group can be regarded as a representation of itself. This is sometimes called the natural representation.

b) Adjoint representation. Let G be a matrix Lie group. Let $\mathfrak{g}$ denote its Lie algebra. We have seen that $\mathfrak{g}$ is a real vector space. Given $A \in G$, let $\Gamma(A) = \mathrm{Ad}_A$. That is, for $\alpha \in \mathfrak{g}$, we take $\Gamma(A)\alpha = A\alpha A^{-1}$. We have seen that $\mathrm{Ad}_A \mathrm{Ad}_B = \mathrm{Ad}_{AB}$, so that Γ is a representation of G, called the Adjoint representation. The Adjoint representation is a real representation. A particular example is the map

$$R : \mathrm{SU}(2) \to \mathrm{SO}(3); u \mapsto R(u),$$

which we discussed in Section 4.8. R may be regarded as a real representation of $SU(2)$ on $\mathbb{R}^3$.

c) Representations and group actions. Here is a very general context for representations, which helps explain why they are important. Groups often arise as groups of bijections on some given set. For example, the permutation group may be thought of as the group of bijections of a finite set. More generally, let G be a group, and S as set. Let $B(S)$ denote the set of bijections from S to itself. $B(S)$ is itself a group under composition of maps. An **action** of G on S is a homomorphism $\Phi : G \to B(S); g \mapsto \Phi_g$. That is,

$$\Phi_g \circ \Phi_h = \Phi_{gh}.$$

A representation is a special kind of action, in which the set S is a vector space and the bijections on S are restricted to be linear.

However, given any action of G, whether it is a representation or not, there is a natural way to construct an associated (complex) representation of G. This is achieved by taking the vector space V to be the set of complex-vaued functions on S, which we denote $\mathcal{F}(S, \mathbb{C})$ (these naturally constitute a complex vector space). Given $g \in G$, we define $\Gamma(g)$ as follows: For $f \in \mathcal{F}(S, \mathbb{C})$, we take

$$(\Gamma(g)f)(x) := f(\Phi_{g^{-1}}(x)), \quad \text{for all } x \in S.$$

It is easy to see that Γ is a representation. Representations often arise in this way.

Of course, if S is infinite, then $\mathcal{F}(S, \mathbb{C})$ is infinite dimensional. One can consider infinite-dimensional representations, too, but we shall confine our attention mostly to finite-dimensional representations.

Let (Γ, V) be a representation, and let $\phi : V \to W$ be an invertible linear map from V to another vector space W. Let

$$\Delta(g) = \phi\Gamma(g)\phi^{-1}.$$

It is easy to see that (Δ, W) is also a representation, which is essentially the same as (Γ, V) (in effect, they differ by a change of basis). This motivates the following definition:

Definition 6.3 (Intertwining maps and equivalent representations.). Let (Γ_1, V_1) and (Γ_2, V_2) be representations of G. An **intertwining map** for (Γ_1, V_1) and (Γ_2, V_2) is a linear map $\phi : V_1 \to V_2$ such that

$$\Gamma_2(g)\phi = \phi\Gamma_1(g)$$

for all $g \in G$. The zero map, $\phi = 0$, is trivially an intertwining map. At the other extreme, if ϕ is invertible, we say that (Γ_1, V_1) and (Γ_2, V_2) are **equivalent**.

A representation Γ is **faithful** if Γ is 1-1; equivalently, $\ker\Gamma = e$, where $e \in G$ is the identity. In this case, G is isomorphic to its image under Γ; nothing is lost, as far as the group structure is concerned, in passing from the group to the representation. Obviously, the natural representation Example 6.2 a) is faithful. The representation R of Example 6.2 b) is not; its kernel is $\{I_2, -I_2\}$.

6.2 Irreducible Representations

Let (Γ, V) be a representation of G.

A subspace $W \subset V$ is a **proper** subspace if $W \neq \{0\}$ and $W \neq V$. A subspace $W \subset V$ is **Γ-invariant**, or **invariant**, if

$$\Gamma(g)W = W, \quad \text{for all } g \in G.$$

That is, for all $w \in W$,

$$\Gamma(g)w \in W, \quad \text{for all } g \in G.$$

Trivially, the 0-subspace and V itself are Γ-invariant. Of more interest are proper invariant subspaces. If $W \subset V$ is a proper Γ-invariant subspace and a basis for V is chosen so that the first m elements span W, then with respect to this basis, $\Gamma(g)$ has the following form:

$$\Gamma(g) = \left(\begin{array}{c|c} A(g) & B(g) \\ \hline 0 & C(g) \end{array}\right),$$

where the partition into blocks is given by $n = m + (n - m)$. It is straightforward to show that $A(g)$ and $C(g)$ constitute representations in their own right, of dimension m and $n - m$ respectively.

Definition 6.4 (Reducible and Irreducible representations.)**.** Let (Γ, V) be a real or complex representation. If there exists a proper Γ-invariant subspace $W \subset V$, then Γ is said to be **reducible**. Reducible representations give rise to representations of smaller dimension, as the example above shows. If V contains no proper invariant subspaces, then (Γ, V) is said to be **irreducible**.

Irreducible representations are basic elements from which other representations may be constructed. A particularly simple situation is the following:

Definition 6.5 (Completely reducible representation.)**.** A (real or complex) representation (Γ, V) is **completely reducible** if there exists Γ-invariant subspaces $W_1, \dots, W_r$ such that

i) $V = W_1 \oplus \cdots \oplus W_r$, i.e. $W_1, \dots, W_r$ span V and $W_j \cap W_k = \{0\}$ for $j \neq k$,

ii) Γ restricted to W_j is irreducible.

If (Γ, V) is completely reducible, there exists a basis for V in which Γ is block-diagonal,

$$\left(\begin{array}{c|c|c|c} \Gamma_1(g) & 0 & 0 & 0 \\ \hline 0 & \Gamma_2(g) & 0 & 0 \\ \hline 0 & 0 & \ddots & 0 \\ \hline 0 & 0 & 0 & \Gamma_r(g) \end{array}\right),$$

and $\Gamma_j(g)$ is an irreducible representation of G.

For certain groups, every representation (including infinite-dimensional representations, properly defined) is completely reducible. Then the study of representations is reduced to the study of (inequivalent) irreducible representations. This is the case for compact matrix Lie groups, which we will define towards the end of this set of notes. For compact groups, it turns out that the irreducible representations are all finite dimensional.

Example 6.6.

Here is an example of a representation which is reducible but not completely reducible. Let $G = \mathbb{R}$, the real numbers under addition. Let

$$\Gamma(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

It is easy to check that $\Gamma(s)\Gamma(t) = \Gamma(t)\Gamma(s) = \Gamma(s+t)$. We regard $\Gamma(t)$ as a representation on $V = \mathbb{C}^2$. Clearly

$$W_1 = \operatorname{span}\{e_1\}, \quad e_1 = (1,0)^\dagger$$

is a proper invariant subspace, since $\Gamma(t)e_1 = e_1$ for all t. However, $\Gamma(t)$ is not completely reducible. If it were, there would be a complementary invariant one-dimensional subspace W_2, spanned by a vector $f \in \mathbb{C}^2$. W_2 invariant is equivalent to f being an eigenvector of $\Gamma(t)$. But for $t \neq 0$, $\Gamma(t)$ has just one linearly independent eigenvector, e_1 (that is, $\Gamma(t)$ is not diagonalisable).

The following shows that there are no nontrivial intertwinings between irreducible representations.

Theorem 6.7 (Schur's Lemma)**.**

Let (Γ_1, V_1), (Γ_2, V_2) be representations of G. Let $\phi : V_1 \to V_2$ be an intertwining map, i.e. a linear map satisfying

$$\Gamma_2(g)\,\phi = \phi\,\Gamma_1(g), \quad \forall g \in G.$$

If Γ_1 is irreducible, then either $\phi = 0$ or ϕ is injective. If Γ_2 is irreducible, then either $\phi = 0$ or ϕ is surjective. If both Γ_1 and Γ_2 are irreducible, then either ϕ is invertible, in which case Γ_1 and Γ_2 are equivalent, or else $\phi = 0$.

Proof.

Let $\ker\phi \subset V_1$ denote the kernel of ϕ. We claim that $\ker\phi$ is Γ_1-invariant. To see this, let $v \in \ker\phi$. We need to show that $\Gamma_1(A)v \in \ker\phi$ for all $A \in G$, i.e. that $\phi(\Gamma_1(A)v) = 0$. We have that

$$\phi\,\Gamma_1(A)\,v = \Gamma_2(A)\phi\,v = 0,$$

as required. As Γ_1 is irreducible, it follows that either $\ker\phi = \{0\}$, in which case ϕ is 1-1, or else $\ker\phi = V_1$, in which case $\phi = 0$.

Let $\operatorname{im}\phi \subset V_2$ denote the image of ϕ. We claim that $\operatorname{im}\phi$ is Γ_2-invariant. To see this, let $w \in \operatorname{im}\phi$. We need to show that $\Gamma_2(A)w \in \operatorname{im}\phi$ for all $A \in G$. That is, given that $w = \phi v$ for some $v \in V_1$ (this is what it means for w to be in the image of ϕ), we need to show that $\Gamma_2(A)w = \phi v'$ for some $v' \in V_1$. We have that

$$\Gamma_2(A)w = \Gamma_2(A)\phi\,v = \phi\Gamma_1(A)\,v = \phi v',$$

where $v' = \Gamma_1(A)v$. Thus, $\operatorname{im}\phi$ is Γ_2-invariant, as required. As Γ_2 is irreducible, it follows that either $\operatorname{im}\phi = \{0\}$, in which case either $\phi = 0$, or else $\operatorname{im}\phi = V_2$, in which case ϕ is onto.

If both Γ_1 and Γ_2 are irreducible, then either $\phi = 0$ or else ϕ is both 1-1 and onto, in which case ϕ is invertible. □

6.3 *Criteria for irreducibility

For this discussion, it will sometimes be convenient to assume that V has a hermitian inner product, denoted $\langle , \rangle$ (this assumption isn't essential). Given a hermitian inner product and $A \in L(V)$, we define $A^\dagger \in L(V)$, the hermitian conjugate of A, as follows:

$$\left\langle u, A^\dagger v \right\rangle = \langle Au, v\rangle, \quad \text{for all } u, v \in \mathbb{C}.$$

For $V = \mathbb{C}^n$, this coincides with $(A^\dagger)_{jk} = A^*_{kj}$.

A **unital algebra** $\mathcal{A} \subset L(V)$ is a complex subspace of $L(V)$ that contains the identity I_V and is closed under multiplication. That is, $I_v \in \mathcal{A}$ and if $A, B \in \mathcal{A}$ and $c \in \mathbb{C}$, then cA, $A + B$ and AB all belong to $\mathcal{A}$.

We can define invariant subspaces and reducibility for unital algebras of $L(V)$ in analogy with the definitions for representations: A subspace $W \subset V$ is **invariant** under $\mathcal{A}$ if $\mathcal{A}W = W$, ie

$$x \in W \implies Ax \in W, \quad \text{for all } A \in \mathcal{A}.$$

$\mathcal{A}$ is **irreducible** if it has no proper invariant subspaces.

The principal example of a unital algebra that we shall have in view is the following: Let G be a group, and (Γ, V) a complex representation of G. The **representation algebra** of Γ, denoted $\mathcal{A}_\Gamma$, is the set of finite linear combinations of the $\Gamma(g)$'s; that is,

$$\mathcal{A}_\Gamma = \left\{ \sum_{g_j \in G} c_{g_j} \Gamma(g_j) \,|\, c_{g_j} \in \mathbb{C} \right\}.$$

The fact that $\Gamma(g)$ is a representation implies that $\mathcal{A}_\Gamma$ is a unital subalgebra of $L(V)$ – the main point is that $\Gamma(g_j)\Gamma(g_k) = \Gamma(g_j g_k)$.

It is easy to see that W is invariant under Γ if and only if it is invariant under $\mathcal{A}_\Gamma$. Therefore, Γ is irreducible if and only if $\mathcal{A}_\Gamma$ is irreducible.

Given $\mathcal{A} \subset L(V)$ be unital algebra, let

$$\mathcal{A}^\dagger = \{A^\dagger \,|\, A \in \mathcal{A}\}.$$

Proposition 6.8.

i) $\mathcal{A}^\dagger$ is a unital algebra of $L(V)$.

ii) If $\mathcal{A}$ is irreducible, then $\mathcal{A}^\dagger$ is irreducible.

Proof.

i) It is clear that $A^\dagger$ is a vector subspace of $L(V)$ closed under multiplication. Indeed, suppose $F, G \in \mathcal{A}^\dagger$, so that $F^\dagger, G^\dagger \in \mathcal{A}$. Then $F + G \in \mathcal{A}^\dagger$, since $F + G = (F^\dagger + G^\dagger)^\dagger$, and $FG \in \mathcal{A}^\dagger$, since $FG = (G^\dagger F^\dagger)^\dagger$.

ii) Suppose $W \subset V$ is invariant under $\mathcal{A}^\dagger$. Let $W^\perp$ denote the orthogonal complement of W, i.e.

$$W^\perp = \{x \in V \,|\, \langle x, w\rangle = 0, \quad \forall w \in W\}.$$

We claim that $W^\perp$ is invariant under $\mathcal{A}$. That is, if $x \in W^\perp$, then $A\,x \in W^\perp$ for all $A \in \mathcal{A}$. To see this, take $w \in W$, and consider

$$\langle Ax, w\rangle = \left\langle x, A^\dagger w\right\rangle = \left\langle x, w'\right\rangle = 0,$$

since $w' := A^\dagger w \in W$.
Therefore, if $\mathcal{A}$ is irreducible and W is invariant under $\mathcal{A}^\dagger$, it follows that $W^\perp$ is invariant under $\mathcal{A}$, which implies that $W^\perp = \{0\}$ or V, which implies that $W^\perp = V$ or $\{0\}$.

□

Theorem 6.9 (Critieria for irreducibility)**.**

Let $\mathcal{A} \subset L(V)$ be a unital algebra. Then the following are equivalent:

(a) $\mathcal{A}$ is irreducible.

(b) Every $v \in V$ is *cyclic*, i.e.

$$\mathcal{A}\,v := \{A\,v \,|\, A \in \mathcal{A}\} = V.$$

(c)

$$\mathcal{A} = L(V).$$

For the proof, it is convenient to introduce **rank-one linear maps**. Let $r, s \in V$, and define $Q_{rs} \in L(V)$ by

$$Q_{rs}\,v = \langle s, v\rangle\, r.$$

Proof.

(a) $\Longleftrightarrow$ (b). If W is a proper invariant subspace under $\mathcal{A}$, then vectors in W are not cyclic. Conversely, if v is not cyclic, that $\mathcal{A}\,v = \{A\,v \,|\, A \in \mathcal{A}\}$ is a proper invariant subspace of V.

(c) $\Longrightarrow$ (b). Let $v \in V$, $v \neq 0$. To show that v is cyclic, it suffices to show that, given $w \in V$, there exists a linear map $T \in \mathcal{A}$ such that $Tv = w$. Such a map is given by

$$T = \frac{1}{\langle v, v\rangle} Q_{wv},$$

and since $\mathcal{A} = L(V)$ by assumption, it follows that $T \in \mathcal{A}$.

(a) $\Longleftrightarrow$ (c). It is clear that (c) implies (a). We show that (a) implies (c). It will be enough to show that if $\mathcal{A}$ is irreducible, then $\mathcal{A}$ contains all rank-one linear maps Q_{rs} for $r, s \in V$. This is because every linear map can be expressed as a linear combination of rank-one maps. Indeed, if e_j is an orthonormal basis for V, then for $M \in L(V)$,

$$M = \sum_{j,k=1}^{n} \langle e_j, Me_k\rangle\, Q_{e_k e_j}.$$

We proceed by induction on $\dim V$. If $\dim V = 1$, then $L(V) = \mathbb{C}$, and the result is clear. Let us suppose that $\dim V > 1$, and that (a) implies (c) for all vector spaces of dimension less than $\dim V$. Suppose that $\mathcal{A}$ is irreducible. Then $\mathcal{A}$ contains elements besides those of the form cI_V. Let T be such an element.

Let $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ denote the distinct eigenvalues of T, i.e. the roots of the characteristic polynomial of T. Since $T \neq cI_V$, then without loss of generality, we may assume that $r \geq 2$. Here is the argument : If $T \neq cI_V$ and T has a single eigenvalue c, then $T = cI + N$, where $N \neq 0$ is nilpotent. We construct a new map, $S \in \mathcal{A}$, which has at least two distinct eigenvalues as follows: Choose $v \in V$ so that $w := Nv \neq 0$. Since $\mathcal{A}$ is irreducible, w is cyclic, so there exists $A \in \mathcal{A}$ such that $Aw = v$. Let $S = AN$. Since $A, N \in \mathcal{A}$, it follows that $S \in \mathcal{A}$. But 0 is an eigenvalue of S, since $\det N = 0$, and 1 is an eigenvalue of S, since $Sv = v$.

It is a basic result in linear algebra (primary decomposition theorem – we give details in the appendix) that there exist maps $P_1, \ldots, P_r \in L(V)$ and associated subspaces $V_j := P_j V$ with the following properties:

(i) The P_j's are **projectors**, i.e. $P_j^2 = P_j$.

(ii) $P_j P_k = 0$ for $j \neq k$; equivalently, $V_j \cap V_k = \{0\}$ for $j \neq k$.

(iii) $\sum_{j=1}^{r} P_j = I_V$; equivalently, $V = V_1 \oplus \cdots \oplus V_r$.

(iv) $(T - \lambda_j)^p V_j = 0$ for large enough p; equivalently, the V_j's are the generalised eigenspaces of T.

(v) The P_j's can be expressed as polynomials in T (this property will be crucial).

Thus, the P_j's are projectors onto the disjoint generalised eigenspaces $V_j = P_j V$ of T, and the generalised eigenspaces span V.

Since P_1 is a polynomial in T, it follows that $P_1 \in \mathcal{A}$ (note that a polynomial in T is a linear combination of I_V and powers of T – since $\mathcal{A}$ is a unital algebra, it contains such a polynomial). Consider the subalgebra $\mathcal{A}_1$ of $L(V_1)$ that consists of elements of the form

$$\mathcal{A}_1 = \{P_1 A P_1 \mid A \in \mathcal{A}\}.$$

We claim that $\mathcal{A}_1$ is irreducible on V_1. This follows from that fact that every $v_1 \in V_1$ is cyclic under $\mathcal{A}_1$, as we now show: We have that

$$\mathcal{A}_1 \, v_1 = P_1 \mathcal{A} P_1 \, v_1 = P_1 \mathcal{A} \, v_1.$$

Since we assume $\mathcal{A}$ is irreducible, (b) implies that $\mathcal{A} v_1 = V$, so that

$$\mathcal{A}_1 \, v_1 = P_1 V = V_1,$$

as required.

By the induction hypothesis, it follows that $\mathcal{A}_1 = L(V_1)$. In particular, take $t \in V_1$, and let Q_{1tt} denote the rank-one projector in $L(V_1)$,

$$Q_{1tt} \, v = \langle t, v \rangle \, t.$$

Then $Q_{1tt} \in \mathcal{A}_1$. Q_{1tt} can be extended to the rank-one projector Q_{tt} defined on all of V by setting $Q_{tt} V_j = 0$ for $j \neq 1$. It follows that $Q_{tt} = P_1 A P_1$ for some $A \in \mathcal{A}$. Since by property v), $P_1 \in \mathcal{A}$, it follows that $Q_{tt} \in \mathcal{A}$.

Since $\mathcal{A}$ is irreducible, t is cyclic. Given $r \in V$, there exists $A \in \mathcal{A}$ such that $A\, t = r$. From Proposition 6.8, $\mathcal{A}^\dagger$ is irreducible; therefore, there exists $B \in \mathcal{A}$ such that $B^\dagger t = s$. Then $A Q_{tt} B \in \mathcal{A}$. Claim that $A Q_{tt} B = T_{rs}$. Indeed, we have that

$$A Q_{tt} B \, v = \langle t, Bv \rangle A \, t = \left\langle B^\dagger \, t, v \right\rangle r = \langle s, v \rangle r = Q_{rs} v.$$

Thus, $Q_{rs} \in \mathcal{A}$, as we wanted to show.

□

6.4 *Appendix. Primary Decomposition Theorem

The primary decomposition is a basic result in linear algebra. References may be readily found, either in the library or on the web. Here is a brief account for the case of matrices in $\mathbb{C}^{n \times n}$. More general versions are formulated for matrices over fields other than $\mathbb{C}$. A simplifying feature of the complex case is that $\mathbb{C}$ is algebraically closed; every polynomial over $\mathbb{C}$ has at least one root (and, therefore, as many roots as its degree, provided roots are counted with multiplicity).

Let $A \in \mathbb{C}^{n \times n}$ be a complex matrix. The matrices I, A, $A^2, \ldots, A^j, \ldots$ all belong to $\mathbb{C}^{n \times n}$, and hence cannot all be linearly independent. Thus, there exists some linear combination of powers of A which vanishes. Equivalently, there exists a polynomial

$$p(z) = c_0 + c_1 z + \cdots + c_m z^m$$

such that $p(A) = 0$. In fact, if $P(z)$ is the characteristic polynomial of A, i.e.

$$P(z) = \det(A - zI),$$

then $P(A) = 0$; this is the Cayley-Hamilton theorem. Let $M(z)$ be the monic polynomial of lowest degree for which $M(A) = 0$ (monic means the coefficient of the highest-order terms is equal to one). $M(z)$ is called the **minimum polynomial** of A. Since the characteristic polynomial has degree n, it follows that $\deg M \leq n$. Also, $P(z)$ must be divisible by $M(z)$ (otherwise, we could write $P(z) = f(z)M(z) + r(z)$, where $\deg r < \deg M$ and $r(A) = 0$, contradicting the assumption that M is the minimum polynomial).

Let $d = \deg M(z)$. Then $M(z)$ has $r \leq d$ roots, $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$, with multiplicities $d_1, \ldots, d_r$, such that $\sum_j d_j = d$ and

$$M(z) = \prod_{j=1}^{r} (z - \lambda_j)^{d_j}.$$

Let

$$f_j(z) = \prod_{k=1, k\neq j}^{r} (z-\lambda_k)^{d_k}.$$

It is clear that the $f_j(z)$'s have no common root. It follows (an argument is given below) that there exist polynomials $g_1, \ldots, g_r$ such that

$$\sum_{j=1}^{r} g_j(z) f_j(z) = 1. \tag{22}$$

Let

$$P_j = g_j(A) f_j(A).$$

Clearly P_j is a polynomial in A. From (22),

$$\sum_{j=1}^{r} P_j(M) = \sum_{j=1}^{r} g_j(M) f_j(M) = I_n. \tag{23}$$

Also, since $f_j(z) f_k(z)$ is divisible by $M(z)$ for $j \neq k$, ie $f_j g_k = qM$ for some polynomial q, it follows that

$$P_j P_k = g_j(A) g_k(A) f_j(A) f_k(A) = g_j(A) g_k(A) q(A) M(A) = 0, \quad \text{if } j \neq k. \tag{24}$$

Multiplying by P_j in (23) and taking account of (24), we get that

$$P_j^2 = P_j.$$

Since $(z-\lambda_j)^{d_j} f_j(z) = M(z)$, it follows that

$$(A - \lambda_j I_n)^{d_j} P_j = (A - \lambda_j)^{d_j} f_j(A) g_j(A) = M(A) g_j(A) = 0.$$

This establishes the results of the Primary Decomposition Theorem.

It remains to establish (22). This follows from the following fact: Let p and q be monic polynomials, and let d denote the monic polynomial of highest degree that divides both p and q. We call d the greatest common factor of p and q, and write

$$d = (p, q).$$

We claim that there exist monic polynomials a and b such that

$$ap + bq = d.$$

We proceed using Euclid's algorithm. Given polynomials p_j and q_j with $\deg q_j \leq p_j$, we may write

$$p_j = c_j q_j + \alpha_j r_j,$$

c_j is a monic polynomial, $\alpha_j \in \mathbb{C}$, and the remainder r_j is a monic polynomial of degree less than $\deg q_j$. Define

$$p_{j+1} = q_j, \quad q_{j+1} = r_j.$$

It is clear that

$$(p_{j+1}, q_{j+1}) = (p_j, q_j),$$

since any polynomial which divides p_j and q_j also divides p_{j+1} and q_{j+1}, and vice versa. It is clear that

$$\deg q_{j+1} < \deg q_j.$$

Iterating this procedure, starting with $p_1 = p$, $q_1 = q$ (without loss of generality, we may assume that $\deg q \leq \deg p$), we eventually get to an iteration n where $q_n = 0$. Then

$$(p_n, q_n) = (p_n, 0) = d.$$

We may write that

$$\begin{pmatrix} d \\ 0 \end{pmatrix} = \begin{pmatrix} p_n \\ q_n \end{pmatrix} = D_{n-1} \cdots D_1 \begin{pmatrix} p \\ q \end{pmatrix},$$

where the D_j's are 2×2 matrices with polynomial entries of the form

$$D_j = \begin{pmatrix} 0 & 1 \\ 1/\alpha_j & -c_j \end{pmatrix}.$$

We note that D_j^{-1} is given by

$$D_j^{-1} = \begin{pmatrix} \alpha_j c_j & \alpha_j \\ 1 & 0 \end{pmatrix},$$

and therefore is also a matrix with polynomial entries. Then

$$\begin{pmatrix} p \\ q \end{pmatrix} = D_1^{-1} \cdots D_{n-1}^{-1} \begin{pmatrix} d \\ 0 \end{pmatrix}$$

gives $d = (p, q)$ in the form $ap + bq$ for some polynomials a and b.

Given a set of, say, three monic polynomials f_1, f_2, f_3 with no common root, let $d_{12} = (f_1, f_2)$ We can find polynomials a_1 and a_2 such that

$$d_{12} = a_1 f_1 + a_2 f_2.$$

Since $(d_{12}, f_3) = 1$, we can also find polynomials b_{12} and b_3 such that

$$1 = b_{12} d_{12} + b_3 d_3.$$

Then

$$1 = b_{12} a_1 f_1 + b_{12} a_2 f_2 + b_3 f_3$$

yields the required resolution of unity. The argument generalises by induction to n monic polynomials with no common root.

Corollary 6.10.

Let $\mathcal{A}$ be an irreducible subalgebra of $L(V)$. If $M \in L(V)$ commutes with every element of $\mathcal{A}$, then M is a multiple of the identity.

Proof.

From Theorem 7.5, $\mathcal{A} = L(V)$. In particular, for arbitrary $r \in V$, $Q_{rr} \in \mathcal{A}$ (recall that Q_{rs} is defined by $Q_{rs} v = \langle s, v \rangle r$). Then $M Q_{rr} = Q_{rr} M$ implies that Mr is proportional to r, so that every vector is an eigenvector of M. This implies that $M = cI_V$ for some $c \in \mathbb{C}$.

□

Corollary 6.11.

Let (Γ, V) be an irreducible representation of a group G on a complex vector space V. Then the following are equivalent:

(a) Γ is irreducible.

(b) Every $v \in V$ is *cyclic*, i.e. $\{\Gamma(g)v \,|\, g \in G\}$ spans V.

(c)
$$\mathcal{A}_\Gamma = L(V),$$
where
$$\mathcal{A}_\Gamma = \left\{ \sum_{g_j \in G} c_{g_j} \Gamma(g_j) \,|\, c_{g_j} \in \mathbb{C} \right\}.$$

Thus, if Γ is irreducible and if $\phi \in L(V)$ commutes with every $\Gamma(g)$, then ϕ is a multiple of the identity.

Note this last result may be regarded as a strengthening of Schur's Lemma for the particular case of complex irreducible representations. If ϕ commutes with $\Gamma(A)$, it may be regarded as an intertwining map from (Γ, V) to itself. If Γ is irreducible and $\phi \neq 0$, then Schur's Lemma (Theorem 6.7) implies that ϕ is invertible. Corollary 6.11, which uses the fact that V is complex, implies that ϕ is a multiple of the identity map. (In Schur's Lemma, we do not assume that the vector space is complex.)

The following example shows that Corollary 6.11 need not hold for real representations (i.e., representations defined on a real vector space).

Example 6.12. Let

$$\mathrm{SO}(2) = \left\{ R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \,\middle|\, 0 \leq \theta < 2\pi \right\}.$$

SO(2) is the group of rotations in $\mathbb{R}^2$. It may be regarded as a one-dimensional matrix Lie group, and as such constitutes a representation on $\mathbb{R}^2$. It is clear SO(2) is irreducible; there is no line through the origin in the plane that is invariant under rotations. However, SO(2) is abelian; indeed,

$$R(\theta_1)R(\theta_2) = R(\theta_2)R(\theta_1) = R(\theta_1 + \theta_2).$$

Thus, every $R(\theta)$ commutes with all of SO(2), but clearly $R(\theta)$ is not a multiple of the identity in general. Likewise, $\mathcal{A}_{\mathrm{SO}(2)} \neq L(\mathbb{R}^2)$. Indeed,

$$\mathcal{A}_{\mathrm{SO}(2)} = \left\{ aI_2 + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \,\middle|\, a, b \in \mathbb{R} \right\}.$$

Note, too, that if SO(2) is regarded as a complex representation over $V = \mathbb{C}^2$, then it is reducible. Indeed, $(1, \pm i)^\dagger$ are eigenvectors of every $R(\theta)$, and therefore span invariant subspaces.

6.5 Representations of compact groups

Definition 6.13 (Compact matrix Lie groups.)**.** A compact matrix Lie group G is <u>**compact**</u> if, for some $M > 0$,

$$||A|| \leq M, \quad \forall A \in G.$$

In particular, a finite matrix Lie group is compact. The unitary groups $U(n)$ are compact, since for $u \in U(n)$,

$$||u||^2 = \mathrm{Tr}(u^\dagger u) = \mathrm{Tr}(I_N) = N.$$

Subgroups of $U(n)$ are also compact. These include $O(n)$, the group of real orthogonal $n \times n$ matrices, since $O(n)$ may be regarded as the subgroup of real unitary matrices.

Definition 6.13 is compatible with the usual notion of compactness for a subset of a finite-dimensional vector space (in this case, the vector space is $\mathbb{C}^{n\times n}$), namely that the subset be closed and bounded. The point is that a matrix Lie group G whose elements are bounded in norm is automatically closed. To see this, we recall that G is closed relative to $GL(n, \mathbb{C})$, from Proposition 1.10. To show that G is closed, and not just closed relative to $GL(n, \mathbb{C})$, it suffices to show that for all $A \in G$,

$$\frac{1}{C} \leq |\det A| \leq C \tag{25}$$

for some $C > 0$.

We have that

$$|\det A|^2 = \overline{\det A}\, \det A = \det A^\dagger\, \det A = \det A^\dagger A.$$

The matrix $A^\dagger A$ is hermitian and positive semidefinite, and therefore has nonnegative real eigenvalues $\lambda_1, \ldots, \lambda_n$. Noting that $\det A^\dagger A$ is given by the product of the λ_j's while $||A||^2 = \mathrm{Tr}\, A^\dagger A$ is given by the sum of the λ_j's, we get that

$$|\det A|^2 = \prod_{j=1}^n \lambda_j \leq \left(\frac{1}{n}\sum_{j=1}^n \lambda_j\right)^n = \left(\frac{1}{n}||A||^2\right)^n \leq \left(\frac{M^2}{n}\right)^n,$$

where we have used the <u>inequality of arithmetic and geometric means</u>. Since $A^{-1} \in G$, we also have that

$$|\det A^{-1}|^2 \leq \left(\frac{M}{\sqrt{n}}\right)^{2n}.$$

But $\det A^{-1} = 1/\det A$, so that

$$\frac{1}{|\det A|^2} \leq \left(\frac{M}{\sqrt{n}}\right)^{2n},$$

or

$$|\det A|^2 \geq \left(\frac{\sqrt{n}}{M}\right)^{2n}.$$

So we can take $C = (M/\sqrt{n})^n$ in (25).

There is a very nice, complete representation theory for compact groups.

The essential property of compact groups is that they have a left- and right-invariant Haar measure that can be normalised. "Normalised" means that the measure can be chosen so that the integral of the function which is everywhere equal to 1 can be taken to be 1. In this case, integration of a function on the group can be viewed as computing its average value.

In what follows, we consider complex representations (Γ, V). Without loss of generality, we may assume that V has a hermitian inner product.

Definition 6.14 (Hermitian inner product and adjoint.)**.** A **hermitian inner product** on a complex vector space V is a sesquilinear form

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}; (u, v) \mapsto \langle u, v \rangle$$

which satisfies the following:

i) $\langle u, v \rangle = \langle v, u \rangle^*$

ii) $\langle u, u \rangle \geq 0$, and $\langle u, u \rangle = 0$ if and only $u = 0$

Note: "Sesquilinear" means linear with respect to one argument and antilinear with respect to the other. That is,

$$\langle u, c_1 v_1 + c_2 v_2 \rangle = c_1 \langle u, v_1 \rangle + c_2 \langle u, v_2 \rangle,$$
$$\langle c_1 u_1 + c_2 u_2, v \rangle = c_1^* \langle u_1, v \rangle + c_2^* \langle u_2, v \rangle.$$

Given a linear map $L \in L(V)$, we define its **adjoint**, or **hermitian conjugate**, denoted $L^\dagger$, by

$$\langle u, L^\dagger v \rangle := \langle Lu, v \rangle.$$

It is straightforward to verify that these definitions coincide with the ones given in Section 1.1 when $V = \mathbb{C}^n$.

Definition 6.15 (Unitary representations)**.** Let V be a complex inner product with hermitian inner product $\langle , \rangle$. A linear invertible map $U \in GL(V)$ is **unitary** if

$$\langle Uv, Uw \rangle = \langle v, w \rangle, \quad \text{for all } v, w \in V.$$

It follows that

$$U^\dagger U = I_V,$$

since

$$\left\langle u, U^\dagger U v \right\rangle = \langle Uu, Uv \rangle = \langle u, v \rangle.$$

If $V = \mathbb{C}^n$ and $U \in GL(n, \mathbb{C})$, then the definition of unitary coincides with the definition of unitary matrix. A representation (Γ, V) of G is **unitary** if $\Gamma(A)$ is unitary for all $A \in G$.

Proposition 6.16. Let G be a compact matrix Lie group with representation (Γ, V). Then Γ is equivalent to a unitary representation.

Proof.

We will write the argument for finite-dimensional matrix Lie groups, but it carries over directly to finite matrix groups (zero-dimensional matrix Lie groups) with the Haar measure $\int_G d\mu(A)$ replaced by $|G|^{-1}\sum_{A\in G}$.

Define
$$T = \int_G \Gamma^\dagger(A)\Gamma(A)\, d\mu(A).$$
Then $T \in GL(V)$ is positive definite hermitian, i.e.

i) $T^\dagger = T$

ii) $\langle v, Tv\rangle \geq 0$,

iii) $\langle v, Tv\rangle = 0 \iff v = 0$.

Properties ii) and iii) follow from the fact that
$$\langle v, Tv\rangle = \int_G \left\langle v, \Gamma^\dagger(A)\Gamma(A)v\right\rangle d\mu(A) = \int_G ||\Gamma(A)v||^2\, d\mu(A),$$
and $\Gamma(A)v = 0$ iff $v = 0$.

A positive definite hermitian element $T \in GL(V)$ has a unique positive definite square root S, i.e.
$$S = S^\dagger, \quad S^2 = T.$$
(One way to see this is to note that T has a complete set of orthonormal eigenvectors e_j with positive eigenvalues λ_j, so that
$$Tv = \sum_{j=1} \lambda_j \langle e_j, v\rangle e_j.$$
Then we take
$$Sv = \sum_{j=1} \lambda_j^{1/2}\langle e_j, v\rangle e_j.)$$

Let
$$\Delta(A) = S\Gamma(A)S^{-1}.$$
Clearly $\Delta(A)$ is a representation of G equivalent to S. Claim that $\Delta(A_0)$ is unitary for all $A_0 \in G$. Indeed,
$$\begin{aligned}
&\Delta(A_0)^\dagger\Delta(A_0)\\
&= S^{-1}\Gamma(A_0)^\dagger S^2\Gamma(A_0)S^{-1} = S^{-1}\Gamma(A_0)^\dagger\, T\,\Gamma(A_0)S^{-1} = S^{-1}\Gamma(A_0)^\dagger\left(\int_G \Gamma^\dagger(A)\Gamma(A)\right)\Gamma(A_0)S^{-1}\, d\mu(A)\\
&= S^{-1}\left(\int_G \Gamma^\dagger(AA_0)\Gamma(AA_0)\, d\mu(A)\right)S^{-1}(\text{using the representation property})\\
&= S^{-1}\left(\int_G \Gamma^\dagger(A)\Gamma(A)\, d\mu(A)\right)S^{-1}\ (\text{using the invariance of the measure})\\
&= S^{-1}TS^{-1} = S^{-1}S^2S^{-1} = I_V.
\end{aligned}$$
□

Proposition 6.17. If (Γ, V) is a unitary representation, then Γ is completely reducible.

Proof.

We proceed by induction on $\dim V$. The assertion is trivial for $\dim V = 1$ (any 1-dimensional representation is irreducible). We suppose it holds for all representations (Γ, W) with $\dim W < \dim V$.

If (Γ, V) is already irreducible, then there is nothing more to show. If not, let $W \subset V$ be a proper Γ-invariant subspace. Let $W^\perp$ denote the orthogonal complement of W, i.e.
$$W^\perp = \{x \in V \mid \langle x, w\rangle = 0, \quad \forall w \in W.$$

We claim that $W^{\perp}$ is also Γ-invariant. To see this, let $x \in W^{\perp}$. We want to show that $\Gamma(g)x \in W^{\perp}$ for all g, i.e. that $\langle \Gamma(g)x, w \rangle = 0$ for all $w \in W$. But, since $\Gamma(g)$ is unitary, $\langle \Gamma(g)x, w \rangle = \left\langle x, \Gamma^{\dagger}(g)w \right\rangle = \left\langle x, \Gamma^{-1}(g)w \right\rangle = \left\langle x, \Gamma(g^{-1})w \right\rangle$. As W is Γ-invariant, $\Gamma(g^{-1})w \in W$, hence $\left\langle x, \Gamma(g^{-1})w \right\rangle = 0$, as required.

Clearly, $V = W \oplus W^{\perp}$; every $v \in V$ has a unique decomposition into components in W and $W^{\perp}$. The restrictions of Γ to W and $W^{\perp}$ are also unitary. Since $\dim W, \dim W^{\perp} < \dim V$, the complete reducibility of (Γ, W) and $(\Gamma, W^{\perp})$, and therefore of (Γ, W), follows by induction. □

Thus, from Propositions 6.16 and 6.17, it follows that every representation of a compact matrix Lie group is equivalent to a direct sum of irreducible unitary representations (there is an analogous decomposition for infinite-dimensional representations). Thus, the representation theory for compact matrix Lie groups reduces to the study of their unitary irreducible representations.

Proposition 6.18.

Let (Γ, V) and (Δ, W) be unitary irreducible representations of a compact matrix Lie group G. Let $T : V \to W$ be linear. Define the linear map $\phi : V \to W$ by

$$\phi = \int_G \Delta(A^{-1})\, T\, \Gamma(A)\, d\mu(A).$$

If (Γ, V) and (Δ, W) are inequivalent, then

$$\phi = 0.$$

If $(\Delta, W) = (\Gamma, V)$, then

$$\phi = \frac{\operatorname{Tr} T}{\dim V}\, I_V.$$

Proof.
Below we show that ϕ is an intertwining map for (Γ, V) and (Δ, W), i.e.

$$\Delta(A_0)\phi = \phi\Gamma(A_0).$$

for all $A_0 \in G$. Let us take this as given for now. By Schur's Lemma (Theorem 6.7), it follows that ϕ is either invertible or else $\phi = 0$.

Suppose that Δ and Γ are inequivalent. Then ϕ cannot be invertible, so that $\phi = 0$.

Suppose that $\Delta = \Gamma$. Then ϕ commutes with $\Gamma(A)$. By Corollary 6.11, $\phi = cI_V$. Taking traces, we get that $\operatorname{Tr}\phi = c \dim V$. But

$$\operatorname{Tr}\phi = \int_G \operatorname{Tr}\left(\Gamma(A^{-1})\, T\, \Gamma(A)\right)\, d\mu(A) = \int_G \operatorname{Tr} T\, d\mu(A) = \operatorname{Tr} T.$$

It follows that

$$\phi = \frac{\operatorname{Tr} T}{\dim V}\, I_V.$$

To proceed, we must establish that ϕ is indeed an intertwining. That is, we show that for all $A_0 \in G$,

$$\Delta(A_0)\phi = \phi\Gamma(A_0).$$

But

$$\begin{aligned}\Delta(A_0)\phi = \int_G \Delta(A_0)\Delta(A^{-1})\, T\, \Gamma(A)\, d\mu(A) &= \int_G \Delta((AA_0^{-1})^{-1})\, T\, \Gamma(A)\, d\mu(A) \\ &= \int_G \Delta(B^{-1})\, T\, \Gamma(BA_0)\, d\mu(B) = \int_G \Delta(B^{-1})\, T\, \Gamma(B)\Gamma(A_0)\, d\mu(B) = \phi\Delta(A_0).\end{aligned}$$

□

Given $u, v \in V$, we define the **matrix element** $\Gamma_{uv}(A)$ by

$$\Gamma_{uv}(A) = \langle u, \Gamma(A)v \rangle.$$

Proposition 6.18 implies that matrix elements of inequivalent unitary representations, regarded as functions on G, are orthogonal – this is the content of the next result.

Corollary 6.19.

Let (Γ, V) and (Δ, W) above be unitary. For all $u, v \in V$ and $x, y \in W$,

$$\int_G \Delta^*_{xy}(A)\Gamma_{uv}(A)\,d\mu(A) = \begin{cases} 0, & \text{if } \Delta \text{ and } \Gamma \text{ are inequivalent,} \\ (\dim V)^{-1}\langle x, u\rangle\langle y, v\rangle, & \text{if } \Delta = \Gamma. \end{cases}$$

Proof.

In the expression for ϕ in Proposition 6.18, let $T = Q_{xu}$, where

$$Q_{xu} r = \langle u, r\rangle x.$$

Then

$$\langle y, \phi\, v\rangle = \int_G \left\langle y, \Delta(A^{-1})\, Q_{xu}\, \Gamma(A) v\right\rangle d\mu(A).$$

Since Δ is unitary, $\Delta(A^{-1}) = \Delta^\dagger(A)$, and the summand is given by

$$\begin{aligned} \left\langle y, \Delta^\dagger(A)\, Q_{xu}\, \Gamma(A) v\right\rangle &= \langle \Delta(A) y, Q_{xu}\, \Gamma(A) v\rangle \\ &= \langle \Delta(A) y, x\rangle\, \langle u, \Gamma(A) v\rangle = \langle x, \Delta(A) y\rangle^*\, \langle u, \Gamma(A) v\rangle = \Delta^*_{xy}(A)\Gamma_{uv}(A). \end{aligned}$$

Thus,

$$\langle y, \phi\, v\rangle = \int_G \Delta^*_{xy}(A)\Gamma_{uv}(A)\,d\mu(A).$$

Referring to Proposition 6.18, if Γ and Δ are inequivalent, then $\phi = 0$. On the other hand, if $\Gamma = \Delta$, then

$$\langle y, \phi\, v\rangle = \langle y, c I_V v\rangle = \frac{1}{\dim V}\langle x, u\rangle\langle y, v\rangle.$$

since $\operatorname{Tr} Q_{xu} = \langle x, u\rangle$.

$\square$

So far, we have established properties of irreducible representations assuming that they exist. But are there any, and how would you construct them? It turns out that the inequivalent irreducible representations for a compact matrix Lie group can be constructed from polynomials in the matrix components, and that the irreducible representations span the space of square-integrable functions on the group.

Theorem 6.20 (*Peter-Weyl)**.**

Let $G \subset \mathbb{C}^{n\times n}$ be a compact matrix Lie group.

a) The irreducible representations of G are finite dimensional.

b) The set of inequivalent irreducible representations of G is countable.

c) Let (Γ^μ, V^μ), $\mu = 1, 2, 3, \ldots$ denote the set of inequivalent irreducible representations, which we may take to be unitary. Then for all $u, v \in V^\mu$, $\Gamma^\mu_{uv}(A)$ is a polynomial in the components of A and $A^\dagger$.

d) Let

$$\Gamma^\mu_{ij}(A) = \left\langle e^\mu_i, \Gamma^\mu(A) e^\mu_j\right\rangle.$$

Then the $(d^\mu)^{1/2}\Gamma^\mu_{ij}(A)$'s constitute an orthonormal basis for $L^2(G, \mathbb{C})$, where $d^\mu = \dim V^\mu$.

Sketch of proof.

1. Let $V^m \subset C(G, \mathbb{C})$ denote the set of homogenous polynomials of degree m in the components of A and $A^\dagger$ with complex coefficients. V^m is a finite-dimensional complex vector space. We define a hermitian inner product on V^m using Haar measure:

$$\langle p, q\rangle = \int_G \overline{p(A)} q(A)\,d\mu(A).$$

2. If $p \in V^m$, then $\Gamma(A_0)p \in V^m$, where $\Gamma(A_0)p \in V^m$ is given by

$$(\Gamma(A_0)p)(A) = p(AA_0)$$

The map $p \mapsto \Gamma(A_0)p$ defines a representation of G on V^m. (In fact, by the invariance of the Haar measure, it is straightforward to show that Γ is unitary.)

3. By Propositions 6.16 and 6.17, (Γ, V^m) is completely reducible. Let V^μ denote an irreducible subspace of V^m of dimension d. Let p_i denote an orthonormal basis for V^μ. Then

$$\Gamma(A_0)p_i = \sum_{j=1}^{d} \Gamma^\mu_{ji}(A_0)p_j,$$

and $\Gamma^\mu : G \to \mathbb{C}^{d\times d}$ constitutes a representation of G on $\mathbb{C}^d$.

Evaluate both sides at A to get

$$p^\mu_i(AA_0) = \sum_{j=1}^{d} \Gamma^\mu_{ji}(A_0)p_j(A).$$

Since the components of AA_0 are bilinear in the components of A and A_0, $p^\mu_i(AA_0)$ is a homogenous polynomial in the components of A_0 and $A_0^\dagger$. It follows that $\Gamma^\mu_{ji}(A_0)$ is a homogenous polynomial in the components of A_0 and $A_0^\dagger$. Moreover, setting $A = I_n$, we obtain

$$p^\mu_i(A_0) = \sum_{j=1}^{d} \Gamma^\mu_{ji}(A_0)p^\mu_j(I_n).$$

Thus, the Γ^μ_{ji}'s span V^μ. The same argument applies to the other irreducible components of V^m. Thus, V^m is spanned by the matrix elements of (a finite number of) irreducible representations of G.

4. The fact that the $(d^\mu)^{1/2}\Gamma^\mu_{ij}$'s form an orthonormal set follows from Proposition 6.19.

5. $G \subset \mathbb{C}^{n\times n}$ may be regarded as a compact subset of $\mathbb{R}^{2n\times 2n}$. By the Weierstrass Polynomial Approximation Theorem, polynomials on a compact subset of $\mathbb{R}^N$ are dense in the space of continuous functions on that subset with respect to the uniform norm. (Note: it is at this point that we need to make use of the fact that our polynomials p are allowed to be functions of the elements of both A and $A^\dagger$. If the p's were functions of elements of A only, then regarded as a functions on $\mathbb{R}^{2n\times 2n}$, they would not exhaust the space of polynomials.) Continuous functions on G, in turn, are dense in $L^2(G, \mathbb{C})$ with respect to the L^2-norm. It follows that polynomials in A and $A^\dagger$ are dense in $L^2(G, \mathbb{C})$. From the preceding, the space of polynomials is spanned by matrix elements of irreducible representations.

6. The preceding establishes that the matrix elements of a countable set of inequivalent irreducible representations of G, which are polynomials in the matrix elements of A and $A^\dagger$, span $L^2(G)$. It remains to show that there are no other inequivalent irreducible representations, either finite or infinite dimensional. This follows by establishing that the matrix elements of such a representation would be functions in $L^2(G)$ orthogonal to the Γ^μ_{ij}'s in $L^2(G)$, and there are no such functions.

□

Example 6.21 (Fourier series)**.** As a simple example, let

$$G = U(1) = \{z \in \mathbb{C} \,|\, |z| = 1\}.$$

We have the representation of $U(1)$ on $L^2(U(1))$, namely

$$f \mapsto \Gamma(z_0)f,$$

where

$$(\Gamma(z_0)f)(z) = f(zz_0).$$

For $f_m(z) = z^m$ (homogeneous polynomial in z), we have that

$$\Gamma(z_0) f_m = z_0^m f_m.$$

Thus, we obtain a one-dimensional representation of $U(1)$ on the subspace of $L^2(G)$ spanned by f_m. On this space, we have the irreducible representation

$$\Gamma_m(z_0) = z_0^m,$$

which happens to coincide with f_m.

The orthogonality and completeness of the $\Gamma_m(z)$'s is precisely the fact that the exponential functions $e^{im\theta}$, $m \in \mathbb{Z}$, constitute an orthonormal basis for $L^2(S^1)$ with Haar measure given by $d\theta/(2\pi)$.

7 Representations of Lie algebras

We shall turn to consider representations of Lie algebras.

Let $\mathfrak{g} \subset \mathbb{C}^{n\times n}$ be a matrix Lie algebra. That is, $\mathfrak{g}$ is a real vector space under matrix addition, and if $a, b \in \mathfrak{g}$, then $[a, b] := ab - ba \in \mathfrak{g}$. Let V be a complex finite-dimensional vector space (e.g., $\mathbb{C}^n$), and $L(V)$ the space of linear maps on V (e.g., $\mathbb{C}^{n\times n}$).

Definition 7.1. A **representation** of $\mathfrak{g}$ is a linear map

$$\hat{\Gamma} : \mathfrak{g} \to L(V)$$

such that for all $a, b \in \mathfrak{g}$,

$$\hat{\Gamma}([a, b]) = \hat{\Gamma}(a)\hat{\Gamma}(b) - \hat{\Gamma}(b)\hat{\Gamma}(a) = [\hat{\Gamma}(a), \hat{\Gamma}(b)]. \tag{26}$$

Note that $L(V)$ is a linear space, so it makes sense to require that $\hat{\Gamma}$ is linear. Eq.(26) means that $\hat{\Gamma}$ is a **Lie algebra homomorphism**.

Example 7.2 (adjoint representation)**.** Let $V = \mathfrak{g}$. Given $a \in \mathfrak{g}$, define $\mathrm{ad}_a \in L(\mathfrak{g})$ by

$$\mathrm{ad}_a(b) = [a, b].$$

We claim that this is a representation of $\mathfrak{g}$. Clearly ad_a is a linear map on $\mathfrak{g}$, and also $a \mapsto \mathrm{ad}_a$ is a linear map from $\mathfrak{g}$ to $L(\mathfrak{g})$, both assertions following from the fact that the Lie bracket is linear in each of its arguments. To verify Eq. (26), we note that

$$\begin{aligned}\mathrm{ad}_{[a,b]}(c) = [[a, b], c] &\overset{\text{Jacobi idenity}}{=} [[a, c], b] + [a, [b, c]] = \\ &= [\mathrm{ad}_a(c), b] + [a, \mathrm{ad}_b(c)] = -\mathrm{ad}_b(\mathrm{ad}\, a(c)) + \mathrm{ad}_a(\mathrm{ad}_b(c)) = (\mathrm{ad}_a \mathrm{ad}_b - \mathrm{ad}_b \mathrm{ad}_a)\,(c).\end{aligned}$$

Since c is arbitrary, Eq. (26) follows. Thus, the fact that adjoint action gives a representation of $\mathfrak{g}$ on itself is essentially equivalent to the Jacobi identity.

Notions of reducibility are defined as for group representations. A representation $\hat{\Gamma}$ of $\mathfrak{g}$ is **reducible** if there exists a proper subspace $W \subset V$ that is invariant under $\hat{\Gamma}$; that is, for all $a \in \mathfrak{g}$, $\hat{\Gamma}(a)W \subset W$. $\hat{\Gamma}$ is **irreducible** if V contains no proper invariant subspaces. $\hat{\Gamma}$ is **completely reducible** if V can be expressed as a direct sum, $\oplus_j W_j$, of proper invariant subspaces W_j such that $\hat{\Gamma}$ restricted to each W_j is irreducible.

7.1 From group representations to algebra representations

Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Let (Γ, V) be a representation of G. We assume that Γ is continuous. It turns out (see Problem Sheet 3) that this implies that Γ is actually smooth. More precisely, if Γ is continuous, then if

$$\Phi : P \subset \mathbb{R}^d \to V_I \subset G$$

is a smooth local parameterisation of G, then $\Gamma \circ \Phi : P \to GL(V)$ is smooth.

Let $A(t)$ be a smooth curve in G with $A(0) = I$, $\dot{A}(0) = a \in \mathfrak{g}$. Define

$$\hat{\Gamma}(a) := \left.\frac{d}{dt}\right|_{t=0} \Gamma(A(t)) \in L(V). \tag{27}$$

We will show that $\hat{\Gamma}$ as defined by Eq. (27) is a representation of $\mathfrak{g}$.

Recall from Proposition 2.2 that

$$\xi_\mu := \frac{\partial \Phi}{\partial x^\mu}(0)$$

constitutes a basis for $\mathfrak{g}$, and that, from the properties of the parameterisation Φ, there exists a smooth curve $x(t) \in P$ with $x(0) = 0$ such that $A(t) = \Phi(x(t))$ and

$$a = \dot{A}(0) = \dot{x}^\mu(0)\xi_\mu.$$

Let

$$\hat{\xi}_\mu := \left.\frac{\partial}{\partial x^\mu}\right|_{x=0} \Gamma(\Phi(x)) \in L(V).$$

From the preceding, it follows that

$$\hat{\Gamma}(a) = \left.\frac{d}{dt}\right|_{t=0} \Gamma(\Phi(x(t))) = \dot{x}^\mu(0)\hat{\xi}_\mu.$$

Thus, for arbitrary coefficients $c^\mu \in \mathbb{R}$, we have that

$$\hat{\Gamma}(c^\mu \xi_\mu) = c^\mu \hat{\xi}_\mu.$$

This establishes that $\hat{\Gamma}$ is linear.

Proposition 7.3. Let $a \in \mathfrak{g}$. Then

$$\Gamma(\exp(ta)) = \exp(t\hat{\Gamma}(a)).$$

Proof. We argue below that both sides of the preceding satisfying the same system of 1st-order linear ODEs with the same initial conditions, and therefore are necessarily the same.

Let the left-hand side be given by

$$L(t) := \Gamma(\exp(ta)).$$

Then $L(0) = I_V$ (where I_V denotes the identity map on V). Also,

$$\dot{L}(t) = \left.\frac{d}{ds}\right|_{s=0} \Gamma(\exp((s+t)a)) = \left.\frac{d}{ds}\right|_{s=0} \Gamma(\exp(sa)\exp(ta)) = \left(\left.\frac{d}{ds}\right|_{s=0} \Gamma(\exp(sa)) \right) \Gamma(\exp(ta)).$$

Since $A(s) := \exp(sa)$ describes a smooth curve in G with $A(0) = I$ and $A'(0) = a$, it follows from Eq. (27) that

$$\dot{L}(t) = \hat{\Gamma}(a)L(t).$$

The right-hand side is given by

$$R(t) = \exp(t\hat{\Gamma}(a)).$$

Then $R(0) = I_V$, and

$$\dot{R}(t) = \hat{\Gamma}(a)R(t).$$

□

Recall the Adjoint representation of G on $\mathfrak{g}$ (cf Proposition 3.2); for $A \in G$ and $b \in \mathfrak{g}$, we define

$$\mathrm{Ad}_A\, b = A\, b\, A^{-1}.$$

Proposition 7.4.

$$\hat{\Gamma}(AbA^{-1}) = \Gamma(A)\hat{\Gamma}(b)\Gamma(A^{-1}).$$

Proof. Given $A \in G$ and $b \in \mathfrak{g}$, we have that

$$\Gamma\left(Ae^{tb}A^{-1}\right) = \Gamma(A)\Gamma\left(e^{tb}\right)\Gamma\left(A^{-1}\right). \tag{28}$$

Since $\tilde{B}(t) := A\exp(tb)A^{-1}$ is a curve in G with $\tilde{B}(0) = I_V$ and $\dot{\tilde{B}}(0) = AbA^{-1}$, differentiating the LHS of (28) yields

$$\left.\frac{d}{dt}\right|_{t=0}\Gamma\left(Ae^{tb}A^{-1}\right) = \hat{\Gamma}\left(AbA^{-1}\right).$$

Differentiating the RHS of (28) yields

$$\left.\frac{d}{dt}\right|_{t=0}\Gamma(A)\Gamma\left(e^{tb}\right)\Gamma\left(A^{-1}\right) = \Gamma(A)\hat{\Gamma}(b)\Gamma\left(A^{-1}\right).$$

□

Proposition 7.5.

$$\hat{\Gamma}([a,b]) = \hat{\Gamma}(a)\hat{\Gamma}(b) - \hat{\Gamma}(b)\hat{\Gamma}(a).$$

Proof. From Proposition 2.9,

$$[a,b] = \left.\frac{d}{ds}\right|_{s=0}\exp(sa)b\exp(-sa).$$

Therefore

$$\hat{\Gamma}([a,b]) = \hat{\Gamma}\left(\left.\frac{d}{ds}\right|_{s=0}e^{sa}be^{-sa}\right) = \left.\frac{d}{ds}\right|_{s=0}\hat{\Gamma}\left(e^{sa}be^{-sa}\right)$$

(since $\hat{\Gamma}$ is linear, the derivative may be taken outside its argument). From Propositions 7.4 and 7.3, it follows that

$$\begin{aligned}\hat{\Gamma}([a,b]) = &\left.\frac{d}{ds}\right|_{s=0}\Gamma(\exp(sa))\hat{\Gamma}(b)\Gamma(\exp(-sa))\\ &= \left.\frac{d}{ds}\right|_{s=0}\exp\left(s\hat{\Gamma}(a)\right)\hat{\Gamma}(b)\exp\left(-s\hat{\Gamma}(a)\right) = \hat{\Gamma}(a)\hat{\Gamma}(b) - \hat{\Gamma}(b)\hat{\Gamma}(a).\end{aligned}$$

□

We may summarise the preceding as follows:

Theorem 7.6.

Let $\Gamma : G \to GL(V)$ be a representation of a matrix Lie group G on a complex vector space V. Then $\hat{\Gamma} : \mathfrak{g} \to L(V)$ given by

$$\hat{\Gamma}(a) = \left.\frac{d}{dt}\right|_{t=0}\Gamma\left(e^{ta}\right)$$

is a representation of $\mathfrak{g}$ on V.

Example 7.7.

i) The Adjoint and adjoint representations of G and $\mathfrak{g}$ respectively on $V = \mathfrak{g}$ are related as in Theorem 7.6.

ii) Let $\mathbf{z} = (z_1, z_2)^T \in \mathbb{C}^2$ and consider the representation Γ of SU(2) on $C^\infty(\mathbb{C}^2, \mathbb{C})$, the space of smooth, complex-valued functions on $\mathbb{C}^2$, defined as follows:

$$(\Gamma(u)f)(\mathbf{z}) := f\left(u^{-1}\mathbf{z}\right).$$

The Lie algebra su(2) consists of matrices of the form $i\mathbf{a}\cdot\boldsymbol{\sigma}$ (cf Lecture 5), where $\mathbf{a} \in \mathbb{R}^3$ and

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We have that

$$\left(\hat{\Gamma}(i\mathbf{a}\cdot\boldsymbol{\sigma})f\right)(\mathbf{z}) = \left(\left.\frac{d}{dt}\right|_{t=0}\Gamma\left(e^{it\mathbf{a}\cdot\boldsymbol{\sigma}}\right)f\right)(\mathbf{z}) = \left.\frac{d}{dt}\right|_{t=0}f\left(e^{-it\mathbf{a}\cdot\boldsymbol{\sigma}}\mathbf{z}\right) = -i\sum_{j,k=1}^{2}(\mathbf{a}\cdot\boldsymbol{\sigma})_{jk}z_k\frac{\partial f}{\partial z_j}(\mathbf{z}).$$

It particular,

$$\hat{\Gamma}(i\sigma_1) = iz_1\frac{\partial}{\partial z_2} + iz_2\frac{\partial}{\partial z_1}, \quad \hat{\Gamma}(i\sigma_2) = z_1\frac{\partial}{\partial z_2} - z_2\frac{\partial}{\partial z_1}, \quad \hat{\Gamma}(i\sigma_3) = -iz_1\frac{\partial}{\partial z_1} + iz_2\frac{\partial}{\partial z_2}.$$

Thus, the representation $\hat{\Gamma}$ is by linear differential operators. One can verify directly that, for example,

$$[\hat{\Gamma}(i\sigma_1), \hat{\Gamma}(i\sigma_2)] = 2\hat{\Gamma}(i\sigma_3).$$

Proposition 7.8.

Let Γ and $\hat{\Gamma}$ be related as in Theorem 7.6. Then if $\hat{\Gamma}$ is irreducible, so is Γ.

The argument is straightforward and is an Exercise in Problem Sheet 3.

7.2 *From algebra representation to group representation

Proposition 7.8 provides a link from the representations of a group to representations of its Lie algebra.

The converse association, from a representation of a Lie algebra to a representation of a group, is not as straightforward. Note that a matrix Lie group (trivially a representation of itself) is not uniquely determined by its matrix Lie algebra (also trivially a representation of itself); for example, SO(3) and SU(2) have the same Lie algebra, but are not themselves the same (nor are they isomorphic).

The question of whether a representation of a Lie algebra determines a representation of its associated Lie group involves the topology of the group. We shall confine ourselves here to the following remarks (some related material is developed in Problem Sheet 3, and is discussed in more detail Brian Hall's book in the references).

A matrix Lie group G is **connected** if every element $A \in G$ can be connected to the identity by a continuous curve $A(t) \in G$, with, say $A(0) = I$ and $A(1) = A$ (if this is the case, we can assume that $A(t)$ is smooth). Examples of matrix Lie groups that are not connected are finite matrix groups (whose connected components are singletons), and direct products of a matrix Lie group G with a finite matrix group. A matrix Lie group G is **simply connected** if every continuous closed curve $A(t) \in G$ based at the identity, so that $A(0) = A(1) = I$, can be continuously contracted to the constant curve $A_0(t) = I$. That is, there exists a continuous family of matrices (a homotopy) $H(s,t)$, defined for $0 \le s, t \le 1$, such that $H(s,0) = H(s,1) = H(0,t) = I$ and $H(1,t) = A(t)$. Note that these definitions are not specific to matrix Lie groups; they apply to more general topological spaces.

The Lie algebra $\mathfrak{g}$ determines a unique Lie group (though not necessarily a matrix Lie group) called the *universal covering* of G, denoted $\bar{G}$, which is simply connected and which has $\mathfrak{g}$ as its Lie algebra. For example, the universal covering of SO(3) is SU(2) (recall that SU(2) may be identified with the three-sphere $S^3 \subset \mathbb{R}^4$, and is therefore simply connected). $\bar{G}$ can be regarded as the space of continuous curves $A(t) \in G$, $0 \le t \le 1$, with $A(0) = I$, modulo an equivalence relation; $A(t)$ and $B(t)$ are equivalent if $A(1) = B(1)$ and if the concatenation of $A(t)$ and $B(1-t)$ is contractible.

If G is connected, one can show that $G = \bar{G}/\pi_1(G)$, where $\pi_1(G)$ is the **fundamental group** of G, i.e. the set of closed continuous curves in G based at the identity with product given by concatenation, and with curves which may be continuously deformed into each other regarded as equivalent. The fundamental group $\pi_1(G)$ may be regarded as a normal subgroup of $\bar{G}$. In particular, if G is simply connected, then $\pi_1(G)$ is trivial, and $\bar{G} = G$.

A representation $\hat{\Gamma}$ of $\mathfrak{g}$ determines a representation $\bar{\Gamma}$ of $\bar{G}$, and $\bar{\Gamma}$ and $\hat{\Gamma}$ are related as in Theorem 7.6. If $\pi_1(G)$ is contained in the kernel of $\bar{\Gamma}$, then $\hat{\Gamma}$ determines a representation Γ of G, with Γ and $\hat{\Gamma}$ related as in Theorem 7.6. In particular, if G is simply connected, then this is always the case. For example, since SU(2) is simply connected, every representation of su(2) determines a representation of SU(2), and irreducible representations of SU(2) and su(2) are in 1-1 correspondence. On the other hand, as SO(3) is isomorphic to $\mathrm{SU}(2)/\{I, -I\}$ (the fundamental group of SO(3) is isomorphic to $\mathbb{Z}_2$), a representation $\hat{\Gamma}$ of so(3) $\simeq$ su(2) determines a representation of SO(3) if and only if the corresponding representation Γ of SU(2) satisfies $\Gamma(-I) = I_V$. Such a representation is called a *tensor representation*. By contrast, representations of SU(2) for which $\Gamma(-I) = -I_V$ are called *spinor representations*. In a tensor representation, rotations through 2π are represented by the identity I, and for spinor representations, by $-I$. We compute the irreducible representations of su(2) in the next section. One can show that even dimensional representations are tensor representations, and odd dimensional representations are spinor representations.

8 Representations of su(2)

We computed the Lie algebra su(2) in Lecture 5 and found that

$$\begin{aligned}\mathrm{su}(2) &= \{\text{traceless antihermitian } 2\times 2 \text{ matrices}\}\\ &= \mathrm{span}(i\sigma_1, i\sigma_2, i\sigma_3),\end{aligned}$$

where the Pauli matrices, σ_j, are given by

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

and span the space of traceless *hermitian* 2×2 matrices. Let

$$e_j = -\tfrac{1}{2} i\sigma_j.$$

Then e_j's constitute a basis for su(2), and their Lie brackets are given by

$$[e_1, e_2] = e_3, \quad [e_2, e_3] = e_1, \quad [e_3, e_1] = e_2. \tag{29}$$

In what follows we calculate explicitly all the inequivalent irreducible representations of su(2). The basic idea is to choose a set of basis vectors for the Lie algebra su(2) and a set of basis vectors for the representation space V in terms of which the representation takes a simple canonical form. In fact, the same idea works more generally for compact semi-simple Lie algebras, as we shall see.

First, we note that we can always choose a basis for the carrier space V with respect to which the representation of a particular element of su(2), say e_3, is diagonal (it turns out that representations of su(2) can be taken to be antihermitian, so that $\hat{\Gamma}(a)$ is antihermitian, and therefore diagonalisable, for all $a \in \mathrm{su}(2)$). The trick is to choose a set of elements in su(2) that, together with e_3, span su(2) and which lead to representations of a simple form. The clue comes from considering the adjoint representation, where su(2) is both the Lie algebra and the carrier space. A judicious choice of basis for the adjoint representation leads to a canonical presentation for every irreducible representation.

8.1 Canonical form for the adjoint representation

We look for eigenvectors of ad_{e_3}. That is, we look for solutions of

$$[e_3, a] = \lambda a,$$

where $a \in \mathrm{su}(2)$.

Clearly, one solution is $a = e_3$ and $\lambda = 0$.

Therefore, we may take the remaining eigenvectors to be of the form $a = c_1 e_1 + c_2 e_2$. The eigenvector equation becomes

$$\big[e_3, c_1 e_1 + c_2 e_2\big] = c_1 e_2 - c_2 e_1 = \lambda(c_1 e_1 + c_2 e_2),$$

where we have used (29) for the Lie brackets $[e_j, e_k]$. Equating coefficients, we get that

$$c_2 = -\lambda c_1, \quad c_1 = \lambda c_2$$

which implies that $\lambda^2 = -1$, or $\lambda = \pm i$. It follows that $c_1 = \pm i c_2$. We take $c_2 = -i/\sqrt{2}$ (this choice is convenient, but not necessary). Then the eigenvectors are given by

$$e_\pm = \frac{1}{\sqrt{2}}\left(\pm e_1 - i e_2\right).$$

Thus,

$$[e_3, e_\pm] = \pm i e_\pm.$$

Now we must address the fact that $e_\pm$ do not belong to su(2); $e_\pm$ consist of complex, rather than real, linear combinations of e_1 and e_2. To proceed, we introduce the *complexification* of su(2), which we denote by $\mathrm{su}(2)_{\mathbb{C}}$, and which we may take to consist of all complex linear combinations of e_1, e_2 and e_3:

$$\mathrm{su}(2)_{\mathbb{C}} = \left\{c_1e_1 + c_2e_2 + c_3e_3 \,|\, c_j \in \mathbb{C}\right\} \subset \mathbb{C}^{2\times 2},$$

We extend the Lie bracket to $\mathrm{su}(2)_{\mathbb{C}}$ in the obvious way:

$$\left[\sum_j c_je_j, \sum_k d_ke_k\right] = \sum_{j,k} c_jd_k[e_j, e_k],$$

and it is obvious that $\mathrm{su}(2)_{\mathbb{C}}$ forms a complex Lie subalgebra of $\mathbb{C}^{2\times 2}$. That is, $\mathrm{su}(2)_{\mathbb{C}}$ is a complex vector subspace of $\mathbb{C}^{2\times 2}$ that is closed under the Lie bracket (the Jacobi identity is automatically satisfied for Lie algebras of matrices).

Since $e_j^\dagger = -e_j$ (the e_j's are antihermitian), it follows that

$$e_+^\dagger = e_-.$$

We will regard $e_\pm$, e_3 as a basis for $\mathrm{su}(2)_{\mathbb{C}}$. We know the Lie brackets of e_3 with $e_\pm$. Let us calculate $[e_+, e_-]$, as follows:

$$[e_+, e_-] = \tfrac{1}{2}[e_1 - ie_2, -e_1 - ie_2] = -ie_3.$$

To summarise, we regard $e_\pm, e_3$ as a canonical basis for $\mathrm{su}(2)_{\mathbb{C}}$. In terms of this basis, the Lie bracket is given by

$$[e_3, e_\pm] = \pm ie_\pm, \quad [e_+, e_-] = -ie_3. \tag{30}$$

We have that

$$e_\pm^\dagger = e_\mp, \quad e_3^\dagger = -e_3. \tag{31}$$

The real Lie algebra su(2) consists of the antihermitian subspace of $\mathrm{su}(2)_{\mathbb{C}}$, and is given by

$$\mathrm{su}(2) = \{ze_+ - \bar{z}e_- + c_3e_3 \,|\, z \in \mathbb{C}, c_3 \in \mathbb{R}\}. \tag{32}$$

8.2 Irreducible representations of su(2)

Let $(\hat\Gamma, V)$ denote an irreducible representation of su(2) on a finite-dimensional vector space V. Without loss of generality, it turns out we may assume that V is an inner product space, and that $\hat\Gamma(a)$ is antihermitian for all $a \in \mathrm{su}(2)$. (We have not shown this explicitly, but it follows from arguments in Section 7.2.)

Let $\hat E_j := \hat\Gamma(e_j)$ denote the representatives of the e_j's. Let

$$\hat E_\pm := \frac{1}{\sqrt{2}}\left(\pm\hat E_1 - i\hat E_2\right). \tag{33}$$

From the representation property, in particular

$$[\hat E_1, \hat E_2] = \hat E_3, \quad [\hat E_2, \hat E_3] = \hat E_1, \quad [\hat E_3, \hat E_1] = \hat E_2,$$

it follows that

$$[\hat E_3, \hat E_\pm] = \pm i\hat E_\pm, \quad [\hat E_+, \hat E_-] = -i\hat E_3. \tag{34}$$

Also, since $\hat E_j^\dagger = -\hat E_j$, we have that

$$\hat E_\pm^\dagger = \hat E_\mp, \quad \hat E_3^\dagger = -\hat E_3. \tag{35}$$

In analogy with our treatment of the adjoint representation, we seek a basis for V in which $\hat E_3$ is diagonal. Since V is complex and $\hat E_3$ is antihermitian, this can be found. Note that the eigenvalues of an antihermitian matrix are pure imaginary.

For $\mu \in \mathbb{R}$, let $W(\mu)$ denote the subspace of eigenvectors of $\hat E_3$ with eigenvalue $i\mu$, i.e.

$$W(\mu) = \{v \in V \,|\, (\hat E_3 - i\mu)v = 0\}.$$

Of course, if $i\mu$ is not an eigenvalue of $\hat{E}_3$, then $W(\mu)$ consists only of the zero vector. Let $i\mu_1, \ldots, i\mu_s$ denote the (imaginary) eigenvalues of $\hat{E}_3$. Then

$$V = \oplus_{r=1}^{s} W(\mu_r).$$

In fact, since $\hat{E}_3$ is antihermitian, the subspaces $W(\mu_r)$ are orthogonal to each other. Note that any one of the eigenvalues $i\mu_r$ could be degenerate, in which case the corresponding eigenspace would have dimension greater than one. (In fact, we will show below that because $\hat{\Gamma}$ is irreducible, the eigenspaces $W(\mu_r)$ are necessarily one dimensional.)

Proposition 8.1.

$$\hat{E}_\pm W(\mu) \subset W(\mu \pm 1).$$

Proof. If $W(\mu) = \{0\}$, i.e. if $i\mu$ is not an eigenvalue of $\hat{E}_3$, the statement is trivial. Therefore, we may assume that $i\mu$ is an eigenvalue of $\hat{E}_3$. Let $v \in W(\mu)$ be a nonzero eigenvector of $\hat{E}_3$ with eigenvalue $i\mu$. We need to show that $\hat{E}_\pm\, v \in W(\mu{\pm}1)$, i.e.

$$\hat{E}_3 \hat{E}_\pm\, v = i(\mu \pm 1)\hat{E}_\pm\, v.$$

From the Lie brackets (34), in particular $[\hat{E}_3, \hat{E}_\pm] = \pm i\hat{E}_\pm$, it follows that

$$\hat{E}_3 \hat{E}_\pm = \hat{E}_\pm \hat{E}_3 \pm i\hat{E}_\pm.$$

Then

$$\hat{E}_3 \hat{E}_\pm\, v = \left(\hat{E}_\pm \hat{E}_3 \pm i\hat{E}_\pm\right) v = \hat{E}_\pm(\hat{E}_3 \pm i)v = i(\mu \pm 1)\hat{E}_\pm\, v,$$

as required. □

Let $i\mu_*$ be an eigenvalue of $\hat{E}_3$, and let $v_* \in W(\mu_*)$ be a nonzero eigenvector of $\hat{E}_3$ with eigenvalue $i\mu_*$. From Proposition 8.1, either $\hat{E}_+^k v_*$ is a nonzero eigenvector of $\hat{E}_3$ with eigenvalue $i(\mu_* + k)$, or else $\hat{E}_+^k v_* = 0$. Since $\hat{E}_3$ can have only a finite number of eigenvalues (as V is finite dimensional), it follows that $\hat{E}_+^k v_*$ must vanish for sufficiently large k. Let $p \geq 0$ denote the smallest nonnegative integer for which

$$\hat{E}_+^p\, v_* \neq 0, \quad \hat{E}_+^{p+1}\, v_* = 0. \tag{36}$$

$$v(p) := \hat{E}_+^p\, v_*. \tag{37}$$

Without loss of generality, we may assume that

$$||v(p)|| = 1$$

(just multiply v_* by the appropriate scalar factor to ensure this is the case). Then $v(p) \in W(\mu_* + p)$.

Next, we repeatedly apply $\hat{E}_-$ to $v(p)$, generating a sequence of vectors in V. Since, from Proposition 8.1, $\hat{E}_-^k v(p) \in W(\mu_* + p - k)$, these vectors are orthogonal to each other. The following establishes that the first $p+1$ vectors in this sequence (including $v(p)$ itself) are nonzero.

Proposition 8.2. For $0 \leq k \leq p$,

$$\hat{E}_-^k v(p) \neq 0.$$

Proof. We will show that the inner product of $\hat{E}_-^k v(p)$ with some given vector u does not vanish, which will imply that $\hat{E}_-^k v(p)$ itself cannot vanish. Indeed, let

$$u = \hat{E}_+^{p-k} v_*.$$

From the definition of p, it follows that $u \neq 0$. Note that u belongs to $W(\mu_* + p - k)$. Since $\hat{E}_-^k v(p)$ belongs to $W(\mu_* + p - k)$ as well, it is at least possible that $\left\langle u, \hat{E}_-^k v(p)\right\rangle \neq 0$.

In fact, we have that

$$\left\langle u, \hat{E}_-^k v(p)\right\rangle = \left\langle \hat{E}_+^{p-k} v_*, \hat{E}_-^k v(p)\right\rangle.$$

Recalling that $\hat{E}_-^\dagger = \hat{E}_+$, we get that

$$\left\langle u, \hat{E}_-^k v(p)\right\rangle = \left\langle \hat{E}_+^k \hat{E}_+^{p-k} v_*, v(p)\right\rangle = \left\langle \hat{E}_+^p v_*, v(p)\right\rangle = \langle v(p), v(p)\rangle = 1,$$

as required. □

Since $\hat{E}_-^{p+r}v(p) \in W(\mu_* - r)$, it follows that $\hat{E}_-^{p+r}v(p)$ must vanish for some $r > 0$. Let $q \geq 0$ be the smallest integer for which

$$\hat{E}_-^{p+q}v(p) \neq 0, \quad \hat{E}_-^{p+q+1}v(p) = 0. \tag{38}$$

Define a sequence of vectors $v(a)$ inductively as follows. $v(p)$ is given by (37). Given $v(a)$, define $v(a-1)$ by

$$v(a-1) = \frac{1}{N(a)}\,\hat{E}_-\,v(a), \quad N(a) = \left(\left\langle \hat{E}_-v(a), \hat{E}_-v(a)\right\rangle\right)^{1/2}. \tag{39}$$

By construction, the $v(a)$'s are normalised. From the definition of q, $v(a) \neq 0$ for $-q \leq a \leq p$. From Proposition 8.1, $v(a) \in W(\mu_* + a)$. Therefore,

$$\begin{aligned}\hat{E}_3v(a) &= i(\mu_* + a)v(a),\\ \hat{E}_-v(a) &= N(a)v(a-1).\end{aligned}$$

Thus, we have constructed a set of $p + q + 1$ orthonormal vectors in V whose span is invariant under $\hat{E}_3$ and $\hat{E}_-$, and with respect to which $\hat{E}_3$ and $\hat{E}_-$ assume a canonical form. By convention, we define

$$v(p+1) := 0, \quad N(p+1) := 0.$$

It remains to consider the action of $\hat{E}_+$ on the $v(a)$'s. This is given by the following:

Proposition 8.3. For $-q \leq a \leq p$,

$$\hat{E}_+v(a) = N(a+1)v(a+1).$$

Note that Proposition 8.3 does not follow automatically from Proposition 8.1, which implies only that $\hat{E}_+\,v(a)$ belongs to $W(\mu_* + a + 1)$, and does not imply that $\hat{E}_+\,v(a)$ is proportional to $v(a+1)$ (a priori, $W(\mu_* + a + 1)$ may have dimension greater than 1).

Proof. There are two things to show, namely that (i) $\hat{E}_+\,v(a)$ is proportional to $v(a+1)$ and that (ii) the constant of proportionality is given by $N(a+1)$. First, we show that (ii) follows from (i). Suppose that

$$\hat{E}_+\,v(a) = C(a+1)\,v(a+1)$$

for some scalar factor $C(a+1)$. Taking the inner product of the preceding with $v(a+1)$ and using the fact that the $v(a)$'s are normalised, we have that

$$C(a+1) = \left\langle v(a+1), \hat{E}_+\,v(a)\right\rangle.$$

Since $\hat{E}_+^\dagger = \hat{E}_-$, we get

$$C(a+1) = \left\langle \hat{E}_-v(a+1), v(a)\right\rangle = N(a+1)\langle v(a), v(a)\rangle = N(a+1), \tag{40}$$

where we have used the definition (39) of $v(a)$.

To establish (i), we proceed by backward induction on a. The assertion holds trivially for $a = p$, since $\hat{E}_+v(p) = 0$. Let us assume that (i) holds for a . We show that it holds for $a - 1$, i.e.

$$\hat{E}_+\,v(a-1) = C(a)v(a).$$

From the definition (39) of $v(a-1)$, we have that

$$\hat{E}_+\,v(a-1) = \hat{E}_+\left(\frac{1}{N(a)}\hat{E}_-\,v(a)\right) = \frac{1}{N(a)}\hat{E}_+\hat{E}_-\,v(a).$$

From the Lie bracket relation

$$[\hat{E}_+, \hat{E}_-] = -i\hat{E}_3,$$

it follows that

$$\hat{E}_+\,v(a-1) = \frac{1}{N(a)}\left(\hat{E}_-\hat{E}_+ - i\hat{E}_3\right)v(a).$$

But

$$\hat{E}_3\,v(a) = i(\mu_* + a)v(a),$$

and by the induction hypothesis and the definition of the $v(a)$'s,

$$\hat{E}_- \hat{E}_+ \, v(a) = N(a+1)\hat{E}_- \, v(a+1) = N^2(a+1)\, v(a).$$

Combining these results, we obtain

$$\hat{E}_+ \, v(a-1) = C(a)\, v(a),$$

where

$$C(a) = \frac{N^2(a+1) + (\mu_* + a)}{N(a)} \tag{41}$$

which shows that $\hat{E}_+ v(a-1)$ is indeed proportional to $v(a)$, as required. □

Comparison of (40) and (41) gives

$$C(a) = N(a) = \frac{N(a+1)^2 + \mu_* + a}{N(a)},$$

or

$$N^2(a+1) - N^2(a) = -(\mu_* + a).$$

This is a first-order difference equation for $N^2(a)$ with two boundary conditions, since we know that $N(a)$ vanishes for $a = p+1$ and $a = -q$. Therefore, in order for a solution to exist, there is necessarily a constraint on the (single) parameter $'mu_*$ in the equation. The solution is necessarily inhomogeneous quadratic in a. The boundary conditions are conveniently incorporated by taking $N^2(a)$ to be of the form

$$N^2(a) = A(p+1-a)(a+q).$$

Substituting this form into the recurrence relation, we get that

$$A(p - q - 2a) = -\mu_* - a$$

which implies that

$$A = \tfrac{1}{2}, \tag{42}$$

$$\mu_* = \tfrac{1}{2}(q-p). \tag{43}$$

The vector space spanned by the $v(a)$'s, $-q \le a \le p$, is invariant under $\hat{E}_3$ and $\hat{E}_\pm$, and therefore is invariant under $\hat{\Gamma}$. Since $\hat{\Gamma}$ is irreducible, it follows that $V = \operatorname{span}\{v(a)\}$, and that

$$n = \dim V = p + q + 1.$$

The largest eigenvalue of $\hat{E}_3$ is given by $\mu_* + p = (p+q)/2 = (n-1)/2$, and the smallest is given by $\mu_* - q = -(p+q)/2 = -(n-1)/2$. Without loss of generality, we may take $\mu_* = -(n-1)/2$ (after all, μ_* is only required to be one of the eigenvalues of $\hat{E}_3$). Then $q = 0$, $p = n-1$, and

$$N^2(a) = \tfrac{1}{2}a(n-a).$$

The index a takes values between 0 and $n-1$. The actions of $\hat{E}_3$ and $\hat{E}_\pm$ on $v(a)$ are given by

$$\begin{aligned} \hat{E}_3 \, v(a) &= \left(-\tfrac{1}{2}(n-1) + a\right) v(a), \\ \hat{E}_+ \, v(a) &= \left(\tfrac{1}{2}(a+1)(n-1-a)\right)^{1/2} v(a+1), \\ \hat{E}_- \, v(a) &= \left(\tfrac{1}{2}a(n-a)\right)^{1/2} v(a-1). \end{aligned} \tag{44}$$

We note that the actions in (44) are determined entirely by n, the dimension of the carrier space. We may summarise as follows:

Theorem 8.4 (Irreducible representations of su(2))**.** Up to equivalence, there is precisely one irreducible representation $\hat{\Gamma}^n$ of su(2) of dimension n for $n \ge 0$. The linear maps $\hat{E}_3 = \hat{\Gamma}^n(e_3)$ and $\hat{E}_\pm = \hat{\Gamma}^n(e_\pm)$, which determine the representation, are given by (44), where $v(a)$, $0 \le a \le (n-1)$, denotes a basis for the carrier space. $\hat{E}_1$ and $\hat{E}_2$ are given by (cf (33))

$$\hat{E}_1 = \frac{1}{\sqrt{2}}\left(\hat{E}_+ - \hat{E}_-\right), \quad \hat{E}_2 = -\frac{1}{\sqrt{2}i}\left(\hat{E}_+ + \hat{E}_-\right).$$

9 Compact simple Lie algebras and Cartan subalgebras

Let G be a group. A subgroup $H \subset G$ is **normal** if $\forall g \in G, h \in H$, we have that

$$ghg^{-1} \in H.$$

$\{I\}$ and G are trivially normal subgroups of G. G is **simple** if it has no nontrivial normal subgroups.

Let G be a matrix Lie group with matrix Lie algebra $\mathfrak{g}$. An **ideal** $I \subset \mathfrak{g}$ is a subspace of $\mathfrak{g}$ such that $\forall a \in \mathfrak{g}, b \in I$, we have that

$$[a, b] \in I.$$

$\{0\}$ and $\mathfrak{g}$ are trivial ideals of $\mathfrak{g}$. $\mathfrak{g}$ is **simple** if it has no nontrivial ideals.

Example 9.1. $\mathrm{U}(n)$, the unitary group, is not simple. The subgroup of multiples of the identity, $e^{i\theta}I_n$, is a normal subgroup. $\mathrm{u}(n)$, the Lie algebra of $U(n)$, is given by $\mathbb{C}_{-}^{n\times n}$, the space of antihermitian matrices. $\mathrm{u}(n)$ is not simple, as the subspace $\{i\theta I_n\}$, is a nontrivial ideal.

Exercise 9.2.

i) Show that $\mathrm{SU}(n)$ is a not simple group, but $\mathrm{su}(n)$ is a simple Lie algebra.

ii) Suppose that $H \subset G$ is a normal subgroup and that H is a matrix Lie group in its own right, with Lie algebra $\mathfrak{h}$. Then show that $\mathfrak{h} \subset \mathfrak{g}$ is an ideal.

Proposition 9.3. Simple properties of simple Lie algebras

a) If $\mathfrak{g}$ is simple, then $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$, where

$$[\mathfrak{g}, \mathfrak{g}] := \operatorname{span}\left\{[a, b] \,\middle|\, a, b \in \mathfrak{g}\right\}.$$

b) If $\mathfrak{g}$ is simple, then $\mathrm{Ad} : G \to GL(\mathfrak{g})$ is irreducible.

c) If $\mathfrak{g}$ is simple and $(\hat{\Gamma}, V)$ is a nontrivial representation of $\mathfrak{g}$ (i.e., $\hat{\Gamma}(a) \neq 0$ for some a), then $\ker \hat{\Gamma} = 0$, and $\mathfrak{g}$ is isomorphic to its image in $L(V)$. In particular, $\mathfrak{g}$ is isomorphic to its adjoint representation.

Proof.

a) $[\mathfrak{g}, \mathfrak{g}]$ is an ideal.

b) An Ad-invariant subspace I is an ideal, since for all $a \in \mathfrak{g}$ and $b \in I$, $[a, b] = \dot{c}(0)$, where

$$c(t) = e^{ta} b e^{-ta}.$$

$c(t)$ is a curve in I, since I is Ad-invariant.

c) The kernel of any Lie algebra homomorphism $\hat{\Gamma} : \mathfrak{g} \to \mathfrak{h}$ is an ideal, since for all $a \in \mathfrak{g}, n \in \ker \hat{\Gamma}$, we have that

$$\hat{\Gamma}([a, n]) = [\hat{\Gamma}(a), \hat{\Gamma}(n)] = 0.$$

□

If $G \subset GL(n, \mathbb{C})$ is a compact matrix Lie group, we say that its Lie algebra, $\mathfrak{g}$, is **compact**. From Proposition 6.16 (representations of compact matrix Lie groups are equivalent to unitary representations), without loss of generality (applying a conjugation if necessary) we may assume that $G \subset \mathrm{U}(n)$ (since G may be regarded as a representation of itself). Then $\mathfrak{g} \subset \mathrm{u}(n) = \mathbb{C}_{-}^{n\times n}$; that is, $\mathfrak{g}$ is comprised of antihermitian matrices. This fact makes the study of compact simple Lie algebras a bit simpler.

We introduce the usual inner product on $\mathbb{C}^{n\times n}$ and restrict to $\mathfrak{g}$. That is, given $a, b \in \mathfrak{g}$, we define

$$\langle a, b \rangle := \operatorname{Tr}(a^{\dagger} b) = -\operatorname{Tr}(ab).$$

Proposition 9.4 (Ad and ad invariance of inner product)**.**

a) For all $C \in G$, $\mathrm{Ad}_C : \mathfrak{g} \to \mathfrak{g}$ is real orthogonal, ie

$$\langle \mathrm{Ad}_C a, \mathrm{Ad}_C b \rangle = \langle a, b \rangle$$

for all $a, b \in \mathfrak{g}$.

b) For all $c \in \mathfrak{g}$, $\mathrm{ad}_c a : \mathfrak{g} \to \mathfrak{g}$ is real antisymmetric, ie

$$\langle \mathrm{ad}_c a, b \rangle + \langle a, \mathrm{ad}_c b \rangle = 0$$

for all $a, b \in \mathfrak{g}$,

Proof.

a) We show that Ad preserves the (real) inner product $\langle \cdot, \cdot \rangle$. Since C is unitary,

$$\mathrm{Ad}_C b = CbC^{-1} = CbC^{\dagger}.$$

Then

$$\langle \mathrm{Ad}_C a, \mathrm{Ad}_C b \rangle = \left\langle CaC^{\dagger}, CbC^{\dagger} \right\rangle = \mathrm{Tr}(Ca^{\dagger}C^{\dagger}CbC^{\dagger}) = \mathrm{Tr}(a^{\dagger}b) = \langle a, b \rangle.$$

b) Let $C = e^{tc}$ in the preceding and differentiate with respect to t at $t = 0$.

$\square$

*There is an intrinsic real bilinear form on a Lie algebra, called the **Killing form**, denoted $K : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$. The Killing form is defined by

$$K(a, b) := \mathrm{Tr}(\mathrm{ad}_a \mathrm{ad}_b).$$

Proposition 9.5 (Ad and ad invariance of Killing form)**.**

a) For all $C \in G$ and all $a, b \in \mathfrak{g}$,

$$K(\mathrm{Ad}_C a, \mathrm{Ad}_C b) = K(a, b).$$

b) For all $c \in \mathfrak{g}$ and all $a, b \in \mathfrak{g}$,

$$K(\mathrm{ad}_c a, b) + K(a, \mathrm{ad}_c b) = 0.$$

Proof.

a) Since $\mathrm{ad}_{\mathrm{Ad}_C a} = \mathrm{Ad}_C \mathrm{ad}_a \mathrm{Ad}_{C^{-1}}$ (check!), we have that

$$K(\mathrm{Ad}_C a, \mathrm{Ad}_C b) = \mathrm{Tr}(\mathrm{ad}_{\mathrm{Ad}_C a} \mathrm{ad}_{\mathrm{Ad}_C b}) = \mathrm{Tr}(\mathrm{Ad}_C \mathrm{ad}_a \mathrm{Ad}_{C^{-1}} \mathrm{Ad}_C \mathrm{ad}_b \mathrm{Ad}_{C^{-1}}) = \mathrm{Tr}(\mathrm{ad}_a \mathrm{ad}_b) = \langle a, b \rangle.$$

b) Set $C = \exp(tc)$ and differentiate at $t = 0$.

$\square$

Proposition 9.6. If $\mathfrak{g}$ is compact simple, then its Killing form is negative definite. In fact,

$$K(a, b) = \lambda \langle a, b \rangle$$

for some $\lambda < 0$.

Proof. Ad is irreducible, from Proposition 9.3. K commutes with Ad. Therefore, K is multiple of the identity. $\square$

9.1 Cartan subalgebra

The Cartan subalgebra generalises the role played by e_3 in our discussion of su(2).

A **subalgebra** of $\mathfrak{g}$ is a subspace $\mathfrak{h}$ which is itself a Lie algebra, i.e. it is closed under the Lie bracket (ideals are subalgebras, but subalgebras are not necessarily ideals). A subalgebra is **abelian** if all of its elements commute; that is, the Lie bracket between its elements vanishes.

Given a subset $S \subset \mathfrak{g}$, the **commutant** of S, denoted S', is the set of all elements of $\mathfrak{g}$ that commute with S; that is, $a \in S'$ if and only if $[a, s] = 0$ for all $s \in S$. Clearly, if $\mathfrak{a}$ is an abelian subalgebra, then $\mathfrak{a} \subset \mathfrak{a}'$.

Definition 9.7. Let $\mathfrak{g} \subset \mathbb{C}_{-}^{n\times n}$ by a compact simple Lie algebra. A **Cartan subalgebra** $\mathfrak{h}$ is an abelian subalgebra of $\mathfrak{g}$ such that

$$\mathfrak{h}' = \mathfrak{h}.$$

That is, a Cartan subalgebra is a maximal abelian subalgebra; if $[h_*, h] = 0$ for all $h \in \mathfrak{h}$, then $h_* \in \mathfrak{h}$ (maximality).

Example 9.8. For su(n), the Lie algebra of traceless antihermitian matrices, a Cartan subalgebra is given by the subset of traceless imaginary diagonal matrices.

A Cartan subalgebra may be constructed as follows. Let $\mathfrak{a}$ denote a nontrivial abelian subalgebra of $\mathfrak{g}$ of dimension m; for example we could take $\mathfrak{a} = \mathrm{span}(a)$, for some nonzero $a \in \mathfrak{g}$, in which case $m = 1$. If $\mathfrak{a} = \mathfrak{a}'$, then $\mathfrak{a}$ is a Cartan subalgebra. Otherwise, choose $b \in \mathfrak{a}' - \mathfrak{a}$. Then $\mathfrak{b} = \mathrm{span}(\mathfrak{a}, b)$ is an abelian subalgebra of dimension $m + 1$. We may repeat this procedure until we obtain an abelian subalgebra $\mathfrak{h}$ with $\mathfrak{h} = \mathfrak{h}'$, i.e. a Cartan subalgebra.

Let $\mathfrak{g}$ be a compact simple Lie algebra, and let $\mathfrak{h} \subset \mathfrak{g}$ be a Cartan subalgebra. Clearly $\mathfrak{h} \neq \mathfrak{g}$; otherwise, $\mathfrak{g}$ would be abelian, and therefore not simple.

Proposition 9.9. If $\mathfrak{g}$ is simple and $\mathfrak{h}$ is a Cartan subalgebra, then so is $\mathrm{Ad}_A \mathfrak{h}$. The Cartan subalgebra is not unique.

Proof. See exercises. First show that if $\mathfrak{h}$ is a Cartan subalgebra, then so is $\mathrm{Ad}_A \mathfrak{h}$ for all $A \in G$. Then show that if $\mathrm{Ad}_A \mathfrak{h} = \mathfrak{h}$ for all $A \in G$, then $\mathfrak{h}$ is a nontrivial ideal. □

However, for compact simple Cartan subalgebras, the Cartan subalgebra is determined up to the Adjoint action:

Theorem 9.10. Let $\mathfrak{g}$ be simple compact. If $\mathfrak{h}_1$ and $\mathfrak{h}_2$ are two Cartan subalgebras of $\mathfrak{g}$, then $\mathfrak{h}_1 = \mathrm{Ad}_A \mathfrak{h}_2$ for some $A \in G$.

Proof. See Problem Sheet 3, where a variational argument is outlined; one can establish that $\mathrm{Ad}_A \mathfrak{h}_1 \cap \mathfrak{h}_2 \neq 0$ for some $A \in G$ (which in turn implies that $\mathrm{Ad}_A \mathfrak{h}_1 = \mathfrak{h}_2$) by minimising the distance $||\,\mathrm{Ad}_A h_2 - h_1||^2$ over A for appropriately chosen nonzero $h_1 \in \mathfrak{h}$ and $h_2 \in \mathfrak{h}_2$. □

Example 9.11. For su(n), the Cartan subalgebras are unitary conjugates of traceless diagonal imaginary matrices; that is, they are sets of matrices of the form $U_0 d U_0^\dagger$, where $U_0 \in \mathrm{SU}(n)$ is fixed and d ranges over the set of traceless diagonal imaginary matrices.

Definition 9.12. The **rank** of a compact simple Lie algebra is the dimension of its Cartan subalgebra.

Clearly, if r is the rank of $\mathfrak{g}$ and d is the dimension of $\mathfrak{g}$, then $r < d$ (otherwise, $\mathfrak{g}$ would be abelian).

10 Weights and roots

10.1 Definitions and basic properties

Let $(\hat{\Gamma}, V)$ denote a representation of a compact simple Lie algebra $\mathfrak{g} \subset \mathbb{C}_{-}^{n\times n}$. It turns out that without loss of generality, we may assume that V is an inner product space and that $\hat{\Gamma}$ is antihermitian. (We won't prove this, but it can be shown by constructing a corresponding representation of a compact group. We know that representations of compact groups are equivalent to unitary representations, and therefore representations of the Lie algebra are equivalent to antihermitian representations.) Let d denote the dimension of $\mathfrak{g}$, and let r denote its rank.

Let $\mathfrak{h} \subset \mathfrak{g}$ be a Cartan subalgebra of $\mathfrak{g}$. It is convenient to introduce a basis $h_1, \ldots, h_r$ that is orthonormal with respect to the inner product on $\mathbb{C}^{n\times n}$:

$$\operatorname{tr}\left(h_j^\dagger h_k\right) = -\operatorname{tr}\left(h_j h_k\right) = \delta_{jk}. \tag{45}$$

Let

$$\hat{H}_j = \hat{\Gamma}(h_j). \tag{46}$$

We will use the fact that if a set of diagonalisable matrices commute with each other, then they can be simultaneously diagonalised. That is, there exists a basis in which the matrices are all diagonal. Moreover, if the matrices are antihermitian, the basis can be taken to be orthonormal. We will apply this result to the basis h_j for the Cartan subalgebra and to its representations $\hat{H}_j$, but we will state it a bit more neutrally, and generally, as follows.

Proposition 10.1. Let $a_1, \ldots, a_n \in L(V)$ be antihermitian maps on a complex inner product space V, and suppose that

$$[a_j, a_k] = 0, \text{ for all } 1 \le j, k \le n.$$

Let $\boldsymbol{\mu} \in \mathbb{R}^n$, and let $W(\boldsymbol{\mu}) \subset V$ be the subspace of simultaneous eigenvectors of the a_j's with eigenvalues $i\mu_j$. That is,

$$W(\boldsymbol{\mu}) = \{v \in V \,|\, a_j v = i\mu_j v\}.$$

Note that if $i\mu_j$ is not an eigenvalue of a_j for some j, then $W(\boldsymbol{\mu})$ is the trivial subspace consisting of 0. Then

$$V = \sum_{\boldsymbol{\mu}} \oplus W(\boldsymbol{\mu}),$$

where the sum is taken over $\boldsymbol{\mu}$'s for which $i\mu_j$ is an eigenvalue of a_j for all j. Morever, the distinct $W(\boldsymbol{\mu})$'s are orthogonal.

Proof. By induction on n. For $n = 1$, this is just the statement that eigenvectors of an antihermitian map a_1 can be chosen to form an orthogonal basis for V. Suppose it holds for $n-1$, and let

$$V = \sum_{\boldsymbol{\nu}} \oplus X(\boldsymbol{\nu}),$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n-1}$ and $a_j v = i\nu_j v$ for $v \in X(\boldsymbol{\nu})$ and $1 \le j \le n-1$. The distinct $X(\boldsymbol{\nu})$'s are orthogonal.

Let $v \in X(\boldsymbol{\nu})$. Claim that $a_n v \in X(\boldsymbol{\nu})$. To verify, we check that

$$a_j a_n v = a_n a_j v = i\nu_j a_n v,$$

as required. Then $X(\boldsymbol{\nu})$ is invariant under a_n, so that $a_n \in L(X(\boldsymbol{\nu}))$. As a_n is antihermitian on V, it follows that $\langle u, a_n v\rangle = -\langle a_n u, v\rangle$, and this relation continues to hold if $u, v \in X(\boldsymbol{\nu})$, so that a_n is antihermitian on $X(\boldsymbol{\nu})$. Therefore, $X(\boldsymbol{\nu})$ can be decomposed into a direct sum of orthogonal eigenspaces of a_n, as follows:

$$X(\boldsymbol{\nu}) = \sum_{\mu_n} \oplus X(\boldsymbol{\nu}; \mu_n),$$

so that $a_n v = i\mu_n v$ for $v \in X(\boldsymbol{\nu}; \mu_n)$. Let $\boldsymbol{\mu} := (\boldsymbol{\nu}, \mu_n)$ and $W(\boldsymbol{\mu}) := X(\boldsymbol{\nu}; \mu_n)$, and the result follows. □

We may apply this to the representation of the Cartan subalgebra. For $\boldsymbol{\mu} \in \mathbb{R}^r$, let

$$W(\boldsymbol{\mu}) = \{v \in V \,|\, \hat{H}_j v = i\mu_j v\}. \tag{47}$$

If $W(\boldsymbol{\mu}) \ne \{0\}$, we say that $\boldsymbol{\mu}$ is a *weight* of the representation $\hat{\Gamma}$.

Let us consider in particular the adjoint representation. In this case, we denote the representatives of the Cartan subalgebra basis by $\hat{h}_j$, and the weights by $\boldsymbol{\alpha} \in \mathbb{R}^r$. Clearly, $0 \in \mathbb{R}^r$ is a weight of the adjoint representation, since $h_j \in W(0)$. Nonzero weights of the adjoint representation are called *roots.* Let us denote eigenvectors of the $\hat{h}_j$'s by $e_{\boldsymbol{\alpha}}$. These satisfy

$$\hat{h}_j e_{\boldsymbol{\alpha}} = [h_j, e_{\boldsymbol{\alpha}}] = i\alpha_j e_{\boldsymbol{\alpha}}. \tag{48}$$

In general, the $e_{\boldsymbol{\alpha}}$'s do not belong to $\mathfrak{g}$, but rather to the complexification of $\mathfrak{g}$, which we denote by $\mathfrak{g}_{\mathbb{C}}$. We can see this as follows: Taking hermitian conjugates in the preceding, and noting that h_j is antihermitian, we get that

$$[h_j, e_{\boldsymbol{\alpha}}^\dagger] = -i\alpha_j e_{\boldsymbol{\alpha}}^\dagger, \tag{49}$$

which implies that $e^\dagger_{\boldsymbol\alpha}$ is an eigenvector of the adjoint representation of the Cartan subalgebra with root $-\boldsymbol\alpha$, and is therefore orthogonal to $e_{\boldsymbol\alpha}$. Thus $e^\dagger_{\boldsymbol\alpha} \neq -e_{\boldsymbol\alpha}$. Without loss of generality, we may normalise the $e_{\boldsymbol\alpha}$'s so that

$$||e_{\boldsymbol\alpha}||^2 = ||e^\dagger_{\boldsymbol\alpha}||^2 = \mathrm{Tr}(e^\dagger_{\boldsymbol\alpha} e_{\boldsymbol\alpha}) = 1.$$

The following shows that the Lie bracket of $e_{\boldsymbol\alpha}$ and $e^\dagger_{\boldsymbol\alpha}$ belongs to the Cartan subalgebra, and provides an explicit formula:

Proposition 10.2.

$$[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}] = -i\boldsymbol\alpha\cdot\mathbf{h}.$$

Proof. From the Jacobi identity,

$$[h_j,[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}]] = [[h_j, e_{\boldsymbol\alpha}]e^\dagger_{\boldsymbol\alpha}] + [e_{\boldsymbol\alpha},[h_j, e^\dagger_{\boldsymbol\alpha}]] = i\alpha_j[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}] - i\alpha_j[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}] = 0.$$

Since $[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}]$ commutes $\mathfrak{h}$, it must belong to $\mathfrak{h}$. We may write that

$$[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}] = \sum_j c_j h_j.$$

Taking inner products with h_k, we obtain

$$c_k = \left\langle h_k,[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}]\right\rangle = -\,\mathrm{Tr}\left(h_k[e_{\boldsymbol\alpha}, e^\dagger_{\boldsymbol\alpha}]\right) = -\,\mathrm{Tr}([h_k, e_{\boldsymbol\alpha}]e^\dagger_{\boldsymbol\alpha}) = -i\alpha_k.$$

□

Let us return to considering a general representation $\hat\Gamma$. Let $\hat E_{\boldsymbol\alpha}$ denote the representative of $e_{\boldsymbol\alpha}$ (obtained from the complexification of $\hat\Gamma$ – that is, if $a,b\in\mathfrak{g}$, we define $\hat\Gamma(a+ib) := \hat\Gamma(a) + i\hat\Gamma(b)$). Then

$$[\hat H_j, \hat E_{\boldsymbol\alpha}] = i\alpha_j\hat E_{\boldsymbol\alpha}, \tag{50}$$

and

$$[\hat E_{\boldsymbol\alpha}, \hat E^\dagger_{\boldsymbol\alpha}] = -i\boldsymbol\alpha\cdot\hat{\mathbf{H}}. \tag{51}$$

Proposition 10.3.

$$\hat E_{\boldsymbol\alpha} W(\boldsymbol\mu) \subset W(\boldsymbol\mu+\boldsymbol\alpha)$$

Proof. Let $v\in W(\boldsymbol\mu)$. Then

$$\hat H_j\hat E_{\boldsymbol\alpha} v = \left(\left[\hat H_j, \hat E_{\boldsymbol\alpha}\right] + \hat E_{\boldsymbol\alpha}\hat H_j\right)v = \left(i\alpha_j\hat E_{\boldsymbol\alpha} + \hat E_{\boldsymbol\alpha} i\mu_j\right)v = i(\alpha_j+\mu_j)\hat E_{\boldsymbol\alpha} v, \tag{52}$$

which implies that $v\in W(\boldsymbol\mu+\boldsymbol\alpha)$. □

Note: It could be that $\hat E_{\boldsymbol\alpha} W(\boldsymbol\mu) = \{0\}$, even if $W(\boldsymbol\mu)\neq\{0\}$ and $W(\boldsymbol\mu+\boldsymbol\alpha)\neq\{0\}$.

Corollary 10.4.

$$\hat E^\dagger_{\boldsymbol\alpha} W(\boldsymbol\mu) \subset W(\boldsymbol\mu-\boldsymbol\alpha) \tag{53}$$

Proof. This follows from taking hermitian conjugates in Proposition 10.3. □

Proposition 10.3 has the following consequence when applied to the adjoint representation:

Corollary 10.5. If $\boldsymbol\alpha$, $\boldsymbol\beta$ are roots of a compact simple Lie algebra $\mathfrak{g}$, then either $\boldsymbol\alpha+\boldsymbol\beta$ is also a root of $\mathfrak{g}$, or else

$$[e_{\boldsymbol\alpha}, e_{\boldsymbol\beta}] = \hat e_{\boldsymbol\alpha}(e_{\boldsymbol\beta}) = 0.$$

Proof. We give a direct argument as follows:

$$[h_j,[e_{\boldsymbol\alpha}, e_{\boldsymbol\beta}]] = [[h_j, e_{\boldsymbol\alpha}], e_{\boldsymbol\beta}] + [e_{\boldsymbol\alpha},[h_j, e_{\boldsymbol\beta}]] = i(\alpha_j+\beta_j)[e_{\boldsymbol\alpha}, e_{\boldsymbol\beta}].$$

Thus, if $[e_{\boldsymbol\alpha}, e_{\boldsymbol\beta}]$ is nonvanishing, then $\boldsymbol\alpha+\boldsymbol\beta$ is a root. □

For fixed $\boldsymbol{\alpha}$, the three elements of the representation, $\hat{E}_{\boldsymbol{\alpha}}$, $\hat{E}^{\dagger}_{\boldsymbol{\alpha}}$, and $\boldsymbol{\alpha}\cdot\hat{\mathbf{H}}$, form a subalgebra of $\hat{\Gamma}(\mathfrak{g})$ isomorphic to su(2). Explicitly, if we let $\alpha = ||\boldsymbol{\alpha}||$ and define

$$\hat{E}_{+} := \frac{1}{\alpha}\hat{E}_{\boldsymbol{\alpha}}, \quad \hat{E}_{-} := \frac{1}{\alpha}\hat{E}^{\dagger}_{\boldsymbol{\alpha}}, \quad \hat{E}_{3} := \frac{1}{\alpha^2}\boldsymbol{\alpha}\cdot\hat{\mathbf{H}}, \tag{54}$$

then it is straightforward to show that

$$[\hat{E}_3, \hat{E}_{\pm}] = \pm i\hat{E}_{\pm}, \quad [\hat{E}_{+}, \hat{E}_{-}] = -i\hat{E}_3. \tag{55}$$

Therefore, by following the construction of irreducible representations of su(2) in Section 8, we can decompose V into a direct sum of subspaces each of which is invariant under $\hat{E}_{\boldsymbol{\alpha}}$, $\hat{E}^{\dagger}_{\boldsymbol{\alpha}}$, and $\boldsymbol{\alpha}\cdot\mathbf{H}$, as follows: Take nonvanishing $v_* \in W(\boldsymbol{\mu})$. Then $\hat{E}_3 v_* = i\mu_* v_*$, where

$$\mu_* = \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\mu}}{\alpha^2}.$$

Let p be the smallest nonnegative integer for which $\hat{E}_{+}^{p+1}v_* = 0$, and let $v(p) = \hat{E}_{+}^{p+1}v_*$, which we may assume to have unit norm. From Proposition 8.2, $\hat{E}_{-}^{j}v(p) \neq 0$ for $0 \leq j \leq p$. Let q be the smallest nonnegative integer for which $\hat{E}_{-}^{p+q+1}v(p) = 0$. We define $v(a)$, $-q \leq a \leq p$ inductively, starting with $v(p)$, by

$$v(a-1) = \frac{1}{N(a)}\hat{E}_{-}v(a), \quad N(a) := \left(\left\langle \hat{E}_{-}v(a), \hat{E}_{-}v(a)\right\rangle\right)^{1/2},$$

where $N(a)$ is chosen so that $v(a)$ is normalised. Then from Proposition 8.3

$$\begin{aligned}
\hat{E}_{-}v(a) &= N(a)v(a-1), \\
\hat{E}_{+}v(a) &= N(a+1)v(a+1), \\
\hat{E}_{3}v(a) &= \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\mu} + a\alpha^2}{\alpha^2}v(a).
\end{aligned}$$

The coefficients $N^2(a)$ satisfy the recurrence relation

$$N^2(a+1) = N^2(a) - \mu_* - a.$$

A necessary condition for this relation to have a solution is that

$$\boxed{q - p = 2\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\mu}}{\alpha^2}.} \tag{56}$$

Eq. (56) turns out to be very important, as it leads to strong constraints on the structure of compact simple Lie algebras and their representations. Given Eq. (56) , $N^2(a)$ is given by

$$N^2(a) = \tfrac{1}{2}(p+1-a)(q+a). \tag{57}$$

One consequence of Eq. (56) is the following:

Proposition 10.6. The roots of a compact simple Lie algebra $\mathfrak{g}$ are nondegenerate. That is, if $\boldsymbol{\alpha}$ is a root of $\mathfrak{g}$ and $W(\boldsymbol{\alpha}) \subset \mathfrak{g}_{\mathbb{C}}$ is the corresponding eigensubspace of the complexified Lie algebra, then $\dim W(\boldsymbol{\alpha}) = 1$.

Proof. Suppose to the contrary that $\dim W(\boldsymbol{\alpha}) > 1$. Then we may choose linearly independent vectors $e_{\boldsymbol{\alpha}}$ and v_* in $W(\boldsymbol{\alpha})$. Without loss of generality we may assume that $e_{\boldsymbol{\alpha}}$ and v_* are orthogonal. Choose $p \geq 0$ such that

$$\hat{e}^{p}_{\boldsymbol{\alpha}}v_* \neq 0, \quad \hat{e}^{p+1}_{\boldsymbol{\alpha}}v_* = 0,$$

and let

$$f_{\boldsymbol{\alpha}} = \left(\hat{e}^{\dagger}_{\boldsymbol{\alpha}}\right)^{p}\hat{e}^{p}_{\boldsymbol{\alpha}}v_*.$$

By Proposition 8.2, $f_{\boldsymbol{\alpha}} \in W(\boldsymbol{\alpha})$ and $f_{\boldsymbol{\alpha}} \neq 0$. Also,

$$\langle e_{\boldsymbol{\alpha}}, f_{\boldsymbol{\alpha}}\rangle = \langle \hat{e}^{p}_{\boldsymbol{\alpha}}\, e_{\boldsymbol{\alpha}}, \hat{e}^{p}_{\boldsymbol{\alpha}}\, v_*\rangle = 0, \tag{58}$$

since, if $p = 0$, this follows from the assumption that $e_{\boldsymbol{\alpha}}$ and v_* are orthogonal, while if $p > 0$, this follows from the fact that $\hat{e}_{\boldsymbol{\alpha}}(e_{\boldsymbol{\alpha}}) = [e_{\boldsymbol{\alpha}}, e_{\boldsymbol{\alpha}}] = 0$. Choose $q \geq 0$ so that

$$\left(\hat{e}_{\boldsymbol{\alpha}}^{\dagger}\right)^{q} (f_{\boldsymbol{\alpha}}) \neq 0, \quad \left(\hat{e}_{\boldsymbol{\alpha}}^{\dagger}\right)^{q+1} (f_{\boldsymbol{\alpha}}) = 0.$$

Then from (56) with $\boldsymbol{\mu} = \boldsymbol{\alpha}$, we get that

$$q - p = 2.$$

In what follows, we show that $q = 0$, which yields the contradiction that establishes the required result.

Let

$$a = [e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}].$$

Noting that

$$[h_j, e_{\boldsymbol{\alpha}}^{\dagger}] = -i\alpha_j e_{\boldsymbol{\alpha}}^{\dagger},$$

which follows from simple calculation, we can use the Jacobi identity to show that

$$[h_j, a] = 0,$$

which holds for all j. Indeed,

$$[h_j, a] = [h_j, [e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}]] = [[h_j, e_{\boldsymbol{\alpha}}^{\dagger}], f_{\boldsymbol{\alpha}}] + [e_{\boldsymbol{\alpha}}^{\dagger}, [h_j, f_{\boldsymbol{\alpha}}]] = i\alpha_j [e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}] - i\alpha_j [e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}] = 0.$$

This implies that a commutes with all elements of the Cartan subalgebra $\mathfrak{h}$, which in turn implies that a belongs to $\mathfrak{h}$, i.e.

$$a = \mathbf{c} \cdot \mathbf{h}$$

for some $\mathbf{c} \in \mathbb{R}^r$. Since the h_j's constitute an orthonormal basis for $\mathfrak{h}$, it follows that

$$c_j = \langle h_j, a \rangle = -\operatorname{tr}(h_j a) = -\operatorname{tr}\left(h_j [e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}]\right) = -\operatorname{tr}\left(e_{\boldsymbol{\alpha}}^{\dagger}[h_j, f_{\boldsymbol{\alpha}}]\right) = i\alpha_j \operatorname{tr}(e_{\boldsymbol{\alpha}}^{\dagger} f_{\boldsymbol{\alpha}}) = i\alpha_j \langle e_{\boldsymbol{\alpha}}, f_{\boldsymbol{\alpha}} \rangle = 0,$$

since, from (58), $e_{\boldsymbol{\alpha}}$ and $f_{\boldsymbol{\alpha}}$ are orthogonal. Thus,

$$[e_{\boldsymbol{\alpha}}^{\dagger}, f_{\boldsymbol{\alpha}}] = \hat{e}_{\boldsymbol{\alpha}}^{\dagger}(f_{\boldsymbol{\alpha}}) = 0,$$

or $q = 0$, as claimed. □

Thus, given a root $\boldsymbol{\alpha}$, the associated eigenvector $e_{\boldsymbol{\alpha}} \in \mathfrak{g}_{\mathbb{C}}$ is unique up to a scalar multiple. It follows that $e_{\boldsymbol{\alpha}}^{\dagger}$ and $e_{-\boldsymbol{\alpha}}$ are proportional. Without loss of generality, we may assume that

$$e_{-\boldsymbol{\alpha}} = e_{\boldsymbol{\alpha}}^{\dagger}. \tag{59}$$

Finally, suppose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are distinct roots of a compact simple Lie algebra. Let p, q be the smallest nonnegative integers such that

$$(\hat{e}_{\boldsymbol{\alpha}})^{p+1} e_{\boldsymbol{\beta}} = 0, \quad (\hat{e}_{\boldsymbol{\alpha}}^{\dagger})^{q+1} e_{\boldsymbol{\beta}} = 0.$$

From (56),

$$m := q - p = \frac{2\boldsymbol{\alpha} \cdot \boldsymbol{\beta}}{\alpha^2}. \tag{60}$$

Similarly, let $p', q' \geq 0$ be the smallest nonnegative integers such that

$$(\hat{e}_{\boldsymbol{\beta}})^{p'+1} e_{\boldsymbol{\alpha}} = 0, \quad (\hat{e}_{\boldsymbol{\beta}}^{\dagger})^{q'+1} e_{\boldsymbol{\alpha}} = 0,$$

From (56),

$$n := q' - p' = \frac{2\boldsymbol{\beta} \cdot \boldsymbol{\alpha}}{\beta^2}. \tag{61}$$

From the product of (60) and (61), we get

$$\frac{(\boldsymbol{\alpha} \cdot \boldsymbol{\beta})^2}{\alpha^2 \beta^2} = \frac{(q-p)(q'-p')}{4}. \tag{62}$$

This may be understood to say that $\cos^2\theta$, where θ denotes the angle between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, is constrained to be an integer multiple mm' of $1/4$. The possible values of $\cos\theta$ and θ are summarised in Table 1.

Table 1: Angles between roots

mn	$\cos\theta$	θ
0	0	$\pi/2$
1	$\pm 1/2$	$\pi/3$ or $2\pi/3$
2	$\pm 1/\sqrt{2}$	$\pi/4$ or $3\pi/4$
3	$\pm\sqrt{3}/2$	$\pi/6$ or $5\pi/6$
4	± 1	0 or π

Example 10.7 (Roots of su(3))**.** su(3) is the space of traceless antihermitian 3×3 matrices. We may take a Cartan subalgebra, $\mathfrak{h}$, to consist of traceless imaginary diagonal 3×3 matrices. The space of such matrices is two dimensional, so that the rank r of su(3) is equal to 2. An orthonormal basis is given by

$$h_1 = \frac{i}{\sqrt{6}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad h_2 = \frac{i}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then

$$\langle h_j, h_k \rangle = \delta_{jk}.$$

su(3) may be regarded as a representation of itself on $V = \mathbb{C}^3$; this is the natural, or fundamental representation. The simultaneous eigenvectors of h_1 and h_2 are just the basis vectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The weights $\boldsymbol{\mu} \in \mathbb{R}^2$ are obtained from the diagonal elements of h_1 and h_2:

$$\boldsymbol{\mu}_1 = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{2}}\right), \quad \boldsymbol{\mu}_2 = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}}\right), \quad \boldsymbol{\mu}_3 = \left(-\frac{2}{\sqrt{6}}, 0\right).$$

Next, we consider the adjoint representation of su(3). Eigenvectors of the adjoint representation of $\mathfrak{h}$ are traceless 3×3 matrices $e_{\boldsymbol{\alpha}_p} \in \mathrm{su}_{\mathbb{C}}(3)$ which satisfy

$$\hat{h}_j e_{\boldsymbol{\alpha}_p} = [h_j, e_{\boldsymbol{\alpha}_p}] = i\boldsymbol{\alpha}_{pj} e_{\boldsymbol{\alpha}_p}. \tag{63}$$

Since h_j is diagonal, the (r,s)th element of the preceding equation has the following simple form:

$$\left((h_j)_{rr} - (h_j)_{ss} - i\alpha_{pj}\right)\left(e_{\boldsymbol{\alpha}_p}\right)_{rs} = 0.$$

As this must hold for all $1 \le r, s \le 3$, we may conclude that solutions $e_{\boldsymbol{\alpha}_p}$ are either diagonal matrices (in which case, $\boldsymbol{\alpha}_p = 0$ and $e_{\boldsymbol{\alpha}_p}$ belongs to $\mathfrak{h}_{\mathbb{C}}$), or else $e_{\boldsymbol{\alpha}_p}$ contains a single nonzero off-diagonal element. Three solutions – the upper triangular ones – with nonzero eigenvalues (i.e. roots) are therefore given by

$$e_{\boldsymbol{\alpha}_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_{\boldsymbol{\alpha}_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_{\boldsymbol{\alpha}_3} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The associated roots are

$$\begin{aligned} \boldsymbol{\alpha}_1 &= \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3 = \left(\sqrt{\frac{3}{2}}, -\frac{1}{\sqrt{2}}\right), \\ \boldsymbol{\alpha}_2 &= \boldsymbol{\mu}_1 - \boldsymbol{\mu}_3 = \left(\sqrt{\frac{3}{2}}, \frac{1}{\sqrt{2}}\right), \\ \boldsymbol{\alpha}_3 &= \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \left(0, \sqrt{2}\right). \end{aligned}$$

The lower triangular matrices $e^\dagger_{\boldsymbol{\alpha}_1}$, $e^\dagger_{\boldsymbol{\alpha}_2}$, $e^\dagger_{\boldsymbol{\alpha}_3}$ are also solutions of (63) with roots $\boldsymbol{\alpha}_1$, $-\boldsymbol{\alpha}_2$ and $-\boldsymbol{\alpha}_3$.

The roots of the adjoint representation and the weights of the fundamental representation are shown below.

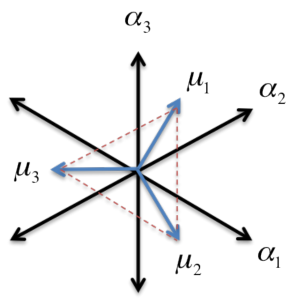


Figure 8: Weights of the fundamental representation and roots of the adjoint representation of su(3). The differences (dotted red) between pairs of weights (blue) gives the roots (black).

10.2 Simple roots

We recall that if $\boldsymbol{\alpha}$ is a root, then so is $-\boldsymbol{\alpha}$ (since if $e_{\boldsymbol{\alpha}}$ is an eigenvector of $\hat{h}_j$ with eigenvalue $i\alpha_j$, then $e_{\boldsymbol{\alpha}}^\dagger$ is an eigenvector of $\hat{h}_j$ with eigenvalue $-i\alpha_j$). Thus, roots come in signed pairs.

We introduce a lexicographic ordering on the roots as follows. Given $\boldsymbol{\alpha} \in \mathbb{R}^r$, we say that $\boldsymbol{\alpha}$ is **positive** if the first nonzero component of $\boldsymbol{\alpha}$ is positive. $\boldsymbol{\alpha}$ is **negative** if $-\boldsymbol{\alpha}$ is positive. We say that $\boldsymbol{\alpha} > \boldsymbol{\beta}$ if $\boldsymbol{\alpha} - \boldsymbol{\beta}$ is positive.

Note that the definition of positivity depends on the choice of basis for $\mathfrak{h}$. Nevertheless, using this definition, we can derive intrinsic (i.e., basis-independent) properties of $\mathfrak{g}$.

We note that since roots are nonzero, they are either positive or negative; from the observation above, they come in positive/negative pairs.

Definition 10.8 (Simple root)**.** A root $\boldsymbol{\alpha}$ is **simple** if $\boldsymbol{\alpha}$ is positive and $\boldsymbol{\alpha}$ cannot be expressed as the some of positive roots.

It is easily seen that every positive root can be expressed as a sum of simple roots. For suppose $\boldsymbol{\alpha}$ is positive. If $\boldsymbol{\alpha}$ is simple, we are done. If not, $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$, where $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ are positive. Then $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 < \boldsymbol{\alpha}$. If either $\boldsymbol{\alpha}_1$ or $\boldsymbol{\alpha}_2$ are not simple, we may express them as sums of smaller, positive roots. This reduction must terminate with simple roots.

Example 10.9 (Simple roots of su(3))**.** Referring to Example 10.9, $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ are the positive roots of su(3). $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_3$ are simple, while $\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_3$, so that $\boldsymbol{\alpha}_2$ is not simple.

Proposition 10.10. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are simple, then $\boldsymbol{\alpha} - \boldsymbol{\beta}$ and $\boldsymbol{\beta} - \boldsymbol{\alpha}$ are not roots.

Proof. Since roots come in signed pairs, if either $\boldsymbol{\alpha} - \boldsymbol{\beta}$ or $\boldsymbol{\beta} - \boldsymbol{\alpha}$ is a root, then so is the other, and one of them will be positive. Suppose for definiteness that $\boldsymbol{\alpha} - \boldsymbol{\beta}$ is positive (otherwise, interchange $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the following). Then $\boldsymbol{\alpha}$ is given by the sum of positive roots, namely $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} - \boldsymbol{\beta}$. This contradicts the assumption that $\boldsymbol{\alpha}$ is simple.

□

Proposition 10.11. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are distinct simple roots, then

$$[e_{\boldsymbol{\alpha}}, e_{-\boldsymbol{\beta}}] = [e_{-\boldsymbol{\alpha}}, e_{\boldsymbol{\beta}}] = 0.$$

Proof.

$$\hat{e}_{-\boldsymbol{\alpha}} e_{\boldsymbol{\beta}} = [e_{-\boldsymbol{\alpha}}, e_{\boldsymbol{\beta}}] \in W(\boldsymbol{\beta} - \boldsymbol{\alpha}).$$

From Proposition 10.10, $W(\boldsymbol{\beta} - \boldsymbol{\alpha}) = \{0\}$. A similar argument shows that $[e_{\boldsymbol{\alpha}}, e_{-\boldsymbol{\beta}}] = 0$. □

Proposition 10.12. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are distinct simple roots, then

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} \leq 0.$$

Proof. Let p be the smallest nonnegative integer for which $e_{\boldsymbol{\alpha}}^{p+1} e_{\boldsymbol{\beta}} = 0$, and q the smallest nonnegative integer for which $\hat{e}_{-\boldsymbol{\alpha}}^{q+1} e_{\boldsymbol{\beta}} = 0$. From (56),

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \alpha^2 \, \frac{q-p}{2}.$$

But since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are simple, it follows from Proposition 10.11 that $q = 0$. The result follows. □

Proposition 10.13. The simple roots are linearly independent.

Proof. Suppose that

$$\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} \boldsymbol{\alpha} = 0$$

for $c_{\boldsymbol{\alpha}} \in \mathbb{R}$. We can partition the sum into terms with positive and negative coefficients $c_{\boldsymbol{\alpha}}$. Let

$$P = \sum_{\substack{\boldsymbol{\beta} \\ c_{\boldsymbol{\beta}} > 0}} c_{\boldsymbol{\beta}} \boldsymbol{\beta}, \quad N = \sum_{\substack{\boldsymbol{\gamma} \\ c_{\boldsymbol{\gamma}} < 0}} |c_{\boldsymbol{\gamma}}| \boldsymbol{\gamma}.$$

Then $P - N = 0$, or

$$P = N.$$

It follows that

$$P^2 = N \cdot P = \sum_{\substack{\boldsymbol{\beta} \\ c_{\boldsymbol{\beta}} > 0}} \sum_{\substack{\boldsymbol{\gamma} \\ c_{\boldsymbol{\gamma}} < 0}} c_{\boldsymbol{\beta}} \, |c_{\boldsymbol{\gamma}}| \boldsymbol{\beta} \cdot \boldsymbol{\gamma}.$$

The left-hand side is nonnegative, while the right-hand side is nonpositive. It follows that both must vanish. But a linear combination of positive roots with positive coefficients cannot vanish, since each term has its first nonzero element positive, so that $c_{\boldsymbol{\beta}} = 0$ and $|c_{\boldsymbol{\gamma}}| = 0$. □

Proposition 10.14. The simple roots span $\mathbb{R}^r$.

Proof. Suppose the contrary. Let $\boldsymbol{\gamma} \in \mathbb{R}^r$ be a nonzero vector orthogonal to all the simple roots. Since all roots can be written as linear combinations of simple roots, it follows that $\boldsymbol{\gamma}$ is orthogonal to every root. Consider the element h_* in the Cartan subalgebra given by

$$h_* = \boldsymbol{\gamma} \cdot \mathbf{h} = \sum_j \gamma_j h_j.$$

We have that

$$[h_*, e_{\boldsymbol{\alpha}}] = i \boldsymbol{\gamma} \cdot \boldsymbol{\alpha} = 0,$$

for all $\boldsymbol{\alpha}$. Since $\mathfrak{g}_{\mathbb{C}}$ is spanned by the $e_{\boldsymbol{\alpha}}$'s and h_j's, it follows that

$$[\xi, h_*] = 0,$$

for all $\xi \in \mathfrak{g}$. Therefore, the one-dimensional subspace spanned by h_* is a nontrivial ideal in $\mathfrak{g}$, which contradicts the assumption that $\mathfrak{g}$ is simple. □

It follows that the number of simple roots coincides with the rank r of $\mathfrak{g}$.

It turns out that the structure of $\mathfrak{g}$ is determined by the simple roots.

Proposition 10.15. Let $\boldsymbol{\gamma}$ be a positive root of $\mathfrak{g}$. Then

$$e_{\boldsymbol{\gamma}} = N \hat{e}_{\boldsymbol{\alpha}_{j_1}} \cdots \hat{e}_{\boldsymbol{\alpha}_{j_{s-1}}} e_{\boldsymbol{\alpha}_{j_s}} \tag{64}$$

for some sequence of simple roots $\boldsymbol{\alpha}_{j_1}, \ldots, \boldsymbol{\alpha}_{j_s}$ and some coefficient N, which will depend on the $\boldsymbol{\alpha}_{j_l}$'s.

Proof. Every positive root may be expressed as a sum of simple roots, so that

$$\boldsymbol{\gamma} = \sum_{l=1}^{s} \boldsymbol{\alpha}_{j_l}. \tag{65}$$

Moreover, since simple roots are linearly independent, the expression (65) is unique up to reordering.

We claim that (65) implies (64) up to a reordering of the roots in (65) (since, in (64), the order matters).

We proceed by induction on s. Clearly (64) holds if $s = 1$; in this case, $\boldsymbol{\gamma}$ is simple.

Take $s > 1$ and suppose the claim holds for $s - 1$. Let $\boldsymbol{\gamma}$ satisfy (65). We must have

$$\hat{e}_{-\boldsymbol{\beta}}\, e_{\boldsymbol{\gamma}} \neq 0$$

for some simple root $\boldsymbol{\beta}$. For if not, we would have, from Eq. (15.21), that

$$\frac{\boldsymbol{\beta} \cdot \boldsymbol{\gamma}}{\beta^2} = -\tfrac{1}{2}p$$

for some nonnegative p. Recalling the argument in Proposition 10.13, this would imply that $\boldsymbol{\gamma}$ is linearly independent of the simple roots, which cannot be.

Choose $\boldsymbol{\beta}$ so that

$$\hat{e}_{-\boldsymbol{\beta}}\, e_{\boldsymbol{\gamma}} = C e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} \neq 0.$$

Thus, $\boldsymbol{\gamma} - \boldsymbol{\beta}$ is a root, and since $s > 1$, it is a positive root. (Note that roots are either positive, in which case they can be expressed as a sum of simple roots with positive integral coefficients, or else negative, in which case they can be expressed as a sum of simple roots with negative integral coefficients. Thus, $s > 1$ rules out $\boldsymbol{\gamma} - \boldsymbol{\beta}$ being negative.) It follows that $\boldsymbol{\gamma} - \boldsymbol{\beta}$ has a unique expression as a sum of t simple roots. Since $(\boldsymbol{\gamma} - \boldsymbol{\beta}) + \boldsymbol{\beta} = \boldsymbol{\gamma}$, and $\boldsymbol{\gamma}$ is a sum of s simple roots, it follows that $t = s - 1$.

By the induction hypothesis, we may write $e_{\boldsymbol{\gamma}-\boldsymbol{\beta}}$ as

$$e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} = N_{-} \hat{e}_{\boldsymbol{\beta}_{j_1}} \cdots \hat{e}_{\boldsymbol{\beta}_{j_{s-2}}} e_{\boldsymbol{\alpha}_{j_{s-1}}}.$$

We claim that

$$\hat{e}_{\boldsymbol{\beta}} e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} \neq 0.$$

Indeed, the inner product of $\hat{e}_{\boldsymbol{\beta}} e_{\boldsymbol{\gamma}-\boldsymbol{\beta}}$ and $e_{\boldsymbol{\gamma}}$ is nonzero, as the following shows:

$$\left\langle e_{\boldsymbol{\gamma}}, \hat{e}_{\boldsymbol{\beta}} e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} \right\rangle = \left\langle \hat{e}_{-\boldsymbol{\beta}} e_{\boldsymbol{\gamma}}, e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} \right\rangle = C \left\langle e_{\boldsymbol{\gamma}-\boldsymbol{\beta}}, e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} \right\rangle = C.$$

Since by Proposition 10.6, roots are nondegenerate, it follows that

$$e_{\boldsymbol{\gamma}} = C \hat{e}_{\boldsymbol{\beta}} e_{\boldsymbol{\gamma}-\boldsymbol{\beta}} = C N_{-} \hat{e}_{\boldsymbol{\beta}} \hat{e}_{\boldsymbol{\beta}_{j_1}} \cdots \hat{e}_{\boldsymbol{\beta}_{j_{s-2}}} e_{\boldsymbol{\alpha}_{j_{s-1}}}.$$

□

Thus, the structure of $\mathfrak{g}$ is determined by the simple roots.

10.3 *Highest weight

We return to considering a general representation $(\hat{\Gamma}, V)$ of a compact simple Lie algebra $\mathfrak{g}$. We assume that V has a hermitian inner product, and that $\hat{\Gamma}$ is antihermitian with respect to this inner product. We let

$$\hat{H}_j = \hat{\Gamma}(h_j),$$
$$\hat{E}_{\boldsymbol{\alpha}} = \hat{\Gamma}(e_{\boldsymbol{\alpha}}).$$

Then
$$\hat{E}^{\dagger}_{\boldsymbol{\alpha}} = \hat{E}_{-\boldsymbol{\alpha}}.$$
Let $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r$ denote simple roots. From Proposition 10.15, since $\mathfrak{g}$ is generated by brackets of simple roots, it follows that $\hat{\Gamma}$ is completely determined by the representatives $\hat{\Gamma}_{\boldsymbol{\alpha}_j}$ of the simple roots.

Let
$$V = \oplus_{\boldsymbol{\mu}} W(\boldsymbol{\mu})$$
denote the decomposition of V into weight spaces $W(\boldsymbol{\mu})$. Thus, $\hat{H}_j$ restricted to $W(\boldsymbol{\mu})$ is just multiplication by $i\mu_j$, and $\hat{E}_{\boldsymbol{\alpha}} W(\boldsymbol{\mu}) \subset W(\boldsymbol{\mu} + \boldsymbol{\alpha})$. We can order weights lexicographically, as we order roots. That is, $\boldsymbol{\mu}$ is positive if its first nonzero component is positive. Let $\boldsymbol{\mu}_*$ denote the **highest weight** of $(\hat{\Gamma}, V)$. That is, $\boldsymbol{\mu}_* - \boldsymbol{\mu}$ is positive for all $\boldsymbol{\mu} \neq \boldsymbol{\mu}_*$ (it is clear that $\boldsymbol{\mu}_*$ exists and is unique).

Let $v_* \in W(\boldsymbol{\mu}_*)$. Let W_{*s} denote the subspace spanned by vectors obtained by applying up to s linear maps $\hat{E}_{-\boldsymbol{\alpha}_{jm}}$ (that is, hermitian conjugates of the representatives of the simple roots) to v_*. That is,
$$W_{*s} = \text{span}\left\{ E_{-\boldsymbol{\alpha}_{j_1}} \cdots E_{-\boldsymbol{\alpha}_{j_t}} v_* \,|\, 0 \le t \le s \right\}. \tag{66}$$
Clearly, W_{*s} is a subspace of $W_{*(s+1)}$. For sufficiently large s, we must have $W_{*s} = W_{*(s+1)}$ (otherwise, V would be infinite dimensional). Let W_* denote the space so obtained for large enough s.

We claim that W_* is invariant under $\hat{\Gamma}$. It suffices to show that W_* is invariant under $\hat{E}_{\boldsymbol{\alpha}_j}$, $\hat{H}_j$ and $\hat{E}_{-\boldsymbol{\alpha}_j}$. By construction, W_* is invariant under $\hat{E}_{-\boldsymbol{\alpha}_j}$. To establish invariance under $\hat{H}_j$ and $\hat{E}_{\boldsymbol{\alpha}_j}$, we proceed by induction on s. We have that $\hat{E}_{\boldsymbol{\alpha}_j} v_* = 0$ (since $\boldsymbol{\mu}_*$ is the highest weight) while $\hat{H}_j v_* = i\mu_{*j} v_*$. Therefore, W_{*s} is invariant under $\hat{H}_j$ and $\hat{E}_{\boldsymbol{\alpha}_j}$ for $s = 0$. Suppose this holds for s. Elements of $W_{*(s+1)}$ not in W_{*s} are of the form $w = \hat{E}_{-\boldsymbol{\alpha}_k} v$ for some $v \in W_{s*}$. We have that
$$\hat{E}_{\boldsymbol{\alpha}_j} w = \hat{E}_{\boldsymbol{\alpha}_j} \hat{E}_{-\boldsymbol{\alpha}_k} v = \hat{E}_{-\boldsymbol{\alpha}_k} \hat{E}_{\boldsymbol{\alpha}_j} v + \left[\hat{E}_{\boldsymbol{\alpha}_j}, \hat{E}_{-\boldsymbol{\alpha}_k}\right] v.$$
In the first term on the right-hand side, we note that by the induction hypothesis, $\hat{E}_{\boldsymbol{\alpha}_j} v \in W_{*s}$, so that $\hat{E}_{-\boldsymbol{\alpha}_k} \hat{E}_{\boldsymbol{\alpha}_j} v \in W_{*(s+1)}$. In the second term, we note that
$$\left[\hat{E}_{\boldsymbol{\alpha}_j}, \hat{E}_{-\boldsymbol{\alpha}_k}\right] = -i\delta_{jk} \boldsymbol{\alpha}_j \cdot \hat{\mathbf{H}} = -i\delta_{jk} \sum_{m=1}^{r} \alpha_{jm} \hat{H}_m.$$
(cf Proposition 10.2). By the induction hypothesis, $\hat{H}_m v \in W_{*s}$, so that the second term belongs to W_{s*}, and hence to $W_{*(s+1)}$. The argument that $\hat{H}_j w$ belongs to $W_{*(s+1)}$ is similar; we have that
$$\hat{H}_j w = \hat{H}_j \hat{E}_{-\boldsymbol{\alpha}_k} v = \hat{E}_{-\boldsymbol{\alpha}_k} \hat{H}_j v + \left[\hat{H}_j, \hat{E}_{-\boldsymbol{\alpha}_k}\right] v.$$
In the first term on the right-hand side, the induction hypothesis gives $\hat{H}_j v \in W_{*s}$, so that $\hat{E}_{-\boldsymbol{\alpha}_k} \hat{H}_j v \in W_{*(s+1)}$. In the second term, we note that
$$\left[\hat{H}_j, \hat{E}_{-\boldsymbol{\alpha}_k}\right] = -i\alpha_{jk} \hat{E}_{-\boldsymbol{\alpha}_k}.$$
We have that $\hat{E}_{-\boldsymbol{\alpha}_k} v \in W_{*(s+1)}$, so that the second term belongs to $W_{*(s+1)}$, too.

Thus, W_* is invariant under $\hat{\Gamma}$. If $\hat{\Gamma}$ is irreducible, it follows that $V = W_*$. In this case, it is straightforward to show the highest weight $\boldsymbol{\mu}_*$ is nondegenerate (if the highest weight were degenerate, W would be reducible), and that $\hat{E}_{\boldsymbol{\alpha}_j}$ can be explicitly determined from $\boldsymbol{\mu}_*$ (along with knowledge of the simple roots) using Equations (56) and (56).

The highest weight $\boldsymbol{\mu}_*$ may be characterised as follows. Since $\hat{E}_{\boldsymbol{\alpha}_j} v_* = 0$, it follows from (56) that
$$\boldsymbol{\alpha}_j \cdot \boldsymbol{\mu}_* = \frac{\alpha_j^2}{2} q_j.$$
where q_j is a nonnegative integer. Since the simple roots form a basis, $\boldsymbol{\mu}_*$ is uniquely determined by the inner products $\boldsymbol{\alpha}_j \cdot \boldsymbol{\mu}_*$, and hence by $(q_1, \dots, q_r)$. Thus, the irreducible representations of $\mathfrak{g}$ are in 1-1 correspondence with r-tuples of nonnegative integers, and may be constructed explicitly.

10.4 Dynkin diagrams

The set of simple roots of a compact Lie algebra can be described by a graph, or really a multigraph, called a Dynkin diagram (a multigraph is a graph with multiple edges between pairs of nodes). The nodes of a Dynkin diagram correspond to the simple roots of the Lie algebra. The edges between pairs of nodes are determined by the angles between the corresponding simple roots. Recall that the angle between two simple roots can be $90^\circ = \pi/2$, $120^\circ = 2\pi/3$, $135^\circ = 3\pi/4$ or $150^\circ = 5\pi/6$. If the angle between two simple roots is 90°, the corresponding nodes are not connected in the Dynkin diagram. If the angle is $120^\circ = 2\pi/3$, $135^\circ = 3\pi/4$ or $150^\circ = 5\pi/6$, the nodes are connected by one, two or three edges respectively.

In addition to the angles between simple roots, we would like to encode information about their lengths in the Dynkin diagram. Information about ratios of lengths can be deduced, like information about angles, from Eq. (62). Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be simple roots. We have that

$$\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{\alpha^2} = -\frac{m}{2}, \qquad \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{\beta^2} = -\frac{n}{2},$$

where m, n are nonnegative integers such that $0 \le mn \le 3$. We have that

$$\cos\theta = -\frac{\sqrt{mn}}{2},$$

where θ is the angle between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and

$$\frac{\alpha^2}{\beta^2} = \frac{n}{m},$$

provided $mn \neq 0$ (if $mn = 0$, we cannot deduce anything about α^2/β^2).

The possible values of m and n along with the associated values of θ and α^2/β^2 are shown in the table below. We also indicate the link in the Dynkin diagram between the nodes corresponding to $\boldsymbol{\alpha}$ (left) and $\boldsymbol{\beta}$ (right). We follow the convention of colouring the node corresponding to the longer of the two simple roots.

mn	m	n	θ	α^2/β^2	Dynkin diagram $\boldsymbol{\alpha}\cdots\boldsymbol{\beta}$
0	0	0	$\pi/2$	–	$\circ \quad \circ$
1	1	1	$2\pi/3$	1	$\circ - \circ$
2	1	2	$3\pi/4$	2	$\bullet = \circ$
2	2	1	$3\pi/4$	1/2	$\circ = \bullet$
3	1	3	$5\pi/6$	3	$\bullet \equiv \circ$
3	3	1	$5\pi/6$	1/3	$\circ \equiv \bullet$

Table 2: Relations between pairs of simple roots.

We note that roots with an angle of 120° between them (corresponding to $m = n = 1$ above) have the same length. It will turn out that compact simple Lie algebras contain simple roots of at most two different lengths. Hence, two colours will suffice.

10.5 The classical Lie algebras

- su(n)

su(n) is the Lie algebra of traceless antihermitian $n\times n$ matrices. The real dimension of su(n) (its dimension as a real vector space) is n^2-1. One can take the Cartan subalgebra to consist of diagonal matrices. The rank of su(n) is equal to $n-1$. The eigenvectors of the adjoint representation of the Cartan subalgebra with nonzero eigenvalues may be taken to be off-diagonal matrices with a single nonzero element. Details are given in Problem 3.10. The Dynkin diagram is shown in Figure 10.5.

- so($2n$)

  so($2n$) is the Lie algebra of real antisymmetric matrices of even dimension, which we take to be $2n$. The real dimension of so($2n$) is $n(2n-1)$. It is convenient to partition elements of so($2n$) into $n \times n$ matrices of 2×2 blocks (this is equivalent to taking $\mathbb{R}^{2n}$ to be the tensor product $\mathbb{R}^n \otimes \mathbb{R}^2$). One can take the Cartan subalgebra to be spanned by block-diagonal matrices with a single nonzero 2×2 block equal to $i\sigma_2$. The rank of so($2n$) is equal to n. Details are given in Problem 3.11. The Dynkin diagram is shown in Figure 10.5.

- so($2n+1$)

  so($2n+1$) is the Lie algebra of real antisymmetric matrices of odd dimension, which we take to be $2n+1$. The real dimension of so($2n+1$) is $n(2n+1)$. It is convenient to identify so($2n+1$) with the subset of so($2n+2$) consisting of real antisymmetric matrices of dimension $(2n+2)\times(2n+2)$ whose last column and last row are both equal to zero. One can take the Cartan subalgebra to be spanned by block-diagonal matrices with a single nonzero 2×2 block equal to $i\sigma_2$, where

$$i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

  The rank of so($2n+1$) is equal to n. Details are given in Problem 3.12. The Dynkin diagram is shown in Figure 10.5.

- usp($2n$)

  The unitary group is the group of linear transformations which preserve a hermitian inner product on a complex vector space. The orthogonal group is the group of linear transformations which preserve a real inner product on a real vector space. The complex symplectic group, $\mathrm{Sp}_{\mathbb{C}}(2n)$, is the group of linear transformations which preserve a real antisymmetric nondegenerate bilinear form on a complex vector space of even dimension $2n$. By way of explanation, a bilinear form $B : V \times V \to \mathbb{C}$ is real if $B(u^*, v^*) = (B(u,v))^*$, antisymmetric if $B(u,v) = -B(v,u)$, and nondegenerate if $B(u,v) = 0$ for all v implies that $u = 0$. Equivalently, $\mathrm{Sp}_{\mathbb{C}}(2n)$ consists of complex $2n \times 2n$ matrices S for which

$$S^T J S = J,$$

  where J is the $2n \times 2n$ matrix, which, when decomposed into 2×2 blocks, is block diagonal with $i\sigma_2$ as its diagonal blocks. As a matrix Lie group, $\mathrm{Sp}_{\mathbb{C}}(2n)$ has dimension $2n(2n+1)$.

  Two important subgroups of $\mathrm{Sp}_{\mathbb{C}}(2n)$ are i) the real symplectic group, $\mathrm{Sp}_{\mathbb{R}}(2n)$ and ii) the unitary symplectic group USp($2n$). The real symplectic group is the subgroup of linear transformations in $\mathrm{Sp}_{\mathbb{C}}(2n)$ which preserves the complex structure on the underlying complex vector space V; that is, if $v \in V$ and v^* denotes its complex conjugate, then $S \in \mathrm{Sp}_{\mathbb{C}}(2n)$ is real if $(Sv)^* = Sv^*$. Equivalently, in terms of matrices, the real symplectic group is just the subgroup of real matrices in $\mathrm{Sp}_{\mathbb{C}}(2n)$. The real symplectic group is basic to symplectic geometry, which underlies the theory of Hamiltonian systems. The dimension of $\mathrm{Sp}_R(2n)$ is $n(2n+1)$.

  The unitary symplectic group USp($2n$) is the group of linear transformations on a complex vector space V which preserve both a hermitian inner product and a real antisymmetric bilinear form. Equivalently, USp($2n$) consists of $2n \times 2n$ matrices U which satisfy

$$U^\dagger U = I_{2n}, \qquad U^T J U = J.$$

  Its dimension is $n(2n+1)$. USp($2n$) can also be regarded as $n \times n$ matrices over the quarternions $\mathbb{H}$ which preserve the quarternionic inner product on $\mathbb{H}^n$. As it is a closed subgroup of U($2n$), USp($2n$) is compact.

  The Lie algebra of USp($2n$), denoted usp($2n$), may be regarded as the space of ($2n \times 2n$)-dimensional traceless antihermitian matrices s which satisfy

$$s^T J + Js = 0.$$

  Its dimension is $n(2n+1)$. Equivalently, usp($2n$) may be regarded as the space of $n \times n$ matrices Q over the quaternions which satisfy $Q_{ab} = \overline{Q_{ab}}$, where $\overline{q}$ denotes the quaternion conjugate; if $q = aI_2 + i\mathbf{b} \cdot \boldsymbol{\sigma}$, then $\overline{q} = aI_2 - i\mathbf{b} \cdot \boldsymbol{\sigma}$. The rank of usp($2n$) is n, and its Dynkin diagram is shown in Figure 10.5. Details are given in Problem 3.13.

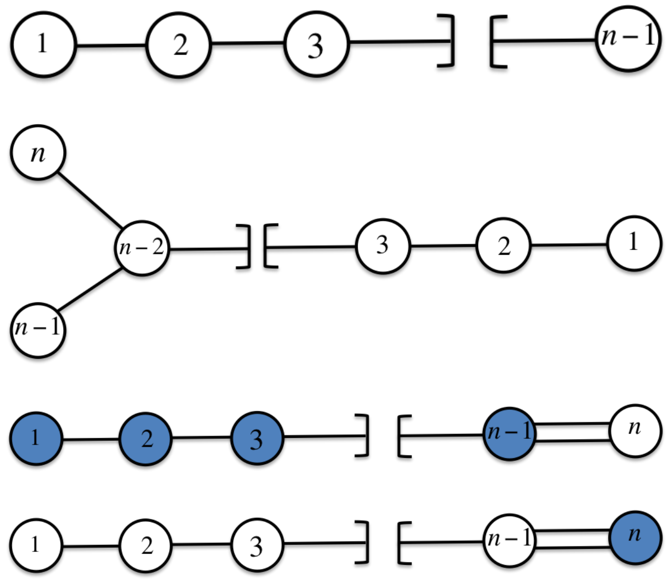


Figure 9: (a) Dynkin diagram for su(n) (b) Dynkin diagram for so($2n$) (c) Dynkin diagram for so($2n+1$) (d) Dynkin diagram for usp($2n$).

11 *Classification of compact simple Lie algebras

We begin with a simple observation, another consequence of Equation (60)

Proposition 11.1. Let $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ be simple roots of a compact simple Lie algebra. If $\boldsymbol{\beta}\cdot\boldsymbol{\gamma}=0$, then

$$[e_{\boldsymbol{\beta}}, e_{\boldsymbol{\gamma}}] = [e^\dagger_{\boldsymbol{\beta}}, e_{\boldsymbol{\gamma}}] = [e_{\boldsymbol{\beta}}, e^\dagger_{\boldsymbol{\gamma}}] = [e^\dagger_{\boldsymbol{\beta}}, e^\dagger_{\boldsymbol{\gamma}}] = 0.$$

Proof. The fact that

$$[e^\dagger_{\boldsymbol{\beta}}, e_{\boldsymbol{\gamma}}] = 0$$

is essentially Proposition 10.11, with $e_{-\boldsymbol{\beta}}$ replaced by $e^\dagger_{\boldsymbol{\beta}}$; this is really a consequence of the roots being simple. From Eq. (60), we have the general relation

$$\frac{\boldsymbol{\beta}\cdot\boldsymbol{\gamma}}{\beta^2} = \frac{q-p}{2},$$

where p is the smallest positive integer such that $\hat{e}^{p+1}_{\boldsymbol{\beta}} e_{\boldsymbol{\gamma}} = 0$ and q is the smallest positive integer such that $\left(\hat{e}^\dagger_{\boldsymbol{\beta}}\right)^{q+1} e_{\boldsymbol{\gamma}} = 0$. Setting $q=0$ (as the roots $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are simple) and $\boldsymbol{\beta}\cdot\boldsymbol{\gamma}=0$ (by assumption), we conclude that $p=0$, i.e.

$$[e_{\boldsymbol{\beta}}, e_{\boldsymbol{\gamma}}] = 0.$$

The remaining relations follow by taking hermitian conjugates of the preceding two. □

Definition 11.2. A set of vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r \in \mathbb{R}^n$ is **decomposable** if it can be decomposed into two nonempty subsets of mutually orthogonal vectors. If such a decomposition does not exist, the set is **indecomposable**.

Proposition 11.3. The set of simple roots of a compact simple Lie algebra is indecomposable.

Proof. Suppose the simple roots of $\mathfrak{g}$ are decomposable. We will show that $\mathfrak{g}$ contains a nontrivial ideal, and therefore is not simple.

Let $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_s, \boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_t$ denote the simple roots of $\mathfrak{g}$, and suppose that $\boldsymbol{\beta}_j \cdot \boldsymbol{\gamma}_k = 0$ for all $1 \le j \le s$ and $1 \le k \le t$. Let $I_{\mathbb{C}}$ denote the subalgebra of $\mathfrak{g}_{\mathbb{C}}$ generated by the $e_{\boldsymbol{\beta}_j}$'s and the $e^\dagger_{\boldsymbol{\beta}_j}$'s. That is, $I_{\mathbb{C}}$ is the smallest set containing complex linear combinations of the $e_{\boldsymbol{\beta}_j}$'s and the $e^\dagger_{\boldsymbol{\beta}_j}$'s that is closed under the Lie bracket; equivalently, it is the span of repeated Lie brackets of the $e_{\boldsymbol{\beta}_j}$'s and the $e^\dagger_{\boldsymbol{\beta}_j}$'s).

We claim that $I_{\mathbb{C}}$ is an ideal. That is, for all $a \in \mathfrak{g}_{\mathbb{C}}$ and for all $b \in I_{\mathbb{C}}$,

$$[a, b] \in I_{\mathbb{C}}. \tag{67}$$

It suffices to show this for b equal to $e_{\boldsymbol{\beta}_l}$ or $e^\dagger_{\boldsymbol{\beta}_l}$, and for a to be one of $e_{\boldsymbol{\beta}_j}$, $e^\dagger_{\boldsymbol{\beta}_j}$, $e_{\boldsymbol{\gamma}_k}$, or $e^\dagger_{\boldsymbol{\gamma}_k}$. This follows from the fact that $I_{\mathbb{C}}$ is generated by the $e_{\boldsymbol{\beta}_j}$'s and $e^\dagger_{\boldsymbol{\beta}_j}$'s, while $\mathfrak{g}_{\mathbb{C}}$ is generated by the $e_{\boldsymbol{\beta}_k}$'s, $e^\dagger_{\boldsymbol{\beta}_k}$'s, $e_{\boldsymbol{\gamma}_k}$'s and $e^\dagger_{\boldsymbol{\gamma}_k}$. From the Jacobi identity, it follows that if (67) holds for the generators of $\mathfrak{g}_{\mathbb{C}}$ and $I_{\mathbb{C}}$, then it holds for all $a \in \mathfrak{g}_{\mathbb{C}}$ and $b \in I_{\mathbb{C}}$.

By definition, $[e_{\boldsymbol{\beta}_j}, e_{\boldsymbol{\beta}_l}]$, $[e^\dagger_{\boldsymbol{\beta}_j}, e_{\boldsymbol{\beta}_l}]$, $[e_{\boldsymbol{\beta}_j}, e^\dagger_{\boldsymbol{\beta}_l}]$ and $[e^\dagger_{\boldsymbol{\beta}_j}, e^\dagger_{\boldsymbol{\beta}_l}]$ all belong to I. From Proposition 11.1, the remaining brackets of the generators, in which $e_{\boldsymbol{\beta}_j}$ is replaced by $e_{\boldsymbol{\gamma}_k}$, all vanish.

Thus, $I_{\mathbb{C}}$ is a nontrivial ideal in $\mathfrak{g}_{\mathbb{C}}$. Let I denote the subalgebra of $\mathfrak{g}$ generated by brackets of the antihermitian elements $(e_{\boldsymbol{\alpha}_j} - e^\dagger_{\boldsymbol{\alpha}_j})$ and $i(e_{\boldsymbol{\alpha}_j} + e_{\boldsymbol{\alpha}_j})^\dagger$. The preceding argument shows that I is a nontrivial ideal in $\mathfrak{g}$. □

Definition 11.4. A set of vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r \in \mathbb{R}^n$ is a **Π-system** if it satisfies the following:

i) the set of $\boldsymbol{\alpha}_j$'s is indecomposable

ii) the $\boldsymbol{\alpha}_j$'s are linearly independent

iii) $2\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_k = -m\alpha_j^2, \quad m \in \{0, 1, 2, 3\}$

It is clear that any indecomposable subset of a Π-system is itself a Π-system. We will use this fact often below. A typical argument is to establish that a class of sets of vectors cannot be a Π-system by showing that a certain small, indecomposable subset of the sets in this class cannot be a Π-system.

Proposition 11.5. The simple roots of a compact simple Lie algebra $\mathfrak{g}$ is a Π-system.

Proof. Property i) follows from Proposition 10.13; ii) was noted in Section 10.4; iii) follows from Proposition 11.3. □

We may represent Π-systems by Dynkin diagrams, as for the simple roots of the classical Lie algebras in Figures 10.5–10.5.

Proposition 11.6. The only 3-element Π-systems are as follows:

$$\circ - \circ - \circ \qquad \circ - \circ = \bullet \qquad \bullet - \bullet = \circ$$

Proof. Let θ_{12}, θ_{23} and θ_{31} denote the angles between the vectors Let $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3 \in \mathbb{R}^n$ denote the vectors in the Π-system, and let

$$\hat{\boldsymbol{\alpha}} = \frac{\boldsymbol{\alpha}_j}{\alpha_j}$$

denote the associated unit vectors. Let

$$\theta_{jk} = \cos^{-1}(\hat{\boldsymbol{\alpha}}_j \cdot \hat{\boldsymbol{\alpha}}_k)$$

denote the angle between $\hat{\boldsymbol{\alpha}}_j$ and $\hat{\boldsymbol{\alpha}}_k$, with $0 < \theta_{jk} < \pi$.

A simple geometric argument shows that $\theta_{12} + \theta_{23} + \theta_{31} < 2\pi$. Indeed, we can choose an orthonormal basis for $\mathbb{R}^n$ so that only the first three components of the $\hat{\boldsymbol{\alpha}}_j$'s are nonvanishing. Then the $\hat{\boldsymbol{\alpha}}_j$'s may be regarded as points on the two-sphere comprising the vertices a spherical triangle, whose sides are arcs of great circles with arc lengths given by the θ_{jk}'s. The sum of the arc lengths is maximised if the vertices are coplanar, in which case the sum is 2π. But if the vertices were coplanar, the $\hat{\boldsymbol{\alpha}}_j$'s would be linearly dependent, which would contradict the fact that they comprise a Π-system. Therefore,

$$\theta_{12} + \theta_{23} + \theta_{31} < 360^\circ.$$

This argument is elaborated in Figure 10

The allowed values of θ_{jk} (in degrees) are 90°, 120°, 135° and 150°. We cannot have two of the θ_{jk}'s being orthogonal, for then the corresponding Π-system would be decomposable, with one vector being

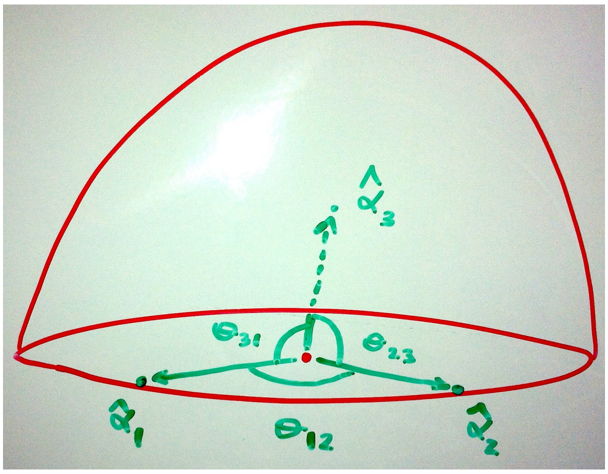


Figure 10: Pictorial proof of the fact that the sum of the angles between three linearly independent unit vectors must be less than 360°. Without loss of generality, we may take two of the vectors to lie on the equator of the sphere, and the third to lie in the northern hemisphere. If the third vector is rotated toward the equator, the sum of the three angles increases (this can be shown analytically, too). When the third vector lies on the equator, the sum of the three angles is 360°.

orthogonal to the other two. Therefore, the allowed values of the θ_{jk}'s are $(90°, 120°, 120°)$, whose sum is 330°; $(90°, 120°, 135°)$, whose sum is 345°. But these values correspond precisely to the three Dynkin diagrams shown above.

□

Corollary 11.7. The only Π-system with a triple edge contains two elements, and has Dynkin diagram

$$\circ \equiv \bullet \tag{68}$$

Proof. The pair of vectors described by the Dynkin diagram above is clearly a Π-system, and no Π-system with more than two elements can have a Dynkin diagram with a triple edge. □

The simple Lie algebra associated with (68) is denoted G_2. It is one of the five so-called exceptional Lie algebras, the first we have encountered. The subscript 2 refers to the fact that G_2 has rank 2. Its structure is obtained in Problem Sheet 3.14.

Proposition 11.8 (Reduction I)**.** In the Dynkin diagram of a Π-system, a pair of nodes connected to each other by a single edge and with no common neighbours can be contracted to a single node. See Figure 11 below.

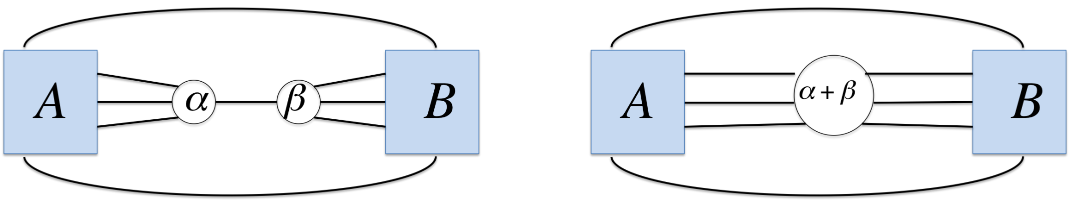


Figure 11: Original Dynkin diagram. Figure 12: Reduced Dynkin diagram

A pair of nodes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, connected by an edge replaced by a single node, $\boldsymbol{\alpha}+\boldsymbol{\beta}$.

Proof. The Dynkin diagram in Figure 11 describes a Π-system of the form $S = A \cup \{\boldsymbol{\alpha}, \boldsymbol{\beta}\} \cup B$, where $\boldsymbol{\alpha}$ is orthogonal to the vectors in B and $\boldsymbol{\beta}$ is orthogonal to the vectors in A (note that A and B need not be disjoint). We claim that $S' = A \cup \{\boldsymbol{\alpha}+\boldsymbol{\beta}\} \cup B$ is also a Π-system. Note that the Dynkin diagram of S' is given in Figure 12.

It is clear that if S is indecomposable, then so is S', and that if the elements of S are linearly independent, then so are those of S'.

It remains to verify property iii) in Definition 11.4 for inner products involving $\boldsymbol{\alpha}+\boldsymbol{\beta}$ (inner products which don't involve $\boldsymbol{\alpha}+\boldsymbol{\beta}$ already satisfy property iii), since S is a Π-system). We observe that

$$||\boldsymbol{\alpha}+\boldsymbol{\beta}|| = \alpha = \beta.$$

Indeed, the fact that $\alpha = \beta$ follows from the fact that $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = -\alpha^2 = -\beta^2$. This in turn implies that

$$(\boldsymbol{\alpha}+\boldsymbol{\beta})\cdot(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \alpha^2 + 2\boldsymbol{\alpha}\cdot\boldsymbol{\beta} + \beta^2 = \alpha^2 = \beta^2.$$

Therefore, if $\mathbf{u} \in A$ and $2\mathbf{u}\cdot\boldsymbol{\alpha} = -m\alpha^2 = -nu^2$ for some m, n with $0 \le m, n \le 3$, it follows that

$$2\mathbf{u}\cdot(\boldsymbol{\alpha}+\boldsymbol{\beta}) = 2\mathbf{u}\cdot\boldsymbol{\alpha} = -m||\boldsymbol{\alpha}+\boldsymbol{\beta}||^2 = -nu^2,$$

since $\mathbf{u}\cdot\boldsymbol{\beta} = 0$. Similarly, if $\mathbf{v} \in B$ and $2\mathbf{v}\cdot\boldsymbol{\beta} = -m\beta^2 = -nv^2$ for some m, n with $0 \le m, n \le 3$, it follows that

$$2\mathbf{v}\cdot(\boldsymbol{\alpha}+\boldsymbol{\beta}) = 2\mathbf{v}\cdot\boldsymbol{\beta} = -m||\boldsymbol{\alpha}+\boldsymbol{\beta}||^2 = -nv^2,$$

since $\mathbf{v}\cdot\boldsymbol{\alpha} = 0$.

□

Proposition 11.8 has the following consequences:

Proposition 11.9. The Dynkin diagram of a Π-system

a) can contain at most one double edge, and

b) cannot contain any cycles.

Proof.
We argue by contradiction.

a) Suppose the Dynkin diagram of a Π-system S contains two or more pairs of nodes with the nodes in each pair connected by a double edge. Consider two such pairs of nodes. Since S is indecomposable, the Dynkin diagram is connected, and there exists a path joining one pair of double-bonded nodes to another such pair. Without loss of generality, we may assume that this path consists of a sequence of nodes connected by single edges (starting from one double-bonded pair of nodes, we terminate the path as soon as we reach another pair of double-bonded nodes). Let S' denote the subset whose nodes belong to this path. S' is clearly indecomposable, so it constitutes a Π-system in its own right. By Proposition 11.8, the single bonds from S' can be contracted to produce a Π-system of 3 nodes containing two double bonds. But this contradicts the classification of 3-node Π-systems given in Proposition 11.6. This is illustrated by the following symbolic reduction, in which we do not need to distinguish nodes of different length by colourings.

$$\bigcirc = \bigcirc - \bigcirc - \cdots - \bigcirc = \bigcirc \qquad \Longrightarrow \qquad \bigcirc = \bigcirc = \bigcirc.$$

b) Suppose the Dynkin diagram of a Π-system S contains a cycle C. By the previous result, there is at most one double edge in this cycle. By contracting single edges, we can, according to Proposition 11.8, produce a Π-system which is a cycle of just 3 nodes, with either all single edges or a single double edge. But this contradicts the classification of 3-node Π-systems given in Proposition 11.6. See Figure 13.

□

Proposition 11.10 (Reduction II)**.** In the Dynkin diagram of a Π-system, a pair of nodes connected by single edges to a third node and with no other neighbours can be replaced by a single node connected by a double edge to that third node. See Figure 11.

Proof. The Dynkin diagram in Figure 14 describes a Π-system of the form $S = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\} \cup C$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are orthogonal to the vectors in C. We claim that $S' = \{\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\gamma}\} \cup C$ is also a Π-system. Note that the Dynkin diagram of S' is given in Figure 15.

It is clear that if S is indecomposable, then so is S', and that if the elements of S are linearly independent, then so are those of S'.

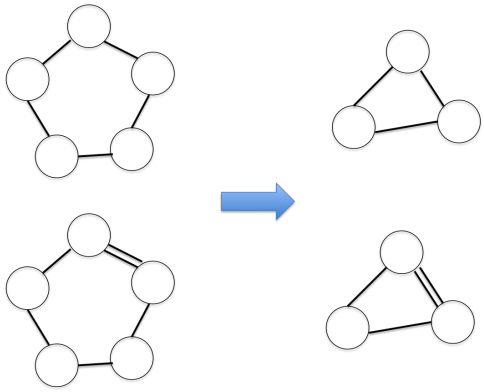
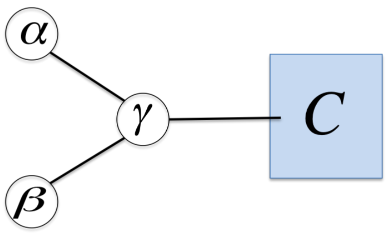


Figure 13: An n-cycle can be contracted to a 3-cycle, which cannot be the Dynkin diagram of a Π-system.



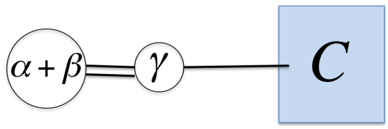


Figure 15: Reduced Dynkin diagram

Figure 14: Original Dynkin diagram.

A pair of nodes connected by single edges to a third node replaced by a single node connected to the third node by a double edge.

It remains to verify property iii) in Definition 11.4 for inner products involving $\boldsymbol{\alpha}+\boldsymbol{\beta}$ (inner products which don't involve $\boldsymbol{\alpha}+\boldsymbol{\beta}$ already satisfy property iii), since S is a Π-system). We observe that $\alpha = \gamma = \beta$, as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are connected to $\boldsymbol{\gamma}$ by a single edge. Therefore

$$||\boldsymbol{\alpha}+\boldsymbol{\beta}||^2 = 2\alpha^2 = 2\beta^2,$$

since $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 0$. We have that $2(\boldsymbol{\alpha}+\boldsymbol{\beta})\cdot\boldsymbol{\gamma} = -\alpha^2 - \beta^2 = -||\boldsymbol{\alpha}+\boldsymbol{\beta}||^2 = -2\gamma^2$, which shows that property iii) is satisfied. If $\mathbf{u} \in C$, then $\mathbf{u}\cdot(\boldsymbol{\alpha}+\boldsymbol{\beta}) = 0$, which satisfies also property iii). $\square$

The *degree* of a node in a graph is the number of its neighbours. Proposition 11.10 has the following consequences for the degrees of nodes in a Dynkin diagram:

Proposition 11.11.

In the Dynkin diagram of a Π-system,

a) every node has degree less than 4,

b) there cannot be a node of degree 3 and a double edge,

c) there is at most one node of degree 3.

Proof. We argue by contradiction.

a) Suppose the Dynkin diagram D of a Π-system S contains a node of degree 4 or more, and let $\boldsymbol{\alpha}$ denote the vector in S associated to that node. Let $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d$, where $d \geq 4$, denote the vectors of S whose inner product with $\boldsymbol{\alpha}$ is nonzero (thus, d is the degree of the node corresponding to $\boldsymbol{\alpha}$). Then $S' = \{\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d\}$ is a Π-system. Consider its Dynkin diagram D'. By Proposition 11.8, D' contains either all single edges or one double edge with the rest single edges. Suppose the former. Then there are at least 4 single edges, which may be taken as two pairs of edges. By Proposition 11.10, these two pairs of single edges can be reduced to two double edges and still constitute the Dynkin diagram of a Π-system. But this contradicts the fact (Proposition 11.8) that such a Dynkin diagram cannot contain two double edges.

For the other case, suppose D' contains a double edge. Then it contains at least 3 single edges. Two of these can be reduced to a single double edge. As with the preceding case, this leads to a contradiction: the Dynkin diagram of a Π-system cannot contain two double edges.

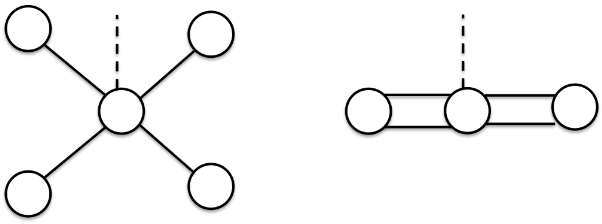


Figure 16:

b) Suppose the Dynkin diagram D of a Π-system S contains a node $\boldsymbol{\alpha}$ of degree 3 as well as a double edge between nodes $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Since the Dynkin diagram is connected (a Π-system is indecomposable), there is a path from $\boldsymbol{\alpha}$ to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The subset S' containing this path as well as all vectors in S which are not orthogonal to $\boldsymbol{\alpha}$ is itself a Π-system. Let D' denote its Dynkin diagram. By assumption, the node corresponding to $\boldsymbol{\alpha}$ in D' has degree at least 3. Therefore, there are at least two edges not contained in the path to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. By Proposition 11.10, these can be reduced to a double edge. The resulting Dynkin diagram contains two double edges, which contradicts Proposition 11.8.

c) The argument that there can be at most one node of degree 3 in the Dynkin diagram of a Π-system is similar to the preceding two arguments, and is indicated in Figure 17. By reducing pairs of nodes connected by a single edge as in Proposition 11.8, a Dynkin diagram with two nodes of degree 3 can be reduced to a Dynkin diagram with a node of degree 4.

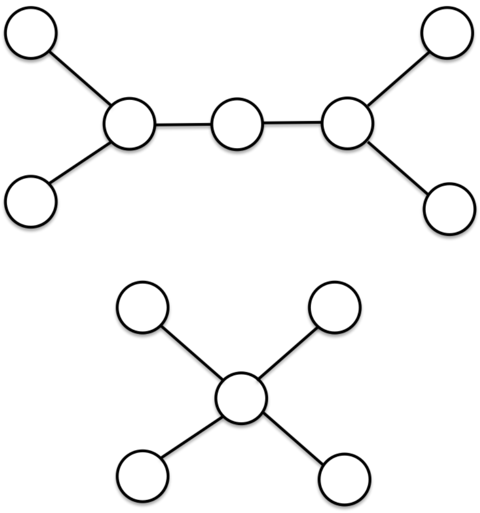


Figure 17:

□

Propositions 11.9 and 11.11 drastically constrain the Dynkin diagrams that correspond to Π-systems. Indeed, apart from G_2, there are just three cases, namely III. Dynkin diagrams which contain a degree-3 node, II. Dynkin diagrams which contain a double edge, and I. Dynkin diagrams which contain neither. This leaves the following possibilities for Dynkin diagrams of Π-systems, apart from G_2.

I. No double edge or degree-three node

In this case, the Dynkin diagram consists of a single chain of n nodes connected by single edges – this is just the Dynkin diagram of su(n), as shown in Figure 10.5.

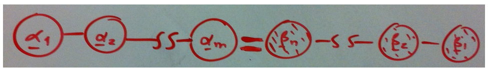


Figure 18:

II. Double edge

In this case, the Dynkin diagram $F(m,n)$ consists of a two chains of nodes, say of length m and n, connected by a double edge. The number of nodes is $m+n$. Let $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ denote the vectors associated with one chain, and $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ denote the vectors associated with the other. Without loss of generality, we may assume that the $\boldsymbol{\beta}_k$'s are longer than the $\boldsymbol{\alpha}_j$'s, by a factor of $\sqrt{2}$. We then have the following inner products:

$$\boldsymbol{\alpha}_j\cdot\boldsymbol{\alpha}_{j'}=\begin{cases}L^2, & j=j',\\ -\frac{1}{2}L^2, & |j-j'|=1,\\ 0, & \text{otherwise,}\end{cases}\qquad \boldsymbol{\beta}_k\cdot\boldsymbol{\beta}_{k'}=\begin{cases}2L^2, & k=k',\\ -L^2, & |k-k'|=1,\\ 0, & \text{otherwise,}\end{cases}\qquad \boldsymbol{\alpha}_j\cdot\boldsymbol{\beta}_k=\begin{cases}-L^2, & j=m, k{=}n\\ 0, & \text{otherwise.}\end{cases} \tag{69}$$

Proposition 11.12. The Dynkin diagram $F(m,n)$ corresponds to a Π-system if and only if either m or n =1, or else $m=n=2$.

Proof. From (69), the set

$$S=\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n\}$$

satisfies properties i) and iii) of Definition 11.4. We will show that property ii) is satisfied, i.e. that the $\boldsymbol{\alpha}_j$'s and $\boldsymbol{\beta}_k$'s are linearly independent, if and only if either m or n =1, or else $m=n=2$.

Suppose

$$\sum_{j=1}^{m}c_j\boldsymbol{\alpha}_j+\sum_{k=1}^{n}d_k\boldsymbol{\beta}_k=0 \tag{70}$$

for $c_j,d_k\in\mathbb{R}$. We take the inner product of the preceding relations with the vectors $\boldsymbol{\alpha}_1$ through $\boldsymbol{\alpha}_{m-1}$ to obtain

$$-c_{j-1}+2c_j-c_{j+1}=0,\quad 1\le j<m,$$

where we take $c_0:=0$ and we have multiplied through by $2/L^2$. This two-term recurrence relation has as its solution

$$c_j=Aj$$

for some undetermined constant A. Similarly, we take the inner product of the preceding relations with the vectors $\boldsymbol{\beta}_1$ through $\boldsymbol{\beta}_{n-1}$ to obtain the

$$-d_{k-1}+2d_k-d_{k+1}=0,\quad 1\le k<n,$$

where we take $d_0:=0$. This two-term recurrence relation has as its solution

$$d_k=Bk$$

for some undetermined constant B.

Finally, we take the inner product of (70) with $\boldsymbol{\alpha}_m$ to obtain

$$-c_{m-1}+2c_m-2d_n=0,$$

and with $\boldsymbol{\beta}_n$ to obtain

$$-d_{n-1}+2d_n-c_m=0.$$

Substituting $c_j=Aj$ and $d_k=Bk$ as above, we get a pair of homogeneous equations for A and B, which may be written in matrix form as

$$\begin{pmatrix} m+1 & -2n\\ -m & n+1\end{pmatrix}\begin{pmatrix} A\\ B\end{pmatrix}=0.$$

This system has a nontrivial solution if and only if the determinant of the 2×2 matrix in the preceding vanishes, i.e.

$$mn - m - n - 1 = 0. \tag{71}$$

If $m = 1$ or $n = 1$, (71) has no solutions. In this case, the vectors $\boldsymbol{\alpha}_j$'s and $\boldsymbol{\beta}_k$'s are linearly independent, and S is a Π-system. $F(m, 1)$ corresponds to the Dynkin diagram of $\mathrm{so}(2m + 1)$, and $F(1, n)$ corresponds to the Dynkin diagram of $\mathrm{usp}(2n + 1)$.

If $m = 2$, (71) becomes

$$n - 3 = 0.$$

Therefore, for $n = 1$ or $n = 2$, the vectors $\boldsymbol{\alpha}_j$'s and $\boldsymbol{\beta}_k$'s are linearly independent, while for $n = 3$, they are not.

For $n > 3$, $F(n, 2)$ cannot be a Π-system, since $F(n, 2)$ contains $F(3, n)$ as an indecomposable subset, and $F(3, n)$ is not a Π-system. Likewise, $F(m, n)$ for $n \geq m > 2$ cannot be a Π-system, since it contains $F(3, n)$ as an indecomposable subset. $\square$

The case $n = 1$ corresponds to $F(2, 1)$, or $\mathrm{usp}(2)$. The case $n = 2$ corresponds to $F(2, 2)$, which has Dynkin diagram

$$\bigcirc - \bigcirc = \bullet - \bullet.$$

This is usually denoted as F_4. It is one of the five exceptional Lie algebras, the second we have encountered. The subscript 4 indicates that the Lie algebra is four simple roots, or has rank 4.

III. Degree-three node

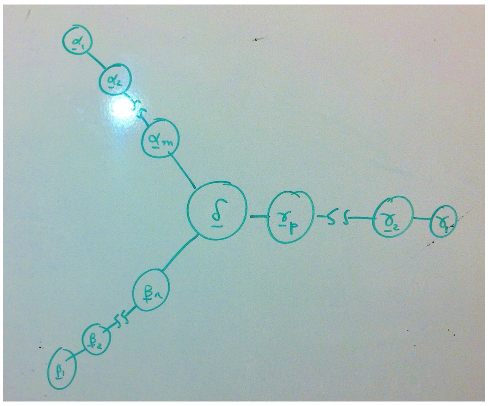


Figure 19:

In this case, the Dynkin diagram $E(m, n, p)$ consists of a three chains of nodes lengths m, n, p connected to a central node. All edges are single edges. The number of nodes is $m + n + p + 1$. Without loss of generality, we may assume that

$$m \leq n \leq p.$$

Let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ and $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ denote the vectors in the three chains, and let $\boldsymbol{\delta}$ denote the vector associated with the central node.

We then have the following inner products:

$$
\begin{aligned}
\boldsymbol{\alpha}_j \cdot \boldsymbol{\alpha}_{j'} &= \begin{cases} L^2, & j = j', \\ -\frac{1}{2}L^2, & |j-j'| = 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \boldsymbol{\alpha}_j \cdot \boldsymbol{\delta} = \begin{cases} -\frac{1}{2}L^2, & j = m, \\ 0, & \text{otherwise}, \end{cases} \\
\boldsymbol{\beta}_k \cdot \boldsymbol{\beta}_{k'} &= \begin{cases} L^2, & k = k', \\ -\frac{1}{2}L^2, & |k-k'| = 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \boldsymbol{\beta}_k \cdot \boldsymbol{\delta} = \begin{cases} -\frac{1}{2}L^2, & k = n, \\ 0, & \text{otherwise}, \end{cases} \\
\boldsymbol{\gamma}_l \cdot \boldsymbol{\gamma}_{l'} &= \begin{cases} L^2, & l = l', \\ -\frac{1}{2}L^2, & |l-l'| = 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \boldsymbol{\gamma}_l \cdot \boldsymbol{\delta} = \begin{cases} -\frac{1}{2}L^2, & l = p, \\ 0, & \text{otherwise}, \end{cases} \\
&\boldsymbol{\alpha}_j \cdot \boldsymbol{\beta}_k = \boldsymbol{\beta}_k \cdot \boldsymbol{\gamma}_l = \boldsymbol{\gamma}_l \cdot \boldsymbol{\alpha}_j = 0, \\
&\boldsymbol{\delta} \cdot \boldsymbol{\delta} = L^2.
\end{aligned} \tag{72}
$$

Proposition 11.13. The Dynkin diagram $E(m, n, p)$ corresponds to a Π-system if and only if either $m = p = 1$ and $n \geq 1$, or else $m = 1$, $n = 2$ and $p = 2$, 3 or 4.

Proof. From (72), the set

$$S = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p\}$$

satisfies properties i) and iii) of Definition 11.4. We will show that property ii) is satisfied, i.e. that the $\boldsymbol{\alpha}_j$'s, $\boldsymbol{\beta}_k$'s and $\boldsymbol{\gamma}_l$'s are linearly independent, if and only if either m or $n = 1$, or else $m = 1$, $n = 2$ and $2 \leq p \leq 4$.

Suppose

$$\sum_{j=1}^{m} c_j \boldsymbol{\alpha}_j + \sum_{k=1}^{n} d_k \boldsymbol{\beta}_k + \sum_{l=1}^{p} e_l \boldsymbol{\gamma}_l + D\boldsymbol{\delta} = 0 \tag{73}$$

for $c_j, d_k, e_l, D \in \mathbb{R}$. We take the inner product of the preceding relations with the vectors $\boldsymbol{\alpha}_1$ through $\boldsymbol{\alpha}_{m-1}$ to obtain

$$-c_{j-1} + 2c_j - c_{j+1} = 0, \quad 1 \leq j \leq m-1,$$

where we take $c_0 := 0$ and we have multiplied through by $2/L^2$. This two-term recurrence relation has as its solution

$$c_j = Aj$$

for some undetermined constant A. Similarly, we take the inner product of the preceding relations with the vectors $\boldsymbol{\beta}_1$ through $\boldsymbol{\beta}_{n-1}$ to obtain

$$d_k = Bk$$

for some undetermined constant B, and the inner product with the vectors $\boldsymbol{\gamma}_1$ through $\boldsymbol{\gamma}_{p-1}$ to obtain

$$e_l = Cl$$

for some undetermined constant C.

Next, we take the inner product of (73) with $\boldsymbol{\alpha}_m$ to obtain

$$-c_{m-1} + 2c_m - D = 0.$$

Substituting $c_m = Am$, this yields

$$(m+1)A - D = 0.$$

Similarly, taking the inner product with $\boldsymbol{\beta}_n$ yields

$$(n+1)B - D = 0,$$

and with $\boldsymbol{\gamma}_n$ yields

$$(p+1)C - D = 0.$$

Finally, taking the inner product with $\boldsymbol{\delta}$ yields

$$-c_m - d_n - e_l + 2D = 0,$$

or

$$-Am - Bn - Cl + 2D = 0.$$

In this way, we obtain four homogeneous equations for A, B, C and D, which we may write in matrix form as

$$\begin{pmatrix} m+1 & 0 & 0 & -1 \\ 0 & n+1 & 0 & -1 \\ 0 & 0 & p+1 & -1 \\ m & n & p & -2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = 0.$$

This system has a nontrivial solution if and only if the determinant of the 4×4 matrix in the preceding vanishes. A straightforward calculation yields the condition

$$mnp - (m+n+p) - 2 = 0. \tag{74}$$

If $m = n = 1$, then (74) has no solutions for any p. In this case, the vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ and $\boldsymbol{\delta}$ are linearly independent, and S is a Π-system. If $m = 1$ and $n = 2$, (74) becomes

$$p - 5 = 0.$$

Therefore, for $2 \leq p \leq 4$, the vectors $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_l$'s, and $\boldsymbol{\delta}$ are linearly independent, while for $p = 5$, they are not. For $p > 5$, $F(1,2,p)$ cannot be a Π-system, since it contains $E(1,2,5)$ as an indecomposable subset, and $E(1,2,5)$ is not a Π-system.

For $m = n = 2$, (74) becomes $3p - 6 = 0$. $p = 1$ corresponds to $E(1,2,2)$, which we have considered already, while for $p = 2$, $E(2,2,2)$ is not a Π-system. It follows that $E(2,2,p)$ is not a Π-system for $p \geq 2$, since it contains $E(2,2,2)$ as an indecomposable subset.

From the preceding, it is clear that there can be no Π systems $E(m,n,p)$ with $2 \leq m \leq n \leq p$, since these contain $E(2,2,2)$ as an indecomposable subset.

□

$E(1,1,p)$ corresponds to the Dynkin diagram of $\mathrm{so}(2p)$. $E(1,2,2)$, $E(1,2,3)$ and $E(1,2,4)$ are usually denoted E_6, E_7 and E_8, and correspond to the remaining three exceptional Lie algebras, shown in Figure III.–III..

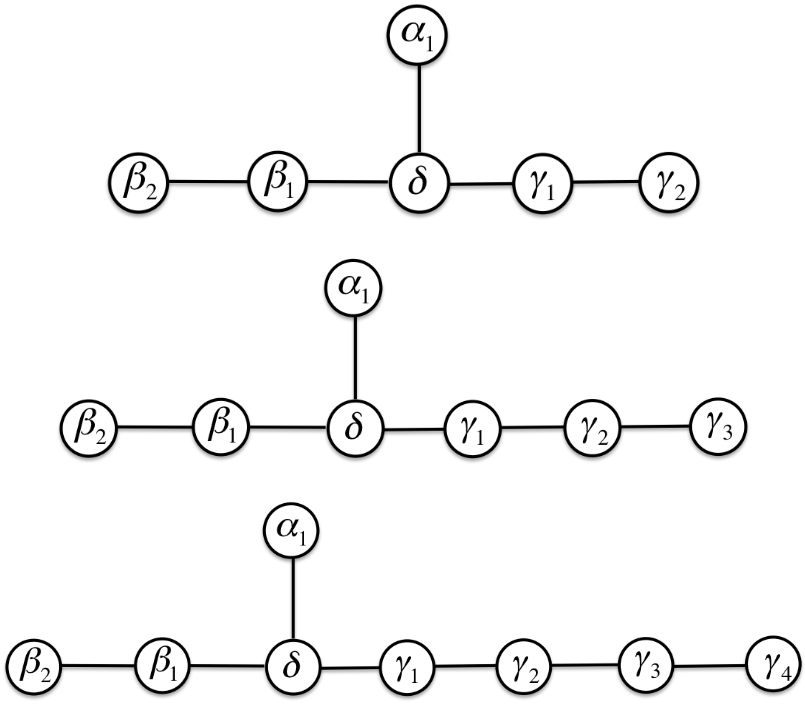


Figure 20: (a) Dynkin diagram for E_6 (b) Dynkin diagram for E_7 (c) Dynkin diagram for E_8